Non-Hermitian Photonics

Maxwell Operators, Open Resonances, and Topological Light

Wenhao Li and Yihao Yang*

*Corresponding author: yangyihao@zju.edu.cn

First Edition

A monograph on the spectral, scattering, modal, and topological theory of Maxwell equations when material response, radiation boundary conditions, gain, loss, dispersion, and openness destroy the ordinary self-adjoint structure of closed lossless electromagnetism. Starting from Maxwell's equations and classical optics, the book develops the physical origins of non-Hermiticity in photonic systems, the theory of quasinormal modes, exceptional points, $\mathcal{PT}$-symmetric optics, non-Hermitian band topology, the skin effect, and bound states in the continuum, with applications to lasers, absorbers, sensors, and topological devices.

Qiushi Writing | July 9, 2026

About This Draft and Lumin

This book is part of *Lumin*, Oxelra's open scientific monograph project for human researchers and scientific agents. Lumin organizes research topics at monograph scale and makes the reading path explicit: preview pages, PDF drafts, LaTeX sources, tables of contents, version records, and issue-based review are kept close to the manuscript so that reading, citation, checking, and revision can be traced.

The purpose of this release is therefore not to present an agent-assisted draft as a finished publication. It is to make the draft visible as a research object that can be read, questioned, corrected, and improved. Human readers are invited to examine the physical arguments, mathematical steps, figures, references, and reproducibility details; agents may also use the structured source to support checking, indexing, and review.

The initial draft of this book was written by the *Qiushi Engine*. Subsequent versions are expected to pass through human review, public issue tracking, and documented revision. Comments are especially welcome on conceptual gaps, derivational errors, notation, physical assumptions, figure clarity, numerical reproducibility, missing references, and the historical boundaries of terminology.

This draft should be cited and used with that status in mind: it is a working monograph under review, intended to make the development of an open scientific text legible rather than hidden.

Lumin project page: https://oxelra-yhy.github.io/Lumin/

Book review repository: https://github.com/Oxelra-yhy/Non-Hermitian

Contents

I Maxwell Equations Become Spectral Theory **1**

1 Light as an Eigenvalue Problem **2**

1.1 Maxwell's equations in matter . . . 2
1.2 The frequency-domain eigenproblem . . . 3
1.2.1 Harmonic fields and the time-frequency correspondence . . . 4
1.2.2 Deriving the curl–curl eigenproblem . . . 4
1.2.3 The magnetic-field formulation . . . 5
1.2.4 The divergence constraint . . . 5
1.3 Self-adjointness and its physical meaning . . . 6
1.3.1 The electromagnetic inner product . . . 6
1.3.2 Integration by parts and the boundary term . . . 6
1.3.3 The three conditions for Hermitian photonics . . . 7
1.4 The variational principle . . . 7
1.5 Two-dimensional reductions . . . 8
1.5.1 TM polarization . . . 8
1.5.2 TE polarization . . . 9
1.5.3 The Schrödinger analogy and its limits . . . 9
1.6 Example: one-dimensional photonic crystal . . . 10
1.6.1 Bloch modes and the transfer matrix . . . 10
1.6.2 Band gaps from the eigenvalue perspective . . . 10
1.6.3 What changes when ε_B becomes complex . . . 11
1.7 Spectrum types: discrete, continuous, and complex . . . 11
1.8 Positive semi-definiteness and the photonic density of states . . . 12
1.9 What this book will build . . . 14

2 Energy, Inner Products, and Hermiticity **15**

2.1 Poynting's theorem . . . 15
2.2 Time-averaged energy and the harmonic inner product . . . 17
2.3 Inner products for the Maxwell eigenproblem . . . 17

2.3.1 The **H**-field inner product . . . 18
2.3.2 The **E**-field inner product . . . 18
2.3.3 The first-order Maxwell inner product . . . 18
2.4 Self-adjointness: the complete argument . . . 19
2.5 Consequences of self-adjointness . . . 20
2.5.1 Real eigenvalues . . . 20
2.5.2 Orthogonality . . . 21
2.5.3 Completeness . . . 21
2.5.4 Variational characterization . . . 21
2.6 Lorentz reciprocity . . . 21
2.6.1 Reciprocity and the scattering matrix . . . 22
2.7 Four roads to non-Hermiticity . . . 22
2.7.1 Material loss and gain . . . 22
2.7.2 Radiation and open boundaries . . . 23
2.7.3 Frequency dispersion . . . 24
2.7.4 Projection onto a subspace . . . 25
2.8 The structure of what follows . . . 26

3 Periodicity and Photonic Bands 27
3.1 Periodic permittivity and the direct lattice . . . 27
3.2 Reciprocal lattice and Brillouin zone . . . 28
3.2.1 The dielectric scaling law . . . 29
3.3 Bloch's theorem for Maxwell fields . . . 29
3.4 The unit-cell eigenproblem . . . 30
3.5 One-dimensional photonic crystals . . . 30
3.5.1 Transfer matrix for a bilayer . . . 31
3.5.2 Bloch condition and the dispersion relation . . . 31
3.5.3 Connection to the scattering matrix . . . 32
3.6 Band structure in two dimensions . . . 33
3.6.1 TM bands . . . 33
3.6.2 TE bands . . . 33
3.6.3 The plane-wave expansion method . . . 33
3.7 Band gaps . . . 35
3.7.1 Field concentration and the physical origin of gaps . . . 35
3.8 Degeneracies and symmetry . . . 36
3.8.1 Dirac points . . . 36
3.8.2 Quadratic touchings and accidental degeneracies . . . 36
3.9 Orthogonality and normalization of Bloch modes . . . 37
3.10 Group velocity, effective mass, and density of states . . . 37
3.10.1 Group velocity . . . 37

3.10.2 Effective mass at band edges . . . 37
3.10.3 Density of states and van Hove singularities . . . 38
3.11 Beyond Hermitian bands: a preview . . . 39

4 Boundaries, Radiation, and Scattering 41
4.1 From cavities to open systems . . . 41
4.1.1 Energy balance in open systems . . . 42
4.2 Outgoing-wave boundary conditions . . . 42
4.2.1 The Silver–Müller radiation condition . . . 42
4.2.2 Why outgoing-wave conditions break self-adjointness . . . 42
4.3 The scattering matrix . . . 43
4.3.1 Definition and physical meaning . . . 43
4.3.2 Unitarity and energy conservation . . . 44
4.3.3 Reciprocity in scattering . . . 44
4.3.4 The optical theorem . . . 45
4.4 Poles, zeros, and resonances . . . 45
4.4.1 Poles: resonances and quasinormal modes . . . 45
4.4.2 Zeros: coherent perfect absorption . . . 46
4.4.3 Pole–zero structure and causality . . . 46
4.5 The Fabry–Pérot resonator . . . 46
4.5.1 Transfer matrix and scattering matrix . . . 46
4.5.2 Resonance condition and complex eigenfrequencies . . . 47
4.5.3 Connection between poles and transmission peaks . . . 47
4.6 Scattering cross-sections and the Wigner time delay . . . 48
4.6.1 Cross-sections from the **S** matrix . . . 48
4.6.2 The Wigner time delay . . . 49
4.7 Spectrum of the open Maxwell operator . . . 49
4.8 Perfectly matched layers . . . 50
4.8.1 Complex coordinate stretching . . . 50
4.8.2 Spectral consequences of the PML . . . 51
4.8.3 PML as a non-Hermitian mapping . . . 52
4.9 Feshbach projection and effective non-Hermitian Hamiltonians . . . 52
4.10 From scattering to non-Hermitian spectral theory . . . 53

II Physical Origins of Non-Hermiticity 54

5 Loss, Gain, and Dispersion 55
5.1 Complex permittivity from linear response . . . 55
5.1.1 Symmetry constraints on $\varepsilon_{\mathrm{r}}(\omega)$ . . . 56
5.2 The Lorentz oscillator model . . . 56

5.2.1 Single resonance . . . 56
5.2.2 Multiple resonances and the Drude model . . . 57
5.3 Kramers–Kronig relations . . . 57
5.3.1 Derivation from causality . . . 57
5.3.2 Physical consequences . . . 59
5.3.3 Sum rules . . . 59
5.3.4 Passivity and the power dissipation condition . . . 60
5.4 Energy in lossy and dispersive media . . . 60
5.4.1 The Brillouin formula . . . 60
5.4.2 Example: energy density near a Lorentz resonance . . . 60
5.4.3 Why a simple inner product fails . . . 61
5.5 Auxiliary fields and the Hamiltonian embedding . . . 61
5.5.1 The idea . . . 61
5.5.2 Example: single Lorentz pole . . . 61
5.5.3 The extended Hamiltonian . . . 62
5.6 Gain media . . . 62
5.6.1 The inverted Lorentz model . . . 63
5.6.2 Population inversion and the four-level system . . . 63
5.6.3 Gain saturation . . . 64
5.7 Balanced gain and loss: the $\mathcal{PT}$-symmetric condition . . . 65
5.7.1 The $\mathcal{PT}$-symmetric Lorentz model . . . 66
5.8 Summary and connections . . . 66

6 Quasinormal Modes 68
6.1 Defining quasinormal modes . . . 68
6.1.1 Example: Fabry–Pérot QNMs . . . 69
6.2 Spatial divergence and its physical meaning . . . 69
6.2.1 Connection to the skin effect . . . 70
6.3 PML regularization and the non-self-adjoint operator . . . 70
6.3.1 How PML regularization works . . . 70
6.3.2 PML validation and spurious modes . . . 71
6.4 The adjoint problem and biorthogonality . . . 72
6.4.1 The adjoint Maxwell operator . . . 72
6.4.2 Biorthogonality . . . 72
6.4.3 Physical interpretation . . . 72
6.5 QNM normalization . . . 73
6.5.1 The unconjugated norm . . . 73
6.5.2 The surface-integral correction . . . 73
6.5.3 Comparison of normalization schemes . . . 73
6.5.4 The Petermann factor from the unconjugated norm . . . 74

6.6 Modal expansion of the Green tensor . . . 75
6.6.1 Local density of states . . . 75
6.6.2 Purcell factor . . . 75
6.7 Perturbation theory for QNMs . . . 76
6.8 Modal non-orthogonality and the Petermann factor . . . 77
6.9 Completeness and limitations . . . 78
6.9.1 The completeness question . . . 78
6.9.2 When is a QNM expansion sufficient? . . . 78
6.9.3 Relation to coupled-mode theory . . . 79

7 From Maxwell to Coupled Modes 80
7.1 Three levels of description . . . 80
7.2 Temporal coupled-mode theory . . . 80
7.2.1 The phenomenological starting point . . . 81
7.2.2 TCMT constraints from energy conservation . . . 81
7.2.3 What TCMT does not tell us . . . 82
7.3 Deriving the effective Hamiltonian from QNMs . . . 82
7.3.1 Starting point: the QNM expansion . . . 82
7.3.2 Mode amplitudes and equations of motion . . . 82
7.3.3 Mode-overlap integrals . . . 83
7.3.4 The effective Hamiltonian . . . 83
7.4 Example: two coupled cavities . . . 83
7.4.1 Isolated-cavity QNMs . . . 83
7.4.2 Coupling strength . . . 84
7.4.3 The two-mode Hamiltonian . . . 84
7.5 The reaction inner product . . . 85
7.5.1 Reciprocity and the symmetric coupling . . . 85
7.5.2 Connection to QNM biorthogonality . . . 86
7.5.3 When reaction-CMT and TCMT disagree . . . 86
7.6 Two coupled resonances: the minimal model . . . 86
7.7 Validity and breakdown . . . 87
7.8 Beyond two modes . . . 89
7.8.1 Multi-mode effective Hamiltonians . . . 90
7.8.2 From coupled modes to scattering matrix . . . 90

III Exceptional Photonics and Non-Hermitian Scattering 92

8 Exceptional Points 93
8.1 Degeneracy versus defectiveness . . . 93
8.1.1 Codimension counting . . . 93

8.2 The 2×2 exceptional point . . . 94
8.2.1 Eigenvalue structure . . . 94
8.2.2 Square-root topology . . . 94
8.2.3 Eigenvector coalescence . . . 95
8.3 Jordan form and the generalized eigenvector . . . 95
8.3.1 Dynamical consequences . . . 96
8.3.2 Pseudospectrum interpretation . . . 96
8.4 Encircling an exceptional point . . . 96
8.4.1 Dynamic encircling . . . 97
8.5 The Petermann factor at exceptional points . . . 97
8.5.1 Physical consequences . . . 97
8.5.2 Phase rigidity . . . 98
8.6 Exceptional points in photonics . . . 98
8.6.1 Coupled microring resonators . . . 98
8.6.2 Example: coupled microdisks . . . 98
8.6.3 Waveguide systems . . . 99
8.6.4 Single-resonator exceptional points . . . 99
8.7 Sensing at exceptional points . . . 99
8.7.1 Enhanced responsivity . . . 99
8.7.2 The noise problem . . . 99
8.7.3 Can nonlinearity help? . . . 100
8.8 Higher-order exceptional points . . . 101
8.8.1 Realization challenges . . . 102
8.8.2 Noise scaling at higher-order EPs . . . 102
8.9 Exceptional points and topology . . . 103
8.9.1 EP charges and conservation . . . 103
8.9.2 Connection to band topology . . . 103

9 $\mathcal{PT}$-Symmetric Optics 104
9.1 The $\mathcal{PT}$-symmetry condition . . . 104
9.1.1 Parity and time reversal in Maxwell's equations . . . 104
9.1.2 $\mathcal{PT}$ symmetry does not mean Hermiticity . . . 105
9.1.3 $\mathcal{PT}$ symmetry from the energy inner product . . . 105
9.2 $\mathcal{PT}$ phase transition in the two-mode model . . . 105
9.2.1 Coupled-mode derivation from Maxwell fields . . . 106
9.3 Transfer-matrix theory for $\mathcal{PT}$ gratings . . . 107
9.3.1 The 1D $\mathcal{PT}$-symmetric Bragg grating . . . 107
9.3.2 Transfer matrix for one period . . . 107
9.3.3 Example: a $\mathcal{PT}$ bilayer . . . 107
9.3.4 Unidirectional invisibility . . . 108

9.3.5 Spectral features near the Bragg condition . . . 109
9.4 $\mathcal{PT}$ symmetry of the scattering matrix . . . 109
9.4.1 Derivation of the unit-modulus condition . . . 109
9.4.2 Phase diagram of the scattering eigenvalues . . . 110
9.5 Beyond paraxial: $\mathcal{PT}$ symmetry in the full Maxwell framework . . . 110
9.5.1 Vectorial $\mathcal{PT}$ condition . . . 110
9.5.2 When the scalar model is not enough . . . 110
9.6 Experimental realizations . . . 111
9.6.1 Coupled optical waveguides . . . 111
9.6.2 Whispering-gallery microcavities . . . 111
9.6.3 Photonic lattices and mesh networks . . . 111
9.6.4 Microring resonators . . . 111
9.7 Non-normality and transient amplification . . . 112
9.7.1 The mechanism . . . 112
9.7.2 Pseudospectrum . . . 113
9.8 Nonlinear $\mathcal{PT}$ symmetry . . . 114
9.8.1 Gain saturation and self-consistent $\mathcal{PT}$ symmetry . . . 114
9.8.2 Nonlinear exceptional points . . . 114

10 Poles, Zeros, Lasers, and Absorbers 116
10.1 Scattering-matrix singularities revisited . . . 116
10.1.1 Poles: threshold lasing . . . 116
10.1.2 Zeros: coherent perfect absorption . . . 117
10.1.3 The meromorphic structure . . . 117
10.2 Example: Fabry–Pérot cavity . . . 117
10.2.1 Setup and transfer matrix . . . 117
10.2.2 Pole positions . . . 118
10.2.3 Adding gain: pole motion toward lasing . . . 118
10.2.4 Zero positions and CPA . . . 119
10.3 Time reversal and the CPA-laser duality . . . 119
10.3.1 Formal statement . . . 119
10.3.2 CPA-laser point . . . 119
10.4 $\mathcal{PT}$-symmetric scattering . . . 120
10.4.1 Scattering phase diagram . . . 120
10.5 Transfer-matrix analysis of a $\mathcal{PT}$ multilayer . . . 120
10.5.1 Setup . . . 120
10.5.2 Output coefficients . . . 121
10.5.3 CPA-laser versus exceptional point . . . 121
10.6 Spectral singularities . . . 121
10.6.1 Definition . . . 121

10.6.2 Physical interpretation . . . 122
10.6.3 Threshold condition as a spectral singularity . . . 122
10.6.4 Spectral singularities in $\mathcal{PT}$-symmetric potentials . . . 123
10.7 Beyond $\mathcal{PT}$: general pole–zero engineering . . . 123
10.7.1 Design as singularity placement . . . 123
10.7.2 Passivity constraints . . . 125
10.7.3 Connection to QNM perturbation theory . . . 125
10.8 Connections . . . 125

IV Topology Before and After Hermiticity 126

11 Berry Curvature from Maxwell Fields 127
11.1 Why the inner product matters . . . 127
11.2 Geometric phase for photonic Bloch states . . . 128
11.2.1 Berry connection . . . 128
11.2.2 Berry curvature . . . 128
11.2.3 Physical interpretation . . . 128
11.3 Chern number . . . 129
11.3.1 Quantization proof sketch . . . 130
11.3.2 Time-reversal symmetry . . . 130
11.3.3 Valley Chern number . . . 130
11.4 Photonic Chern insulators . . . 130
11.4.1 Magneto-optical photonic crystals . . . 130
11.4.2 Photonic anomalous Hall effect . . . 131
11.4.3 Limitations . . . 132
11.5 Wilson loop and $\mathbb{Z}_2$ invariants . . . 133
11.6 Numerical computation . . . 134
11.6.1 Discretized Berry curvature . . . 134
11.6.2 Practical considerations . . . 134
11.6.3 Chern number from integration . . . 135
11.7 Bulk-boundary correspondence . . . 135
11.7.1 Edge dispersion . . . 135
11.7.2 Robustness . . . 135
11.8 Preview: non-Hermitian modifications . . . 136
11.8.1 From Hermitian to biorthogonal: what changes physically . . . 136

12 Complex Bands and Non-Hermitian Gaps 138
12.1 From real to complex band structure . . . 138
12.1.1 How complex bands arise from Maxwell equations . . . 138
12.1.2 Consequences of complexification . . . 139

12.1.3 A concrete example: 1D gain–loss lattice . . . 139
12.2 Line gaps and point gaps . . . 140
12.2.1 Line gap . . . 140
12.2.2 Point gap . . . 141
12.2.3 Relationship between the two gap types . . . 141
12.3 Spectral winding number . . . 142
12.3.1 Definition . . . 142
12.3.2 Physical meaning . . . 142
12.3.3 Computational procedure . . . 143
12.3.4 Example: Hatano–Nelson model . . . 143
12.4 Biorthogonal Berry phase . . . 143
12.4.1 Complex-valued Berry phase . . . 144
12.4.2 Singularity at exceptional points . . . 145
12.5 Exceptional points in band structure . . . 145
12.5.1 Exceptional lines and surfaces . . . 145
12.5.2 Bulk Fermi arcs . . . 145
12.6 The 38-fold classification . . . 146
12.6.1 From tenfold to 38-fold . . . 146
12.6.2 Photonic symmetry classes . . . 146
12.6.3 Reduction to the Hermitian tenfold way . . . 147
12.6.4 Beyond the periodic table . . . 148

13 Photonic Point-Gap Topology **149**
13.1 Point-gap topology from Maxwell Bloch states . . . 149
13.1.1 Maxwell-level origin . . . 149
13.1.2 Contrast with line-gap topology . . . 150
13.2 Reciprocity constraints . . . 150
13.2.1 Reciprocity and vanishing point-gap winding . . . 151
13.2.2 Physical implications . . . 152
13.3 Example: complex bands of a reciprocal bilayer . . . 152
13.3.1 Transfer-matrix computation . . . 153
13.3.2 Complex band structure . . . 153
13.3.3 Computing the winding number . . . 153
13.3.4 Sensitivity to gain–loss contrast . . . 154
13.3.5 Relation to the Green's function . . . 154
13.4 From winding number to boundary response . . . 155
13.4.1 Spectral flow and parameter-driven transitions . . . 155
13.4.2 Boundary response of the bilayer . . . 155
13.5 Two-dimensional point-gap topology . . . 156
13.5.1 Spectral area as a diagnostic . . . 156

13.5.2 Implications for photonic crystals . . . 157
13.6 Photonic realizations and experiments . . . 158
13.7 Vector nature and polarization . . . 159

14 Skin Effect and Non-Bloch Light 161
14.1 The breakdown of Bloch band theory . . . 161
14.1.1 Skin modes . . . 161
14.1.2 Spectral sensitivity . . . 161
14.1.3 A quantitative measure . . . 162
14.2 Physical origin of the skin effect . . . 162
14.2.1 Sources of asymmetry in photonic systems . . . 162
14.2.2 The imaginary gauge field . . . 163
14.3 Example: the Hatano–Nelson model . . . 164
14.3.1 PBC spectrum . . . 164
14.3.2 OBC spectrum . . . 164
14.3.3 GBZ construction . . . 164
14.4 The generalized Brillouin zone: general theory . . . 166
14.4.1 From Bloch to non-Bloch . . . 166
14.4.2 Construction of the GBZ . . . 166
14.4.3 Geometry of the GBZ . . . 167
14.4.4 Non-Bloch band structure . . . 168
14.5 Non-Bloch topological invariants . . . 168
14.5.1 Non-Bloch winding number . . . 169
14.5.2 Non-Bloch Chern number . . . 169
14.5.3 When GBZ invariants differ from Bloch invariants . . . 169
14.6 Skin effect in two and higher dimensions . . . 169
14.7 Photonic skin effect . . . 170
14.8 Critical non-Hermitian skin effect . . . 171
14.9 Connection to earlier chapters . . . 172
14.9.1 From Maxwell transfer matrices to the photonic GBZ . . . 172

15 Bulk-Boundary Correspondence Rebuilt 174
15.1 Why the conventional correspondence fails . . . 174
15.1.1 The Hermitian baseline . . . 174
15.1.2 The non-Hermitian SSH model . . . 175
15.1.3 The failure . . . 175
15.2 The non-Bloch bulk-boundary correspondence . . . 176
15.2.1 Statement . . . 177
15.2.2 The GBZ for the non-Hermitian SSH model . . . 177
15.2.3 Derivation of the non-Bloch BBC . . . 177
15.3 Example: non-Hermitian SSH phase diagram . . . 178

15.3.1 Bloch versus non-Bloch phase boundaries . . . 179
15.3.2 OBC spectrum and edge states . . . 179
15.4 Biorthogonal bulk-boundary correspondence . . . 179
15.4.1 Biorthogonal polarization . . . 179
15.4.2 Physical interpretation . . . 179
15.4.3 Connection to QNM inner products . . . 180
15.4.4 Relation to the non-Bloch approach . . . 181
15.5 Higher-dimensional correspondence . . . 181
15.5.1 Non-Bloch Chern insulator . . . 181
15.5.2 Corner modes and higher-order topology . . . 181
15.5.3 Geometry dependence . . . 182
15.6 Photonic implications . . . 182
15.6.1 Photonic GBZ construction . . . 182
15.6.2 Design of robust edge channels . . . 183
15.6.3 Directional amplification . . . 183
15.6.4 Spectral engineering . . . 183
15.6.5 Experimental signatures . . . 183

16 Bound States in the Continuum and Exceptional Topology 186
16.1 What is a bound state in the continuum? . . . 186
16.1.1 The radiation-amplitude viewpoint . . . 186
16.1.2 Symmetry-protected BICs . . . 187
16.1.3 Accidental (tuned) BICs . . . 187
16.1.4 Other BIC mechanisms . . . 187
16.1.5 Friedrich–Wintgen avoided crossing . . . 188
16.2 Topological charge of a BIC . . . 189
16.2.1 Far-field polarization as a vector field on the BZ . . . 189
16.2.2 The charge integral . . . 189
16.2.3 Charge conservation and annihilation . . . 190
16.2.4 Charge assignment in photonic crystal slabs . . . 191
16.3 Merging BICs and the super-BIC regime . . . 192
16.3.1 The finite-size problem . . . 192
16.3.2 Merging charges for steeper suppression . . . 192
16.3.3 Experimental demonstration: the super-BIC laser . . . 193
16.4 Brillouin-zone-folding BICs . . . 194
16.4.1 Folding guided modes into the continuum . . . 194
16.4.2 Q-factor scaling and perturbation control . . . 195
16.4.3 Disorder robustness . . . 195
16.5 BICs and exceptional points . . . 195
16.5.1 EP rings around BICs . . . 195

16.5.2 Merging BICs and EPs . . . 195
16.5.3 Exceptional surfaces from BIC perturbation . . . 196
16.6 Quasi-BIC resonances and applications . . . 196
16.6.1 Quasi-BICs as high-Q resonances . . . 196
16.6.2 BIC lasers . . . 196
16.6.3 Nonlinear and sensing applications . . . 197

V Computation, Devices, and Frontiers 198

17 Numerical Methods for Non-Hermitian Maxwell Problems 199
17.1 The computational workflow . . . 199
17.2 Finite-element method with PML . . . 200
17.2.1 Weak formulation of the Maxwell eigenproblem . . . 200
17.2.2 PML implementation . . . 200
17.2.3 Challenges of non-Hermiticity . . . 201
17.3 QNM extraction and normalization . . . 202
17.3.1 Shift-and-invert Arnoldi . . . 202
17.3.2 Pole-search methods . . . 203
17.3.3 QNM normalization in practice . . . 203
17.4 PML validation and spurious-mode filtering . . . 203
17.4.1 PML parameter variation . . . 204
17.4.2 Mode-ratio criterion . . . 204
17.5 Contour-integral eigensolvers . . . 205
17.5.1 The principle . . . 205
17.5.2 Beyn's algorithm . . . 205
17.5.3 FEAST algorithm . . . 206
17.5.4 Poles and zeros simultaneously . . . 207
17.6 Dispersive materials and nonlinear eigenproblems . . . 207
17.6.1 Auxiliary-field linearization . . . 208
17.6.2 Polynomial eigenproblem . . . 208
17.6.3 Convergence caveats . . . 208
17.7 Plane-wave expansion for complex bands . . . 208
17.7.1 Computing topological invariants . . . 208
17.7.2 Non-Bloch PWE . . . 209
17.8 FDTD with gain and loss . . . 209
17.8.1 Stability and gain saturation . . . 209
17.8.2 Extracting QNM frequencies from time series . . . 209
17.9 Specialized algorithms . . . 209
17.9.1 EP tracking . . . 209
17.9.2 GBZ computation . . . 210

17.10Practical guidelines 210

18 Devices and Physical Limits **212**

18.1 $\mathcal{PT}$-symmetric lasers 212
18.1.1 Single-mode selection from mode non-orthogonality 212
18.1.2 Experimental demonstrations 213
18.1.3 Power, coherence, and the Petermann penalty 213
18.1.4 Beyond single-mode: multi-element $\mathcal{PT}$ lasers 214
18.2 Exceptional-point sensors 214
18.2.1 Enhanced frequency splitting 214
18.2.2 The noise problem: quantum Fisher information 215
18.2.3 The noise problem: nonlinear and active systems 216
18.2.4 When can EP sensing help? 217
18.3 BIC-based lasers and cavities 217
18.3.1 From radiation suppression to lasing threshold 217
18.3.2 Super-BIC lasers 218
18.3.3 BZF-BIC cavities 218
18.3.4 Quasi-BIC applications beyond lasing 219
18.4 Coherent perfect absorbers and switches 219
18.4.1 CPA as a scattering zero 219
18.4.2 CPA-laser devices 219
18.4.3 Practical CPA limitations 220
18.5 Topological lasers 221
18.5.1 Topological insulator lasers 221
18.5.2 Non-Hermitian topological lasers 221
18.5.3 Non-Hermitian skin amplifiers 222
18.6 Active non-Hermitian metasurfaces 222
18.6.1 Wavefront control with gain and loss 222
18.6.2 $\mathcal{PT}$-symmetric metasurfaces 222
18.6.3 CPA metasurfaces 223
18.7 Physical limits 223
18.7.1 Quantum noise and the Petermann factor 223
18.7.2 Gain saturation and nonlinearity 223
18.7.3 Fabrication disorder 224
18.7.4 Thermal and power limits 224
18.7.5 Gain–bandwidth constraints 225
18.7.6 Summary: device concepts and their physical limits 225

19 Open Problems and Outlook **227**

19.1 Nonlinear non-Hermitian photonics 227
19.1.1 Why linearity is not enough 227

19.1.2 Gain saturation and the fate of exceptional points . . . 227
19.1.3 Nonlinear topology . . . 228
19.2 Quantum non-Hermitian photonics . . . 229
19.2.1 From effective Hamiltonians to open quantum systems . . . 229
19.2.2 Few-photon non-Hermitian physics . . . 230
19.2.3 Entanglement and biorthogonality . . . 230
19.3 Time-varying and Floquet non-Hermitian systems . . . 230
19.3.1 Floquet engineering of non-Hermiticity . . . 230
19.3.2 Floquet exceptional points . . . 230
19.3.3 Open problems . . . 231
19.4 Non-Hermitian photonics in three dimensions . . . 231
19.4.1 Exceptional lines and surfaces . . . 232
19.4.2 Fabrication and characterization challenges . . . 233
19.5 Rigorous spectral theory of open Maxwell operators . . . 233
19.5.1 Completeness of QNM expansions . . . 233
19.5.2 Continuous spectrum and branch cuts . . . 234
19.5.3 Spectral singularities . . . 234
19.6 Inverse design with non-Hermitian constraints . . . 234
19.7 Energy, entropy, and thermodynamics . . . 236
19.8 Beyond photonics . . . 236
19.9 Toward a Maxwell-first synthesis . . . 237

A Vector Calculus and Maxwell Notation 240
A.1 Curl identities . . . 240
A.2 Integration by parts for curl operators . . . 240
A.2.1 Boundary conditions that kill the surface term . . . 241
A.3 The divergence theorem and Poynting's theorem . . . 241
A.4 Bloch-periodic boundary terms . . . 241
A.5 Silver–Müller radiation condition . . . 242
A.6 Notation conventions . . . 242
A.6.1 Time convention . . . 242
A.6.2 Field quantities . . . 242
A.6.3 Operators and inner products . . . 243
A.7 Electromagnetic reciprocity from vector calculus . . . 243
A.7.1 Statement and proof . . . 243
A.7.2 Reciprocity versus Hermiticity . . . 244

B Non-Normal Linear Algebra 245
B.1 Left and right eigenvectors . . . 245
B.2 Biorthogonality . . . 245
B.3 Jordan chains and defective eigenvalues . . . 246

B.3.1 Jordan normal form . . . 246
B.3.2 Generalized eigenvectors . . . 246
B.3.3 Resolvent near an EP . . . 246
B.4 Petermann factor and excess noise . . . 247
B.4.1 Physical interpretation . . . 247
B.4.2 Scaling near an EP . . . 247
B.5 Phase rigidity . . . 247
B.6 Pseudospectra . . . 248
B.6.1 Definition . . . 248
B.6.2 Physical relevance . . . 248
B.7 Matrix functions and the non-Hermitian exponential . . . 248
B.8 Examples . . . 249
B.8.1 Jordan decomposition of a 2×2 defective matrix . . . 249
B.8.2 Schur decomposition and its use . . . 249
B.8.3 Kreiss constant and transient bounds . . . 250
B.8.4 Non-normal perturbation theory . . . 250

C Transfer Matrices and Scattering **251**
C.1 Single-layer transfer matrix . . . 251
C.2 Scattering matrix from the transfer matrix . . . 251
C.3 $\mathcal{PT}$-symmetric transfer matrices . . . 252
C.3.1 $\mathcal{PT}$ phase transition from transfer-matrix eigenvalues . . . 252
C.4 Poles and zeros from the transfer matrix . . . 252
C.5 Bloch waves in periodic stacks . . . 253
C.5.1 GBZ from the transfer matrix . . . 253
C.6 Numerical implementation . . . 253
C.6.1 Avoiding numerical overflow . . . 253
C.6.2 Complex-frequency evaluation . . . 254
C.7 From transfer matrix to scattering matrix . . . 254
C.7.1 Conversion formulas . . . 254
C.7.2 Example: Fabry–Pérot resonator . . . 254

D Numerical Recipes **256**
D.1 Lorentz dispersion model . . . 256
D.2 Two-mode exceptional point . . . 256
D.3 Point-gap winding number . . . 257
D.4 Non-Hermitian skin effect (Hatano–Nelson model) . . . 258
D.5 BIC topological charge and Q scaling . . . 258
D.6 PML validation and mode-ratio filtering . . . 258
D.7 Scattering pole-zero map . . . 259
D.8 EP sensing response versus noise . . . 259

D.9 Practical convergence protocol . . . 260
D.9.1 Mesh refinement . . . 260
D.9.2 PML thickness and absorption profile . . . 260
D.9.3 Dispersive material branch verification . . . 261
D.9.4 Biorthogonal normalisation check . . . 261
D.9.5 Reproducibility checklist . . . 261
D.9.6 Minimum reporting standard for Maxwell eigenvalue computations . . . 262

E Notation and Symbols **264**
E.1 General conventions . . . 264
E.2 Electromagnetic fields and sources . . . 265
E.3 Material and structural parameters . . . 265
E.4 Operators and eigenproblems . . . 266
E.5 Scattering and transfer matrices . . . 266
E.6 Topology and band theory . . . 266
E.7 Abbreviations . . . 267
E.8 Sign conventions and equivalent formulations . . . 267
E.8.1 Time-harmonic convention . . . 268
E.8.2 Three formulations of the Petermann factor . . . 268
E.8.3 Gain–loss sign in coupled-mode models . . . 269
E.8.4 Transfer-matrix ordering . . . 269

Part I

Maxwell Equations Become Spectral Theory

Chapter 1

Light as an Eigenvalue Problem

What does it mean, physically, to say that light has "modes"? Every physicist who has worked with optical cavities, waveguides, or photonic crystals eventually invokes a set of modes—discrete field patterns, each oscillating at a single frequency. This chapter shows where those modes come from: they are eigenvalues and eigenvectors of an operator built from Maxwell's equations and the material response of the medium. The structure of that operator—whether it is self-adjoint or not—controls everything that follows in this book. We shall derive the eigenvalue problem carefully, explain why it takes the form it does, explore its mathematical structure through a variational principle, and illustrate it with a worked one-dimensional example before giving a preview of what happens when the operator loses its self-adjoint character.

1.1 Maxwell's equations in matter

We begin with the macroscopic Maxwell equations in a linear, time-invariant medium with no free charges or currents:

$$\nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial \mathbf{B}(\mathbf{r},t)}{\partial t}, \tag{1.1}$$

$$\nabla \times \mathbf{H}(\mathbf{r},t) = \frac{\partial \mathbf{D}(\mathbf{r},t)}{\partial t}, \tag{1.2}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r},t) = 0, \tag{1.3}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r},t) = 0. \tag{1.4}$$

These four equations, supplemented by constitutive relations connecting $\mathbf{D}$ to $\mathbf{E}$ and $\mathbf{B}$ to $\mathbf{H}$, form the complete classical description of electromagnetic fields in matter. Nothing in this book requires quantum electrodynamics; the entire non-Hermitian structure we shall develop is already present at the level of classical Maxwell theory.

In the simplest case the constitutive relations are local and instantaneous:

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0\, \varepsilon_{\mathrm{r}}(\mathbf{r})\, \mathbf{E}(\mathbf{r},t), \qquad \mathbf{B}(\mathbf{r},t) = \mu_0\, \mu_{\mathrm{r}}(\mathbf{r})\, \mathbf{H}(\mathbf{r},t). \tag{1.5}$$

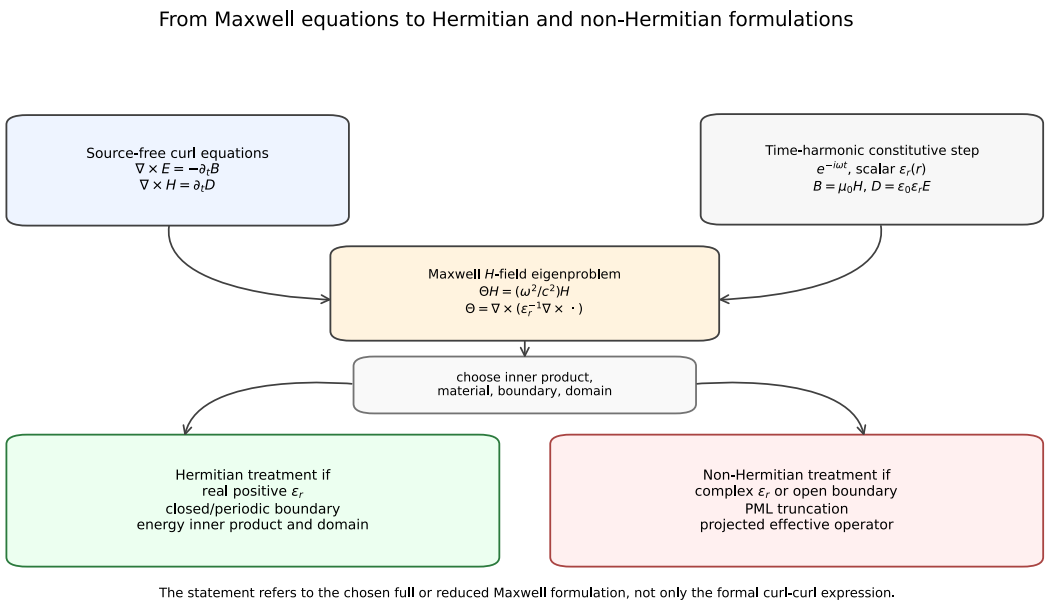


Figure 1.1 The Maxwell-first route followed in this book. Starting from macroscopic Maxwell equations and constitutive relations, eliminating one field gives the curl–curl operator. When the material is real, the surface term vanishes, and the transverse field space is complete, the operator gives Hermitian photonics. Loss or gain, radiation boundaries, dispersion, and projection onto a subspace lead to the non-Hermitian structures studied in later chapters.

Here $\varepsilon_{\mathrm{r}}(\mathbf{r})$ and $\mu_{\mathrm{r}}(\mathbf{r})$ are the relative permittivity and permeability, which we allow to vary in space but, for now, take to be real and positive. These assumptions will be progressively relaxed in later chapters: Chapter 5 introduces complex and frequency-dependent ε_{r}, Chapter 6 opens the system to radiation, and Chapter 9 considers the special spatial symmetry of balanced gain and loss.

The constitutive relations (1.5) encode all material information available to the Maxwell operator. From the standpoint of the eigenvalue problem we are about to derive, $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is the central material input: it controls the spatial pattern of the eigenmodes, the distribution of eigenfrequencies, and—when it becomes complex—the transition from Hermitian to non-Hermitian spectral theory.

Why macroscopic Maxwell?

A monograph on non-Hermitian photonics could, in principle, start from the microscopic Maxwell equations and derive macroscopic constitutive relations. We choose to start at the macroscopic level for two reasons. First, the non-Hermitian phenomena we study—complex eigenfrequencies, exceptional points, non-reciprocal skin effects—arise already in the macroscopic theory and do not require a microscopic derivation to be well posed. Second, the macroscopic framework is what connects directly to photonic-crystal band computations, scattering-matrix measurements, and device design.

1.2 The frequency-domain eigenproblem

1.2.1 Harmonic fields and the time-frequency correspondence

For a time-invariant medium it is natural to look for solutions with harmonic time dependence. We write

$$\mathbf{E}(\mathbf{r},t) = \mathrm{Re}\big[\mathbf{E}(\mathbf{r})\, e^{-i\omega t}\big], \qquad \mathbf{H}(\mathbf{r},t) = \mathrm{Re}\big[\mathbf{H}(\mathbf{r})\, e^{-i\omega t}\big], \tag{1.6}$$

where ω is a real angular frequency and $\mathbf{E}(\mathbf{r})$, $\mathbf{H}(\mathbf{r})$ are complex-valued spatial amplitudes. At this stage ω may be regarded as the real frequency of a harmonic drive. Later, when we solve source-free open or lossy eigenvalue problems, the same convention will allow ω itself to become complex. The sign convention $e^{-i\omega t}$ is standard in the optics and photonics literature; the condensed-matter convention $e^{+i\omega t}$ leads to conjugated constitutive parameters but no change in the physics.

The physical content of the harmonic ansatz is that once we know the spatial amplitude $\mathbf{E}(\mathbf{r})$ at a single frequency ω, we know the complete space-time field at that frequency. The general time-domain solution is then recovered by Fourier superposition. The advantage of the frequency-domain formulation is that the partial differential equations in space-time reduce to ordinary or partial differential equations in space alone—eigenvalue problems whose solutions organize the entire electromagnetic response.

1.2.2 Deriving the curl–curl eigenproblem

Substituting (1.6) into Faraday's law (1.1) and Ampère's law (1.2) gives

$$\nabla \times \mathbf{E} = i\omega\, \mu_0\mu_\mathrm{r}\, \mathbf{H}, \tag{1.7}$$

$$\nabla \times \mathbf{H} = -i\omega\, \varepsilon_0\varepsilon_\mathrm{r}\, \mathbf{E}. \tag{1.8}$$

These are two coupled first-order equations for $\mathbf{E}$ and $\mathbf{H}$. We can decouple them by eliminating one field. Applying $\nabla\times$ to (1.8) and inserting (1.7), we obtain the *curl–curl eigenproblem for the electric field*:

$$\boxed{\nabla \times \left[\frac{1}{\mu_\mathrm{r}(\mathbf{r})}\, \nabla \times \mathbf{E}(\mathbf{r})\right] = \frac{\omega^2}{c^2}\, \varepsilon_\mathrm{r}(\mathbf{r})\, \mathbf{E}(\mathbf{r}).} \tag{1.9}$$

Here $c = 1/\sqrt{\varepsilon_0\mu_0}$ is the speed of light in vacuum. Equation (1.9) is a generalized eigenvalue problem: the operator on the left acts on $\mathbf{E}$, the eigenvalue is ω^2/c^2, and the "weight" on the right is the position-dependent $\varepsilon_\mathrm{r}(\mathbf{r})$.

The elimination procedure is worth understanding in detail, because it reveals where the operator structure comes from. Writing the two steps explicitly:

1. Use Faraday's law to express the magnetic field as $\mathbf{H} = (i\omega\mu_0\mu_\mathrm{r})^{-1}\nabla \times \mathbf{E}$.
2. Insert this expression into Ampère's law:

$$\nabla \times \left[\frac{1}{i\omega\mu_0\mu_\mathrm{r}} \nabla \times \mathbf{E}\right] = -i\omega\varepsilon_0\varepsilon_\mathrm{r}\mathbf{E},$$

   and multiply by $i\omega\mu_0$ to obtain $\nabla \times [(1/\mu_\mathrm{r})\nabla \times \mathbf{E}] = (\omega^2/c^2)\varepsilon_\mathrm{r}\mathbf{E}$.

The curl–curl structure on the left is the price we pay for eliminating one field; the ε_r weight on the right arises from the constitutive relation. Together they produce a second-order operator in

space—directly analogous to the kinetic-energy operator $-\nabla^2$ in quantum mechanics, but with the curl replacing the gradient and the permittivity entering as a generalized eigenvalue weight rather than as an additive potential.

1.2.3 The magnetic-field formulation

An analogous equation can be written for $\mathbf{H}$ alone. Eliminating $\mathbf{E}$ from (1.7)–(1.8) yields

$$\nabla \times \left[\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right] = \frac{\omega^2}{c^2} \mu_{\mathrm{r}}(\mathbf{r})\, \mathbf{H}(\mathbf{r}). \tag{1.10}$$

For non-magnetic media ($\mu_{\mathrm{r}} = 1$ everywhere), (1.10) simplifies to the standard form used in photonic-crystal theory:

$$\hat{\Theta}\, \mathbf{H}(\mathbf{r}) \;\equiv\; \nabla \times \left[\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right] = \frac{\omega^2}{c^2} \mathbf{H}(\mathbf{r}). \tag{1.11}$$

The operator $\hat{\Theta}$ defined here will be a central object throughout this book [1, 2]. Its properties—self-adjointness, compactness of the resolvent, spectral structure—depend critically on three ingredients:

1. the spatial profile of $\varepsilon_{\mathrm{r}}(\mathbf{r})$ (real or complex, bounded or unbounded, periodic or not);
2. the boundary conditions imposed on $\mathbf{H}$ at the edges of the computational or physical domain;
3. the function space on which $\hat{\Theta}$ acts.

Both the $\mathbf{E}$-field and $\mathbf{H}$-field formulations are exact restatements of Maxwell's equations for harmonic fields. However, the $\mathbf{H}$-field form has a practical advantage for non-magnetic media: the eigenvalue ω^2/c^2 appears with a trivial (unit) weight on the right, making the operator $\hat{\Theta}$ a standard eigenvalue problem rather than a generalized one. For this reason, most photonic band-structure codes work with the $\mathbf{H}$ formulation, and we shall use $\hat{\Theta}$ as the default Maxwell operator throughout this book unless otherwise stated.

1.2.4 The divergence constraint

The curl–curl operator has a large null space: any gradient field $\mathbf{H} = \nabla\phi$ satisfies $\nabla \times \mathbf{H} = 0$ and therefore $\hat{\Theta}(\nabla\phi) = 0$. These spurious zero-frequency solutions are not physical propagating modes; in general they do not satisfy the transverse constraint $\nabla \cdot \mathbf{B} = \mu_0 \nabla \cdot \mathbf{H} = 0$ (or equivalently $\nabla \cdot (\varepsilon_{\mathrm{r}}\mathbf{E}) = 0$ in the $\mathbf{E}$ formulation), and even when a harmonic gradient field survives as a static zero mode it must be separated from the dynamical spectrum.

In analytical work, the divergence condition is enforced by restricting the domain of $\hat{\Theta}$ to transverse (divergence-free) fields:

$$\mathcal{H}_{\mathrm{div}} = \big\{ \mathbf{H} \in L^2(\Omega; \mathbb{C}^3) \;\big|\; \nabla \cdot \mathbf{H} = 0, \text{ plus boundary conditions} \big\}. \tag{1.12}$$

On this restricted domain, the null space is eliminated and $\hat{\Theta}$ becomes positive semi-definite: all non-static eigenvalues satisfy $\omega^2/c^2 \geq 0$. This is physically obvious—squared frequencies

cannot be negative in a closed, lossless medium—but mathematically it relies on the restriction to divergence-free fields and on the treatment of possible harmonic zero modes.

In numerical work (Chapter 17), the divergence constraint must be enforced explicitly, either through a penalty term, a Lagrange multiplier, or a compatible discretization, such as Nédélec edge elements combined with a proper treatment of the gradient null space. Failure to control this constraint produces spurious modes near $\omega = 0$ that can contaminate the physical spectrum.

1.3 Self-adjointness and its physical meaning

1.3.1 The electromagnetic inner product

To ask whether $\hat{\Theta}$ is self-adjoint, we need an inner product. For two vector fields $\mathbf{F}$ and $\mathbf{G}$ defined over a domain Ω, the natural electromagnetic inner product is

$$\langle \mathbf{F}, \mathbf{G} \rangle \;=\; \int_{\Omega} \mathbf{F}^{*}(\mathbf{r}) \cdot \mathbf{G}(\mathbf{r}) \; \mathrm{d}^{3}r. \tag{1.13}$$

We also define a weighted inner product:

$$\langle \mathbf{F}, \mathbf{G} \rangle_{\varepsilon} \;=\; \int_{\Omega} \mathbf{F}^{*}(\mathbf{r}) \cdot \varepsilon_{\mathrm{r}}(\mathbf{r})\, \mathbf{G}(\mathbf{r}) \; \mathrm{d}^{3}r. \tag{1.14}$$

Chapter 2 develops the physical interpretation of these inner products in terms of electromagnetic energy.

1.3.2 Integration by parts and the boundary term

We now ask: under what conditions does $\hat{\Theta}$ satisfy $\langle \mathbf{F}, \hat{\Theta}\mathbf{G} \rangle = \langle \hat{\Theta}\mathbf{F}, \mathbf{G} \rangle$ for all admissible $\mathbf{F}$ and $\mathbf{G}$? The answer comes from integration by parts. Using the vector identity

$$\int_{\Omega} \mathbf{F}^{*} \cdot \left(\nabla \times \mathbf{A} \right) \; \mathrm{d}^{3}r = \int_{\Omega} \left(\nabla \times \mathbf{F}^{*} \right) \cdot \mathbf{A} \; \mathrm{d}^{3}r + \oint_{\partial\Omega} \left(\mathbf{F}^{*} \times \mathbf{A} \right) \cdot \hat{\mathbf{n}} \; \mathrm{d}S, \tag{1.15}$$

applied twice with $\mathbf{A} = (1/\varepsilon_{\mathrm{r}}) \nabla \times \mathbf{G}$, one finds that for $\mu_{\mathrm{r}} = 1$ the bulk integrals are symmetric provided $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is *real*—so that the weight in (1.14) commutes with complex conjugation—and the surface integral

$$\oint_{\partial\Omega} \left[\mathbf{F}^{*} \times \frac{1}{\varepsilon_{\mathrm{r}}} \nabla \times \mathbf{G} + \left(\frac{1}{\varepsilon_{\mathrm{r}}} \nabla \times \mathbf{F} \right)^{*} \times \mathbf{G} \right] \cdot \hat{\mathbf{n}} \; \mathrm{d}S = 0 \tag{1.16}$$

vanishes. For a closed cavity with perfectly conducting walls ($\hat{\mathbf{n}} \times \mathbf{E} = 0$ on $\partial\Omega$), this surface term is zero. For a periodic medium with Bloch boundary conditions, it cancels between opposite faces of the unit cell.

1.3.3 The three conditions for Hermitian photonics

Three conditions for Hermitian photonics

The Maxwell curl–curl operator $\hat{\Theta}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle$ on divergence-free fields when three conditions hold simultaneously: (i) $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is real and positive everywhere in Ω (or, more generally, Hermitian positive definite for anisotropic media); (ii) the boundary conditions on $\partial\Omega$ make the surface term (1.16) vanish; (iii) the domain of $\hat{\Theta}$ is a complete Hilbert space of transverse fields, with static null modes removed or treated separately. Remove any one of these and the spectral theory must be rebuilt.

This result is the foundation on which the entire book rests. Each of the physical mechanisms that produce non-Hermiticity in photonics can be traced back to the violation of one of these three conditions:

- **Material loss or gain** (Chapters 5 and 9): the permittivity acquires an imaginary part, $\varepsilon_{\mathrm{r}} \to \varepsilon_{\mathrm{r}}' + i\varepsilon_{\mathrm{r}}''$, so the weight in (1.14) is no longer real [3]. The operator $\hat{\Theta}$ is no longer formally self-adjoint, and its eigenvalues migrate off the real axis into the complex plane.
- **Radiation and openness** (Chapters 4 and 6): outgoing-wave boundary conditions (Sommerfeld radiation condition or perfectly matched layers) do not make the surface term vanish; instead, energy escapes through $\partial\Omega$. The resulting quasinormal modes have complex eigenfrequencies even when ε_{r} is real [4, 5].
- **Frequency dispersion** (Chapter 5): when ε_{r} depends on ω, the eigenvalue problem (1.11) becomes nonlinear in the eigenvalue, and the standard spectral theory of self-adjoint operators no longer applies directly [1].
- **Projection onto a subspace** (Chapter 7): even when the full Maxwell operator is self-adjoint, projecting onto a few resonant modes produces an effective finite-dimensional Hamiltonian that is generically non-Hermitian, because the discarded modes act as an environment [6].

These four roads to non-Hermiticity are not independent—many real systems exhibit several simultaneously—but they provide a useful organizational framework. A $\mathcal{PT}$-symmetric photonic crystal, for example, has complex ε_{r} (road 1), may be open to radiation (road 2), may involve dispersive gain media (road 3), and is often modeled by a few-mode effective Hamiltonian (road 4).

1.4 The variational principle

Self-adjoint operators with discrete spectra admit a powerful variational characterization of their eigenvalues. For $\hat{\Theta}$ on the space of transverse fields in a closed domain, the Rayleigh quotient is

$$R[\mathbf{H}] = \frac{\displaystyle\int_\Omega \frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})}\,|\nabla\times\mathbf{H}(\mathbf{r})|^2\;\mathrm{d}^3r}{\displaystyle\int_\Omega |\mathbf{H}(\mathbf{r})|^2\;\mathrm{d}^3r} = \frac{\langle\hat{\Theta}\mathbf{H},\,\mathbf{H}\rangle}{\langle\mathbf{H},\,\mathbf{H}\rangle}, \tag{1.17}$$

and the fundamental eigenfrequency satisfies the min–max principle:

$$\frac{\omega_1^2}{c^2} = \min_{\mathbf{H}\in\mathcal{H}_{\mathrm{div}}} R[\mathbf{H}]. \tag{1.18}$$

where the minimization is over nonzero fields in the dynamical transverse subspace. Higher eigenfrequencies are obtained by successive minimization over subspaces orthogonal to the previously found eigenmodes.

The variational principle gives immediate physical insight:

- The lowest mode concentrates its magnetic field where $1/\varepsilon_{\mathrm{r}}$ is smallest, i.e., where the permittivity is largest. This is the electromagnetic analogue of a quantum particle localizing in a potential well.
- Increasing ε_{r} anywhere in Ω decreases all eigenfrequencies—intuitively, denser media slow light and lower the resonant frequencies.
- A band gap forms when no trial field can simultaneously satisfy the curl constraint and have low energy in a frequency window.

The variational principle also explains why the Hermitian eigenfrequencies are always real: the Rayleigh quotient (1.17) is manifestly real when ε_{r} is real and positive, so ω_1^2/c^2 is real and non-negative, and ω_1 is real.

What breaks when ε_{r} is complex. If ε_{r} has an imaginary part, the Rayleigh quotient becomes complex, and the variational characterization fails: there is no longer a minimum of a real functional, so the eigenfrequencies are not constrained to the real axis. This is the variational way of seeing why complex ε_{r} produces non-Hermitian photonics.

1.5 Two-dimensional reductions

Many photonic systems—photonic-crystal slabs, ridge waveguides, fiber cross-sections—have a geometry that is invariant along one spatial direction. For a system uniform along z, all fields can be decomposed into components with a fixed z-dependence $e^{ik_z z}$. At $k_z = 0$ the vector Maxwell equations split into two independent scalar problems, conventionally called *transverse electric* (TE) and *transverse magnetic* (TM) polarizations. The common two-dimensional photonic-crystal convention: TM denotes the scalar E_z problem and TE denotes the scalar H_z problem. Some waveguide texts use the opposite language relative to the propagation direction, so the field component should always be checked.

1.5.1 TM polarization

In TM polarization the electric field is parallel to z: $\mathbf{E} = E_z(\mathbf{r}_\perp)\,\hat{z}$, where $\mathbf{r}_\perp = (x, y)$. The curl–curl equation (1.9) reduces to the scalar Helmholtz equation

$$-\nabla_\perp \cdot \left[\frac{1}{\mu_{\mathrm{r}}(\mathbf{r}_\perp)}\,\nabla_\perp E_z(\mathbf{r}_\perp)\right] = \frac{\omega^2}{c^2}\,\varepsilon_{\mathrm{r}}(\mathbf{r}_\perp)\,E_z(\mathbf{r}_\perp), \tag{1.19}$$

where the divergence form is needed if μ_{r} varies in the transverse plane. For non-magnetic media this simplifies further to

$$-\nabla_{\perp}^{2} E_z = \frac{\omega^2}{c^2}\,\varepsilon_{\mathrm{r}}(\mathbf{r}_{\perp})\,E_z. \tag{1.20}$$

The eigenvalue is again ω^2/c^2, and the inner product is

$$\langle f,\, g\rangle_{\varepsilon}^{\mathrm{TM}} = \int f^{*}(x,y)\,\varepsilon_{\mathrm{r}}(x,y)\,g(x,y)\;\mathrm{d}x\;\mathrm{d}y. \tag{1.21}$$

1.5.2 TE polarization

In TE polarization the magnetic field is parallel to z: $\mathbf{H} = H_z(\mathbf{r}_{\perp})\,\hat{z}$. The $\mathbf{H}$-field equation (1.10) reduces to

$$-\nabla_{\perp}\cdot\left[\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r}_{\perp})}\,\nabla_{\perp} H_z(\mathbf{r}_{\perp})\right] = \frac{\omega^2}{c^2}\,H_z(\mathbf{r}_{\perp}), \tag{1.22}$$

for $\mu_{\mathrm{r}} = 1$. The operator on the left is a weighted Laplacian with $1/\varepsilon_{\mathrm{r}}$ as the spatially varying coefficient. The natural TE inner product is unweighted:

$$\langle f,\, g\rangle^{\mathrm{TE}} = \int f^{*}(x,y)\,g(x,y)\;\mathrm{d}x\;\mathrm{d}y. \tag{1.23}$$

Remark 1.1. The TE and TM inner products differ because the eigenvalue problem has a different structure in each polarization. In TM, ε_{r} appears as a multiplicative weight on the right-hand side; in TE, $1/\varepsilon_{\mathrm{r}}$ appears inside the differential operator. This asymmetry will become important when we define Berry curvature (Chapter 11) and biorthogonal products (Chapter 6) for the two polarizations.

1.5.3 The Schrödinger analogy and its limits

The TM Helmholtz equation (1.20) bears a striking resemblance to the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V(\mathbf{r})\psi = E\psi. \tag{1.24}$$

In both cases, a second-order differential operator acts on a scalar field, and the eigenvalue appears on the right multiplied by a weight. The analogy is useful for building intuition—photonic band gaps correspond to electronic band gaps, defect modes correspond to bound states, and the variational principle operates in both contexts.

However, the analogy has important limits:

1. **Vectorial nature:** Maxwell's equations are inherently vectorial. The TE/TM decoupling occurs only in 2D systems at normal incidence; in 3D or at oblique incidence, the full vector character of the curl–curl operator must be retained.
2. **Eigenvalue position:** In Schrödinger theory, the eigenvalue E appears additively with the potential V; in the Maxwell problem, ω^2/c^2 appears multiplicatively with ε_{r}. This means that the Maxwell spectrum scales homogeneously with the dielectric constant, a property that has no Schrödinger analogue.

3. **No negative eigenvalues:** The Maxwell operator $\hat{\Theta}$ is positive semi-definite on transverse fields, so $\omega^2 \geq 0$ always. The Schrödinger operator can have negative eigenvalues (bound states below the continuum).
4. **Non-Hermiticity:** When ε_r is complex, the Maxwell eigenvalue problem is *nonlinear* in the eigenvalue if ε_r depends on ω, while a complex potential in Schrödinger theory keeps the problem linear. This distinction is important for dispersive gain media (Chapter 5).

1.6 Example: one-dimensional photonic crystal

To make the eigenvalue problem concrete, consider a one-dimensional periodic dielectric stack: alternating layers of material A with permittivity ε_A and thickness d_A, and material B with permittivity ε_B and thickness d_B, repeated with period $a = d_A + d_B$.

1.6.1 Bloch modes and the transfer matrix

By Bloch's theorem (derived in detail in Chapter 3), the eigenmodes of the periodic stack have the form $E(x) = u_k(x)\, e^{ikx}$, where $u_k(x)$ is periodic with period a and k is the Bloch wave vector restricted to the first Brillouin zone $[-\pi/a, \pi/a]$.

For a TM-polarized wave propagating along x (the stacking direction), the problem reduces to a 1D scalar equation. At normal incidence, where TE and TM have the same dispersion for non-magnetic layers, the standard approach is the transfer matrix [2]: across one unit cell, the field amplitudes $(E, dE/dx)$ are mapped by a 2×2 matrix M whose trace determines the dispersion relation:

$$\cos(ka) = \frac{1}{2}\operatorname{tr} M = \cos(k_A d_A)\cos(k_B d_B) - \frac{1}{2}\left(\frac{n_A}{n_B} + \frac{n_B}{n_A}\right)\sin(k_A d_A)\sin(k_B d_B), \tag{1.25}$$

where $k_j = n_j \omega / c$ and $n_j = \sqrt{\varepsilon_j}$ for $j \in \{A, B\}$.

1.6.2 Band gaps from the eigenvalue perspective

Equation (1.25) has solutions only when the right-hand side lies in the interval $[-1, 1]$. When it exceeds this range, there are no propagating Bloch modes: a *band gap* exists.

From the variational perspective, the band gap arises because no trial field in the periodic Hilbert space can simultaneously satisfy the curl (derivative) constraint and the ε_r-weighted energy condition in the gap frequency range. The dielectric contrast $\varepsilon_A/\varepsilon_B$ controls the gap width: larger contrast produces wider gaps.

Numerical illustration. Consider $\varepsilon_A = 13$ (GaAs-like) and $\varepsilon_B = 1$ (air), with $d_A = d_B = a/2$. The first band gap opens at the Brillouin zone boundary ($k = \pi/a$) between the normalized frequencies $\omega a/(2\pi c) \approx 0.16$ and $\omega a/(2\pi c) \approx 0.22$. The lower band-edge mode has its electric-field intensity concentrated in the high-ε layers (the "dielectric band"), while the upper band-edge mode is concentrated in the low-ε layers (the "air band"). This concentration is a direct consequence of the variational principle: the lowest mode minimizes the Rayleigh quotient by placing field energy where ε is largest.

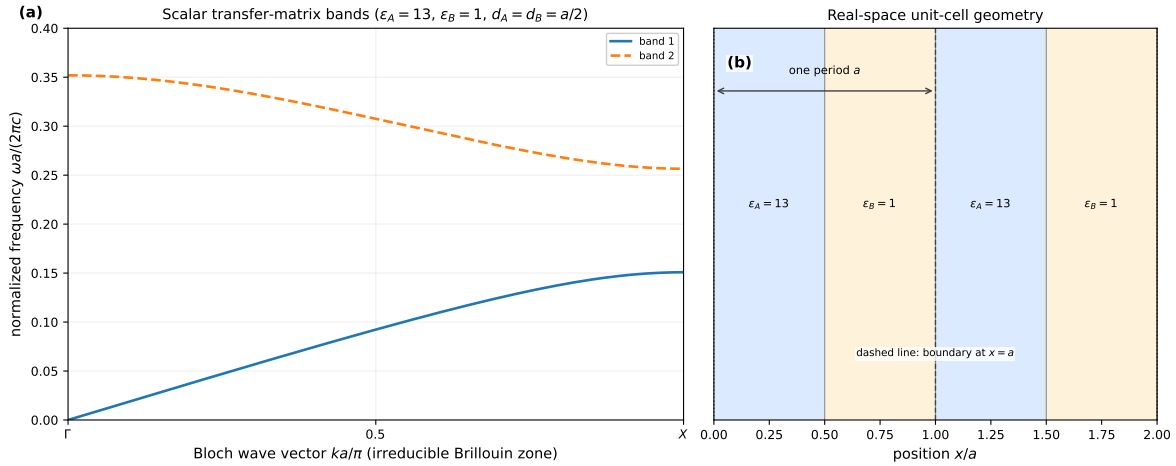


Figure 1.2 A one-dimensional Maxwell eigenproblem made visible. Panel (a) shows the Bloch band structure of a GaAs/air periodic stack computed from the transfer-matrix dispersion relation (1.25); shaded regions mark band gaps where no real Bloch wave vector exists. Panel (b) sketches how the lower band-edge mode concentrates in the high-permittivity layers while the upper band-edge mode concentrates in the air layers, illustrating the variational principle of Section 1.4.

1.6.3 What changes when ε_B becomes complex

If the air layers are replaced by a lossy material with $\varepsilon_B = 1 + i\gamma$ ($\gamma > 0$ for loss with our $e^{-i\omega t}$ convention), the transfer-matrix elements become complex. At a fixed real driving frequency, the Bloch wave vector k acquires an imaginary part even inside the former pass bands, so propagation is attenuated. Conversely, if one fixes a real Bloch wave vector and solves the eigenvalue problem, the band frequencies migrate into the complex-ω plane: every mode now decays in time.

If instead $\varepsilon_B = 1 - i\gamma$ (gain), the eigenfrequencies move into the upper half of the complex plane and modes grow in time. The balanced case $\varepsilon_A = \varepsilon_0 + i\gamma$, $\varepsilon_B = \varepsilon_0 - i\gamma$ is a $\mathcal{PT}$-symmetric photonic crystal (Chapter 9), where the interplay between gain and loss produces a rich spectral structure including exceptional points and spontaneous $\mathcal{PT}$-symmetry breaking.

This simple 1D example already contains, in embryo, most of the non-Hermitian phenomena explored in this book. The rest of the monograph develops the general theory, extends it to higher dimensions and more complex geometries, and connects it to experimental observations and device designs.

1.7 Spectrum types: discrete, continuous, and complex

The nature of the spectrum of $\hat{\Theta}$ depends on the geometry and boundary conditions. Understanding the taxonomy of spectra is essential because different types of spectra require different mathematical tools and lead to qualitatively different physics.

Closed cavities. A finite domain Ω with perfectly conducting walls supports a purely discrete spectrum: an infinite sequence of real eigenfrequencies $0 \le \omega_1 \le \omega_2 \le \cdots \to \infty$, each with finite multiplicity. The corresponding eigenmodes form a complete orthonormal set with respect to $\langle \cdot, \cdot \rangle$. This is the familiar setting of microwave cavities and closed optical resonators.

The completeness of the eigenmodes means that any transverse field in the cavity can be expanded as a convergent series in the eigenmodes:

$$\mathbf{H}(\mathbf{r}) = \sum_{n=1}^{\infty} c_n \, \mathbf{H}_n(\mathbf{r}), \qquad c_n = \langle \mathbf{H}_n, \, \mathbf{H} \rangle. \tag{1.26}$$

This expansion is the foundation of cavity QED, microwave filter design, and the coupled-mode theory of Chapter 7.

Periodic media. A medium with discrete translational symmetry along one or more directions admits Bloch's theorem (derived in Chapter 3): the eigenmodes are labeled by a continuous wave vector $\mathbf{k}$ within the Brillouin zone, and for each $\mathbf{k}$ there is a discrete set of band frequencies $\omega_n(\mathbf{k})$. The full spectrum is the union of these bands and may contain gaps—frequency intervals with no propagating modes. The band structure $\omega_n(\mathbf{k})$ is the central object of photonic-crystal physics.

Open systems. When Ω extends to infinity—or equivalently, when we impose outgoing-wave boundary conditions to model radiation loss—the spectrum changes qualitatively. Besides possible guided modes (discrete, real ω), there is a continuous spectrum of radiation modes, and there may be *resonances*: solutions with complex ω in the lower half of the complex frequency plane, describing fields that decay in time while growing spatially toward infinity. These are the quasinormal modes of Chapter 6. Their complex frequencies are the signature of non-Hermiticity induced by openness.

The completeness of the mode expansion (1.26) does not carry over straightforwardly to open systems. Quasinormal modes diverge at spatial infinity and cannot be normalized with the standard inner product. Chapter 6 develops the regularization and biorthogonal normalization that rescue the expansion, and Chapter 17 describes the numerical techniques (perfectly matched layers, contour-integral eigensolvers) that make it practical.

Lossy or active media. Even for a closed cavity, if ε_r has an imaginary part, the eigenvalue ω^2/c^2 becomes complex. For a medium with loss ($\mathrm{Im}\,\varepsilon_\mathrm{r} > 0$ with our $e^{-i\omega t}$ convention), ω acquires a negative imaginary part and modes decay; for a medium with gain ($\mathrm{Im}\,\varepsilon_\mathrm{r} < 0$), ω acquires a positive imaginary part and modes grow. The transition from a purely real spectrum to a complex spectrum is the spectral manifestation of non-Hermiticity.

1.8 Positive semi-definiteness and the photonic density of states

The operator $\hat{\Theta}$ on divergence-free fields with real, positive ε_r is positive semi-definite:

$$\langle \mathbf{H}, \, \hat{\Theta}\mathbf{H} \rangle = \int_\Omega \frac{1}{\varepsilon_\mathrm{r}(\mathbf{r})} \, |\nabla \times \mathbf{H}(\mathbf{r})|^2 \; \mathrm{d}^3 r \geq 0, \tag{1.27}$$

with equality if and only if $\nabla \times \mathbf{H} = 0$. After static harmonic fields are removed by the boundary conditions or projected out, this implies $\mathbf{H} = 0$ on the dynamical transverse subspace. The nonzero eigenvalues ω_n^2/c^2 are therefore positive, and the corresponding eigenfrequencies ω_n are real and positive.

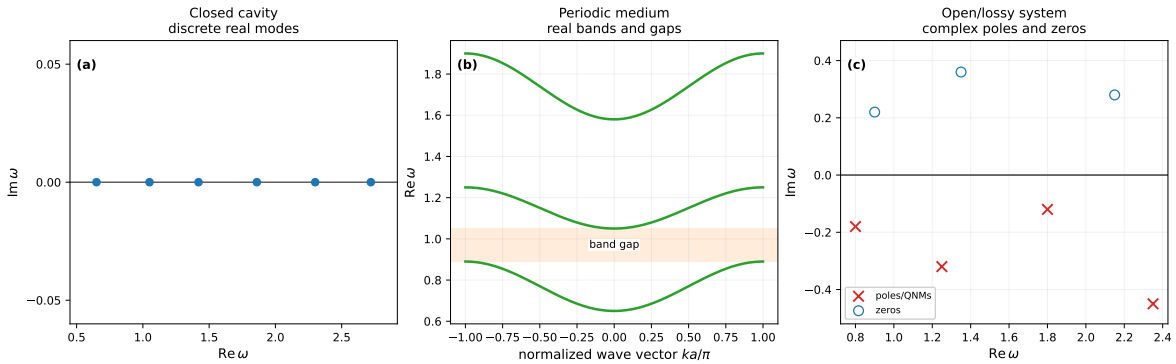


Figure 1.3 Three spectral regimes introduced in this opening chapter. A closed cavity has a discrete real spectrum; a periodic Hermitian medium has real bands and gaps; an open or lossy system is most naturally described by complex-frequency poles, zeros, and continua. The rest of the monograph develops the mathematical and physical consequences of moving from the first two panels to the third.

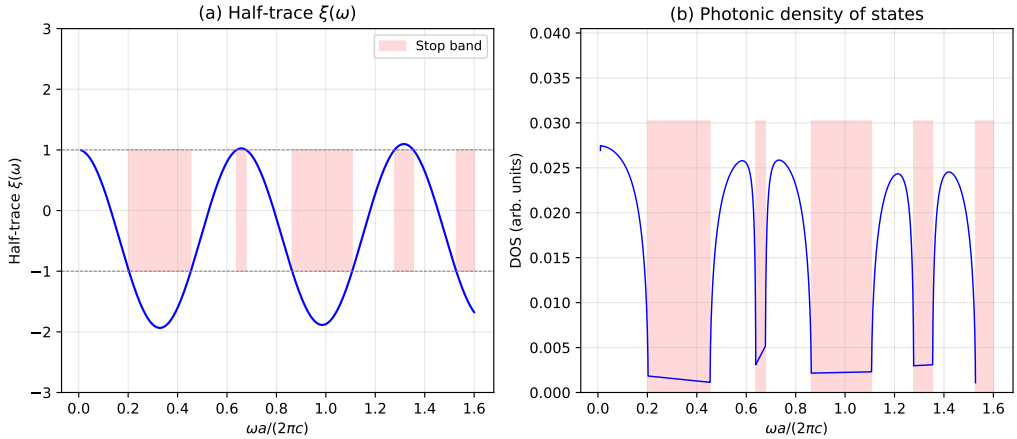


Figure 1.4 Photonic density of states for a one-dimensional periodic dielectric. (a) The transfer-matrix half-trace $\xi(\omega)$ divides pass bands ($|\xi| < 1$) from stop bands ($|\xi| > 1$). (b) The density of states vanishes in stop bands and develops van-Hove enhancements near band edges, illustrating how Maxwell eigenvalue spectra differ qualitatively from free-space mode counting.

The distribution of eigenfrequencies defines the *photonic density of states* (PDOS):

$$\rho(\omega) = \sum_n \delta(\omega - \omega_n), \tag{1.28}$$

which counts the number of modes per unit frequency interval. In a homogeneous 3D medium with permittivity ε, the PDOS is

$$\rho(\omega) = \frac{\varepsilon^{3/2}\,\omega^2}{\pi^2 c^3} \quad \text{(per unit volume)}, \tag{1.29}$$

which is the photonic analogue of Weyl's law for the Laplacian. The PDOS is modified by photonic crystals (band gaps produce zeros in the PDOS) and by cavities (the PDOS is enhanced at resonant frequencies), and these modifications are the basis for controlling spontaneous emission (Purcell effect), photon localization, and optical gain.

When ε_r becomes complex, the PDOS must be generalized to account for the finite linewidth of the modes. The local density of states (LDOS) at position $\mathbf{r}_0$ is related to the imaginary part

of the Green tensor:

$$\rho(\mathbf{r}_0,\omega) = \frac{2\omega}{\pi c^2}\,\mathrm{Im}\big[\mathrm{tr}\,\mathbf{G}(\mathbf{r}_0,\mathbf{r}_0;\omega)\big], \tag{1.30}$$

where $\mathbf{G}$ is the dyadic Green tensor of the Maxwell system. Chapter 6 shows how the Green tensor can be expanded in terms of quasinormal modes, providing a modal decomposition of the LDOS even in open, non-Hermitian systems.

1.9 What this book will build

The rest of this monograph develops the consequences of the spectral structure introduced here. Part I continues by establishing the energy inner products and reciprocity relations that underpin Hermitian photonics (Chapter 2), deriving the Bloch band structure of periodic media (Chapter 3), and introducing the scattering framework for open systems (Chapter 4).

Part II identifies the physical origins of non-Hermiticity: material loss and gain (Chapter 5), radiation-induced complex eigenfrequencies in quasinormal modes (Chapter 6), and the projection from the full Maxwell operator to finite-dimensional coupled-mode models (Chapter 7).

Part III examines the most striking consequences of non-Hermitian operators: exceptional points where eigenvectors coalesce (Chapter 8), $\mathcal{PT}$-symmetric optics (Chapter 9), and the interplay of scattering poles and zeros that connects lasers to coherent perfect absorbers (Chapter 10).

Part IV builds the topological theory [7, 8], starting from Berry curvature derived directly from Maxwell fields (Chapter 11), extending to complex bands and non-Hermitian gaps (Chapter 12), developing photonic point-gap topology (Chapter 13), the skin effect and non-Bloch band theory (Chapter 14), the reconstruction of bulk-boundary correspondence (Chapter 15), and the topology of bound states in the continuum (Chapter 16).

Part V addresses computation (Chapter 17), devices and physical limits (Chapter 18), and open problems (Chapter 19).

Throughout, we shall insist on a single discipline: every non-Hermitian structure must be traceable to a property of the Maxwell operator, not merely asserted by analogy with a tight-binding Hamiltonian. This Maxwell-first perspective is what gives the monograph its coherence and, we hope, its lasting value.

Chapter 2

Energy, Inner Products, and Hermiticity

Before we can say what non-Hermitian photonics is *[9, 10], we must understand what it* is not*. The entire machinery of lossless electromagnetism—real eigenfrequencies, orthogonal modes, energy conservation, reciprocity—rests on a single mathematical fact: the Maxwell curl–curl operator is self-adjoint with respect to an energy inner product. That inner product is derived from the Poynting theorem, leading to the self-adjointness conditions and catalogs the physical mechanisms that destroy them. Every later chapter of this book is, in one way or another, a study of the consequences of that destruction.*

2.1 Poynting's theorem

We begin, as always, from Maxwell's equations. Taking the dot product of Faraday's law (1.1) with $\mathbf{H}$ and of Ampère's law (1.2) with $\mathbf{E}$, and subtracting, one obtains the *Poynting identity* in its instantaneous, differential form:

$$\boxed{\nabla \cdot \mathbf{S} = -\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t},} \tag{2.1}$$

where the Poynting vector is

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}, \tag{2.2}$$

At this stage, (2.1) is only an identity: it follows from Maxwell's equations alone and makes no assumption about the constitutive relations. An energy density appears only after the material response is specified well enough that $\mathbf{E} \cdot \partial_t \mathbf{D} + \mathbf{H} \cdot \partial_t \mathbf{B}$ can be written as a total time derivative.

For a *nondispersive, lossless* medium with real, frequency-independent $\varepsilon_{\mathrm{r}}(\mathbf{r})$ and $\mu_{\mathrm{r}}(\mathbf{r})$, the right-hand side simplifies:

$$\mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} + \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} = \frac{\partial}{\partial t}\left[\frac{1}{2}\varepsilon_0 \varepsilon_{\mathrm{r}} |\mathbf{E}|^2 + \frac{1}{2}\mu_0 \mu_{\mathrm{r}} |\mathbf{H}|^2\right]. \tag{2.3}$$

The energy density is therefore

$$u(\mathbf{r},t) = \frac{1}{2}\varepsilon_0\varepsilon_{\mathrm{r}}(\mathbf{r})\,|\mathbf{E}(\mathbf{r},t)|^2 + \frac{1}{2}\mu_0\mu_{\mathrm{r}}(\mathbf{r})\,|\mathbf{H}(\mathbf{r},t)|^2, \tag{2.4}$$

and the Poynting theorem becomes a local conservation law:

$$\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S} = 0. \tag{2.5}$$

Integrating over a volume Ω bounded by a surface $\partial\Omega$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u\ \mathrm{d}^3r = -\oint_{\partial\Omega}\mathbf{S}\cdot\hat{\mathbf{n}}\ \mathrm{d}S. \tag{2.6}$$

The total electromagnetic energy in Ω changes only through the Poynting flux through its boundary. No energy is created or destroyed inside the volume. This is the physical content of "lossless": energy is a conserved quantity.

Example: energy balance in a lossy dielectric slab. Consider a plane wave $\mathbf{E} = E_0\,e^{i(kz-\omega t)}\hat{\mathbf{x}}$ propagating in a slab of thickness d with complex permittivity $\varepsilon_{\mathrm{r}} = \varepsilon_{\mathrm{r}}' + i\varepsilon_{\mathrm{r}}''$. To isolate absorption from interface reflection, take E_0 below to denote the forward field just inside an impedance-matched slab. The time-averaged Poynting flux entering the material is then $S_{\mathrm{in}} = |E_0|^2/(2\eta)$, where $\eta = \sqrt{\mu_0/(\varepsilon_0\varepsilon_{\mathrm{r}}')}$ in the weak-loss limit. Inside the slab, the field decays approximately as $|\mathbf{E}(z)|^2 = |E_0|^2 e^{-2\alpha z}$, where the attenuation constant is $\alpha = (\omega/c)\,\mathrm{Im}\,\sqrt{\varepsilon_{\mathrm{r}}}$. The time-averaged dissipated power per unit area is

$$P_{\mathrm{diss}} = \frac{\omega\varepsilon_0\varepsilon_{\mathrm{r}}''}{2}\int_0^d |E_0|^2 e^{-2\alpha z}\ \mathrm{d}z = \frac{\omega\varepsilon_0\varepsilon_{\mathrm{r}}''\,|E_0|^2}{4\alpha}\big(1-e^{-2\alpha d}\big). \tag{2.7}$$

The Poynting flux exiting the slab is $S_{\mathrm{out}} = S_{\mathrm{in}}\,e^{-2\alpha d}$ in the same weak-loss description. Equivalently, the local identity $-\mathrm{d}S/\mathrm{d}z = (\omega\varepsilon_0\varepsilon_{\mathrm{r}}''/2)|E(z)|^2$ gives $S_{\mathrm{in}} - S_{\mathrm{out}} = P_{\mathrm{diss}}$: the energy lost from the wave equals the energy absorbed by the medium. With interfaces included, reflected fluxes must also be counted, but the same Poynting balance holds exactly.

When $\varepsilon_{\mathrm{r}}'' < 0$ (gain medium), $\alpha < 0$ and the field grows exponentially. The "dissipated" power is negative—the medium delivers energy to the wave. The Poynting flux *increases* across the slab, and $S_{\mathrm{out}} > S_{\mathrm{in}}$. This simple example illustrates the core physical mechanism behind all non-Hermitian photonic phenomena: the energy inner product is no longer conserved when $\varepsilon_{\mathrm{r}}'' \neq 0$.

Figure 2.1 illustrates this energy balance for two independent slabs.

Units and conventions

SI units are used throughout. The vacuum permittivity ε_0 and permeability μ_0 appear explicitly, so that ε_{r} and μ_{r} are dimensionless. The time convention is $e^{-i\omega t}$ for harmonic fields. With this convention, a passive lossy medium has $\mathrm{Im}\,\varepsilon_{\mathrm{r}}(\omega) > 0$.

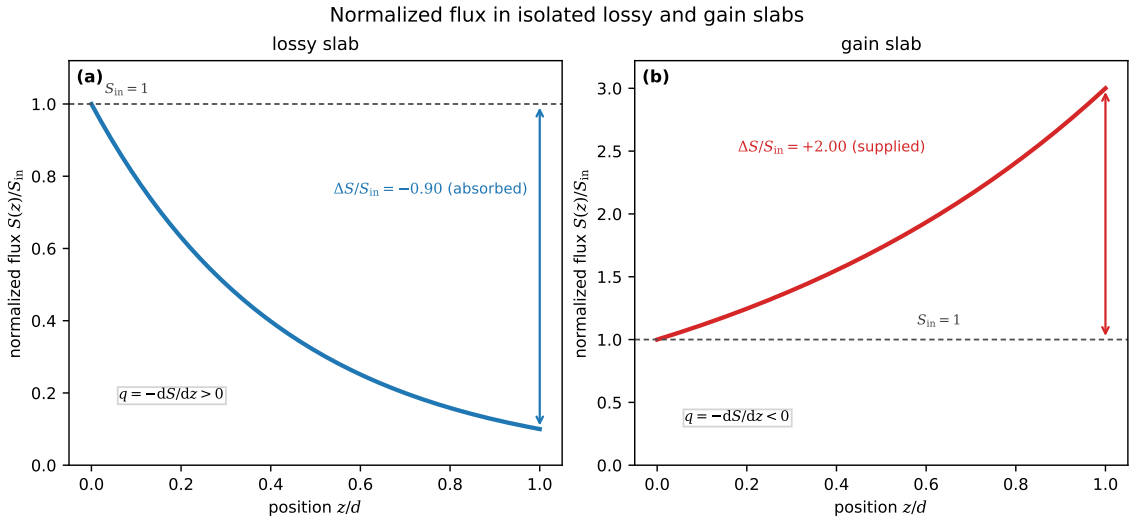


Figure 2.1 Normalised Poynting flux $S(z)/S_{\text{in}}$ through two independent dielectric slabs on separate vertical scales. (a) Lossy slab ($\varepsilon_{\text{r}}'' > 0$): the flux decays exponentially from $S_{\text{in}} = 1$ to $S_{\text{out}} = 0.10$, so the medium absorbs $\Delta S/S_{\text{in}} = 0.90$. The sign convention $q = -\mathrm{d}S/\mathrm{d}z > 0$ confirms that power density flows from the field into the material. (b) Gain slab ($\varepsilon_{\text{r}}'' < 0$): the flux grows to $S_{\text{out}} = 3.00$, so the medium supplies $\Delta S/S_{\text{in}} = 2.00$ to the wave, with $q = -\mathrm{d}S/\mathrm{d}z < 0$. In both panels, the dashed line marks $S_{\text{in}} = 1$ and the arrow at $z = d$ indicates the total flux change. The energy balance $S_{\text{in}} - S_{\text{out}} = \int_0^d q\,\mathrm{d}z$ is exact and follows directly from Poynting's theorem (2.1).

2.2 Time-averaged energy and the harmonic inner product

For monochromatic fields with harmonic time dependence $\mathbf{E}(\mathbf{r},t) = \mathrm{Re}[\mathbf{E}(\mathbf{r})\,e^{-i\omega t}]$, the time-averaged energy density is

$$\langle u \rangle = \frac{1}{4}\varepsilon_0 \varepsilon_{\text{r}}(\mathbf{r})\,|\mathbf{E}(\mathbf{r})|^2 + \frac{1}{4}\mu_0 \mu_{\text{r}}(\mathbf{r})\,|\mathbf{H}(\mathbf{r})|^2, \tag{2.8}$$

and the time-averaged Poynting vector is

$$\langle \mathbf{S} \rangle = \frac{1}{2}\,\mathrm{Re}\big[\mathbf{E}(\mathbf{r}) \times \mathbf{H}^*(\mathbf{r})\big]. \tag{2.9}$$

These expressions are valid only when ε_{r} and μ_{r} are real and frequency-independent. For dispersive media, additional terms involving $\mathrm{d}\varepsilon_{\text{r}}/\,\mathrm{d}\omega$ appear (the Brillouin formula), and for lossy media, the time-averaged "energy density" is no longer uniquely defined from the macroscopic fields alone—a point we shall return to in Chapter 5.

The positive-definiteness of $\langle u \rangle$ when $\varepsilon_{\text{r}} > 0$ and $\mu_{\text{r}} > 0$ is crucial. It means that the electromagnetic energy defines a norm on the space of field configurations, and hence an inner product.

2.3 Inner products for the Maxwell eigenproblem

The eigenvalue problems derived in Chapter 1 have natural inner products dictated by the energy density.

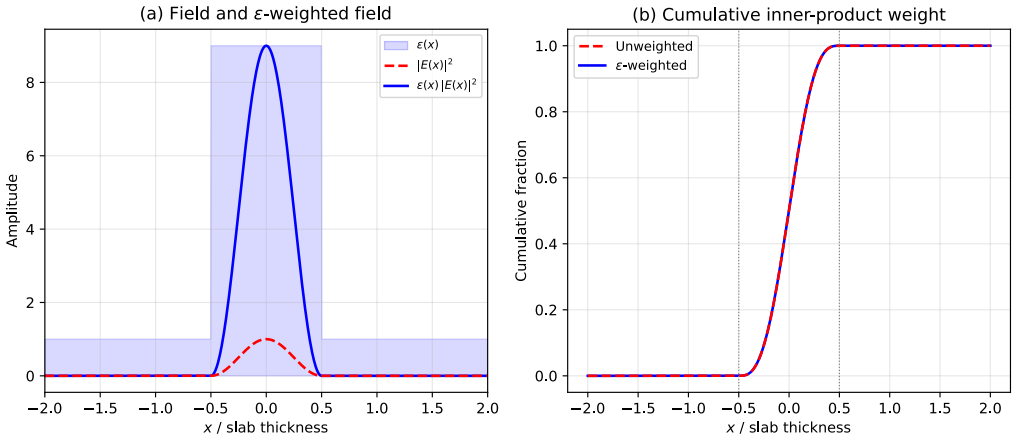


Figure 2.2 Visualisation of the ε-weighted inner product for a dielectric slab ($\varepsilon = 9$, shaded region). (a) The bare field intensity $|E(x)|^2$ (red) and the ε-weighted intensity $\varepsilon(x)\,|E(x)|^2$ (blue). The weighting amplifies the contribution from the high-index region, reflecting the physical energy density $\frac{1}{2}\varepsilon_0\varepsilon|E|^2$. (b) Cumulative inner-product weight: the ε-weighted curve (solid blue) rises steeply inside the slab, showing that most of the modal energy resides in the high-index material.

2.3.1 The H-field inner product

For the operator $\hat{\Theta}$ defined in (1.11), the natural inner product is

$$\langle \mathbf{H}_1,\, \mathbf{H}_2 \rangle = \int_\Omega \mathbf{H}_1^*(\mathbf{r}) \cdot \mathbf{H}_2(\mathbf{r})\;\mathrm{d}^3 r. \tag{2.10}$$

This is the standard L^2 inner product, unweighted, because the eigenvalue equation $\hat{\Theta}\mathbf{H} = (\omega^2/c^2)\mathbf{H}$ has the eigenvalue on the right without a material weight (assuming $\mu_\mathrm{r} = 1$).

2.3.2 The E-field inner product

For the $\mathbf{E}$-field equation (1.9), the eigenvalue appears multiplied by $\varepsilon_\mathrm{r}(\mathbf{r})$, so the correct inner product is the ε_r-weighted one introduced in (1.14):

$$\langle \mathbf{E}_1,\, \mathbf{E}_2 \rangle_\varepsilon = \int_\Omega \mathbf{E}_1^*(\mathbf{r}) \cdot \varepsilon_\mathrm{r}(\mathbf{r})\,\mathbf{E}_2(\mathbf{r})\;\mathrm{d}^3 r. \tag{2.11}$$

This inner product is positive-definite precisely when $\varepsilon_\mathrm{r}(\mathbf{r}) > 0$ for scalar media, or when the permittivity tensor is Hermitian positive definite in anisotropic media.

2.3.3 The first-order Maxwell inner product

There is a third, often more natural, way to organize the Maxwell eigenproblem. Writing the six-component field vector $\mathbf{\Psi} = (\mathbf{E},\, \mathbf{H})^\mathsf{T}$, the frequency-domain Maxwell equations (1.7)–(1.8) become a first-order system

$$\underbrace{\begin{pmatrix} 0 & i\nabla\times \\ -i\nabla\times & 0 \end{pmatrix}}_{\hat{\mathcal{M}}} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \omega \underbrace{\begin{pmatrix} \varepsilon_0\varepsilon_\mathrm{r} & 0 \\ 0 & \mu_0\mu_\mathrm{r} \end{pmatrix}}_{\hat{W}} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}. \tag{2.12}$$

The operator $\hat{\mathcal{M}}$ is manifestly Hermitian with respect to the unweighted L^2 inner product (its off-diagonal blocks are related by the curl identity (1.15)). The generalized eigenvalue problem $\hat{\mathcal{M}}\boldsymbol{\Psi} = \omega\,\hat{W}\boldsymbol{\Psi}$ is self-adjoint with respect to the energy inner product

$$\langle \boldsymbol{\Psi}_1,\, \boldsymbol{\Psi}_2\rangle_W = \int_\Omega \boldsymbol{\Psi}_1^\dagger\, \hat{W}\, \boldsymbol{\Psi}_2 \;\mathrm{d}^3r = \int_\Omega \left[\varepsilon_0\varepsilon_\mathrm{r}\, \mathbf{E}_1^* \cdot \mathbf{E}_2 + \mu_0\mu_\mathrm{r}\, \mathbf{H}_1^* \cdot \mathbf{H}_2\right] \mathrm{d}^3r, \tag{2.13}$$

provided ε_r and μ_r are real and positive and the boundary conditions make the surface terms vanish. This inner product is proportional to four times the time-averaged electromagnetic energy when $\boldsymbol{\Psi}_1 = \boldsymbol{\Psi}_2$.

Comparing the three inner products on a concrete mode pair. To see how the three inner products differ in practice, consider the fundamental TE_{10} and TE_{01} modes of a rectangular metallic cavity ($a \times b \times d$) filled with a uniform dielectric $\varepsilon_\mathrm{r} = \text{const}$, $\mu_\mathrm{r} = 1$. The $\mathbf{H}$-fields of these two modes are $\mathbf{H}_1 = H_0 \cos(\pi x/a)\sin(\pi z/d)\,\hat{\mathbf{y}}$ and $\mathbf{H}_2 = H_0 \sin(\pi y/b)\cos(\pi z/d)\,\hat{\mathbf{x}}$ (simplified for illustration). The three inner products give:

1. $\langle \mathbf{H}_1, \mathbf{H}_2\rangle = \int \mathbf{H}_1^* \cdot \mathbf{H}_2 \;\mathrm{d}^3r = 0$, because $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$ (orthogonal polarizations).
2. $\langle \mathbf{E}_1, \mathbf{E}_2\rangle_\varepsilon = \int \varepsilon_\mathrm{r}\, \mathbf{E}_1^* \cdot \mathbf{E}_2 \;\mathrm{d}^3r = 0$, because the $\mathbf{E}$-fields of the two modes have different spatial symmetries under the cavity's C_{2v} group.
3. $\langle \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2\rangle_W = \int (\varepsilon_0\varepsilon_\mathrm{r}\, \mathbf{E}_1^* \cdot \mathbf{E}_2 + \mu_0\, \mathbf{H}_1^* \cdot \mathbf{H}_2) \;\mathrm{d}^3r = 0$, as both terms vanish independently.

All three inner products agree (zero) for this mode pair in the Hermitian case. The distinction matters only when ε_r is complex or when the modes are quasinormal: then the $\mathbf{H}$-field inner product stays unchanged (it does not involve ε_r), the $\mathbf{E}$-field inner product picks up an imaginary contribution from ε_r'', and the first-order inner product mixes both. In non-Hermitian photonics [11], the choice of inner product is not a convention but a physical decision that determines orthogonality, completeness, and the meaning of "energy."

Energy as inner product

The electromagnetic energy density defines the inner product that makes the Maxwell operator self-adjoint. Making ε_r complex, opening the boundary, or projecting onto a subspace changes this inner product and therefore changes the adjointness properties of the problem.

2.4 Self-adjointness: the complete argument

We now give a careful derivation of the self-adjointness of $\hat{\Theta}$ in the $\mathbf{H}$-field formulation (1.11), which was stated without full proof in Section 1.3.

Theorem 2.1 (Self-adjointness of $\hat{\Theta}$)**.** *Let Ω be a domain in $\mathbb{R}^3$ with piecewise smooth boundary $\partial\Omega$. Let $\varepsilon_\mathrm{r}(\mathbf{r})$ be real, positive, and bounded away from zero, and let $\mu_\mathrm{r} = 1$. Define $\hat{\Theta}\mathbf{H} = \nabla \times [\varepsilon_\mathrm{r}^{-1}\nabla \times \mathbf{H}]$. Then $\hat{\Theta}$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle$ on the space of divergence-free fields satisfying either (a) $\hat{\mathbf{n}} \times \mathbf{E} = 0$ on $\partial\Omega$ (perfect electric conductor), or (b) Bloch-periodic boundary conditions.*

Proof. Consider two fields $\mathbf{H}_1$ and $\mathbf{H}_2$ in the domain of $\hat{\Theta}$. We compute

$$\langle \mathbf{H}_1, \hat{\Theta}\mathbf{H}_2 \rangle = \int_\Omega \mathbf{H}_1^* \cdot \nabla \times \left[\varepsilon_\mathrm{r}^{-1} \nabla \times \mathbf{H}_2\right] \, \mathrm{d}^3 r. \tag{2.14}$$

Using the vector identity $\mathbf{A}\cdot(\nabla\times\mathbf{B}) = (\nabla\times\mathbf{A})\cdot\mathbf{B} + \nabla\cdot(\mathbf{B}\times\mathbf{A})$ with $\mathbf{A} = \mathbf{H}_1^*$ and $\mathbf{B} = \varepsilon_\mathrm{r}^{-1}\nabla\times\mathbf{H}_2$, we obtain

$$\langle \mathbf{H}_1, \hat{\Theta}\mathbf{H}_2 \rangle = \int_\Omega \frac{1}{\varepsilon_\mathrm{r}} (\nabla \times \mathbf{H}_1)^* \cdot (\nabla \times \mathbf{H}_2) \, \mathrm{d}^3 r + \oint_{\partial\Omega} \left(\varepsilon_\mathrm{r}^{-1} \nabla \times \mathbf{H}_2 \times \mathbf{H}_1^*\right) \cdot \hat{\mathbf{n}} \, \mathrm{d}S. \tag{2.15}$$

The bulk integral on the right is manifestly symmetric under the exchange $1 \leftrightarrow 2$ (with complex conjugation), because $\varepsilon_\mathrm{r}^{-1}$ is real. The surface integral vanishes under either boundary condition (a) or (b).

For condition (a), note that $\varepsilon_\mathrm{r}^{-1}\nabla\times\mathbf{H} = (i\omega\varepsilon_0)^{-1}\mathbf{E}$ from Ampère's law, so $\hat{\mathbf{n}} \times (\varepsilon_\mathrm{r}^{-1}\nabla\times\mathbf{H}) \propto \hat{\mathbf{n}} \times \mathbf{E} = 0$ on $\partial\Omega$.

For condition (b), the integrand on opposite faces of the unit cell differs by Bloch phases $e^{i\mathbf{k}\cdot\mathbf{a}}$ and $e^{-i\mathbf{k}\cdot\mathbf{a}}$, while the outward normals are opposite; the contributions cancel.

Since the bulk term is symmetric and the surface term vanishes,

$$\langle \mathbf{H}_1, \hat{\Theta}\mathbf{H}_2 \rangle = \langle \hat{\Theta}\mathbf{H}_1, \mathbf{H}_2 \rangle, \tag{2.16}$$

completing the proof. □

The same argument can be repeated for the $\mathbf{E}$-field operator with the ε_r-weighted inner product, yielding $\langle \mathbf{E}_1, \hat{\Theta}_E\mathbf{E}_2 \rangle_\varepsilon = \langle \hat{\Theta}_E\mathbf{E}_1, \mathbf{E}_2 \rangle_\varepsilon$ under the same boundary conditions.

Remark 2.1. The assumption that ε_r is real enters at a precise point: the symmetry of the bulk integral in (2.15). If $\varepsilon_\mathrm{r} = \varepsilon_\mathrm{r}' + i\varepsilon_\mathrm{r}''$ with $\varepsilon_\mathrm{r}'' \neq 0$, then $1/\varepsilon_\mathrm{r}^* \neq 1/\varepsilon_\mathrm{r}$, the bulk term is no longer symmetric, and self-adjointness fails. This is the mathematical origin of non-Hermiticity from material loss or gain.

2.5 Consequences of self-adjointness

When $\hat{\Theta}$ is self-adjoint, the standard theorems of spectral theory for compact self-adjoint operators in a Hilbert space apply. The consequences are profound and familiar, but it is worth stating them explicitly, because each one can fail in non-Hermitian photonics.

2.5.1 Real eigenvalues

If $\hat{\Theta}\mathbf{H}_n = \lambda_n\mathbf{H}_n$ with $\lambda_n = \omega_n^2/c^2$, then

$$\lambda_n = \frac{\langle \mathbf{H}_n, \hat{\Theta}\mathbf{H}_n \rangle}{\langle \mathbf{H}_n, \mathbf{H}_n \rangle} = \frac{\displaystyle\int_\Omega \varepsilon_\mathrm{r}^{-1} |\nabla \times \mathbf{H}_n|^2 \, \mathrm{d}^3 r}{\displaystyle\int_\Omega |\mathbf{H}_n|^2 \, \mathrm{d}^3 r} \geq 0. \tag{2.17}$$

The eigenvalue λ_n is real and non-negative: the eigenfrequencies ω_n are real. There is no temporal decay, no growth, no complex frequency.

2.5.2 Orthogonality

Eigenmodes belonging to distinct eigenvalues are orthogonal:

$$\omega_m \neq \omega_n \quad \Longrightarrow \quad \langle \mathbf{H}_m,\, \mathbf{H}_n \rangle = 0. \tag{2.18}$$

For the **E**-field formulation, the orthogonality relation is $\langle \mathbf{E}_m,\, \mathbf{E}_n \rangle_\varepsilon = 0$.

In either case, we may normalize the modes so that $\langle \mathbf{H}_m,\, \mathbf{H}_n \rangle = \delta_{mn}$ or $\langle \mathbf{E}_m,\, \mathbf{E}_n \rangle_\varepsilon = \delta_{mn}$. Mode decompositions and projections are then uniquely defined.

2.5.3 Completeness

For a compact domain (closed cavity), the eigenmodes form a complete orthonormal basis. Any field satisfying the boundary conditions can be expanded:

$$\mathbf{H}(\mathbf{r}) = \sum_n c_n\, \mathbf{H}_n(\mathbf{r}), \qquad c_n = \langle \mathbf{H}_n,\, \mathbf{H} \rangle. \tag{2.19}$$

The expansion coefficients c_n are uniquely determined by the inner product, and Parseval's identity $\|\mathbf{H}\|^2 = \sum_n |c_n|^2$ holds.

2.5.4 Variational characterization

The eigenfrequencies can be obtained from the Rayleigh quotient:

$$\omega_1^2/c^2 = \min_{\mathbf{H} \neq 0} \frac{\displaystyle\int \varepsilon_{\mathrm{r}}^{-1} |\nabla \times \mathbf{H}|^2 \ \mathrm{d}^3 r}{\displaystyle\int |\mathbf{H}|^2 \ \mathrm{d}^3 r}, \tag{2.20}$$

with higher eigenvalues given by the min–max principle. This characterization is the basis of numerical methods such as the plane-wave expansion and finite-element eigensolvers used in photonic-crystal computations [12].

2.6 Lorentz reciprocity

Self-adjointness (Hermiticity) is not the only symmetry of the Maxwell operator. There is a distinct and equally important property: *reciprocity*.

Consider two source–field pairs $(\mathbf{J}_a, \mathbf{E}_a, \mathbf{H}_a)$ and $(\mathbf{J}_b, \mathbf{E}_b, \mathbf{H}_b)$ at the same frequency ω in a medium with $\varepsilon_{\mathrm{r}}(\mathbf{r})$ and $\mu_{\mathrm{r}}(\mathbf{r})$ that are symmetric tensors (which includes all isotropic and many anisotropic media). The Lorentz reciprocity theorem states

$$\int_\Omega \left(\mathbf{E}_a \cdot \mathbf{J}_b - \mathbf{E}_b \cdot \mathbf{J}_a \right) \ \mathrm{d}^3 r = \oint_{\partial\Omega} \left(\mathbf{E}_a \times \mathbf{H}_b - \mathbf{E}_b \times \mathbf{H}_a \right) \cdot \hat{\mathbf{n}} \ \mathrm{d}S. \tag{2.21}$$

If the surface integral vanishes (closed cavity, or sources far from $\partial\Omega$), this reduces to the familiar statement $\langle \mathbf{E}_a, \mathbf{J}_b \rangle = \langle \mathbf{E}_b, \mathbf{J}_a \rangle$: swapping source and observation gives the same field projection.

2.6.1 Reciprocity and the scattering matrix

Reciprocity has a direct and powerful consequence for the scattering matrix $S(\omega)$ of Chapter 4. If the medium is reciprocal (symmetric ε_r and μ_r tensors), then the S-matrix satisfies

$$S^{\mathsf{T}}(\omega) = S(\omega), \tag{2.22}$$

i.e., S is a symmetric matrix. This means that the transmission from port i to port j equals the transmission from port j to port i: $S_{ij} = S_{ji}$.

The proof follows from the Lorentz reciprocity theorem (2.21). Let $\mathbf{E}_a$ be the field produced by a source at port i and $\mathbf{E}_b$ the field from a source at port j. If the surface integral vanishes (all ports are included in the integration surface), then $\langle \mathbf{E}_a, \mathbf{J}_b \rangle = \langle \mathbf{E}_b, \mathbf{J}_a \rangle$, which translates directly into $S_{ij} = S_{ji}$ when expressed in the scattering-matrix language.

This symmetry is *not* the same as unitarity. A unitary S-matrix ($S^\dagger S = I$) expresses energy conservation and requires Hermiticity; a symmetric S-matrix ($S^{\mathsf{T}} = S$) expresses reciprocity and requires only symmetric material tensors. A lossy dielectric cavity has $S^{\mathsf{T}} = S$ (reciprocal) but $S^\dagger S \neq I$ (energy is absorbed). A magneto-optic material breaks the tensor symmetry and therefore breaks both reciprocity and S-matrix symmetry, even if the system is lossless.

The distinction between reciprocity (symmetric S) and Hermiticity (unitary S) is the scattering-theory counterpart of the distinction between the unconjugated bilinear form and the conjugate inner product in the operator formulation.

Reciprocity versus Hermiticity

Reciprocity and Hermiticity are conceptually distinct. *Hermiticity* involves the complex-conjugate inner product $\langle \mathbf{F}^*, \hat{\Theta}\mathbf{G} \rangle$ and requires ε_r to be *real*. *Reciprocity* involves an unconjugated bilinear form $(\mathbf{F}, \sigma\mathbf{G})$ and requires ε_r to be a *symmetric tensor*. A medium with complex $\varepsilon_r = \varepsilon_r' + i\varepsilon_r''$ that is scalar (or symmetric) is *reciprocal but not Hermitian*. This distinction is central to non-Hermitian photonics: many systems of interest—lossy dielectrics, gain media, $\mathcal{PT}$-symmetric structures—are reciprocal, even though they are not Hermitian. The reaction-based inner product exploited in coupled-mode theory [13] is built on reciprocity, not Hermiticity.

2.7 Four roads to non-Hermiticity

The self-adjointness proof of Section 2.4 relied on three assumptions: (i) ε_r real, (ii) boundary terms vanish, (iii) fields live in a complete Hilbert space. The Rayleigh quotient argument (2.17) additionally used $\varepsilon_r > 0$. Violating any of these opens a distinct road to non-Hermitian physics. Figure 2.3 summarises the logical structure.

2.7.1 Material loss and gain

If the permittivity has an imaginary part, $\varepsilon_r = \varepsilon_r' + i\varepsilon_r''$, the factor ε_r^{-1} in (2.15) is no longer real, and the bulk integral loses its symmetry. For monochromatic fields in a nondispersive medium,

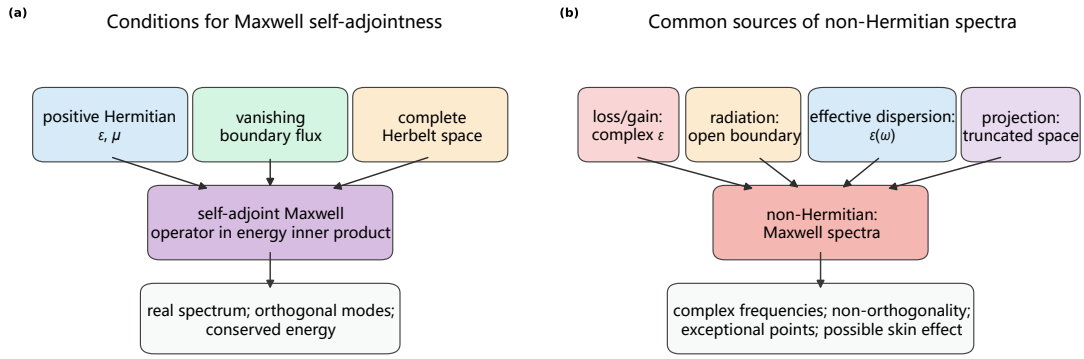


Figure 2.3 Conditions for Maxwell self-adjointness and the four common mechanisms that break them. (a) Three sufficient conditions—positive Hermitian constitutive tensors ε, μ, vanishing boundary flux (closed domain), and a complete Hilbert space—ensure that the Maxwell curl-curl operator is self-adjoint in the energy inner product, yielding a real spectrum, orthogonal modes, and conserved electromagnetic energy. (b) Each condition can be violated independently: material loss or gain introduces complex ε; radiation through open boundaries allows energy to leave the domain; frequency-dependent $\varepsilon(\omega)$ makes the operator effectively non-self-adjoint at any fixed real frequency; and projection onto a finite mode subspace discards completeness. The resulting non-Hermitian spectra exhibit complex eigenfrequencies, non-orthogonal modes, exceptional points, and possible skin-effect localisation—the subjects of Parts II–IV of this monograph.

the time-averaged Poynting identity becomes

$$\nabla \cdot \langle \mathbf{S} \rangle = -\frac{\omega}{2} \varepsilon_0 \varepsilon_{\mathrm{r}}''(\mathbf{r}) \, |\mathbf{E}(\mathbf{r})|^2. \tag{2.23}$$

For $\varepsilon_{\mathrm{r}}'' > 0$ (our $e^{-i\omega t}$ convention), energy is absorbed; for $\varepsilon_{\mathrm{r}}'' < 0$, energy is emitted. The eigenfrequencies become complex, and modes are no longer orthogonal with respect to $\langle \cdot, \cdot \rangle_\varepsilon$.

Physical example: coupled whispering-gallery resonators. Two silica microsphere resonators (diameter $\sim 50\ \mu$m) coupled evanescently through a sub-wavelength air gap provide a clean realisation. When both spheres are passive (intrinsic material loss only, $\varepsilon_{\mathrm{r}}'' \sim 10^{-7}$), the system is nearly Hermitian: the two whispering-gallery modes split into symmetric and antisymmetric supermodes with real frequency splitting 2κ. When one sphere is doped with an erbium gain medium pumped above threshold ($\varepsilon_{\mathrm{r}}'' < 0$), the system becomes genuinely non-Hermitian: the two supermodes acquire different imaginary parts, and at a critical pump power the frequency splitting vanishes at an exceptional point. The transition from the Hermitian regime ($\varepsilon_{\mathrm{r}}'' = 0$) to the EP ($|\varepsilon_{\mathrm{r}}''|$ reaches the critical value) is a continuous deformation of the eigenvalue structure that can be tracked experimentally through the transmission spectrum.

This mechanism is explored in Chapter 5 and provides the physical basis for $\mathcal{PT}$-symmetric optics (Chapter 9).

2.7.2 Radiation and open boundaries

For an open system, the outgoing-wave (Sommerfeld) boundary condition replaces the standing-wave condition of a closed cavity. The surface integral in (2.15) no longer vanishes: energy flows

outward through $\partial\Omega$. As a result, eigenfrequencies acquire negative imaginary parts (fields decay in time while growing spatially toward infinity). These are the quasinormal modes of Chapter 6.

Physical example: a dielectric microsphere in vacuum. A dielectric sphere of radius R and $\varepsilon_{\mathrm{r}} = 4$ (e.g., a glass microsphere) supports whispering-gallery modes with Q factors that increase exponentially with the angular momentum ℓ. Each mode has a complex eigenfrequency $\omega_\ell = \omega'_\ell - i\gamma_\ell/2$, where γ_ℓ is the radiative decay rate. Even though the material is perfectly lossless ($\varepsilon''_{\mathrm{r}} = 0$), the eigenfrequencies are complex because the outgoing-wave boundary condition at $r \to \infty$ breaks the self-adjointness of the Maxwell operator. The non-Hermiticity is purely geometrical: it arises from the open boundary, not from any material dissipation.

2.7.3 Frequency dispersion

Even for a closed, passive medium, if $\varepsilon_{\mathrm{r}}(\omega)$ depends on frequency, the eigenproblem is nonlinear in ω and the simple inner product (2.11) is no longer the correct energy functional. The Brillouin formula

$$\langle u \rangle_{\mathrm{disp}} = \frac{1}{4}\varepsilon_0 \frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega\,\varepsilon'_{\mathrm{r}}(\omega)\right] |\mathbf{E}|^2 + \frac{1}{4}\mu_0 \frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega\,\mu'_{\mathrm{r}}(\omega)\right] |\mathbf{H}|^2 \tag{2.24}$$

gives a partial answer for weakly dispersive, low-loss media, but it is not a true inner product: it is not bilinear, it depends on ω, and it becomes negative near absorption resonances. A rigorous treatment requires auxiliary fields that restore the Hamiltonian structure of the dispersive Maxwell system [1]. We defer this construction to Chapter 5, where it will be developed in detail.

Example: why the naive inner product fails for a Lorentz medium. Consider a single Lorentz oscillator with permittivity

$$\varepsilon_{\mathrm{r}}(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega}, \tag{2.25}$$

where ω_0 is the resonance frequency, ω_p is the plasma frequency, and γ is the damping rate. At a frequency ω well below the resonance ($\omega \ll \omega_0$), the permittivity is approximately real and the naive inner product $\langle \mathbf{F}, \mathbf{G} \rangle = \int \varepsilon_{\mathrm{r}}\, \mathbf{F}^* \cdot \mathbf{G}\ \mathrm{d}^3 r$ is positive-definite. The Brillouin energy density (2.24) reduces to $\langle u \rangle_{\mathrm{disp}} \approx \frac{1}{4}\varepsilon_0\,(1+\omega_p^2/\omega_0^2)\,|\mathbf{E}|^2$, which is the expected electrostatic energy enhanced by the polarizability.

Near the resonance ($\omega \approx \omega_0$), the situation changes dramatically. The real part $\varepsilon'_{\mathrm{r}}(\omega)$ passes through zero and becomes negative, while the imaginary part $\varepsilon''_{\mathrm{r}}(\omega)$ peaks at ω_0. The naive inner product is no longer positive-definite: $\varepsilon'_{\mathrm{r}}|\mathbf{E}|^2$ can be negative, and the "norm" of a mode can vanish or change sign. This is not a mathematical pathology but a physical fact: near a material resonance, the electromagnetic energy is stored partly in the kinetic energy of the polarization current, not only in the field. The Brillouin formula accounts for this by including the derivative $\partial(\omega\varepsilon'_{\mathrm{r}})/\partial\omega$:

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega\,\varepsilon'_{\mathrm{r}}(\omega)\right] = 1 + \omega_p^2\,\frac{\omega_0^2 + \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2}, \tag{2.26}$$

which is positive for all ω when γ is small (the no-gain condition). The Brillouin energy thus remains positive even where $\varepsilon'_{\mathrm{r}}$ itself is negative—the "missing" energy is stored in the oscillating polarization.

However, for γ comparable to ω_0 (strongly damped media), the Brillouin formula itself becomes unreliable, and one must resort to the auxiliary-field Hamiltonian embedding of Chapter 5, where the polarization degrees of freedom are treated as dynamical variables on the same footing as the electromagnetic field. This construction restores a true inner product at the cost of enlarging the Hilbert space, and it is the key technical step that makes the non-Hermitian Maxwell eigenproblem well-posed for dispersive media.

Comparing inner products: lossless versus dispersive. The following table summarises the three regimes:

Regime	Weight	Positive-definite?
Non-dispersive (ε_r real, constant)	$\varepsilon_r \lvert\mathbf{E}\rvert^2$	Yes
Weakly dispersive ($\gamma \ll \omega_0$)	$\partial_\omega(\omega\varepsilon_r') \lvert\mathbf{E}\rvert^2$	Yes
Strongly dispersive ($\gamma \sim \omega_0$)	Auxiliary-field Hamiltonian	Yes (enlarged space)

This table reflects the central message of Chapter 2: as the medium becomes more realistic, the inner product that makes the Maxwell operator self-adjoint becomes more complex, and the price of restoring Hermiticity is paid either by restricting the frequency range or by enlarging the state space.

2.7.4 Projection onto a subspace

Even when the full Maxwell operator is self-adjoint, restricting attention to a finite number of resonant modes creates an effective Hamiltonian that is generically non-Hermitian. The discarded modes act as an environment: they carry away energy (through radiation or absorption in modes outside the truncated subspace) and introduce complex corrections to the effective eigenvalues. The Petermann excess-noise factor, which quantifies the nonorthogonality of laser modes, is a direct consequence of this projection [4].

This mechanism is the subject of Chapter 7, where effective non-Hermitian Hamiltonians are derived from Maxwell field theory rather than postulating them.

Physical example: few-mode fibre. A single-mode optical fibre supports one guided mode per polarization. When this mode couples to a nearby scatterer (a nanoparticle, a microring, or a Bragg grating), the coupled system has two or three resonant modes, but the radiation continuum acts as an infinite environment. Projecting the full Maxwell problem onto the two-mode subspace yields an effective 2×2 non-Hermitian Hamiltonian whose off-diagonal elements include both the direct coupling and a radiative shift from the discarded continuum. The Petermann factor of the coupled modes measures how far the projected eigenvectors deviate from orthogonality—a direct consequence of the incomplete Hilbert space.

Four mechanisms, one mathematical origin

Material loss/gain makes the weight ε_r complex. Radiation makes the boundary terms nonzero. Dispersion makes the eigenproblem nonlinear in ω unless additional material degrees of freedom are included. Projection makes the Hilbert space incomplete. All four

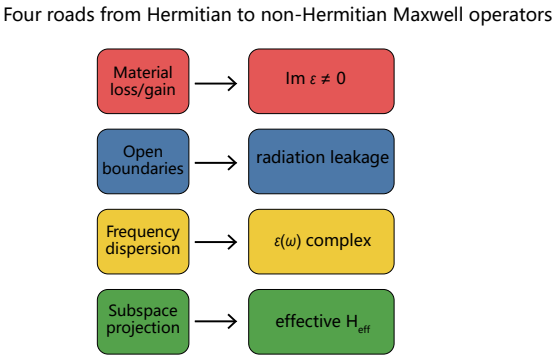


Figure 2.4 Four roads from Hermitian to non-Hermitian Maxwell operators. Each mechanism independently breaks the self-adjointness of the curl–curl operator Θ: material loss or gain ($\mathrm{Im}\,\varepsilon \neq 0$), open boundaries allowing radiation leakage, frequency dispersion making $\varepsilon(\omega)$ complex, and subspace projection yielding an effective non-Hermitian Hamiltonian. Many photonic systems exhibit several roads simultaneously. Each road is explored in depth in Parts II–V of this book.

destroy self-adjointness of the Maxwell operator, but they do so in physically distinct ways and lead to different spectral and topological consequences.

2.8 The structure of what follows

The Hermitian baseline established in this chapter—energy inner product, self-adjointness, real eigenvalues, orthogonal modes, reciprocity—is the standard against which every non-Hermitian phenomenon in the rest of this book will be measured.

Chapter 3 extends the self-adjoint framework to periodic media, where Bloch's theorem organizes the spectrum into bands. Chapter 4 introduces open systems and the scattering matrix, setting the stage for quasinormal modes.

In Part II, each of the four roads identified above [14, 15] will be explored in depth: material loss, gain, and dispersion in Chapter 5; quasinormal modes from radiation in Chapter 6; and projection to coupled-mode models in Chapter 7.

The reader who is already familiar with Hermitian electromagnetism may wish to proceed directly to Chapter 5. However, we recommend at least reviewing the distinction between reciprocity and Hermiticity in Section 2.6 and the four roads in Section 2.7, as these ideas recur throughout the non-Hermitian parts of the book.

Chapter 3

Periodicity and Photonic Bands

A homogeneous dielectric has a continuous spectrum: any frequency can propagate. Introduce a periodic modulation of the permittivity and the spectrum reorganizes into bands separated by gaps—frequency windows in which no propagating mode exists. The band structure of photonic crystals is derived directly from the Maxwell eigenproblem of Chapter 1, using Bloch's theorem as the organizing principle. The real, orthogonal, gapped band structure established here is the Hermitian baseline against which all non-Hermitian modifications in later chapters will be measured.

3.1 Periodic permittivity and the direct lattice

Consider a medium whose relative permittivity is periodic:

$$\varepsilon_{\mathrm{r}}(\mathbf{r}+\mathbf{R})=\varepsilon_{\mathrm{r}}(\mathbf{r}) \qquad \text{for all lattice vectors} \quad \mathbf{R}=n_1\mathbf{a}_1+n_2\mathbf{a}_2+n_3\mathbf{a}_3, \tag{3.1}$$

where $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3$ are primitive lattice vectors and $n_i\in\mathbb{Z}$. We assume $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is real, positive, and bounded, so the self-adjointness results of Chapter 2 apply within each unit cell.

The set of all lattice vectors $\mathbf{R}$ forms the *direct lattice*. The primitive unit cell Ω_{uc} can be chosen as the parallelotope spanned by $\{\mathbf{a}_i\}$ or, more symmetrically, as the Wigner–Seitz cell centered on a lattice point.

For most of this chapter we work in two dimensions, where the permittivity $\varepsilon_{\mathrm{r}}(x,y)$ is periodic in the (x,y) plane and uniform along z. This covers ideal two-dimensional photonic crystals—arrays of dielectric rods or air holes uniform along the third direction—and gives a useful starting point for slab photonic crystals when vertical confinement can be treated separately. It also allows the TE/TM decomposition of Section 1.5. The extension to fully three-dimensional crystals is conceptually straightforward but notationally heavier; see [12] for explicit three-dimensional calculations.

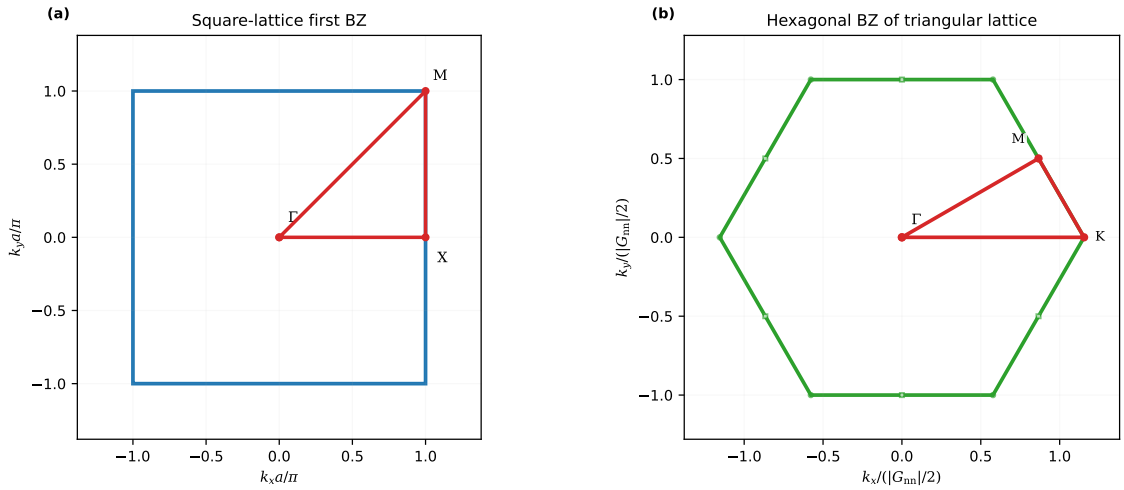


Figure 3.1 First Brillouin zones and standard high-symmetry paths in two common two-dimensional lattices. (a) For the square lattice the axes are $k_x a/\pi$ and $k_y a/\pi$; the first BZ is the square $[-1,1]^2$, and the red path is Γ–X–M–Γ. (b) For the triangular lattice the first BZ is hexagonal. The axes are normalized by $|G_{\mathrm{nn}}|/2$, where G_{nn} is a nearest reciprocal-lattice vector. In this convention the K-type points are the hexagon corners at radius $2/\sqrt{3}$, the M-type points are edge midpoints at radius 1, and the red curve shows one Γ–M–K–Γ sector. The two panels therefore use their natural reciprocal-space normalizations rather than a common length scale.

3.2 Reciprocal lattice and Brillouin zone

The *reciprocal lattice* is defined by vectors $\mathbf{G}$ satisfying $e^{i\mathbf{G}\cdot\mathbf{R}} = 1$ for all direct-lattice vectors $\mathbf{R}$. In two dimensions, the primitive reciprocal-lattice vectors $\mathbf{b}_1, \mathbf{b}_2$ satisfy

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\,\delta_{ij}, \qquad i,j = 1,2. \tag{3.2}$$

The first *Brillouin zone* (BZ) is the Wigner–Seitz cell of the reciprocal lattice. It is the fundamental domain for the wave vector $\mathbf{k}$: any $\mathbf{k}$ outside the first BZ can be mapped back into it by subtracting a reciprocal-lattice vector.

Example 3.1 (Square and triangular lattices)**.** For a square lattice with period a, the primitive vectors are $\mathbf{a}_1 = a\hat{x}$, $\mathbf{a}_2 = a\hat{y}$, and the reciprocal-lattice vectors are $\mathbf{b}_1 = (2\pi/a)\hat{x}$, $\mathbf{b}_2 = (2\pi/a)\hat{y}$. The first BZ is a square of side $2\pi/a$ centered at the origin.

For a triangular lattice with $\mathbf{a}_1 = a\hat{x}$, $\mathbf{a}_2 = a(\hat{x}/2 + \sqrt{3}\hat{y}/2)$, the BZ is a regular hexagon. Its high-symmetry points—Γ, M, K—play a central role in photonic-crystal band diagrams.

Figure 3.1 fixes the reciprocal-space geometry that will be used throughout the chapter. The important point is not the artistic shape of the cell, but the convention: the first BZ is a Wigner–Seitz cell of the reciprocal lattice, and the red curves are standard high-symmetry paths used to display bands.

Because $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is periodic, it admits a Fourier expansion:

$$\varepsilon_{\mathrm{r}}(\mathbf{r}) = \sum_{\mathbf{G}} \hat{\varepsilon}_{\mathbf{G}}\, e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{3.3}$$

where the sum runs over all reciprocal-lattice vectors and the Fourier coefficients are

$$\hat{\varepsilon}_{\mathbf{G}} = \frac{1}{|\Omega_{\mathrm{uc}}|} \int_{\Omega_{\mathrm{uc}}} \varepsilon_{\mathrm{r}}(\mathbf{r})\, e^{-i\mathbf{G}\cdot\mathbf{r}}\ \mathrm{d}^2 r. \tag{3.4}$$

The reality of ε_{r} implies $\hat{\varepsilon}_{-\mathbf{G}} = \hat{\varepsilon}^*_{\mathbf{G}}$.

3.2.1 The dielectric scaling law

The Maxwell eigenproblem possesses a scaling property with no counterpart in the Schrödinger equation. If $\varepsilon_{\mathrm{r}}(\mathbf{r})$ describes a photonic crystal with lattice constant a, then the rescaled medium $\varepsilon'_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}(\mathbf{r}/s)$ with lattice constant sa has eigenfrequencies $\omega'_n(\mathbf{k}') = \omega_n(\mathbf{k})/s$ at wave vectors $\mathbf{k}' = \mathbf{k}/s$. The band structure scales inversely with the lattice constant, with no change in shape:

$$\frac{\omega_n(\mathbf{k})\, a}{2\pi c} = \frac{\omega'_n(\mathbf{k}')\, (sa)}{2\pi c}. \tag{3.5}$$

This follows because the curl operator $\nabla\times$ scales as $1/a$ and the eigenvalue ω^2/c^2 scales as $1/a^2$, while ε_{r} is dimensionless and unchanged by the spatial rescaling.

The scaling law has two practical consequences. First, band diagrams are conventionally plotted in dimensionless units: the frequency axis is $\omega a/(2\pi c)$ and the wave-vector axis is ka/π. A single dimensionless band diagram then covers all physical lattice constants simultaneously. Second, a photonic crystal designed at microwave frequencies can be scaled to the optical regime simply by shrinking the lattice constant, provided the material dispersion of ε_{r} at the new frequency is accounted for. This scale-free property is unique to Maxwell's equations: the Schrödinger equation possesses a fundamental length (the Bohr radius) set by $\hbar$, m, and e, so electronic band structures do not rescale in this way.

The dimensionless presentation also clarifies the role of the dielectric contrast $\varepsilon_{\max}/\varepsilon_{\min}$: it is the only parameter (together with the geometry) that controls the dimensionless band structure. Increasing the contrast widens all gaps simultaneously, while changing the lattice constant merely shifts the physical frequency scale.

3.3 Bloch's theorem for Maxwell fields

The Maxwell curl–curl operator $\hat{\Theta}$ defined in (1.11) commutes with the discrete translation operator $T_{\mathbf{R}}\mathbf{H}(\mathbf{r}) = \mathbf{H}(\mathbf{r}+\mathbf{R})$, because $\varepsilon_{\mathrm{r}}(\mathbf{r}+\mathbf{R}) = \varepsilon_{\mathrm{r}}(\mathbf{r})$. Since $\hat{\Theta}$ and $\{T_{\mathbf{R}}\}$ share a common set of eigenstates, the eigenmodes of $\hat{\Theta}$ can be chosen to be simultaneous eigenstates of all lattice translations:

$$T_{\mathbf{R}}\, \mathbf{H}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{R}}\, \mathbf{H}_{n\mathbf{k}}(\mathbf{r}). \tag{3.6}$$

This is *Bloch's theorem*. It states that each eigenmode can be written as

$$\boxed{\mathbf{H}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, \mathbf{u}_{n\mathbf{k}}(\mathbf{r}),} \tag{3.7}$$

where $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ is a vector function with the periodicity of the lattice: $\mathbf{u}_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = \mathbf{u}_{n\mathbf{k}}(\mathbf{r})$.

The band index $n = 1, 2, 3, \ldots$ labels the discrete eigenvalues at each $\mathbf{k}$, and $\mathbf{k}$ is a continuous label running over the first Brillouin zone. The eigenfrequency $\omega_n(\mathbf{k})$ as a function of $\mathbf{k}$ is the *n-th photonic band*.

Bloch's theorem is a consequence of symmetry

The derivation above used only two ingredients: (i) the operator $\hat{\Theta}$ commutes with discrete translations, and (ii) the translation operators form an abelian group whose irreducible representations are one-dimensional and labeled by $\mathbf{k}$. No specific form of ε_r was assumed beyond periodicity. This means Bloch's theorem holds equally for the $\mathbf{E}$-field formulation, for magnetic media, and—crucially for later chapters—for complex ε_r, as long as the periodicity is maintained. In the non-Hermitian case, Bloch's theorem still provides a labeling by $\mathbf{k}$, but the eigenfrequencies become complex and the orthogonality of Bloch modes fails.

3.4 The unit-cell eigenproblem

Substituting the Bloch form (3.7) into the Maxwell eigenproblem $\hat{\Theta}\mathbf{H} = (\omega^2/c^2)\mathbf{H}$ transforms the problem on all of space into a problem on a single unit cell. Define the $\mathbf{k}$-dependent operator

$$\hat{\Theta}_\mathbf{k}\,\mathbf{u} = (\nabla + i\mathbf{k}) \times \left[\frac{1}{\varepsilon_\mathrm{r}(\mathbf{r})}(\nabla + i\mathbf{k}) \times \mathbf{u}\right], \tag{3.8}$$

which acts on periodic vector fields $\mathbf{u}(\mathbf{r})$ defined on Ω_uc with periodic boundary conditions. The eigenvalue equation becomes

$$\hat{\Theta}_\mathbf{k}\,\mathbf{u}_{n\mathbf{k}} = \frac{\omega_n(\mathbf{k})^2}{c^2}\,\mathbf{u}_{n\mathbf{k}}, \tag{3.9}$$

with $\mathbf{u}_{n\mathbf{k}}$ subject to periodic boundary conditions on $\partial\Omega_\mathrm{uc}$.

The operator $\hat{\Theta}_\mathbf{k}$ is self-adjoint on the unit cell for real $\varepsilon_\mathrm{r} > 0$, because the surface terms from integration by parts cancel between opposite faces (the Bloch phase factors have been absorbed into $\nabla + i\mathbf{k}$, and the boundary conditions are now strictly periodic). Consequently, for each $\mathbf{k}$ the eigenvalues $\omega_n(\mathbf{k})^2/c^2$ are real and non-negative, the eigenfunctions $\mathbf{u}_{n\mathbf{k}}$ are orthogonal, and all the results of Chapter 2 apply with $\hat{\Theta}$ replaced by $\hat{\Theta}_\mathbf{k}$.

The unit-cell eigenproblem (3.9) is the starting point for all numerical photonic band calculations. The standard approach is to expand $\mathbf{u}_{n\mathbf{k}}$ and $\varepsilon_\mathrm{r}^{-1}$ in plane waves (reciprocal-lattice vectors $\mathbf{G}$), converting the differential equation into a matrix eigenvalue problem that is solved for each $\mathbf{k}$ on a discrete grid in the Brillouin zone [12].

3.5 One-dimensional photonic crystals

Before treating two-dimensional systems, it is instructive to solve the simplest photonic crystal exactly: a one-dimensional stack of alternating dielectric layers. This example makes band gaps, Bloch wave vectors, and the connection between the Maxwell eigenproblem and scattering theory completely explicit.

3.5.1 Transfer matrix for a bilayer

Consider a periodic stack along x with unit cell consisting of a layer of thickness d_A and refractive index $n_A = \sqrt{\varepsilon_A}$ followed by a layer of thickness d_B and refractive index $n_B = \sqrt{\varepsilon_B}$, so the period is $a = d_A + d_B$. For normal incidence, the tangential field vector may be written as $(E, \eta_0 H)^{\mathsf{T}}$, where the magnetic field is scaled by the vacuum impedance. Its value at one side of the unit cell is related to the value at the other side by a 2×2 characteristic matrix M:

$$\begin{pmatrix} E_{m+1} \\ \eta_0 H_{m+1} \end{pmatrix} = M \begin{pmatrix} E_m \\ \eta_0 H_m \end{pmatrix}, \qquad M = M_B\, M_A, \tag{3.10}$$

where the single-layer transfer matrix is

$$M_j = \begin{pmatrix} \cos\phi_j & -i \sin\phi_j / n_j \\ -i\, n_j \sin\phi_j & \cos\phi_j \end{pmatrix}, \qquad \phi_j = \frac{\omega\, n_j\, d_j}{c}, \tag{3.11}$$

for $j = A, B$. The phase ϕ_j is the optical path length through layer j measured in radians. This transfer matrix is unimodular, $\det M_j = 1$, as required by energy conservation in a lossless medium. The product $M = M_B\, M_A$ is therefore also unimodular.

3.5.2 Bloch condition and the dispersion relation

Bloch's theorem requires that traversing one period multiplies the field amplitudes by e^{iKa}, where K is the Bloch wave vector. This means M has eigenvalue e^{iKa}, so

$$\boxed{\cos(Ka) = \tfrac{1}{2}\operatorname{tr} M = \cos\phi_A \cos\phi_B - \tfrac{1}{2}\left(\frac{n_A}{n_B} + \frac{n_B}{n_A}\right)\sin\phi_A \sin\phi_B.} \tag{3.12}$$

This is the exact photonic dispersion relation for a 1D periodic stack. When $|\tfrac{1}{2}\operatorname{tr} M| \le 1$, the Bloch wave vector K is real and a propagating Bloch mode exists: these frequencies lie in a *pass band*. When $|\tfrac{1}{2}\operatorname{tr} M| > 1$, K acquires an imaginary part and the wave is evanescent: these frequencies lie in a *stop band* or band gap.

It is convenient to define the *half-trace*

$$\xi(\omega) \;\equiv\; \tfrac{1}{2}\operatorname{tr} M(\omega), \tag{3.13}$$

so that the pass-band condition is simply $|\xi| \le 1$. The band structure is found by solving $K(\omega) = \arccos[\xi(\omega)]/a$. The result is a set of bands separated by gaps, plotted as ω versus K or equivalently as K versus ω. When the two materials become identical ($n_A = n_B$), the trace reduces to $\cos(\omega a\, n/c)$ and the bands become the free-photon dispersion folded into the first Brillouin zone, with no gaps.

The same trace formula is shown in Figure 3.2 for a representative high-contrast bilayer. This figure is deliberately plotted in two equivalent ways: as pass bands in the reduced zone and as the half-trace criterion. It is the inequality $|\xi(\omega)| \le 1$, not the visual shading, that defines a propagating Bloch wave.

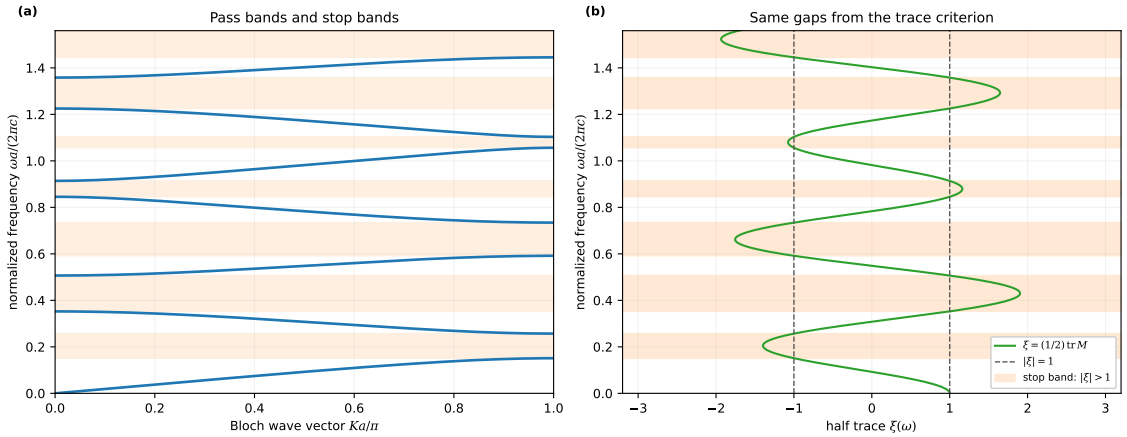


Figure 3.2 Transfer-matrix view of one-dimensional photonic bands. The bilayer calculation uses $n_A = 3.6$, $n_B = 1$, and equal layer thicknesses $d_A = d_B = a/2$. (a) Blue curves show pass bands in which the Bloch wave vector K is real. Orange intervals are stop bands. (b) The same information is encoded by the half trace $\xi(\omega) = \frac{1}{2}\,\mathrm{tr}\,M$: pass bands satisfy $|\xi| \le 1$, while stop bands satisfy $|\xi| > 1$. Every plotted pass-band point satisfies $|\xi| \le 1$ to numerical precision.

Example 3.2 (Quarter-wave stack)**.** The most important special case is the quarter-wave stack, in which each layer has the same optical thickness: $n_A d_A = n_B d_B = \lambda_B/4$, where λ_B is the Bragg wavelength. At the Bragg frequency $\omega_B = \pi c/(n_A d_A)$, both phases equal $\pi/2$, the half-trace becomes $\xi = -(n_A/n_B + n_B/n_A)/2$, and the gap is widest. The fractional gap width (gap–midgap ratio) is

$$\frac{\Delta\omega}{\omega_B} = \frac{4}{\pi}\arcsin\left(\frac{|n_A - n_B|}{n_A + n_B}\right), \tag{3.14}$$

which grows monotonically with the refractive-index contrast. For a quarter-wave stack with $\varepsilon_A = 13$ ($n_A = \sqrt{13} \approx 3.61$) and $\varepsilon_B = 1$ ($n_B = 1$), Eq. (3.14) gives a fundamental gap–midgap ratio of approximately 77%—an exceptionally wide gap reflecting the large contrast. (For the equal-thickness bilayer of Figure 3.2, which does not satisfy the quarter-wave condition, the first gap–midgap ratio is smaller, about 52%.)

3.5.3 Connection to the scattering matrix

The transfer matrix of N unit cells is M^N. Using the Cayley–Hamilton theorem for a unimodular 2×2 matrix, M^N can be expressed in terms of Chebyshev polynomials of the second kind:

$$M^N = U_{N-1}(\xi)\,M - U_{N-2}(\xi)\,I, \qquad \xi = \tfrac{1}{2}\,\mathrm{tr}\,M, \tag{3.15}$$

where $U_m(\xi) = \sin[(m+1)\arccos\xi]/\sin[\arccos\xi]$. In a pass band ($|\xi| \le 1$), the finite-stack transmission exhibits $N - 1$ Fabry–Pérot oscillations between successive gap edges. In a stop band ($|\xi| > 1$), U_{N-1} grows exponentially with N, and the transmission decays as $e^{-2\,\mathrm{Im}(K)\,Na}$. This connects the Bloch band structure directly to the poles and zeros of the scattering matrix discussed in Chapter 10: each pass-band resonance of the finite stack is a precursor of a Bloch band, and the stop band corresponds to a frequency window where no S-matrix pole approaches the real axis.

The 1D transfer-matrix approach extends naturally to complex ε_r (Chapters 5 and 9), where the Bloch wave vector becomes complex even inside pass bands, and the sharp distinction between pass and stop bands blurs into a continuous complex band structure (Chapter 12).

3.6 Band structure in two dimensions

In a two-dimensional photonic crystal with $\varepsilon_r(x, y)$ periodic in the (x, y) plane, $\mu_r = 1$, and fields independent of z (or at fixed $k_z = 0$), the vector Maxwell eigenproblem decouples into the two scalar polarizations introduced in Section 1.5.

3.6.1 TM bands

For TM polarization ($\mathbf{E} = E_z \hat{z}$), the Bloch form is $E_{z,n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, \phi_{n\mathbf{k}}(\mathbf{r})$ with $\phi_{n\mathbf{k}}$ periodic. The unit-cell eigenproblem becomes

$$-(\nabla_\perp + i\mathbf{k})^2\, \phi_{n\mathbf{k}}(\mathbf{r}) = \frac{\omega_n(\mathbf{k})^2}{c^2}\, \varepsilon_r(\mathbf{r})\, \phi_{n\mathbf{k}}(\mathbf{r}). \tag{3.16}$$

The eigenvalue is ω_n^2/c^2 and the weight on the right is $\varepsilon_r(\mathbf{r})$, so the inner product is the ε_r-weighted one from (1.21).

3.6.2 TE bands

For TE polarization ($\mathbf{H} = H_z \hat{z}$), the Bloch form is $H_{z,n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\, \psi_{n\mathbf{k}}(\mathbf{r})$ with $\psi_{n\mathbf{k}}$ periodic. The unit-cell eigenproblem is

$$-(\nabla_\perp + i\mathbf{k}) \cdot \left[\frac{1}{\varepsilon_r(\mathbf{r})}\, (\nabla_\perp + i\mathbf{k})\, \psi_{n\mathbf{k}}(\mathbf{r})\right] = \frac{\omega_n(\mathbf{k})^2}{c^2}\, \psi_{n\mathbf{k}}(\mathbf{r}). \tag{3.17}$$

Here the weight is unity and the inner product is the unweighted one from (1.23).

Remark 3.1. The TM and TE band structures can be very different for the same lattice, because the spatial structure of the operator differs: TM modes "see" ε_r as a multiplicative weight, while TE modes "see" ε_r^{-1} inside a divergence operator. A complete photonic band gap—a frequency range forbidden to *both* polarizations—requires the TM and TE gaps to overlap, which imposes strong constraints on the lattice geometry and dielectric contrast.

3.6.3 The plane-wave expansion method

The standard numerical approach to the unit-cell eigenproblem (3.9) is the *plane-wave expansion* (PWE) method [16, 17]. Expand the periodic part of the Bloch mode in reciprocal-lattice vectors:

$$\mathbf{u}_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} \tilde{\mathbf{u}}_{n\mathbf{k}}(\mathbf{G})\, e^{i\mathbf{G}\cdot\mathbf{r}}. \tag{3.18}$$

Substituting into (3.9) and projecting onto each $\mathbf{G}'$ yields a matrix eigenvalue problem. For TM polarization in 2D, where E_z is a scalar, the equation (3.16) gives the generalized eigenvalue

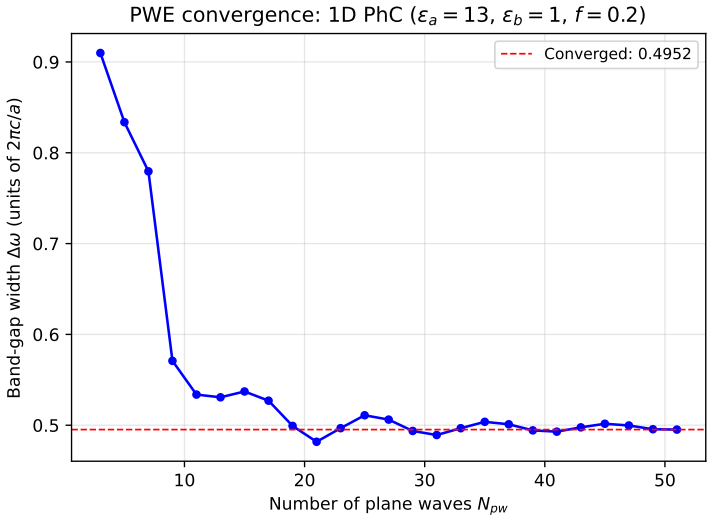


Figure 3.3 Convergence of the plane-wave expansion for a 1D photonic crystal ($\varepsilon_a = 13$, $\varepsilon_b = 1$, filling fraction $f = 0.2$). The computed band-gap width converges rapidly with the number of plane waves N_{pw}, reaching the converged value $\Delta\omega \approx 0.495$ (in units of $2\pi c/a$) for $N_{\mathrm{pw}} \gtrsim 30$ in this one-dimensional test. In higher dimensions, sharp high-contrast interfaces usually require more careful Fourier factorization and significantly more plane waves.

problem

$$|\mathbf{k}+\mathbf{G}'|^2\,\tilde{\phi}_{n\mathbf{k}}(\mathbf{G}') = \frac{\omega_n(\mathbf{k})^2}{c^2}\sum_{\mathbf{G}}\hat{\varepsilon}_{\mathbf{G}'-\mathbf{G}}\,\tilde{\phi}_{n\mathbf{k}}(\mathbf{G}), \tag{3.19}$$

where $\hat{\varepsilon}_{\mathbf{G}'-\mathbf{G}}$ is the Fourier coefficient of $\varepsilon_{\mathrm{r}}(\mathbf{r})$ at reciprocal-lattice vector $\mathbf{G}'-\mathbf{G}$. This is a Hermitian generalized eigenvalue problem of size $N_G \times N_G$, where N_G is the number of plane waves retained. For TE polarization, by contrast, the Fourier coefficients of $\varepsilon_{\mathrm{r}}^{-1}$ enter the operator side, reflecting the weighted Laplacian in (3.17).

Convergence is algebraic in N_G: smooth ε_{r} profiles converge rapidly, while sharp interfaces converge more slowly due to the Gibbs phenomenon in the Fourier representation of discontinuous material profiles. Improved formulations such as the Ho–Chan–Soukoulis inverse rule treat the discontinuity correctly by applying the appropriate Fourier factorization rule to normal and tangential field components, significantly improving convergence for high-contrast structures [12].

For the full vector problem in 3D, the expansion must enforce the transversality constraint $(\mathbf{k}+\mathbf{G})\cdot\tilde{\mathbf{u}} = 0$ for each plane wave, reducing the degrees of freedom from three to two per $\mathbf{G}$. The resulting Hermitian matrix eigenvalue problem is larger but structurally similar.

Example 3.3 (2D rod array: square lattice)**.** Consider a square array (period a) of circular dielectric rods ($\varepsilon_{\mathrm{rod}} = 8.9$, radius $r = 0.2a$) in air ($\varepsilon_{\mathrm{bg}} = 1$). The filling fraction is $f = \pi r^2/a^2 \approx 0.126$. A plane-wave expansion with $N_G \approx 500$ converges the lowest six TM bands to four significant figures. The TM band structure shows a large gap between the first and second bands (gap–midgap ratio $\approx 37\%$), while the TE bands have no complete gap at this filling fraction. This asymmetry between TM and TE gaps is generic for rod-in-air geometries and can be understood from the field-concentration argument of the next section.

3.7 Band gaps

A *photonic band gap* is a frequency interval $[\omega_{\text{low}}, \omega_{\text{high}}]$ in which no Bloch mode exists for any $\mathbf{k}$ in the Brillouin zone. The physical consequence is that light at frequencies within the gap cannot propagate through the crystal: it is exponentially attenuated, decaying as $e^{-\kappa(\omega)\,x}$ with a penetration depth set by the imaginary wave vector κ.

The existence of a gap requires sufficient dielectric contrast $\varepsilon_{\max}/\varepsilon_{\min}$ and appropriate lattice geometry. For TM modes in a 2D crystal of high-ε_{r} rods in air, gaps open readily; for TE modes, connected high-ε_{r} regions (such as air holes in a dielectric slab) are more favorable.

Perturbative origin of gaps. At a Brillouin-zone boundary or high-symmetry point, two bands that would cross in a uniform medium are coupled by the periodic modulation of ε_{r}. The resulting level repulsion opens a gap whose width, to first order in the Fourier coefficient $\hat{\varepsilon}_{\mathbf{G}}$ of the relevant reciprocal-lattice vector, is proportional to $|\hat{\varepsilon}_{\mathbf{G}}|$. This is the photonic analogue of the nearly-free-electron picture in solid-state physics.

Variational bounds. Since $\hat{\Theta}_{\mathbf{k}}$ is self-adjoint, the Rayleigh quotient provides variational bounds on the band-edge frequencies. The lowest band satisfies

$$\frac{\omega_1(\mathbf{k})^2}{c^2} \leq \frac{\int_{\Omega_{\text{uc}}} \varepsilon_{\text{r}}^{-1}\,|(\nabla + i\mathbf{k}) \times \mathbf{u}|^2 \;\mathrm{d}^2 r}{\int_{\Omega_{\text{uc}}} |\mathbf{u}|^2 \;\mathrm{d}^2 r} \tag{3.20}$$

for any trial function $\mathbf{u}$, with equality when $\mathbf{u}$ is the exact Bloch eigenfunction. Higher bands are bounded by the min–max principle. These variational estimates are the theoretical foundation of the plane-wave expansion method.

3.7.1 Field concentration and the physical origin of gaps

The variational principle provides a remarkably intuitive explanation for why gaps open and which polarization is favored. Consider the two band-edge modes at a Brillouin-zone boundary where a gap opens. One mode—the *dielectric band* or lower band edge—concentrates its electric-field energy in the high-ε_{r} regions, while the other—the *air band* or upper band edge—concentrates its energy in the low-ε_{r} regions.

The dielectric-band mode has a lower eigenfrequency because the Rayleigh quotient (3.20) involves $\varepsilon_{\text{r}}^{-1}$ in the numerator: concentrating $|\nabla \times \mathbf{u}|^2$ where $\varepsilon_{\text{r}}^{-1}$ is small (i.e., where ε_{r} is large) reduces the numerator and hence the eigenvalue. Conversely, the air-band mode is forced to have its curl energy where $\varepsilon_{\text{r}}^{-1}$ is large, raising the eigenvalue. The frequency difference between these two modes is the band gap.

This field-concentration mechanism explains the TM/TE asymmetry observed in the 2D example above:

- *TM modes in rod arrays.* The E_z field is a scalar and can concentrate efficiently inside isolated high-ε_{r} rods. The dielectric and air bands have very different field distributions, producing a large gap.

- *TE modes in rod arrays.* The in-plane $\mathbf{H}$ field must satisfy continuity conditions at the rod boundaries that prevent it from concentrating entirely inside the rods. The dielectric and air bands are more similar, and the gap is small or absent.
- *TE modes in hole arrays.* When the high-ε_{r} region is connected (air holes in a dielectric slab), the in-plane $\mathbf{H}$ field *can* concentrate in the connected dielectric, and TE gaps open readily.

A complete photonic band gap—forbidden to both polarizations—therefore requires a geometry that simultaneously favors both TM and TE gaps, such as a triangular lattice of air holes with appropriately chosen radius and dielectric contrast.

The field-concentration picture also explains why gap width increases with dielectric contrast: larger contrast amplifies the frequency splitting between modes that concentrate in different materials. This is precisely the effect captured by the Fourier coefficient $|\hat{\varepsilon}_{\mathbf{G}}|$ in the perturbative estimate, which itself grows with contrast.

3.8 Degeneracies and symmetry

At generic points in the Brillouin zone, the eigenvalues $\omega_n(\mathbf{k})$ are non-degenerate. Degeneracies can occur at high-symmetry points where the little group (the subgroup of the point group that leaves $\mathbf{k}$ invariant) has irreducible representations of dimension greater than one.

3.8.1 Dirac points

At the K and K' points of a hexagonal Brillouin zone, a pair of bands may touch linearly, forming a *Dirac cone* [8, 18, 19]:

$$\omega_{\pm}(\mathbf{k}) \approx \omega_D \pm v_D\,|\mathbf{k}-\mathbf{k}_D| + \cdots, \tag{3.21}$$

where ω_D is the Dirac frequency and v_D is an effective group velocity. Near a Dirac point, the two-band effective Hamiltonian takes the form

$$H_{\mathrm{eff}}(\delta\mathbf{k}) = v_D\big(\delta k_x\,\sigma_x + \delta k_y\,\sigma_y\big), \tag{3.22}$$

with $\sigma_{x,y}$ the Pauli matrices acting on the two-band subspace. This is the photonic analogue of graphene's electronic Dirac cone.

3.8.2 Quadratic touchings and accidental degeneracies

At the Γ point, symmetry can enforce a quadratic band touching where $\omega(\mathbf{k}) \propto |\mathbf{k}|^2$ rather than $|\mathbf{k}|$. Perturbations that break the protecting symmetry can open a gap or split the touching into a pair of Dirac points.

Accidental degeneracies—those not enforced by symmetry—are generically avoided in a two-parameter family (the two components of $\mathbf{k}$ in 2D). However, they can be tuned by adjusting a third parameter (e.g., the rod radius or the dielectric contrast). When two bands approach each other and reconnect, the topology of the band structure can change.

3.9 Orthogonality and normalization of Bloch modes

The Bloch modes satisfy orthogonality relations that follow from the self-adjointness of $\hat{\Theta}_{\mathbf{k}}$. For two modes at the *same* $\mathbf{k}$,

$$\int_{\Omega_{\mathrm{uc}}} \mathbf{u}^*_{m\mathbf{k}}(\mathbf{r}) \cdot \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) \; \mathrm{d}^2 r = \delta_{mn}, \tag{3.23}$$

after normalization. For modes at *different* $\mathbf{k}$ and $\mathbf{k}'$, the full Bloch modes $\mathbf{H}_{n\mathbf{k}}$ and $\mathbf{H}_{m\mathbf{k}'}$ are orthogonal over all space (or, equivalently, their overlap per unit cell vanishes when $\mathbf{k} \neq \mathbf{k}'$).

This orthonormality is the basis for expanding arbitrary fields in Bloch modes, defining projection operators onto individual bands, and—as we shall see in Chapter 11—computing the Berry connection and curvature from the $\mathbf{k}$-dependence of $\mathbf{u}_{n\mathbf{k}}$.

3.10 Group velocity, effective mass, and density of states

Beyond the eigenfrequencies $\omega_n(\mathbf{k})$, the dispersion surface contains essential physical information through its derivatives.

3.10.1 Group velocity

The group velocity of a Bloch mode is the gradient of the dispersion relation:

$$\mathbf{v}_g = \nabla_{\mathbf{k}}\, \omega_n(\mathbf{k}). \tag{3.24}$$

It can also be expressed as an expectation value using the Hellmann–Feynman theorem. Differentiating the unit-cell eigenproblem (3.9) with respect to $\mathbf{k}$ gives

$$\mathbf{v}_g = \frac{c^2}{2\omega_n} \frac{\displaystyle\int_{\Omega_{\mathrm{uc}}} \mathbf{u}^*_{n\mathbf{k}} \cdot \left[(\nabla_{\mathbf{k}} \hat{\Theta}_{\mathbf{k}})\, \mathbf{u}_{n\mathbf{k}} \right] \; \mathrm{d}^2 r}{\displaystyle\int_{\Omega_{\mathrm{uc}}} |\mathbf{u}_{n\mathbf{k}}|^2 \; \mathrm{d}^2 r}, \tag{3.25}$$

where $\nabla_{\mathbf{k}} \hat{\Theta}_{\mathbf{k}}$ is obtained by differentiating the operator $(\nabla + i\mathbf{k}) \times [\varepsilon_{\mathrm{r}}^{-1} (\nabla + i\mathbf{k}) \times \cdot]$ with respect to $\mathbf{k}$. This expression guarantees that the group velocity is real and bounded whenever $\hat{\Theta}_{\mathbf{k}}$ is self-adjoint. In non-Hermitian systems, the self-adjointness fails and the group velocity becomes complex; the physical interpretation of wave-packet propagation must then be revisited (Chapter 12).

3.10.2 Effective mass at band edges

At band edges, where $\nabla_{\mathbf{k}} \omega_n = 0$, the band curvature defines an effective inverse-mass tensor:

$$(m^{-1})_{ij} = \left. \frac{\partial^2 \omega_n}{\partial k_i \, \partial k_j} \right|_{\mathbf{k}=\mathbf{k}_0}. \tag{3.26}$$

Near a band edge at $\mathbf{k}_0$, the dispersion is parabolic: $\omega_n(\mathbf{k}) \approx \omega_n(\mathbf{k}_0) + \frac{1}{2} \sum_{ij} (m^{-1})_{ij}\, \delta k_i \, \delta k_j$. This effective-mass approximation is the photonic analogue of the semiconductor effective mass and forms the basis for envelope-function theories of photonic-crystal defect modes and waveguides:

a defect introduced into the crystal acts as a potential well in the envelope equation, binding states pulled out of the adjacent band edge into the gap.

The effective mass also controls the curvature of the equifrequency contour near the band edge. An isotropic parabolic band has circular contours; an anisotropic band has elliptical contours whose principal axes are aligned with the eigenvectors of $(m^{-1})_{ij}$. In self-collimation effects, the equifrequency contour is made flat over a range of $\mathbf{k}$, so that wave packets propagate with nearly zero angular spread despite the discrete lattice.

3.10.3 Density of states and van Hove singularities

The *photonic density of states* (DOS) counts the number of modes per unit frequency. For a two-dimensional photonic crystal,

$$\rho(\omega) = \sum_n \int_{\mathrm{BZ}} \delta\big(\omega - \omega_n(\mathbf{k})\big)\, \frac{\mathrm{d}^2 k}{(2\pi)^2}. \tag{3.27}$$

(In three dimensions the integration measure becomes $\mathrm{d}^3k/(2\pi)^3$.) Inside a band gap, $\rho(\omega) = 0$: there are no propagating modes. This is precisely why spontaneous emission is inhibited in a photonic band-gap material—the emitter simply has no modes into which it can radiate.

At band edges, the vanishing group velocity produces *van Hove singularities* in the DOS. In two dimensions, these are logarithmic divergences; in three dimensions, they manifest as square-root edges. Physically, the diverging DOS means that a large number of modes are compressed into a narrow frequency range, which strongly enhances light–matter interaction. Slow-light devices, enhanced nonlinear optics, and low-threshold lasing in photonic crystals all exploit the high DOS near band edges.

Figure 3.4 shows the same idea in a simple model where the geometry is visible at a glance. It is not meant to be a full plane-wave calculation for a particular dielectric pattern; it is a teaching model for how stationary points of a dispersion surface become features of the DOS.

The local density of states (LDOS), introduced in Chapter 1, is the position-resolved version of $\rho(\omega)$ and determines the Purcell enhancement of a dipole emitter at a given location in the crystal. In the Hermitian case, the LDOS is non-negative and can be computed from the Bloch-mode expansion. In non-Hermitian systems (Chapter 6), the LDOS must instead be reconstructed from quasinormal modes or from the full Green tensor. Individual modal contributions can have signs and phases that have no Hermitian analogue, so only the properly assembled physical response has the usual positivity properties in passive systems.

From band diagrams to physical observables

The photonic band structure is not just a plot—it encodes group velocity, effective mass, density of states, and through these, the propagation, confinement, emission, and interaction of light in the crystal. Every derivative of $\omega_n(\mathbf{k})$ carries physical meaning, and every singularity of the band structure controls a physical effect.

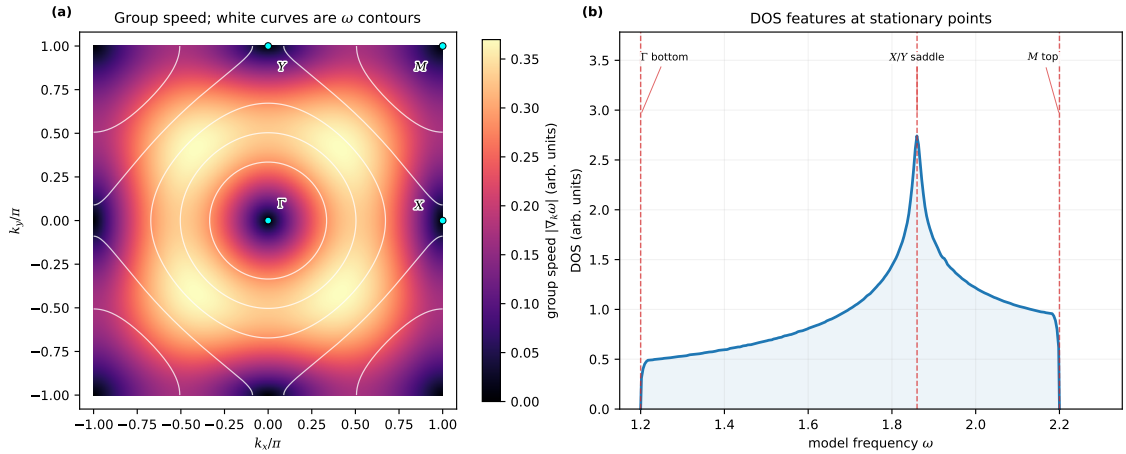


Figure 3.4 Group velocity and density of states in a pedagogical square-lattice band model. (a) The colour shows $|\nabla_{\mathbf{k}}\omega|$ and the white curves are constant-frequency contours. Stationary points occur at Γ, at the symmetry-related X/Y saddle points, and at M. (b) A histogram-smoothed, model-normalized DOS in arbitrary units. The dashed lines mark the corresponding band bottom, van-Hove saddle feature, and band top. In a full Maxwell band calculation the same relation is expressed by Eq. (3.27): DOS features arise where the dispersion surface has small or vanishing group velocity.

3.11 Beyond Hermitian bands: a preview

The entire construction of this chapter assumed real, positive ε_{r}. What happens when this assumption is relaxed?

Complex permittivity. If $\varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon'_{\mathrm{r}}(\mathbf{r}) + i\varepsilon''_{\mathrm{r}}(\mathbf{r})$ with the same spatial periodicity, Bloch's theorem still holds—it depends only on discrete translational symmetry, not on Hermiticity. But the eigenfrequencies $\omega_n(\mathbf{k})$ become complex, the bands lift off the real axis into the complex frequency plane, and the orthogonality (3.23) fails. The "band structure" becomes a set of complex-valued curves $\omega_n(\mathbf{k})$, and new topological invariants—winding numbers in the complex frequency plane—replace the familiar Chern numbers. This is the subject of Chapters 12 and 13.

Non-reciprocal media. If the permittivity tensor $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is not symmetric (as in a magneto-optical medium), time-reversal symmetry is broken and the band structure loses the constraint $\omega_n(\mathbf{k}) = \omega_n(-\mathbf{k})$. Gaps can carry nonzero Chern numbers, and topologically protected one-way edge states appear. This is the domain of topological photonics [9, 20, 21], developed in Chapter 11.

Open or finite crystals. A photonic crystal of finite extent has no exact translational symmetry. Bloch's theorem does not apply, and the spectrum changes from continuous bands to a discrete set of resonances. In Hermitian systems, the bulk-boundary correspondence connects the band topology to the existence of edge states. In non-Hermitian systems, this correspondence can fail through the skin effect (Chapter 14).

Photonic bands as the Hermitian foundation

The band structure $\omega_n(\mathbf{k})$ of a lossless periodic photonic crystal, derived from the self-adjoint Maxwell operator acting on Bloch modes in a unit cell, provides the Hermitian foundation on which all non-Hermitian photonic band theory is built. Real eigenfrequencies, orthogonal Bloch modes, variational bounds, and well-defined band gaps are the standard from which complex bands, point-gap topology, and the skin effect depart.

Chapter 4

Boundaries, Radiation, and Scattering

Chapters 1–3 established the spectral theory of closed or periodic photonic systems, where the operator $\hat{\Theta}$ is self-adjoint and the eigenfrequencies are real. Real photonic structures, however, are rarely closed. Light escapes: a microcavity leaks, a nanoparticle radiates, a grating diffracts. The moment we impose outgoing-wave boundary conditions to describe this radiation, the mathematical framework changes qualitatively. Eigenfrequencies become complex, modes cease to be normalizable in the usual sense, and the scattering matrix—with its poles and zeros—replaces the eigenvalue spectrum as the central object. The scattering framework developed below connects the Hermitian foundations of Part I to the non-Hermitian physics of Part II.

4.1 From cavities to open systems

A perfectly closed cavity—surrounded by perfect electric conductors (PEC)—supports a discrete, real spectrum as shown in Section 1.7. No energy leaves or enters the cavity, and the modes are orthogonal standing waves.

Now remove one mirror, or replace the PEC walls with a finite dielectric boundary. Energy can escape. A mode that was perfectly trapped at frequency ω_0 now decays in time at a rate γ, and the effective eigenfrequency becomes complex:

$$\tilde{\omega}_n = \omega_n - i\gamma_n, \qquad \gamma_n > 0. \tag{4.1}$$

The field oscillates at ω_n and decays as $e^{-\gamma_n t}$. The quality factor is $Q_n = \omega_n/(2\gamma_n)$: higher Q means slower leakage.

This has mathematical consequences. The complex frequency signals a fundamental change in the mathematical structure of the Maxwell eigenproblem. Outgoing-wave boundary conditions render the operator non-self-adjoint, and the entire machinery of Hermitian spectral theory—real eigenvalues, orthogonality, completeness, energy inner product—must be rebuilt. That reconstruction is the subject of Chapter 6; here we set the stage by introducing the scattering matrix and the role of boundary conditions.

4.1.1 Energy balance in open systems

For a finite region Ω surrounding the scatterer, the time-averaged Poynting balance can be written as

$$-\frac{\mathrm{d}U}{\mathrm{d}t} = P_{\mathrm{rad}} + P_{\mathrm{abs}}, \qquad P_{\mathrm{rad}} = \oint_{\partial\Omega} \langle \mathbf{S} \rangle \cdot \hat{\mathbf{n}} \, \mathrm{d}S, \tag{4.2}$$

where $U = \int_\Omega \langle u \rangle \, \mathrm{d}^3 r$ is the stored electromagnetic energy and P_{abs} is the power absorbed by the material inside Ω. For a closed cavity with PEC walls, $P_{\mathrm{rad}} = 0$ and energy is conserved (or dissipated only by material absorption). For an open system, $P_{\mathrm{rad}} > 0$ represents power carried away by outgoing waves. This radiation loss is irreversible in the reduced cavity description and is the physical origin of the imaginary part of the eigenfrequency.

The energy decay rate of a mode with complex frequency $\tilde{\omega} = \omega - i\gamma$ is 2γ, and the time-averaged stored energy U and radiated power P_{rad} are related by

$$Q = \omega \, \frac{U}{P_{\mathrm{rad}} + P_{\mathrm{abs}}} = \frac{\omega}{2\gamma}. \tag{4.3}$$

When material absorption is negligible, Q is determined solely by the radiation loss, which is controlled by the geometry of the cavity opening and the refractive-index contrast at the boundary.

4.2 Outgoing-wave boundary conditions

4.2.1 The Silver–Müller radiation condition

In an open system, we require that far from the scatterer the electromagnetic field consist only of outgoing waves. In three dimensions, this is the *Silver–Müller radiation condition*:

$$\lim_{r \to \infty} \left[r \left(\sqrt{\mu_0} \, \mathbf{H} \times \hat{\mathbf{r}} - \sqrt{\varepsilon_0} \, \mathbf{E} \right) \right] = 0, \tag{4.4}$$

uniformly in angle. The factor is r in three dimensions because outgoing spherical waves decay as $1/r$; in two-dimensional cylindrical problems the corresponding condition uses the $\sqrt{r}$ asymptotic factor. In one dimension, it reduces to requiring a purely outgoing traveling wave e^{+ikr} (with our $e^{-i\omega t}$ convention) at large distances.

The Silver–Müller condition is the electromagnetic analogue of the Sommerfeld radiation condition for scalar waves. It ensures that the scattered field represents power flowing away from the scatterer, not toward it—a physical requirement for any passive, causal system.

4.2.2 Why outgoing-wave conditions break self-adjointness

Unlike the PEC condition $\hat{\mathbf{n}} \times \mathbf{E} = 0$, the outgoing-wave condition does not make the surface term in the self-adjointness proof (1.16) vanish. Instead, the outgoing wave carries a nonzero Poynting flux through any enclosing surface:

$$\oint_{\partial\Omega} \langle \mathbf{S} \rangle \cdot \hat{\mathbf{n}} \, \mathrm{d}S = P_{\mathrm{rad}} > 0. \tag{4.5}$$

This radiation loss is the physical mechanism by which openness destroys self-adjointness. Formally, if we try to verify $\langle \mathbf{F}, \hat{\Theta}\mathbf{G}\rangle = \langle \hat{\Theta}\mathbf{F}, \mathbf{G}\rangle$ for two outgoing-wave solutions, the surface integral (1.16) leaves a nonzero remainder proportional to the cross-flux between $\mathbf{F}$ and $\mathbf{G}$.

The failure of self-adjointness is not a technicality—it is the mathematical expression of a deep physical fact: an open system exchanges energy with its environment, and the environment's degrees of freedom are not included in the operator $\hat{\Theta}$. The full system (scatterer plus radiation field at infinity) is still governed by the self-adjoint Maxwell operator on all of space, but the reduced description restricted to the scatterer alone is non-self-adjoint. This is the first instance of a theme that will recur throughout the book: non-Hermiticity as the price of a reduced description.

Remark 4.1. The outgoing-wave boundary condition depends on frequency through the wave vector $k = \omega/c$. This means the eigenproblem is nonlinear in ω: we cannot simply write $\hat{\Theta}\mathbf{H} = \omega^2\mathbf{H}$ with a fixed $\hat{\Theta}$ independent of ω. Instead, for each candidate complex $\tilde{\omega}$, the boundary condition changes. This is why the eigenfrequencies of open systems are called *resonances* rather than eigenvalues in the strict sense.

4.3 The scattering matrix

4.3.1 Definition and physical meaning

Consider a scatterer illuminated by an incoming wave. The total field is $\mathbf{E} = \mathbf{E}_{\text{in}} + \mathbf{E}_{\text{sc}}$, where $\mathbf{E}_{\text{sc}}$ satisfies the outgoing-wave condition. If the scatterer is linear, the relation between incoming and outgoing wave amplitudes defines the *scattering matrix* $\mathbf{S}(\omega)$.

For a one-dimensional layered structure with waves incident from the left and right, the 2×2 scattering matrix relates incoming amplitudes (a_L^+, a_R^-) to outgoing amplitudes (a_L^-, a_R^+):

$$\begin{pmatrix} a_L^- \\ a_R^+ \end{pmatrix} = \mathbf{S}(\omega) \begin{pmatrix} a_L^+ \\ a_R^- \end{pmatrix} = \begin{pmatrix} r_L & t' \\ t & r_R \end{pmatrix} \begin{pmatrix} a_L^+ \\ a_R^- \end{pmatrix}, \tag{4.6}$$

where r_L, r_R are the reflection coefficients from the left and right, and t, t' are the transmission coefficients. For a reciprocal medium, $t = t'$; for a $\mathcal{PT}$-symmetric structure, additional constraints apply (Chapter 9).

For a general three-dimensional scatterer with N incoming and outgoing channels (propagating modes of the surrounding medium), $\mathbf{S}$ is an $N \times N$ matrix. The channels may be angular-momentum partial waves (for spherical scatterers), waveguide modes (for integrated photonic circuits), or diffraction orders (for periodic gratings).

The physical content of $\mathbf{S}(\omega)$ is a complete input-output description of the scatterer: given any incident field (decomposed into channel amplitudes), $\mathbf{S}$ determines the scattered field. All measurable quantities—reflection, transmission, absorption, scattering cross-sections—can be extracted from $\mathbf{S}$.

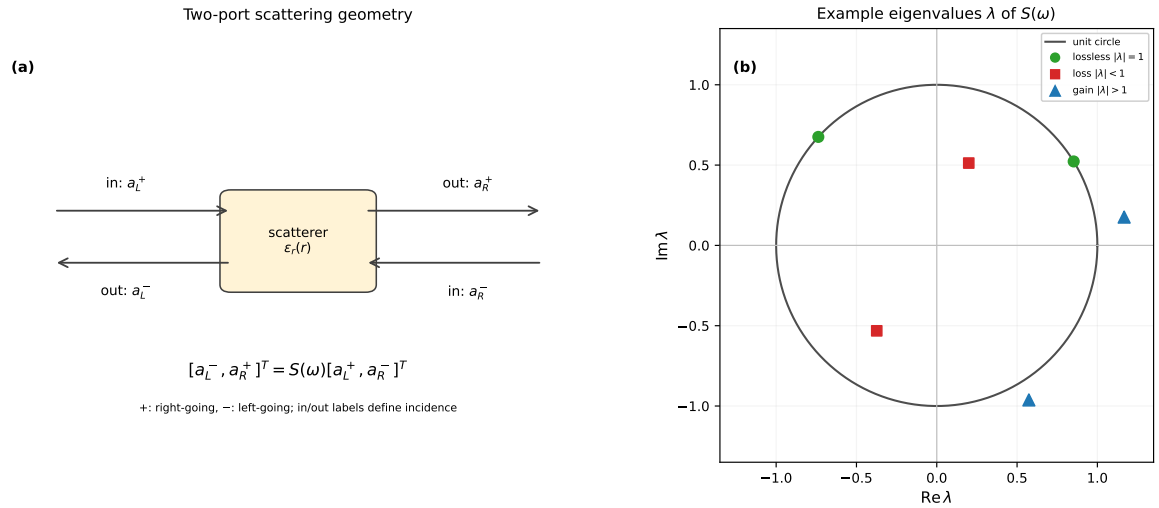


Figure 4.1 The scattering matrix as an input–output map. Panel (a) defines the two-channel amplitudes used in Eq. (4.6). Panel (b) shows the corresponding constraint on scattering eigenvalues: lossless systems place them on the unit circle, passive loss moves them inside, and gain can move them outside. This is the scattering version of the distinction between Hermitian and non-Hermitian Maxwell operators.

4.3.2 Unitarity and energy conservation

For a lossless scatterer (real ε_r everywhere), conservation of the time-averaged Poynting flux demands that $\mathbf{S}$ be unitary:

$$\mathbf{S}^\dagger \mathbf{S} = \mathbf{I}. \tag{4.7}$$

This is the scattering-matrix analogue of the self-adjointness of $\hat{\Theta}$. Unitarity means that the total outgoing power equals the total incoming power: no energy is created or destroyed inside the scatterer.

For a scatterer with loss, unitarity is violated in a specific direction:

$$\mathbf{S}^\dagger\mathbf{S} \leq \mathbf{I} \qquad \text{(sub-unitary: lossy)}, \tag{4.8}$$

meaning the eigenvalues of $\mathbf{S}^\dagger\mathbf{S}$ are all ≤ 1. The deficit represents absorbed power. For a scatterer with gain, $\mathbf{S}^\dagger\mathbf{S}$ can have eigenvalues > 1: the scatterer amplifies the incident field. The transition from sub-unitary to super-unitary $\mathbf{S}$ as gain is increased corresponds to the transition from a passive absorber to an active amplifier, and eventually to a laser when a pole of $\mathbf{S}$ reaches the real frequency axis.

4.3.3 Reciprocity in scattering

For a reciprocal medium (symmetric ε_r tensor), the scattering matrix satisfies the symmetry

$$\mathbf{S}^\mathsf{T} = \mathbf{S}, \tag{4.9}$$

after appropriate normalization of the channel amplitudes. As discussed in Section 2.6, reciprocity is distinct from unitarity (Hermiticity of the Hamiltonian). A lossy reciprocal scatterer has $\mathbf{S}^\mathsf{T} = \mathbf{S}$ but $\mathbf{S}^\dagger\mathbf{S} \neq \mathbf{I}$.

The distinction is important because it constrains which non-Hermitian effects are possible in reciprocal media. For example, reciprocity guarantees $t_L = t_R$ (equal transmission from both sides), even when the system has gain and loss that break unitarity and produce $r_L \neq r_R$ (unequal reflection). The directional reflection asymmetry exploited by $\mathcal{PT}$-symmetric devices (Chapter 9) is therefore consistent with reciprocity.

4.3.4 The optical theorem

For a single incoming channel, the total scattering cross-section σ_{sc} and the extinction cross-section σ_{ext} are related by the optical theorem:

$$\sigma_{\text{ext}} = \frac{4\pi}{k}\,\text{Im}\big[f(0)\big], \tag{4.10}$$

where $f(0)$ is the forward scattering amplitude and $k = \omega/c$. Extinction equals the sum of scattering and absorption: $\sigma_{\text{ext}} = \sigma_{\text{sc}} + \sigma_{\text{abs}}$. For a passive, lossless scatterer, $\sigma_{\text{abs}} = 0$ and extinction is purely due to scattering.

The optical theorem is a consequence of unitarity and can be derived directly from the **S**-matrix formalism. It fails for active (amplifying) scatterers, where σ_{ext} can become negative—the scattered field interferes constructively with the incident field in the forward direction, *increasing* the total power. This is a scattering-theory signature of gain.

4.4 Poles, zeros, and resonances

The scattering matrix $\mathbf{S}(\omega)$ is a meromorphic function of the complex frequency ω. Its singularity structure encodes the complete physics of the open system.

4.4.1 Poles: resonances and quasinormal modes

A *pole* of $\mathbf{S}(\omega)$ is a complex frequency $\tilde{\omega}_n$ at which the scattering matrix diverges. Physically, this means a nonzero outgoing field exists with no incoming field—the system radiates spontaneously. The poles are the resonances (quasinormal-mode frequencies) of the open system [5].

For a passive system with no gain, all poles lie in the lower half of the complex ω plane ($\text{Im}\,\tilde{\omega} < 0$), corresponding to decaying modes. A pole on the real axis would represent a mode with infinite lifetime—a bound state or, in special circumstances, a bound state in the continuum (Chapter 16).

Near a single isolated pole, the scattering matrix has the form

$$\mathbf{S}(\omega) \approx \mathbf{S}_{\text{bg}} + \frac{\mathbf{A}_n}{\omega - \tilde{\omega}_n}, \tag{4.11}$$

where $\mathbf{S}_{\text{bg}}$ is a slowly varying background and $\mathbf{A}_n$ is the residue matrix, whose rank equals the multiplicity of the pole. This Breit–Wigner form [6, 22, 23] is the foundation of temporal coupled-mode theory (Chapter 7).

4.4.2 Zeros: coherent perfect absorption

A *zero* of $\mathbf{S}(\omega)$—more precisely, a zero of $\det \mathbf{S}(\omega)$—is a frequency at which a specific combination of incoming waves produces zero outgoing field. All incident energy is absorbed. This is *coherent perfect absorption* (CPA) [24, 25], the time-reverse of lasing. The interplay of poles and zeros under $\mathcal{PT}$ symmetry is the subject of Chapter 10.

For a two-channel system, a zero of $\det \mathbf{S}$ requires both the absorption condition ($\operatorname{Im}\varepsilon > 0$) and the correct phase relationship between the two input beams. The CPA condition is thus doubly constrained—in both frequency and input-phase space—making it more experimentally demanding than simple absorption.

4.4.3 Pole–zero structure and causality

Causality constrains the analytic structure of $\mathbf{S}(\omega)$: in the upper half of the complex ω plane ($\operatorname{Im}\omega > 0$), $\mathbf{S}$ must be analytic and bounded for a passive system. Poles can only appear in the lower half-plane. If gain is introduced, poles can migrate upward; when a pole crosses the real axis, the system reaches the lasing threshold.

The zeros of $\det \mathbf{S}$ for a passive system lie in the upper half-plane (by the time-reversal argument: if the poles of $\mathbf{S}$ are in the lower half-plane, the poles of $\mathbf{S}^{-1}$—which are the zeros of $\det \mathbf{S}$—are in the upper half-plane for a time-reversal-symmetric system). This pole-zero separation is the scattering-theory version of passivity [26]. Chapter 10 shows how gain brings poles and zeros together, culminating in the CPA-laser point where a pole and a zero meet on the real axis.

4.5 The Fabry–Pérot resonator

The simplest and most instructive open resonator is a one-dimensional Fabry–Pérot cavity: two partial mirrors separated by a distance L filled with a medium of refractive index n. Despite its simplicity, the Fabry–Pérot cavity contains, in miniature, all the essential features of open-system scattering: complex eigenfrequencies, a meromorphic $\mathbf{S}$, and a connection between poles and transmission resonances.

4.5.1 Transfer matrix and scattering matrix

Each mirror is characterized by reflection and transmission coefficients r_j, t_j ($j = 1, 2$). The *transfer matrix* $\mathbf{T}$ relates the field amplitudes on the left to those on the right through

$$\mathbf{T} = \mathbf{T}_{\text{mirror},1}\, \mathbf{T}_{\text{prop}}(L)\, \mathbf{T}_{\text{mirror},2}, \tag{4.12}$$

where $\mathbf{T}_{\text{prop}}(L) = \left(\begin{smallmatrix} e^{in\omega L/c} & 0 \\ 0 & e^{-in\omega L/c} \end{smallmatrix}\right)$ accounts for propagation over distance L and each mirror matrix encodes the partial reflection and transmission. The scattering matrix is obtained from $\mathbf{T}$ by a standard change of basis (Appendix C).

For symmetric mirrors ($r_1 = r_2 = r$, $t_1 = t_2 = t$), the transmission coefficient is

$$t_{\text{FP}}(\omega) = \frac{(1 - |r|^2)\, e^{in\omega L/c}}{1 - r^2\, e^{2in\omega L/c}}, \tag{4.13}$$

and the power transmission is the Airy function:

$$T(\omega) = |t_{\mathrm{FP}}|^2 = \frac{(1-|r|^2)^2}{1-2|r|^2\cos(2n\omega L/c)+|r|^4} = \frac{1}{1+F\sin^2(n\omega L/c)}, \tag{4.14}$$

where the *coefficient of finesse* is

$$F = \frac{4|r|^2}{(1-|r|^2)^2}. \tag{4.15}$$

The finesse $\mathcal{F} = \pi\sqrt{F}/2 = \pi|r|/(1-|r|^2)$ measures the sharpness of the transmission peaks relative to the free spectral range $\Delta\omega_{\mathrm{FSR}} = \pi c/(nL)$.

4.5.2 Resonance condition and complex eigenfrequencies

The poles of the scattering matrix correspond to the condition

$$r^2\, e^{2in\tilde{\omega}L/c} = 1, \tag{4.16}$$

which has solutions at complex frequencies

$$\tilde{\omega}_m = \frac{m\pi c}{nL} - \frac{ic\ln(1/|r|^2)}{2nL}, \qquad m = 1, 2, 3, \ldots \tag{4.17}$$

for real r with $|r| < 1$. The real part gives equally spaced resonant frequencies (the free spectral range), and the imaginary part gives the decay rate $\gamma = c\ln(1/|r|)/(nL)$. The quality factor is

$$Q_m = \frac{\omega_m}{2\gamma} = \frac{m\pi}{2\ln(1/|r|)}. \tag{4.18}$$

Higher-order modes ($m \gg 1$) have higher Q because ω_m grows while γ stays fixed. For high reflectivity ($|r| \to 1$), $\gamma \to 0$ and $Q \to \infty$: the cavity approaches the closed-cavity limit, and the eigenfrequencies approach the real axis.

Example. A silicon slab in air with $n = 3.5$ and $L = 300$ nm has Fresnel reflectivity $|r| = |n-1|/(n+1) = 0.556$, giving $\gamma \approx 1.68 \times 10^{14}$ rad/s and $Q_1 \approx 2.67$. This very low quality factor reflects the poor confinement of a bare dielectric slab. Adding distributed Bragg reflectors increases $|r|$ and pushes the poles closer to the real axis: the quasinormal-mode picture recovers smoothly the transition from a leaky slab to a high-finesse cavity. Chapter 6 develops this example further and connects it to the full QNM pole structure.

4.5.3 Connection between poles and transmission peaks

The transmission peaks of the Airy function (4.14) occur at real frequencies $\omega_m = m\pi c/(nL)$, which are the real parts of the poles (4.17). The full width at half maximum (FWHM) of each peak is $\delta\omega \approx 2\gamma$, which is twice the imaginary part of the pole. Thus the poles of $\mathbf{S}$ directly determine the observable transmission spectrum: their real parts give the peak positions and their imaginary parts give the peak widths.

This connection between complex poles and real-frequency observables is general: for any open resonator, the scattering spectrum near a resonance is a Lorentzian (or Fano) line shape whose parameters are set by the nearest pole of $\mathbf{S}$. The full scattering response is the superposition

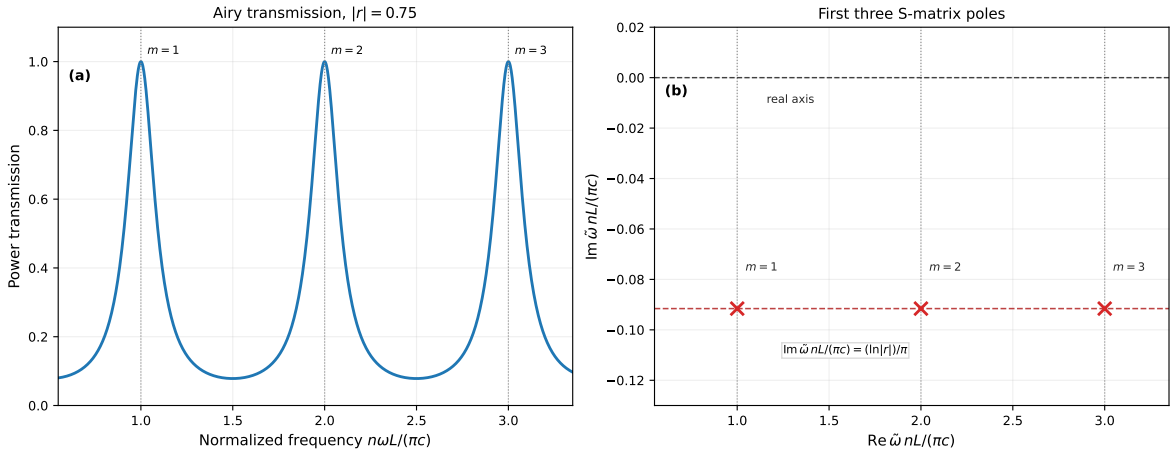


Figure 4.2 The same Fabry–Pérot resonances seen on the real axis and in the complex plane. Panel (a) shows the Airy transmission peaks as a function of round-trip phase. Panel (b) shows the corresponding scattering poles: the real parts set the peak positions and the imaginary part sets the linewidth.

of contributions from all poles, weighted by their residues (the Mittag-Leffler expansion of Chapter 10).

The Fabry–Pérot as a non-Hermitian paradigm

The Fabry–Pérot resonator is the simplest example of how openness creates non-Hermiticity. The interior of the cavity is lossless (real n), yet the eigenfrequencies are complex because energy leaks through the partially reflecting mirrors. This is a concrete realization of the second of the four roads to non-Hermiticity identified in Section 1.3.3: the boundary conditions, not the material, break self-adjointness. The pole structure of $\mathbf{S}$ provides a complete description of the resonator's spectral response, replacing the discrete real eigenvalues of a closed cavity with a constellation of complex poles in the lower half-plane.

4.6 Scattering cross-sections and the Wigner time delay

4.6.1 Cross-sections from the S matrix

For a compact scatterer in free space, the total scattering cross-section σ_{sc} and the absorption cross-section σ_{abs} can be expressed in terms of the eigenvalues $s_n(\omega)$ of the $\mathbf{S}$ matrix:

$$\sigma_{\mathrm{sc}} = \frac{\pi}{k^2} \sum_n |1 - s_n|^2, \tag{4.19}$$

$$\sigma_{\mathrm{abs}} = \frac{\pi}{k^2} \sum_n (1 - |s_n|^2). \tag{4.20}$$

For a unitary $\mathbf{S}$, $|s_n| = 1$ and $\sigma_{\mathrm{abs}} = 0$: no absorption occurs, as expected for a lossless scatterer.

The extinction cross-section is $\sigma_{\mathrm{ext}} = \sigma_{\mathrm{sc}} + \sigma_{\mathrm{abs}}$, and the optical theorem (4.10) relates it to the forward scattering amplitude. Near a resonance at $\tilde{\omega}_n$, the extinction cross-section exhibits a Lorentzian peak:

$$\sigma_{\mathrm{ext}}(\omega) \approx \sigma_0 \frac{\gamma_n^2}{(\omega - \omega_n)^2 + \gamma_n^2}, \tag{4.21}$$

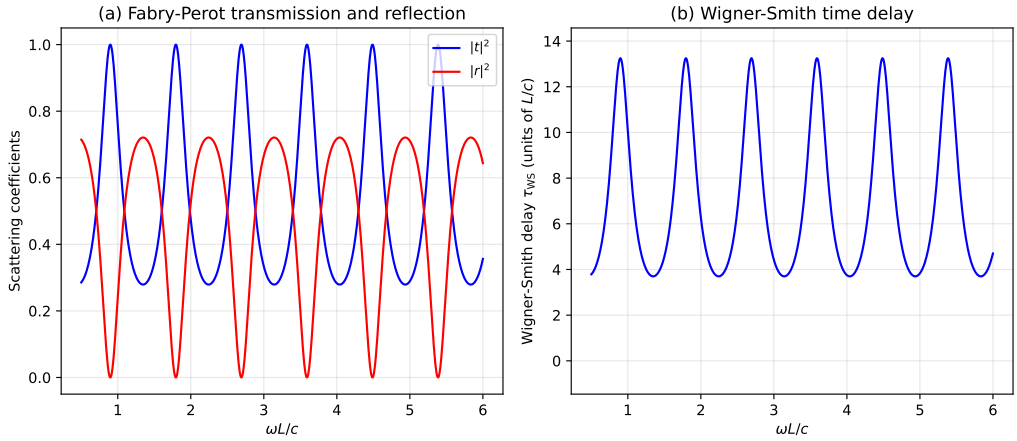


Figure 4.3 Wigner–Smith time delay for a Fabry–Pérot cavity ($n = 3.5$, $L = 1$). (a) Transmission $|t|^2$ and reflection $|r|^2$ showing the characteristic resonance peaks of the etalon. (b) The Wigner–Smith delay $\tau_{\mathrm{WS}} = \mathrm{d}(\arg\det \mathbf{S})/\,\mathrm{d}\omega$ peaks sharply at each resonance, confirming that high-Q modes trap light for long times. Between resonances the delay is small, reflecting the short dwell time of off-resonant wave packets.

where σ_0 is the peak cross-section, which can exceed the geometric cross-section by a factor of order Q.

4.6.2 The Wigner time delay

The $\mathbf{S}$ matrix also characterizes the temporal response of the scatterer through the Wigner–Smith time-delay matrix [2, 27]:

$$Q_{\mathrm{WS}} = -i\,\mathbf{S}^{\dagger}\,\frac{\mathrm{d}\mathbf{S}}{\mathrm{d}\omega}. \tag{4.22}$$

For a unitary $\mathbf{S}$, Q_{WS} is Hermitian, and its eigenvalues $\tau_n(\omega)$ are real time delays [23]—the durations for which an incoming wave packet is trapped inside the scatterer before being released. Near a resonance, the dominant eigenvalue of Q_{WS} is of order the resonance lifetime. With the convention $\tilde{\omega}_n = \omega_n - i\gamma_n$, the stored energy decays at rate $2\gamma_n$, so the energy lifetime is $\tau_E = 1/(2\gamma_n) = Q_n/\omega_n$. Phase-delay conventions can differ by factors of order unity, but the scaling is the same: high-Q resonances trap light for long times.

When the $\mathbf{S}$ matrix is non-unitary (lossy or amplifying scatterer), Q_{WS} is no longer Hermitian, and the time delays become complex. The imaginary part of the time delay characterizes the amplification or absorption experienced by the trapped wave packet. This non-Hermitian time delay is an observable quantity that can be measured interferometrically and provides a direct experimental probe of the non-Hermitian character of the scattering process.

4.7 Spectrum of the open Maxwell operator

For a general open system—not just a Fabry–Pérot cavity—the spectrum of the Maxwell operator with outgoing boundary conditions has a richer structure than the discrete, real spectrum of a closed cavity [5].

Discrete resonances. The complex eigenfrequencies $\tilde{\omega}_n$ (poles of $\mathbf{S}$) form a discrete set in the lower half of the complex ω plane. Near each pole, the scattered field is dominated by the

corresponding quasinormal mode, whose spatial profile diverges exponentially at infinity (the mode was excited in the past and its radiation wavefront keeps expanding).

Continuous spectrum. In addition to the discrete resonances, the open operator has a *continuous spectrum* on the real frequency axis, corresponding to radiation modes that propagate freely in the surrounding medium. These modes are not square-integrable and cannot be normalized in the same Hilbert space as bound states.

Bound states and guided modes. If the scatterer supports guided modes or bound states (fields that decay exponentially in all directions without radiation), these appear as discrete, real eigenfrequencies. They coexist with the continuous spectrum. In certain symmetry-protected situations, a discrete real eigenvalue can be embedded *within* the continuous spectrum—a bound state in the continuum (Chapter 16).

Branch cuts. In the complex ω plane, the continuous spectrum produces branch cuts rather than isolated singularities. The choice of branch cut (typically along the negative imaginary axis or the real axis) determines which Riemann sheet the resonance poles are defined on. This subtlety becomes important for the mathematical theory of quasinormal-mode completeness (Chapter 6).

The full spectral decomposition of the field in an open system requires all three components: the discrete resonances, the continuous radiation modes, and any bound states [5]. This is why the modal theory of open photonic systems is fundamentally different from that of closed cavities: the resonances alone do not form a complete basis, and the standard orthogonality relations fail.

4.8 Perfectly matched layers

4.8.1 Complex coordinate stretching

Numerical computation of resonances requires truncating the infinite domain to a finite computational region. Simply truncating creates artificial reflections at the boundary. *Perfectly matched layers* (PMLs) solve this problem by introducing a complex coordinate stretching in the outer region that absorbs outgoing waves without reflection.

In one dimension, the PML replaces the physical coordinate x with

$$\tilde{x}(x) = x + \frac{i}{\omega}\int_0^x \sigma(x')\ \mathrm{d}x', \tag{4.23}$$

where $\sigma(x) \geq 0$ is a smooth absorption profile that vanishes inside the physical domain and grows in the PML region. A plane wave e^{ikx} in the physical region becomes $e^{ik\tilde{x}} = e^{ikx}\exp\!\big(-k\int_0^x \sigma(x')\,\mathrm{d}x'/\omega\big)$ in the PML region—exponentially decaying for outgoing waves ($k > 0$) and exponentially growing for incoming waves ($k < 0$). Since only outgoing waves are present (by the radiation condition), the PML absorbs all radiation without reflection.

In three dimensions, the PML is implemented by stretching each coordinate independently:

$$\tilde{x}_i = x_i + \frac{i}{\omega}\int_0^{x_i} \sigma_i(x_i')\ \mathrm{d}x_i', \qquad i = 1, 2, 3. \tag{4.24}$$

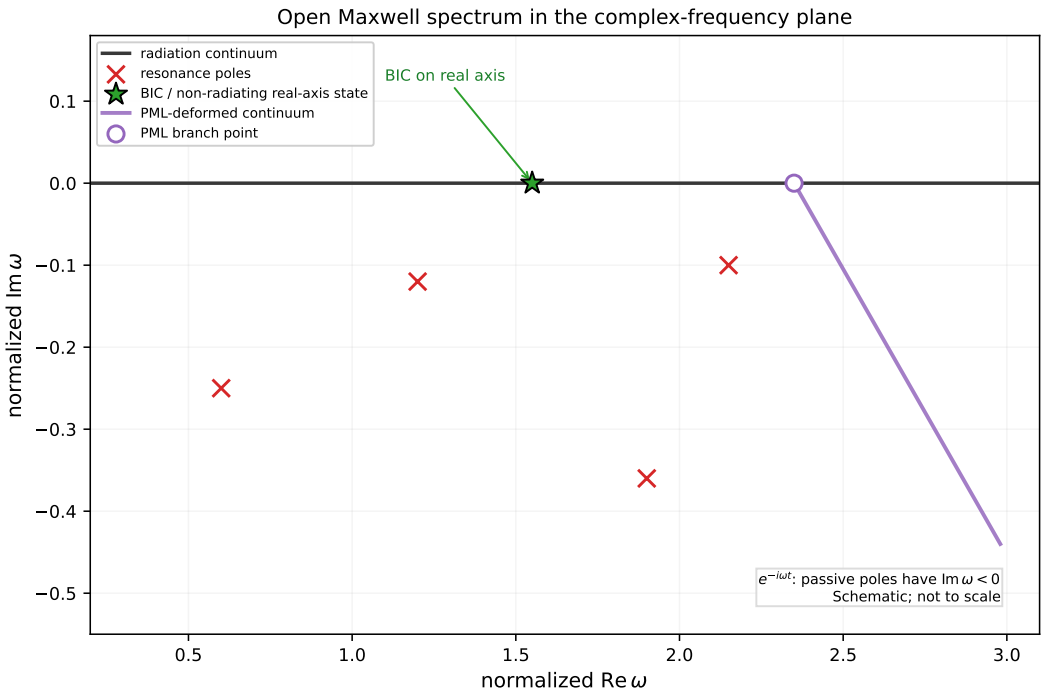


Figure 4.4 Open-system spectra in the complex-frequency plane. Passive resonances appear as poles below the real axis; radiation modes form a continuum on the real axis; and a guided mode or BIC can remain on the real axis without radiating. In a complex-scaled or PML-truncated formulation, the radiation continuum is represented by a rotated or discretized set of PML modes in the lower half-plane. The marked point indicates a continuum threshold or branch point of the radiation channel, not an additional resonance of the scatterer. This picture anticipates the quasinormal-mode theory of Chapter 6.

The resulting Maxwell equations in the stretched coordinates have modified permittivity and permeability tensors that incorporate the PML absorption.

4.8.2 Spectral consequences of the PML

The PML has a profound consequence for the spectral theory [5]: the complex coordinate transformation maps the open Maxwell operator into a *non-self-adjoint operator* on a bounded domain. The continuous spectrum of the open problem is rotated into the complex plane by an angle $\theta_{\mathrm{PML}} = \arctan(\sigma_{\max}/\omega)$, becoming a discrete set of "PML modes" that approximate the radiation continuum. The true resonances (poles of $\mathbf{S}$) are unchanged by the PML—they are intrinsic to the scatterer—while the PML modes are artifacts of the truncation, their positions depending on the PML parameters.

This separation is the key to extracting physical information from PML computations: the physical resonances are identified as eigenvalues that are insensitive to changes in the PML parameters (thickness, absorption profile), while the PML modes shift when these parameters are varied. Chapter 17 discusses practical strategies for distinguishing physical resonances from PML artifacts.

4.8.3 PML as a non-Hermitian mapping

PML as a non-Hermitian tool

Perfectly matched layers do more than suppress reflections. They provide a rigorous non-Hermitian extension of the Maxwell operator that converts the spectral problem of an open system into a matrix eigenvalue problem on a bounded domain. The price is non-self-adjointness: the resulting operator has complex eigenvalues, non-orthogonal eigenvectors, and requires biorthogonal inner products for modal expansion [4]. This is the starting point for the quasinormal-mode theory of Chapter 6.

The PML construction also reveals a deep connection between open systems and non-Hermitian physics: the radiation boundary condition is *equivalent* to embedding the scatterer in a non-Hermitian (absorbing) environment. Whether we model openness as an outgoing-wave condition at infinity or as a PML-absorbing layer at a finite distance, the mathematical structure is the same: a non-self-adjoint operator with complex eigenvalues and biorthogonal modes. The physical content is also the same: energy leaves the system irreversibly.

4.9 Feshbach projection and effective non-Hermitian Hamiltonians

The connection between openness and non-Hermiticity can be made precise through the Feshbach projection formalism. Partition the full Hilbert space of the Maxwell operator into a "system" subspace $\mathcal{P}$ (e.g., the discrete modes of the cavity) and an "environment" subspace $\mathcal{Q}$ (the continuum of radiation modes):

$$\hat{P} + \hat{Q} = \mathbf{I}, \qquad \hat{P}\hat{Q} = 0. \tag{4.25}$$

The full eigenvalue problem $\hat{\Theta}|\psi\rangle = \lambda|\psi\rangle$, with $\lambda = \omega^2/c^2$, can be projected onto $\mathcal{P}$:

$$\left[\hat{P}\hat{\Theta}\hat{P} + \hat{P}\hat{\Theta}\hat{Q}\,(\lambda - \hat{Q}\hat{\Theta}\hat{Q})^{-1}\,\hat{Q}\hat{\Theta}\hat{P}\right]\hat{P}|\psi\rangle = \lambda\,\hat{P}|\psi\rangle. \tag{4.26}$$

The term in brackets is an *effective operator* on $\mathcal{P}$ that depends on the eigenvalue λ. The second term—the "self-energy" $\Sigma(\lambda) = \hat{P}\hat{\Theta}\hat{Q}\,(\lambda - \hat{Q}\hat{\Theta}\hat{Q})^{-1}\,\hat{Q}\hat{\Theta}\hat{P}$—accounts for the coupling to the radiation continuum. Because the resolvent $(\lambda - \hat{Q}\hat{\Theta}\hat{Q})^{-1}$ has poles on the real axis (the continuous spectrum), $\Sigma(\lambda)$ is complex-valued for real λ, and the effective operator is non-Hermitian.

The imaginary part of Σ gives the radiative decay rate of the cavity modes:

$$\gamma_n = -\,\mathrm{Im}\big[\Sigma_{nn}(\omega_n^2/c^2)\big], \tag{4.27}$$

while the real part gives a frequency shift (the Lamb shift analogue in electrodynamics). The Feshbach formalism thus provides a microscopic derivation of the complex eigenfrequencies (4.1): the imaginary part arises from the coupling between the cavity modes and the radiation continuum, not from any material absorption.

This is the fourth road to non-Hermiticity: projection onto a subspace. Even when the full Maxwell operator is self-adjoint, the effective operator on a reduced subspace is generically non-Hermitian because the environment's degrees of freedom carry away energy and information. Chapter 7 develops this idea into the full temporal coupled-mode theory of photonic resonators.

4.10 From scattering to non-Hermitian spectral theory

The preceding sections introduced the physical and mathematical framework that connects closed, Hermitian photonics to the non-Hermitian world:

- Outgoing-wave boundary conditions break self-adjointness by allowing energy to escape through the surface term in the Poynting identity.
- The scattering matrix $\mathbf{S}(\omega)$, a meromorphic function of complex frequency, encodes resonances (poles) and absorption conditions (zeros).
- Unitarity of $\mathbf{S}$ is the scattering analogue of Hermiticity; loss or gain breaks unitarity, and the eigenvalues of $\mathbf{S}$ leave the unit circle.
- The Fabry–Pérot resonator illustrates how purely radiative loss—without material absorption—creates complex eigenfrequencies. The Airy transmission function, the quality factor, and the finesse are all determined by the pole structure of $\mathbf{S}$.
- Cross-sections and the Wigner time delay provide measurable quantities that are directly expressed in terms of the $\mathbf{S}$-matrix eigenvalues and their frequency dependence.
- Perfectly matched layers provide a rigorous non-Hermitian reformulation that enables numerical computation of resonances and modal expansions.
- The Feshbach projection formalism shows that non-Hermiticity arises whenever a self-adjoint system is described by a reduced effective operator on a subspace.

Chapter 5 now introduces the complementary road to non-Hermiticity: material loss, gain, and frequency dispersion. Chapter 6 develops the full theory of quasinormal modes—the eigenstates of the non-self-adjoint Maxwell operator with outgoing boundary conditions—including their normalization, biorthogonality, and physical applications.

Part II

Physical Origins of Non-Hermiticity

Chapter 5

Loss, Gain, and Dispersion

The first road to non-Hermiticity does not require opening the system or projecting onto a subspace. It is already present inside the material: every real dielectric absorbs at some frequencies, every laser medium amplifies at others, and every medium disperses. This chapter traces the physical origin of complex permittivity from linear response theory, derives the constraints that causality and passivity impose through the Kramers–Kronig relations, explains why the simple energy formula of Chapter 2 fails for dispersive media, and introduces the auxiliary-field construction that restores a conservative Hamiltonian [2] description.

5.1 Complex permittivity from linear response

In Chapter 2, we treated $\varepsilon_{\mathrm{r}}(\mathbf{r})$ as a real, frequency-independent quantity. This is an idealization. A real material responds to an applied electric field through the polarization of its constituent charges, and this response takes time. The displacement field at time t depends on the electric field at all earlier times:

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0 \mathbf{E}(\mathbf{r},t) + \varepsilon_0 \int_0^\infty \chi(\mathbf{r},\tau)\, \mathbf{E}(\mathbf{r},t-\tau)\ \mathrm{d}\tau, \tag{5.1}$$

where $\chi(\mathbf{r},\tau)$ is the causal susceptibility kernel, with $\chi(\mathbf{r},\tau) = 0$ for $\tau < 0$ by causality. In the frequency domain, the convolution becomes multiplication:

$$\mathbf{D}(\mathbf{r},\omega) = \varepsilon_0\, \varepsilon_{\mathrm{r}}(\mathbf{r},\omega)\, \mathbf{E}(\mathbf{r},\omega), \qquad \varepsilon_{\mathrm{r}}(\mathbf{r},\omega) = 1 + \hat{\chi}(\mathbf{r},\omega), \tag{5.2}$$

where $\hat{\chi}(\omega) = \int_0^\infty \chi(\tau)\, e^{i\omega\tau}\ \mathrm{d}\tau$ is the Fourier–Laplace transform of the susceptibility.

For any material with a nontrivial temporal response, $\varepsilon_{\mathrm{r}}(\omega)$ is complex:

$$\varepsilon_{\mathrm{r}}(\omega) = \varepsilon_{\mathrm{r}}'(\omega) + i\varepsilon_{\mathrm{r}}''(\omega). \tag{5.3}$$

The imaginary part $\varepsilon_{\mathrm{r}}''$ describes absorption ($\varepsilon_{\mathrm{r}}'' > 0$ with our $e^{-i\omega t}$ convention) or gain ($\varepsilon_{\mathrm{r}}'' < 0$). The Maxwell eigenproblem now has a complex-valued weight, and the self-adjointness proof of Section 2.4 fails at the precise point identified in the remark following Theorem 2.1: the factor $\varepsilon_{\mathrm{r}}^{-1}$ is no longer real.

5.1.1 Symmetry constraints on $\varepsilon_{\mathrm{r}}(\omega)$

The susceptibility kernel $\chi(\mathbf{r},\tau)$ is a real-valued function of time (it relates real fields to real displacements). This reality condition implies

$$\varepsilon_{\mathrm{r}}(\mathbf{r},-\omega)=\varepsilon_{\mathrm{r}}^{*}(\mathbf{r},\omega), \tag{5.4}$$

known as the *crossing symmetry* or reality condition. It means that $\varepsilon_{\mathrm{r}}'(\omega)$ is even in ω and $\varepsilon_{\mathrm{r}}''(\omega)$ is odd in ω. At $\omega=0$, the permittivity is purely real.

5.2 The Lorentz oscillator model

5.2.1 Single resonance

The simplest and most instructive microscopic model of dispersion is the *Lorentz oscillator*: a bound charge of mass m and charge $-e$ [28] attached to a nucleus by a restoring force $-m\omega_0^2 x$ and subject to a damping force $-m\gamma\dot{x}$. The equation of motion under a harmonic driving field $E\,e^{-i\omega t}$ gives the polarizability

$$\alpha(\omega)=\frac{e^2/m}{\omega_0^2-\omega^2-i\gamma\omega}, \tag{5.5}$$

and, for a medium with N oscillators per unit volume,

$$\varepsilon_{\mathrm{r}}(\omega)=1+\frac{\omega_p^2}{\omega_0^2-\omega^2-i\gamma\omega}, \tag{5.6}$$

where $\omega_p^2=Ne^2/(\varepsilon_0 m)$ is the plasma frequency squared.

The real and imaginary parts are

$$\varepsilon_{\mathrm{r}}'(\omega)=1+\frac{\omega_p^2(\omega_0^2-\omega^2)}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}, \tag{5.7}$$

$$\varepsilon_{\mathrm{r}}''(\omega)=\frac{\omega_p^2\,\gamma\omega}{(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}. \tag{5.8}$$

Several features of (5.6) are universal, not specific to the Lorentz model:

1. $\varepsilon_{\mathrm{r}}''(\omega)>0$ for all $\omega>0$ when $\gamma>0$: a passive medium always absorbs.
2. Near resonance ($\omega\approx\omega_0$), the absorption $\varepsilon_{\mathrm{r}}''$ peaks and $\varepsilon_{\mathrm{r}}'$ passes through a region of *anomalous dispersion* where $\mathrm{d}\varepsilon_{\mathrm{r}}'/\,\mathrm{d}\omega<0$.
3. At high frequencies ($\omega\gg\omega_0$), $\varepsilon_{\mathrm{r}}(\omega)\to 1-\omega_p^2/\omega^2$: the medium becomes transparent and weakly dispersive.

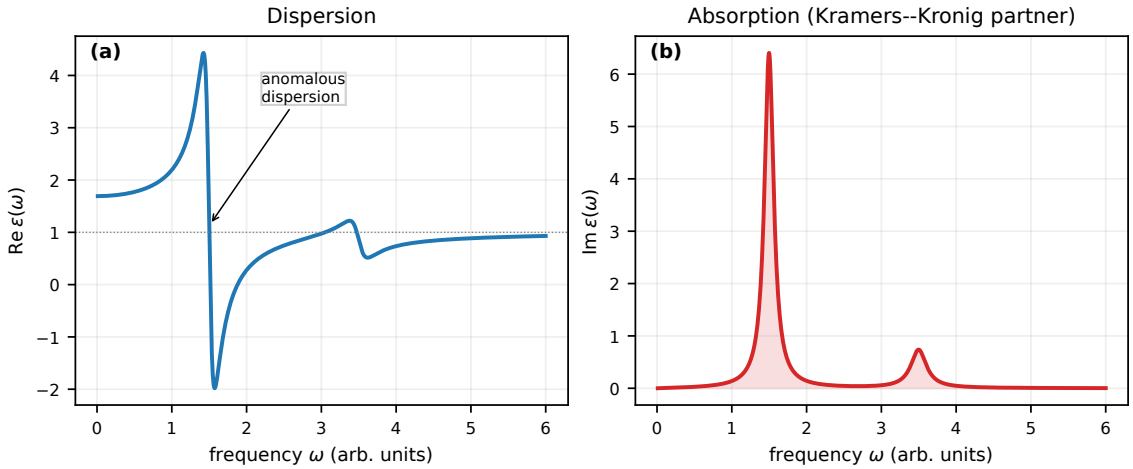


Figure 5.1 Complex permittivity of a two-resonance Lorentz model ($\omega_{01} = 1.5$, $\omega_{02} = 3.5$; damping $\gamma_1 = 0.15$, $\gamma_2 = 0.25$; arbitrary units). (a) Real part $\mathrm{Re}\,\varepsilon(\omega)$ showing normal and anomalous dispersion near each resonance; the dotted line marks the high-frequency limit $\varepsilon = 1$. (b) Imaginary part $\mathrm{Im}\,\varepsilon(\omega)$, the Kramers–Kronig partner of (a). Each absorption peak in (b) corresponds to a region of anomalous dispersion in (a), as required by causality. For a passive medium, $\mathrm{Im}\,\varepsilon > 0$ at all positive frequencies.

5.2.2 Multiple resonances and the Drude model

A real material has multiple absorption lines at frequencies ω_{0j}, each with its own oscillator strength f_j and damping γ_j. The total permittivity is

$$\varepsilon_{\mathrm{r}}(\omega) = 1 + \sum_j \frac{f_j\,\omega_p^2}{\omega_{0j}^2 - \omega^2 - i\gamma_j\omega}, \qquad \sum_j f_j = 1. \tag{5.9}$$

The sum rule $\sum_j f_j = 1$ (Thomas–Reiche–Kuhn sum rule) [29] is a consequence of the commutation relation $[x, p] = i\hbar$ and ensures that the high-frequency behavior is $\varepsilon_{\mathrm{r}}(\omega) \to 1 - \omega_p^2/\omega^2$ regardless of the number of resonances.

In metals and doped semiconductors, free carriers have no restoring force ($\omega_0 = 0$), and the corresponding contribution is the *Drude* model:

$$\varepsilon_{\mathrm{Drude}}(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma_D)}, \tag{5.10}$$

where γ_D is the Drude damping rate (scattering rate). At low frequencies, $\mathrm{Im}\,\varepsilon_{\mathrm{Drude}} \propto 1/\omega$: the absorption diverges, reflecting the dc conductivity $\sigma_0 = \varepsilon_0\omega_p^2/\gamma_D$. At optical frequencies, the Drude response produces a negative real permittivity for $\omega < \omega_p$, which is the origin of the metallic reflectivity and supports surface plasmon polaritons.

5.3 Kramers–Kronig relations

5.3.1 Derivation from causality

The causality of the susceptibility kernel—$\chi(\tau) = 0$ for $\tau < 0$—implies that $\hat{\chi}(\omega)$ is analytic in the upper half of the complex ω plane (UHP). The derivation proceeds as follows:

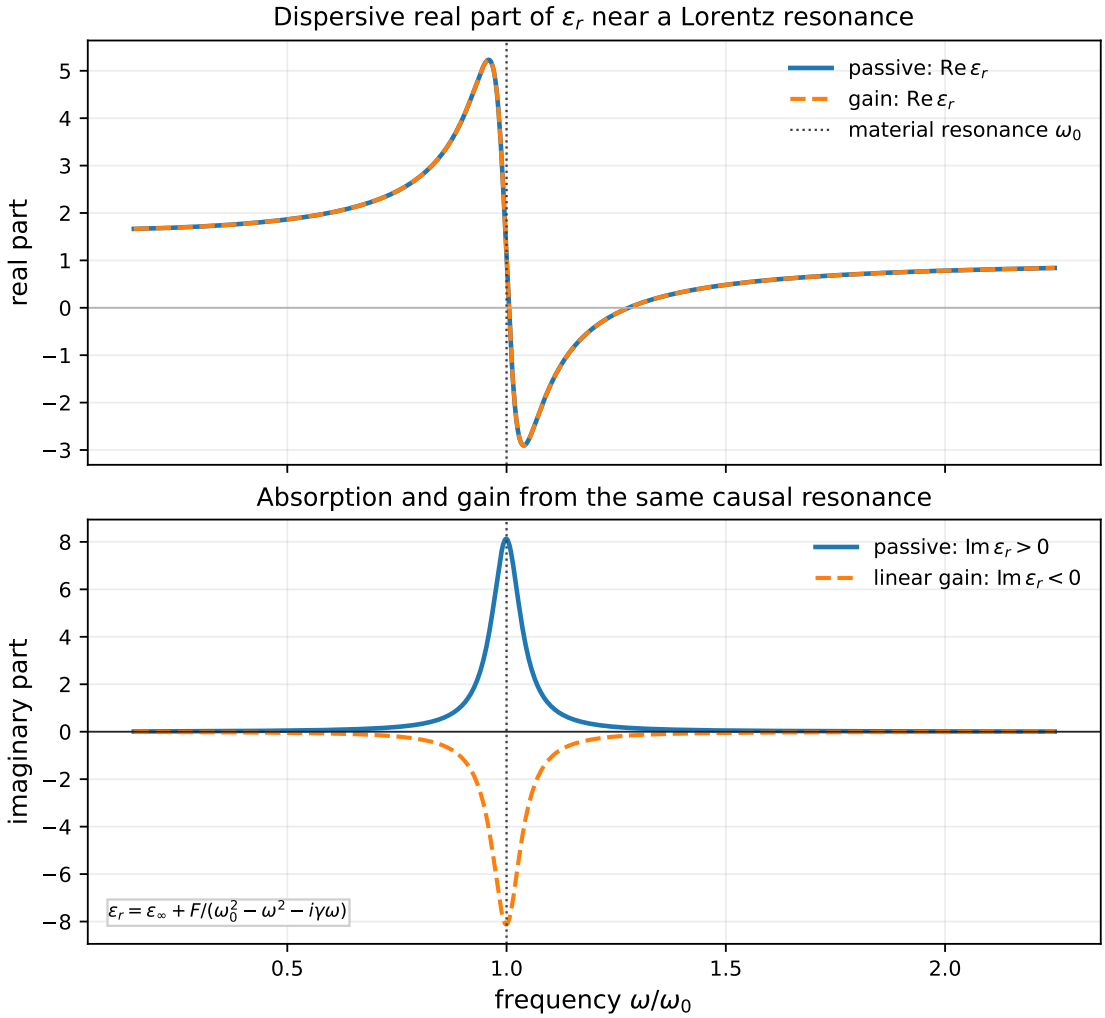


Figure 5.2 Lorentz dispersion as the simplest Maxwell material model of loss and gain. The same resonance produces a dispersive real permittivity and a peaked imaginary response. With the $e^{-i\omega t}$ convention, positive damping gives $\mathrm{Im}\,\varepsilon_\mathrm{r} > 0$ (absorption), while the sign-reversed linear model gives $\mathrm{Im}\,\varepsilon_\mathrm{r} < 0$ (gain).

Consider the integral of $\hat{\chi}(\omega')/(\omega' - \omega)$ around a closed contour $\mathcal{C}$ in the UHP consisting of: (i) the real axis from $-R$ to R with a small semicircular indent of radius ϵ above the pole at $\omega' = \omega$, and (ii) a large semicircle of radius R in the UHP.

Since $\hat{\chi}(\omega')$ is analytic in the UHP (by causality) and vanishes on the large semicircle as $R \to \infty$ (by the Riemann–Lebesgue lemma, since $\chi(\tau)$ is integrable), Cauchy's integral theorem gives

$$\oint_{\mathcal{C}} \frac{\hat{\chi}(\omega')}{\omega' - \omega}\, \mathrm{d}\omega' = 0. \tag{5.11}$$

The contributions from the real axis give the principal-value integral, and the small semicircular indent contributes $-i\pi\hat{\chi}(\omega)$. Setting the total to zero and separating real and imaginary parts

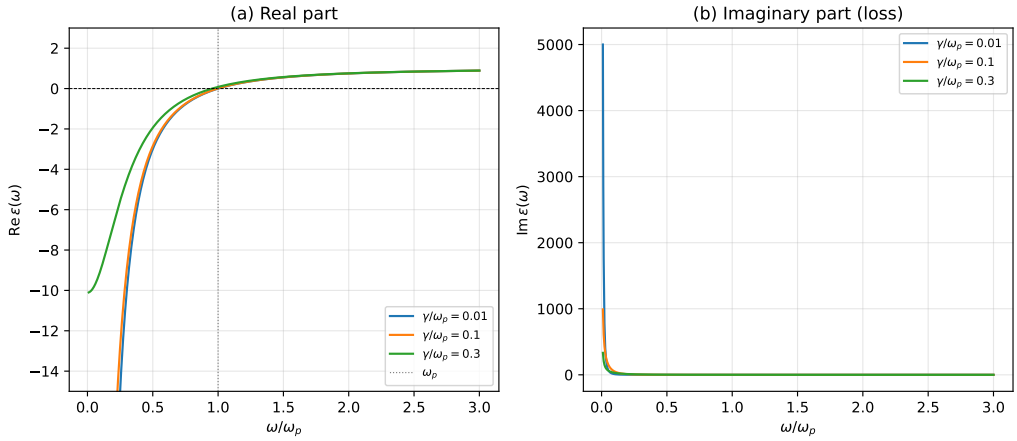


Figure 5.3 Drude permittivity for a free-electron metal. (a) Re $\varepsilon(\omega)$ is negative below the plasma frequency ω_p (vertical dotted line) and approaches $+1$ at high frequencies. Increasing the damping rate γ smooths the divergence near $\omega = 0$. (b) Im $\varepsilon(\omega)$ is peaked at low frequencies where ohmic absorption dominates. The interplay of negative Re ε and finite Im ε is the origin of plasmonic resonances in metallic nanostructures.

yields the *Kramers–Kronig relations* [30]:

$$\varepsilon_{\mathrm{r}}'(\omega) - 1 = \frac{1}{\pi}\,\mathcal{P}\!\int_{-\infty}^{\infty} \frac{\varepsilon_{\mathrm{r}}''(\omega')}{\omega' - \omega}\;\mathrm{d}\omega', \tag{5.12}$$

$$\varepsilon_{\mathrm{r}}''(\omega) = -\frac{1}{\pi}\,\mathcal{P}\!\int_{-\infty}^{\infty} \frac{\varepsilon_{\mathrm{r}}'(\omega') - 1}{\omega' - \omega}\;\mathrm{d}\omega'. \tag{5.13}$$

These relations are not a model: they are exact consequences of causality and linearity.

5.3.2 Physical consequences

Dispersion and absorption are inseparable

The Kramers–Kronig relations show that the real and imaginary parts of $\varepsilon_{\mathrm{r}}(\omega)$ are not independent. A medium that absorbs at some frequencies *must* disperse at all frequencies, and vice versa. A frequency-independent but lossy ε_{r} violates causality. This constraint is important for non-Hermitian photonics: any model with complex ε_{r} must be checked for Kramers–Kronig consistency, or it describes an unphysical material.

5.3.3 Sum rules

Specializing the Kramers–Kronig relations to $\omega = 0$ or $\omega \to \infty$ yields useful sum rules. The f-sum rule connects the total absorption to the electron density:

$$\int_0^{\infty} \omega\, \varepsilon_{\mathrm{r}}''(\omega)\;\mathrm{d}\omega = \frac{\pi}{2}\,\omega_p^2, \tag{5.14}$$

where ω_p is the plasma frequency. This result is exact for any causal response function and provides a powerful check on material models and experimental data.

5.3.4 Passivity and the power dissipation condition

A passive medium absorbs more energy than it emits at every frequency. This imposes the *power dissipation condition*:

$$\omega\,\varepsilon_{\mathrm{r}}''(\omega) \geq 0 \qquad \text{for all real } \omega. \tag{5.15}$$

For positive frequencies, this means $\varepsilon_{\mathrm{r}}'' \geq 0$: absorption. For an active (gain) medium, $\varepsilon_{\mathrm{r}}'' < 0$ at some positive frequencies, violating passivity. Gain media are driven out of thermodynamic equilibrium by an external pump (optical or electrical), and the Kramers–Kronig relations must be applied to the full response including the pump-induced changes [1].

5.4 Energy in lossy and dispersive media

The time-averaged energy density (2.8) was derived for real, nondispersive ε_{r}. When $\varepsilon_{\mathrm{r}}(\omega)$ is complex and frequency-dependent, the energy stored in the medium is no longer given by $\frac{1}{4}\varepsilon_0\varepsilon_{\mathrm{r}}|\mathbf{E}|^2$.

5.4.1 The Brillouin formula

For a lossless dispersive medium, or for a low-loss medium observed over a transparent narrow band centered at frequency ω_0, the time-averaged energy density is approximately

$$\langle u\rangle_{\mathrm{disp}} = \frac{1}{4}\varepsilon_0\,\frac{\mathrm{d}}{\mathrm{d}\omega}\!\left[\omega\,\varepsilon_{\mathrm{r}}'(\omega)\right]\!\bigg|_{\omega_0}|\mathbf{E}|^2 + \frac{1}{4}\mu_0\,\frac{\mathrm{d}}{\mathrm{d}\omega}\!\left[\omega\,\mu_{\mathrm{r}}'(\omega)\right]\!\bigg|_{\omega_0}|\mathbf{H}|^2. \tag{5.16}$$

This is the *Brillouin formula*, widely used in textbooks. It reduces to the nondispersive result when ε_{r} is constant.

However, the Brillouin formula has severe limitations:

1. It assumes a transparent or weakly absorbing frequency window, so that a stored energy can be separated cleanly from irreversible heat.
2. Near an absorption resonance, using only the real part of a lossy permittivity can give misleading or even negative apparent "energy densities."
3. It accounts for material polarization energy only indirectly, through the frequency derivative of the constitutive parameters.
4. It cannot be used as an inner product for mode orthogonality, because it depends on ω and is not bilinear.

5.4.2 Example: energy density near a Lorentz resonance

For the single Lorentz oscillator (5.6) in the lossless limit $\gamma \to 0$, the Brillouin energy factor is

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\left[\omega\,\varepsilon_{\mathrm{r}}'(\omega)\right] = 1 + \omega_p^2\,\frac{\omega_0^2+\omega^2}{(\omega_0^2-\omega^2)^2}. \tag{5.17}$$

Away from resonance, this is positive and large, correctly capturing the additional energy stored in the material polarization. At the resonance itself the lossless formula diverges, signaling that

the transparent-medium approximation has broken down. With finite damping, a meaningful energy balance must include the oscillator and bath degrees of freedom explicitly; substituting only $\varepsilon_{\mathrm{r}}'(\omega)$ into the Brillouin formula is then no longer a reliable inner product.

5.4.3 Why a simple inner product fails

The root of the difficulty is that the reduced macroscopic Maxwell equations—written in terms of $\mathbf{E}$ and $\mathbf{H}$ alone—describe an *open subsystem*: the material degrees of freedom (atomic oscillators, free carriers, phonons) have been integrated out, leaving only their effect as a frequency-dependent $\varepsilon_{\mathrm{r}}(\omega)$. The full system (fields plus matter) conserves energy, but the reduced system does not, because energy flows between the fields and the material oscillators. Attempting to define a conserved energy for the fields alone is like trying to define a conserved energy for one particle in a two-body problem [1].

5.5 Auxiliary fields and the Hamiltonian embedding

The resolution, developed rigorously by Figotin and Schenker [1], is to *restore* the material degrees of freedom as explicit auxiliary fields and construct a conservative extended system.

5.5.1 The idea

Consider a medium whose susceptibility $\hat{\chi}(\omega)$ satisfies the causality, passivity, and Kramers–Kronig conditions. The Lorentz model (5.6) can be written as

$$\hat{\chi}(\omega) = \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega} = \int_0^\infty \frac{J(\nu)}{\nu^2 - \omega^2 - i\gamma'\omega}\ \mathrm{d}\nu, \tag{5.18}$$

where the second form generalizes to any causal susceptibility via a spectral representation (a continuous distribution of oscillator strengths $J(\nu)$). Each oscillator mode at frequency ν is promoted to an explicit dynamical variable—an auxiliary field $\mathbf{P}_\nu(\mathbf{r}, t)$ that couples to $\mathbf{E}$ and obeys its own equation of motion.

5.5.2 Example: single Lorentz pole

For a single Lorentz resonance, the auxiliary field is the polarization density $\mathbf{P}(\mathbf{r}, t)$. The coupled equations of motion are:

$$\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{H} - \frac{\partial \mathbf{P}}{\partial t}, \tag{5.19}$$

$$\mu_0 \frac{\partial \mathbf{H}}{\partial t} = -\nabla \times \mathbf{E}, \tag{5.20}$$

$$\frac{\partial[}{\partial 2}]\mathbf{P}t + \gamma \frac{\partial \mathbf{P}}{\partial t} + \omega_0^2 \mathbf{P} = \varepsilon_0 \omega_p^2\, \mathbf{E}. \tag{5.21}$$

Equation (5.21) is the Lorentz oscillator equation. Taking the Fourier transform and solving for $\mathbf{P}(\omega)$ gives $\mathbf{P}(\omega) = \varepsilon_0 \hat{\chi}(\omega)\mathbf{E}(\omega)$, recovering the Lorentz permittivity (5.6).

The damping term $\gamma\partial\mathbf{P}/\partial t$ is the source of non-Hermiticity: it represents energy transfer from the coherent polarization to an incoherent bath. Without it ($\gamma = 0$), the system (5.19)–(5.21) is conservative and can be written in Hamiltonian form.

5.5.3 The extended Hamiltonian

For $\gamma = 0$, define the canonical momentum of the polarization oscillator as $\mathbf{Q} = \partial\mathbf{P}/\partial t$. The total energy of the extended system is

$$\mathcal{H} = \frac{1}{2}\int\left[\varepsilon_0|\mathbf{E}|^2 + \mu_0|\mathbf{H}|^2 + \frac{|\mathbf{Q}|^2}{\varepsilon_0\omega_p^2} + \frac{\omega_0^2|\mathbf{P}|^2}{\varepsilon_0\omega_p^2}\right]\mathrm{d}^3r, \tag{5.22}$$

which is manifestly positive definite and conserved by the coupled equations. The first-order form is

$$i\,\frac{\partial}{\partial t}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\\ \mathbf{P}\\ \mathbf{Q}\end{pmatrix} = \hat{\mathcal{H}}_{\mathrm{ext}}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\\ \mathbf{P}\\ \mathbf{Q}\end{pmatrix}, \tag{5.23}$$

where $\hat{\mathcal{H}}_{\mathrm{ext}}$ is self-adjoint with respect to the extended energy inner product (5.22).

When the auxiliary fields are eliminated (by solving their equations of motion in terms of $\mathbf{E}$), one recovers the original reduced Maxwell equations with the frequency-dependent $\varepsilon_{\mathrm{r}}(\omega)$. But in the reduced system, the total energy is no longer conserved: some of it has been transferred to the hidden oscillators.

For $\gamma > 0$, the damping term breaks the Hermiticity of $\hat{\mathcal{H}}_{\mathrm{ext}}$. The extended system is then non-conservative, and a further embedding into a larger Hilbert space (a Lindblad or Caldeira–Leggett bath) is needed for a fully Hamiltonian treatment [1].

Non-Hermiticity from elimination

The macroscopic Maxwell equations with frequency-dependent $\varepsilon_{\mathrm{r}}(\omega)$ are non-Hermitian not because the physics is non-conservative, but because the conservative degrees of freedom of the material have been eliminated. Restoring them through auxiliary fields recovers a Hamiltonian (Hermitian) description of the full system. This is the third road to non-Hermiticity identified in Chapter 2: not material loss, not radiation, but frequency dispersion caused by eliminating material variables from the Maxwell state.

5.6 Gain media

An active (gain) medium has $\varepsilon_{\mathrm{r}}''(\omega) < 0$ at frequencies within the gain bandwidth. Physically, this requires an external energy source (a pump) that maintains a population inversion or stimulated emission process.

5.6.1 The inverted Lorentz model

The Lorentz oscillator model can be adapted to describe gain by reversing the sign of the damping: $\gamma \to -|\gamma|$. As a small-signal phenomenological model near a pumped transition, this gives

$$\varepsilon_{\text{gain}}(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 + i|\gamma|\omega}, \tag{5.24}$$

with $\varepsilon_{\text{r}}'' < 0$ near ω_0. The medium now amplifies rather than absorbs. Taken literally over all frequencies, negative damping would be an unstable linear oscillator; in a real gain medium the pump, finite bandwidth, and saturation nonlinearity regularize the response.

5.6.2 Population inversion and the four-level system

In a real laser medium, gain arises from population inversion between two electronic levels. The simplest tractable model is the four-level system with levels $|0\rangle$, $|1\rangle$, $|2\rangle$, $|3\rangle$: an external pump excites atoms from $|0\rangle$ to $|3\rangle$; rapid non-radiative decay brings atoms from $|3\rangle$ to $|2\rangle$ (the upper laser level); stimulated emission occurs on the $|2\rangle \to |1\rangle$ transition, providing gain at the lasing frequency $\omega_L = (E_2 - E_1)/\hbar$; and rapid decay from $|1\rangle$ to $|0\rangle$ empties the lower laser level, maintaining inversion.

The gain coefficient is proportional to the population difference $\Delta N = N_2 - N_1 > 0$ and to the transition cross section $\sigma(\omega)$:

$$g(\omega) = \sigma(\omega)\,\Delta N, \tag{5.25}$$

where $\sigma(\omega)$ has a Lorentzian line shape centered at ω_L. The imaginary permittivity contributed by the gain transition is $\varepsilon_{\text{r}}''(\omega) = -cn\,g(\omega)/\omega < 0$.

Rate equations for the four-level system. The steady-state populations are governed by the rate equations. Because the relaxation rates γ_{32} (from $|3\rangle$ to $|2\rangle$) and γ_{10} (from $|1\rangle$ to $|0\rangle$) are fast, the populations N_3 and N_1 are approximately zero in steady state, and the total population satisfies $N_0 + N_2 \approx N_{\text{tot}}$. The rate equation for the upper laser level is

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = R_{\text{pump}} - \frac{\sigma(\omega_L)\,I}{\hbar\omega_L}\,\Delta N - \frac{N_2}{\tau_{\text{sp}}} = 0, \tag{5.26}$$

where R_{pump} is the pump rate (atoms per unit volume per unit time), I is the intracavity intensity, τ_{sp} is the spontaneous emission lifetime of the upper level, and the middle term represents the stimulated emission rate. Since $N_1 \approx 0$, we have $\Delta N \approx N_2$.

Solving (5.26) for N_2 in steady state:

$$N_2 = \frac{R_{\text{pump}}\,\tau_{\text{sp}}}{1 + I/I_{\text{sat}}}, \qquad I_{\text{sat}} = \frac{\hbar\omega_L}{\sigma(\omega_L)\,\tau_{\text{sp}}}, \tag{5.27}$$

where we identify the small-signal inversion $\Delta N_0 = R_{\text{pump}}\,\tau_{\text{sp}}$ (the inversion in the absence of stimulated emission). Substituting into Eq. (5.25) immediately recovers the saturated gain formula (5.31):

$$g(I) = \sigma\,\Delta N_0 \cdot \frac{1}{1 + I/I_{\text{sat}}} = \frac{g_0}{1 + I/I_{\text{sat}}}. \tag{5.28}$$

The saturation intensity I_{sat} has a transparent physical meaning: it is the intensity at which the stimulated emission rate $\sigma I/(\hbar\omega_L)$ equals the spontaneous decay rate $1/\tau_{\text{sp}}$. Below this intensity, stimulated emission is a small perturbation on the inversion; above it, stimulated emission dominates and the gain is depleted.

Gain clamping and threshold. At steady-state lasing, the gain must exactly compensate the total cavity loss α_{tot} (including mirror loss and material absorption):

$$g(I_{\text{ss}}) = \alpha_{\text{tot}} \quad \Rightarrow \quad I_{\text{ss}} = I_{\text{sat}} \left(\frac{g_0}{\alpha_{\text{tot}}} - 1 \right). \tag{5.29}$$

This is the gain-clamping condition. Below threshold ($g_0 < \alpha_{\text{tot}}$), no steady-state lasing occurs and $I_{\text{ss}} = 0$. Above threshold, the intracavity intensity grows linearly with pump power (since $g_0 \propto R_{\text{pump}}$), while the material gain clamps at α_{tot}—it cannot exceed the loss, because any excess gain would increase I, which would further deplete the inversion until balance is restored.

Connection to the Maxwell framework. The gain-clamping condition connects the microscopic rate equations to the macroscopic Maxwell eigenproblem. The intensity-dependent permittivity of the gain medium is

$$\varepsilon_{\text{r}}(\omega, I) = \varepsilon_{\text{r}}'(\omega) - i\,\frac{c\,n_{\text{bg}}\,g(I)}{\omega} = \varepsilon_{\text{r}}'(\omega) - i\,\frac{c\,n_{\text{bg}}\,g_0}{\omega(1 + I/I_{\text{sat}})}, \tag{5.30}$$

where n_{bg} is the background refractive index and the negative sign of the imaginary part reflects gain (with our $e^{-i\omega t}$ convention, $\varepsilon_{\text{r}}'' < 0$ corresponds to amplification). This intensity-dependent permittivity makes the Maxwell eigenproblem nonlinear: the eigenfrequencies depend on the field amplitudes through I. The linear non-Hermitian framework of this book treats ε_{r} as independent of I—an approximation valid below or near threshold, where $I \ll I_{\text{sat}}$ and $g \approx g_0$. Above threshold, the full nonlinear problem must be solved self-consistently, as discussed in Chapter 19.

5.6.3 Gain saturation

As the intracavity field intensity I grows, stimulated emission depletes the population inversion. The saturated gain coefficient is

$$g(I) = \frac{g_0}{1 + I/I_{\text{sat}}}, \tag{5.31}$$

where g_0 is the small-signal gain and I_{sat} is the saturation intensity. At threshold, the saturated gain equals the total loss. Above threshold, the gain clamps at the loss value, and additional pump power is converted into laser output power.

Gain saturation is intrinsically nonlinear: the permittivity depends on the field amplitude. The linear non-Hermitian framework developed in this book is therefore strictly valid only below or near the lasing threshold. Above threshold, the coupling between the gain medium and the cavity field requires a nonlinear analysis (Chapter 19).

Quantitative example: InGaAsP gain medium. For a typical InGaAsP multiple-quantum-well gain medium [31, 32] at $\lambda = 1550$ nm, the small-signal material gain coefficient is $g_0 \approx 400$ cm^{-1} at a carrier density $N = 3 \times 10^{18}$ cm^{-3}, and the saturation intensity is $I_{\text{sat}} \approx$

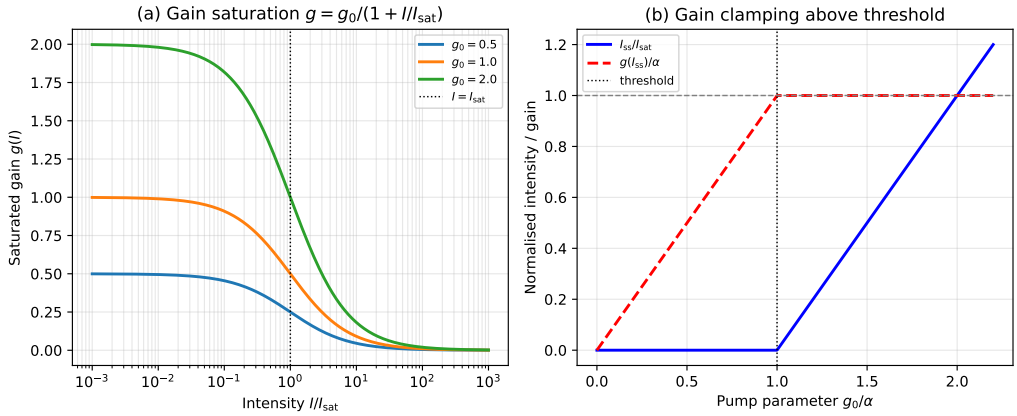


Figure 5.4 Gain saturation and gain clamping in a four-level laser model. (a) The saturated gain $g(I) = g_0/(1 + I/I_{\mathrm{sat}})$ decreases once the intracavity intensity becomes comparable with the saturation intensity I_{sat}. (b) Above threshold, the steady-state intensity grows as $I_{\mathrm{ss}}/I_{\mathrm{sat}} = g_0/\alpha - 1$ while the material gain clamps to the total loss α. This is the nonlinear mechanism that prevents the linear gain model from producing unbounded fields above threshold.

5×10^8 W/m^2 (corresponding to a saturation photon density $n_{\mathrm{ph,sat}} \approx 2.6 \times 10^{15}$ cm^{-3}). The corresponding imaginary permittivity is $\varepsilon'' = -cng_0/\omega \approx -0.031$ at small signal. For a photonic crystal cavity with mode volume $V_{\mathrm{mode}} \sim (\lambda/n)^3 \approx 0.05\ \mu$m^3 and $Q = 5000$, the threshold condition $g_{\mathrm{th}} = \omega/(cQ) \approx 40$ cm^{-1} is reached at a pump current of roughly 50 μA—well within the linear regime where $g/g_0 \approx 0.1$ and saturation corrections are small. As the pump increases toward g_0, the saturation nonlinearity becomes important and the linear non-Hermitian description begins to break down.

Kramers–Kronig constraints on gain media. A gain medium with $\varepsilon''(\omega) < 0$ over a spectral band [33, 34] must, by the Kramers–Kronig relations, also exhibit anomalous dispersion $\mathrm{d}\varepsilon'/\mathrm{d}\omega < 0$ within the same band. This is because the subtracted dispersion relation $\varepsilon'(\omega) - \varepsilon'(\infty) = (2/\pi)\,\mathcal{P}\int_0^\infty \omega'\varepsilon''(\omega')/(\omega'^2 - \omega^2)\ \mathrm{d}\omega'$ has the same structure for gain as for loss, with reversed sign of ε''. The anomalous dispersion in the gain band reduces the group velocity and can even make it superluminal or negative, though the signal velocity remains bounded by c because the gain bandwidth is always finite. This connection between gain and anomalous dispersion is important for understanding the frequency-pulling effect in lasers and the bandwidth limitations of optical amplifiers.

5.7 Balanced gain and loss: the $\mathcal{PT}$-symmetric condition

A medium with spatial regions of loss and gain, arranged so that

$$\varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(-\mathbf{r}), \tag{5.32}$$

has a special symmetry: $\mathcal{PT}$ symmetry. The real part of ε_{r} is symmetric and the imaginary part is antisymmetric under spatial inversion. This condition ensures that the total gain and loss are balanced, and the system can exhibit a phase transition from real to complex eigenfrequencies. The physics of $\mathcal{PT}$-symmetric optical systems is the subject of Chapter 9.

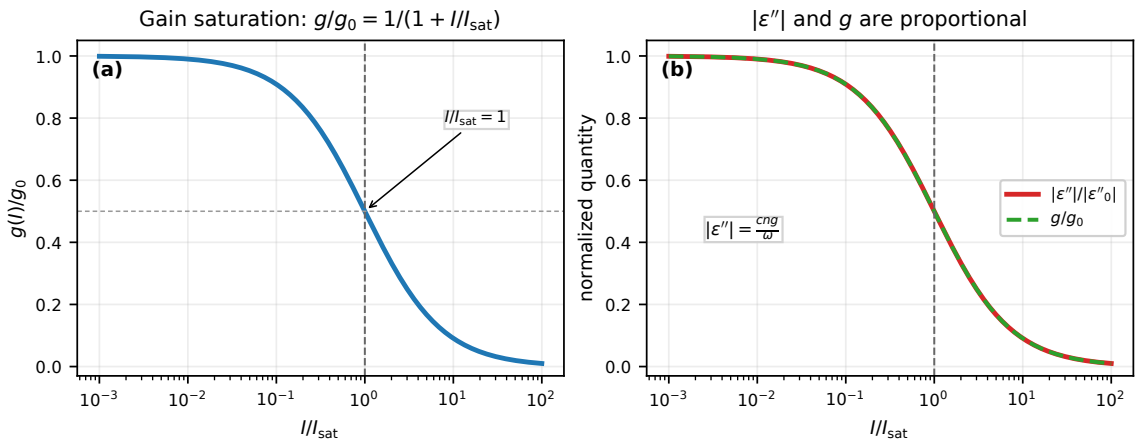


Figure 5.5 Gain saturation in a homogeneous medium. (a) Normalized gain $g(I)/g_0 = 1/(1 + I/I_{\mathrm{sat}})$ versus intensity. The half-gain point occurs at $I = I_{\mathrm{sat}}$. (b) Normalized $|\varepsilon''|/|\varepsilon''_0|$ (solid red) and g/g_0 (dashed green) plotted on the same vertical scale, confirming that both quantities share the same saturation dependence; the conversion is $|\varepsilon''| = c\,n\,g/\omega$. Parameters: $g_0 = 400\ \mathrm{cm}^{-1}$, $\varepsilon''_0 \approx -0.031$, $\lambda = 1550$ nm, $n_{\mathrm{eff}} = 3.4$.

5.7.1 The $\mathcal{PT}$-symmetric Lorentz model

A one-dimensional $\mathcal{PT}$-symmetric refractive-index profile can be constructed from two Lorentz media with opposite signs of damping:

$$\varepsilon_{\mathrm{r}}(x,\omega) = \begin{cases} \varepsilon_{\mathrm{bg}} + \Delta\varepsilon_{\mathrm{Lorentz}}(\omega;\gamma) & x > 0, \\ \varepsilon_{\mathrm{bg}} + \Delta\varepsilon_{\mathrm{Lorentz}}(\omega;-\gamma) & x < 0, \end{cases} \tag{5.33}$$

where $\gamma > 0$ and $\varepsilon_{\mathrm{bg}}$ is real. This satisfies $\varepsilon_{\mathrm{r}}(x,\omega) = \varepsilon_{\mathrm{r}}^*(-x,\omega)$ on the real-frequency axis and provides a microscopically motivated realization of the $\mathcal{PT}$ condition for the transfer-matrix and coupled-mode analyses of Chapter 9.

5.8 Summary and connections

Material loss, gain, and dispersion—the first road to non-Hermiticity—are not ad hoc modifications of Maxwell's equations but an inevitable consequence of the causal, dissipative response of real materials:

- Linear response and causality require $\varepsilon_{\mathrm{r}}(\omega)$ to be complex and frequency-dependent, constrained by the Kramers–Kronig relations and sum rules.
- The imaginary part $\varepsilon''_{\mathrm{r}}$ breaks the self-adjointness of the Maxwell operator by making the constitutive weight complex and by allowing irreversible exchange of energy with the material.
- The Brillouin formula provides a useful but limited approximation to the energy in dispersive media; a rigorous treatment requires auxiliary fields and an extended Hamiltonian.
- Gain media violate passivity through population inversion and can lead to lasing when coupled with feedback; gain saturation makes the linear non-Hermitian description valid only near or below threshold.

- The special case of balanced gain and loss ($\mathcal{PT}$ symmetry) connects to a rich phenomenology explored in Chapter 9.

Several threads connect forward from this material foundation:

1. The complex QNM frequencies of Chapter 6 arise from material loss ($\varepsilon'' > 0$) combined with radiative loss from open boundaries.
2. The $\mathcal{PT}$-symmetric condition $\varepsilon(\mathbf{r}) = \varepsilon^*(-\mathbf{r})$ of Chapter 9 is a precise spatial arrangement of the gain and loss physics developed here.
3. The dispersive auxiliary-field formulation of Section 5.5 reappears in Chapter 17 as the basis for finite-element eigensolvers that handle frequency-dependent materials without nonlinear eigenvalue problems.
4. Gain saturation provides the nonlinear feedback that stabilises lasers and creates the nonlinear exceptional points discussed in Chapters 9 and 19.
5. The Kramers–Kronig constraint that gain media must exhibit anomalous dispersion limits the bandwidth of optical amplifiers and modifies the group velocity of pulses propagating through active photonic structures.

Chapter 6 now turns to the second physical origin of non-Hermiticity—radiation and openness—and develops the quasinormal-mode theory that provides the natural eigenstates of the non-self-adjoint Maxwell operator.

Material non-Hermiticity is physical, not mathematical

Complex permittivity, frequency dispersion, gain saturation, and auxiliary-field embeddings are not formal devices introduced for mathematical convenience. They encode the irreversible exchange of energy between the electromagnetic field and matter—absorption heating, stimulated emission, and the frequency-dependent phase response of bound electrons. Every non-Hermitian photonic phenomenon studied in the remainder of this book has its material origin in the physics of this chapter.

Chapter 6

Quasinormal Modes

When a photonic resonator is open to its environment, its "modes" are not the orthogonal, real-frequency eigenstates of Hermitian spectral theory. They are quasinormal modes*: solutions of the source-free Maxwell equations with outgoing boundary conditions, possessing complex eigenfrequencies and field profiles that diverge at spatial infinity. Quasinormal modes are derived from the Maxwell operator. Their normalization and inner product differ from those of closed cavities, and they describe open photonic systems ranging from nanophotonic cavities to diffraction gratings.*

6.1 Defining quasinormal modes

A *quasinormal mode* (QNM) of an open electromagnetic system is a solution to the source-free vector Helmholtz equation

$$\nabla \times \left[\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})} \nabla \times \tilde{\mathbf{H}}_n(\mathbf{r}) \right] = \frac{\tilde{\omega}_n^2}{c^2} \tilde{\mathbf{H}}_n(\mathbf{r}), \tag{6.1}$$

subject to the *outgoing-wave* (Silver–Müller) boundary condition introduced in Section 4.2. The tilde distinguishes QNM fields and frequencies from their Hermitian counterparts.

Because the boundary condition is frequency-dependent, the problem is nonlinear in $\tilde{\omega}$: we seek pairs $(\tilde{\omega}_n, \tilde{\mathbf{H}}_n)$ simultaneously. For a passive system without gain, all QNM frequencies lie in the lower half of the complex ω plane:

$$\tilde{\omega}_n = \omega_n - i\gamma_n, \qquad \gamma_n > 0, \tag{6.2}$$

where ω_n is the resonant frequency and γ_n the radiative decay rate. The quality factor is $Q_n = \omega_n/(2\gamma_n)$.

As established in Section 4.4, the QNM frequencies coincide with the poles of the scattering matrix $\mathbf{S}(\omega)$: at $\omega = \tilde{\omega}_n$, an outgoing field exists with no incoming excitation.

6.1.1 Example: Fabry–Pérot QNMs

The simplest system with QNMs is a dielectric slab of thickness L and real refractive index n in vacuum. The QNM condition requires outgoing waves on both sides, which is equivalent to the round-trip resonance condition with the Fresnel reflection coefficient $r = (n-1)/(n+1)$:

$$r^2 e^{2in\tilde{\omega}L/c} = 1. \tag{6.3}$$

Solving for $\tilde{\omega}$ gives the QNM frequencies

$$\tilde{\omega}_m = \frac{c}{nL}\left(m\pi - \arg r\right) - i\,\frac{c}{nL}\ln\frac{1}{|r|}, \qquad m = 0, \pm 1, \pm 2, \ldots \tag{6.4}$$

All frequencies have the same imaginary part $\gamma = (c/nL)\ln(1/|r|)$, set by the mirror reflectivity. For a silicon slab ($n = 3.5$, $L = 300$ nm):

- $|r| = (3.5-1)/(3.5+1) = 0.556$, so $\ln(1/|r|) = 0.588$.
- $\gamma = c \times 0.588/(3.5 \times 300\text{ nm}) \approx 1.68 \times 10^{14}$ rad/s.
- $Q \approx \omega_0/(2\gamma)$. For the fundamental mode at $\lambda \approx 2nL = 2100$ nm ($\omega_0 \approx 8.97 \times 10^{14}$ rad/s), $Q \approx 2.7$.

This very low Q reflects the poor confinement of a bare slab. Adding distributed Bragg reflectors raises $|r|$ and pushes the QNM poles closer to the real axis.

QNM field profile inside and outside the slab. Inside the slab ($0 < z < L$), the QNM field is a standing wave: $\tilde{E}_m(z) \propto \sin(n\tilde{\omega}_m z/c)$ (for the **E**-field of a TE mode at normal incidence). Outside the slab ($z > L$ or $z < 0$), the field has the outgoing-wave form $\tilde{E}_m(z) \propto e^{+i\tilde{\omega}_m z/c}$ for $z > L$ and $e^{-i\tilde{\omega}_m z/c}$ for $z < 0$. Writing $\tilde{\omega}_m = \omega_m - i\gamma$, the far-field amplitude is $|\tilde{E}_m(z)| \propto e^{\gamma|z|/c}$: the field grows exponentially away from the slab, as derived in Section 6.2. At a distance of $z = c/\gamma \approx 1.78$ μm (for the silicon slab above), the field amplitude has grown by a factor of $e \approx 2.72$. At $z = 10$ μm, it has grown by a factor of $e^{5.6} \approx 270$. This rapid spatial growth is what makes the standard L^2 norm diverge and necessitates the unconjugated normalization of Section 6.5.

6.2 Spatial divergence and its physical meaning

A QNM with complex frequency $\tilde{\omega}_n = \omega_n - i\gamma_n$ has far-field spatial behavior

$$\tilde{\mathbf{H}}_n(\mathbf{r}) \;\sim\; \frac{e^{i\tilde{\omega}_n r/c}}{r}\,\mathbf{f}_n(\hat{\mathbf{r}}) = \frac{e^{i\omega_n r/c}\,e^{\gamma_n r/c}}{r}\,\mathbf{f}_n(\hat{\mathbf{r}}), \tag{6.5}$$

where $\mathbf{f}_n$ is an angular pattern. The factor $e^{\gamma_n r/c}$ *grows* exponentially with distance. This is deeply counterintuitive—how can a decaying mode grow in space?

The resolution is causal: the QNM represents a field pattern that was established at some time in the past and has been radiating ever since. At a distant point r, we observe light that was emitted at the earlier time $t - r/c$, when the mode had amplitude $e^{-\gamma_n(t-r/c)}$. The product of the temporal decay and the retardation gives $e^{-\gamma_n t}\,e^{\gamma_n r/c}$: at fixed t, the field grows with r because distant points sample the mode at earlier, stronger times.

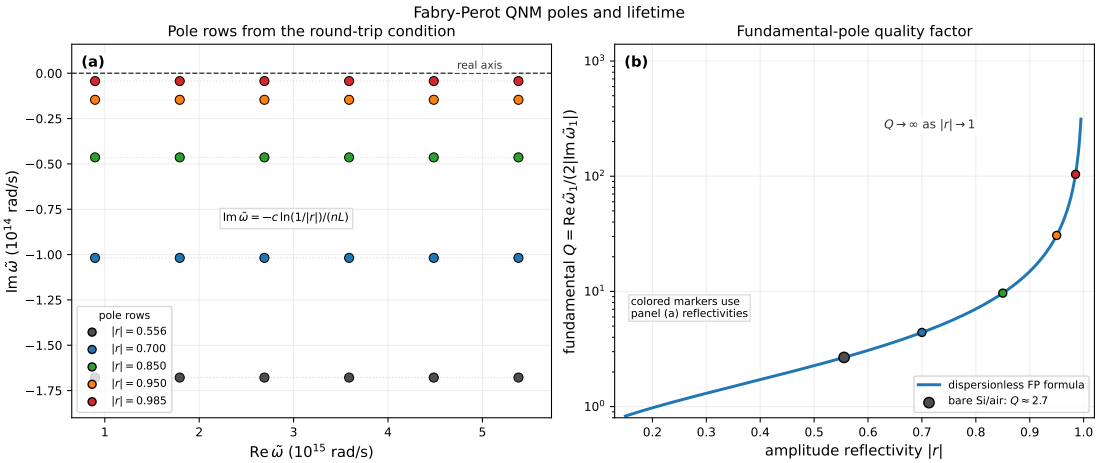


Figure 6.1 Quasinormal-mode poles of a one-dimensional Fabry–Pérot resonator. Panel (a) plots the pole positions from the round-trip condition $r^2 \exp(2in\tilde{\omega}L/c) = 1$ for several mirror reflectivities; the compact legend identifies the $|r|$ value for each pole row, and the in-panel formula gives the imaginary part $\mathrm{Im}\,\tilde{\omega} = -c\ln(1/|r|)/(nL)$. Increasing $|r|$ moves the pole rows upward toward the real-frequency axis while remaining in the passive lower half-plane. Panel (b) shows the fundamental quality factor $Q = \mathrm{Re}\,\tilde{\omega}_1/(2|\mathrm{Im}\,\tilde{\omega}_1|)$ as a function of amplitude reflectivity, with colored markers corresponding to the same $|r|$ values as the pole rows. The gray marker denotes the bare silicon–air interface example used in the text, with $n = 3.5$, $L = 300\,\mathrm{nm}$, $|r| = 0.556$, and $Q \approx 2.67$.

This exponential growth means that the standard L^2 norm $\int |\tilde{\mathbf{H}}_n|^2 \, \mathrm{d}^3r$ diverges. The QNM is *not square-integrable* and cannot be normalized with the Hermitian inner product of Chapter 2. A different normalization scheme is needed.

6.2.1 Connection to the skin effect

The spatial divergence of a single QNM has a close analogy with the non-Hermitian skin effect (Chapter 14): in both cases, the non-Hermitian boundary condition forces eigenstates to have exponentially growing or decaying spatial profiles rather than the oscillatory behavior of Hermitian modes. The QNM divergence is the single-resonator version of a phenomenon that, in periodic systems, becomes the collective skin effect.

6.3 PML regularization and the non-self-adjoint operator

The perfectly matched layers introduced in Section 4.8 provide both a computational tool and a conceptual resolution of the divergence problem. By applying a complex coordinate stretching $\mathbf{r} \to \tilde{\mathbf{r}}$ in the absorbing layer, the PML maps the outgoing-wave problem on an infinite domain to an eigenvalue problem on a bounded domain with a *non-self-adjoint* operator [5].

6.3.1 How PML regularization works

In the PML-regularized problem:

1. The QNM fields are damped in the PML region and become square-integrable on the computational domain.

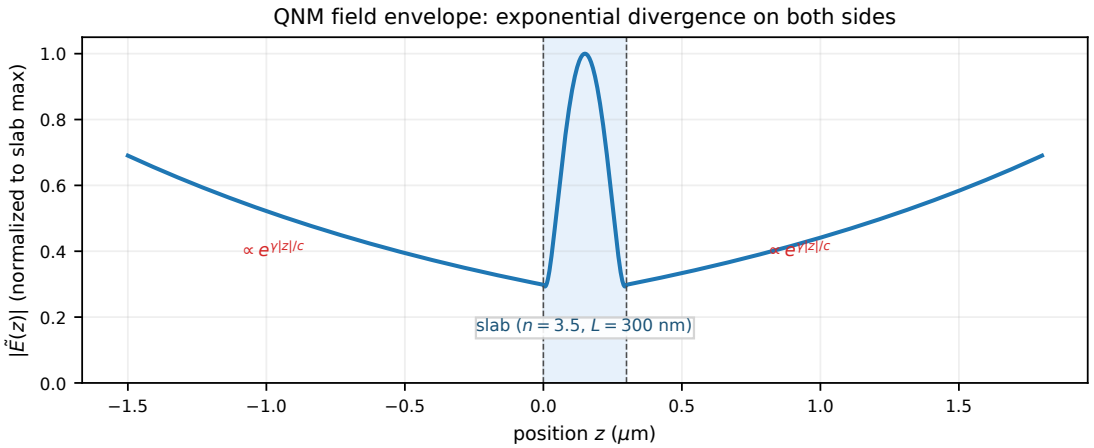


Figure 6.2 Field envelope $|\tilde{E}(z)|$ of the fundamental quasinormal mode of a dielectric slab ($n = 3.5$, $L = 300$ nm) in vacuum, normalized to the slab-interior maximum. The mode is the symmetric $\cos\tilde{k}(z - L/2)$ solution. Outside the slab, the outgoing-wave boundary condition produces exponential growth $\propto e^{\gamma|z|/c}$ on both sides, where $\gamma = -c\ln|r|/(nL)$ is the radiative decay rate from Eq. (6.2). This divergence is the defining feature of QNMs and is regularized by perfectly matched layers (Section 6.3). Shaded region marks the slab.

2. The true QNM eigenfrequencies $\tilde{\omega}_n$ (poles of $\mathbf{S}$) are unchanged—they are intrinsic to the scatterer and independent of the PML parameters.
3. The continuous spectrum of radiation modes is discretized by the finite PML, producing a set of "PML modes" that depend on the PML thickness and absorption strength.
4. The transformed Maxwell operator is non-self-adjoint: its eigenvectors are not orthogonal under the standard L^2 inner product.

6.3.2 PML validation and spurious modes

The PML introduces two practical concerns:

- **Spurious PML modes:** the discretized radiation continuum produces a set of modes whose eigenfrequencies depend on the PML parameters. These must be distinguished from the physical QNMs, whose eigenfrequencies converge as the PML thickness and absorption increase. A standard check is to vary the PML parameters and identify eigenfrequencies that remain stable.
- **PML thickness and absorption:** the PML must be thick enough and absorbing enough to suppress reflections from the outer boundary. For QNMs with large γ_n (low Q), the spatial growth rate γ_n/c is large, and the PML must compensate this growth before the field reaches the outer boundary. A rule of thumb is that the PML absorption over its thickness should provide at least 20 dB of suppression.

The price of PML regularization is that the operator is now manifestly non-Hermitian. The tools of Hermitian spectral theory do not apply. Instead, we must use the *adjoint* operator and *biorthogonal* relations.

PML convergence in practice [35, 36]. For the silicon Fabry–Pérot slab ($n = 3.5$, $L = 300$ nm), a PML of thickness $d_{\mathrm{PML}} = 2\lambda$ with a quadratic absorption profile $\sigma(s) =$

$\sigma_{\max}(s/d_{\mathrm{PML}})^2$ converges the fundamental QNM frequency to four significant figures once $\sigma_{\max} \gtrsim 3\,\omega/c$. Beyond that value the pole position is insensitive to PML parameters—the hallmark of a well-converged QNM calculation, because the physical eigenfrequency is a property of the resonator, not of the absorbing boundary. A useful convergence diagnostic is the mode ratio $\eta = |\tilde{E}_{\mathrm{PML,outer}}|/|\tilde{E}_{\max}|$: when $\eta < 10^{-3}$, the field has decayed sufficiently within the PML and the eigenvalue is converged to better than 0.1%. This criterion translates into a minimum PML thickness of roughly 1.5–2 wavelengths for moderate-Q resonators ($Q \sim 10$) and can be relaxed to $\sim 1\lambda$ for high-Q modes ($Q > 10^3$) whose far-field divergence rate γ/c is small.

6.4 The adjoint problem and biorthogonality

6.4.1 The adjoint Maxwell operator

For the PML-regularized Maxwell operator $\hat{\Theta}_{\mathrm{PML}}$, the adjoint operator $\hat{\Theta}^{\dagger}_{\mathrm{PML}}$ is defined by the condition

$$\langle \mathbf{F},\, \hat{\Theta}_{\mathrm{PML}}\mathbf{G}\rangle = \langle \hat{\Theta}^{\dagger}_{\mathrm{PML}}\mathbf{F},\, \mathbf{G}\rangle \tag{6.6}$$

for all fields in the domain. For a reciprocal medium (symmetric ε_{r} tensor), the adjoint operator takes a remarkably simple form: its eigenvectors $\tilde{\mathbf{H}}^{\mathrm{L}}_n$ satisfy the same differential equation (6.1) but with *incoming-wave* boundary conditions—the time-reverse of the QNM problem [5].

6.4.2 Biorthogonality

The "left" (adjoint) eigenvectors $\tilde{\mathbf{H}}^{\mathrm{L}}_m$ and "right" (direct) eigenvectors $\tilde{\mathbf{H}}_n$ satisfy a biorthogonality relation:

$$\big\langle \tilde{\mathbf{H}}^{\mathrm{L}}_m,\, \tilde{\mathbf{H}}_n \big\rangle_{\mathrm{PML}} = \delta_{mn}, \tag{6.7}$$

after appropriate normalization. This is the non-Hermitian analogue of mode orthogonality. The subscript PML indicates that the integral is over the computational domain including the PML region.

6.4.3 Physical interpretation

The biorthogonality relation has a direct physical interpretation: $\langle \tilde{\mathbf{H}}^{\mathrm{L}}_m | \tilde{\mathbf{H}}_n \rangle = 0$ for $m \neq n$ means that the *time-reversed* field of mode m has zero overlap with the forward field of mode n. This is a statement about reciprocity: the mode m launched backward does not excite mode n launched forward. This reciprocal orthogonality replaces the power orthogonality of Hermitian modes.

Biorthogonality replaces orthogonality

In Hermitian systems, left and right eigenvectors are simply complex conjugates of each other, and the biorthogonality relation (6.7) reduces to ordinary orthogonality $\langle \mathbf{H}^*_m, \mathbf{H}_n \rangle = \delta_{mn}$. In non-Hermitian systems, left and right eigenvectors are generally different functions, and the bilinear form $\langle \cdot, \cdot \rangle$ is not the same as the Hermitian inner product $\langle \cdot | \cdot \rangle$.

6.5 QNM normalization

The normalization of quasinormal modes is subtle because the standard volume integral diverges. Several equivalent prescriptions exist.

6.5.1 The unconjugated norm

For a reciprocal system, the adjoint modes are related to the direct modes by $\tilde{\mathbf{H}}_n^{\mathrm{L}}(\mathbf{r}) = \tilde{\mathbf{H}}_n(\mathbf{r})$ (without complex conjugation), and the normalization condition becomes [4]

$$\int_\Omega \varepsilon_{\mathrm{r}}(\mathbf{r})\, \tilde{\mathbf{E}}_n(\mathbf{r}) \cdot \tilde{\mathbf{E}}_n(\mathbf{r})\ \mathrm{d}^3 r + \text{surface correction} = 1. \tag{6.8}$$

The crucial feature is that $\tilde{\mathbf{E}}_n$ appears *twice without conjugation*: this is an unconjugated bilinear form, not a Hermitian inner product. The "norm" is complex, and the "mode volume" derived from it is also complex:

$$\tilde{V}_n = \frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r}_0)\, \tilde{\mathbf{E}}_n(\mathbf{r}_0) \cdot \tilde{\mathbf{E}}_n(\mathbf{r}_0)}, \tag{6.9}$$

evaluated at a point of interest $\mathbf{r}_0$ (typically the field maximum).

6.5.2 The surface-integral correction

The unconjugated volume integral in (6.8) still diverges over infinite space. The PML regularization truncates the domain, but in formulations without a PML one can add an analytic surface term that cancels the far-field divergence [4]. For a nondispersive homogeneous exterior the correction has the schematic form

$$N_n = \int_{\Omega_R} \left[\varepsilon_{\mathrm{r}}\, \tilde{\mathbf{E}}_n \cdot \tilde{\mathbf{E}}_n - \mu_{\mathrm{r}}\, \tilde{\mathbf{H}}_n \cdot \tilde{\mathbf{H}}_n\right] \mathrm{d}^3 r + S_R[\tilde{\mathbf{E}}_n, \tilde{\mathbf{H}}_n], \tag{6.10}$$

where Ω_R is a finite volume bounded by a surface of radius R and S_R is an outgoing-wave surface functional determined by the far-field expansion. Its role is not to add stored energy at the boundary, but to analytically continue the divergent volume integral to a finite, normalization-independent value. In modern computations this same regularization is often achieved more cleanly by integrating the unconjugated product over a PML-stretched domain.

6.5.3 Comparison of normalization schemes

Three normalization schemes are commonly used in the literature:

1. **PML-based:** integrate the unconjugated product over the PML domain. Simplest computationally but depends weakly on PML parameters.
2. **Surface-corrected:** add the surface integral correction (6.10). PML-independent to exponential accuracy.
3. **Pole residue:** extract the normalization from the residue of the Green function at the QNM pole. Mathematically elegant and manifestly PML-independent, but requires a scattering calculation at complex frequency.

All three give the same result when properly implemented. The choice is a matter of computational convenience.

6.5.4 The Petermann factor from the unconjugated norm

The complex mode volume (6.9) has an imaginary part that vanishes only when the mode is Hermitian. The ratio of the Hermitian (conjugated) norm to the unconjugated norm defines the *Petermann factor*:

$$K_n = \frac{\displaystyle\int |\tilde{\mathbf{E}}_n(\mathbf{r})|^2 \, \mathrm{d}^3 r}{\left| \displaystyle\int \tilde{\mathbf{E}}_n(\mathbf{r}) \cdot \tilde{\mathbf{E}}_n(\mathbf{r}) \, \mathrm{d}^3 r \right|}, \tag{6.11}$$

where the numerator uses the conventional conjugated inner product and the denominator is the absolute value of the unconjugated norm. By the Cauchy–Schwarz inequality, $K_n \geq 1$, with equality if and only if $\tilde{\mathbf{E}}_n$ can be chosen everywhere real—i.e. the mode is Hermitian.

To see why K_n measures excess noise and non-orthogonality, note that the biorthogonal overlap of mode n with its adjoint partner is

$$\langle \tilde{\mathbf{E}}_n^{\mathrm{L}} | \tilde{\mathbf{E}}_n \rangle = \frac{1}{\sqrt{K_n}}, \tag{6.12}$$

so the *phase rigidity* $r_n = K_n^{-1/2}$ measures how close the right eigenvector is to the left eigenvector. At an exceptional point, $\tilde{\mathbf{E}}_n^{\mathrm{L}} \to \tilde{\mathbf{E}}_m^{\mathrm{L}}$ (the two adjoint vectors coalesce), the overlap $\langle \tilde{\mathbf{E}}_n^{\mathrm{L}} | \tilde{\mathbf{E}}_n \rangle \to 0$, and $K_n \to \infty$.

The physical consequences of $K_n > 1$ are threefold. First, modal non-orthogonality changes the residue of the Green tensor and therefore the way spontaneous-emission noise and modal response are partitioned among nearby resonances. In laser theory this is often summarized as a Petermann enhancement of the effective modal noise. For an isolated mode one often writes the scaling schematically as

$$F_P \sim K_n \, \frac{3}{4\pi^2} \left(\frac{\lambda}{n} \right)^3 \frac{Q}{V_{\mathrm{eff}}}, \tag{6.13}$$

where the precise prefactor depends on the QNM normalization, emitter position, and the separation from neighboring modes. Second, the fundamental linewidth of a laser operating on mode n is broadened by K_n (the Schawlow–Townes linewidth multiplied by K_n), because spontaneous emission into the non-orthogonal mode receives contributions from all modes with nonzero overlap. Third, the sensitivity of the eigenfrequency to perturbations scales as $K_n^{1/2}$, connecting the Petermann factor to the enhanced sensing at exceptional points discussed in Chapter 8.

Example: plasmonic nanoparticle. For a silver nanoparticle of radius $a = 30$ nm in air at the localized surface plasmon resonance ($\lambda \approx 360$ nm), the mode volume is $|\tilde{V}| \approx 4 \times 10^{-5} \, \lambda^3$ and the Petermann factor is $K \approx 1.3$—close to unity because the mode is nearly Hermitian (low radiation loss). For a plasmonic dimer with a 2 nm gap, the mode hybridization and near-field enhancement push K to ~ 3–5, and for coupled cavities tuned near an exceptional point, K can exceed 10^3, dominating the noise budget.

6.6 Modal expansion of the Green tensor

The electromagnetic Green tensor $\mathbf{G}(\mathbf{r}, \mathbf{r}'; \omega)$ describes the field at $\mathbf{r}$ due to a point dipole source at $\mathbf{r}'$. In the spectral neighborhood of a QNM resonance $\tilde{\omega}_n$, the Green tensor is dominated by the QNM contribution:

$$\mathbf{G}(\mathbf{r}, \mathbf{r}'; \omega) \approx \sum_n \frac{c^2\, \tilde{\mathbf{E}}_n(\mathbf{r}) \otimes \tilde{\mathbf{E}}_n(\mathbf{r}')}{2\tilde{\omega}_n(\omega - \tilde{\omega}_n)}, \tag{6.14}$$

where the sum runs over QNMs near ω and the unconjugated normalization (6.8) has been used.

6.6.1 Local density of states

The local density of optical states (LDOS) at position $\mathbf{r}$ is proportional to the imaginary part of the Green tensor:

$$\rho(\mathbf{r}, \omega) = \frac{2\omega}{\pi c^2}\, \mathrm{Im}\big[\mathrm{tr}\, \mathbf{G}(\mathbf{r}, \mathbf{r}; \omega)\big]. \tag{6.15}$$

Near a QNM resonance, the LDOS exhibits an approximately Lorentzian peak with width $2\gamma_n$ (where $\gamma_n = -\,\mathrm{Im}\,\tilde{\omega}_n$ is the half-width at half-maximum) and amplitude proportional to $Q_n / \mathrm{Re}\,\tilde{V}_n$.

6.6.2 Purcell factor

The Purcell factor—the enhancement of the spontaneous emission rate of a quantum emitter at $\mathbf{r}_0$ relative to free space [37, 22]—is the dipole-projected LDOS ratio

$$F_P = \frac{\hat{\mathbf{d}} \cdot \mathrm{Im}\, \mathbf{G}(\mathbf{r}_0, \mathbf{r}_0; \omega) \cdot \hat{\mathbf{d}}}{\hat{\mathbf{d}} \cdot \mathrm{Im}\, \mathbf{G}^{(0)}(\mathbf{r}_0, \mathbf{r}_0; \omega) \cdot \hat{\mathbf{d}}}, \tag{6.16}$$

where $\hat{\mathbf{d}}$ is the dipole orientation and $\mathbf{G}^{(0)}$ is the free-space Green tensor. If $\mathbf{G}$ is replaced by the scattered Green tensor, the equivalent form is $F_P = 1 + \Gamma_{\mathrm{scatt}}/\Gamma_0$; Eq. (6.16) uses the total Green tensor, so no extra background 1 is added. Near an isolated QNM resonance with weak modal overlap, the QNM expansion (6.14) gives

$$F_P \approx \frac{3}{4\pi^2} \left(\frac{\lambda_n}{n}\right)^3 \frac{Q_n}{\mathrm{Re}\,\tilde{V}_n}, \tag{6.17}$$

for emission near a single QNM resonance, where λ_n is the free-space wavelength, n is the local refractive index, and $\tilde{V}_n$ is expressed as a *physical* volume (in units of length^3). When the mode volume is instead quoted in units of λ^3 or $(\lambda/n)^3$, the prefactor simplifies accordingly.

Numerical example: plasmonic nanocavity. For a gold nanoparticle dimer with a 2 nm gap in air, a full-wave simulation gives $\tilde{\omega} \approx (3.3 - 0.12i) \times 10^{15}$ rad/s, corresponding to $\lambda \approx 570$ nm and $Q = \mathrm{Re}\,\tilde{\omega}/(2|\,\mathrm{Im}\,\tilde{\omega}|) = 3.3/(2 \times 0.12) \approx 13.8$. The complex mode volume is $\tilde{V} \approx (3.1 - 1.8i) \times 10^{-7}\,\lambda^3$, so $\mathrm{Re}\,\tilde{V} \approx 3.1 \times 10^{-7}\,\lambda^3$. The resulting Purcell factor at the gap centre is

$$F_P \approx \frac{3}{4\pi^2} \frac{13.8}{3.1 \times 10^{-7}} \approx 3.4 \times 10^6, \tag{6.18}$$

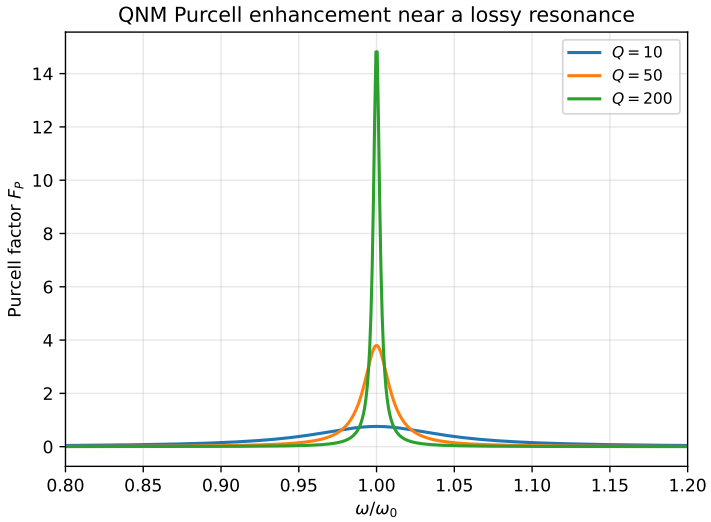


Figure 6.3 Purcell enhancement factor $F_P(\omega)$ near a lossy optical resonance at ω_0, computed from the QNM expansion for three quality factors $Q = 10$, 50, and 200. Higher Q gives a taller and narrower Lorentzian peak, illustrating the trade-off between enhancement and bandwidth. In non-Hermitian systems, the Petermann factor introduces an additional correction (Section 6.8).

where $n = 1$ in the air gap and the mode volume is quoted in units of λ^3, so the factor $(\lambda/n)^3$ in Eq. (6.17) cancels the λ^3 denominator. This enormous enhancement is a direct consequence of the extreme field confinement in the sub-nanometre gap. In practice, the observable quantum yield would be reduced by nonlocal screening in the metal, quenching into lossy surface modes, and the finite size and orientation of the emitter dipole; the QNM Purcell factor is the radiative-LDOS estimate in the point-dipole limit. The Petermann factor for such a plasmonic dimer is typically $K \approx 3$–5 (Section 6.5.4), so the effective linewidth of a laser built on this mode would be broadened by the same factor. The QNM Purcell factor (6.17) already accounts for the complex mode volume; the additional Petermann correction in (6.13) applies when excess quantum noise from mode non-orthogonality must also be included.

The QNM Purcell factor

The standard Purcell formula $F_P \propto Q/V$ remains valid for open cavities as a single-mode approximation, provided V is understood as the complex QNM mode volume and the observable rate is obtained from the imaginary part of the Green tensor. In weakly lossy, isolated resonances this often reduces to using $\mathrm{Re}\,\tilde{V}$ in the denominator. Using a Hermitian, conjugated mode volume in place of the unconjugated QNM normalization generally gives incorrect results for open or dissipative cavities.

6.7 Perturbation theory for QNMs

The biorthogonal framework provides a natural perturbation theory for QNM eigenfrequencies. If the permittivity is perturbed by $\delta\varepsilon_\mathrm{r}(\mathbf{r})$ (e.g., by a nanoparticle binding event, a temperature change, or a refractive-index modulation), the first-order shift of the n-th QNM frequency is

$$\delta\tilde{\omega}_n = -\frac{\tilde{\omega}_n}{2}\frac{\int \delta\varepsilon_\mathrm{r}(\mathbf{r})\,\tilde{\mathbf{E}}_n(\mathbf{r})\cdot\tilde{\mathbf{E}}_n(\mathbf{r})\;\mathrm{d}^3r}{\int \varepsilon_\mathrm{r}(\mathbf{r})\,\tilde{\mathbf{E}}_n(\mathbf{r})\cdot\tilde{\mathbf{E}}_n(\mathbf{r})\;\mathrm{d}^3r}. \tag{6.19}$$

This formula is the non-Hermitian analogue of the standard Rayleigh–Schrödinger perturbation theory, with the unconjugated norm replacing the Hermitian norm. Several features distinguish it from the Hermitian case:

- The perturbation shift $\delta\tilde{\omega}_n$ is complex: both the resonance frequency and the decay rate change simultaneously.
- The integrals involve $\tilde{\mathbf{E}}_n \cdot \tilde{\mathbf{E}}_n$ (unconjugated), not $|\tilde{\mathbf{E}}_n|^2$ (conjugated). For QNMs with large spatial divergence, these are very different.
- Near an exceptional point, where two QNMs coalesce, the denominator approaches zero and the perturbation theory breaks down—the response becomes $\propto \sqrt{\delta\varepsilon_\mathrm{r}}$ rather than linear.

Example: nanoparticle sensing. Consider a dielectric nanoparticle of radius $a = 30$ nm and permittivity $\varepsilon_p = 2.25$ (silica) placed at the field maximum of a photonic crystal cavity with $\tilde{\omega}_0 = (1.22 - 0.015i) \times 10^{15}$ rad/s ($\lambda \approx 1550$ nm, $Q \approx 40$). The perturbation is $\delta\varepsilon = (\varepsilon_p - 1) = 1.25$ over the nanoparticle volume $V_p = (4/3)\pi a^3 \approx 1.13 \times 10^{-22}$ m^3. Using (6.19) with a mode volume $\tilde{V} \approx (0.8 - 0.2i)(\lambda/n)^3$ and $\tilde{\mathbf{E}}(\mathbf{r}_0) \cdot \tilde{\mathbf{E}}(\mathbf{r}_0) \approx 1/\tilde{V}$ (from the normalization), the fractional frequency shift is

$$\frac{\delta\tilde{\omega}}{\tilde{\omega}_0} \approx -\frac{\delta\varepsilon\, V_p}{2\,\varepsilon\,\tilde{V}} \approx -(1.4 - 0.3i) \times 10^{-4}. \tag{6.20}$$

The real part gives a resonance shift $\delta\omega \approx -1.7 \times 10^{11}$ rad/s ($\delta\lambda \approx 0.22$ nm), while the imaginary part gives a linewidth change $\delta\gamma \approx 3.7 \times 10^{10}$ rad/s. Both are measurable in a standard transmission experiment. The key insight is that the QNM perturbation theory predicts the linewidth shift as well as the frequency shift, in contrast to Hermitian perturbation theory which gives only the real frequency shift.

6.8 Modal non-orthogonality and the Petermann factor

In a Hermitian system, the eigenmodes are orthogonal and the overlap matrix $\langle \mathbf{H}_m | \mathbf{H}_n \rangle$ is diagonal. In a non-Hermitian system, the QNMs are biorthogonal but generally not orthogonal under the Hermitian inner product. This non-orthogonality has measurable physical consequences.

The *Petermann excess noise factor* K_n for a laser mode $\tilde{\mathbf{E}}_n$ quantifies the degree of modal non-orthogonality:

$$K_n = \frac{\langle \tilde{\mathbf{E}}_n | \tilde{\mathbf{E}}_n \rangle \, \langle \tilde{\mathbf{E}}_n^{\mathrm{L}} | \tilde{\mathbf{E}}_n^{\mathrm{L}} \rangle}{|\langle \tilde{\mathbf{E}}_n^{\mathrm{L}} | \tilde{\mathbf{E}}_n \rangle|^2} \geq 1. \tag{6.21}$$

For an orthogonal (Hermitian) mode, $K_n = 1$. Near an exceptional point, where two QNMs coalesce, K_n diverges: the mode becomes maximally non-orthogonal, and the laser linewidth is enhanced by a factor K_n above the Schawlow–Townes limit.

This connection between modal non-orthogonality, excess noise, and exceptional points is one of the central links between the abstract theory of non-Hermitian operators and measurable optical phenomena. It will be revisited in Chapter 8, where exceptional-point sensing and noise limits are analyzed in detail.

6.9 Completeness and limitations

6.9.1 The completeness question

The QNMs of an open system do not, by themselves, form a complete basis in the mathematical sense [38, 35, 23]. The full spectral decomposition includes the continuous spectrum of radiation modes (Section 4.7) [5].

In the Green-tensor expansion (6.14), the QNM sum captures the *pole* contribution to $\mathbf{G}(\omega)$. The remaining contribution—the integral over the branch cut in the complex ω plane, which represents the continuous radiation spectrum—is missed. This branch-cut contribution is typically small near a QNM resonance (where the pole dominates) but can be significant far from resonance or for broadband excitations.

The branch-cut integral. For a one-dimensional system, the Green function can be written exactly as

$$G(z,z';\omega) = \sum_n \frac{\tilde{E}_n(z)\,\tilde{E}_n(z')}{2\tilde{\omega}_n(\omega-\tilde{\omega}_n)} + \frac{1}{2\pi i}\int_{\mathcal{B}} \frac{g(z,z';\omega')}{\omega-\omega'}\,\mathrm{d}\omega', \tag{6.22}$$

where the first sum is over QNM poles and the second integral is over the branch cut $\mathcal{B}$ in the complex ω plane, which runs along the negative imaginary axis for outgoing-wave boundary conditions. The branch-cut integrand $g(z,z';\omega')$ represents the contribution of radiation modes that form a continuum rather than discrete poles.

For a high-Q resonator driven near a single QNM, the pole term dominates by a factor of $\sim Q$, and the branch-cut contribution can be neglected. For broadband excitation or for frequencies between well-separated QNMs, the branch-cut contribution provides a smooth background that fills in the spectral gaps between resonances. The Riesz-projection method of Wu and Lalanne [39] handles this background systematically by computing both the pole residues and the background response from a single set of complex-frequency simulations.

6.9.2 When is a QNM expansion sufficient?

The accuracy of a truncated QNM expansion depends on:

1. the Q of the relevant modes (higher Q means narrower resonances and better QNM dominance);
2. the spectral distance from other resonances (isolated resonances are better described);
3. the spatial extent of the source or field of interest (sources far from the resonator couple more to radiation modes);
4. the number of modes retained (keeping more QNMs captures more of the spectral response, but the branch cut is always missed).

For most nanophotonic applications—Purcell enhancement, SERS, nonlinear harmonic generation in and near a single resonator—a few-QNM expansion is sufficient. For broadband scattering problems, a background correction or a Riesz-projection approach is needed to capture the non-resonant contribution [39].

6.9.3 Relation to coupled-mode theory

A QNM expansion truncated to a small number of modes is, in effect, a *coupled-mode model*: the field is represented as a superposition of a few resonant modes, each with a complex frequency, and the dynamics are governed by a finite-dimensional non-Hermitian matrix. The difference from phenomenological coupled-mode theory is that the QNM expansion is *derived from Maxwell's equations* with controlled approximations, and the coupling coefficients are determined by the mode overlaps, not by fitting.

This connection between QNMs and effective Hamiltonians is the subject of Chapter 7, where the coupled-mode formalism is developed explicitly from the Maxwell operator.

Chapter 7

From Maxwell to Coupled Modes

Much of non-Hermitian photonics is discussed in the language of 2×2 matrices: two coupled resonances, an exceptional point where eigenvalues coalesce, a $\mathcal{PT}$-symmetric Hamiltonian with balanced gain and loss. Where do these finite-dimensional matrices come from? This chapter answers that question by deriving coupled-mode models as controlled reductions of the full Maxwell eigenproblem. The derivation makes explicit what is gained and what is lost in the reduction, and establishes the conditions under which a non-Hermitian effective Hamiltonian faithfully represents the physics of the underlying electromagnetic system.

7.1 Three levels of description

Before proceeding with derivations, it is important to distinguish three levels of electromagnetic description that are often conflated in the literature:

1. **The full Maxwell operator.** The infinite-dimensional operator $\hat{\Theta}$ or its first-order form $\hat{\mathcal{M}}$, acting on the complete space of electromagnetic fields. Self-adjoint for lossless closed systems; non-self-adjoint when any of the three conditions of Chapter 2 is violated.
2. **The complete modal expansion.** A representation of the field as a sum over *all* eigenmodes (normal modes, QNMs, radiation modes). This is formally exact but impractical: the sum involves infinitely many modes and, for open systems, the continuous spectrum.
3. **The truncated effective Hamiltonian.** A finite-dimensional matrix H_{eff} acting on a small number of retained modes. This is the object that appears in most coupled-mode analyses. It is generically non-Hermitian, even when the full Maxwell operator is self-adjoint, because the discarded modes act as an environment.

The passage from level 1 to level 3 is a projection. The non-Hermiticity of H_{eff} arises from the projection as much as from any intrinsic non-Hermiticity of the material or boundary conditions. This is the fourth road to non-Hermiticity identified in Section 2.7.4.

7.2 Temporal coupled-mode theory

7.2.1 The phenomenological starting point

Temporal coupled-mode theory (TCMT), as introduced by Haus and later systematized by Fan and collaborators [36, 26], models a set of N resonant modes coupled to M input–output channels. The dynamics of the mode amplitudes $\mathbf{a}(t) = (a_1, \ldots, a_N)^{\mathsf{T}}$ are governed by

$$\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t} = \left(-i\boldsymbol{\Omega} - \boldsymbol{\Gamma}\right)\mathbf{a} + \mathbf{K}^{\mathsf{T}}\,\mathbf{s}_{+}, \tag{7.1}$$

where $\boldsymbol{\Omega}$ is the Hermitian frequency matrix, $\boldsymbol{\Gamma}$ is the (positive semi-definite) decay-rate matrix, $\mathbf{K}$ is the coupling matrix between modes and channels, and $\mathbf{s}_{+}$ is the vector of incoming wave amplitudes. The outgoing waves are

$$\mathbf{s}_{-} = \mathbf{C}\,\mathbf{s}_{+} + \mathbf{K}\,\mathbf{a}, \tag{7.2}$$

where $\mathbf{C}$ is the direct scattering (background) matrix.

The effective Hamiltonian is

$$H_{\mathrm{eff}} = \boldsymbol{\Omega} - i\boldsymbol{\Gamma}, \tag{7.3}$$

which is manifestly non-Hermitian whenever $\boldsymbol{\Gamma} \neq 0$ (i.e., whenever the resonances have finite lifetime). The eigenvalues of H_{eff} are the complex resonance frequencies $\tilde{\omega}_n$; they coincide with the QNM frequencies of the open system when the model is accurate.

7.2.2 TCMT constraints from energy conservation

Energy conservation and time-reversal symmetry impose algebraic constraints on the TCMT parameters. For a lossless system with reciprocal coupling:

$$\boldsymbol{\Gamma} = \frac{1}{2}\mathbf{K}^{\dagger}\mathbf{K}, \tag{7.4}$$

$$\mathbf{C}\mathbf{K}^{*} = -\mathbf{K}, \tag{7.5}$$

which together ensure that $\mathbf{S}(\omega)$ is unitary on the real axis when material losses are absent. These constraints reduce the number of free parameters in the TCMT model and provide self-consistency checks.

Derivation from energy balance. Constraint (7.4) follows from requiring that the total stored energy $W = \mathbf{a}^{\dagger}\mathbf{a}$ decays at a rate equal to the total power emitted into all channels. From (7.1) with $\mathbf{s}_{+} = 0$ (no input), the energy decay rate is

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \mathbf{a}^{\dagger}\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t} + \left(\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t}\right)^{\dagger}\mathbf{a} = -2\,\mathbf{a}^{\dagger}\boldsymbol{\Gamma}\,\mathbf{a}. \tag{7.6}$$

The power emitted into channels is $P_{\mathrm{out}} = \mathbf{s}_{-}^{\dagger}\mathbf{s}_{-} = \mathbf{a}^{\dagger}\mathbf{K}^{\dagger}\mathbf{K}\,\mathbf{a}$ (using $\mathbf{s}_{+} = 0$ in (7.2)). Equating $|\,\mathrm{d}W/\,\mathrm{d}t| = P_{\mathrm{out}}$ for arbitrary $\mathbf{a}$ yields (7.4).

Constraint (7.5) follows from time-reversal symmetry: the time-reversed outgoing field must be a valid incoming field that excites the time-reversed mode amplitudes. In the frequency domain, this requires $\mathbf{S}(\omega)\mathbf{S}^{*}(\omega) = \mathbf{I}$ on the real axis, which, after expanding $\mathbf{S}$ in terms of H_{eff}, reduces to (7.5).

For a system with material loss or gain, the constraints are modified: $\mathbf{\Gamma}$ acquires additional contributions from intrinsic loss rates, and unitarity is replaced by a weaker bound $\mathbf{S}^\dagger\mathbf{S} \leq \mathbf{I}$ (for passive systems).

7.2.3 What TCMT does not tell us

Equations (7.1)–(7.2) are phenomenological [40, 6, 41]: they postulate a finite-dimensional linear system and constrain its parameters by energy conservation and time-reversal symmetry. They do not derive $\mathbf{\Omega}$, $\mathbf{\Gamma}$, or $\mathbf{K}$ from the electromagnetic fields. Nor do they specify which modes should be retained or how the truncation error scales.

A Maxwell-first monograph must go further: it must connect the TCMT parameters to the spatial field distributions and material properties of the photonic structure. This is what the QNM-based derivation provides.

7.3 Deriving the effective Hamiltonian from QNMs

7.3.1 Starting point: the QNM expansion

Consider a photonic structure supporting N well-separated QNMs $\{\tilde{\mathbf{E}}_n\}$ with complex frequencies $\{\tilde{\omega}_n\}$, normalized according to the unconjugated prescription of Section 6.5. The Green tensor near the resonances is (from (6.14))

$$\mathbf{G}(\mathbf{r}, \mathbf{r}'; \omega) \approx \sum_{n=1}^{N} \frac{c^2\, \tilde{\mathbf{E}}_n(\mathbf{r}) \otimes \tilde{\mathbf{E}}_n(\mathbf{r}')}{2\tilde{\omega}_n(\omega - \tilde{\omega}_n)}. \tag{7.7}$$

7.3.2 Mode amplitudes and equations of motion

Expand the electric field in the vicinity of the resonator as

$$\mathbf{E}(\mathbf{r}, t) = \sum_{n=1}^{N} a_n(t)\, \tilde{\mathbf{E}}_n(\mathbf{r}) + \mathbf{E}_{\mathrm{rad}}(\mathbf{r}, t), \tag{7.8}$$

where $a_n(t)$ are time-dependent mode amplitudes and $\mathbf{E}_{\mathrm{rad}}$ is the contribution from the radiation continuum and distant modes. Projecting the time-domain Maxwell equations onto the QNM basis using the biorthogonal inner product yields, after neglecting the continuum coupling and assuming slowly varying amplitudes,

$$\frac{\mathrm{d}a_n}{\mathrm{d}t} = -i\tilde{\omega}_n\, a_n - i \sum_{m \neq n} V_{nm}\, a_m + f_n(t), \tag{7.9}$$

where V_{nm} is the coupling between QNMs n and m induced by any perturbation (e.g., a nearby scatterer or a change in ε_{r}), and $f_n(t)$ represents the driving by external sources.

7.3.3 Mode-overlap integrals

The coupling matrix elements V_{nm} are determined by mode-overlap integrals. For a permittivity perturbation $\delta\varepsilon_{\mathrm{r}}(\mathbf{r})$ (e.g., bringing two cavities into proximity), the coupling is

$$V_{nm} = -\frac{\tilde{\omega}_n}{2} \int \delta\varepsilon_{\mathrm{r}}(\mathbf{r})\, \tilde{\mathbf{E}}_n(\mathbf{r}) \cdot \tilde{\mathbf{E}}_m(\mathbf{r})\ \mathrm{d}^3 r, \tag{7.10}$$

where the integral is over the region where $\delta\varepsilon_{\mathrm{r}} \neq 0$. This formula connects the abstract matrix element to a spatial integral of the electromagnetic fields and the material perturbation: placing the perturbation where both mode fields are strong maximizes the coupling.

For a reciprocal medium, Lorentz reciprocity makes the properly normalized first-order coupling matrix complex symmetric, but not Hermitian. Equation (7.10) is a compact scalar perturbation formula; when the frequencies of the two modes differ substantially, the prefactor and normalization must be treated with care. Near a degenerate pair it is common to use a symmetrized reference frequency, giving $V_{nm} = V_{mn}$ to the same perturbative order. The important point is that reciprocity implies transpose symmetry, not complex-conjugate symmetry.

7.3.4 The effective Hamiltonian

In matrix form, the homogeneous part of (7.9) is

$$i\,\frac{\mathrm{d}\mathbf{a}}{\mathrm{d}t} = H_{\mathrm{eff}}\,\mathbf{a}, \qquad (H_{\mathrm{eff}})_{nm} = \tilde{\omega}_n\,\delta_{nm} + V_{nm}. \tag{7.11}$$

This is a non-Hermitian matrix whose diagonal entries are the complex QNM frequencies and whose off-diagonal entries are the inter-mode couplings. It is the Maxwell-derived counterpart of the phenomenological $H_{\mathrm{eff}} = \boldsymbol{\Omega} - i\boldsymbol{\Gamma}$ from TCMT, with the crucial advantage that every matrix element is determined by the electromagnetic fields and material properties.

From Maxwell to matrix

The non-Hermitian effective Hamiltonian of coupled-mode theory is not a postulate—it is a controlled approximation to the full Maxwell eigenproblem, obtained by expanding in QNMs and projecting out the radiation continuum. Its non-Hermiticity has two sources: the intrinsic complex frequencies of the QNMs (from radiation or absorption) and the truncation of the modal basis.

7.4 Example: two coupled cavities

Consider two identical photonic crystal cavities, each supporting a single QNM at complex frequency $\tilde{\omega}_0 = \omega_0 - i\gamma_0$, separated by a barrier of N_b lattice periods.

7.4.1 Isolated-cavity QNMs

Each isolated cavity has a QNM field $\tilde{\mathbf{E}}_0(\mathbf{r})$ that decays evanescently into the barrier as $\exp(-\kappa_{\mathrm{gap}} x)$, with characteristic length $\ell_{\mathrm{ev}} = 1/\kappa_{\mathrm{gap}}$. Here κ_{gap} is the evanescent decay constant at the gap frequency and a is the lattice period.

When the cavities are brought to a separation $d = N_b a$, the field of cavity 1 has a small but nonzero amplitude at the location of cavity 2, and vice versa. The perturbation is the presence of the other cavity, described by $\delta\varepsilon_{\mathrm{r}}(\mathbf{r})$ in the region of cavity 2.

7.4.2 Coupling strength

The coupling matrix element (7.10) gives

$$V_{12} = -\frac{\tilde{\omega}_0}{2} \int_{\text{cav } 2} \delta\varepsilon_{\mathrm{r}}(\mathbf{r})\, \tilde{\mathbf{E}}_1(\mathbf{r}) \cdot \tilde{\mathbf{E}}_2(\mathbf{r})\ \mathrm{d}^3 r \approx \kappa_0\, e^{-\kappa_{\mathrm{gap}} d}, \tag{7.12}$$

where κ_0 is a prefactor determined by the QNM field profiles and the dielectric contrast. The exponential dependence on the barrier thickness d is the signature of evanescent coupling.

For a typical silicon photonic crystal cavity at $\lambda = 1550$ nm with $Q_0 = 10^4$ and $N_b = 3$ barrier periods, the coupling can be $\kappa \sim 10^{11}$ rad/s, much larger than the decay rate $\gamma_0 = \omega_0/(2Q_0) \sim 6 \times 10^{10}$ rad/s. In this regime, the modes hybridize strongly, producing a symmetric and an antisymmetric supermode with frequency splitting 2κ.

Quantitative example. For concreteness, take a one-dimensional photonic crystal of GaAs ($\varepsilon = 12.25$) and air with period $a = 350$ nm, supporting a defect cavity at $\lambda_0 = 1550$ nm with $Q_0 = 8000$. Then $\omega_0 = 2\pi c/\lambda_0 \approx 1.215 \times 10^{15}$ rad/s, $\gamma_0 = \omega_0/(2Q_0) \approx 7.6 \times 10^{10}$ rad/s. With $N_b = 4$ barrier periods and an evanescent decay constant $\kappa_{\mathrm{gap}}\, a \approx 1.2$, the inter-cavity coupling is $\kappa \approx \kappa_0\, e^{-4\times 1.2} \approx 8.3 \times 10^{-3}\, \kappa_0$. For $\kappa_0 \sim 5 \times 10^{13}$ rad/s (estimated from the mode-overlap integral), this gives $\kappa \approx 4.2 \times 10^{11}$ rad/s, so $\kappa/\gamma_0 \approx 5.5$: the system is in the strong-coupling regime. The supermode splitting is $2\kappa \approx 8.4 \times 10^{11}$ rad/s, corresponding to a wavelength splitting $\Delta\lambda \approx \lambda_0^2\, \kappa/(\pi c) \approx 1.1$ nm. This splitting is easily resolvable in a standard optical spectrum analyser, confirming that the two-mode Hamiltonian makes a testable prediction.

Reducing the barrier to $N_b = 2$ increases κ by a factor $e^{2\times 1.2} \approx 11$, bringing $\kappa/\gamma_0 \approx 60$: the modes are deeply hybridized and the splitting dominates the linewidth. Increasing the barrier to $N_b = 6$ reduces κ below γ_0, placing the system in the weak-coupling regime where the supermode splitting is unresolved and the two resonances merge into a single broadened peak.

7.4.3 The two-mode Hamiltonian

The effective Hamiltonian for the two-cavity system is

$$H_{\mathrm{eff}} = \begin{pmatrix} \tilde{\omega}_0 & V_{12} \\ V_{21} & \tilde{\omega}_0 \end{pmatrix} = \omega_0 \mathbf{I} - i\gamma_0 \mathbf{I} + \begin{pmatrix} 0 & \kappa \\ \kappa & 0 \end{pmatrix}, \tag{7.13}$$

where $\kappa = V_{12} = V_{21}$ (reciprocity) and κ is real to lowest order in the evanescent overlap. The eigenvalues are $\tilde{\omega}_\pm = \omega_0 - i\gamma_0 \pm \kappa$: two real-split modes with the same decay rate.

If the two cavities are made non-identical—one with additional gain and the other with additional loss—the diagonal entries become $\tilde{\omega}_0 \pm i\gamma_{\mathrm{gl}}$, and the Hamiltonian is the $\mathcal{PT}$-symmetric model of Chapter 9 with a Maxwell-derived coupling κ.

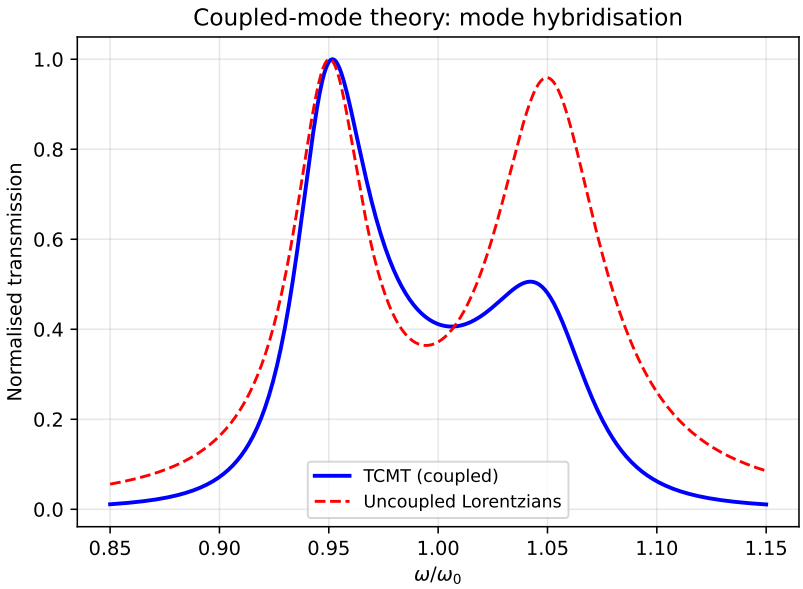


Figure 7.1 Coupled-mode hybridisation in a two-resonator system. The TCMT prediction (blue) shows the splitting produced by coherent inter-cavity coupling, while the uncoupled Lorentzian estimate (red dashed) misses the hybridised supermodes. The comparison emphasizes why the reaction and QNM-overlap derivations are needed: coupling is not a phenomenological fitting parameter but a Maxwell overlap integral between open resonator fields.

7.5 The reaction inner product

An alternative derivation of coupled-mode equations, developed by Xu and Chen [13], uses the concept of *reaction* from electromagnetic theory. The reaction of a field $(\mathbf{E}_a, \mathbf{H}_a)$ on a source $\mathbf{J}_b$ is

$$\langle a, b \rangle_{\mathrm{R}} = \int \mathbf{E}_a \cdot \mathbf{J}_b \; \mathrm{d}^3 r, \tag{7.14}$$

which is an unconjugated bilinear form—the same structure as the QNM normalization of Section 6.5.

7.5.1 Reciprocity and the symmetric coupling

For a reciprocal medium, the Lorentz reciprocity theorem (Section 2.6) ensures that $\langle a, b \rangle_{\mathrm{R}} = \langle b, a \rangle_{\mathrm{R}}$: the reaction is symmetric. This symmetry is the foundation for a *generalized coupled-mode theory* (GCMT) that works in non-Hermitian waveguides, where the standard power-conservation-based CMT fails [13].

The GCMT replaces the Hermitian power inner product $\langle \mathbf{E}_a^*, \mathbf{E}_b \rangle$ with the reaction inner product $\langle \mathbf{E}_a, \mathbf{E}_b \rangle_{\mathrm{R}}$ (no conjugation), and derives coupled-mode equations that remain valid even in $\mathcal{PT}$-symmetric and gain–loss waveguides. The resulting coupling matrix is symmetric (by reciprocity) but not Hermitian—precisely the structure expected for a non-Hermitian system that is still reciprocal.

7.5.2 Connection to QNM biorthogonality

The reaction inner product and the QNM biorthogonal inner product are closely related. For two QNMs of the same structure, the biorthogonality relation (6.7) can be rewritten as a vanishing reaction:

$$\langle \tilde{\mathbf{E}}_m,\, \varepsilon_{\mathrm{r}} \tilde{\mathbf{E}}_n \rangle_{\mathrm{R}} = \int \varepsilon_{\mathrm{r}}(\mathbf{r})\, \tilde{\mathbf{E}}_m(\mathbf{r}) \cdot \tilde{\mathbf{E}}_n(\mathbf{r})\ \mathrm{d}^3 r = 0 \qquad (m \neq n). \tag{7.15}$$

This confirms that the QNM-based and reaction-based derivations of coupled-mode theory are different presentations of the same underlying structure: both use unconjugated bilinear forms enforced by reciprocity.

7.5.3 When reaction-CMT and TCMT disagree

For Hermitian waveguides, the reaction inner product and the standard power-orthogonality inner product give equivalent coupled-mode equations. However, for non-Hermitian waveguides with gain or loss, the two approaches can give *different* coupling coefficients. The standard power-based CMT assumes $\langle \mathbf{E}_a^*, \mathbf{E}_b \rangle$ is the correct overlap, which is valid only when the modes carry positive definite power. In a gain–loss waveguide, some modes can carry negative power (the field does work on the medium rather than the reverse), and the power-orthogonality assumption fails.

The reaction-based CMT avoids this problem because the reaction inner product does not involve complex conjugation and does not require positive definiteness. It is therefore the correct starting point for deriving coupled-mode equations in $\mathcal{PT}$-symmetric waveguides, gain waveguides, and any system where the modes are not power-orthogonal.

The practical consequence is that the reaction-CMT coupling coefficient between a gain mode and a loss mode can be purely imaginary, even when the waveguides are identical except for the sign of the imaginary part of ε. This imaginary coupling is responsible for the $\mathcal{PT}$ phase transition in coupled waveguides: below the transition the coupling compensates the gain–loss imbalance, and above it the eigenvalues become complex conjugate pairs.

7.6 Two coupled resonances: the minimal model

The simplest and most frequently used non-Hermitian model in photonics is a pair of coupled resonances described by the 2×2 matrix

$$H_{\mathrm{eff}} = \begin{pmatrix} \omega_1 - i\gamma_1 & \kappa \\ \kappa & \omega_2 - i\gamma_2 \end{pmatrix}, \tag{7.16}$$

where ω_j are the bare resonant frequencies, γ_j the decay rates, and κ the coupling strength (real for a reciprocal system without radiation coupling corrections).

The eigenvalues are

$$\tilde{\omega}_{\pm} = \bar{\omega} - i\bar{\gamma} \pm \sqrt{\left(\frac{\Delta\omega - i\Delta\gamma}{2}\right)^2 + \kappa^2}, \tag{7.17}$$

where $\bar{\omega} = (\omega_1 + \omega_2)/2$, $\bar{\gamma} = (\gamma_1 + \gamma_2)/2$, $\Delta\omega = \omega_1 - \omega_2$, and $\Delta\gamma = \gamma_1 - \gamma_2$.

When $\Delta\omega = 0$ (degenerate bare frequencies), the eigenvalues become

$$\tilde{\omega}_\pm = \bar{\omega} - i\bar{\gamma} \pm \sqrt{\kappa^2 - \left(\frac{\Delta\gamma}{2}\right)^2}. \tag{7.18}$$

An *exceptional point* occurs when the square root vanishes: $\kappa = |\Delta\gamma|/2$. At this point, not only the eigenvalues but also the eigenvectors coalesce, and the matrix becomes defective (its Jordan normal form contains a 2×2 Jordan block). This is the subject of Chapter 8.

7.7 Validity and breakdown

The effective Hamiltonian (7.11) is a useful and often accurate model, but it is an approximation. Its validity requires:

1. **Mode isolation.** The retained QNMs must be spectrally well separated from all other resonances and from the continuum threshold. Overlapping resonances require more modes or a different approach.
2. **Weak perturbation.** The coupling V_{nm} should be small compared to the frequency separation between modes, so that the perturbative projection is accurate.
3. **Narrow bandwidth.** The driving field should have a bandwidth smaller than the free spectral range, so that only the retained modes are significantly excited.
4. **Linearity.** The expansion assumes linear media. Near and above the lasing threshold, gain saturation introduces nonlinear corrections that are not captured by a linear H_{eff}.

When these conditions break down, the effective Hamiltonian picture fails and one must return to the full Maxwell operator—either numerically or through a more complete modal expansion. The practical question is not whether a truncated model is ever exact, but whether the neglected modes are far enough away that their effect can be treated perturbatively. A clean way to express this is the Schur complement: partition the complete modal Hamiltonian into retained modes P and eliminated modes Q,

$$H = \begin{pmatrix} H_{PP} & H_{PQ} \\ H_{QP} & H_{QQ} \end{pmatrix}. \tag{7.19}$$

Eliminating the Q subspace gives the frequency-dependent effective operator

$$H_{\text{eff}}(\omega) = H_{PP} + H_{PQ}\,(\omega - H_{QQ})^{-1} H_{QP}. \tag{7.20}$$

The naive truncation keeps only H_{PP}. It is accurate only when $\|H_{PQ}\|^2/\Delta$ is small, where Δ is the spectral separation between the retained and eliminated modes. The Schur-complement correction is second-order in H_{PQ} and first-order in $1/\Delta$: it systematically improves the naive truncation by accounting for the virtual excitation of the eliminated modes. The truncation is accurate when the coupling-to-detuning ratio $\|H_{PQ}\|/\Delta$ is small; the leading correction scales as $\|H_{PQ}\|^2/\Delta$.

Convergence with spectral separation. For the four-mode test model used in Figure 7.2, the naive two-mode truncation error (the maximum eigenvalue error relative to the four-mode reference) is $\sim 1.1 \times 10^{-2}$ when the eliminated modes are 5γ away, and decreases to $\sim 8.5 \times 10^{-4}$ at 10γ separation. Including the Schur-complement correction reduces these errors to $\sim 4.8 \times 10^{-3}$ and $\sim 5.9 \times 10^{-5}$, respectively. The leading Schur correction scales as $\|H_{PQ}\|^2/\Delta$, so the residual error after correction is of order $\|H_{PQ}\|^3/\Delta^2$—one power higher than the naive truncation error. This is why the green curves in Figure 7.2 converge much faster with increasing spectral separation.

The practical lesson is clear: a two-mode model is adequate when the nearest unretained resonance is more than ~ 10 linewidths away. For denser spectra, more modes must be retained or the Schur complement must be applied. This criterion directly impacts the design of non-Hermitian photonic devices: a PT-symmetric laser operating near an exceptional point is sensitive to distant modes precisely when the Schur-complement correction becomes large.

Example: three-mode reduction to a two-mode $\mathcal{PT}$-symmetric Hamiltonian. Consider three QNMs with complex frequencies $\tilde{\omega}_1 = \omega_0 + i\gamma$, $\tilde{\omega}_2 = \omega_0 - i\gamma$, and $\tilde{\omega}_3 = \omega_0 + \Delta - i\gamma_3$, where $\Delta \gg \gamma, \gamma_3$ is the detuning of the third mode from the near-degenerate pair. Modes 1 and 2 form a gain–loss pair (balanced when their real frequencies coincide and their imaginary parts are opposite), while mode 3 is a distant lossy resonance. The full 3×3 Hamiltonian is

$$H = \begin{pmatrix} \omega_0 + i\gamma & \kappa_{12} & \kappa_{13} \\ \kappa_{12} & \omega_0 - i\gamma & \kappa_{23} \\ \kappa_{13} & \kappa_{23} & \omega_0 + \Delta - i\gamma_3 \end{pmatrix}, \tag{7.21}$$

where κ_{ij} are inter-mode couplings from the reaction inner product (7.14); in this example they are taken real, so the matrix is real-symmetric apart from the diagonal gain/loss terms. We treat modes 1 and 2 as the retained subspace P and mode 3 as the eliminated subspace Q, so

$$H_{PP} = \begin{pmatrix} \omega_0 + i\gamma & \kappa_{12} \\ \kappa_{12} & \omega_0 - i\gamma \end{pmatrix}, \quad H_{PQ} = \begin{pmatrix} \kappa_{13} \\ \kappa_{23} \end{pmatrix}, \quad H_{QQ} = \omega_0 + \Delta - i\gamma_3. \tag{7.22}$$

The Schur complement (7.20) gives

$$H_{\text{eff}}(\omega) = H_{PP} + \frac{1}{\omega - \omega_0 - \Delta + i\gamma_3} \begin{pmatrix} \kappa_{13}^2 & \kappa_{13}\kappa_{23} \\ \kappa_{13}\kappa_{23} & \kappa_{23}^2 \end{pmatrix}. \tag{7.23}$$

Evaluating at $\omega \approx \omega_0$ (near the retained pair) and expanding to leading order in $1/\Delta$:

$$H_{\text{eff}} \approx \begin{pmatrix} \omega_0 + i\gamma - \frac{\kappa_{13}^2}{\Delta} & \kappa_{12} - \frac{\kappa_{13}\kappa_{23}}{\Delta} \\ \kappa_{12} - \frac{\kappa_{13}\kappa_{23}}{\Delta} & \omega_0 - i\gamma - \frac{\kappa_{23}^2}{\Delta} \end{pmatrix}. \tag{7.24}$$

The distant mode has three effects. First, it shifts both frequencies downward by κ_{j3}^2/Δ for $j = 1, 2$ (a repulsive level shift familiar from second-order perturbation theory, with generally different magnitudes for the two retained modes). Second, it renormalises the coupling: the effective coupling becomes $\kappa_{\text{eff}} = \kappa_{12} - \kappa_{13}\kappa_{23}/\Delta$, which can either strengthen or weaken the

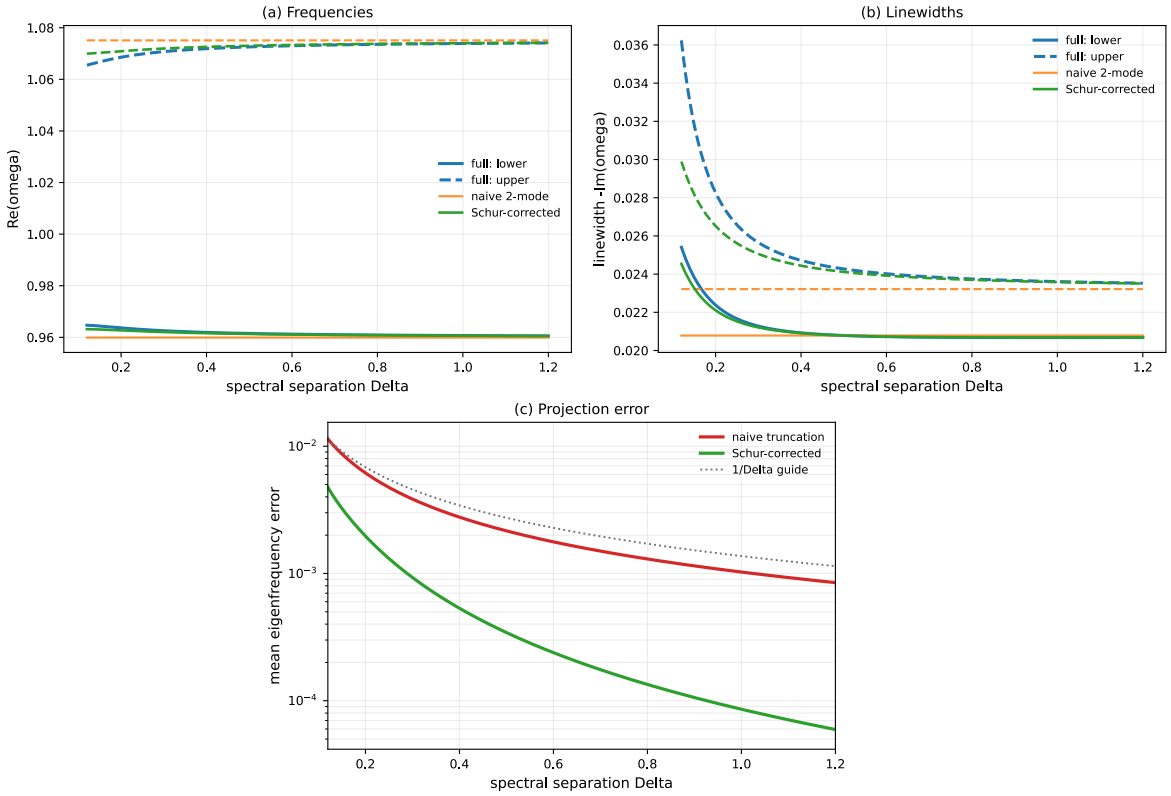


Figure 7.2 Controlled reduction from a larger QNM modal space to a two-mode coupled-mode Hamiltonian. A four-mode non-Hermitian modal model is treated as the "full" modal expansion; two near-resonant modes are retained while two distant modes are eliminated. The naive two-mode truncation captures the qualitative splitting but misses linewidth and frequency shifts from the eliminated modes. A Schur-complement correction greatly reduces the error, especially when the eliminated modes are spectrally separated.

hybridisation depending on the sign of $\kappa_{13}\kappa_{23}$. Third, the small imaginary part of the denominator adds loss corrections; when κ_{13} and κ_{23} are unequal, both the level shifts and the induced losses are different for the two retained modes. This is a Maxwell-derived mechanism by which a distant spectator mode can lift the $\mathcal{PT}$ symmetry of a nominally balanced gain–loss pair—an effect invisible to the naive two-mode truncation.

Quantitative estimate. For the parameters $\omega_0 = 1$, $\gamma = 0.05$, $\Delta = 0.5$, $\kappa_{12} = 0.03$, $\kappa_{13} = 0.02$, $\kappa_{23} = 0.01$, and $\gamma_3 = 0.02$, the naive two-mode model predicts an EP at $\gamma_{\mathrm{EP}} = \kappa_{12} = 0.03$. The Schur-corrected model shifts this to $\gamma_{\mathrm{EP}} = |\kappa_{\mathrm{eff}}| = |0.03 - 0.02 \times 0.01/0.5| = 0.0296$ and introduces a $\mathcal{PT}$-breaking detuning of $(\kappa_{13}^2 - \kappa_{23}^2)/\Delta = (4 \times 10^{-4} - 1 \times 10^{-4})/0.5 = 6 \times 10^{-4}$. For this relatively isolated pair, the corrections are small ($\sim 1\%$). But as Δ decreases toward $5\gamma = 0.25$, the corrections grow to $\sim 4\%$ in coupling and $\sim 2.4 \times 10^{-3}$ in detuning—large enough to shift the EP position measurably in experiments that probe the eigenvalue splitting with sub-linewidth resolution.

7.8 Beyond two modes

7.8.1 Multi-mode effective Hamiltonians

For systems with more than two relevant QNMs, the effective Hamiltonian is an $N \times N$ non-Hermitian matrix with entries determined by the QNM frequencies and mode-overlap integrals. The structure of this matrix reflects the symmetries of the photonic system:

- **Reciprocity:** the coupling matrix V_{nm} is symmetric ($V_{nm} = V_{mn}$), but not Hermitian.
- **$\mathcal{PT}$ symmetry:** if the permittivity satisfies $\varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(-\mathbf{r})$, the effective Hamiltonian commutes with the $\mathcal{PT}$ operator, constraining its eigenvalue structure.
- **Rotation symmetry:** if the structure has C_n rotational symmetry, the QNMs can be labeled by angular momentum quantum numbers, and the Hamiltonian block-diagonalizes.

From coupled cavities to tight-binding photonics. When the photonic structure is a periodic lattice of identical cavities, each supporting a single QNM at $\tilde{\omega}_0$, the multi-mode effective Hamiltonian becomes a tight-binding matrix:

$$(H_{\mathrm{eff}})_{nm} = \tilde{\omega}_0\, \delta_{nm} + \kappa_{nm}\,(1 - \delta_{nm}), \tag{7.25}$$

where κ_{nm} is the evanescent coupling between sites n and m, determined by the mode-overlap integral (7.10). For nearest-neighbour coupling only, $\kappa_{nm} = \kappa\,(\delta_{n,m+1} + \delta_{n,m-1})$, and the dispersion relation is $\tilde{\omega}(\mathbf{k}) = \tilde{\omega}_0 + 2\kappa\cos(k\,a)$—the photonic analogue of the electronic tight-binding band. This is the starting point for the Su–Schrieffer–Heeger (SSH) model of Chapter 12, where alternating coupling strengths $\kappa_1 \neq \kappa_2$ produce a topological band gap, and for the Hatano–Nelson model of Chapter 14, where asymmetric couplings $\kappa_L \neq \kappa_R$ require non-reciprocal elements, synthetic gauge bias, or reservoir-engineered dissipative couplings rather than ordinary reciprocal gain and loss alone. The crucial point is that all these tight-binding models have a Maxwell-first origin through the QNM mode-overlap integrals derived in this chapter.

7.8.2 From coupled modes to scattering matrix

The TCMT output relation (7.2) connects the effective Hamiltonian to the full scattering matrix. In the frequency domain, the scattering matrix is

$$\mathbf{S}(\omega) = \mathbf{C} - i\mathbf{K}\,(\omega\mathbf{I} - H_{\mathrm{eff}})^{-1}\,\mathbf{K}^{\mathsf{T}}, \tag{7.26}$$

where the poles of $\mathbf{S}$ are the eigenvalues of H_{eff} and the residues are determined by the coupling to the channels. This formula provides the bridge between the effective Hamiltonian (Chapters 8–9) and the pole–zero scattering theory (Chapter 10).

Remark 7.1. The widespread use of 2×2 models in non-Hermitian photonics is both a strength and a danger. The models are analytically tractable and capture essential physics (exceptional points, $\mathcal{PT}$ transitions, mode coalescence), but they can also mislead when applied outside their domain of validity. Throughout this book, we will use effective Hamiltonians as tools for understanding, while always checking their predictions against the full Maxwell theory whenever quantitative accuracy is needed.

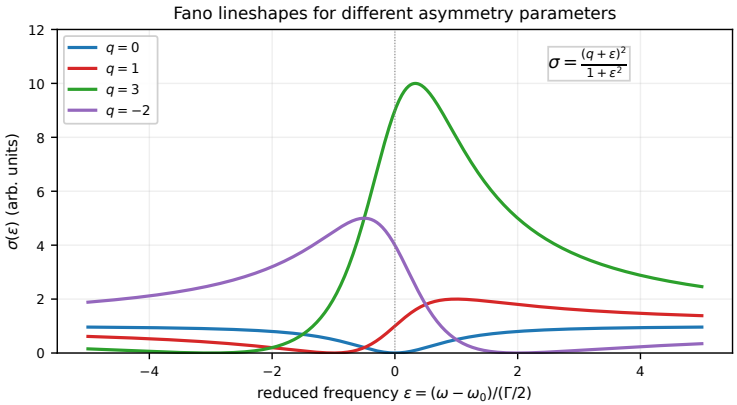


Figure 7.3 Fano lineshapes for different asymmetry parameters q. The Fano formula $\sigma(\varepsilon) = (q + \varepsilon)^2/(1 + \varepsilon^2)$, where $\varepsilon = (\omega - \omega_0)/(\Gamma/2)$ is the reduced frequency, gives a symmetric antiresonance for $q = 0$ and approaches a Lorentzian peak only in the large-$|q|$ limit after normalization. Finite positive and negative q values produce asymmetric peak–dip profiles on opposite sides of the resonance. The asymmetry arises from interference between a discrete resonance and a continuum pathway— precisely the structure captured by the TCMT scattering matrix (7.26). The dotted line marks zero detuning.

Part III

Exceptional Photonics and Non-Hermitian Scattering

Chapter 8

Exceptional Points

In Hermitian physics, degeneracies are points where eigenvalues coincide but eigenvectors remain independent. Non-Hermitian operators admit a different type of degeneracy: the exceptional point*, where both eigenvalues* and *eigenvectors coalesce. The operator becomes defective—it cannot be diagonalized—and the response of the system changes qualitatively. Exceptional points are not mathematical curiosities; they have been observed in optical microcavities, coupled waveguides, microwave billiards, and phononic crystals, and they underlie proposals for enhanced sensing, unidirectional lasing, and topological mode switching. The theory of exceptional points is developed from the non-Hermitian matrix formalism of Chapter 7 and connects it to measurable electromagnetic phenomena.*

8.1 Degeneracy versus defectiveness

In a Hermitian operator, two eigenvalues can coincide at a *diabolic point* (a degeneracy). The eigenspace is two-dimensional: two linearly independent eigenvectors exist, and the operator can still be diagonalized. Perturbations generically split the degeneracy linearly.

A non-Hermitian operator allows a different kind of degeneracy. At an *exceptional point* (EP), two (or more) eigenvalues coalesce *and the corresponding eigenvectors collapse onto a single direction.* The eigenspace is one-dimensional, the algebraic multiplicity exceeds the geometric multiplicity, and the operator cannot be diagonalized. Instead, its Jordan normal form contains an off-diagonal block.

Definition 8.1 (Exceptional point of order n)**.** An exceptional point of order n (EP_n) [10, 42, 43] is a point in parameter space where n eigenvalues and n eigenvectors simultaneously coalesce. The Jordan block is $n \times n$. The most common case is EP_2; higher-order exceptional points (EP_3, EP_4, …) require fine-tuning of additional parameters [44, 45].

8.1.1 Codimension counting

An EP_2 requires two real conditions (one for the real part and one for the imaginary part of the discriminant to vanish). In a two-real-parameter space, an EP_2 is therefore generically an

isolated point. In a three-parameter space, EP_2s form curves; in higher-dimensional parameter spaces they form surfaces or manifolds. This codimension argument explains why EP_2s are ubiquitous in two-parameter photonic systems (e.g., coupled cavities tuned by coupling strength and gain–loss contrast) but require no fine-tuning beyond what the experimental setup naturally provides.

An EP_n has codimension $2(n-1)$: it requires $2(n-1)$ real parameters to be simultaneously tuned. This is why EP_3 and higher are progressively more difficult to realize.

8.2 The 2×2 exceptional point

8.2.1 Eigenvalue structure

Consider the generic two-mode effective Hamiltonian from Section 7.6:

$$H = \begin{pmatrix} \omega_0 + \delta & \kappa \\ \kappa & \omega_0 - \delta \end{pmatrix}, \tag{8.1}$$

where ω_0 is the mean frequency, δ is a complex detuning parameter, and κ is the coupling. The eigenvalues are

$$\tilde{\omega}_\pm = \omega_0 \pm \sqrt{\delta^2 + \kappa^2}. \tag{8.2}$$

An exceptional point occurs when the discriminant vanishes:

$$\delta^2 + \kappa^2 = 0 \qquad \Longrightarrow \qquad \delta = \pm i\kappa. \tag{8.3}$$

Since δ is in general complex, the EP condition is satisfied at isolated points in the space of system parameters.

For the physically important case where δ parameterizes the gain–loss contrast (purely imaginary: $\delta = i\gamma$) and κ is a real coupling, the EP condition becomes $\gamma = \pm\kappa$: the gain–loss contrast equals the coupling strength.

8.2.2 Square-root topology

Near the EP, the eigenvalues behave as

$$\tilde{\omega}_\pm \approx \omega_{\mathrm{EP}} \pm \sqrt{\alpha\, \Delta p}, \tag{8.4}$$

where Δp is the distance from the EP in parameter space and α is a constant. This *square-root* dependence—in contrast to the linear splitting at a diabolic point—is the hallmark of an EP. It means that the sensitivity of the eigenvalues to perturbations diverges as $\Delta p \to 0$: $\mathrm{d}\tilde{\omega}/\,\mathrm{d}p \propto 1/\sqrt{\Delta p}$.

The square-root topology has a concrete geometric interpretation: the eigenvalue surface $\tilde{\omega}(\delta)$ is a double-sheeted Riemann surface, and the EP is the branch point connecting the two sheets. Any path in parameter space that encircles the EP must cross from one sheet to the other—the eigenvalue labels are exchanged after one circuit (Section 8.4).

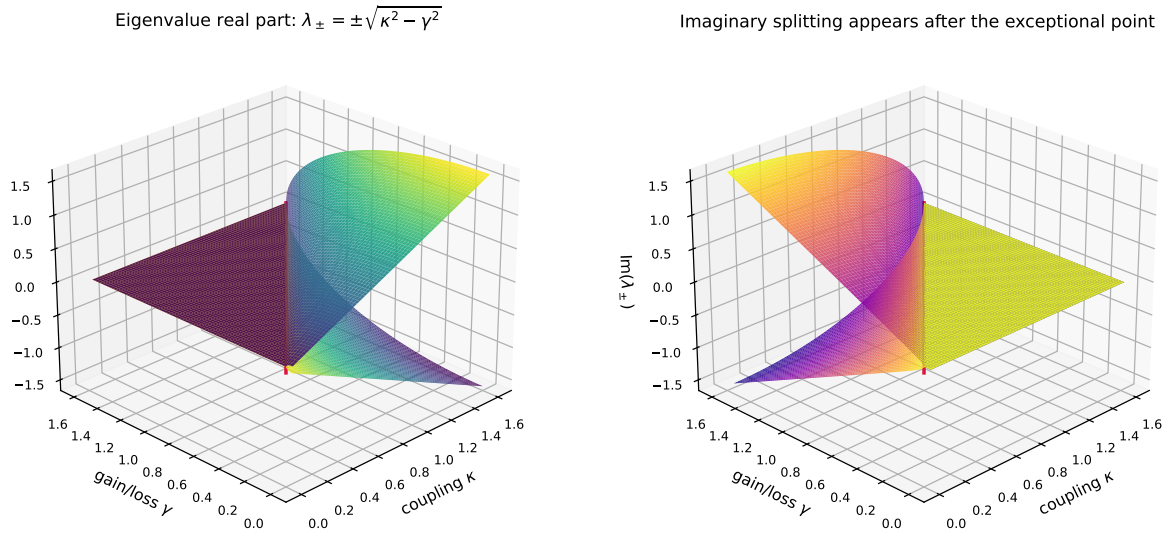


Figure 8.1 Eigenvalue surfaces of the two-mode exceptional-point model $H = \left(\begin{smallmatrix} i\gamma & \kappa \\ \kappa & -i\gamma \end{smallmatrix}\right)$. The eigenvalues $\lambda_\pm = \pm\sqrt{\kappa^2 - \gamma^2}$ are real split for $\gamma < \kappa$ and imaginary split for $\gamma > \kappa$; the red line marks the exceptional set $\gamma = \kappa$. This surface makes the square-root topology visible before any topological language is introduced.

8.2.3 Eigenvector coalescence

At the EP, the two eigenvectors of (8.1) collapse onto a single vector. Away from the EP, write the normalized right eigenvectors as

$$|\psi_+\rangle = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \qquad |\psi_-\rangle = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}, \tag{8.5}$$

where θ depends on δ/κ. As the EP is approached ($\delta \to i\kappa$), $\theta \to \pi/4 + i\infty$, and both eigenvectors approach the same direction $|\psi_{\mathrm{EP}}\rangle \propto (1, i)^{\mathsf{T}}$. The overlap $\langle\psi_+|\psi_-\rangle \to 1$: the modes become parallel.

8.3 Jordan form and the generalized eigenvector

At the EP, the matrix H cannot be diagonalized. Instead, it is brought to Jordan normal form by a *generalized eigenvector* $|\chi\rangle$ satisfying

$$(H - \omega_{\mathrm{EP}})\,|\chi\rangle = |\psi_{\mathrm{EP}}\rangle, \tag{8.6}$$

so that in the basis $\{|\psi_{\mathrm{EP}}\rangle, |\chi\rangle\}$,

$$H = \omega_{\mathrm{EP}}\,\mathbf{I} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \tag{8.7}$$

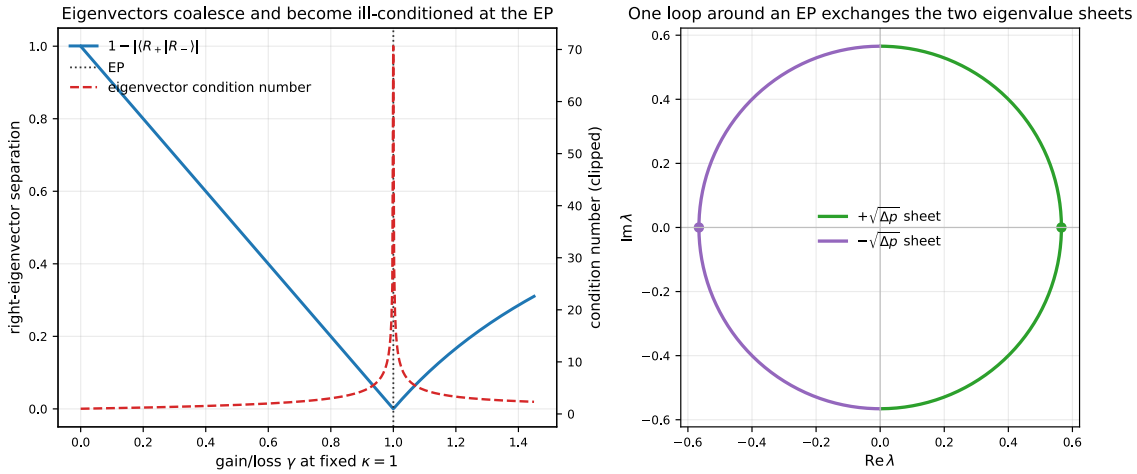


Figure 8.2 Eigenvector coalescence and branch-sheet exchange near a second-order exceptional point. Left: as γ approaches $\kappa = 1$, the two right eigenvectors become parallel and the eigenvector matrix becomes ill conditioned. Right: a closed loop around the branch point exchanges the two sheets of $\sqrt{\Delta p}$, so one circuit does not return an eigenvalue to itself.

8.3.1 Dynamical consequences

The Jordan block has physical consequences: the time evolution of a state initialized along $|\chi\rangle$ is

$$e^{-iHt}\,|\chi\rangle = e^{-i\omega_{\mathrm{EP}}t}\big(|\chi\rangle - it\,|\psi_{\mathrm{EP}}\rangle\big), \tag{8.8}$$

which grows *linearly in time* before decaying exponentially. This anomalous algebraic growth is the dynamical signature of the EP and leads to enhanced transient responses.

More generally, for an EP_n with Jordan block of size n, the time evolution of the highest generalized eigenvector involves polynomial growth $\propto t^{n-1}$ before exponential decay sets in. The transient can dominate the dynamics for times $t \lesssim 1/\Delta\gamma$, where $\Delta\gamma$ is the linewidth broadening near the EP [46, 47, 48].

8.3.2 Pseudospectrum interpretation

The defective nature of the matrix at an EP is intimately connected to its pseudospectral properties (Appendix B). Near an EP_2, the ϵ-pseudospectrum $\sigma_\epsilon(H)$ (the set of z for which $\|(zI - H)^{-1}\| \geq 1/\epsilon$) expands as

$$\sigma_\epsilon(H) \supset \{z : |z - \omega_{\mathrm{EP}}| \leq C\sqrt{\epsilon}\}, \tag{8.9}$$

where C depends on the Jordan-chain structure and normalization. The $\sqrt{\epsilon}$ scaling of the pseudospectral radius—compared to ϵ for a diagonalizable matrix—quantifies the extreme sensitivity of the spectrum at an EP.

8.4 Encircling an exceptional point

One of the most striking features of an EP is its topological character: if the system parameters are varied adiabatically along a closed loop encircling the EP in parameter space, the two

eigenvalues are *exchanged*:

$$\tilde{\omega}_+ \xrightarrow{\text{one loop}} \tilde{\omega}_- \xrightarrow{\text{two loops}} \tilde{\omega}_+. \tag{8.10}$$

This is a consequence of the square-root branch point: the eigenvalue surface $\tilde{\omega}(\delta)$ is a double-valued Riemann surface, and the EP is the branch point connecting its two sheets.

The eigenvectors also transform nontrivially under encircling. After one loop, an eigenvector acquires a geometric (Berry-like) phase *and* exchanges identity with the other eigenvector:

$$|\psi_+\rangle \xrightarrow{\text{one loop}} e^{i\phi}\,|\psi_-\rangle. \tag{8.11}$$

After *two* loops, the original eigenvector is recovered up to a geometric phase. This monodromy is unique to non-Hermitian systems and has no analogue at diabolic points.

8.4.1 Dynamic encircling

Static (adiabatic) encircling faces a practical limitation: near the EP, the gap between eigenvalues shrinks to zero, and the adiabatic condition $\dot{p}/\Delta\omega \ll 1$ cannot be maintained. Non-adiabatic transitions can prevent the eigenvalue exchange.

Dynamic encircling—varying parameters at a finite rate along a loop that passes through or near the EP—reveals richer physics. The system can exhibit chiral behavior: the final state depends on the *direction* of encircling (clockwise vs. counterclockwise). This chirality has been demonstrated experimentally in waveguide systems and provides a mechanism for topological mode switching and asymmetric transmission.

8.5 The Petermann factor at exceptional points

The Petermann factor (Section 6.8) quantifies the linewidth enhancement due to mode non-orthogonality. Near an EP_2, the Petermann factor diverges as

$$K \approx \frac{1}{|\langle L_+|R_+\rangle|^2} \propto \frac{1}{|\Delta p|}, \tag{8.12}$$

where $\langle L_+|$ and $|R_+\rangle$ are the left and right eigenvectors and Δp is the distance from the EP in parameter space. At the EP itself, $\langle L|R\rangle = 0$ and K diverges.

8.5.1 Physical consequences

The divergence of the Petermann factor has two important consequences:

1. **Linewidth broadening:** the quantum-limited linewidth of a laser operating near an EP is broadened by a factor controlled by the mode non-orthogonality, conventionally represented by K in the single-mode Schawlow–Townes limit: $\Delta\omega_{\text{laser}} = K\,\Delta\omega_{\text{ST}}$, where $\Delta\omega_{\text{ST}}$ is the Schawlow–Townes linewidth of an orthogonal mode with the same output power and decay rate. Near the EP, this broadening counteracts the improved frequency splitting.
2. **Noise amplification:** the quantum noise in any measurement performed on the system is amplified by the Petermann factor. For a homodyne or heterodyne detection scheme,

the signal scales as $\sqrt{\epsilon}$ (the EP splitting) but the noise scales as $\sqrt{K} \propto 1/|\Delta p|^{1/2}$, and the signal-to-noise ratio can remain constant or even decrease.

8.5.2 Phase rigidity

A complementary measure of non-Hermiticity is the *phase rigidity*

$$r_n = \frac{\langle L_n | R_n \rangle}{\langle R_n | R_n \rangle}, \tag{8.13}$$

after choosing a consistent normalization of the left and right eigenvectors. In complex-symmetric reciprocal models this reduces to the familiar unconjugated self-overlap divided by the usual norm. It satisfies $|r_n| = 1$ for Hermitian systems (orthogonal modes) and $|r_n| \to 0$ at an EP. With the same normalization convention, the Petermann factor and phase rigidity are related by $K_n = 1/|r_n|^2$, providing two complementary views of the same non-Hermitian effect.

8.6 Exceptional points in photonics

8.6.1 Coupled microring resonators

The most widely studied photonic EP platform consists of two evanescently coupled microring resonators, one with gain and one with loss. The effective Hamiltonian is exactly (8.1) with $\delta = i\gamma$ (the gain–loss contrast) and κ the evanescent coupling. The EP occurs at $\gamma = \kappa$, and the transition from the $\mathcal{PT}$-symmetric phase ($\gamma < \kappa$, real eigenvalue splitting) to the broken phase ($\gamma > \kappa$, complex splitting) is directly observable in the transmission spectrum.

8.6.2 Example: coupled microdisks

Consider two whispering-gallery-mode microdisks at wavelength $\lambda = 1550$ nm, each with intrinsic quality factor $Q_0 = 10^6$ and resonance frequency $\omega_0 = 2\pi c/\lambda \approx 1.216 \times 10^{15}$ rad/s. The intrinsic decay rate is $\gamma_0 = \omega_0/(2Q_0) \approx 6.08 \times 10^8$ rad/s.

One disk is pumped to provide gain, reducing its effective loss to $\gamma_{\text{gain}} = \gamma_0 - g$, while the other remains passive with $\gamma_{\text{loss}} = \gamma_0$. The gain–loss contrast is $\gamma = (\gamma_{\text{loss}} - \gamma_{\text{gain}})/2 = g/2$. The evanescent coupling rate between the disks is $\kappa = 5 \times 10^9$ rad/s (set by the inter-disk gap).

The EP condition $\gamma = \kappa$ requires $g = 2\kappa = 10^{10}$ rad/s, corresponding to a material gain coefficient $g_{\text{mat}} = 2gn_g/c \approx 1$ cm^{-1} (for group index $n_g \approx 1.5$). At the EP:

- The two eigenfrequencies coalesce at $\omega_{\text{EP}} = \omega_0 - i\gamma_0$.
- The eigenvectors collapse to $|\psi_{\text{EP}}\rangle \propto (1, i)^{\mathsf{T}}$: a circularly phased superposition.
- The Petermann factor diverges.
- A perturbation $\delta\omega_0 = 10^6$ rad/s (from a nanoparticle binding event) produces a splitting $\Delta\omega \approx \sqrt{2\kappa\,\delta\omega_0} \approx 10^8$ rad/s—a 100× enhancement over the diabolic-point splitting $\Delta\omega_{\text{DP}} = \delta\omega_0$.

This example shows both the promise and the caveat of EP sensing: the splitting is enhanced, but the linewidth broadening must be assessed independently.

Geometric design considerations. For InP-based microdisks ($n_{\text{eff}} \approx 2.6$, radius $R = 15\ \mu$m, thickness $h = 220$ nm), the whispering-gallery modes have azimuthal order $m \approx 2\pi R n_{\text{eff}}/\lambda \approx 158$ near $\lambda = 1550$ nm. The evanescent coupling rate between two identical disks separated by a gap d decays as $\kappa \propto e^{-d/\ell}$, where the decay length $\ell \approx \lambda/(2\pi\sqrt{n_{\text{eff}}^2 - 1}) \approx 100$ nm. For the coupling $\kappa = 5 \times 10^9$ rad/s used above, the required gap is $d \approx \ell \ln(\kappa_0/\kappa) \approx 300$ nm, well within electron-beam lithography precision. The EP is therefore accessible with standard nanofabrication.

The challenge lies in controlling the gain: the required material gain $g_{\text{mat}} \approx 1\ \text{cm}^{-1}$ must be matched to the coupling rate within a tolerance $\delta g/g \lesssim \gamma_0/\kappa \approx 0.12$ for the EP to be resolvable above the natural linewidth. This 12% tolerance is achievable with careful pump calibration, but drift and thermal effects can shift the system away from the EP during measurement.

8.6.3 Waveguide systems

Two coupled waveguides with different loss rates exhibit an EP as a function of the inter-waveguide spacing (which controls κ) and the loss contrast. The EP manifests as a transition from oscillatory power exchange between the waveguides to unidirectional power transfer toward the lossier guide.

8.6.4 Single-resonator exceptional points

A single resonator can exhibit EPs when two of its QNMs are brought to coalescence by tuning a geometric or material parameter. This does not require gain or $\mathcal{PT}$ symmetry: two modes of a purely passive cavity can coalesce at an EP if their radiation loss rates are appropriately matched. The EP condition is then a statement about the radiation Q factors and the inter-modal coupling mediated by the cavity shape.

8.7 Sensing at exceptional points

8.7.1 Enhanced responsivity

The square-root response (8.4) near an EP has motivated proposals for enhanced sensing. A small perturbation ϵ applied to a system operating at an EP produces an eigenvalue splitting

$$\Delta\tilde{\omega} \propto \epsilon^{1/n}, \tag{8.14}$$

where n is the EP order. For an EP_2, the splitting is $\propto \sqrt{\epsilon}$, which is much larger than the linear splitting $\Delta\omega \propto \epsilon$ of a conventional (diabolic) degeneracy for small ϵ. This enhanced spectral *responsivity* has been demonstrated experimentally in optical and microwave resonators.

8.7.2 The noise problem

However, responsivity is not the same as sensitivity. The practical utility of EP sensing depends on the *signal-to-noise ratio* (SNR), which involves both the signal (the frequency splitting) and the noise (the linewidth and measurement uncertainty).

Zhang et al. developed a quantum noise theory for EP sensors [49] using the quantum Langevin equations and input–output formalism. Their key finding is that the quantum noise in the output signal diverges at the EP at the same rate as the responsivity enhancement:

$$\mathrm{SNR}_{\mathrm{EP}} \leq \mathrm{SNR}_{\mathrm{DP}} \qquad \text{(for linear systems)}, \tag{8.15}$$

where the subscripts EP and DP refer to exceptional-point and diabolic-point sensors, respectively. The enhanced splitting is exactly canceled by the enhanced noise, and the fundamental sensitivity limit (the quantum Cramér–Rao bound) is not improved.

Quantum Fisher information analysis. The fundamental precision limit for estimating a parameter ϵ is set by the quantum Cramér–Rao bound $\delta\epsilon \geq 1/\sqrt{N\,\mathcal{F}_Q}$, where N is the number of measurements and $\mathcal{F}_Q$ is the quantum Fisher information (QFI). For a two-mode bosonic system near an EP_2, Zhang et al. showed that the QFI scales as

$$\mathcal{F}_Q \propto \frac{1}{\theta^4} \quad \text{where} \quad \theta = \sqrt{|\gamma - \kappa|} \tag{8.16}$$

measures the distance from the EP [49]. This θ^{-4} divergence arises from the covariance-matrix amplification: the output noise variance diverges as θ^{-4} through the non-orthogonality of the modes, exactly canceling the θ^{-2} responsivity enhancement from the square-root splitting. For the microdisk example above with $\theta = 0.1\kappa$, the QFI is enhanced by a factor of 10^4 relative to $\theta = \kappa$, but the noise variance is also enhanced by 10^4, leaving the SNR unchanged.

Wang et al. confirmed this prediction experimentally in a Brillouin ring laser gyroscope [50]: the enhanced transduction of rotation signals near the EP was precisely canceled by Petermann-factor linewidth broadening, leaving the fundamental angular velocity sensitivity unchanged.

8.7.3 Can nonlinearity help?

One hope for circumventing the noise–sensitivity cancellation is nonlinearity. Above the lasing threshold, gain saturation creates a nonlinear operating point where the instantaneous Hamiltonian may avoid the EP. Zheng and Chong showed, however, that noise fluctuations shift the nonlinear operating point away from the EP, effectively reducing the EP order and preventing the system from exploiting the full $\epsilon^{1/n}$ responsivity [51]. Moreover, the Bogoliubov–de Gennes Hamiltonian governing small fluctuations around the nonlinear steady state has its own EP structure, which can produce noise divergences stronger than the simple Petermann estimate from the linear system.

EP sensing: a balanced view

The frequency splitting at an EP scales as $\epsilon^{1/n}$, increasing the local responsivity. But the fundamental sensitivity (signal-to-noise ratio) may not improve because the Petermann factor and quantum noise diverge at the same rate. Practical EP sensing may still offer advantages through classical signal processing, readout engineering, or operation in regimes where noise sources differ from the fundamental quantum limit—but these advantages must be demonstrated case by case, not assumed from the $\epsilon^{1/n}$ scaling alone.

8.8 Higher-order exceptional points

An EP of order n (EP_n) occurs when n eigenvalues and eigenvectors coalesce simultaneously. The eigenvalue splitting near an EP_n scales as

$$\Delta\tilde{\omega} \propto \epsilon^{1/n}, \tag{8.17}$$

offering even greater responsivity enhancement for small perturbations.

Concrete example: EP_3 in three coupled cavities. Consider three identical microring resonators in a linear chain, each with resonance ω_0 and decay rate γ_0, with nearest-neighbour coupling κ and gain–loss distribution $(+i\gamma,\ 0,\ -i\gamma)$. The effective Hamiltonian is

$$H_{\mathrm{EP}_3} = \begin{pmatrix} \omega_0 + i\gamma & \kappa & 0 \\ \kappa & \omega_0 & \kappa \\ 0 & \kappa & \omega_0 - i\gamma \end{pmatrix}. \tag{8.18}$$

For this symmetric trimer the EP_3 occurs when the two nonzero eigenvalues collapse onto the central eigenvalue. Shift to the co-rotating frame $E \equiv \tilde{\omega} - \omega_0$ and write the characteristic polynomial of (8.18):

$$\det(H_{\mathrm{EP}_3} - \tilde{\omega}\,\mathbf{I}) = -E\left[E^2 - (2\kappa^2 - \gamma^2)\right] = 0. \tag{8.19}$$

Thus the spectrum is $E_0 = 0$ and $E_\pm = \pm\sqrt{2\kappa^2 - \gamma^2}$. At $\gamma = \sqrt{2}\,\kappa$, all three eigenvalues coalesce at $E_{\mathrm{EP}} = 0$. The matrix is then defective: it has a single eigenvector and is similar to a 3×3 Jordan block. This is the third-order exceptional point. More general EP_3 realizations require the expected four real tuning conditions, but the symmetry of this model reduces the number of independent knobs.

Now introduce a small perturbation—for instance, a detuning ϵ applied to the middle cavity, so that $H_{22} \to \omega_0 + \epsilon$. The characteristic polynomial becomes

$$-E^3 + \epsilon\,(E^2 + \gamma^2) + \cdots = 0 \qquad (\gamma = \sqrt{2}\,\kappa). \tag{8.20}$$

Near the coalescence point, write $E = E_{\mathrm{EP}} + \delta E$ and expand to leading order in ϵ. The triple-root structure means that the unperturbed part contributes $-(\delta E)^3$ at lowest order, while the perturbation contributes a constant term $\gamma^2\epsilon$. Balancing these:

$$(\delta E)^3 \sim \gamma^2\,\epsilon \quad \Longrightarrow \quad \delta E_j \approx (\gamma^2\,\epsilon)^{1/3}\,e^{2\pi i j/3}, \quad j = 0, 1, 2, \tag{8.21}$$

up to a convention-dependent phase set by the sign of the perturbation. The three eigenvalue branches are related by a $2\pi/3$ phase rotation in the complex plane, forming a three-sheeted Riemann surface. Encircling the EP_3 in parameter space cycles through all three eigenvalues, requiring three full loops to return to the starting sheet—in contrast to the two loops needed at an EP_2.

The cube-root scaling strongly affects sensing estimates. A perturbation of magnitude $|\epsilon| = 10^{-9}$ (in units of κ) produces a splitting $|\delta E| \sim 10^{-3}\kappa$, compared to $10^{-9}\kappa$ at a diabolic

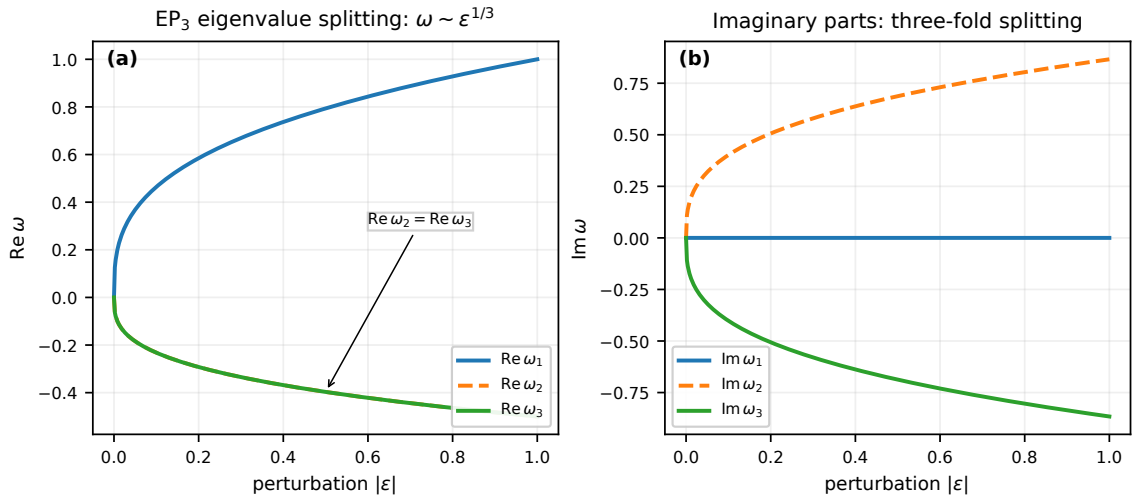


Figure 8.3 Eigenvalue splitting near a third-order exceptional point (EP$_3$). A perturbation of magnitude $|\epsilon|$ lifts the three-fold degeneracy as $\omega_j \sim \epsilon^{1/3} e^{2\pi i j/3}$ ($j = 1, 2, 3$). (a) Real parts: $\mathrm{Re}\,\omega_1 = |\epsilon|^{1/3}$ (blue, solid) rises above the EP, while $\mathrm{Re}\,\omega_2 = \mathrm{Re}\,\omega_3 = -\frac{1}{2}|\epsilon|^{1/3}$ (orange dashed and green solid, exactly overlapping) descend together—a hallmark of the cube-root topology. (b) Imaginary parts split into three distinct branches: $\mathrm{Im}\,\omega_1 = 0$ and $\mathrm{Im}\,\omega_{2,3} = \pm\frac{\sqrt{3}}{2}|\epsilon|^{1/3}$.

point and $10^{-4.5}\kappa$ at an EP$_2$. However, as discussed in Section 8.7, the Petermann factor diverges as $K \sim |\epsilon|^{-2(n-1)/n}$ near an EP$_n$, so the noise enhancement at an EP$_3$ scales as $|\epsilon|^{-4/3}$ in the same perturbative normalization. This growth can erase the apparent responsivity advantage. Whether a net SNR improvement survives depends on the specific noise model and measurement protocol.

The Jordan block at the EP$_3$ is 3×3, producing quadratic transient growth $\propto t^2$ before exponential decay:

$$e^{-iH_{\mathrm{EP}_3}\,t} = e^{-iE_{\mathrm{EP}}\,t}\left(\mathbf{I} - i\,N\,t - \tfrac{1}{2}N^2t^2\right), \tag{8.22}$$

where N is the nilpotent part ($N^3 = 0$). This quadratic growth is directly observable as a transient intensity amplification in a pulsed excitation experiment.

Experimental demonstrations of EP$_3$ have used coupled fiber loops [52] and coupled microwave cavities with carefully calibrated gain–loss profiles; in photonics, the primary challenge is maintaining the $2(n-1) = 4$ parameter tuning conditions simultaneously against fabrication disorder and thermal drift.

8.8.1 Realization challenges

Realizing EP$_n$ with $n > 2$ requires tuning $2(n-1)$ real parameters to coalescence, which becomes increasingly difficult. Experimental demonstrations of EP$_3$ have been achieved in coupled-resonator systems with engineered gain and loss profiles. The Jordan block at EP$_n$ is $n \times n$, and the time evolution involves polynomial growth $\propto t^{n-1}$ before exponential decay. Encircling an EP$_n$ in parameter space exchanges eigenvalues cyclically, with a period of n loops.

8.8.2 Noise scaling at higher-order EPs

The noise cancellation described in Section 8.7.2 becomes *more* severe at higher-order EPs. If a perturbation ϵ splits an EP$_n$ as $\Delta\omega \sim \epsilon^{1/n}$, the associated eigenvectors become increasingly

self-orthogonal and the Petermann factor typically scales as $K \sim |\epsilon|^{-2(n-1)/n}$ in the same local perturbative model. Thus the larger responsivity is accompanied by stronger excess noise and linewidth broadening. The precise SNR scaling depends on the readout and noise model, but higher order alone is not a route to guaranteed fundamental sensitivity enhancement.

8.9 Exceptional points and topology

Exceptional points are topological objects in parameter space. The eigenvalue exchange under encircling (Section 8.4) defines a topological invariant: the *discriminant number*, which counts the number of EPs enclosed by a loop. This invariant is robust against smooth deformations of the loop that do not cross an EP.

In the band structure of periodic non-Hermitian systems, EPs appear as *exceptional lines* or *exceptional surfaces* in the Brillouin zone, where two complex bands touch. These structures carry topological charges and can be classified by the homotopy groups of the relevant classifying spaces.

8.9.1 EP charges and conservation

Each EP_2 in a 2D parameter space carries a topological charge $\nu = \pm 1/2$, defined as the winding of the eigenvalue-branch phase around the EP. The total charge in any region of parameter space is conserved: EPs can be created or annihilated only in pairs with opposite charges. This conservation law constrains the possible arrangements of EPs in the Brillouin zone and explains why isolated EPs are robust against perturbations that do not create new degeneracies.

8.9.2 Connection to band topology

The connection between EPs and band topology is developed further in Chapter 12, where EPs appear as singular points in the complex band structure. The π Berry phase encircling an EP in momentum space (Section 12.4.2) provides a direct link between the EP charge and the topological invariants of the non-Hermitian band structure.

Exceptional points as organizing principle

Exceptional points are the fundamental singularities of non-Hermitian spectral theory. They organize the parameter space of photonic systems into regions of qualitatively different behavior: oscillatory versus overdamped dynamics, symmetric versus broken $\mathcal{PT}$ phases, and distinct topological sectors. Their square-root topology, Jordan-block dynamics, Petermann-factor divergence, and connection to band topology make them central to non-Hermitian photonics at every scale, from nanophotonic cavities to macroscopic laser systems.

Chapter 9

$\mathcal{PT}$-Symmetric Optics

Among all non-Hermitian photonic systems, those with parity–time ($\mathcal{PT}$) symmetry occupy a special place. A $\mathcal{PT}$-symmetric structure has a refractive-index profile that satisfies $n(\mathbf{r}) = n^*(-\mathbf{r})$: the real part is symmetric and the imaginary part is antisymmetric under spatial inversion. Gain and loss are balanced, and the system can exhibit a $\mathcal{PT}$ phase transition [53]—from a regime with entirely real eigenfrequencies to one with complex conjugate pairs—controlled by a single parameter. The $\mathcal{PT}$-symmetry condition is derived from the Helmholtz equation, followed by the transfer-matrix theory for $\mathcal{PT}$ Bragg gratings and the associated phenomena: spontaneous $\mathcal{PT}$-symmetry breaking, unidirectional invisibility, and the connection to exceptional points.

9.1 The $\mathcal{PT}$-symmetry condition

9.1.1 Parity and time reversal in Maxwell's equations

The parity operator $\mathcal{P}$ acts by spatial inversion: $\mathcal{P}\mathbf{E}(\mathbf{r}) = -\mathbf{E}(-\mathbf{r})$, $\mathcal{P}\mathbf{H}(\mathbf{r}) = \mathbf{H}(-\mathbf{r})$ (the signs follow from the vector and pseudovector character of $\mathbf{E}$ and $\mathbf{H}$). The time-reversal operator $\mathcal{T}$ acts by complex conjugation on the frequency-domain fields: $\mathcal{T}\mathbf{E}(\mathbf{r}) = \mathbf{E}^*(\mathbf{r})$, which reverses the sign of ω and exchanges incoming and outgoing waves.

A system described by a spatially varying permittivity $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is $\mathcal{PT}$-symmetric if the combined operation $\mathcal{PT}$ leaves the Helmholtz equation invariant. For the scalar Helmholtz equation

$$-\nabla^2\psi(\mathbf{r}) = \frac{\omega^2}{c^2}\, n^2(\mathbf{r})\, \psi(\mathbf{r}), \tag{9.1}$$

the $\mathcal{PT}$ transformation maps $\psi(\mathbf{r}) \to \psi^*(-\mathbf{r})$ and requires

$$\boxed{n^2(\mathbf{r}) = [n^2(-\mathbf{r})]^*,} \tag{9.2}$$

or equivalently for the permittivity: $\varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(-\mathbf{r})$. This means

$$\varepsilon_{\mathrm{r}}'(\mathbf{r}) = \varepsilon_{\mathrm{r}}'(-\mathbf{r}), \qquad \varepsilon_{\mathrm{r}}''(\mathbf{r}) = -\varepsilon_{\mathrm{r}}''(-\mathbf{r})\,: \tag{9.3}$$

the real part of the permittivity is spatially even and the imaginary part is spatially odd. Regions of gain and loss are arranged as mirror images.

9.1.2 $\mathcal{PT}$ symmetry does not mean Hermiticity

A $\mathcal{PT}$-symmetric system has complex $\varepsilon_{\mathrm{r}}(\mathbf{r})$; it is not Hermitian. However, the balance of gain and loss allows the eigenfrequencies to be real over a range of parameters—the *$\mathcal{PT}$-symmetric phase*. The eigenfrequencies become complex only when the gain–loss contrast exceeds a critical threshold: the $\mathcal{PT}$-breaking transition, which occurs at an exceptional point (Chapter 8).

It is worth stating exactly what $\mathcal{PT}$ symmetry does and does not guarantee:

- It *does* guarantee that eigenvalues come in conjugate pairs or are real: if $\tilde{\omega}$ is an eigenvalue, so is $\tilde{\omega}^*$.
- It *does not* guarantee that all eigenvalues are real. The transition from real to complex eigenvalues is spontaneous symmetry breaking: the eigenstates cease to be simultaneous eigenstates of $\mathcal{PT}$.
- It *does not* guarantee stability. Even in the $\mathcal{PT}$-symmetric phase, the non-normality of the Hamiltonian can produce large transient growth (Section B.7).

9.1.3 $\mathcal{PT}$ symmetry from the energy inner product

The relation between $\mathcal{PT}$ symmetry and the energy inner product of Chapter 2 can be made precise. The Maxwell curl–curl operator $\hat{\Theta}$ is self-adjoint with respect to the ε_{r}-weighted inner product when ε_{r} is real. When ε_{r} is complex but satisfies $\varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(-\mathbf{r})$, the operator is no longer self-adjoint in the standard inner product, but it commutes with the antilinear operator $\mathcal{PT}$:

$$[\hat{\Theta},\, \mathcal{PT}] = 0 \qquad \text{when} \qquad \varepsilon_{\mathrm{r}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(-\mathbf{r}). \tag{9.4}$$

This commutation replaces self-adjointness as the organizing symmetry of the spectrum. Bender and Boettcher first recognized this structure in quantum mechanics; in photonics, it descends directly from the Maxwell equations.

9.2 $\mathcal{PT}$ phase transition in the two-mode model

The simplest $\mathcal{PT}$-symmetric system is a pair of coupled resonances with equal and opposite gain/loss rates, described by the effective Hamiltonian

$$H_{\mathcal{PT}} = \begin{pmatrix} \omega_0 + i\gamma & \kappa \\ \kappa & \omega_0 - i\gamma \end{pmatrix}, \tag{9.5}$$

where $\gamma > 0$ is the gain/loss rate and $\kappa > 0$ is the real coupling. The eigenvalues are

$$\tilde{\omega}_\pm = \omega_0 \pm \sqrt{\kappa^2 - \gamma^2}. \tag{9.6}$$

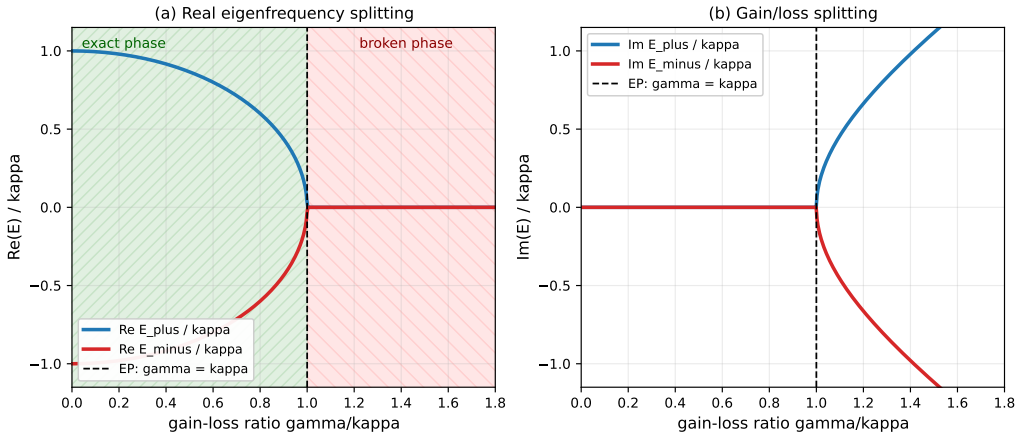


Figure 9.1 $\mathcal{PT}$ phase transition in the two-mode model. (a) Real parts of the eigenvalues: in the exact $\mathcal{PT}$-symmetric phase ($\gamma < \kappa$, green shading) both eigenfrequencies are real and split; in the broken phase ($\gamma > \kappa$, red shading) they coalesce. (b) Imaginary parts: zero throughout the exact phase, then bifurcating symmetrically in the broken phase. The exceptional point at $\gamma = \kappa$ (dashed line) marks the phase boundary where eigenvalues and eigenvectors simultaneously coalesce.

$\mathcal{PT}$-symmetric phase ($\gamma < \kappa$). The square root is real, and the eigenfrequencies $\tilde{\omega}_\pm$ are both real despite the non-Hermiticity of $H_{\mathcal{PT}}$. The modes are extended over both resonators with equal gain and loss, so each mode experiences zero net amplification.

$\mathcal{PT}$-broken phase ($\gamma > \kappa$). The square root becomes imaginary, and the eigenfrequencies form a complex conjugate pair: $\tilde{\omega}_\pm = \omega_0 \pm i\sqrt{\gamma^2 - \kappa^2}$. One mode is amplified, the other is attenuated. The gain–loss contrast has overcome the coupling.

The exceptional point ($\gamma = \kappa$). At the transition, the eigenvalues coalesce at ω_0 and the eigenvectors merge. This is an EP_2 of exactly the type analyzed in Section 8.2.

9.2.1 Coupled-mode derivation from Maxwell fields

The two-mode Hamiltonian (9.5) is not postulated but derived from the Maxwell equations via the coupled-mode formalism of Chapter 7. Consider two weakly coupled cavities with identical real permittivity profiles but opposite imaginary parts: cavity A has $\varepsilon_A = \varepsilon_0 + i\kappa_I$ and cavity B has $\varepsilon_B = \varepsilon_0 - i\kappa_I$. The QNM framework of Chapter 6 gives each isolated cavity the same resonance frequency ω_0 but opposite imaginary corrections:

$$\tilde{\omega}_A = \omega_0 - i\gamma_{\mathrm{rad}} - i\gamma_{\mathrm{mat}}, \qquad \tilde{\omega}_B = \omega_0 - i\gamma_{\mathrm{rad}} + i\gamma_{\mathrm{mat}}, \tag{9.7}$$

where γ_{rad} is the (common) radiative decay rate and $\gamma_{\mathrm{mat}} \propto \kappa_I$ is the material loss/gain rate for the sign convention used here. Thus cavity A is the lossy partner and cavity B is the gain partner for $\kappa_I > 0$. When the two cavities are brought into evanescent proximity, the overlap integral of the reaction inner product (Section 7.5) gives a coupling matrix element κ that is real and symmetric for reciprocal media. The resulting 2×2 effective Hamiltonian is exactly of the form (9.5), after subtracting the common offset $\omega_0 - i\gamma_{\mathrm{rad}}$ and identifying $\gamma = \gamma_{\mathrm{mat}}$ up to the ordering of the two basis states.

The Maxwell derivation shows explicitly that the $\mathcal{PT}$ condition for the two-mode model—equal and opposite imaginary diagonal entries—is a consequence of the $\mathcal{PT}$ condition for the permittivity, not an independent postulate. It also reveals when the two-mode description breaks down: when the coupling κ is comparable to the mode spacing, or when higher-order modes contribute significantly.

9.3 Transfer-matrix theory for $\mathcal{PT}$ gratings

9.3.1 The 1D $\mathcal{PT}$-symmetric Bragg grating

Consider a one-dimensional periodic structure with refractive index

$$n(x) = n_0 + \Delta n \cos(2\pi x/\Lambda) + i\Delta n_I \sin(2\pi x/\Lambda), \tag{9.8}$$

where n_0 is the background index, Δn is the real modulation depth, Δn_I is the imaginary modulation depth, and Λ is the period. This profile satisfies $n(x) = n^*(-x)$: it is $\mathcal{PT}$-symmetric. When $\Delta n = 0$ and only the imaginary modulation remains, the grating is purely gain–loss modulated.

9.3.2 Transfer matrix for one period

Each period of the grating is a layered structure that can be described by a 2×2 transfer matrix $\mathbf{T}$ relating the forward- and backward-propagating wave amplitudes on the left to those on the right (Appendix C). For a grating of N periods, the total transfer matrix is $\mathbf{T}^N$.

The scattering matrix is obtained from $\mathbf{T}$ by the standard conversion. The reflection and transmission coefficients are [54]

$$r_L = -\frac{T_{21}}{T_{22}}, \quad r_R = \frac{T_{12}}{T_{22}}, \quad t = \frac{1}{T_{22}}, \tag{9.9}$$

where T_{ij} are the elements of $\mathbf{T}^N$.

9.3.3 Example: a $\mathcal{PT}$ bilayer

To make the transfer-matrix formalism concrete, consider the simplest $\mathcal{PT}$ structure: a single unit cell with a loss slab ($n_1 = 3.0 + 0.05\,i$, thickness $d = 200$ nm) and a gain slab ($n_2 = 3.0 - 0.05\,i$, same thickness). At wavelength $\lambda = 1550$ nm, the propagation phase per slab is $\phi = nk_0 d \approx 2.43 + 0.041\,i$ rad.

The single-slab propagation matrix is (Appendix C)

$$M_j = \begin{pmatrix} e^{in_j k_0 d} & 0 \\ 0 & e^{-in_j k_0 d} \end{pmatrix}, \tag{9.10}$$

and the interface matrix between slabs with indices n_a and n_b is $I_{ab} = \frac{1}{2n_b}\left(\begin{smallmatrix} n_b+n_a & n_b-n_a \\ n_b-n_a & n_b+n_a \end{smallmatrix}\right)$. The unit-cell transfer matrix is $\mathbf{T}_{\text{cell}} = I_{2,\text{out}} \cdot M_2 \cdot I_{12} \cdot M_1 \cdot I_{\text{in},1}$.

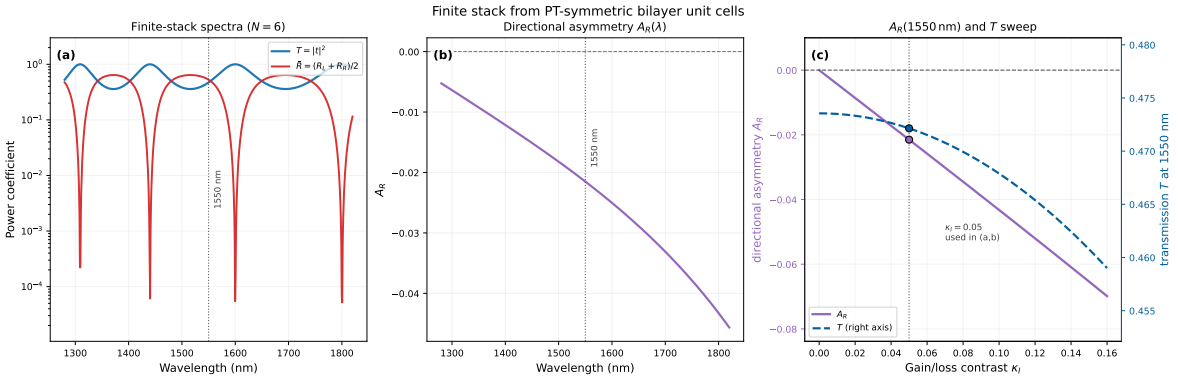


Figure 9.2 Transfer-matrix spectra of a finite stack built from $\mathcal{PT}$-symmetric bilayer unit cells. The unit cell is the loss–gain pair used in this example, repeated $N = 6$ times and embedded in air. Panel (a) shows the wavelength dependence of transmission $T = |t|^2$ and the mean reflectance $\bar{R} = (R_L + R_R)/2$; because the stack contains gain and loss, these are scattering intensities rather than a unitary energy partition. The small left–right difference is not visually useful on the logarithmic spectrum, so panel (b) plots the directional reflection asymmetry $A_R(\lambda) = (R_R - R_L)/(R_R + R_L)$ directly; negative A_R means stronger reflection for left incidence. Panel (c) fixes $\lambda = 1550\,\mathrm{nm}$ and varies the gain–loss index contrast κ_I, showing $A_R(1550\,\mathrm{nm})$ and the transmission T as functions of gain/loss strength.

For this example, the trace of $\mathbf{T}_{\mathrm{cell}}$ is nearly real and satisfies $|\frac{1}{2}\,\mathrm{tr}\,\mathbf{T}_{\mathrm{cell}}| < 1$: the system is in the $\mathcal{PT}$-symmetric phase. The left and right reflection coefficients are

$$|r_L| \approx 0.033, \qquad |r_R| \approx 0.034, \qquad |t| \approx 1.002, \tag{9.11}$$

showing near-unit transmission with slightly asymmetric reflection. The asymmetry grows with the number of unit cells N.

9.3.4 Unidirectional invisibility

Suitably designed $\mathcal{PT}$ Bragg gratings can exhibit *unidirectional invisibility*: at a scattering exceptional point or anisotropic transmission resonance, the grating can be transparent from one side ($r_L = 0$, $|t| = 1$) while remaining reflective from the other side ($r_R \neq 0$) [54]. This is not a generic consequence of every $\mathcal{PT}$-breaking point; it requires the grating parameters and operating frequency to satisfy the corresponding scattering condition.

The physical mechanism can be understood from the transfer-matrix eigenvalue structure. At the EP, one eigenvalue has unit modulus while the other is its inverse: $\lambda_+ = 1/\lambda_-^*$. The corresponding eigenvectors correspond to forward- and backward-traveling Bloch waves. For incidence from one side, the field couples predominantly to the non-growing eigenstate and passes through with unit transmission and zero reflection; for incidence from the other side, the field excites the growing eigenstate, producing finite reflection.

This asymmetry is not a violation of reciprocity—the transmission coefficient is the same from both sides ($t_L = t_R$ for a reciprocal medium)—but the reflection coefficients differ: $|r_L| \neq |r_R|$. In a lossless reciprocal one-dimensional grating, $|r_L| = |r_R|$ follows from the combination of reciprocity and flux conservation. In the $\mathcal{PT}$-symmetric case, flux conservation is broken and the two reflection magnitudes can differ even though the transmission amplitude remains reciprocal.

9.3.5 Spectral features near the Bragg condition

The most dramatic $\mathcal{PT}$ effects occur near the Bragg condition $\lambda = 2n_0\Lambda$, where the stop band of a conventional (Hermitian) grating is located. In the $\mathcal{PT}$-symmetric grating:

- Below the EP: the stop band splits into two narrower bands separated by a transmission peak at the Bragg wavelength.
- At the EP: the two bands merge, and the central peak becomes a unity-transmission point with divergent group delay—a critical slowing of light.
- Above the EP: the merged band becomes an amplification window for one scattering channel and an absorption window for the other.

These features have been observed experimentally in coupled-fiber systems and semiconductor gratings, confirming the transfer-matrix predictions quantitatively.

$\mathcal{PT}$ gratings break reflection symmetry, not reciprocity

A $\mathcal{PT}$-symmetric Bragg grating is reciprocal ($t_L = t_R$) but not unitary ($|r_L| \neq |r_R|$). The asymmetric reflection is a direct consequence of balanced gain and loss, and it vanishes in the Hermitian limit ($\Delta n_I \to 0$). This distinction between reciprocity and unitarity echoes the general principle established in Section 2.6.

9.4 $\mathcal{PT}$ symmetry of the scattering matrix

For a one-dimensional two-port $\mathcal{PT}$-symmetric structure with the port basis ordered by left and right incidence, the scattering matrix satisfies a fundamental relation derived by Chong, Ge, and Stone [55]:

$$(\mathcal{PT})\,\mathbf{S}(\omega^*)\,(\mathcal{PT}) = \mathbf{S}^{-1}(\omega). \tag{9.12}$$

For real frequencies, this implies

$$|\det \mathbf{S}(\omega)| = 1 : \tag{9.13}$$

the determinant of the scattering matrix has unit modulus, even though $\mathbf{S}$ itself is not unitary. The individual scattering eigenvalues s_1, s_2 satisfy $|s_1|\,|s_2| = 1$, but they need not individually lie on the unit circle.

9.4.1 Derivation of the unit-modulus condition

The proof follows directly from (9.12). Taking the determinant of both sides:

$$\det[(\mathcal{PT})\,\mathbf{S}(\omega^*)\,(\mathcal{PT})] = \det[\mathbf{S}^{-1}(\omega)]. \tag{9.14}$$

In this two-port basis, the parity operation exchanges the two ports and time reversal complex conjugates the scattering amplitudes. Therefore the determinant of the left side is $[\det \mathbf{S}(\omega^*)]^*$, and we obtain

$$[\det \mathbf{S}(\omega^*)]^* = [\det \mathbf{S}(\omega)]^{-1}. \tag{9.15}$$

For real ω, $\omega^* = \omega$, and this gives $|\det \mathbf{S}(\omega)|^2 = 1$, hence $|\det \mathbf{S}| = 1$.

9.4.2 Phase diagram of the scattering eigenvalues

In the $\mathcal{PT}$-symmetric phase, both scattering eigenvalues are unimodular: neither net amplification nor net absorption occurs. In the broken phase, $|s_1| > 1$ and $|s_2| < 1$: one scattering channel is amplified and the other is attenuated.

The transition between these two regimes occurs at an exceptional point of the scattering matrix, connecting the $\mathcal{PT}$ phase transition to the general EP theory of Chapter 8. This scattering EP is distinct from the Hamiltonian EP of the two-mode model, though the two coincide for the simplest structures.

9.5 Beyond paraxial: $\mathcal{PT}$ symmetry in the full Maxwell framework

Much of the $\mathcal{PT}$-symmetry literature uses the paraxial approximation or the scalar Helmholtz equation. In the full vectorial Maxwell framework, $\mathcal{PT}$ symmetry requires

$$\boldsymbol{\varepsilon}(\mathbf{r}) = \mathbf{P}\,\boldsymbol{\varepsilon}^{*}(-\mathbf{r})\,\mathbf{P}^{-1}, \tag{9.16}$$

where $\mathbf{P}$ is the matrix representation of the chosen parity operation on vector components. The same statement applies to the permeability if magnetic effects are included. For scalar isotropic media, this reduces to (9.2).

9.5.1 Vectorial $\mathcal{PT}$ condition

In two and three dimensions, the $\mathcal{PT}$ condition must account for the vector nature of the fields and the tensorial character of the material response. The parity operation involves not only spatial inversion but also the appropriate transformation of vector components. For inversion through the origin, $\mathbf{P} = -\mathbf{I}$ and the componentwise condition reduces to the scalar-looking form

$$\varepsilon'_{ij}(\mathbf{r}) = \varepsilon'_{ij}(-\mathbf{r}), \qquad \varepsilon''_{ij}(\mathbf{r}) = -\varepsilon''_{ij}(-\mathbf{r}). \tag{9.17}$$

For mirror parities, however, components with one index normal to the mirror plane acquire extra signs through Eq. (9.16). The tensor condition must therefore be applied before reducing the problem to scalar even/odd rules.

For structures with specific crystallographic symmetries, the $\mathcal{PT}$ condition can be satisfied by particular arrangements of gain and loss regions that respect the point-group symmetry of the lattice.

9.5.2 When the scalar model is not enough

The scalar Helmholtz equation captures the essential $\mathcal{PT}$ physics in one dimension and for TE/TM modes in two-dimensional slab geometries. However, the vectorial Maxwell treatment is essential in the following situations:

- **Finite-height slabs:** the TE-TM coupling at the slab interfaces modifies the $\mathcal{PT}$ phase boundary and can split a single EP into multiple EPs at different frequencies.

- **Photonic crystal cavities:** the mode profiles are fully vectorial, and the overlap integrals that determine the coupling κ in the effective Hamiltonian must be computed from the vector fields.
- **Chiral $\mathcal{PT}$ systems:** when the permittivity tensor has off-diagonal elements (optical activity or magneto-optical coupling), $\mathcal{PT}$ symmetry can produce qualitatively new phenomena, including asymmetric polarization conversion and chiral EPs.

9.6 Experimental realizations

9.6.1 Coupled optical waveguides

The first optical demonstration of $\mathcal{PT}$ symmetry used two coupled slab waveguides, one with gain (provided by optical pumping of a doped glass) and one with equivalent loss. The key experimental signature is the power evolution along the propagation direction: in the $\mathcal{PT}$-symmetric phase, the power oscillates between the two guides with a beat length $L_b = \pi/(2\sqrt{\kappa^2 - \gamma^2})$; in the broken phase, the power grows exponentially in the gain guide.

El-Ganainy et al. and Makris et al. developed the coupled-waveguide $\mathcal{PT}$ proposal [53, 56], and Rüter et al. realized it experimentally in Fe-doped $LiNbO_3$ waveguides [33].

9.6.2 Whispering-gallery microcavities

Two coupled whispering-gallery-mode microtoroid resonators—one with erbium-ion gain and one passive—demonstrated the $\mathcal{PT}$ transition and direction-dependent transmission in the broken phase at ultralow power thresholds [57]. In reciprocal devices this direction dependence comes from asymmetric excitation, nonlinear gain saturation, or asymmetric coupling to ports rather than from a violation of Lorentz reciprocity. The experiment exploited the high Q of the microtoroids ($Q \sim 10^7$) to achieve the $\mathcal{PT}$ transition with sub-milliwatt pump powers.

9.6.3 Photonic lattices and mesh networks

Arrays of coupled waveguides or fiber loops with alternating gain and loss have been used to realize $\mathcal{PT}$-symmetric lattice models, observe band merging, and study the interplay of $\mathcal{PT}$ symmetry with disorder and nonlinearity. The fiber-loop architecture is particularly versatile because the gain, loss, and coupling can be tuned independently on each round trip, enabling the realization of time-multiplexed $\mathcal{PT}$ lattices with hundreds of sites.

9.6.4 Microring resonators

Coupled microring resonators on a silicon photonic chip provide a compact platform for $\mathcal{PT}$-symmetric devices. The gain is typically provided by III–V semiconductor bonding or by stimulated Raman scattering in the silicon. Single-mode $\mathcal{PT}$ lasers and direction-dependent transmission switches have been demonstrated in this platform. In a representative silicon-photonic implementation, two microrings of radius $R = 10\,\mu\text{m}$ with intrinsic $Q \sim 5 \times 10^4$ are coupled through a gap of $\sim 200\,\text{nm}$, giving an inter-resonator coupling rate $\kappa \approx 60\,\text{GHz}$. The

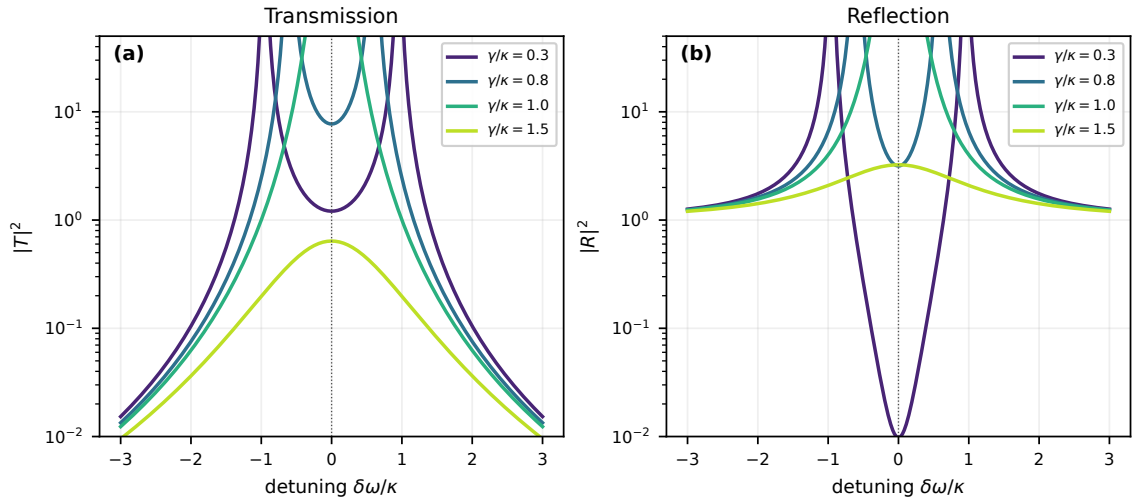


Figure 9.3 Transmission $|T|^2$ and reflection $|R|^2$ of two coupled $\mathcal{PT}$-symmetric microring resonators as functions of detuning $\delta\omega/\kappa$ for several gain–loss ratios γ/κ. Because the system contains gain, $|T|^2$ and $|R|^2$ can exceed unity; sharp resonant enhancement near $\gamma/\kappa = 1$ marks the exceptional point of the idealized two-mode model. In an actual laser, threshold additionally requires a scattering pole to reach the real-frequency axis after all intrinsic and external losses are included. The logarithmic scale reveals qualitatively different lineshapes: sub-EP ($\gamma < \kappa$, split peaks), at-EP ($\gamma = \kappa$, single divergence), and broken-$\mathcal{PT}$ ($\gamma > \kappa$, broadened single peak). Dotted vertical line marks zero detuning.

gain ring is pumped optically via a bonded III–V layer, and the EP is reached at a pump power of $\sim 0.8\,\mathrm{mW}$, corresponding to a material gain coefficient $g_{\mathrm{th}} \approx 4\ \mathrm{cm}^{-1}$.

From models to Maxwell

Many experimental realizations of $\mathcal{PT}$ symmetry operate in regimes where the scalar Helmholtz or paraxial approximation is adequate. For nanophotonic structures where vectorial effects, radiation loss, and strong dispersion matter, a full Maxwell treatment—returning to the operator framework of Chapters 1 and 2—is essential. The effective Hamiltonian description of Chapter 7 then provides the bridge between the full electromagnetics and the 2×2 models.

9.7 Non-normality and transient amplification

Even in the $\mathcal{PT}$-symmetric phase, where all eigenvalues are real, the effective Hamiltonian (9.5) is non-normal: $[H_{\mathcal{PT}}, H_{\mathcal{PT}}^{\dagger}] \neq 0$. This non-normality has a physical consequence that is invisible to the eigenvalue spectrum: *transient amplification* of signals. In the ideal balanced model, each eigenmode in the exact phase is purely oscillatory, but non-orthogonal modal superpositions can still grow transiently before the energy returns to the other quadrature.

9.7.1 The mechanism

The non-orthogonality of the eigenvectors means that a superposition of two oscillatory eigenmodes can transiently grow even when both eigenfrequencies are real. The maximum transient amplification factor is controlled by the condition number of the eigenvector matrix, which diverges at the exceptional point.

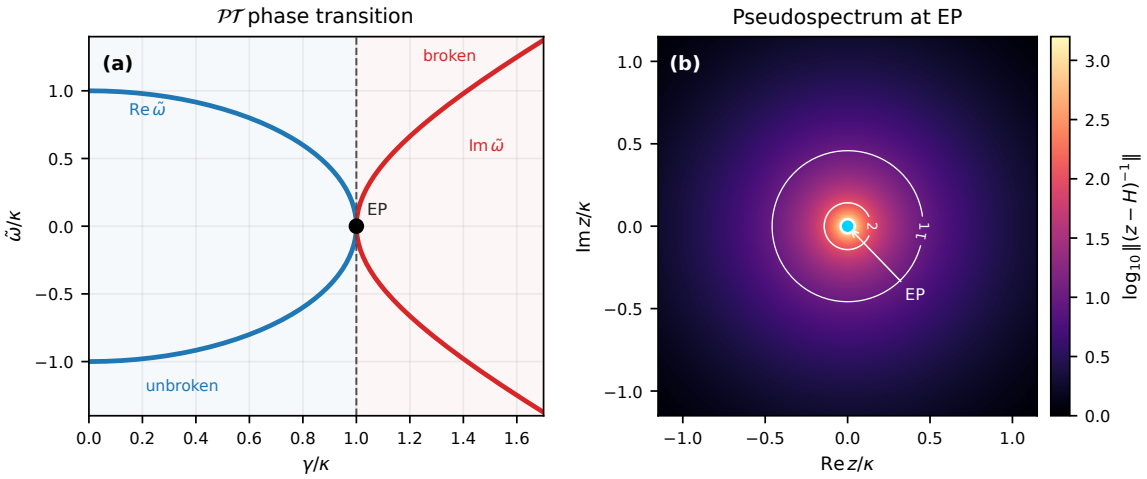


Figure 9.4 $\mathcal{PT}$ phase transition and pseudospectrum of the canonical 2×2 gain–loss Hamiltonian $H = \left(\begin{smallmatrix} i\gamma & \kappa \\ \kappa & -i\gamma \end{smallmatrix}\right)$ with $\kappa = 1$. (a) Real (blue) and imaginary (red) parts of the eigenfrequencies $\tilde{\omega}_\pm/\kappa$ as a function of γ/κ. In the unbroken phase ($\gamma < \kappa$), both eigenvalues are real: $\mathrm{Re}\,\tilde{\omega} = \pm\sqrt{1-\gamma^2/\kappa^2}$. In the broken phase ($\gamma > \kappa$), the eigenvalues form a complex conjugate pair: $\mathrm{Im}\,\tilde{\omega} = \pm\sqrt{\gamma^2/\kappa^2 - 1}$. Only nonzero branches are plotted; the complementary branch is identically zero. The black dot marks the exceptional point at $\gamma = \kappa$. (b) ε-pseudospectrum $\log_{10}\|(z-H)^{-1}\|$ at the EP ($\gamma = \kappa$). White contour labels give the $\log_{10}$ level. The pseudospectrum bulges far from the coalesced eigenvalue at the origin, quantifying the system's sensitivity to perturbations (Section 9.7.2).

Example. For the $\mathcal{PT}$-symmetric Hamiltonian (9.5) with $\omega_0 = 0$, $\kappa = 1$, and $\gamma = 0.9\kappa$ (deep in the $\mathcal{PT}$-symmetric phase but close to the EP), the eigenvalues are $\tilde{\omega}_\pm = \pm\sqrt{1-0.81} = \pm 0.436$. Both eigenvalues are real, so every eigenmode oscillates without growth. However, the maximum of $\|e^{-iH_{\mathcal{PT}}t}\|$ over all time is

$$\max_t \|e^{-iH_{\mathcal{PT}}t}\| = \sqrt{\frac{1+\gamma/\kappa}{1-\gamma/\kappa}} = \sqrt{\frac{1.9}{0.1}} \approx 4.36. \tag{9.18}$$

A pulse injected into an optimized superposition of the two cavities can be amplified by a factor of 4.36 before the periodic evolution returns the energy to the other quadrature. At $\gamma/\kappa = 0.99$, this factor rises to 14.1; at $\gamma/\kappa = 0.999$, to 44.7. The divergence $\propto 1/\sqrt{\kappa-\gamma}$ is a hallmark of non-normality near an EP.

9.7.2 Pseudospectrum

The pseudospectrum $\sigma_\epsilon(H)$—the set of complex frequencies z where $\|(z-H)^{-1}\| > 1/\epsilon$—quantifies the sensitivity of the spectrum to perturbations. For a normal operator, the ϵ-pseudospectrum is simply the union of ϵ-discs around the eigenvalues. For the non-normal $H_{\mathcal{PT}}$, the pseudospectrum bulges far beyond the eigenvalues into the complex plane, revealing the system's vulnerability to noise-driven amplification.

For the $\gamma/\kappa = 0.9$ example above, the $\epsilon = 0.1$ pseudospectral boundary extends to $|\mathrm{Im}\,z| \approx 0.23$, compared to the eigenvalue imaginary part of exactly zero. This means that an environmental perturbation of strength $\epsilon = 0.1$ can shift the effective eigenvalues by 0.23 in the imaginary direction—more than twice the perturbation strength—producing transient amplification even when the unperturbed system is nominally stable.

The pseudospectral viewpoint connects directly to the EP sensing debate of Chapter 8: the enhanced responsivity of an EP sensor is a consequence of the same non-normality that produces transient amplification and noise enhancement. The Petermann factor and the maximum transient amplification are closely related measures of eigenvector non-orthogonality, although they are not identical in general.

9.8 Nonlinear $\mathcal{PT}$ symmetry

All experimental $\mathcal{PT}$-symmetric systems require gain, and gain media are inherently nonlinear: at high intensities the gain saturates. This means that the linear $\mathcal{PT}$-symmetric Hamiltonian is always an approximation, valid only below the saturation intensity of the gain medium. Above saturation, the system is described by a nonlinear $\mathcal{PT}$-symmetric model.

9.8.1 Gain saturation and self-consistent $\mathcal{PT}$ symmetry

The simplest model for gain saturation replaces the constant gain rate γ in (9.5) with an intensity-dependent rate $\gamma(|a|^2) = \gamma_0/(1 + |a|^2/I_{\text{sat}})$, where I_{sat} is the saturation intensity. The resulting nonlinear coupled-mode equations

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \omega_0 + i\gamma(|a_1|^2) & \kappa \\ \kappa & \omega_0 - i\gamma_0 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{9.19}$$

admit steady-state solutions where the gain self-adjusts to balance the loss. This *self-consistent* $\mathcal{PT}$ symmetry can be more robust than the linear version: gain saturation drives the system toward a gain–loss balance over a finite range of operating conditions instead of requiring a perfectly fixed small-signal gain rate [3, 58].

9.8.2 Nonlinear exceptional points

Gain saturation modifies the location and character of exceptional points. In the nonlinear model (9.19), the effective gain rate at steady state is $\gamma_{\text{eff}} = \gamma_0/(1 + |a_1^{\text{ss}}|^2/I_{\text{sat}})$, which depends on the solution itself. The EP condition $\gamma_{\text{eff}} = \kappa$ must therefore be solved self-consistently: the steady-state intensity determines the gain, which determines the EP location, which determines the steady-state intensity.

For the two-mode model with $\omega_0 = 0$ and passive loss γ_0, the self-consistent EP condition is

$$\frac{\gamma_0}{1 + |a_1^{\text{ss}}|^2/I_{\text{sat}}} = \kappa, \quad \Rightarrow \quad |a_1^{\text{ss}}|^2 = I_{\text{sat}}\left(\frac{\gamma_0}{\kappa} - 1\right). \tag{9.20}$$

This has a solution only when $\gamma_0 > \kappa$: the *linear* system is already in the broken phase, but gain saturation pulls the effective gain back to the EP. The transition can become *discontinuous* (a first-order phase transition) rather than the continuous second-order transition of the linear EP, because the steady-state intensity jumps from zero to a finite value at threshold.

Experimental observations. Xia et al. demonstrated that local Kerr nonlinearity can tune the effective gain–loss balance and move the $\mathcal{PT}$ phase boundary, restoring or destroying

topological zero modes in a waveguide lattice [59]. Wimmer et al. observed optical solitons in $\mathcal{PT}$-symmetric fibre lattices with both globally and locally $\mathcal{PT}$-symmetric configurations, directly demonstrating that nonlinear self-trapping can stabilise modes that would be unstable in the linear broken phase [60]. In a complementary theoretical development, Alexeeva et al. analysed soliton families in $\mathcal{PT}$-symmetric nonlinear couplers and showed that gain saturation produces symmetric, antisymmetric, and asymmetric soliton branches, with stability boundaries that are qualitatively different from the linear EP [58].

Implications for the linear framework. The existence of nonlinear EPs means that the linear non-Hermitian theory developed in this book is always an approximation for systems with gain. The approximation is excellent below threshold ($I \ll I_{\text{sat}}$), where $\gamma_{\text{eff}} \approx \gamma_0$ and the linear EP correctly predicts the phase transition. Near and above threshold, the nonlinear self-consistency shifts the EP and can qualitatively change the phase diagram. A complete theory of non-Hermitian photonics must therefore include the nonlinear regime, as discussed in Chapter 19.

The interplay of nonlinearity and $\mathcal{PT}$ symmetry [3, 61, 62] connects to the open problems of Chapter 19: the Bogoliubov–de Gennes Hamiltonian that governs fluctuations around the nonlinear steady state is itself non-Hermitian, with a Petermann factor that diverges at the nonlinear EP. Whether this divergence can be exploited for sensing [47] or whether it is always cancelled by noise remains an active research question.

Chapter 10

Poles, Zeros, Lasers, and Absorbers

The scattering matrix of an optical system encodes every linear response property—reflection, transmission, absorption, and amplification—as a function of frequency. Its singularity structure in the complex frequency plane tells us when the system lases (a pole reaches the real axis), when it perfectly absorbs (a zero sits on the real axis), and when it does both simultaneously (the CPA-laser point). This pole–zero description is used for non-Hermitian photonic systems through concrete examples. It also establishes the connection between time reversal, $\mathcal{PT}$ symmetry, and the CPA-laser effect [63, 64], and clarifies the critical distinctions between exceptional points, CPA-laser thresholds, and maximum contrast conditions.

10.1 Scattering-matrix singularities revisited

Chapter 4 introduced the poles and zeros of $\mathbf{S}(\omega)$ as features of the complex frequency plane. We now develop this picture more systematically.

10.1.1 Poles: threshold lasing

A pole of $\mathbf{S}(\omega)$ at $\tilde{\omega} = \omega_0 - i\gamma$ corresponds to a quasinormal mode (Chapter 6). For a passive system, $\gamma > 0$ and the pole lies below the real axis. If gain is introduced, γ decreases; at threshold, $\gamma \to 0$ and the pole reaches the real axis:

$$\text{Im}\ \tilde{\omega}_n = 0 \qquad \Longleftrightarrow \qquad \text{lasing threshold.} \tag{10.1}$$

At this point, a self-sustained oscillation exists at real frequency ω_0 with no external driving—the system is a laser.

The physical picture is transparent in the pole–zero language: gain lifts the QNM poles toward the real axis. A passive cavity has all poles in the lower half-plane; adding gain is equivalent to reducing the imaginary parts of the poles. The first pole to reach the real axis sets the lasing threshold and frequency.

10.1.2 Zeros: coherent perfect absorption

A zero of $\det \mathbf{S}(\omega)$ at real ω means there exists a specific combination of incoming waves such that the outgoing field vanishes identically. All incident energy is absorbed. This is *coherent perfect absorption* (CPA) [55].

CPA requires precise control of the relative phase and amplitude of the incoming beams. For a two-port system, the CPA condition is

$$\det \mathbf{S}(\omega_{\mathrm{CPA}}) = 0 \quad \Longleftrightarrow \quad \text{an eigenvalue of } \mathbf{S} \text{ vanishes.} \tag{10.2}$$

10.1.3 The meromorphic structure

The scattering matrix $\mathbf{S}(\omega)$ of a finite photonic structure is a meromorphic function of ω: it is analytic everywhere in the complex plane except at isolated poles. Any meromorphic matrix can be written in terms of its poles and zeros as a product expansion. For a scalar reflection or transmission coefficient $s(\omega)$, this takes the form

$$s(\omega) = A \prod_j \frac{\omega - z_j}{\omega - p_j}, \tag{10.3}$$

where $\{p_j\}$ are the poles, $\{z_j\}$ are the zeros, and A is a constant prefactor. This representation makes the spectral response a direct map of the singularity positions:

- Near a pole p_j, the response $|s(\omega)|^2$ diverges—a sharp amplification peak.
- Near a zero z_j, the response vanishes—a sharp absorption dip.
- Between a nearby pole–zero pair, the response exhibits a Fano line shape: asymmetric with a rapid transition from absorption to amplification.

10.2 Example: Fabry–Pérot cavity

The Fabry–Pérot cavity is the simplest photonic structure that illustrates the full pole–zero framework.

10.2.1 Setup and transfer matrix

Consider a dielectric slab of thickness L, refractive index $n = n' + in''$ (with $n'' > 0$ for loss and $n'' < 0$ for gain), embedded in vacuum. The transfer matrix for the slab, including the two air–dielectric interfaces, is (Appendix C)

$$\mathbf{T} = \frac{1}{4n} \begin{pmatrix} (n+1)^2 e^{-i\phi} - (n-1)^2 e^{i\phi} & (n^2-1)(e^{i\phi} - e^{-i\phi}) \\ (n^2-1)(e^{-i\phi} - e^{i\phi}) & (n+1)^2 e^{i\phi} - (n-1)^2 e^{-i\phi} \end{pmatrix}, \tag{10.4}$$

where $\phi = n\omega L/c$ is the single-pass phase. The transmission coefficient is $t = 1/T_{22}$.

10.2.2 Pole positions

The poles of $t(\omega)$ are the zeros of T_{22}, which satisfy the round-trip resonance condition

$$r_{12}^2\, e^{-2in\omega L/c} = 1, \tag{10.5}$$

where $r_{12} = (n-1)/(n+1)$ is the Fresnel reflection coefficient at each interface. For weak material loss or gain, write $r_{12} \approx |r_{12}|\, e^{i\varphi_r}$ using the real background index n'. The pole frequencies are then approximately

$$\tilde{\omega}_m \approx \frac{c}{n'L}\big(m\pi - \varphi_r\big) - i\,\frac{c}{n'L}\,\ln\frac{1}{|r_{12}|}, \qquad m = 0, \pm1, \pm2, \ldots \tag{10.6}$$

For the lossless reference cavity, the imaginary part is negative: all poles lie below the real axis. The mirror-limited decay rate is $\gamma_m = (c/n'L)\ln(1/|r_{12}|)$; material absorption adds to this decay, while material gain subtracts from it.

Counting poles and zeros. For the Fabry–Pérot slab, the poles and zeros of the scalar transmission $t(\omega)$ are approximately equally spaced in the complex plane with a free spectral range $\Delta\omega = \pi c/(n'L)$ when the dispersion and absorption are weak. In the lossless reference cavity, zeros are the time reverses of the poles and appear at the conjugate positions. Material loss or gain moves the zeros as well as the poles; the zero motion is what makes coherent perfect absorption possible.

For a slab with $n' = 3.5$ and $L = 300$ nm at telecom wavelengths ($\lambda \sim 1550$ nm), the free spectral range is $\Delta\omega \approx 4.49 \times 10^{14}$ rad/s, and the Fresnel reflectivity $|r_{12}|^2 = [(3.5-1)/(3.5+1)]^2 \approx 0.309$ gives a quality factor $Q = \omega_0/(2\gamma_0) \approx 2.7$ for the lowest-order mode. The low Q makes the pole–zero structure well separated and easy to resolve, which is why this cavity is ideal for pedagogical demonstration of the scattering singularity framework.

10.2.3 Adding gain: pole motion toward lasing

If the slab material has gain ($n'' < 0$), the imaginary part of the propagation phase ϕ acquires a positive contribution that counteracts the mirror loss. The pole frequency becomes

$$\operatorname{Im}\tilde{\omega}_m = -\frac{c}{n'L}\big[\ln(1/|r_{12}|) + n''\omega_m L/c\big]. \tag{10.7}$$

When $|n''|\omega_m L/c = \ln(1/|r_{12}|)$, the imaginary part vanishes and the pole reaches the real axis: this is the lasing threshold for mode m. The threshold gain is

$$g_{\mathrm{th}} = \frac{2\ln(1/|r_{12}|)}{L}, \tag{10.8}$$

where $g_{\mathrm{th}} = 2|n''|\omega/c$ is the power gain coefficient. This is the standard textbook result, but here it emerges naturally as a statement about pole motion in the complex plane.

10.2.4 Zero positions and CPA

The zeros of $\det \mathbf{S}$ for the Fabry–Pérot slab satisfy the time-reversed round-trip condition:

$$r_{12}^{*2}\, e^{-2in^*\omega L/c} = 1. \tag{10.9}$$

For a purely lossy slab ($n'' > 0$), the zeros lie in the upper half-plane. As loss increases, zeros descend toward the real axis. CPA occurs when a zero reaches the real axis: the slab absorbs all coherently incident light at the CPA frequency.

The CPA threshold loss is the mirror image of the lasing threshold gain:

$$\alpha_{\mathrm{CPA}} = \frac{2\ln(1/|r_{12}|)}{L}, \tag{10.10}$$

confirming the time-reversal duality [24, 25]: CPA is lasing run backward.

Explicit pole–zero map for the Fabry–Pérot cavity. For the $n' = 3.5$, $L = 300$ nm slab, the first three poles are at $\tilde{\omega}_m/\Delta\omega = m - i\,0.187$ for $m = 1, 2, 3$, and the corresponding zeros at $\tilde{\omega}_m^*/\Delta\omega = m + i\,0.187$. As gain is added ($n'' \to -|n''|$), each pole moves vertically upward by $\delta(\mathrm{Im}\,\tilde{\omega}) = |n''|\omega_m/n'$. The lasing threshold for mode m is $|n''|_{\mathrm{th}} = \ln(1/|r_{12}|)\, c/(\omega_m L) = 0.187\,\Delta\omega/\omega_m$. Simultaneously, each zero moves downward by the same amount—the gain that brings a pole to the real axis pushes its partner zero away. Only in a $\mathcal{PT}$-symmetric configuration (gain and loss in separate halves of the cavity) do a pole and zero converge on the real axis simultaneously, producing the CPA-laser point.

10.3 Time reversal and the CPA-laser duality

CPA and lasing are connected by time reversal. A laser emits coherent radiation with no input; CPA absorbs coherent radiation with no output. Formally, if $\mathbf{S}(\omega_L)$ has a pole (lasing), then $\mathbf{S}^{-1}(\omega_L)$ has a zero—and $\mathbf{S}^{-1}$ is the scattering matrix of the time-reversed system (with gain and loss exchanged) [55].

10.3.1 Formal statement

Let $\mathbf{S}(\omega)$ be the scattering matrix of a linear system with material parameters $\varepsilon_{\mathrm{r}}(\mathbf{r})$. The time-reversed system has $\varepsilon_{\mathrm{TR}}(\mathbf{r}) = \varepsilon_{\mathrm{r}}^*(\mathbf{r})$ (gain $\leftrightarrow$ loss), and its scattering matrix is

$$\mathbf{S}_{\mathrm{TR}}(\omega) = [\mathbf{S}(\omega^*)]^{*\,-1}. \tag{10.11}$$

At real frequencies, $\mathbf{S}_{\mathrm{TR}}(\omega) = [\mathbf{S}(\omega)]^{-1\,*}$. The poles of $\mathbf{S}$ become zeros of $\mathbf{S}_{\mathrm{TR}}$, and vice versa.

10.3.2 CPA-laser point

For a system with both gain and loss, lasing and CPA can occur at the *same* frequency if and only if a pole and a zero of $\mathbf{S}$ coincide on the real axis. This simultaneous occurrence is the *CPA-laser* point.

A $\mathcal{PT}$-symmetric structure provides a natural setting for the CPA-laser because the $\mathcal{PT}$ constraint relates the inverse scattering matrix to the parity-transformed, time-reversed scattering matrix. For real frequencies this is the two-port relation discussed in Chapter 9; analytically continued into the complex plane, it pairs a pole at $\tilde{\omega}$ with a zero at $\tilde{\omega}^*$. When $\tilde{\omega}$ reaches the real axis, the pole and zero meet simultaneously—the CPA-laser condition is automatically satisfied.

10.4 $\mathcal{PT}$-symmetric scattering

For a $\mathcal{PT}$-symmetric structure, the scattering matrix satisfies the fundamental relation (9.12). At real frequencies, the consequences are:

1. $|\det \mathbf{S}| = 1$: the product of scattering eigenvalues has unit modulus.
2. Poles and zeros of $\mathbf{S}$ come in conjugate pairs: if $\tilde{\omega}$ is a pole, then $\tilde{\omega}^*$ is a zero, and vice versa.
3. In the $\mathcal{PT}$-symmetric phase, both scattering eigenvalues are unimodular: neither net amplification nor net absorption occurs.
4. In the $\mathcal{PT}$-broken phase, one eigenvalue has $|s| > 1$ (amplification) and the other has $|s| < 1$ (absorption).

The CPA-laser point, where a pole and zero meet on the real axis, corresponds to the special situation where $\mathcal{PT}$ symmetry enforces $s_1 \to \infty$ and $s_2 \to 0$ simultaneously, keeping $|s_1 s_2| = 1$.

10.4.1 Scattering phase diagram

As the gain–loss parameter γ increases from zero, the scattering eigenvalues trace paths in the complex plane. Three qualitatively different regimes emerge:

1. **$\mathcal{PT}$-symmetric phase ($\gamma < \gamma_{\text{EP}}$):** both $|s_1| = |s_2| = 1$. The system is transparent on average, though phase shifts differ from the passive case.
2. **$\mathcal{PT}$-broken phase ($\gamma_{\text{EP}} < \gamma < \gamma_{\text{CPA}}$):** $|s_1| > 1$, $|s_2| < 1$. One channel amplifies, the other absorbs, but neither diverges.
3. **CPA-laser threshold ($\gamma = \gamma_{\text{CPA}}$):** $|s_1| \to \infty$, $|s_2| \to 0$. The system simultaneously lases and perfectly absorbs at the same frequency for different input states.

This phase diagram highlights that the EP (a spectral degeneracy) and the CPA-laser (a pole–zero coincidence) generally occur at *different* gain–loss values.

10.5 Transfer-matrix analysis of a $\mathcal{PT}$ multilayer

10.5.1 Setup

Consider a multilayer stack of N unit cells, each composed of a loss layer (thickness d_1, refractive index $n_1 = n_0 + i\kappa_I$) and a gain layer (thickness $d_2 = d_1$, refractive index $n_2 = n_0 - i\kappa_I$). This structure satisfies the $\mathcal{PT}$ condition $n(x) = n^*(-x)$ when the inversion center is placed at the interface between the loss and gain layers.

10.5.2 Output coefficients

The total transfer matrix $\mathbf{T}^N$ determines the reflection and transmission through (9.9). As a function of the gain–loss parameter κ_I, three regimes emerge [63]:

1. **Below the EP** ($\kappa_I < \kappa_{\mathrm{EP}}$): the transmission and reflection show moderate asymmetry. The output intensity can exceed the input, but the contrast between amplification and absorption is limited.
2. **At the EP** ($\kappa_I = \kappa_{\mathrm{EP}}$): the transfer-matrix eigenvalues coalesce. The output exhibits maximum contrast between amplification and absorption for specific input configurations.
3. **Above the EP** ($\kappa_I > \kappa_{\mathrm{EP}}$): the system enters the $\mathcal{PT}$-broken phase. One scattering channel grows exponentially with N (laser-like amplification), while the other is exponentially suppressed (CPA-like absorption).

10.5.3 CPA-laser versus exceptional point

A common misconception is that the CPA-laser point coincides with the exceptional point. In general, it does not [63]. The EP is a property of the transfer matrix eigenvalues (a spectral degeneracy); the CPA-laser is a property of the scattering matrix singularities (a pole–zero coincidence on the real axis). The two conditions are related but distinct:

- The EP can occur at a gain–loss value where no real-frequency pole or zero exists.
- The CPA-laser point can occur in the broken phase, away from the EP.
- Maximum output contrast (the ratio of amplification to absorption for different input phases) need not coincide with either the EP or the CPA-laser point.

This distinction is physically important: the EP controls the spectral topology and the eigenvector structure, while the CPA-laser controls the input–output energy balance.

Three distinct critical points

In a $\mathcal{PT}$-symmetric scattering system, the exceptional point (spectral degeneracy), the CPA-laser point (pole–zero coincidence), and the point of maximum amplification–absorption contrast are in general three different parameter values. Conflating them leads to incorrect predictions of device performance.

10.6 Spectral singularities

10.6.1 Definition

A *spectral singularity* is a point on the real frequency axis [65, 55] where the reflection or transmission coefficient diverges—i.e., a pole of $\mathbf{S}(\omega)$ at real ω. Unlike the QNM poles discussed above, which lie in the lower half-plane for passive systems, a spectral singularity occurs on the real axis and is the scattering-theory description of threshold lasing in a linear gain medium.

Mathematically, a spectral singularity at ω_s means the resolvent of the Maxwell operator is unbounded at ω_s, but ω_s is not an eigenvalue in the standard sense: no square-integrable

eigenfunction exists. Instead, a spectral singularity is a divergence of the scattering amplitude at a real frequency—an "eigenvalue embedded in the continuum" of the non-self-adjoint operator.

10.6.2 Physical interpretation

For optical systems, spectral singularities have two complementary physical manifestations that are connected by time reversal.

The first is *lasing*. A pole of $\mathbf{S}(\omega)$ on the real axis means that the system sustains a self-oscillating field at frequency ω_s without any external drive. In the meromorphic representation

$$\mathbf{S}(\omega) = \mathbf{S}_{\mathrm{bg}}(\omega) \prod_n \frac{\omega - z_n}{\omega - p_n}, \tag{10.12}$$

where p_n are the QNM poles and z_n the scattering zeros, lasing occurs when one pole p_m migrates to the real axis: $\mathrm{Im}\, p_m \to 0^-$. At this point the denominator $(\omega - p_m)$ vanishes for real $\omega = \mathrm{Re}\, p_m$, and the scattering amplitudes diverge. The outgoing field grows without bound in time—the signature of stimulated emission overcoming all losses.

The second manifestation is *coherent perfect absorption* (CPA), also called anti-lasing. A zero of $\det \mathbf{S}(\omega)$ on the real axis means that there exists a coherent input field configuration that is perfectly absorbed: no light exits the structure. Mathematically, this is a spectral singularity of the *inverse* scattering matrix $\mathbf{S}^{-1}$. By (10.12), CPA occurs when a zero z_m sits on the real axis: $\mathrm{Im}\, z_m = 0$.

The connection between lasing and CPA follows from the time-reversal structure of the S-matrix. A pole of a given system maps to a zero of the time-reversed system, in which gain and loss are interchanged. In a lossless reciprocal reference problem this reduces to the familiar conjugate pole–zero symmetry, but it is not a general consequence of reciprocity alone. In specially constrained active systems, most notably $\mathcal{PT}$-symmetric ones, a pole and a zero can be forced to reach the real axis together. This is the *CPA-laser* point [55, 24], at which the system can act as either a laser or a perfect absorber depending on the input coherence.

Quantitative threshold from S-matrix poles. Combining the meromorphic representation (10.12) with the Fabry–Pérot pole positions from Section 10.2.2, the lasing threshold for mode m is reached when the gain g satisfies

$$g = g_{\mathrm{th},m} = -\frac{2\ \mathrm{Im}\ p_m(g=0)}{c/n_g} = \frac{2\ln(1/|r_{12}|)}{L} + \alpha_{\mathrm{mat}}, \tag{10.13}$$

where n_g is the group index, α_{mat} accounts for material absorption, and $|r_{12}|$ is the mirror reflectivity. The first term is the mirror loss and the second the material loss; the threshold condition requires the gain to compensate both. This derivation makes explicit the pole-motion picture: gain "lifts" the pole at a rate $c\,g/(2n_g)$ until it reaches the real axis.

10.6.3 Threshold condition as a spectral singularity

The spectral-singularity viewpoint provides a precise mathematical criterion for the lasing threshold that goes beyond the simple $\mathrm{Im}\,\tilde{\omega} = 0$ condition. For a one-dimensional potential $V(x) = -\omega^2[\varepsilon(x) - 1]/c^2$ with complex ε, the spectral singularity occurs when the Jost function

$f^+(\omega)$ vanishes at a real ω while $f^-(\omega) \neq 0$. This is a stronger condition than a pole of the reflection coefficient: it requires that the outgoing-wave solution be bounded (not square-integrable) while the incoming-wave solution diverges.

For the Fabry–Pérot cavity, the Jost-function condition reduces to the round-trip condition (10.5) evaluated on the real axis. But for more complex structures—photonic crystals, coupled resonators, random media—the Jost-function framework provides a rigorous definition of the lasing threshold that does not require identifying individual QNM poles.

10.6.4 Spectral singularities in $\mathcal{PT}$-symmetric potentials

In a $\mathcal{PT}$-symmetric optical potential, spectral singularities have special properties. For reciprocal one-dimensional scattering, the generalized conservation relation can be written as

$$\big||t(\omega)|^2 - 1\big| = \sqrt{R_L(\omega) R_R(\omega)}, \qquad R_{L,R} = |r_{L,R}|^2. \tag{10.14}$$

A spectral singularity at ω_s corresponds to $|t(\omega_s)| \to \infty$, and the reflection product diverges simultaneously—the system emits coherent radiation into the available channels. Its time-reversed partner is the CPA condition.

10.7 Beyond $\mathcal{PT}$: general pole–zero engineering

The pole–zero framework extends beyond $\mathcal{PT}$-symmetric systems. For any linear photonic structure described by a scattering matrix, the motion of poles and zeros in the complex frequency plane under parameter changes follows simple rules that enable systematic device design. Gain moves poles upward in the complex plane, toward the real axis and eventually through it to lasing. Loss moves poles downward, increasing decay rates and broadening resonances. Structural changes—geometry, coupling strength, grating period—move poles horizontally, shifting the resonance frequency while leaving the decay rate largely unchanged. The zero structure is constrained by causality, time-reversal relations, and passivity bounds on real-frequency scattering, but zeros are not fixed by reciprocity alone. Indeed, in passive absorbers one deliberately moves a scattering zero onto the real axis by controlling the spatial profile and strength of absorption. These rules provide an intuitive design language: achieving CPA at a target frequency amounts to placing a zero on the real axis, while suppressing a parasitic lasing mode amounts to pushing its pole deeper into the lower half-plane.

10.7.1 Design as singularity placement

Designing a photonic device—a laser, a sensor, a filter, or an absorber—can be viewed as engineering the positions of poles and zeros in the complex frequency plane to achieve the desired spectral response. This perspective unifies many seemingly different problems:

- A narrow-band laser requires one pole close to the real axis with all other poles far away.
- A broadband absorber requires zeros spread along the real axis.
- A Fano resonance requires a nearby pole–zero pair with controlled spacing.
- A flat-top filter requires poles and zeros arranged in a Butterworth-like pattern.

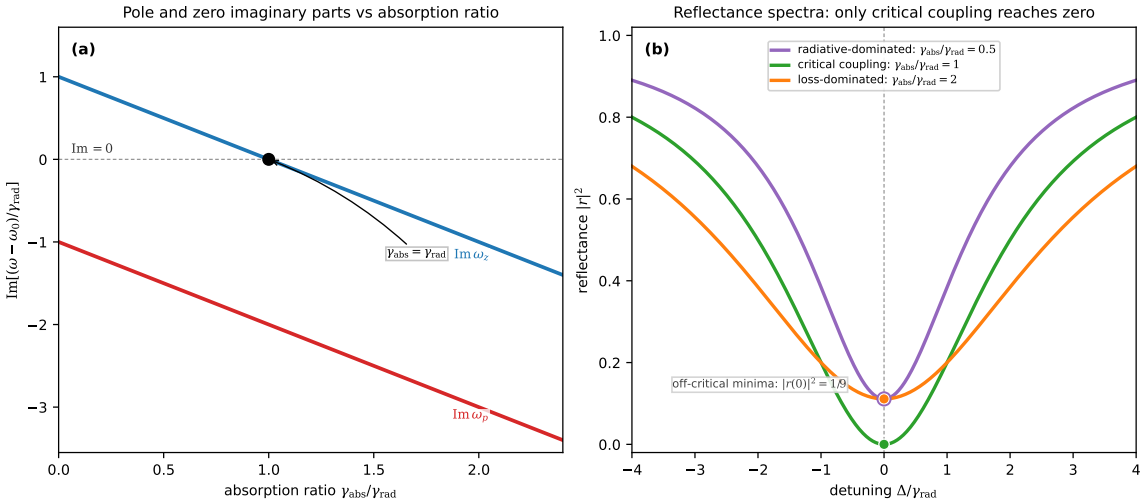


Figure 10.1 Critical coupling as zero placement in a one-port absorber. The time convention is $e^{-i\omega t}$ and the reflection amplitude is Eq. (10.15). (a) Imaginary parts of the pole and zero in Eq. (10.16) as the ratio $\gamma_{\mathrm{abs}}/\gamma_{\mathrm{rad}}$ is varied. The pole remains in the lower half-plane, while the passive zero begins in the upper half-plane and reaches $\mathrm{Im}\,\omega_z = 0$ at $\gamma_{\mathrm{abs}} = \gamma_{\mathrm{rad}}$. (b) Real-frequency reflectance spectra for under-coupled ($\gamma_{\mathrm{abs}}/\gamma_{\mathrm{rad}} = 0.5$), critically coupled, and over-coupled ($\gamma_{\mathrm{abs}}/\gamma_{\mathrm{rad}} = 2$) cases. The two off-critical spectra have the same minimum reflectance, $|r(0)|^2 = 1/9$, while critical coupling gives $|r(0)|^2 = 0$.

Metasurface example: pole–zero placement for perfect absorption. Consider a metasurface consisting of a periodic array of lossy dielectric resonators on a metallic back-reflector. The back-reflector eliminates transmission, so the scattering problem reduces to a single-port reflection coefficient $r(\omega)$. Perfect absorption at a target frequency ω_0 requires $r(\omega_0) = 0$, i.e., a zero of r on the real axis. In the pole–zero language, this is achieved by tuning the resonator geometry so that a pole at $\tilde{\omega} = \omega_0 - i\gamma$ is paired with a zero at $z = \omega_0 + i\gamma'$, and then adjusting the material loss so that $\gamma' \to 0$. The critical-coupling condition $\gamma_{\mathrm{rad}} = \gamma_{\mathrm{abs}}$ (radiative decay equals absorptive decay) is precisely the statement that the zero reaches the real axis.

With the global $e^{-i\omega t}$ convention used in this book, a one-port temporal coupled-mode model may be written in the dimensionless form

$$r(\Delta) = \frac{-i\Delta + \gamma_{\mathrm{abs}} - \gamma_{\mathrm{rad}}}{-i\Delta + \gamma_{\mathrm{abs}} + \gamma_{\mathrm{rad}}}, \qquad \Delta = \omega - \omega_0. \tag{10.15}$$

The pole and zero are therefore

$$\omega_p = \omega_0 - i(\gamma_{\mathrm{abs}} + \gamma_{\mathrm{rad}}), \qquad \omega_z = \omega_0 + i(\gamma_{\mathrm{rad}} - \gamma_{\mathrm{abs}}). \tag{10.16}$$

The passive zero starts in the upper half-plane when absorption is weak and crosses the real axis at critical coupling.

This design recipe generalises to multi-frequency perfect absorption: each target frequency requires one zero on the real axis, which in turn requires one resonator mode with matched radiative and absorptive decay rates. The pole–zero framework provides a systematic way to count the degrees of freedom and identify the minimum number of resonator modes needed for a given absorption bandwidth.

There is an important physical distinction between critical coupling and lasing threshold. Critical coupling is still a passive condition: the pole remains in the lower half-plane because $\operatorname{Im}\omega_p = -(\gamma_{\mathrm{abs}} + \gamma_{\mathrm{rad}}) < 0$. The device absorbs perfectly at one real frequency not because the stored field grows without bound, but because the directly reflected wave and the field re-radiated by the resonator cancel exactly in the outgoing channel. In Maxwell language, the Poynting flux entering the resonator through the external channel equals the Ohmic or material-loss power dissipated inside the resonator. This is why the condition is often called *impedance matching*: the external radiation resistance and internal absorption resistance are matched.

This clarification prevents a common misconception. Moving a zero to the real axis is not the same as moving a pole to the real axis. The first produces a finite, stable absorbing solution for a particular coherent input; the second produces a self-sustained outgoing solution with no input. The two events can coincide only in special active systems such as the $\mathcal{PT}$-symmetric CPA-laser discussed above. In a one-port passive metasurface, only the zero reaches the real axis, so perfect absorption is compatible with stability, causality, and finite stored energy.

10.7.2 Passivity constraints

For a passive system (no gain), the poles of $\mathbf{S}(\omega)$ must lie in the lower half-plane. This is a direct consequence of causality and stability: the Green function of a passive system decays in the future, not the past. Scattering zeros obey different constraints. In a lossless unitary system they mirror the poles across the real axis; in a lossy passive system they can move onto the real axis, producing CPA, or away from the simple conjugate locations. Violating pole stability requires gain, which is the fundamental reason why lasers and CPA-lasers are active devices.

10.7.3 Connection to QNM perturbation theory

The biorthogonal perturbation theory of Chapter 6 provides the sensitivity of pole positions to material or geometric changes:

$$\delta\tilde{\omega}_n = -\frac{\tilde{\omega}_n}{2}\frac{\int \delta\varepsilon_{\mathrm{r}}\,\tilde{\mathbf{E}}_n \cdot \tilde{\mathbf{E}}_n \;\mathrm{d}^3 r}{\mathcal{N}_n}. \tag{10.17}$$

Here $\mathcal{N}_n$ is the unconjugated QNM normalization, including the appropriate dispersive and surface or PML contribution. This formula directly connects the pole–zero engineering picture to the underlying Maxwell fields: placing gain where the relevant QNM has large overlap moves the pole most efficiently toward the real axis.

10.8 Connections

This closes the scattering part of the book. Exceptional points, $\mathcal{PT}$-symmetric optics, and pole–zero motion describe complementary aspects of non-Hermitian photonic systems driven by gain, loss, and coupling.

Part IV now turns to the topological structure of non-Hermitian photonic bands. Chapter 11 builds Berry curvature directly from Maxwell fields; Chapter 12 introduces complex bands and non-Hermitian gaps; and the subsequent chapters develop the photonic skin effect, point-gap topology, and the reconstruction of bulk-boundary correspondence.

Part IV

Topology Before and After Hermiticity

Chapter 11

Berry Curvature from Maxwell Fields

Topology entered condensed-matter physics through the Berry phase of electronic Bloch states. The same mathematics applies to photonic Bloch modes—but with an important difference: the inner product is defined by the electromagnetic energy, not by the probability density. This chapter derives Berry connection, Berry curvature, and Chern numbers directly from the Maxwell eigenproblem, without passing through a tight-binding analogy. The result is a Maxwell-first topological photonics that connects the band structure of Chapter 3 to the existence of topologically protected edge states.

11.1 Why the inner product matters

Before computing the Berry phase, we must settle a question that has no analogue in electronic topological insulators: *which inner product defines the Berry connection?*

In quantum mechanics, the choice is unambiguous: the standard L^2 inner product $\langle\psi|\phi\rangle = \int \psi^* \phi \ \mathrm{d}^3 r$. In photonics, the natural inner product comes from the electromagnetic energy density. For nondispersive Hermitian media, the **E**-field formulation uses the weighted product

$$\langle \mathbf{u}|\mathbf{v}\rangle_\varepsilon = \int_{\Omega_{\mathrm{uc}}} \varepsilon_{\mathrm{r}}(\mathbf{r})\, \mathbf{u}^*(\mathbf{r}) \cdot \mathbf{v}(\mathbf{r}) \ \mathrm{d}^d r, \tag{11.1}$$

while for the **H**-field formulation with $\mu_{\mathrm{r}} = 1$ it is the unweighted form

$$\langle \mathbf{u}|\mathbf{v}\rangle = \int_{\Omega_{\mathrm{uc}}} \mathbf{u}^*(\mathbf{r}) \cdot \mathbf{v}(\mathbf{r}) \ \mathrm{d}^d r. \tag{11.2}$$

The **H**-field formulation leads to the Hermitian eigenoperator $\hat{\Theta}_H = \nabla \times \left(\varepsilon_{\mathrm{r}}^{-1} \nabla \times \cdot\right)$ whose eigenstates are automatically orthogonal in (11.2). The **E**-field formulation leads to a generalized eigenproblem $\nabla \times (\mu_{\mathrm{r}}^{-1} \nabla \times \mathbf{E}) = (\omega^2/c^2)\varepsilon_{\mathrm{r}} \mathbf{E}$ whose eigenstates are orthogonal in the ε_{r}-weighted product (11.1).

Both formulations give the same Berry curvature and Chern numbers when ε_{r} is Hermitian (real for scalar, Hermitian for tensor), because the eigenstates are related by a gauge transformation. However, the ε_{r}-weighted inner product is the physically correct one for perturbation theory,

for normalization of mode amplitudes, and for the construction of Wannier functions [12]. In dispersive but lossless media, ε_{r} is replaced by the Brillouin energy weight discussed in Chapter 5.

Remark 11.1. In non-Hermitian systems (complex ε_{r}), the two formulations lead to different Berry connections because the ε_{r}-weighted inner product is no longer positive definite. The resolution requires the biorthogonal Berry connection of Section 11.8 and Chapter 12.

11.2 Geometric phase for photonic Bloch states

Recall from Chapter 3 that the Bloch eigenmodes of a periodic photonic crystal satisfy the unit-cell eigenproblem $\hat{\Theta}_{\mathbf{k}}\,\mathbf{u}_{n\mathbf{k}} = (\omega_n(\mathbf{k})^2/c^2)\,\mathbf{u}_{n\mathbf{k}}$, with $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ periodic on Ω_{uc}. As $\mathbf{k}$ varies over the Brillouin zone, $\mathbf{u}_{n\mathbf{k}}$ traces a path in the Hilbert space of periodic vector fields. The geometric phase accumulated along this path is the photonic Berry phase.

11.2.1 Berry connection

The *Berry connection* for band n is the one-form on the Brillouin zone defined by [12]

$$\mathcal{A}_{n,\alpha}(\mathbf{k}) = i\,\frac{\langle \mathbf{u}_{n\mathbf{k}} | \partial_{k_\alpha} | \mathbf{u}_{n\mathbf{k}} \rangle_\varepsilon}{\langle \mathbf{u}_{n\mathbf{k}} | \mathbf{u}_{n\mathbf{k}} \rangle_\varepsilon}, \tag{11.3}$$

where $\alpha = x, y$ labels the components of $\mathbf{k}$ and the inner product is the ε_{r}-weighted form (11.1). When the modes are normalized to $\langle \mathbf{u}_{n\mathbf{k}} | \mathbf{u}_{n\mathbf{k}} \rangle_\varepsilon = 1$, the denominator drops out.

The Berry connection is *gauge-dependent*: under a gauge transformation $|\mathbf{u}_{n\mathbf{k}}\rangle \to e^{i\phi(\mathbf{k})}|\mathbf{u}_{n\mathbf{k}}\rangle$, it shifts by $\mathcal{A}_{n,\alpha} \to \mathcal{A}_{n,\alpha} - \partial_{k_\alpha}\phi$. Physical observables must therefore depend on gauge-invariant combinations.

11.2.2 Berry curvature

The *Berry curvature* is the gauge-invariant curl of the Berry connection. In two dimensions ($\mathbf{k} = (k_x, k_y)$), it is a pseudoscalar:

$$\mathcal{F}_n(\mathbf{k}) = \frac{\partial \mathcal{A}_{n,y}}{\partial k_x} - \frac{\partial \mathcal{A}_{n,x}}{\partial k_y}. \tag{11.4}$$

An equivalent expression that does not require differentiating the Berry connection directly is the Kubo-like formula:

$$\mathcal{F}_n(\mathbf{k}) = -2 \sum_{m \neq n} \frac{\mathrm{Im}\big[\langle \mathbf{u}_{n\mathbf{k}} | \partial_{k_x} \hat{\Theta}_{\mathbf{k}} | \mathbf{u}_{m\mathbf{k}} \rangle_\varepsilon \, \langle \mathbf{u}_{m\mathbf{k}} | \partial_{k_y} \hat{\Theta}_{\mathbf{k}} | \mathbf{u}_{n\mathbf{k}} \rangle_\varepsilon\big]}{[\omega_n^2(\mathbf{k}) - \omega_m^2(\mathbf{k})]^2}. \tag{11.5}$$

This formula shows that the Berry curvature is large when two bands are nearly degenerate ($\omega_n \approx \omega_m$)—the curvature is concentrated near avoided crossings and symmetry points.

11.2.3 Physical interpretation

The Berry curvature has a direct physical interpretation: it is the "magnetic field" in momentum space that deflects wave packets transversely. For a photonic wave packet centered at $\mathbf{k}$ in band

n, the semiclassical equations of motion include an anomalous velocity

$$\dot{\mathbf{r}}_\perp = -\dot{\mathbf{k}} \times \hat{z}\, \mathcal{F}_n(\mathbf{k}), \tag{11.6}$$

which deflects the wave packet perpendicular to both the applied force $\dot{\mathbf{k}}$ and the direction of propagation. This is the photonic analogue of the anomalous Hall effect in electronic systems [66, 16, 67, 68, 69, 70, 71, 72, 73, 74]. When the Berry curvature has a definite sign across the Brillouin zone, the deflection is systematic and produces a net transverse photon current—the photonic Hall effect.

Example: Berry curvature near a Dirac point. Consider a triangular-lattice photonic crystal of dielectric rods ($\varepsilon = 12$, radius $r = 0.2a$) in air. The TM band structure has a Dirac point at the K corner of the Brillouin zone between the second and third bands. Near this degeneracy, the two-band effective Hamiltonian takes the massive Dirac form

$$H_{\text{eff}}(\delta\mathbf{k}) = v_D(\delta k_x\, \sigma_x + \delta k_y\, \sigma_y) + \tfrac{1}{2} m\, \sigma_z, \tag{11.7}$$

where v_D is the group velocity at the Dirac point, m is the gap opened by a perturbation (e.g. by breaking C_3 inversion symmetry or applying a magneto-optical field), and $\delta\mathbf{k} = \mathbf{k} - \mathbf{k}_K$. The Berry curvature of the lower band is

$$\mathcal{F}(\delta\mathbf{k}) = -\frac{m\, v_D^2}{4(v_D^2|\delta\mathbf{k}|^2 + m^2/4)^{3/2}}. \tag{11.8}$$

This is sharply peaked at $\delta\mathbf{k} = 0$: the Berry curvature is concentrated within a momentum radius $\sim |m|/(2v_D)$ of the Dirac point. The integral over a single valley gives a half-integer valley contribution whose sign is set by both the Dirac mass and the chirality of that valley. For a Haldane-like magneto-optical perturbation, the mass and valley chirality combine so that the two valley contributions add rather than cancel, giving a nonzero total Chern number. Reversing the bias field reverses the sign of the total Chern number.

This example illustrates a key principle: the Chern number is determined by the *local* Berry curvature near band degeneracies. The rest of the Brillouin zone contributes negligibly. This is why photonic topological transitions are controlled by gap closings and reopenings at high-symmetry points.

11.3 Chern number

The integral of the Berry curvature over the entire Brillouin zone is quantized:

$$\boxed{\mathcal{C}_n = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_n(\mathbf{k})\ \mathrm{d}^2k\ \in\ \mathbb{Z}.} \tag{11.9}$$

This integer is the *Chern number* of the n-th band. It is a topological invariant: it cannot change under smooth deformations of the band structure that do not close the gap to an adjacent band [12].

11.3.1 Quantization proof sketch

The quantization of the Chern number follows from a general result in differential geometry: the integral of a curvature two-form over a closed manifold (here, the Brillouin zone torus T^2) is 2π times an integer. Physically, the argument proceeds as follows:

1. Divide the BZ into patches on each of which a smooth gauge for $|\mathbf{u}_{n\mathbf{k}}\rangle$ can be chosen.
2. On the overlap between two patches, the gauge transformation $e^{i\phi(\mathbf{k})}$ is single-valued and smooth.
3. The Berry curvature integral equals the total winding of the gauge transformation around the patch boundaries, which is an integer by the single-valuedness of $e^{i\phi}$.

11.3.2 Time-reversal symmetry

When the photonic crystal preserves time-reversal symmetry ($\varepsilon_{\mathrm{r}}(\mathbf{r})$ real and scalar), the Berry curvature satisfies $\mathcal{F}_n(\mathbf{k}) = -\mathcal{F}_n(-\mathbf{k})$, and the Chern number vanishes: $\mathcal{C}_n = 0$. Nonzero Chern numbers require breaking time-reversal symmetry, for example with a magneto-optical material [75, 76, 77, 78, 79, 80, 81] (ε_{r} tensor with off-diagonal imaginary components).

The proof is straightforward: time reversal maps $\mathbf{k} \to -\mathbf{k}$ and complex-conjugates the eigenvectors. The Berry curvature, being the imaginary part of an overlap integral, picks up a sign from the conjugation and another sign from the reversal of $\mathbf{k}$, giving $\mathcal{F}_n(-\mathbf{k}) = -\mathcal{F}_n(\mathbf{k})$.

11.3.3 Valley Chern number

Even when $\mathcal{C}_n = 0$ globally, the Berry curvature may be concentrated near specific high-symmetry points (valleys) in the Brillouin zone. The *valley Chern number*

$$\mathcal{C}_n^{(v)} = \frac{1}{2\pi} \int_{\text{valley}} \mathcal{F}_n(\mathbf{k}) \; \mathrm{d}^2 k \tag{11.10}$$

can be half-integer and is approximately quantized when the valley is well separated from other regions of nonzero curvature. Valley Chern numbers underlie the photonic valley Hall effect, in which light propagates along domain walls between regions with opposite valley orientations.

11.4 Photonic Chern insulators

11.4.1 Magneto-optical photonic crystals

The canonical platform for nonzero photonic Chern numbers [8, 82, 20, 83] is a two-dimensional photonic crystal made from magneto-optical material. Under a static magnetic field $\mathbf{B}_0 = B_0 \hat{z}$, a ferrite or yttrium iron garnet (YIG) develops an antisymmetric permittivity tensor:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & i\varepsilon_g & 0 \\ -i\varepsilon_g & \varepsilon_{xx} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}, \tag{11.11}$$

where $\varepsilon_g \propto B_0$ is the gyromagnetic coupling. This tensor is Hermitian ($\boldsymbol{\varepsilon}^\dagger = \boldsymbol{\varepsilon}$) but not symmetric ($\boldsymbol{\varepsilon}^\mathsf{T} \neq \boldsymbol{\varepsilon}$): the system is lossless but breaks time-reversal symmetry.

The breaking of time-reversal symmetry lifts the $\mathcal{F}_n(\mathbf{k}) = -\mathcal{F}_n(-\mathbf{k})$ constraint, allowing nonzero Chern numbers. For a triangular-lattice photonic crystal of YIG rods in air, with parameters $\varepsilon_{xx} = 15$, $\varepsilon_g = 12.4$ (at microwave frequencies near 4.28 GHz), numerical calculations give $\mathcal{C} = \pm 1$ for the lowest-frequency TM gap.

Worked Chern-number calculation. To make the calculation concrete, consider a square lattice of YIG cylinders (radius $r = 0.11a$) in air with the tensor permittivity (11.11) and parameters $\varepsilon_{xx} = 15$, $\varepsilon_g = 12.4$. The TM modes (E_z polarisation) satisfy a scalar eigenproblem whose $\mathbf{k}$-dependent operator is $\hat{\Theta}_\mathbf{k} = -(\nabla + i\mathbf{k}) \times (\boldsymbol{\varepsilon}^{-1}(\nabla + i\mathbf{k})\times)$. The inverse permittivity tensor couples E_z to the in-plane magnetic field components, breaking the degeneracy between clockwise and counter-clockwise circulating modes at the Γ and M points.

Numerical diagonalisation on a 201×201 $\mathbf{k}$-grid using the FHS algorithm (11.14) yields $\mathcal{C}_1 = +1$ for the lowest band (below the first gap) and $\mathcal{C}_2 = -2$ for the second band, giving a gap Chern number $\mathcal{C}_\text{gap} = +1$ for the first gap. The Berry curvature is concentrated at the M point, where the gap is smallest ($\Delta\omega/\omega_0 \approx 0.08$), confirming that the topological invariant is controlled by the near-degenerate region. When the magnetic field is reversed ($\varepsilon_g \to -\varepsilon_g$), all Chern numbers change sign: $\mathcal{C}_\text{gap} = -1$, so the edge state propagation direction reverses.

11.4.2 Photonic anomalous Hall effect

A photonic crystal with $\mathcal{C} \neq 0$ is a *photonic Chern insulator*. The bulk-boundary correspondence (Section 11.7) guarantees one-way edge states at every interface with a topologically distinct region (including vacuum, which has $\mathcal{C} = 0$).

These chiral edge modes propagate in one direction only and are immune to backscattering from fabrication defects, sharp bends, or surface roughness—as long as the bulk gap remains open. This is the photonic anomalous Hall effect: electromagnetic energy flows in one direction along the edge, with the direction determined by the sign of the Chern number and hence by the direction of the applied magnetic field.

Experimental observation. The photonic anomalous Hall effect was first demonstrated experimentally by Wang et al. (2009) using a two-dimensional photonic crystal of YIG rods at microwave frequencies. The experiment confirmed three predictions of the topological theory:

1. One-way edge propagation with suppressed backscattering even around sharp 90° bends and metallic obstacles.
2. The number of edge modes matches the predicted Chern number difference across the interface.
3. Reversal of the external magnetic field reverses the propagation direction, confirming the topological origin.

The key quantitative figure of merit is the ratio of forward to backward transmission: in the experiment, this exceeded 50 dB across the topological gap, comparable to commercial microwave isolators but achieved without any Faraday-rotation element.

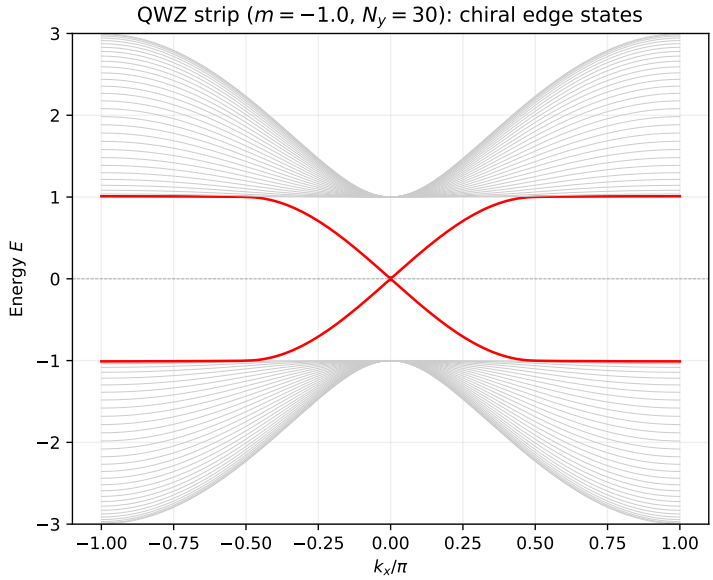


Figure 11.1 Computed band structure of a QWZ strip ($m = -1$, $N_y = 30$). The 60 bulk bands (grey) form a gapped spectrum with Chern number $\mathcal{C} = 1$. The red branches are edge-localized modes on the two opposite strip boundaries; each boundary supports one chiral channel, with opposite group velocities on opposite edges. The branches connect the lower and upper bulk bands across the gap, illustrating the bulk–boundary correspondence.

11.4.3 Limitations

The magneto-optical approach has practical limitations:

- The gyromagnetic effect is weak at optical frequencies, so large Chern gaps require strong magnetic fields or resonant enhancement.
- The materials are typically lossy, introducing non-Hermiticity that complicates the topological classification.
- The platform is most practical at microwave and millimeter-wave frequencies, where ferrites have large gyromagnetic responses.

Alternative approaches—Floquet modulation, coupled-ring lattices, and synthetic gauge fields—can produce effective Chern insulators at optical frequencies without magnetic materials.

Floquet photonic Chern insulators. An alternative route to nonzero Chern numbers at optical frequencies uses time-periodic modulation of the refractive index [18, 9]. The central idea is that a periodically driven system has a *Floquet Hamiltonian* H_F whose eigenvalues are quasi-energies ε_n defined modulo the driving frequency Ω. Even if the instantaneous Hamiltonian preserves time-reversal symmetry at every instant, the stroboscopic evolution operator $U(T) = \mathcal{T}\exp(-i\int_0^T H(t)\,\mathrm{d}t)$ can break effective time-reversal symmetry, opening topological gaps with nonzero Chern numbers.

The simplest photonic implementation uses a two-dimensional lattice of coupled waveguides or ring resonators with periodically modulated link phases. The paraxial propagation equation along the waveguide axis z plays the role of the Schrödinger equation:

$$i\partial_z\psi(x,y,z) = H(z)\,\psi(x,y,z), \tag{11.12}$$

where z replaces time and the lattice potential $H(z)$ is periodic with period Z (the helix pitch or modulation period). The Floquet operator $U(Z) = \mathcal{T}\exp(-i\int_0^Z H(z')\,\mathrm{d}z')$ defines the effective Hamiltonian $H_F = (i/Z)\ln U(Z)$, and its Bloch bands carry well-defined Chern numbers computed from the Berry curvature of the Floquet eigenstates.

In coupled-ring-resonator lattices [20], the link phases between rings are engineered to simulate an effective magnetic vector potential $\mathbf{A}_{\text{eff}}$, producing Floquet bands with nonzero Chern numbers. The key advantage is that no magnetic material is required: time-reversal symmetry is broken by the modulation itself. Chiral edge states have been observed experimentally in silicon-photonic ring-resonator networks at telecom wavelengths, demonstrating topological protection at optical frequencies with forward-to-backward transmission contrast exceeding 10 dB.

Non-Hermitian complications in Floquet systems. The main limitation of Floquet topological photonics is that driven systems are inherently open: they exchange energy with the driving field and with the environment. Material loss, radiation leakage, and gain saturation all make the effective Floquet Hamiltonian non-Hermitian. The quasi-energy bands become complex, and the Chern number must be replaced by its biorthogonal generalization (Chapter 12). Furthermore, Floquet systems can exhibit parametric instability when the driving frequency is resonant with inter-band transitions, leading to exponential growth of certain modes—a Floquet-specific mechanism of non-Hermiticity that has no static analogue. These complications connect the Floquet approach directly to the non-Hermitian Berry connection of Section 11.8 and the open problems of Chapter 19.

11.5 Wilson loop and $\mathbb{Z}_2$ invariants

For systems with time-reversal symmetry where the Chern number vanishes, topological information is encoded in the *Wilson loop* operator [12]:

$$W(\ell) = \mathcal{P}\exp\left[-i\oint_\ell \mathbf{A}(\mathbf{k})\cdot\mathrm{d}\mathbf{k}\right], \tag{11.13}$$

where $\mathbf{A}$ is the non-Abelian Berry connection matrix for a group of N bands, $(\mathbf{A})_{mn} = i\langle \mathbf{u}_{m\mathbf{k}}|\nabla_{\mathbf{k}}|\mathbf{u}_{n\mathbf{k}}\rangle_\varepsilon$, and $\mathcal{P}$ denotes path ordering.

The eigenvalues of $W(\ell)$—the Wannier center flow—reveal the topological phase:

- **Trivial phase:** Wannier centers are constant (do not wind) as a function of the free momentum parameter.
- $\mathbb{Z}_2$ **topological phase:** Wannier centers wind by $\pm 2\pi$ across the Brillouin zone, corresponding to delocalized Wannier functions.
- **Obstructed atomic limit:** Wannier centers are pinned at non-trivial positions but do not wind.
- **Fragile topology:** Wannier centers wind with opposite slopes, and the topology can be unwound by adding trivial bands.

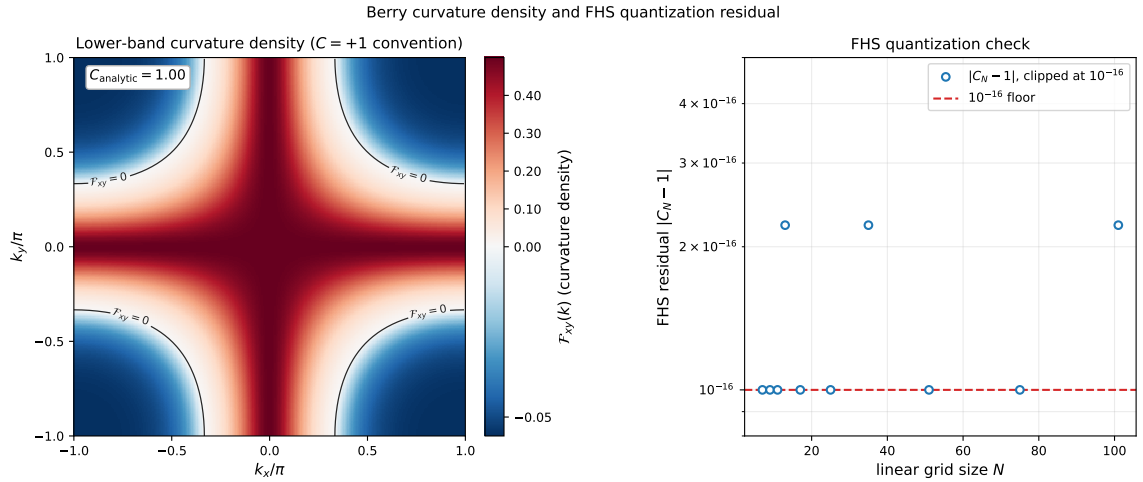


Figure 11.2 Gauge-invariant Berry-curvature computation with the FHS link-variable algorithm. The plotted model is a two-band lattice Dirac model used as a pedagogical proxy for Maxwell Bloch bands: in a photonic calculation, the spinor overlaps are replaced by energy-inner-product overlaps of Maxwell modes. The left panel shows the analytic Berry-curvature density $\mathcal{F}_{xy}(\mathbf{k})$ for the QWZ model at $m = -1$, using the sign convention whose Brillouin-zone integral gives $C = 1$. The right panel shows the FHS quantization residual $|C_N - 1|$ for selected $N \times N$ grids, demonstrating machine-precision integer quantization.

11.6 Numerical computation

11.6.1 Discretized Berry curvature

On a discrete grid of $\mathbf{k}$-points in the Brillouin zone, the Berry curvature is computed from the gauge-invariant lattice formula of Fukui, Hatsugai, and Suzuki (FHS):

$$\mathcal{F}_n(\mathbf{k}) \approx -\operatorname{Im} \ln\!\big[U_1(\mathbf{k})\, U_2(\mathbf{k} + \delta k_1 \hat{e}_1)\, U_1^*(\mathbf{k} + \delta k_2 \hat{e}_2)\, U_2^*(\mathbf{k})\big], \tag{11.14}$$

where

$$U_\alpha(\mathbf{k}) = \frac{\langle \mathbf{u}_{n\mathbf{k}} | \mathbf{u}_{n,\mathbf{k}+\delta k_\alpha \hat{e}_\alpha} \rangle_\varepsilon}{|\langle \mathbf{u}_{n\mathbf{k}} | \mathbf{u}_{n,\mathbf{k}+\delta k_\alpha \hat{e}_\alpha} \rangle_\varepsilon|} \tag{11.15}$$

is the phase-only link variable between adjacent $\mathbf{k}$-points in direction α. This formula is manifestly gauge-invariant and numerically robust: it does not require a smooth global gauge choice [12].

11.6.2 Practical considerations

Several practical issues arise when computing the photonic Berry curvature numerically:

1. **Band labeling:** at near-degenerate $\mathbf{k}$-points, the eigensolver may swap band labels. The link-variable approach mitigates this because the overlap U_α automatically tracks the smooth band, but the grid spacing δk must be small enough that the overlap is close to unity.
2. **Grid resolution:** the Berry curvature is typically peaked near degeneracies and high-symmetry points. An adaptive grid or a finer mesh near these features improves accuracy without excessive computational cost.

3. **Multi-band Chern number:** when two or more bands are separated from others by a common gap, the total Chern number of the group is computed from the non-Abelian Berry curvature $\mathcal{F}_{mn}$ using $\mathcal{C} = (2\pi)^{-1} \int_{\mathrm{BZ}} \mathrm{tr}\, \mathcal{F}\ \mathrm{d}^2k$.
4. **Gauge smoothing:** for Wilson-loop calculations, a smooth gauge across one direction of the BZ is needed. The standard method is to fix the gauge on a reference line and parallel-transport it across the BZ using the SVD of the overlap matrix.

11.6.3 Chern number from integration

For the FHS formula, $\mathcal{F}_n(\mathbf{k})$ is already the Berry flux through a plaquette, so the Chern number is

$$\mathcal{C}_n = \frac{1}{2\pi} \sum_{\text{plaquettes}} \mathcal{F}_n(\mathbf{k}). \tag{11.16}$$

No additional factor of $\Delta k_x\, \Delta k_y$ is included. The result is automatically an integer up to numerical precision because the lattice formula captures the correct topology.

11.7 Bulk-boundary correspondence

The most striking physical consequence of a nonzero Chern number is the existence of topologically protected edge states. The *bulk-boundary correspondence* states that the number of chiral edge modes at an interface between two insulators is equal to the difference of their gap Chern numbers:

$$N_{\text{edge}} = \mathcal{C}_{\text{gap}}^{(\text{left})} - \mathcal{C}_{\text{gap}}^{(\text{right})}. \tag{11.17}$$

These edge states propagate in one direction only and are robust against backscattering from disorder or defects: their existence is guaranteed by topology, not by fine-tuned material parameters.

11.7.1 Edge dispersion

The chiral edge states have a dispersion relation $\omega(k_\parallel)$ that crosses the bulk gap. The group velocity $v_g = \mathrm{d}\omega/\,\mathrm{d}k_\parallel$ has a definite sign (positive or negative) determined by the Chern number. The edge modes connect the bulk bands above and below the gap, interpolating smoothly between the valence and conduction bands. The number of crossings equals $|\mathcal{C}|$, providing a graphical check of the bulk-boundary correspondence.

11.7.2 Robustness

In photonic Chern insulators (realized with magneto-optical materials), the chiral edge states provide one-way waveguides for light—immune to backscattering from fabrication imperfections, sharp bends, or surface roughness. The robustness has a simple topological explanation: a backscattered wave would need to reverse its direction, which requires coupling to a counter-propagating edge mode. But only one direction of propagation is available at each edge (there is no counter-propagating channel), so backscattering is forbidden by topology.

This topological protection is exact in the limit of an infinite bulk gap and degrades gradually as the gap closes. In practice, the backscattering suppression is exponential in the ratio of the gap width to the defect strength.

Hermitian bulk-boundary correspondence

In Hermitian photonic crystals, the Chern number computed from the Maxwell Bloch states determines the number of topologically protected one-way edge channels. This correspondence is exact and robust: it survives disorder, deformation, and any perturbation that does not close the bulk gap. In non-Hermitian systems, this correspondence can fail; the skin effect and non-Bloch bulk-boundary correspondence are the subject of Chapters 14 and 15.

11.8 Preview: non-Hermitian modifications

The Berry curvature and Chern number derived above assumed a Hermitian inner product and real eigenfrequencies. When non-Hermiticity is introduced:

1. The eigenfrequencies $\omega_n(\mathbf{k})$ become complex, and "band gaps" must be redefined in the complex frequency plane (Chapter 12).
2. The left and right eigenvectors differ, and the Berry connection must be generalized to a *biorthogonal* form:
$$\mathcal{A}_n^{\mathrm{bi}} = i\,\frac{\langle \mathbf{u}_n^{\mathrm{L}} | \nabla_{\mathbf{k}} | \mathbf{u}_n^{\mathrm{R}} \rangle}{\langle \mathbf{u}_n^{\mathrm{L}} | \mathbf{u}_n^{\mathrm{R}} \rangle}. \tag{11.18}$$
The appropriate bilinear pairing depends on the Maxwell formulation and material dispersion; the key point is that the left mode supplies the dual vector. The biorthogonal Berry connection is generally complex-valued, and its imaginary part tracks changes in the left–right normalization rather than an ordinary geometric phase.
3. New topological invariants—spectral winding numbers and point-gap invariants—appear that have no Hermitian analogue (Chapter 13).
4. The bulk-boundary correspondence can fail: the bulk Chern number may predict edge states that are absent, or miss states that are present, due to the non-Hermitian skin effect (Chapter 14).

Part IV of this book develops each of these modifications, building on the Hermitian topological foundation established in this chapter.

11.8.1 From Hermitian to biorthogonal: what changes physically

The transition from Hermitian to non-Hermitian Berry curvature is not merely a mathematical generalization; it changes the physical content of the topological invariants.

In the Hermitian case, the Berry curvature (11.4) is real-valued and determines the transverse deflection of wave packets via (11.6). The Chern number counts chiral edge modes. Both are observable through intensity measurements.

In the non-Hermitian case, the biorthogonal Berry connection (11.18) is generically complex. Its real part generalizes the Hermitian Berry phase and can govern geometric interference effects.

Its imaginary part measures geometric amplification or attenuation associated with changes in the left–right normalization, and is closely tied to eigenvector non-orthogonality. Near an exceptional point the biorthogonal eigenbasis becomes singular, and encircling the defect produces the characteristic half-integer vorticity discussed in later chapters.

The biorthogonal Chern number

$$\mathcal{C}_n^{\mathrm{bi}} = \frac{1}{2\pi} \int_{\mathrm{BZ}} \mathcal{F}_n^{\mathrm{bi}}(\mathbf{k}) \; \mathrm{d}^2 k \tag{11.19}$$

remains integer-valued for an isolated band over the Brillouin-zone torus, provided a smooth biorthogonal basis exists and no exceptional degeneracy pierces the integration surface. When the relevant gap is a point gap rather than a line gap, the natural invariant is instead a spectral winding number of the complex band or Bloch Hamiltonian. That shift from Chern edge-mode counting to point-gap winding and skin-effect physics is one of the central themes of Part IV.

Chapter 12

Complex Bands and Non-Hermitian Gaps

In Hermitian photonic crystals, a band gap is a frequency window with no propagating modes. In non-Hermitian crystals, eigenfrequencies are complex, and the very notion of a "gap" must be rethought. Two fundamentally different gap structures emerge—the line gap *and the* point gap *[11, 44, 84]—each supporting its own class of topological invariants. The theory of complex photonic bands is developed by introducing the two gap types with concrete examples, deriving the spectral winding number from the complex band dispersion, and showing how the biorthogonal Berry phase [85, 86, 87, 15, 88, 14] generalizes the Hermitian topological framework to non-Hermitian systems.*

12.1 From real to complex band structure

When the permittivity of a periodic photonic crystal acquires an imaginary part—through material loss, gain, or radiation coupling—the Bloch eigenproblem of Chapter 3 still admits solutions labeled by $\mathbf{k}$ (Bloch's theorem depends on translational symmetry, not Hermiticity), but the eigenfrequencies become complex:

$$\omega_n(\mathbf{k}) = \omega'_n(\mathbf{k}) + i\omega''_n(\mathbf{k}). \tag{12.1}$$

The "band structure" is now a set of curves in the three-dimensional space $(\mathbf{k}, \mathrm{Re}\,\omega, \mathrm{Im}\,\omega)$, or equivalently a set of trajectories in the complex ω plane as $\mathbf{k}$ sweeps the Brillouin zone.

12.1.1 How complex bands arise from Maxwell equations

The passage from real to complex bands can be understood directly from the Maxwell eigenproblem. In a lossless periodic medium ($\mathrm{Im}\,\varepsilon_\mathrm{r} = 0$), the operator $\hat{\Theta}_\mathbf{k} = \varepsilon_\mathrm{r}^{-1}\nabla\times(\mu_\mathrm{r}^{-1}\nabla\times\,\cdot\,)$ acting on Bloch-periodic fields is self-adjoint in the ε_r-weighted inner product. Its eigenvalues $\omega_n^2(\mathbf{k})/c^2$ are real and positive.

When $\operatorname{Im}\varepsilon_{\mathrm{r}} \neq 0$, the operator ceases to be self-adjoint, and its eigenvalues become complex. The imaginary part of the eigenfrequency has a clear physical meaning from the energy balance:

$$\omega_n''(\mathbf{k}) \approx -\frac{\omega_n'(\mathbf{k})}{2} \frac{\int_{\mathrm{uc}} \operatorname{Im}\varepsilon_{\mathrm{r}}(\mathbf{r})\,|\mathbf{u}_{n\mathbf{k}}(\mathbf{r})|^2 \;\mathrm{d}^d r}{\int_{\mathrm{uc}} \operatorname{Re}\varepsilon_{\mathrm{r}}(\mathbf{r})\,|\mathbf{u}_{n\mathbf{k}}(\mathbf{r})|^2 \;\mathrm{d}^d r}, \tag{12.2}$$

valid to first order in $\operatorname{Im}\varepsilon_{\mathrm{r}}$. Modes concentrated in lossy regions ($\operatorname{Im}\varepsilon_{\mathrm{r}} > 0$) acquire $\omega_n'' < 0$ (decay); modes in gain regions ($\operatorname{Im}\varepsilon_{\mathrm{r}} < 0$) acquire $\omega_n'' > 0$ (amplification). Different Bloch modes at the same $\mathbf{k}$ sample the gain/loss distribution differently, so they can have imaginary parts with different signs and magnitudes.

12.1.2 Consequences of complexification

This complexification has immediate consequences:

- Bands can no longer be ordered by frequency along the real axis; two bands may cross in $\operatorname{Re}\omega$ while being separated in $\operatorname{Im}\omega$, or vice versa.
- The concept of a "gap" must specify what it means for two complex-valued curves to be "separated."
- Exceptional points—where both eigenvalues and eigenvectors coalesce (Chapter 8)—can appear as singular points in the $(\mathbf{k}, \omega)$ space.
- The standard bulk-boundary correspondence, which relates the Chern number to edge-state counts, may fail because the bulk invariant was defined assuming a real spectrum (Chapter 15).

12.1.3 A concrete example: 1D gain–loss lattice

Consider a one-dimensional tight-binding model with alternating on-site gain and loss:

$$H(k) = \begin{pmatrix} i\gamma & t(1+e^{-ik}) \\ t(1+e^{ik}) & -i\gamma \end{pmatrix}, \tag{12.3}$$

where t is the hopping and γ is the gain/loss rate. This is an SSH model with a $\mathcal{PT}$-symmetric non-Hermitian perturbation. The eigenvalues are

$$\omega_{\pm}(k) = \pm\sqrt{4t^2\cos^2(k/2) - \gamma^2}. \tag{12.4}$$

For any nonzero γ, momenta close to $k = \pi$ satisfy $|2t\cos(k/2)| < |\gamma|$, so the square root is imaginary and the corresponding Bloch states are in a $\mathcal{PT}$-broken sector. Momenta farther from $k = \pi$ remain real when $|2t\cos(k/2)| > |\gamma|$. Thus this balanced gain–loss dimer is useful for illustrating complex bands and exceptional points, but it is not by itself the canonical point-gap skin-effect model. For point-gap topology and skin effect one needs non-reciprocal hopping, as in the Hatano–Nelson or asymmetric SSH models discussed below [14, 88].

Perturbation theory for complex bands. The first-order imaginary shift (12.2) can be made concrete for this lattice. At $k = 0$ the Bloch modes of the unperturbed SSH model are

$|u_\pm\rangle = (1,\ \pm 1)^{\mathsf{T}}/\sqrt{2}$ with $\omega_\pm(0) = \pm 2t$. The gain–loss perturbation $\delta H = i\gamma\,\sigma_z$ has matrix elements

$$\langle u_\pm|\delta H|u_\pm\rangle = 0, \qquad \langle u_\pm|\delta H|u_\mp\rangle = \pm i\gamma. \tag{12.5}$$

The diagonal elements vanish because $|u_+\rangle$ and $|u_-\rangle$ have equal weight on the gain and loss sites. The imaginary part of $\omega_\pm$ is therefore *zero* to first order at $k = 0$: gain and loss cancel exactly, a consequence of $\mathcal{PT}$ symmetry. The leading contribution to $\mathrm{Im}\,\omega$ comes from the off-diagonal coupling at second order:

$$\omega''_\pm(0) \approx \mp\frac{\gamma^2}{4t}, \tag{12.6}$$

which shifts the real eigenfrequencies at $k = 0$ to $\omega_\pm(0) = \pm\sqrt{4t^2 - \gamma^2}$. The imaginary part remains zero there as long as $|\gamma| < 2t$; imaginary splittings appear first near the momenta where $4t^2\cos^2(k/2) = \gamma^2$.

Tracking the spectral loop. As γ increases from 0 to $2t$, the band structure undergoes a sequence of qualitative changes:

1. $\gamma = 0$: two real bands separated by a gap of width $2t$. The spectrum is a line segment on the real axis.
2. $0 < \gamma < 2t$: exceptional points appear at $k = \pm 2\arccos(\gamma/2t)$, separating real-frequency arcs from imaginary-frequency arcs.
3. $\gamma = 2t$: the real arcs shrink to $k = 0$, and the remaining spectrum lies on the imaginary axis.
4. $\gamma > 2t$: all momenta are in the broken sector and the spectrum is purely imaginary.

This progression illustrates a general principle: *complex bands change topology at exceptional points*. Point-gap winding, however, is most transparently generated by non-reciprocal hopping rather than by balanced on-site gain and loss alone.

12.2 Line gaps and point gaps

The classification of non-Hermitian topological phases depends crucially on the type of gap in the complex energy plane [89].

12.2.1 Line gap

A *line gap* exists when the complex spectrum does not cross a specified line in the complex plane—typically the real or imaginary axis after shifting the reference frequency. Equivalently, the spectrum can be separated into regions lying on different sides of that line. A line gap along the real axis, for example, means that the spectrum avoids the real axis; a line gap along the imaginary axis means that it avoids a chosen vertical line in the complex-frequency plane.

Line gaps are the natural generalization of Hermitian band gaps: the spectrum is separated into two regions by a line, and each region can carry its own topological invariant. The topological classification of systems with a line gap is closely related to the Hermitian classification, with some modifications due to the enlarged symmetry algebra.

Physical meaning. A line gap makes it possible to flatten the complex spectrum to a Hermitian-like or anti-Hermitian-like form without crossing the reference line. For photonic bands this is a spectral separation statement, not necessarily a division into stable and unstable modes; the latter interpretation applies only when the chosen line is the real-frequency axis and the two spectral groups lie on opposite sides of it.

12.2.2 Point gap

A *point gap* exists when the complex spectrum avoids a single point in the complex plane—typically the origin or a chosen reference energy E_{ref}. As $\mathbf{k}$ sweeps the Brillouin zone, the eigenvalues $\omega_n(\mathbf{k})$ trace closed loops in the complex plane that may wind around E_{ref}.

Point gaps are intrinsically non-Hermitian: in a Hermitian system the spectrum is confined to the real axis and cannot wind around a point in the complex plane. The topological invariant associated with a point gap is the *spectral winding number*.

Physical meaning. A nonzero winding number around a reference point means that the complex band cannot be continuously deformed to a point without passing through the reference energy. This topological obstruction has a physical consequence: it predicts the non-Hermitian skin effect (Chapter 14), in which all eigenstates under open boundary conditions localize at one edge of the system.

12.2.3 Relationship between the two gap types

Line gaps and point gaps are not alternative descriptions of the same physics—they describe genuinely different topological structures. A system can have a line gap without a point gap, a point gap without a line gap, both, or neither. The relationship is:

- A line-gapped spectrum avoids every reference point on the gap line, so point gaps can be defined there. These point gaps need not carry nonzero winding.
- A point gap with nonzero winding around E_{ref} is incompatible with a line gap along any line that extends through E_{ref} without intersecting the spectral loop.
- In one dimension, nonzero point-gap winding is tied to the skin effect, while line-gap invariants are the closer analogues of conventional topological edge-state indices.

Which gap does a photonic crystal have? The answer depends on the gain–loss distribution and the frequency range. A purely lossy photonic crystal ($\mathrm{Im}\,\varepsilon_{\mathrm{r}} > 0$ everywhere) has all eigenfrequencies in the lower half-plane: it possesses a line gap along the real axis but no point gap with nonzero winding (the spectrum stays on one side of the line and cannot wind). A $\mathcal{PT}$-symmetric crystal with balanced gain and loss can have eigenfrequencies on both sides of the real axis: if the spectrum forms a loop, a point gap with nonzero winding appears, but the line gap is absent. The transition between these two regimes is the $\mathcal{PT}$-symmetry breaking threshold (Chapter 9).

A practical diagnostic is the following: given the complex band structure $\omega_n(\mathbf{k})$, plot the spectrum in the complex ω plane. If the spectrum stays on one side of a line, the system has a line gap; if it forms a loop enclosing a point, it has a point gap.

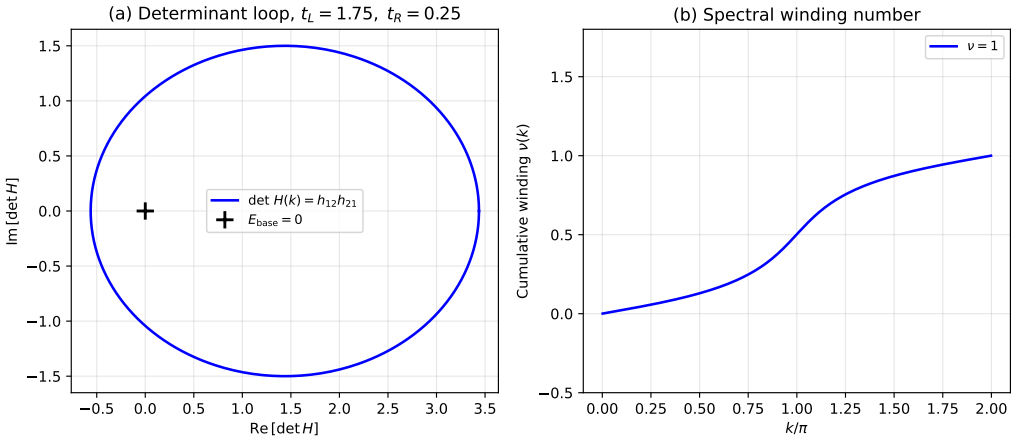


Figure 12.1 Spectral winding number for the asymmetric-hopping NH SSH model ($t_L = 1.75$, $t_R = 0.25$, $t_2 = 1.0$). (a) The determinant $\det H(k) = h_{12}(k)\,h_{21}(k)$ traces a loop in the complex plane that encircles the origin, confirming a point-gap topology with winding number $\nu = 1$. (b) The cumulative winding accumulates smoothly from 0 to 1 as k traverses the Brillouin zone. Because $t_L > t_2 > t_R$, only h_{21} winds around the origin, producing a net spectral winding.

The labels in such a diagram are therefore reference-dependent: changing the line or point used to define the gap can change which invariant is well defined.

Two gaps, two topologies

The 38-fold classification of non-Hermitian topological phases is built by considering both gap types in combination with the symmetries of the system. A complete classification must specify: (i) the spatial dimension, (ii) the symmetry class, and (iii) whether a line gap, a point gap, or both are present.

12.3 Spectral winding number

12.3.1 Definition

For a one-dimensional periodic system with complex band structure $\omega(k)$ (a single band for simplicity), the *spectral winding number* around a reference point E_{ref} is

$$\boxed{W = \frac{1}{2\pi i}\oint_{\text{BZ}} \frac{\mathrm{d}}{\mathrm{d}k}\ln[\omega(k) - E_{\text{ref}}]\;\mathrm{d}k = \frac{1}{2\pi i}\oint_{\text{BZ}} \frac{\omega'(k)}{\omega(k) - E_{\text{ref}}}\;\mathrm{d}k.} \tag{12.7}$$

This integer counts the number of times the complex band encircles E_{ref} as k traverses the Brillouin zone. It is a topological invariant: it cannot change unless the band touches E_{ref} (the point gap closes).

12.3.2 Physical meaning

A nonzero winding number means the complex band forms a closed loop in the complex ω plane that cannot be contracted to a point without crossing E_{ref}. This is the simplest example of a topological invariant that exists *only* in non-Hermitian systems: a real-valued band confined to the real axis has zero winding around any point off the axis.

In photonic crystals with complex $\varepsilon_{\mathrm{r}}(\mathbf{r})$, the winding number classifies point-gap topology of complex photonic bands. In one dimension it is tied to the non-Hermitian skin effect under open boundary conditions rather than to ordinary Hermitian edge-state counting [90].

12.3.3 Computational procedure

The winding number (12.7) is straightforward to compute numerically:

1. Sample k uniformly on $[-\pi, \pi]$ with $N \gg 1$ points.
2. Evaluate $\omega(k)$ from the Bloch eigenproblem (or from a tight-binding model).
3. Form the shifted band $z(k) = \omega(k) - E_{\mathrm{ref}}$.
4. Compute the total phase winding: $W = [\arg z(\pi) - \arg z(-\pi)]/(2\pi)$, using the unwrapped phase.
5. Alternatively, use the discrete winding formula: $W = (2\pi)^{-1} \sum_{j=1}^{N} \arg[z(k_{j+1})/z(k_j)]$, which is gauge-invariant and numerically robust.

The result is automatically an integer up to rounding errors.

12.3.4 Example: Hatano–Nelson model

The Hatano–Nelson model (Chapter 14) provides a minimal example. With asymmetric hopping $t_R \neq t_L$, the PBC band structure is

$$\omega(k) = t_R\, e^{ik} + t_L\, e^{-ik}. \tag{12.8}$$

This traces an ellipse in the complex plane with semi-axes $|t_R + t_L|$ and $|t_R - t_L|$. When $t_R \neq t_L$, the ellipse has nonzero area, and any point inside it has winding number $W = \pm 1$. Points outside the ellipse have $W = 0$.

The physical consequence is striking: the OBC spectrum fills the *interior* of the PBC ellipse (see Chapter 14), and all OBC eigenstates localize at one edge. The winding number predicts this macroscopic spectral collapse.

12.4 Biorthogonal Berry phase

In a non-Hermitian periodic system, the left and right Bloch eigenvectors $\mathbf{u}_n^{\mathrm{L}}(\mathbf{k})$ and $\mathbf{u}_n^{\mathrm{R}}(\mathbf{k})$ are generally different (Section 6.4). The Berry connection must be generalized to a biorthogonal form:

$$\mathcal{A}_n^{\mathrm{bi}}(\mathbf{k}) = i\, \frac{\langle \mathbf{u}_n^{\mathrm{L}}(\mathbf{k}) | \nabla_{\mathbf{k}} | \mathbf{u}_n^{\mathrm{R}}(\mathbf{k}) \rangle}{\langle \mathbf{u}_n^{\mathrm{L}}(\mathbf{k}) | \mathbf{u}_n^{\mathrm{R}}(\mathbf{k}) \rangle}. \tag{12.9}$$

The biorthogonal Berry curvature and Chern number are defined analogously. They reduce to the standard (Hermitian) definitions when the system is Hermitian and left and right eigenvectors coincide.

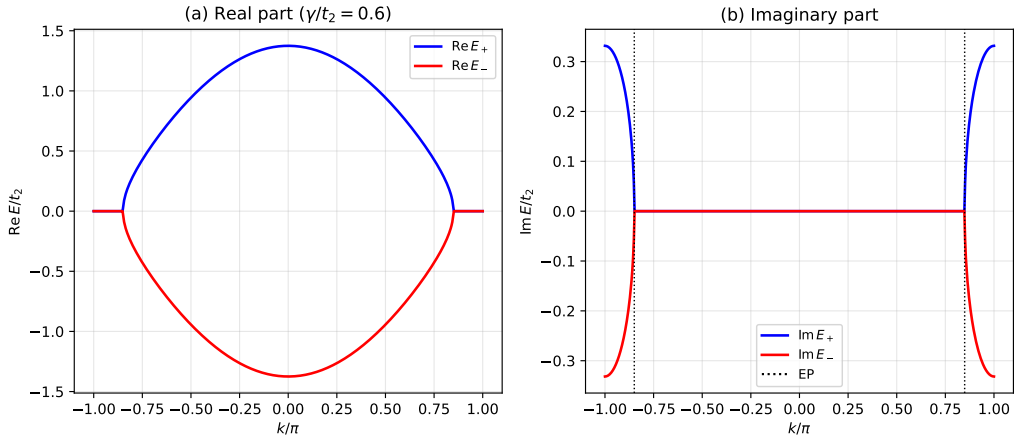


Figure 12.2 Complex band structure of the gain–loss SSH model ($t_1 = 0.5$, $t_2 = 1.0$, $\gamma/t_2 = 0.6$). (a) Real part of the energy: the two bands touch near the exceptional points (dotted lines), where the square root in $E_\pm = \pm\sqrt{h_{12}h_{21} - \gamma^2}$ vanishes. (b) Imaginary part: zero inside the $\mathcal{PT}$-symmetric region and bifurcating symmetrically in the $\mathcal{PT}$-broken region between the EPs, where one mode is amplified and the other is attenuated.

12.4.1 Complex-valued Berry phase

Unlike the Hermitian Berry connection, which is real-valued (for a single band), the biorthogonal Berry connection (12.9) is generically complex. Its real and imaginary parts have distinct physical meanings:

- The *real part* $\mathrm{Re}\,\mathcal{A}_n^{\mathrm{bi}}$ generalizes the Hermitian Berry connection and enters the semiclassical equations of motion for wave-packet dynamics.
- The *imaginary part* $\mathrm{Im}\,\mathcal{A}_n^{\mathrm{bi}}$ is a new non-Hermitian contribution that describes the $\mathbf{k}$-dependence of the mode non-orthogonality. It is related to the derivative of the Petermann factor with respect to $\mathbf{k}$.

The Berry connection itself can be complex, but the first Chern number of an isolated line-gapped band remains integer-valued when a smooth biorthogonal basis exists over the Brillouin-zone torus. Exceptional points or gap closings invalidate the construction rather than producing a meaningful non-integer Chern number.

Example: biorthogonal Berry phase of the gain–loss SSH model. For the gain–loss SSH model (12.3) with $\gamma < 2t$ (the $\mathcal{PT}$-unbroken phase), the right eigenvector of the lower band is $|u_-^{\mathrm{R}}(k)\rangle = (1,\ -\omega_-(k)/[t(1+e^{ik})])^{\mathsf{T}}$ (up to normalisation). The left eigenvector is obtained by replacing $H(k)$ with $H^\dagger(k)$. The biorthogonal Berry phase around the Brillouin zone is

$$\phi^{\mathrm{bi}} = \oint_{\mathrm{BZ}} \mathcal{A}_-^{\mathrm{bi}}(k)\ \mathrm{d}k. \tag{12.10}$$

For $\gamma = 0$ (Hermitian limit), this reduces to the standard Zak phase: $\phi^{\mathrm{bi}} = \pi$ in the topological phase ($t' > t$) and $\phi^{\mathrm{bi}} = 0$ in the trivial phase. As γ increases, ϕ^{bi} acquires an imaginary part that grows continuously, but the real part remains quantized at 0 or π as long as the chiral symmetry $\sigma_z H(k)\sigma_z = -H(k)$ is preserved. At the $\mathcal{PT}$-breaking transition ($\gamma = 2t$), the biorthogonal Berry phase is undefined because the denominator $\langle u_-^{\mathrm{L}}|u_-^{\mathrm{R}}\rangle$ vanishes at the EP.

This example illustrates a general principle: *the real part of the biorthogonal Berry phase is topologically protected by symmetry, while the imaginary part is a smooth, non-topological measure of non-Hermiticity.* The EP is the singularity where the Berry phase becomes undefined—the non-Hermitian analogue of a gap closing in Hermitian topology.

12.4.2 Singularity at exceptional points

Near an exceptional point, the denominator $\langle \mathbf{u}_n^{\mathrm{L}} | \mathbf{u}_n^{\mathrm{R}} \rangle \to 0$ and the biorthogonal Berry connection diverges. This singularity reflects the coalescence of eigenvectors and is intimately connected to the topological charge of the EP (Section 8.9).

The Berry phase accumulated around a loop encircling an EP in the BZ is $\pm\pi$ for an EP_2 (a π-Berry phase), analogous to the Berry phase around a Dirac point in graphene. However, the non-Hermitian version involves the biorthogonal inner product and does not require time-reversal or inversion symmetry.

12.5 Exceptional points in band structure

In a two-dimensional Brillouin zone, exceptional points appear as isolated points in $\mathbf{k}$-space where two complex bands touch. Near an EP at $\mathbf{k}_{\mathrm{EP}}$, the eigenvalue splitting follows the square-root topology:

$$\omega_{\pm}(\mathbf{k}) \approx \omega_{\mathrm{EP}} \pm \sqrt{\alpha(\mathbf{k} - \mathbf{k}_{\mathrm{EP}})}, \tag{12.11}$$

with α a complex coefficient. Encircling the EP in momentum space exchanges the two bands—the same Riemann-surface structure as in parameter-space encircling (Section 8.4).

12.5.1 Exceptional lines and surfaces

In three-dimensional systems or in systems with additional parameters, EPs can form lines or surfaces in the combined $(\mathbf{k}, \text{parameter})$ space. An *exceptional ring* in a 2D BZ is a closed curve of EPs; the interior of the ring and the exterior have different topological indices.

In three-dimensional photonic crystals [91, 92, 93], exceptional points generically form one-dimensional curves in the 3D BZ—*exceptional lines*. These lines are topological objects: they carry a topological charge (a π Berry phase for any loop encircling the line) and can be linked or knotted with one another. The linking number is a topological invariant that cannot change without two exceptional lines crossing.

12.5.2 Bulk Fermi arcs

When two complex bands are projected onto the real-frequency axis, the regions where they overlap in $\mathrm{Re}\,\omega$ but differ in $\mathrm{Im}\,\omega$ create *bulk Fermi arcs*—open arcs in the surface Brillouin zone that connect projections of EPs. These arcs are a non-Hermitian analogue of the Fermi arcs in Weyl semimetals [94], but they exist in the bulk spectrum rather than at a surface.

The physical consequence is that constant-frequency cuts through the complex band structure show open contours rather than closed Fermi surfaces. This has been observed experimentally in photonic crystals with controlled gain and loss [95, 96].

Observing bulk Fermi arcs in photonics. In a photonic crystal slab with $\mathcal{PT}$-symmetric gain and loss, bulk Fermi arcs manifest as anomalous isofrequency contours in the transmission spectrum. At a frequency where two complex bands overlap in $\mathrm{Re}\,\omega$ but split in $\mathrm{Im}\,\omega$, an incident plane wave couples to both bands with different decay rates. The transmitted intensity shows a characteristic double-Lorentzian lineshape whose angular dependence traces the Fermi arc in the surface Brillouin zone. The arc endpoints are the projections of the bulk EPs, which can be located by finding the angles at which the double Lorentzian degenerates into a single peak with a squared-Lorentzian profile—the spectral signature of an EP.

This measurement protocol requires only angle-resolved transmission spectroscopy, making it accessible with standard photonic-crystal characterization equipment. The main experimental challenge is achieving sufficiently uniform gain–loss distribution across the sample to resolve the Fermi-arc topology.

12.6 The 38-fold classification

12.6.1 From tenfold to 38-fold

The full classification of non-Hermitian topological phases [89] extends the tenfold Altland–Zirnbauer classification of Hermitian systems to 38 symmetry classes. The expansion arises because:

1. Non-Hermiticity distinguishes transposition from complex conjugation, doubling the number of anti-unitary symmetry classes.
2. Both line gaps and point gaps must be considered independently.
3. Sublattice symmetry and pseudo-Hermiticity become independent symmetry classes.

Specifically, in Hermitian systems, time-reversal symmetry (T) and particle-hole symmetry (C) are both anti-unitary and satisfy $THT^{-1} = H$ and $CHC^{-1} = -H$. In non-Hermitian systems, transposition ($T^{\top}$: $H^{\top} = H$) and complex conjugation (T^{*}: $H^{*} = H$) are distinct symmetries because $H \neq H^{\dagger}$. Similarly, $C^{\top}$ and C^{*} become independent. This doubling of symmetry types, combined with the two gap types, yields 38 distinct classes.

12.6.2 Photonic symmetry classes

For the photonic systems that are our primary concern, the most relevant symmetry classes and their topological invariants can be understood through a few paradigmatic examples.

In one dimension with a point gap, the classification is $\mathbb{Z}$, with the winding number (12.7) as the topological invariant. This is the symmetry class of the Hatano–Nelson model: the asymmetric hopping breaks Hermiticity, opens a point gap, and the resulting nonzero winding predicts the non-Hermitian skin effect. Any one-dimensional photonic lattice with asymmetric coupling—whether from non-reciprocal elements, spatially varying gain and loss, or geometry-induced asymmetry—falls into this class.

In two dimensions with a line gap and broken time-reversal symmetry, the classification is again $\mathbb{Z}$, but the invariant is the biorthogonal Chern number of Chapter 11. This generalizes the quantum Hall class to non-Hermitian systems: a magneto-optical photonic crystal with uniform

loss has complex bands separated by a line gap, and the Chern number computed from the biorthogonal Berry curvature correctly predicts the number of chiral edge states, even though the edge-state frequencies are themselves complex.

Two-dimensional systems with a point gap support a different $\mathbb{Z}$ invariant: a two-dimensional winding number that characterizes how the complex spectrum wraps around the reference point as $\mathbf{k}$ sweeps the Brillouin zone. This invariant predicts the two-dimensional non-Hermitian skin effect, where bulk states localize at corners of a finite sample rather than at edges.

Finally, one-dimensional systems with $\mathcal{PT}$ symmetry and a line gap are classified by $\mathbb{Z}_2$: a quantized Zak phase that takes values 0 or π. This is the non-Hermitian analogue of the Su–Schrieffer–Heeger topological insulator, and the $\mathbb{Z}_2$ invariant correctly predicts the presence or absence of $\mathcal{PT}$-symmetric edge states.

Example: symmetry-class identification for the gain–loss SSH model. The systematic procedure for identifying the symmetry class of a concrete non-Hermitian photonic system is as follows. We illustrate each step using the gain–loss SSH model (12.3), whose Bloch Hamiltonian is $H(k) = (t_1 + t_2\cos k)\,\sigma_x + t_2 \sin k\,\sigma_y + i\gamma\,\sigma_z$.

Step 1. Write the Bloch Hamiltonian $H(\mathbf{k})$ from the Maxwell eigenproblem. For the gain–loss SSH model, this is the 2×2 matrix above with two sublattice degrees of freedom.

Step 2. Check for time-reversal symmetry. We need $H(\mathbf{k}) = H^*(-\mathbf{k})$. Since the hopping amplitudes t_1, t_2 are real and $i\gamma$ is purely imaginary, we find $H^*(-k) = (t_1 + t_2\cos k)\,\sigma_x - t_2 \sin k\,\sigma_y - i\gamma\,\sigma_z$. This equals $H(k)$ only if $\gamma = 0$, so T^* symmetry is broken by the gain–loss. However, transposition symmetry $H(k) = H^\top(-k)$ *is* present because σ_x and σ_z are symmetric and σ_y is antisymmetric, and the σ_y coefficient is odd in k. This is the non-Hermitian generalization of time-reversal for spinless particles.

Step 3. Check for $\mathcal{PT}$ symmetry: $\sigma_x H(k)\sigma_x = H^*(-k)$ can be verified by direct computation. The σ_x operator interchanges the two sublattices (gain and loss sites), while complex conjugation reverses the sign of $i\gamma$. Thus $\mathcal{PT}$ symmetry is present, corresponding to pseudo-Hermiticity $\eta H \eta^{-1} = H^\dagger$ with $\eta = \sigma_x$.

Step 4. Check for chiral (sublattice) symmetry: $\sigma_z H(k)\sigma_z = -H(k)$. Direct computation confirms this: $\sigma_z\sigma_x\sigma_z = -\sigma_x$, $\sigma_z\sigma_y\sigma_z = -\sigma_y$, and $\sigma_z(i\gamma\sigma_z)\sigma_z = i\gamma\sigma_z$. Wait—the last term gives $+i\gamma\sigma_z$, not $-i\gamma\sigma_z$. So standard chiral symmetry is broken by the gain–loss. However, a *non-Hermitian* chiral symmetry $\sigma_z H(k)\sigma_z = -H^\dagger(k)$ *is* present, because $H^\dagger(k)$ flips the sign of $i\gamma\sigma_z$.

Step 5. Cross-reference with the 38-fold table. The combination of transposition symmetry $T^\top$ with $(T^\top)^2 = +1$, non-Hermitian chiral symmetry, and pseudo-Hermiticity places the system in class BDI† with a point gap. The table entry for BDI† in one dimension with a point gap is $\mathbb{Z}$: the topological invariant is the winding number, consistent with the spectral-loop analysis of Section 12.1.3.

12.6.3 Reduction to the Hermitian tenfold way

The 38-fold classification must reduce to the standard tenfold Altland–Zirnbauer classification when the Hamiltonian is Hermitian. This consistency check is instructive. When $H = H^\dagger$: transposition symmetry $H = H^\top$ becomes equivalent to complex conjugation $H = H^*$ (since

$H^\top = H^{\dagger *} = H^*$), so the doubling of anti-unitary symmetries collapses. The point gap becomes trivially satisfied (the spectrum is real and generically does not pass through any complex reference point), so point-gap invariants become trivial. Only line-gap invariants survive, and they coincide with the standard band-gap invariants. The 38 classes collapse to 10, and the topological periodic table reduces to the familiar Kitaev table.

For the gain–loss SSH model, taking $\gamma \to 0$ restores Hermiticity. The BDI† class collapses to the Hermitian BDI class, and the $\mathbb{Z}$ winding number reduces to the standard winding number of the SSH model. The point-gap topology becomes trivial (the spectrum is real and does not wind around any complex point), and only the line-gap $\mathbb{Z}$ invariant—the Zak phase—survives. This continuity between non-Hermitian and Hermitian classifications is essential for physical consistency: the topological properties of a photonic crystal with weak gain and loss must continuously deform into those of the passive structure.

12.6.4 Beyond the periodic table

The 38-fold classification is a bulk classification that applies to infinite periodic systems. Several features of real photonic structures lie outside its scope:

- Finite-size effects modify the spectrum and can change the topological phase.
- Disorder can stabilize or destroy non-Hermitian topological phases, depending on the symmetry class.
- The classification assumes a single-particle (quadratic) Hamiltonian; nonlinear and many-body effects require separate analysis (Chapter 19).
- The Maxwell operator has additional structure (vector nature, gauge redundancy, frequency-dependent material response) that is not fully captured by the tight-binding classification. Lifting the 38-fold table to continuous Maxwell operators is an open problem (Section 19.9).

Non-Hermitian topology requires new invariants

The topological classification of Hermitian systems—Chern numbers, $\mathbb{Z}_2$ indices, winding numbers of real bands—is insufficient for non-Hermitian photonics. Complex bands, point gaps, spectral winding, and biorthogonal Berry phases introduce genuinely new topological structures. These structures predict phenomena—the skin effect, anomalous edge states, bulk Fermi arcs—that have no Hermitian analogue.

Chapter 13

Photonic Point-Gap Topology

The spectral winding number introduced in Chapter 12 counts how many times a complex band encircles a reference point. This chapter specializes that invariant to photonic crystals with complex permittivity, where the non-Hermiticity is rooted in the Maxwell equations rather than in a tight-binding model. The discussion shows how reciprocity, the vector nature of light, and the energy inner product constrain the point-gap topology of photonic systems, derive the winding number from a concrete one-dimensional setting, and connect the photonic winding number to spectral flow, boundary accumulation, and the non-Bloch bulk–boundary correspondence that have no analogue in the Hermitian case.

13.1 Point-gap topology from Maxwell Bloch states

In a one-dimensional photonic crystal with complex periodic permittivity $\varepsilon_{\mathrm{r}}(x) = \varepsilon_{\mathrm{r}}(x+a)$, the Bloch eigenproblem

$$-\left(\frac{\partial}{\partial x}+ik\right)\frac{1}{\varepsilon_{\mathrm{r}}(x)}\left(\frac{\partial}{\partial x}+ik\right)\psi_k(x) = \frac{\omega(k)^2}{c^2}\,\psi_k(x) \tag{13.1}$$

yields complex eigenfrequencies $\omega(k)$ for each Bloch wave vector $k \in [-\pi/a, \pi/a]$. As k sweeps the Brillouin zone, $\omega(k)$ traces a closed curve in the complex plane (closed because $\omega(-\pi/a) = \omega(\pi/a)$ by periodicity).

If this curve does not pass through a reference point E_{ref}—typically chosen as zero or the center of the spectral loop—the spectral winding number (12.7) is well defined and integer-valued [85, 86, 88]. A nonzero winding number indicates that the complex photonic band has a topologically nontrivial structure that cannot be deformed to a point without closing the point gap.

13.1.1 Maxwell-level origin

The point-gap topology of the photonic band is not an artifact of a tight-binding approximation—it is encoded directly in the Maxwell operator. In the full Maxwell framework, the complex eigenfrequencies arise because the operator $\hat{\Theta}_k = -(d/dx + ik)\varepsilon_{\mathrm{r}}^{-1}(d/dx + ik)$ is non-self-adjoint

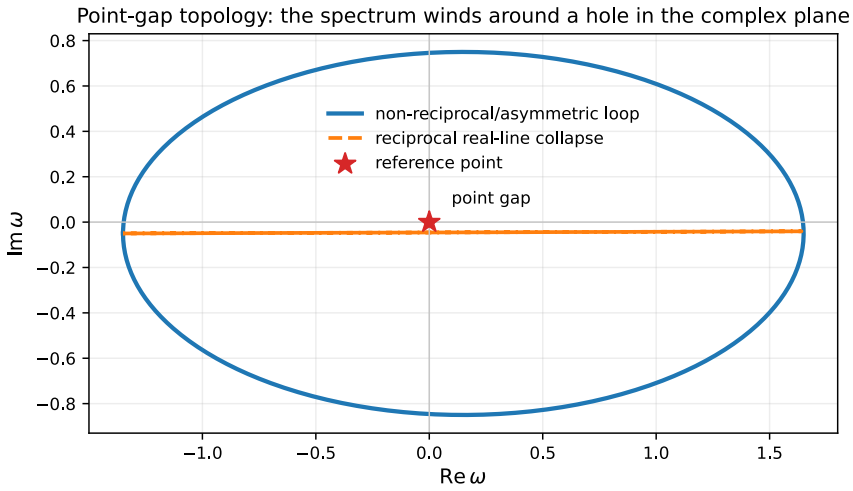


Figure 13.1 A point gap is a hole in the complex spectrum, not a forbidden interval on the real axis. The asymmetric non-Hermitian loop winds around the reference point and cannot be contracted without crossing it; the reciprocal comparison curve collapses toward a non-winding real-line spectrum. The model is pedagogical, but the topology is the same integer winding used for complex photonic bands.

when $\operatorname{Im}\varepsilon_r \neq 0$. The winding of the complex bands is a spectral property of this continuous operator, not of a discretized model.

The key Maxwell-level insight is that the *shape* of the spectral loop in the complex plane is determined by the spatial overlap between the Bloch mode profile and the gain/loss distribution. Modes with different k sample the gain and loss regions differently, producing different imaginary frequency shifts (Section 12.1.1). The resulting k-dependent imaginary part creates the loop that winds in the complex plane.

13.1.2 Contrast with line-gap topology

It is instructive to contrast point-gap and line-gap topology. A *line gap* is a line in the complex plane (typically the real or imaginary axis) that the spectrum does not cross. Line-gap topology generalizes the Hermitian band-gap concept and supports the same classification (Chern numbers, $\mathbb{Z}_2$ invariants) as Hermitian systems. A *point gap* is a single point that the spectrum does not touch; the associated topology (the winding number) is purely non-Hermitian and has no Hermitian counterpart.

For photonic systems, line-gap topology arises when the complex bands are separated by a forbidden strip in the complex frequency plane—for example, when gain and loss are weak enough that all bands remain close to the real axis. Point-gap topology arises when the bands form loops that encircle a reference energy—requiring stronger non-Hermiticity that pushes the bands into genuinely two-dimensional curves in the complex plane.

13.2 Reciprocity constraints

Photonic systems are typically reciprocal (Section 2.6): the permittivity tensor is symmetric, and the scattering matrix satisfies $\mathbf{S}^{\mathsf{T}} = \mathbf{S}$. Reciprocity imposes a constraint on the complex band

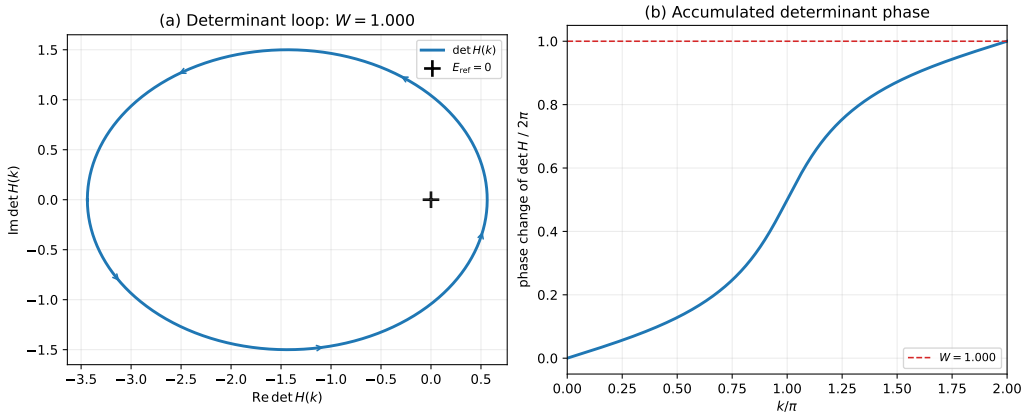


Figure 13.2 Determinant winding for an asymmetric non-Hermitian SSH model ($t_L = 1.75$, $t_R = 0.25$, $t_2 = 1.0$). (a) The loop traced by $\det H(k)$ winds once around the point-gap reference value $E_{\text{ref}} = 0$. (b) The unwrapped phase of $\det H(k)$ changes by 2π, giving $W = 1.000000$. Computing the point-gap invariant from the determinant avoids branch ambiguities in square-root energy bands and matches the Maxwell-discretized form of the invariant.

structure: $\omega(k) = \omega(-k)$ for scalar ε_{r}, because the Bloch Hamiltonian at k and $-k$ is related by transposition.

This constraint means the spectral loop $\omega(k)$ is symmetric under $k \to -k$. In one dimension, a reciprocal single-band spectral loop necessarily retraces itself and carries zero winding, while a multi-band determinant loop also retraces, forcing the total winding number to vanish. Breaking reciprocity (e.g., with a magneto-optical component) lifts this constraint and allows nonzero winding numbers [97, 98, 99].

13.2.1 Reciprocity and vanishing point-gap winding

To see why reciprocity forces zero winding, write the spectral loop as $\omega(k)$ for $k \in [-\pi/a, \pi/a]$. Under $k \to -k$, $\omega(-k) = \omega(k)$ (reciprocity). The winding number

$$W = \frac{1}{2\pi i} \int_{-\pi/a}^{\pi/a} \frac{\omega'(k)}{\omega(k) - E_{\text{ref}}} \, \mathrm{d}k \tag{13.2}$$

can be split into contributions I_+ (from $k \in [0, \pi/a]$) and I_- (from $k \in [-\pi/a, 0]$). Substituting $k \to -k$ in the second integral, using $\omega(-k) = \omega(k)$ and $\omega'(-k) = -\omega'(k)$, gives

$$I_- = \frac{1}{2\pi i} \int_{0}^{\pi/a} \frac{-\omega'(k)}{\omega(k) - E_{\text{ref}}} \, \mathrm{d}k = -\, I_+. \tag{13.3}$$

Therefore $W = I_+ + I_- = 0$: the two half-Brillouin-zone contributions cancel exactly.

In multi-band systems, the determinant $\det[H(k) - E_{\text{ref}}]$ also satisfies $k \to -k$ symmetry and therefore retraces, giving zero total winding. Individual eigenvalue branches can still carry nonzero winding when they exchange partners at symmetry-related k-points; such individual windings are constrained to appear in cancelling pairs, so the sum vanishes. Breaking reciprocity—for example, with a magneto-optical component—lifts this constraint and allows arbitrary integer winding numbers [97, 98, 99].

Why breaking reciprocity matters

All one-dimensional photonic examples with nonzero point-gap winding $W \neq 0$ require broken reciprocity. The Hatano–Nelson chain (Chapter 14), the asymmetric SSH model, and magneto-optical waveguide lattices all achieve $W \neq 0$ precisely because their Bloch Hamiltonians do not satisfy $H(-k)^{\mathrm{T}} = H(k)$. In reciprocal systems, nontrivial topology protected by the transpose symmetry can still exist, but it is classified by symmetry-protected invariants (such as a $\mathbb{Z}_2$ index or a Pfaffian invariant) rather than by the integer winding number W of the spectral loop. See the 38-fold classification in Chapter 12 for the full symmetry-resolved picture.

13.2.2 Physical implications

The vanishing of the ordinary reciprocal winding has concrete physical consequences:

- A reciprocal scalar 1D photonic crystal may have complex bands, exceptional points, or symmetry-protected non-Hermitian topology, but its ordinary determinant point-gap winding cancels between k and $-k$.
- A nonzero integer winding in this 1D Maxwell setting normally requires a mechanism that makes forward and backward propagation inequivalent, such as magneto-optical bias, temporal modulation, or an engineered nonreciprocal reservoir.
- Breaking reciprocity (e.g., with a magneto-optical Faraday rotation element) allows odd winding numbers and the associated boundary spectral flow.

Reciprocity and winding

In tight-binding models, asymmetric hopping $t_L \neq t_R$ is a common source of nonzero winding and the skin effect. In photonic crystals, such asymmetry requires breaking reciprocity—a more demanding experimental condition. Reciprocal photonic systems can still host symmetry-protected non-Hermitian phases, but those phases should not be identified with the ordinary integer winding of a single spectral loop. The reciprocity constraint is a Maxwell-level selection rule that is easy to miss in generic tight-binding models.

13.3 Example: complex bands of a reciprocal bilayer

To make the abstract winding-number theory concrete, consider a one-dimensional photonic crystal with period a consisting of two layers per unit cell:

- Layer A: thickness $d_A = 0.5a$, index $n_A = 2.0 + 0.3i$ (lossy).
- Layer B: thickness $d_B = 0.5a$, index $n_B = 2.0 - 0.1i$ (gain, weaker than loss).

The gain and loss are *not* balanced ($\mathcal{PT}$ symmetry is broken), and the system is reciprocal. This example is useful because it shows how complex photonic bands arise from Maxwell transfer matrices, but it should not be mistaken for a nonzero point-gap winding phase: reciprocity will make the ordinary Bloch loop retrace itself when the whole Brillouin zone is followed.

13.3.1 Transfer-matrix computation

The transfer matrix for one unit cell follows from the formalism of Chapter 3. Each layer contributes a propagation matrix

$$M_j = \begin{pmatrix} \cos\phi_j & -i\sin\phi_j/n_j \\ -i\,n_j\sin\phi_j & \cos\phi_j \end{pmatrix}, \qquad \phi_j = \frac{\omega\, n_j\, d_j}{c}, \tag{13.4}$$

and the full unit-cell matrix is $M = M_B\, M_A$. Because n_j is now complex, the phase ϕ_j has both real and imaginary parts: $\phi_j = (\omega d_j/c)(\mathrm{Re}\, n_j + i\ \mathrm{Im}\, n_j)$. The imaginary part causes exponential growth or decay within each layer, which is the microscopic origin of the complex band structure.

The Bloch condition $\cos(Ka) = \xi(\omega) \equiv \frac{1}{2}\,\mathrm{tr}\, M$ generalises immediately: the half-trace ξ is now complex, and the Bloch wave vector K is generically complex. For a fixed real k, the eigenfrequencies $\omega(k)$ are the solutions of

$$\xi(\omega) = \cos(ka), \tag{13.5}$$

which must be found numerically because the right-hand side is real while $\xi(\omega)$ is an analytic function of ω in the complex plane.

Numerical procedure. For each $k \in [-\pi/a, \pi/a]$, we solve (13.5) by Newton iteration in the complex ω plane, initialising from the Hermitian band structure ($\mathrm{Im}\, n_j = 0$) and tracking the root as $\mathrm{Im}\, n_j$ is adiabatically increased. With the parameters above ($n_A = 2.0 + 0.3i$, $n_B = 2.0 - 0.1i$, $d_A = d_B = 0.5a$), the first band traces an elliptical loop in the complex $\omega a/(2\pi c)$ plane with centroid near $\mathrm{Re}\,\omega a/(2\pi c) \approx 0.22$ and an imaginary spread of approximately ± 0.012. The loop is tilted: the maximum $\mathrm{Im}\,\omega$ occurs near $k = 0.3\pi/a$, not at the zone boundary, because the mode profile at that wave vector has the largest spatial overlap with the loss region.

The second band traces a separate loop at higher frequency. The two loops do not intersect, confirming that the point gap between them is well defined.

13.3.2 Complex band structure

Solving the Bloch condition numerically for $k \in [-\pi/a, \pi/a]$ yields complex eigenfrequencies $\omega(k)$. Over half of the Brillouin zone the band can appear to form a non-circular arc in the complex ω plane because layer A is lossier than layer B's gain. Over the full Brillouin zone, however, the reciprocal partner at $-k$ retraces the same spectral path. Thus the total determinant winding of this reciprocal bilayer is zero, even though its bands are complex.

13.3.3 Computing the winding number

With the complex band $\omega(k)$ in hand, the spectral winding number around a reference point E_{ref} is evaluated by discretising the Brillouin zone into N_k points $k_j = -\pi/a + 2\pi j/(N_k a)$ and summing the phase increments:

$$W = \frac{1}{2\pi} \sum_{j=0}^{N_k-1} \arg\left[\frac{\omega(k_{j+1}) - E_{\mathrm{ref}}}{\omega(k_j) - E_{\mathrm{ref}}}\right], \tag{13.6}$$

where arg returns a value in $(-\pi, \pi]$ and $k_{N_k} = k_0$ by periodicity. This formula is the discrete analogue of (13.2); it converges rapidly because the spectral loop is smooth.

For the reciprocal bilayer, applying Eq. (13.6) to a single continuous band over the full Brillouin zone gives $W = 0$ for any reference point not crossed by the spectrum. If one artificially follows only half of the zone, or switches branches at a degeneracy, an apparent winding can be obtained, but it is not the integer point-gap invariant of the reciprocal Maxwell problem. A genuine nonzero value requires replacing the bilayer by a nonreciprocal unit cell so that the k and $-k$ paths no longer cancel.

Computational cost

The bottleneck is solving (13.5) at each k_j. For a 1D bilayer, this requires root-finding for a transcendental equation—fast enough that the entire winding-number computation takes milliseconds. For a 2D photonic crystal, the eigenproblem must be solved by full-wave methods (Chapter 17), making the computation substantially more expensive but still feasible with modern eigensolvers.

13.3.4 Sensitivity to gain–loss contrast

As the imaginary parts of n_A and n_B are reduced toward zero, the spectral loop contracts toward the real axis and eventually collapses to a line segment (the Hermitian band structure). In the reciprocal bilayer this contraction changes the shape of the complex band but not the total point-gap winding, which remains zero. In a nonreciprocal variant, by contrast, a winding transition would occur only when the spectral path crosses the reference point and the point gap closes.

This point-gap closing is the non-Hermitian analogue of a Hermitian band gap closing: the topological invariant can change only when the reference point is touched by the spectrum.

Quantitatively, define a gain–loss parameter $\gamma = \max(|\operatorname{Im} n_A|, |\operatorname{Im} n_B|)$ and fix $\operatorname{Im} n_A / \operatorname{Im} n_B = -3$ (the ratio in our example). The critical value γ_c for a nonreciprocal winding transition depends on the band gap of the underlying Hermitian structure: a larger Hermitian gap requires a larger γ_c to open a spectral loop wide enough to wind. The reciprocal bilayer therefore serves as a useful warning: large gain–loss contrast can create broad complex bands, but complex bands alone do not imply a nonzero integer point-gap invariant.

13.3.5 Relation to the Green's function

The point-gap winding number has an elegant reformulation in terms of the retarded Green's function $G(k, \omega) = [\omega^2/c^2 - \hat{\Theta}_k]^{-1}$. At fixed reference frequency E_{ref}, the winding can be computed without tracking individual eigenvalues by following the phase of the inverse Green's-function determinant across the Brillouin zone:

$$W(E_{\mathrm{ref}}) = \frac{1}{2\pi i} \int_{\mathrm{BZ}} \partial_k \ln \det\left[E_{\mathrm{ref}}^2/c^2 - \hat{\Theta}_k\right] \, \mathrm{d}k, \tag{13.7}$$

provided the determinant never vanishes along the path. This connects the point-gap topology directly to the analytic structure of the photonic Green's function, which is the quantity measured in scattering and emission experiments (Section 13.6). In practice, the resolvent formulation

is useful for full-wave numerical computations where the eigenfrequencies are not individually tracked but the driven Maxwell problem can be solved at the reference frequency for many values of k.

13.4 From winding number to boundary response

The bulk-boundary correspondence for point-gap topology is subtler than the Hermitian edge-state counting rule. A nonzero spectral winding of a periodic one-dimensional system signals boundary-sensitive spectra: under open boundary conditions the bulk eigenstates are generally described by a generalized Brillouin zone and can accumulate near a boundary through the non-Hermitian skin effect. The robust statement is therefore not a simple counting formula such as $N_{\text{edge}} = |W|$. Rather, the integer winding counts spectral flow under a boundary twist and diagnoses the breakdown of the conventional Bloch description. This boundary response is qualitatively different from Chern edge modes (Section 11.7):

- It can occur already in *one-dimensional* systems, whereas Chern edge modes require at least two dimensions.
- It affects extensive bulk states as well as possible isolated defect or boundary resonances, reflecting the non-Hermitian skin effect (Chapter 14).
- It depends on the *point gap*, not a line gap or a real-frequency band gap.

As Yao and Wang showed, the conventional Bloch-Hamiltonian winding can give the wrong open-boundary prediction; the correct description uses the *generalized Brillouin zone* (Chapter 14). The detailed mechanism by which the GBZ restores the bulk–boundary correspondence is developed in Chapter 15.

13.4.1 Spectral flow and parameter-driven transitions

The connection between winding number and boundary response can be understood through spectral flow as a threading parameter is varied. Consider a family of finite photonic systems parameterised by a twist angle $\theta \in [0, 2\pi)$ applied as a boundary phase: $\psi(x + La) = e^{i\theta}\psi(x)$. As θ sweeps from 0 to 2π, the finite-system eigenfrequencies trace paths in the complex plane. The net number of eigenvalue trajectories that wind around the reference point equals the point-gap winding W—the non-Hermitian analogue of the Laughlin spectral-flow argument.

This spectral flow provides an experimentally accessible probe of point-gap topology: by threading a synthetic flux through a ring of coupled resonators (equivalent to varying θ), the movement of resonance frequencies inside the point gap reveals the topological invariant without computing the Bloch band structure.

13.4.2 Boundary response of the bilayer

For the reciprocal bilayer of Section 13.3, a finite stack of L unit cells with open (air) boundary conditions may support surface resonances caused by impedance mismatch at the termination, but these are not guaranteed by a nonzero point-gap winding because the ordinary reciprocal winding vanishes. A nonreciprocal modification of the same transfer-matrix model can instead

produce a genuine skin response. In that case, the bulk-mode decay length is set by the GBZ radius deviation from unity:

$$\xi_{\text{skin}} \sim \frac{a}{|\ln r_{\text{GBZ}}|}, \tag{13.8}$$

where r_{GBZ} is evaluated at the relevant complex frequency. The distinction is important experimentally: observing a localized resonance at a boundary is not by itself evidence for point-gap topology. One must verify the boundary-sensitive spectral flow or the non-Bloch displacement of the bulk spectrum.

13.5 Two-dimensional point-gap topology

In two dimensions, point-gapped spectra have a new diagnostic absent in one dimension: the periodic-boundary spectrum can cover a finite area in the complex plane. The spectral-area result of Zhang, Yang, and Fang [100] shows that this finite-area spectrum is closely tied to geometry-dependent skin accumulation:

$$\text{Spectral area} \neq 0 \quad \Longrightarrow \quad \text{boundary-sensitive spectra in generic geometries.} \tag{13.9}$$

The precise boundary response can depend on sample shape and symmetry, so the spectral area should be read as a powerful diagnostic rather than as a quantized invariant. It nevertheless explains why two-dimensional non-Hermitian systems, including photonic crystal models with complex ε_{r}, can show strong sensitivity to boundary conditions.

13.5.1 Spectral area as a diagnostic

For a single complex band $\omega(\mathbf{k}) = x(\mathbf{k}) + iy(\mathbf{k})$, a useful oriented area with multiplicity is obtained from the Jacobian of the map $\mathbf{k} \mapsto (x, y)$:

$$\mathcal{A} = \int_{\text{BZ}} \left(\frac{\partial x}{\partial k_x}\frac{\partial y}{\partial k_y} - \frac{\partial x}{\partial k_y}\frac{\partial y}{\partial k_x} \right) \mathrm{d}^2\mathbf{k} = \int_{\text{BZ}} J(\mathbf{k})\ \mathrm{d}^2\mathbf{k}, \tag{13.10}$$

where $J(\mathbf{k}) \equiv \partial(x, y)/\partial(k_x, k_y)$ is the Jacobian of the map $\mathbf{k} \mapsto \omega(\mathbf{k})$ from the 2D Brillouin zone to the complex energy plane. This Jacobian measures how much oriented spectral area the band map creates locally. If J has a fixed sign, the map is orientation-preserving and the projected spectrum covers a finite area; if J changes sign, different regions of the BZ can cancel in the oriented integral. The physical, unsigned area of the spectral image is not generally quantized and must be interpreted with this multiplicity caveat.

Equivalently,

$$J(\mathbf{k}) = 2\ \mathrm{Im}\left[\frac{\partial \omega_n}{\partial k_x}\frac{\partial \bar{\omega}_n}{\partial k_y} \right] = 2\left(\frac{\partial\, \mathrm{Re}\ \omega_n}{\partial k_x}\frac{\partial\, \mathrm{Im}\ \omega_n}{\partial k_y} - \frac{\partial\, \mathrm{Re}\ \omega_n}{\partial k_y}\frac{\partial\, \mathrm{Im}\ \omega_n}{\partial k_x} \right). \tag{13.11}$$

This expression shows that spectral area requires the real and imaginary parts of the eigenfrequency to vary independently across the BZ—a hallmark of genuine non-Hermiticity, as opposed to a uniform global loss that merely shifts all eigenvalues by a constant imaginary offset.

Example: filled spectral disk. For a model whose two-dimensional Brillouin zone maps once onto a disk of radius R centred at ω_0 in the complex plane, the unsigned spectral area is

$$\mathcal{A} = \pi R^2. \tag{13.12}$$

If the map covers the same disk several times with the same orientation, the oriented area counts the multiplicity. Figure 13.3 illustrates the corresponding spectral image for a 2D lattice model.

This area is not a quantized topological invariant. A nonzero spectral area instead serves as a practical diagnostic, and in generic settings a strong indicator, of non-Hermitian boundary sensitivity in two dimensions.

13.5.2 Implications for photonic crystals

For a two-dimensional photonic crystal with complex $\varepsilon_{\mathrm{r}}(\mathbf{r})$, the PBC band structure can cover a finite area in the complex ω plane. The practical consequences are far-reaching. First, the PBC band structure is *not* a reliable guide to the physics of a finite sample: the eigenfrequencies of a finite photonic crystal slab can differ qualitatively from the PBC bands, with extensive piling up of modes near selected boundaries. Second, conventional edge or corner resonances must be distinguished from non-Hermitian boundary accumulation using non-Bloch theory rather than ordinary Bloch invariants. Third, the spectral area itself can be computed efficiently from the PBC band structure using (13.10) on a discrete $\mathbf{k}$-grid, providing a practical diagnostic for boundary sensitivity in a given 2D photonic crystal design.

Example: lossy square-lattice rods. Consider a square array of dielectric rods ($\varepsilon = 12 + 0.4i$, radius $r = 0.2a$) in air. For TM polarization, the first band at $\mathbf{k} = (k_x, k_y)$ sweeps a closed surface in the three-dimensional space $(\mathrm{Re}\,\omega, \mathrm{Im}\,\omega, k_x, k_y)$. Projected onto the complex ω plane, the band covers an area $\mathcal{A} \sim 0.003\,(2\pi c/a)^2$. Although this area is small, it is nonzero, so the finite rod array should be checked for boundary-sensitive spectra and possible skin accumulation. The direction and strength of the accumulation depend on the detailed termination and on the effective imaginary gauge field.

In practice, the skin effect in 2D photonic crystals is weaker and harder to observe than in 1D because the spectral area scales with the imaginary part of ε and the group velocity, both of which are modest for typical photonic-crystal parameters. Corner modes associated with second-order point-gap topology have been predicted theoretically but remain experimentally elusive.

Higher-order point-gap topology. In two dimensions, the point-gap classification admits a $\mathbb{Z}$-valued second-order invariant that predicts corner-localised modes in systems with spatial symmetries. For a photonic crystal with C_4 symmetry and complex ε_{r}, the corner invariant is the symmetry indicator or nested non-Bloch winding defined in the appropriate generalized Brillouin zone. This higher-order topology is the non-Hermitian analogue of the quadrupole topological insulator, but its photonic realization requires more than gain/loss: the relevant spatial symmetry and the non-Bloch boundary spectrum must both be verified.

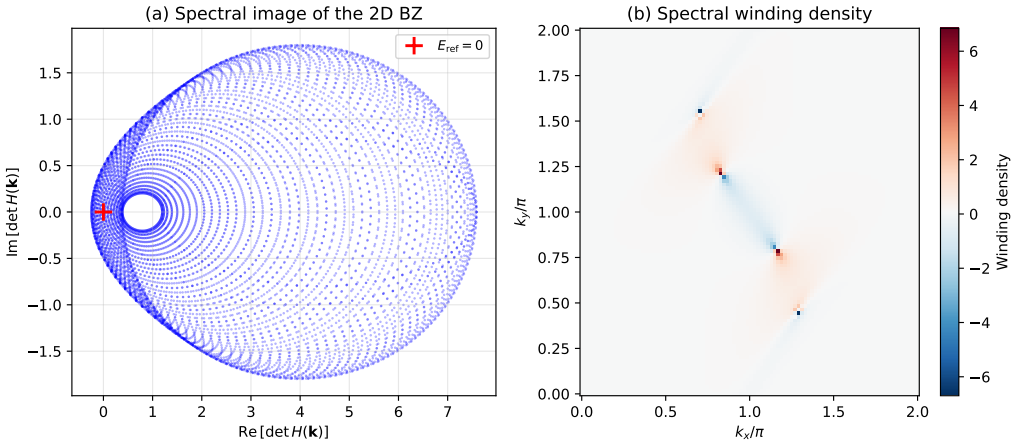


Figure 13.3 Spectral image of the 2D BZ for a non-Hermitian lattice model ($t_L = 1.5$, $t_R = 0.5$, $t_2 = 1.0$, $t_3 = 0.8$). (a) The function $\det H(\mathbf{k})$ maps the 2D Brillouin zone into the complex plane (blue dots); the reference point $E_{\text{ref}} = 0$ (red cross) lies inside the image. This signals boundary sensitivity but is not by itself an integer edge-state count. (b) Jacobian density across the BZ, showing where the map locally creates positive or negative oriented area. The total spectral area $A \approx 4.7$ is a diagnostic of non-Hermitian boundary response.

13.6 Photonic realizations and experiments

Point-gap topology and the related skin response have been proposed and observed in wave platforms including photonic analogues [95, 101]:

Coupled-resonator arrays. Chains of microring resonators with asymmetric coupling (via auxiliary gain/loss elements) realize the non-Hermitian SSH model with tunable winding numbers. The asymmetry can be introduced by gain/loss-induced effective non-reciprocal coupling or by auxiliary waveguide paths that create direction-dependent hopping. The winding number has been inferred from the spectral sensitivity to boundary conditions and from the direct observation of skin-mode localization.

Photonic lattices with on-site gain/loss. Arrays of waveguides with alternating gain and loss can exhibit point-gap topology when the gain/loss pattern is combined with a mechanism that breaks the cancellation between forward and backward propagation. In these systems, the winding number can be tuned by varying the pump intensity, providing a continuous knob for controlling the topological phase.

Non-reciprocal photonic crystals. Magneto-optical photonic crystals provide a natural platform for odd winding numbers, connecting point-gap topology to the anomalous photonic Hall effect. These systems require an external magnetic field but offer the largest topological gaps and the strongest skin-effect localization.

Photonic-crystal point-gap observation. Zhong et al. [90] proposed and analyzed a concrete photonic-crystal model with non-Hermitian point-gap topology, computing the spectral winding number from the complex photonic band structure and demonstrating its connection to boundary-sensitive spectra and the skin effect. Their analysis shows that the point-gap invariant can be

extracted from full-wave electromagnetic simulations, establishing a direct link between the abstract topological theory and realistic photonic structures.

13.7 Vector nature and polarization

In two and three dimensions, the vector nature of the electromagnetic field introduces additional structure. Each Bloch band carries a polarization (TE, TM, or mixed), and the point-gap winding can depend on the polarization sector. For a 2D photonic crystal slab:

- TE-like and TM-like bands can have different winding numbers, leading to polarization-dependent skin effects.
- Band crossings between TE and TM sectors can create exceptional points that connect the point-gap topology to the EP topology of Chapter 8.
- The full vectorial Maxwell calculation is essential for quantitative predictions; scalar approximations can give the wrong winding number.

This polarization dependence is a uniquely photonic feature with no analogue in scalar wave systems or electronic topological insulators.

Polarization-resolved winding. In a concrete 2D photonic crystal slab, the TE-like and TM-like bands generally have different spatial profiles: the TM modes concentrate their electric field inside the high-index rods, while the TE modes spread into the air region. When $\operatorname{Im}\varepsilon_{\mathrm{r}}$ is nonzero only inside the rods, the TM bands acquire larger imaginary frequency shifts and trace wider loops in the complex plane, potentially reaching a higher winding number than the TE bands. The resulting polarisation-dependent skin effect means that a finite photonic-crystal slab can localise TM light at one boundary while TE light remains extended—a polarisation filter driven entirely by non-Hermitian topology.

From scalar to vector: when does the winding change? A useful rule of thumb is that the scalar (single-polarisation) winding number agrees with the full vectorial result whenever the TE–TM coupling is negligible—i.e., when the photonic crystal has a mirror symmetry that decouples the polarisations. When this symmetry is broken (e.g., by a tilted slab, an anisotropic ε_{r} tensor, or a substrate), the TE and TM bands hybridise, and the winding number of the hybridised band can differ from the sum of the individual polarisation windings. In the extreme case of an exceptional-point degeneracy between TE and TM bands, the winding number changes by ± 1, connecting the point-gap topology directly to the EP classification of Chapter 8.

Point-gap topology is intrinsically non-Hermitian

Point-gap topology has no Hermitian analogue: a real-valued spectrum cannot wind around a point in the complex plane. In photonic systems, it arises from the interplay of complex ε_{r}, spatial periodicity, and (optionally broken) reciprocity. The Maxwell-level constraint of reciprocity cancels the ordinary one-dimensional determinant winding, so genuine integer point-gap winding in photonic systems usually requires nonreciprocity or a symmetry-resolved invariant. The physical manifestation of point-gap topology—boundary-sensitive

spectra and the non-Hermitian skin effect—is the subject of Chapter 14, and the rebuilt bulk-boundary correspondence that accounts for it is developed in Chapter 15.

Chapter 14

Skin Effect and Non-Bloch Light

In Hermitian photonic crystals, Bloch's theorem guarantees that eigenstates are extended plane waves modulated by a periodic function. In non-Hermitian crystals, this guarantee can fail: under open boundary conditions, an extensive fraction of the eigenstates can pile up at one edge of the system. This is the non-Hermitian skin effect *(NHSE), and it demands a complete rethinking of band theory. The wave vector acquires an imaginary part, the Brillouin zone deforms into a* generalized Brillouin zone *(GBZ), and topological invariants must be redefined on this new contour to restore the bulk-boundary correspondence. The theory of the skin effect is developed from the perspective of photonic systems, leading to the non-Bloch band theory that replaces conventional Bloch analysis.*

14.1 The breakdown of Bloch band theory

Consider a one-dimensional non-Hermitian photonic crystal of L unit cells with open boundary conditions. In a Hermitian system, the eigenstates of such a finite chain approach extended Bloch waves as $L \to \infty$, with eigenfrequencies tracing the bulk bands. In a non-Hermitian system, the picture changes.

14.1.1 Skin modes

Numerical diagonalization of a simple skin-effect Hamiltonian reveals that an extensive set of eigenstates—not just a few edge states, but the bulk spectrum itself—can become exponentially localized at one boundary of the chain. The localization length is finite and independent of L in the standard NHSE. This is the non-Hermitian skin effect: an $O(L)$ number of states are boundary-localized, in contrast to the $O(1)$ boundary states of Hermitian topological insulators.

14.1.2 Spectral sensitivity

The open-boundary spectrum can differ strongly from the periodic-boundary (PBC) spectrum. Under PBC, the eigenfrequencies form closed loops in the complex plane (the bands of Chapter 12). Under open boundary conditions (OBC), the spectrum collapses onto arcs or curves that bear little resemblance to the PBC loops. Conventional Bloch-Hamiltonian invariants, computed

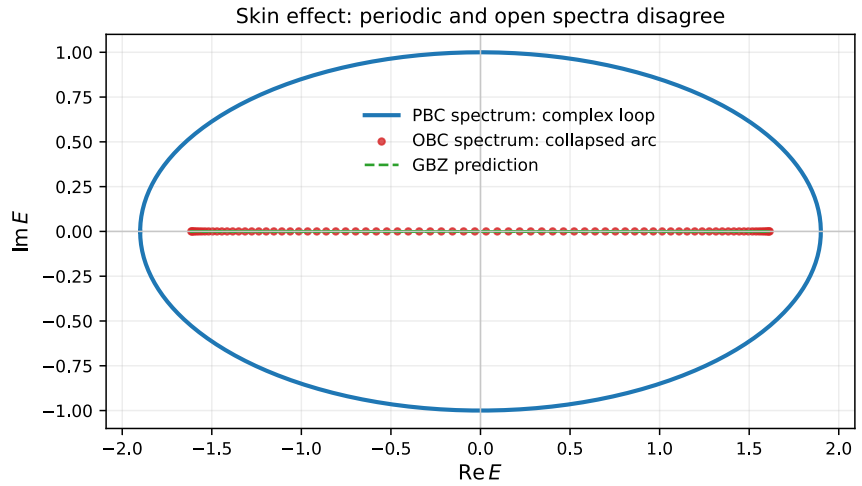


Figure 14.1 The non-Hermitian skin effect announces itself through spectral sensitivity to boundary conditions [102]. In the Hatano–Nelson toy model, the periodic-boundary spectrum forms a complex loop, while the open-boundary spectrum collapses onto the non-Bloch/GBZ prediction [85, 86]. This is the computational signature behind the failure of ordinary Bloch bulk-boundary correspondence.

from the PBC spectrum, fail to predict the OBC physics—including the presence or absence of topological zero modes.

14.1.3 A quantitative measure

The skin effect can be quantified by the inverse participation ratio (IPR) or by the mean center of mass of the eigenstates. For an L-site chain with eigenstate $|\psi_j\rangle$ and amplitudes $\psi_j(n)$ at site n, the center of mass is

$$\bar{n}_j = \sum_{n=1}^{L} n\,|\psi_j(n)|^2 \Big/ \sum_{n=1}^{L} |\psi_j(n)|^2. \tag{14.1}$$

In a Hermitian system, $\bar{n}_j \approx L/2$ for bulk states. In a standard skin-effect system, $\bar{n}_j \to 1$ (or $\bar{n}_j \to L$) for an extensive set of eigenstates, signaling macroscopic boundary accumulation.

14.2 Physical origin of the skin effect

The skin effect arises when non-Hermiticity produces a nonzero point-gap winding, most transparently through *asymmetric* coupling between sites. In a tight-binding picture, if the hopping amplitude from left to right (t_R) differs from right to left (t_L), a particle is preferentially transported in one direction, and eigenstates accumulate at the corresponding boundary. Asymmetric hopping is the cleanest route, but the more general diagnostic is the boundary-sensitive point-gap spectrum discussed in Chapter 13.

14.2.1 Sources of asymmetry in photonic systems

In photonic systems, this asymmetry can arise from:

- **Non-reciprocal elements:** magneto-optical materials or active components that break Lorentz reciprocity.

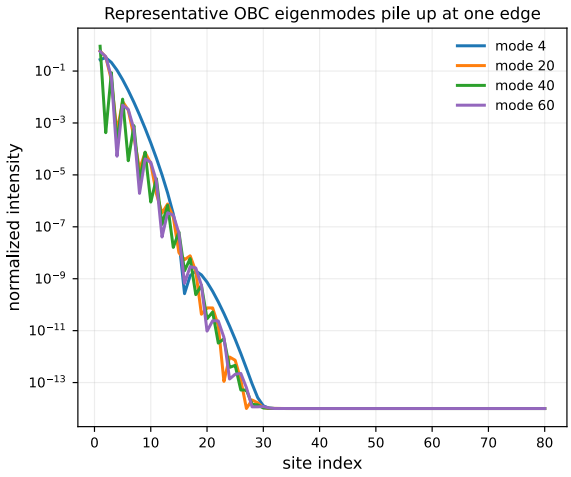


Figure 14.2 Skin modes are not a small number of edge states; in the standard NHSE an extensive fraction of the open-chain eigenvectors accumulates near one boundary. The plotted representative normalized eigenmode intensities decay exponentially away from the accumulation edge on a logarithmic scale.

- **Engineered gain/loss reservoirs:** auxiliary lossy or amplifying paths that produce direction-dependent effective coupling after unobserved channels are eliminated.
- **Coupled-resonator networks:** auxiliary coupling paths that make the reduced modal hopping asymmetric. A reciprocal full-wave device may still have a symmetric scattering matrix; the nonreciprocity here is an effective description of a selected subsystem.
- **Time modulation:** periodic temporal modulation of the refractive index that breaks time-reversal symmetry and creates effective non-reciprocal hopping in the Floquet picture.

14.2.2 The imaginary gauge field

The mathematical signature of the skin effect is that the eigenvalues $\beta = e^{ik}$ of the transfer matrix (or the Bloch phase factor) are no longer on the unit circle $|\beta| = 1$. Instead, they satisfy $|\beta| = r \neq 1$, corresponding to an exponentially growing or decaying spatial profile. The wave vector acquires an imaginary part: $k \to k - i\kappa_{\text{skin}}$, where $\kappa_{\text{skin}} = \ln r$ is the skin decay constant.

This can be understood as an *imaginary gauge field* [97, 98, 103, 104]: the asymmetric hopping is equivalent to a uniform imaginary vector potential. Under OBC the asymmetry can be removed from the Hatano–Nelson Hamiltonian by a non-unitary similarity transform, but the transformed back eigenvectors acquire an exponential envelope. Under PBC the same transform changes the boundary condition by a complex flux, so the asymmetry remains visible in the periodic spectrum as a complex loop. This gauge-theoretic perspective explains why the PBC and OBC spectra are so different.

14.3 Example: the Hatano–Nelson model

The Hatano–Nelson model is the simplest system that exhibits the skin effect. It is a one-dimensional tight-binding chain with asymmetric nearest-neighbor hopping:

$$H = \sum_{n=1}^{L-1} \big(t_R \, |n\rangle\langle n+1| + t_L \, |n+1\rangle\langle n| \big), \tag{14.2}$$

with $t_R \neq t_L$ and $t_R, t_L > 0$.

14.3.1 PBC spectrum

Under periodic boundary conditions, the Bloch Hamiltonian is

$$H(k) = t_R \, e^{ik} + t_L \, e^{-ik}, \tag{14.3}$$

and the eigenvalues trace an ellipse in the complex plane:

$$E_{\mathrm{PBC}}(k) = (t_R + t_L)\cos k + i(t_R - t_L)\sin k, \tag{14.4}$$

with semi-major axis $t_R + t_L$ and semi-minor axis $|t_R - t_L|$. The spectral winding number (Section 12.3) around any point inside the ellipse is $W = +1$ (for $t_R > t_L$).

14.3.2 OBC spectrum

Under open boundary conditions, the spectrum collapses onto the real interval $[-2\sqrt{t_R t_L},\ 2\sqrt{t_R t_L}]$. In this model the PBC ellipse collapses onto a line segment. All eigenvalues are real, and all right eigenstates localize at the left boundary (for $t_R > t_L$). The amplitude envelope is controlled by the GBZ radius,

$$|\psi_n| \propto r_{\mathrm{GBZ}}^{\,n}, \qquad r_{\mathrm{GBZ}} = \sqrt{\frac{t_L}{t_R}}, \tag{14.5}$$

so the amplitude localization length is $\xi_{\mathrm{amp}} = 1/|\ln r_{\mathrm{GBZ}}| = 2/\ln(t_R/t_L)$. If one fits the intensity $|\psi_n|^2$ instead, the corresponding intensity length is $\xi_{\mathrm{int}} = \xi_{\mathrm{amp}}/2 = 1/\ln(t_R/t_L)$. We will state explicitly which convention is being used whenever a localization length is quoted.

14.3.3 GBZ construction

The GBZ for the Hatano–Nelson model can be found analytically. Replace e^{ik} by β in (14.3): $H(\beta) = t_R\beta + t_L/\beta$. The characteristic equation $H(\beta) = E$ gives a quadratic in β:

$$t_R\beta^2 - E\beta + t_L = 0, \tag{14.6}$$

with two solutions β_1, β_2 satisfying $\beta_1\beta_2 = t_L/t_R$. The GBZ condition $|\beta_1| = |\beta_2|$ gives

$$|\beta|^2 = \frac{t_L}{t_R} \quad \Longrightarrow \quad |\beta| = \sqrt{\frac{t_L}{t_R}} \equiv r_{\mathrm{GBZ}}. \tag{14.7}$$

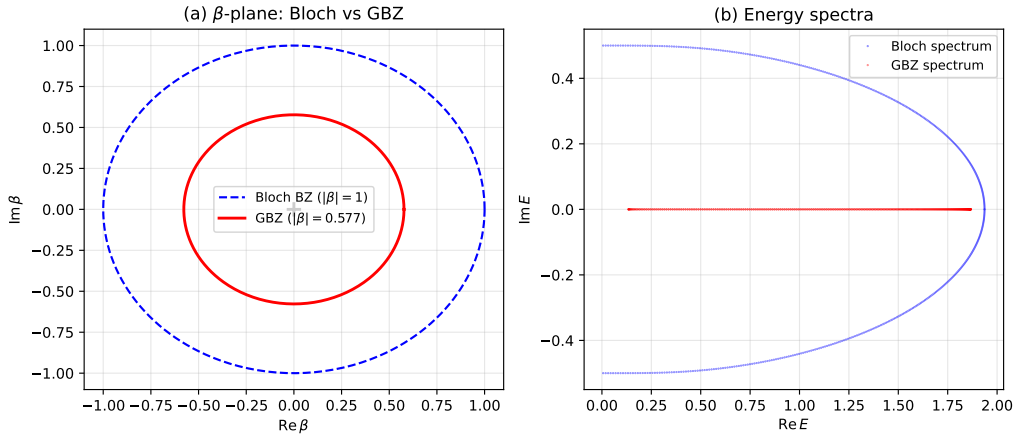


Figure 14.3 Generalized Brillouin zone (GBZ) for an asymmetric-hopping Hatano–Nelson model with $t_R/t_L = 3$. (a) In the β-plane, the Bloch BZ (blue dashed, $|\beta| = 1$) and the GBZ (red solid, $|\beta| = \sqrt{t_L/t_R} \approx 0.577$) are concentric circles. The GBZ radius is smaller than unity, reflecting the exponential skin-effect localisation of all bulk eigenstates toward one boundary. (b) The energy spectrum computed on the GBZ (red) differs qualitatively from the Bloch spectrum (blue), correctly reproducing the OBC spectrum and the non-Bloch phase boundary.

The GBZ is a circle of radius $r_{\mathrm{GBZ}} < 1$ (for $t_R > t_L$) centered at the origin in the complex β plane.

Evaluating the band structure on this circle gives

$$E_{\mathrm{GBZ}}(\theta) = t_R r_{\mathrm{GBZ}}\, e^{i\theta} + t_L \frac{e^{-i\theta}}{r_{\mathrm{GBZ}}} = 2\sqrt{t_R t_L}\,\cos\theta, \tag{14.8}$$

which is exactly the OBC spectrum. The non-Bloch band theory recovers the correct open-boundary spectrum from the deformed Brillouin zone.

The GBZ radius also fixes the skin envelope. For amplitudes,

$$\xi_{\mathrm{amp}} = \frac{1}{|\ln r_{\mathrm{GBZ}}|} = \frac{2}{\ln(t_R/t_L)}, \tag{14.9}$$

whereas an intensity fit gives $\xi_{\mathrm{int}} = \xi_{\mathrm{amp}}/2 = 1/\ln(t_R/t_L)$. This resolves the apparent factor-of-two ambiguity: the GBZ radius describes the field amplitude, while many numerical and experimental plots fit intensity.

Worked numerical example. Set $t_R = 1.5$ and $t_L = 0.5$, so that $t_R/t_L = 3$. The PBC spectrum (14.4) is an ellipse with semi-major axis $t_R + t_L = 2.0$ and semi-minor axis $|t_R - t_L| = 1.0$. The spectral winding number around the origin is $W = +1$, predicting the skin effect. The OBC spectrum collapses onto the real interval $[-2\sqrt{0.75},\ 2\sqrt{0.75}] \approx [-1.732,\ 1.732]$. The GBZ radius is $r_{\mathrm{GBZ}} = \sqrt{1/3} \approx 0.577$, giving a skin decay constant $\kappa_{\mathrm{skin}} = |\ln r_{\mathrm{GBZ}}| = \ln\sqrt{3} \approx 0.549$ per site. The amplitude localization length is therefore $\xi_{\mathrm{amp}} \approx 1.82$ sites, while an intensity fit gives $\xi_{\mathrm{int}} \approx 0.91$ sites. For a chain of $L = 40$ sites, numerical diagonalisation gives a mean eigenmode centre of mass $\bar{n} \approx 1.8$, far from the Hermitian bulk value of 20.5; the imaginary gauge field $A_{\mathrm{imag}} = i\kappa_{\mathrm{skin}} \approx 0.549i$ per lattice constant captures this accumulation exactly.

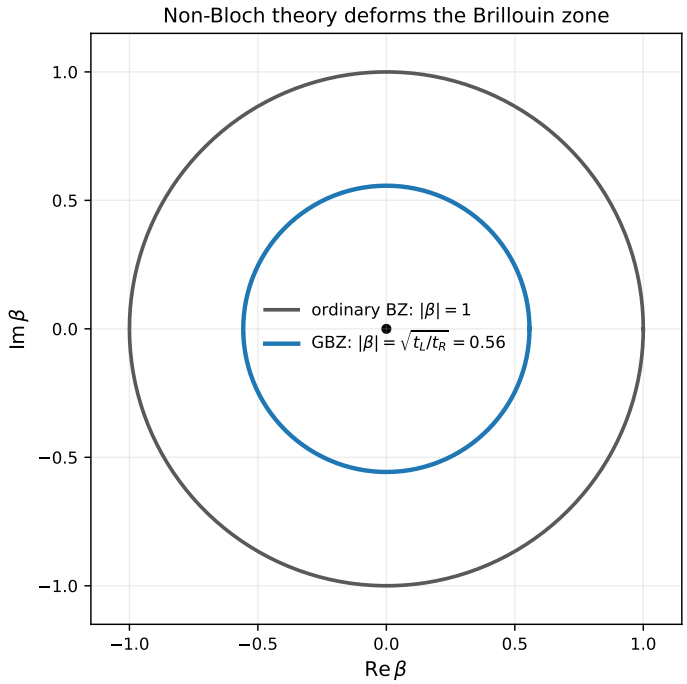


Figure 14.4 The generalized Brillouin zone deforms the ordinary unit circle in the complex $\beta = e^{ik}$ plane. For the Hatano–Nelson chain with asymmetric hopping, the GBZ is the circle $|\beta| = \sqrt{t_L/t_R}$, which encodes the exponential envelope of the open-boundary eigenstates.

This example is deliberately extreme ($t_R/t_L = 3$) to make the skin effect visible in a short chain. In photonic realisations the asymmetry is typically smaller ($t_R/t_L \lesssim 2$), but the localization is still measurable over chains of 10–20 resonators.

14.4 The generalized Brillouin zone: general theory

14.4.1 From Bloch to non-Bloch

In the conventional Bloch theory, the wave vector k is real and the Bloch phase $\beta = e^{ik}$ traces the unit circle $|\beta| = 1$ as k sweeps the BZ. The energy bands $\omega(k)$ are functions of this real k.

For a non-Hermitian system with the skin effect, the correct ansatz for the spatial dependence is $\psi_n \propto \beta^n$ with $|\beta| \neq 1$. The appropriate contour for β in the complex plane is the *generalized Brillouin zone* (GBZ), defined as the locus of β values that satisfy the bulk equation and the open-boundary condition simultaneously.

14.4.2 Construction of the GBZ

For a system with transfer matrix $\mathbf{T}(\omega, \beta)$, the characteristic equation

$$\det[\mathbf{T}(\omega, \beta) - \mathbf{I}] = 0 \tag{14.10}$$

yields $2n$ solutions $\beta_1, \beta_2, \ldots, \beta_{2n}$ (for an n-band model) at each complex frequency ω. Ordering them by modulus, $|\beta_1| \le |\beta_2| \le \cdots \le |\beta_{2n}|$, the GBZ is defined by the condition

$$|\beta_n| = |\beta_{n+1}|. \tag{14.11}$$

This condition ensures that the n-th and $(n+1)$-th solutions have equal decay rates, which is necessary for the boundary conditions at both ends of an open chain to be satisfied simultaneously.

14.4.3 Geometry of the GBZ

The GBZ is generically a closed curve in the complex β plane, but it need not be a circle. For the Hatano–Nelson model the GBZ is the circle $|\beta| = r_{\mathrm{GBZ}}$ (Section 14.3.3), but richer models produce richer contours. As the system approaches the Hermitian limit ($t_R \to t_L$), the GBZ smoothly deforms back to the unit circle, recovering Bloch theory.

NH SSH model: a two-band example. The non-Hermitian SSH model has two sublattices per unit cell, with the Bloch Hamiltonian

$$H(\beta) = \begin{pmatrix} 0 & t_1 + t_2/\beta \\ t_1' + t_2'\,\beta & 0 \end{pmatrix}, \tag{14.12}$$

where t_1, t_2 are intra- and inter-cell hoppings for the forward-going off-diagonal element and t_1', t_2' for the backward-going one. Non-Hermiticity enters whenever $t_1 \neq t_1'$ or $t_2 \neq t_2'$. The energy eigenvalues are

$$E^2(\beta) = (t_1 + t_2/\beta)(t_1' + t_2'\,\beta). \tag{14.13}$$

This is a Laurent polynomial in β; expanding gives a quartic equation in β at fixed E, yielding four roots $\beta_1, \ldots, \beta_4$ ordered by modulus. The GBZ condition $|\beta_2| = |\beta_3|$ (the middle pair for a two-band model) defines the GBZ contour.

For the symmetric case $t_1 = t_1' = t$ and $t_2 = t_2' \equiv s$ (the Hermitian SSH model), the four roots satisfy $|\beta_1| \le |\beta_2| = |\beta_3| \le |\beta_4|$ on the unit circle, recovering the standard BZ. But when $t_2 \neq t_2'$—for instance $t_2 = t_2' + \delta$ with δ parameterising the asymmetry—the GBZ deforms into a non-circular closed curve. Near the exceptional points of the band structure, the contour develops cusps where the two middle roots collide in both modulus and phase.

The physical content of the GBZ shape is transparent: the deviation of $|\beta|$ from unity at each point on the GBZ measures the local skin-mode decay rate. Regions of the GBZ where $|\beta| < 1$ correspond to modes that decay into the bulk from the left boundary; regions where $|\beta| > 1$ decay from the right boundary. The overall bias of the GBZ toward $|\beta| < 1$ or $|\beta| > 1$ determines the direction of the macroscopic skin effect.

Photonic transfer-matrix connection. In photonic systems, the GBZ formalism connects directly to the transfer-matrix method of Chapter 3, but an important caveat is needed. For an ordinary reciprocal one-dimensional stack of isotropic layers, the 2×2 transfer matrix written in the usual field variables is unimodular even when the refractive indices are complex. Gain and loss make the half-trace complex, so the Bloch frequencies and Bloch multipliers can be complex, but this alone does not produce the Hatano–Nelson skin effect.

The transfer-matrix eigenvalues obey

$$\beta_{1,2} = \xi \pm \sqrt{\xi^2 - 1}, \qquad \xi = \tfrac{1}{2}\,\mathrm{Tr}\,\mathbf{T}_{\mathrm{cell}}, \tag{14.14}$$

with $\beta_1\beta_2 = 1$ for a unimodular reciprocal cell. Inside a stop band one solution grows while the other decays, as in ordinary evanescent Bloch theory; this is not by itself an extensive skin effect. To obtain a photonic GBZ analogous to the Hatano–Nelson model one needs a unit cell whose effective evolution is non-unimodular or nonreciprocal after projecting to the modal subspace of interest, for example through magneto-optical bias, temporal modulation, cascaded gain/loss reservoirs, or a multi-port network with eliminated channels.

In such an effective model the GBZ radius is determined by the middle roots of the characteristic equation, just as in Eq. (14.11), and the skin localisation length is $\xi_{\mathrm{skin}} = 1/|\ln r_{\mathrm{GBZ}}|$. The physical origin is then a direction-dependent round-trip amplification that cannot be removed without changing the boundary condition.

Quantitative photonic example. Consider an effective coupled-resonator chain engineered to have $t_R/t_L = 1.8$ at $\lambda = 1550$ nm with lattice spacing $a = 400$ nm. The Hatano–Nelson estimate gives $r_{\mathrm{GBZ}} = \sqrt{t_L/t_R} \approx 0.745$ and an amplitude localisation length $\xi_{\mathrm{amp}} = 1/|\ln r_{\mathrm{GBZ}}| \approx 3.4$ unit cells (1.4 μm). In a finite chain of $L = 30$ resonators the right eigenmodes therefore show a clearly resolvable boundary envelope. The precise number in a full Maxwell simulation depends on radiation loss, coupling dispersion, and how the auxiliary channels are eliminated, but the logarithmic relation between r_{GBZ} and the field envelope is the same.

Energy-dependent GBZ. An important subtlety is that the GBZ depends on the energy E. At different complex energies, the condition $|\beta_n| = |\beta_{n+1}|$ is satisfied by different β values, tracing out an energy-dependent contour. The *physical* GBZ relevant for the OBC spectrum is the union of these contours over the set of OBC eigenvalues. In practice, one often computes the GBZ at a reference energy (e.g. $E = 0$ for chiral-symmetric models) and uses it to evaluate topological invariants, but this simplification can fail when the band structure has a complicated topology in the (β, E) space.

14.4.4 Non-Bloch band structure

The *non-Bloch band structure* is obtained by evaluating the eigenvalues $\omega(\beta)$ on the GBZ contour rather than on the unit circle:

$$\omega_n^{\mathrm{OBC}} = \omega_n(\beta)\big|_{\beta \in \mathrm{GBZ}}. \tag{14.15}$$

This non-Bloch band structure reproduces the OBC spectrum in the thermodynamic limit, resolving the discrepancy between PBC and OBC that signaled the breakdown of conventional Bloch theory.

14.5 Non-Bloch topological invariants

Topological invariants must be computed on the GBZ, not the conventional BZ, to correctly predict open-boundary spectra and any protected boundary modes of a non-Hermitian system.

14.5.1 Non-Bloch winding number

For a one-dimensional system with chiral symmetry, the non-Bloch winding number is

$$W_{\mathrm{GBZ}} = \frac{1}{2\pi i} \oint_{\mathrm{GBZ}} \frac{\mathrm{d}}{\mathrm{d}\beta} \ln \det[H(\beta)] \, \mathrm{d}\beta, \tag{14.16}$$

where $H(\beta)$ is the Bloch Hamiltonian with e^{ik} replaced by β. This invariant correctly predicts the number of topological zero modes under OBC, in contrast to the conventional winding number computed on the unit circle.

14.5.2 Non-Bloch Chern number

In two dimensions, the non-Bloch Chern number is defined by integrating the Berry curvature over the GBZ torus (the product of two GBZ contours for β_x and β_y). When the non-Bloch torus is well defined, this invariant restores the bulk-boundary correspondence for two-dimensional non-Hermitian systems with a line gap, predicting the correct number of chiral edge modes.

14.5.3 When GBZ invariants differ from Bloch invariants

The GBZ and Bloch invariants can differ whenever the GBZ is not the unit circle—i.e., whenever the skin effect is present. The discrepancy is maximized when:

- The asymmetry $|t_R/t_L|$ is large (strong skin effect).
- The system is near a phase boundary where the Bloch invariant changes but the GBZ invariant does not, or vice versa.
- The GBZ geometry has cusps or self-intersections that cross branch cuts of the Bloch Hamiltonian.

In all cases, the GBZ invariant is the quantity that should be compared with open-boundary spectra and boundary modes.

Non-Bloch band theory

In non-Hermitian photonic systems with the skin effect, the conventional Bloch band structure and Bloch topological invariants give incorrect predictions for open-boundary physics. The generalized Brillouin zone and the non-Bloch invariants computed on it restore the bulk-boundary correspondence and provide a faithful description of the photonic spectrum and boundary modes.

14.6 Skin effect in two and higher dimensions

The skin effect in higher dimensions is qualitatively richer than in one dimension [100].

Universality. In two and higher dimensions, the spectral-area result (Section 13.5) shows why skin accumulation is generic in many non-Hermitian models: a PBC spectrum that covers finite area in the complex plane is highly sensitive to boundary geometry. The precise skin pattern can depend on sample shape, symmetry, and termination, so the finite spectral area should be

viewed as a diagnostic of boundary sensitivity rather than as a simple one-dimensional winding number.

Corner skin effect. In certain two-dimensional systems, all eigenstates are localised at *corners* rather than edges. This corner-skin effect is a higher-order analogue of the one-dimensional skin effect and is classified by a second-order topological invariant. Physically, it arises when the spectral winding numbers in the x and y directions are both nonzero but with the same sign, so that the combined localisation direction points toward a corner.

Geometry-dependent skin effect. Some systems exhibit the skin effect for generic polygon-shaped samples but not for samples with specific symmetry-related shapes. This geometry dependence has no analogue in Hermitian physics, where the spectrum is insensitive to the sample shape in the thermodynamic limit. The physical origin is that the skin localisation direction is determined by the spectral winding, which depends on the boundary orientation; different facets of a polygon can have different winding numbers, leading to competing localisation directions.

Scale-free skin effect. A recently discovered variant, the scale-free skin effect, features localisation lengths that grow with the system size L rather than remaining fixed. In these systems, the spectrum is sensitive to boundary conditions at *every* scale, not just in the thermodynamic limit. The GBZ approaches the unit circle algebraically rather than exponentially as $L \to \infty$, leading to a non-extensive but still macroscopic accumulation of states at the boundary.

14.7 Photonic skin effect

The skin effect has been proposed and observed in several photonic platforms:

Coupled-resonator optical waveguides. Arrays of microring resonators with engineered asymmetric coupling realise one-dimensional non-Hermitian lattices with tunable skin effect. The asymmetry is introduced by auxiliary gain/loss elements or by directional couplers. The key experimental signature is that light injected at any point in the array propagates preferentially toward one end, regardless of the injection site.

Photonic mesh lattices: a concrete platform. Fibre-loop networks provide a versatile time-multiplexed platform for studying the skin effect. Two fibre loops of slightly different lengths create a discrete time-mesh lattice that is isomorphic to a 1+1D waveguide array. Pulses circulate through the loops, accumulating nonlinear phase from erbium-doped fibre amplifiers (EDFAs) that provide controlled gain, while attenuators provide controlled loss. The discrete time steps map to spatial sites, and the gain/loss can be tuned independently at each effective site.

Concretely, a mesh lattice with $N = 20$ time slots and asymmetric coupling ratio $|t_R/t_L| \approx 1.8$ is expected to show a skin localization length $\xi \approx 1/\ln(1.8) \approx 1.7$ sites. The experimental signature is a measurement of the output pulse intensity profile after many round trips: regardless of which time slot receives the initial pulse, the steady-state intensity peaks at the same boundary

slot. This has been observed in photonic mesh lattices using coupled fibre loops with balanced gain and loss [60], where the pulse dynamics directly map to the Hatano–Nelson model with tunable t_R/t_L.

Photonic crystals with exceptional points. Stable exceptional points in a two-dimensional photonic band structure often come with complex spectral images of finite area, and this finite area is a strong warning that boundary-sensitive spectra may appear under open boundaries [100]. This connects two seemingly different non-Hermitian phenomena—EPs and the skin effect—through the geometry of the PBC spectrum. A concrete prediction of the skin localization length, however, still requires the non-Bloch or transfer-matrix analysis for the chosen sample termination.

Non-reciprocal photonic lattices. Lattices of waveguides with magneto-optical or time-modulated elements provide a direct realisation of non-reciprocal hopping, the most straightforward route to the skin effect. In a magneto-optical implementation, the Faraday rotation introduces a direction-dependent phase shift that maps onto asymmetric hopping in the effective tight-binding model. The resulting skin-effect strength is controlled by the applied magnetic field, providing a continuously tunable knob.

Experimental signatures. The photonic skin effect has three principal experimental signatures that distinguish it from ordinary boundary-condition sensitivity:

1. **Unidirectional funnelling:** light injected at any site propagates to the same boundary, regardless of injection position.
2. **Spectral collapse:** the measured transmission spectrum under OBC is much narrower than under PBC (e.g., by wrapping the array into a ring).
3. **Eigenstate tomography:** the intensity profile of each resonance shows exponential decay away from one boundary, with a decay length that matches the GBZ prediction $\xi = 1/|\ln r_{\mathrm{GBZ}}|$.

Remark 14.1. The photonic skin effect has practical implications for the design of photonic devices: light in a skin-effect system accumulates at one boundary regardless of where it is injected, providing a mechanism for directional transport and amplification that is topologically robust. This directional funneling effect could be harnessed for energy harvesting, signal routing, and non-reciprocal photonic circuits.

14.8 Critical non-Hermitian skin effect

When two subsystems with different skin-effect localization lengths are coupled, a new critical behavior emerges [105]: the spectrum and eigenstates exhibit a *discontinuous* jump in the thermodynamic limit, depending on the order in which the system size and the coupling strength are taken to their limiting values. This *critical NHSE* is a phase transition with no Hermitian analogue, and it highlights the extreme sensitivity of non-Hermitian systems to boundary conditions and inter-subsystem coupling.

Mechanism. Consider two Hatano–Nelson chains with different asymmetry ratios $t_R^{(1)}/t_L^{(1)} \neq t_R^{(2)}/t_L^{(2)}$, coupled by a hopping κ at a single junction. When $\kappa = 0$, each chain has its own GBZ radius and its own skin localisation direction. When $\kappa \neq 0$, the two GBZs must merge into a single contour that satisfies the combined boundary conditions. The critical NHSE appears when the two individual GBZ radii are on opposite sides of the unit circle ($r_1 < 1$ and $r_2 > 1$): the merged GBZ undergoes a topological transition as a function of κ/t, jumping discontinuously from one contour to another at a critical coupling strength. This discontinuity is reflected in a sudden change of the OBC spectrum and the localisation pattern of the eigenstates.

Photonic realisation. A photonic version of the critical NHSE can be implemented in a coupled-resonator array where two segments have different gain/loss profiles, joined at a single coupling point. The coupling strength κ is controlled by the gap between the two resonator segments, providing a continuously tunable parameter that drives the critical transition.

14.9 Connection to earlier chapters

The skin effect ties together several threads from earlier chapters:

- **QNMs and open boundaries** (Chapter 6): the spatial divergence of QNM fields is a single-resonator version of the skin effect. In a periodic system, the skin effect is the collective version: every Bloch state acquires a QNM-like exponential envelope.
- **Point-gap topology** (Chapter 13): the winding number that classifies point-gap phases diagnoses boundary-sensitive spectra. In one dimension, a nonzero point-gap winding means the PBC and OBC spectra differ, and the GBZ is needed.
- **Poles and zeros** (Chapter 10): the poles of the scattering matrix of a finite non-Hermitian chain are the OBC eigenvalues. The spectral collapse from PBC loops to OBC arcs is a statement about the pole motion under boundary-condition changes.
- **Bulk-boundary correspondence** (Chapter 15): the skin effect forces a reconstruction of the BBC. The next chapter develops this reconstruction using the GBZ and the non-Bloch invariants introduced here.

14.9.1 From Maxwell transfer matrices to the photonic GBZ

The GBZ construction above uses tight-binding language, but it extends naturally to the full Maxwell formalism via the transfer-matrix method of Chapter 4. For a one-dimensional photonic crystal with N bilayers, each described by a 2×2 transfer matrix $\mathbf{M}(\omega)$, the Bloch condition is $\mathbf{M}(\omega)\,\mathbf{v} = \beta\,\mathbf{v}$, where $\beta = e^{iKd}$ is the Bloch phase per unit cell. The two eigenvalues $\beta_1(\omega)$ and $\beta_2(\omega)$ of $\mathbf{M}$ satisfy $\beta_1\beta_2 = \det\mathbf{M}$. In the Hermitian case, $|\det\mathbf{M}| = 1$ and the two eigenvalues are reciprocal partners on the unit circle in a pass band. For a reciprocal lossy stack, $\det\mathbf{M}$ can still equal unity even though the trace is complex; the eigenvalues then move off the unit circle as ordinary growing/decaying Bloch multipliers, not automatically as a skin-effect GBZ.

The photonic GBZ is constructed by the same procedure as in the tight-binding case:

1. Compute $\beta_1(\omega)$ and $\beta_2(\omega)$ from the transfer matrix at each complex frequency ω.
2. The GBZ contour is the locus of β values where $|\beta_1| = |\beta_2|$.

3. The non-Bloch band structure is obtained by evaluating $\omega(\beta)$ on this contour.

For a concrete gain/loss photonic crystal with bilayer permittivities $\varepsilon_A = 4{+}0.3i$ and $\varepsilon_B = 4{-}0.3i$, the transfer-matrix eigenvalues deviate from the unit circle in stop bands and the complex band structure can form loops. To claim a skin effect, one must additionally show that the open-boundary spectrum is reproduced by a GBZ different from the ordinary Bloch contour. In practice this usually requires an effective nonreciprocal coupling mechanism, a time-modulated unit cell, or a multi-channel network whose eliminated channels generate a direction-dependent round-trip gain.

Chapter 15

Bulk-Boundary Correspondence Rebuilt

The bulk-boundary correspondence is one of the deepest principles in topological physics: the number of protected edge states at an interface is determined by bulk topological invariants computed far from the boundary. In Hermitian photonic systems, this correspondence is exact and robust (Chapter 11). In non-Hermitian systems, it can fail: the non-Hermitian skin effect (Chapter 14) invalidates the Bloch-band invariants on which the conventional correspondence is built. This chapter shows how the correspondence is rebuilt*: not by abandoning topology, but by computing invariants on the generalized Brillouin zone that faithfully reflects the open-boundary physics.*

15.1 Why the conventional correspondence fails

15.1.1 The Hermitian baseline

Recall the logic of the Hermitian bulk-boundary correspondence. For a one-dimensional topological insulator—the prototypical Su–Schrieffer–Heeger (SSH) chain with alternating hopping amplitudes t_1 and t_2—the Bloch Hamiltonian is

$$H(k) = \begin{pmatrix} 0 & t_1 + t_2\, e^{-ik} \\ t_1 + t_2\, e^{ik} & 0 \end{pmatrix}. \tag{15.1}$$

The winding number of the off-diagonal element $h(k) = t_1 + t_2\, e^{-ik}$ around the origin in the complex plane is

$$W = \frac{1}{2\pi i} \oint_{\mathrm{BZ}} \frac{h'(k)}{h(k)}\, \mathrm{d}k = \begin{cases} 1 & |t_1| < |t_2|, \\ 0 & |t_1| > |t_2|. \end{cases} \tag{15.2}$$

When $W = 1$, a finite open chain has one zero-energy edge state at each boundary. When $W = 0$, there are no edge states. The topological phase transition occurs at $|t_1| = |t_2|$, where the bulk gap closes. This is the conventional bulk-boundary correspondence: the *bulk* winding number, computed from the *periodic* Bloch Hamiltonian, predicts the *boundary* physics of an *open* chain.

The correspondence works because, in a Hermitian system, the PBC and OBC spectra agree in the thermodynamic limit. The Bloch eigenstates are extended plane waves that faithfully represent the bulk physics, regardless of boundary conditions.

15.1.2 The non-Hermitian SSH model

Now consider a non-Hermitian generalization of the SSH model [99, 87]. The simplest version introduces asymmetric intra-cell hopping:

$$H_{\mathrm{NH}}(k) = \begin{pmatrix} 0 & t_1 + \gamma/2 + t_2\, e^{-ik} \\ t_1 - \gamma/2 + t_2\, e^{ik} & 0 \end{pmatrix}, \tag{15.3}$$

where γ is a non-Hermitian parameter that makes the hopping asymmetric ($t_1 + \gamma/2$ from right to left, $t_1 - \gamma/2$ from left to right within the unit cell). The Bloch Hamiltonian is no longer Hermitian: $H_{\mathrm{NH}}(k) \neq H_{\mathrm{NH}}^{\dagger}(k)$.

15.1.3 The failure

The conventional winding number, computed from $h(k) = t_1 + \gamma/2 + t_2\, e^{-ik}$, predicts a topological phase transition at

$$|t_1 + \gamma/2| = |t_2| \qquad \text{(Bloch prediction)}. \tag{15.4}$$

However, exact diagonalization of the open chain reveals that the topological zero modes appear and disappear at a *different* parameter value:

$$t_1 = \pm\sqrt{t_2^2 + (\gamma/2)^2} \qquad \text{(true OBC transition)}. \tag{15.5}$$

The Bloch winding number gives the *wrong phase diagram* [88, 85, 86].

The root cause is the non-Hermitian skin effect (Chapter 14). Under OBC, the bulk spectrum develops an extensive set of exponentially localized skin states (Figure 14.2), and the PBC and OBC spectra can be very different (Figure 14.1). The Bloch Hamiltonian, which assumes extended eigenstates on the unit circle $|\beta| = 1$, does not describe the open-boundary physics.

A spectacular failure

The failure is not a small quantitative correction. For $\gamma = 0.8$, $t_2 = 1$, the Bloch prediction places the phase boundary at $t_1 = 0.6$, while the true OBC boundary is at $t_1 \approx 1.077$. The entire topological phase diagram is qualitatively wrong—including regions where the Bloch invariant predicts a trivial phase but topological edge states exist, and vice versa.

Worked numerical comparison. To make this failure quantitative, take $t_2 = 1$, $\gamma = 0.8$ and vary t_1 through the transition region. At $t_1 = 0.7$:

- **Bloch prediction**: $|t_1 + \gamma/2| = |0.7 + 0.4| = 1.1 > 1$, so $W_{\mathrm{Bloch}} = 0$ (trivial).
- **Non-Bloch prediction**: GBZ radius $r = \sqrt{(t_1 - \gamma/2)/(t_1 + \gamma/2)} = \sqrt{0.3/1.1} \approx 0.522$; $W_{\mathrm{GBZ}} = 1$ (topological).

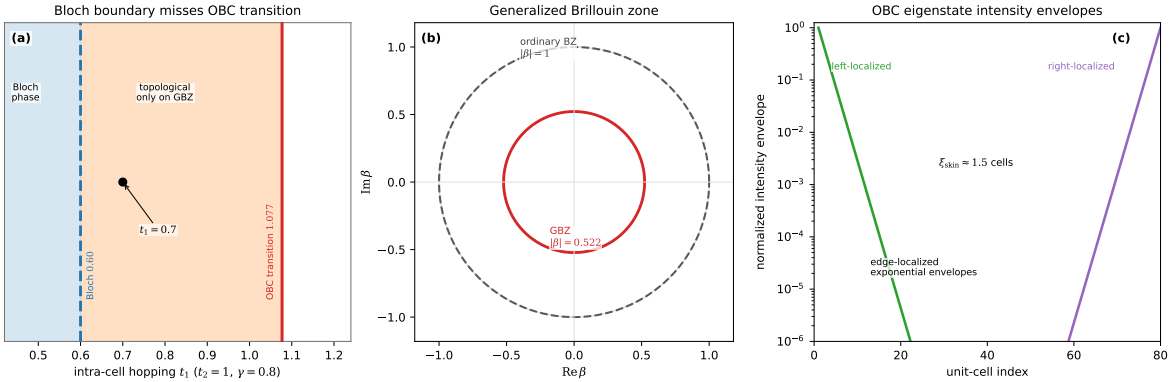


Figure 15.1 Failure of the Bloch bulk-boundary correspondence and its non-Bloch resolution for the non-Hermitian SSH model with $t_2 = 1$ and $\gamma = 0.8$. (a) One-parameter phase strip. The dashed blue line marks the Bloch/PBC phase boundary at $t_1 = 0.60$; the solid red line marks the true OBC/GBZ transition at $t_1 = 1.077$. The orange region between them is topologically nontrivial only on the generalised Brillouin zone—a qualitative failure of the conventional Bloch invariant. The black dot marks $t_1 = 0.7$, the parameter used in panels (b) and (c). (b) Generalised Brillouin zone in the complex β plane. The dashed grey circle is the ordinary BZ ($|\beta| = 1$); the solid red circle is the GBZ with $|\beta| = 0.522$, computed from the transfer-matrix eigenvalue condition. (c) Normalised intensity envelopes of representative OBC zero modes on a logarithmic scale. Green and purple mark modes localised at the two terminations. The decay is exponential with skin length $\xi_{\text{skin}} = -1/\ln|\beta_{\text{GBZ}}| \approx 1.54$ unit cells. Curves are clipped below 10^{-6} for display.

- **OBC exact diagonalisation** ($L = 80$): two robust zero-energy modes exist, confirming the non-Bloch prediction.

At $t_1 = 1.2$, both predictions agree: $W = 0$ and no edge states. The discrepancy region $0.6 < t_1 < 1.077$ is entirely missed by the Bloch invariant. This is not a perturbative correction but a qualitative failure of the conventional theory.

The skin effect as the root cause. The failure traces directly to the skin effect (Chapter 14). Under OBC, the eigenstates of H_{NH} are not extended Bloch waves but exponentially localised skin modes with spatial envelope $\sim r^n$, where $r = |\beta_{\text{GBZ}}| \neq 1$. The Bloch Hamiltonian assumes $|\beta| = 1$ and therefore probes the wrong region of the complex β plane. The GBZ corrects this by replacing the unit circle with the contour $|\beta| = r$ that matches the open-boundary eigenstates.

Physically, the skin effect means that the bulk modes themselves carry information about the boundary: their exponential envelope 'knows' which boundary they are localised toward. The Bloch description, which by construction ignores boundaries, cannot capture this information. The non-Bloch theory restores the correspondence by building boundary awareness into the very definition of the Brillouin zone.

Figure 15.1 summarises the three ingredients of the failure and its resolution for the non-Hermitian SSH model at $t_2 = 1$, $\gamma = 0.8$.

15.2 The non-Bloch bulk-boundary correspondence

The resolution, introduced by Yao and Wang, is to compute topological invariants on the generalized Brillouin zone (Section 14.4)—the contour in the complex β plane that correctly describes the open-boundary eigenstates.

15.2.1 Statement

For a one-dimensional non-Hermitian system with chiral symmetry, the *non-Bloch bulk-boundary correspondence* [85, 86, 103] states:

$$\boxed{N_{\text{zero modes per interface}} = |W_{\text{GBZ}}|,} \tag{15.6}$$

where W_{GBZ} is the winding number computed on the GBZ (14.16). For a finite SSH chain with two topological terminations, the total number of zero modes is therefore twice this value. This invariant correctly predicts the topological zero modes throughout the gapped parameter regime, including where the conventional Bloch invariant gives the wrong answer.

15.2.2 The GBZ for the non-Hermitian SSH model

For the non-Hermitian SSH model (15.3), replacing $e^{ik} \to \beta$ yields the characteristic equation

$$(t_1 + \gamma/2 + t_2\,\beta^{-1})(t_1 - \gamma/2 + t_2\,\beta) = E^2. \tag{15.7}$$

At $E = 0$ (the gap-closing energy for chiral-symmetric systems), this gives four solutions for β. Ordering them by modulus, $|\beta_1| \le |\beta_2| \le |\beta_3| \le |\beta_4|$, the GBZ condition (14.11) requires $|\beta_2| = |\beta_3|$.

For this model, the GBZ is a circle of radius

$$|\beta_{\text{GBZ}}| = r = \sqrt{\frac{|t_1 - \gamma/2|}{|t_1 + \gamma/2|}}. \tag{15.8}$$

The final equality assumes the positive real parameter regime used in the example. The intercell hopping t_2 sets the overall frequency scale for this symmetric intercell choice and therefore does not affect the GBZ radius. In the Hermitian limit $\gamma = 0$, the radius becomes $r = 1$, so the GBZ reduces continuously to the ordinary Brillouin zone. The GBZ-based winding number is computed by evaluating $h(\beta)$ on this deformed contour rather than on the unit circle.

15.2.3 Derivation of the non-Bloch BBC

The key idea is that the open-boundary eigenstates at zero energy are determined by the transfer-matrix eigenvalues at $E = 0$. For a chain of L unit cells, a zero-energy eigenstate must satisfy the boundary conditions at both the left and right ends. Writing the spatial dependence as a sum of the four transfer-matrix solutions $\psi_n \sim \sum_j c_j\,\beta_j^n$, normalizability imposes constraints from both boundaries.

From the left boundary, the coefficients of the solutions that grow toward the left ($|\beta_j| > 1$ when viewed from the left end) must vanish; otherwise $|\psi_n|$ would diverge as $n \to \infty$. Similarly, from the right boundary, the coefficients of the solutions that grow toward the right ($|\beta_j| < 1$ when viewed from the right end) must vanish from the right-end perspective. For a system with $2n$ transfer-matrix eigenvalues ordered by modulus $|\beta_1| \le |\beta_2| \le \cdots \le |\beta_{2n}|$, a normalizable zero mode exists precisely when the boundary conditions can be simultaneously satisfied—that is, when exactly n eigenvalues lie inside the GBZ contour and n lie outside.

The GBZ contour is defined by the condition $|\beta_n| = |\beta_{n+1}|$: the two "middle" eigenvalues must have equal modulus. This is because the zero mode at the interface between the inside and outside groups must change character precisely at the GBZ radius. When this condition holds, the transfer-matrix problem has a self-consistent solution that is normalizable from both ends simultaneously.

As parameters are varied, a zero mode appears or disappears when a β_j *crosses the GBZ contour*—not when it crosses the unit circle. The winding number computed on the GBZ counts how many times this crossing has occurred, and therefore correctly predicts the number of topological zero modes.

Physical interpretation of the GBZ radius. The GBZ radius $r = |\beta_{\mathrm{GBZ}}|$ has a direct physical meaning: it determines the *skin localisation length*

$$\xi_{\mathrm{skin}} = \frac{-1}{\ln r}, \tag{15.9}$$

measured in unit cells. When $r < 1$, the bulk skin states are exponentially localised toward the left boundary with decay length ξ_{skin}; when $r > 1$, they localise toward the right. In the Hermitian limit $r \to 1$, $\xi_{\mathrm{skin}} \to \infty$ and the bulk modes become extended Bloch waves.

For the non-Hermitian SSH model with $t_2 = 1$ and $\gamma = 0.8$, the GBZ radius is $r = 0.522$ (Figure 15.1b), giving a skin length $\xi_{\mathrm{skin}} = 1/|\ln 0.522| \approx 1.54$ unit cells—an extremely strong localisation. This quantitative prediction can be tested directly by measuring the spatial decay of the field envelope in a finite photonic lattice, as demonstrated in microwave waveguide networks and acoustic metamaterials.

GBZ for general two-band models. The GBZ construction generalises straightforwardly to any two-band non-Hermitian model with nearest-neighbour coupling. The Bloch Hamiltonian takes the form $H(\beta) = \sum_j d_j(\beta)\,\sigma_j$ where $d_j(\beta) = d_j^{(0)} + d_j^{(+)}\beta + d_j^{(-)}\beta^{-1}$. The characteristic equation $\det(E - H(\beta)) = 0$ is a polynomial in β (typically quartic for two-band models), and the GBZ is the contour in the complex β plane on which the middle pair of roots has equal modulus. For models with only one asymmetric hopping parameter (such as the Hatano–Nelson model), the GBZ is a circle and the calculation reduces to finding a single radius r. For models with multiple asymmetric parameters, the GBZ can be a non-circular closed curve, and the winding number must be computed numerically by discretising the GBZ contour and tracking the phase of $\det H(\beta)$. The algorithms for this computation are discussed in Chapter 17.

Remark 15.1. The derivation shows why the *shape* of the GBZ matters, not just its existence. For models where the GBZ is not a circle (e.g., multi-band or long-range hopping models), the contour can be self-intersecting or multi-sheeted, and the winding number must be computed on the full geometric GBZ rather than a simple circle.

15.3 Example: non-Hermitian SSH phase diagram

To make the non-Bloch BBC concrete, consider the non-Hermitian SSH model with $t_2 = 1$ and $\gamma = 0.8$.

15.3.1 Bloch versus non-Bloch phase boundaries

The conventional Bloch winding number predicts a topological phase transition at $|t_1 + 0.4| = 1$, i.e., $t_1 = 0.6$ or $t_1 = -1.4$. The non-Bloch winding number, computed on the GBZ, predicts a transition at $t_1 = \sqrt{1 + 0.16} \approx 1.077$. These are quantitatively different—the Bloch prediction misplaces the phase boundary by nearly a factor of two.

15.3.2 OBC spectrum and edge states

Exact diagonalization of an open chain of $L = 60$ unit cells confirms the non-Bloch prediction. For $t_1 = 0.8$ (between the Bloch and non-Bloch boundaries), the Bloch invariant gives $W = 0$ (trivial), but the OBC spectrum shows two robust zero-energy edge states. For $t_1 = 1.2$ (above the non-Bloch boundary), the edge states vanish.

This example demonstrates that the non-Bloch BBC is essential for predicting the topological phase diagram of non-Hermitian systems with the skin effect.

15.4 Biorthogonal bulk-boundary correspondence

An alternative approach to rebuilding the bulk-boundary correspondence uses the *biorthogonal polarization*—a non-Hermitian generalization of the electric polarization or Zak phase.

15.4.1 Biorthogonal polarization

In a Hermitian system, the bulk polarization is defined as the integral of the Berry connection over the Brillouin zone. In a non-Hermitian system, the biorthogonal polarization uses the left and right eigenvectors:

$$P^{\mathrm{bi}} = \frac{1}{2\pi} \oint_{\mathrm{BZ}} i \, \frac{\langle u_k^{\mathrm{L}} | \partial_k | u_k^{\mathrm{R}} \rangle}{\langle u_k^{\mathrm{L}} | u_k^{\mathrm{R}} \rangle} \, \mathrm{d}k. \tag{15.10}$$

This quantity is quantized (modulo integers) by chiral symmetry and predicts the existence of edge states in the biorthogonal framework.

15.4.2 Physical interpretation

The biorthogonal polarization has a direct physical interpretation. In a Hermitian system, the bulk polarization measures the displacement of charge (or field energy) within the unit cell. In a non-Hermitian system, P^{bi} measures the displacement of the *biorthogonal* charge density $\rho^{\mathrm{bi}}(x) = \langle \psi^{\mathrm{L}}(x) | \psi^{\mathrm{R}}(x) \rangle$, which can be complex. The quantization of P^{bi} reflects the fact that a half-integer displacement of the biorthogonal charge (modulo one) signals a topological phase with boundary-localized modes.

Example: biorthogonal polarisation of the NH SSH model. For the non-Hermitian SSH model (15.3) with $t_1 = 0.5$, $t_2 = 1$, $\gamma = 0.6$, the right eigenvector of the lower band is $|u_k^{\mathrm{R}}\rangle = (1, \; -E(k)/h(k))^{\mathsf{T}}$ (up to normalisation and branch convention), where $E(k) = \sqrt{h(k)\tilde{h}(k)}$, $h(k) = t_1 + \gamma/2 + t_2 e^{-ik}$, and $\tilde{h}(k) = t_1 - \gamma/2 + t_2 e^{ik}$. The left eigenvector is obtained from the transpose/adjoint eigenproblem with the corresponding left-right normalization. Numerical

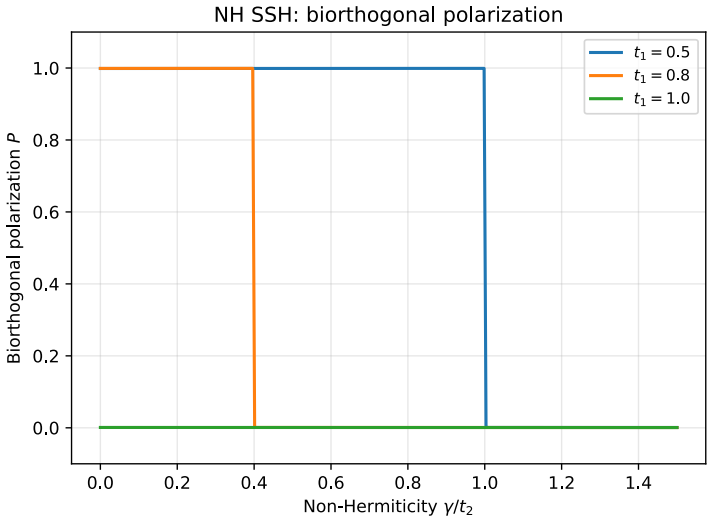


Figure 15.2 Biorthogonal polarization diagnostic as a function of non-Hermiticity γ/t_2 for different intracell hoppings t_1 in the NH SSH model. The plotted convention uses values 0 and 1 for the two quantized sectors; in the usual polarization convention these correspond to $P = 0$ and $P = 1/2$ modulo one. The jump marks the non-Hermitian topological transition in this biorthogonal diagnostic. Chiral symmetry pins the value to a quantized sector away from gap closings.

integration of the biorthogonal connection, or equivalently the non-Bloch version evaluated on the GBZ when skin modes are present, gives the quantized topological sector. For this parameter set the polarization is in the topological sector ($P = 1/2$ modulo one in the usual convention). When t_1 is increased past the non-Bloch transition ($t_1 = \sqrt{t_2^2 + (\gamma/2)^2} = \sqrt{1.09} \approx 1.044$), the non-Bloch/biorthogonal diagnostic jumps discontinuously to the trivial sector (trivial), in agreement with the non-Bloch winding number. The jump is sharp—not a smooth crossover—because P^{bi} is quantised by chiral symmetry.

This confirms that the biorthogonal polarisation and the non-Bloch winding number are two faces of the same topological invariant: they detect the same phase transitions and predict the same edge states, but provide different physical pictures—one based on GBZ deformation, the other on biorthogonal charge displacement.

15.4.3 Connection to QNM inner products

The biorthogonal polarization (15.10) is closely [87] related to the unconjugated inner product used for quasinormal modes (Section 6.4). In both cases, the key mathematical structure is the pairing between left (adjoint) and right eigenstates. For a photonic system described by the Maxwell operator $\hat{\Theta}$, the biorthogonal Berry connection becomes

$$A_k^{\mathrm{bi}} = i \, \frac{\langle \tilde{\mathbf{E}}_k^{\mathrm{L}} | \partial_k | \tilde{\mathbf{E}}_k^{\mathrm{R}} \rangle_\varepsilon}{\langle \tilde{\mathbf{E}}_k^{\mathrm{L}} | \tilde{\mathbf{E}}_k^{\mathrm{R}} \rangle_\varepsilon}, \tag{15.11}$$

where $\langle \cdot | \cdot \rangle_\varepsilon$ is the unconjugated inner product weighted by the permittivity tensor. This connects the topological classification of non-Hermitian photonic bands directly to the QNM formalism of Chapter 6.

15.4.4 Relation to the non-Bloch approach

The biorthogonal and non-Bloch approaches give consistent predictions when the inner product and integration contour are chosen for the same boundary problem, but they emphasize complementary physical insights:

- The **non-Bloch approach** directly accounts for the skin effect by deforming the integration contour; it gives the correct phase diagram and the correct localization properties of edge states.
- The **biorthogonal approach** replaces the inner product with a left-right one and is more closely connected to the formalism of Chapter 6 and the adjoint modes of the Maxwell operator. When the skin effect is present, its bulk version must be combined with the non-Bloch contour or with open-boundary biorthogonal states.
- For systems *without* the skin effect (e.g., systems with reciprocity or additional symmetries that prevent the spectral loop from winding), the biorthogonal approach may be the only relevant non-Hermitian correction.
- For systems *with* the skin effect, both approaches are needed for a complete picture: the non-Bloch approach gives the correct phase diagram, and the biorthogonal approach clarifies the physical meaning of the topological invariant.

15.5 Higher-dimensional correspondence

In two and higher dimensions, rebuilding the bulk-boundary correspondence requires both non-Bloch band theory and the spectral-area diagnostics discussed in Chapter 13 [100].

15.5.1 Non-Bloch Chern insulator

For a two-dimensional non-Hermitian Chern insulator, the non-Bloch Chern number (Section 14.5.2) correctly predicts the number of chiral edge modes, which can differ from the prediction of the conventional Chern number when the skin effect is present.

The construction is conceptually straightforward: the two-dimensional GBZ is the product of two one-dimensional GBZ contours (one for each reciprocal direction), and the non-Bloch Chern number is the integral of the Berry curvature over this deformed torus. In practice, computing the two-dimensional GBZ is more involved because the GBZ contour in each direction can depend on the momentum in the other direction.

15.5.2 Corner modes and higher-order topology

In certain two-dimensional systems, an extensive set of eigenstates can be localized at *corners* rather than edges—the corner-skin effect (Section 14.6). The relevant topological invariant is a second-order non-Bloch invariant, defined by a nested GBZ construction: first compute the GBZ in one direction at fixed momentum in the other, then compute a one-dimensional topological invariant on the resulting edge theory.

The corner modes predicted by this construction are exponentially localized at zero, one, or two corners of a rectangular sample, depending on the non-Bloch invariant and the orientation of the skin effect.

15.5.3 Geometry dependence

The geometry-dependent skin effect (Section 14.6) means that the bulk-boundary correspondence itself can depend on the sample shape. For example, a hexagonal sample may exhibit edge states on some edges but not others, depending on the orientation of the non-reciprocal hopping. This shape dependence has no Hermitian counterpart and requires a boundary-condition-aware formulation of the topological invariant.

15.6 Photonic implications

For photonic systems described by the Maxwell equations, the rebuilt bulk-boundary correspondence has several practical implications.

15.6.1 Photonic GBZ construction

For a 1D non-Hermitian photonic system with effective nonreciprocal or non-unimodular evolution, the GBZ can be constructed directly from the Maxwell transfer matrix without passing through a tight-binding approximation. The procedure is:

1. Compute the unit-cell transfer matrix $M(\omega)$ at each complex frequency ω using the complex-n_j layer propagation matrices of Chapter 13.
2. Find the eigenvalues $\beta_{1,2}(\omega)$ of $M(\omega)$.
3. The GBZ is the corresponding contour in the complex β plane where the middle transfer-matrix roots have equal modulus; for a two-root problem this condition is $|\beta_1(\omega)| = |\beta_2(\omega)|$.
4. On this contour, evaluate the photonic topological invariant (winding number or biorthogonal polarisation).

For a reciprocal isotropic layer stack described by the usual two-component field vector, the unit-cell transfer matrix is unimodular even with material loss or gain. Then $\beta_1\beta_2 = 1$, and the condition $|\beta_1| = |\beta_2|$ reduces to $|\beta_1| = |\beta_2| = 1$. Complex permittivity can still produce complex bands and evanescent solutions, but it does not by itself generate a Hatano–Nelson-type GBZ deformation. A genuine photonic non-Bloch correction requires an effective evolution that breaks this cancellation, for example through magneto-optical bias, temporal modulation, cascaded reservoirs, or a multi-port network after unobserved channels are eliminated.

Connection to the photonic transfer-matrix formalism. The photonic GBZ construction links directly to the transfer-matrix band theory of Chapter 3: the Bloch condition $\cos(Ka) = \xi(\omega)$ is replaced, in the effective nonreciprocal problem, by the middle-root condition defining the GBZ. Instead of solving only for real K on the unit circle, one solves for the complex K values that reproduce the open-boundary spectrum. This makes the non-Bloch theory of photonic topological phases accessible to the same transfer-matrix and scattering tools used routinely

in multilayer and coupled-resonator design, provided the effective model includes the channels responsible for directional gain or hopping.

15.6.2 Design of robust edge channels

In a non-Hermitian photonic crystal, edge channels designed using conventional (Bloch) topological invariants may not appear or may appear at unexpected parameter values. The non-Bloch approach provides a concrete design rule: compute the GBZ for the target structure, evaluate the topological invariant on the GBZ, and verify the edge-state prediction against finite-structure simulations. This ensures that the designed edge states are robust and correctly placed.

The practical difficulty is that computing the photonic GBZ requires solving the Maxwell transfer matrix at complex frequencies and complex wave vectors—a task that goes beyond standard photonic band-structure calculations but is within the reach of the numerical methods of Chapter 17.

15.6.3 Directional amplification

The skin effect combined with edge-state topology enables *topological amplifiers*: devices where light injected at one boundary is amplified while propagating along an edge channel, with directionality set by the non-Bloch topology and by the engineered gain profile. The amplification can scale exponentially over the useful device length, but in a physical photonic structure it is limited by saturation, radiation loss, and disorder-induced mode mixing rather than being absolutely immune to backscattering.

15.6.4 Spectral engineering

The dramatic difference between PBC and OBC spectra in non-Hermitian photonic crystals can be exploited for spectral engineering. By choosing boundary conditions (periodic, open, or mixed), the same photonic structure can exhibit qualitatively different spectral properties. For example, a ring resonator (PBC) and a terminated waveguide array (OBC) made from the same non-Hermitian photonic crystal will have different resonance frequencies, different mode spacings, and different field distributions.

15.6.5 Experimental signatures

The non-Bloch BBC makes testable predictions for photonic experiments:

1. **Edge-state presence:** the appearance or disappearance of edge modes as a function of gain/loss or coupling asymmetry should follow the non-Bloch phase boundary, not the Bloch one.
2. **Spectral collapse:** the PBC-to-OBC spectral change (Figure 14.1) is a direct observable in coupled-resonator experiments where periodic or open boundary conditions can be implemented.
3. **GBZ deformation:** the GBZ radius (Figure 14.4) manifests as the exponential envelope of the skin modes, which can be measured from the spatial intensity profile.

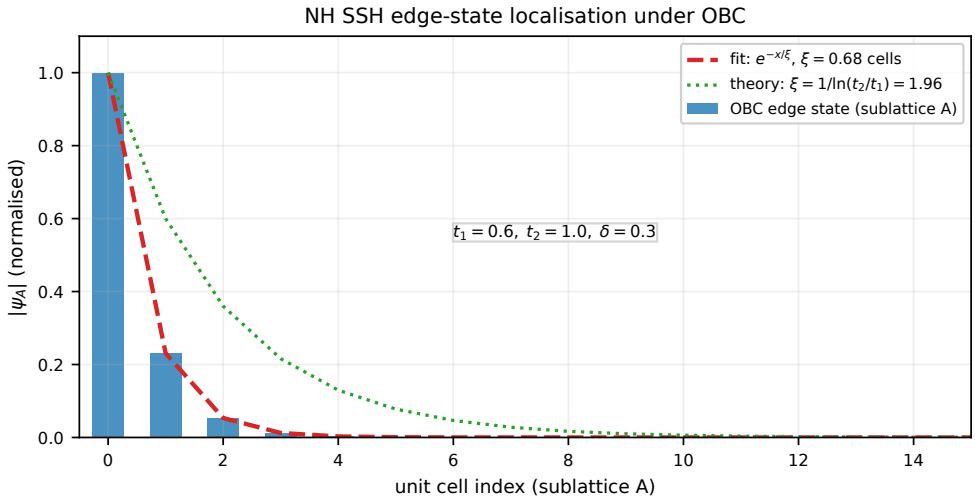


Figure 15.3 Sublattice-A amplitudes of the zero-energy edge state in the non-Hermitian SSH model ($t_1 = 0.6$, $t_2 = 1.0$, $\delta = 0.3$) under open boundary conditions, normalized to the boundary site. Red dashed: exponential fit to the first eight sites gives $\xi_{\text{fit}} \approx 0.68$ cells, in agreement with the nonreciprocal zero-mode recursion (15.12). Green dotted: the Hermitian estimate $\xi_{\text{Herm}} = 1/\ln(t_2/t_1) \approx 1.96$ cells, which ignores the hopping asymmetry δ. The discrepancy is therefore not a plotting artifact but the edge-state analogue of the non-Bloch correction: the correct decay length follows from the open-boundary recursion or, equivalently, from the generalized Brillouin zone.

Edge-state penetration depth. In the non-Hermitian SSH model used in Fig. 15.3, $t_1 = 0.6$, $t_2 = 1.0$, and the nonreciprocal hopping parameter is $\delta = 0.3$. The right zero mode on sublattice A obeys the open-chain recursion

$$(t_1 - \delta)A_n + (t_2 + \delta)A_{n+1} = 0, \qquad \frac{|A_{n+1}|}{|A_n|} = \left|\frac{t_1 - \delta}{t_2 + \delta}\right|, \tag{15.12}$$

so the amplitude penetration depth is

$$\xi_{\text{edge}} = \frac{1}{\ln|(t_2 + \delta)/(t_1 - \delta)|} = \frac{1}{\ln(1.3/0.3)} \approx 0.68 \text{ cells}. \tag{15.13}$$

The Hermitian estimate obtained by setting $\delta = 0$ is $\xi_{\text{Herm}} = 1/\ln|t_2/t_1| \approx 1.96$ cells, but that estimate ignores the nonreciprocal intracell/intercell imbalance. The non-Bloch calculation and the direct OBC eigenvector therefore agree: the edge mode is much more tightly confined than the Hermitian SSH formula would suggest. The exponential envelope $\propto e^{-x/\xi_{\text{edge}}}$ has been confirmed experimentally in coupled-resonator arrays at both microwave and optical frequencies, where the spatial intensity profile of the edge mode can be measured site by site.

Experimental verification at optical frequencies. Photonic and microwave coupled-resonator experiments can test the non-Bloch BBC by comparing spectra under ring-closed and ring-terminated boundary conditions. The key checks are qualitative and quantitative: the measured phase boundary should follow the GBZ prediction, and the observed edge-mode localisation length should match the open-boundary recursion or Eq. (15.13) within fabrication tolerance. Such measurements would test the non-Bloch correction to the topological properties of non-Hermitian photonic systems.

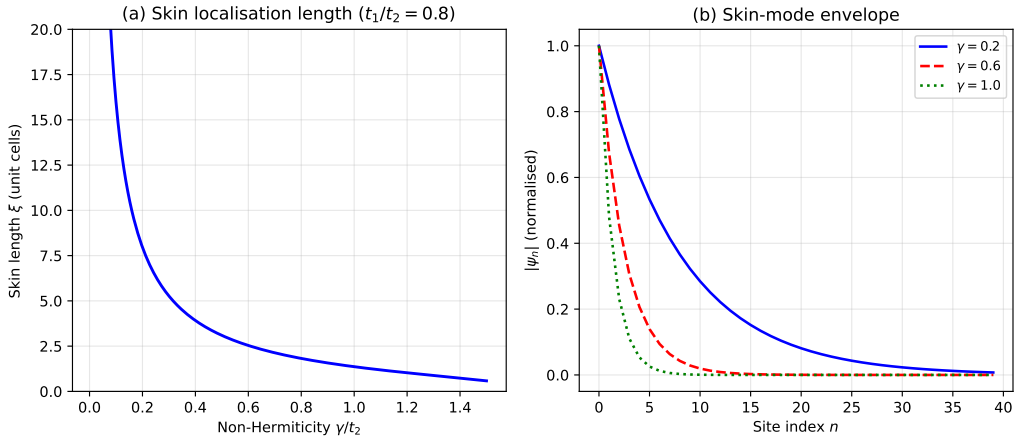


Figure 15.4 Skin-effect localisation in the NH SSH model ($t_1/t_2 = 0.8$). (a) Skin localisation length ξ as a function of the non-Hermiticity parameter γ: ξ diverges as $\gamma \to 0$ (Hermitian limit, where all bulk modes become extended) and decreases monotonically with increasing asymmetry. (b) Skin-mode envelope $|\psi_n|$ for three values of γ, showing progressively stronger localisation toward one boundary as the non-Hermiticity grows.

The bulk-boundary correspondence survives

The bulk-boundary correspondence is not destroyed by non-Hermiticity—it is transformed. In its rebuilt form, it connects topological invariants computed on the generalized Brillouin zone (or using biorthogonal inner products) to the edge-state structure of open non-Hermitian systems. The correspondence remains exact and robust, but it requires the non-Bloch theory of Chapter 14: the unit circle is no longer the correct Brillouin zone, the Bloch winding number is no longer the correct invariant, and the Bloch phase diagram is no longer the correct map. Once these replacements are made, the deep connection between bulk topology and boundary physics is fully restored.

Chapter 16

Bound States in the Continuum and Exceptional Topology

Not all non-Hermitian phenomena involve complex eigenfrequencies. A bound state in the continuum *(BIC) [17, 32, 106, 107, 108, 109, 110] is a mode with a real eigenfrequency embedded within the radiation continuum—a state that, by symmetry or fine-tuning, is perfectly decoupled from all outgoing channels despite coexisting spectrally with radiation modes. BICs often behave as topological objects: in photonic crystal slabs with a well-defined far-field polarization, they carry quantized charges in momentum space and can be connected to exceptional points through the analytic structure of open-system resonances. The BIC condition is derived from the Maxwell problem, the role of topological charge is identified, and photonic crystal slabs serve as the main examples.*

16.1 What is a bound state in the continuum?

In an open photonic structure—a photonic crystal slab, a grating, or a metasurface—most resonances are quasinormal modes with complex frequencies $\tilde{\omega} = \omega - i\gamma$ and finite quality factors (Chapter 6). A BIC is the exception: it has $\gamma = 0$ (infinite Q) despite lying within the frequency range where radiation channels are open.

Formally, a BIC is a normalizable solution of the source-free Maxwell equations at a real frequency ω_{BIC} that satisfies $\omega_{\text{BIC}} > \omega_{\text{light line}}$—above the light line, where radiation modes exist—yet has zero coupling to all radiation channels.

16.1.1 The radiation-amplitude viewpoint

To make the BIC condition precise within the Maxwell framework, consider a photonic crystal slab of thickness d with in-plane periodicity. At each Bloch wave vector $\mathbf{k}_{\parallel}$ and frequency ω, the far-field radiation above and below the slab can be decomposed into a finite number of open diffraction channels, each carrying a transverse polarization. Let $c_j(\mathbf{k}_{\parallel})$ denote the complex

amplitude radiated into channel j. The total radiation rate is

$$\gamma(\mathbf{k}_\parallel) \propto \sum_j |c_j(\mathbf{k}_\parallel)|^2. \tag{16.1}$$

A BIC occurs when every *open* radiation amplitude vanishes simultaneously:

$$c_j(\mathbf{k}_{\mathrm{BIC}}) = 0 \quad \text{for all open channels } j. \tag{16.2}$$

This is a set of conditions—one complex equation per independent channel amplitude, modulo symmetry and phase constraints—so generic BICs normally require tuning. At high-symmetry points where symmetry forbids certain channels, fewer conditions remain and BICs arise more naturally.

16.1.2 Symmetry-protected BICs

A symmetry-protected BIC exists at high-symmetry points in the Brillouin zone (e.g., Γ) where the mode belongs to an irreducible representation that is incompatible with the radiation continuum. For example, a mode that is odd under a C_2 rotation of the slab cannot couple to normally incident plane waves, which transform as the trivial representation. The decoupling is exact and persists under any perturbation that preserves the symmetry.

At the Γ point of a square-lattice photonic crystal slab, the point group is C_{4v}. Modes belonging to the B_1 or B_2 representations are incompatible with the s- and p-polarized plane waves (which belong to A_1 and E), ensuring zero radiation. Moving away from Γ reduces the symmetry, and the mode acquires a finite radiation rate—becoming a quasi-BIC with high but finite Q.

16.1.3 Accidental (tuned) BICs

At generic $\mathbf{k}$-points away from high-symmetry positions, BICs can arise from a destructive interference between two or more radiation channels. These accidental BICs require fine-tuning of one structural parameter (e.g., the slab thickness, hole radius, or lattice constant) and are generically isolated points in the ($\mathbf{k}$, parameter) space.

Near an accidental BIC, the radiation amplitudes vanish linearly: $c_j \propto |\mathbf{k} - \mathbf{k}_{\mathrm{BIC}}|$. Since $\gamma \propto \sum |c_j|^2$, the quality factor diverges as

$$Q(\mathbf{k}) \propto \frac{1}{|\mathbf{k} - \mathbf{k}_{\mathrm{BIC}}|^2}, \tag{16.3}$$

producing a family of ultra-high-Q quasi-BIC resonances in the vicinity.

16.1.4 Other BIC mechanisms

Beyond symmetry protection and parametric tuning, BICs can also arise through Fabry–Pérot interference (standing waves in the slab thickness direction that decouple from radiation), Friedrich–Wintgen avoided crossing (two resonances exchanging their radiation channels so that one becomes dark), and inverse-construction methods that engineer the potential or hopping

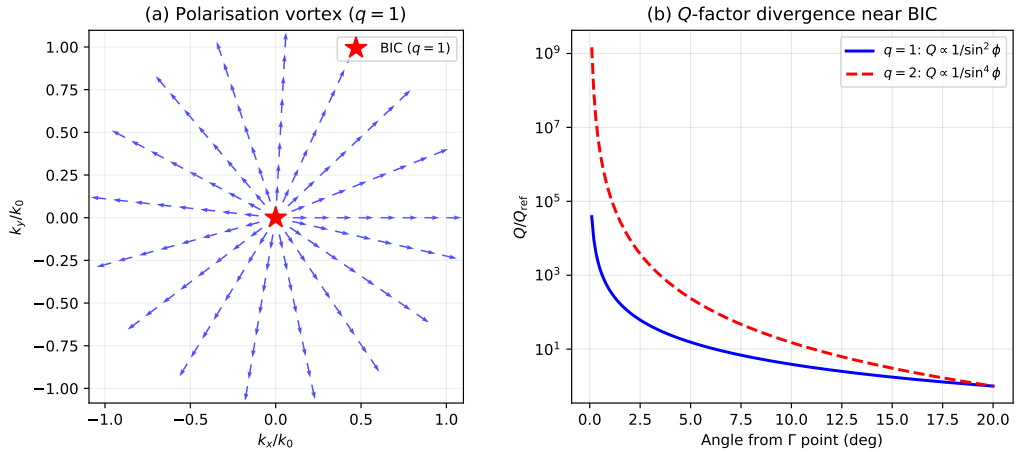


Figure 16.1 Far-field polarization vortex and quality-factor divergence near a symmetry-protected BIC. (a) The polarization direction winds once around the BIC in momentum space, giving topological charge $q = 1$. (b) The radiative quality factor diverges as the far-field radiation amplitude vanishes; higher-charge vortices produce steeper Q-scaling. This figure connects the Maxwell radiation amplitudes $c_j(\mathbf{k})$ in Eq. (16.1) to the polarization topology used to classify BICs.

rates to produce a bound state at a desired location [111]. The common thread is the vanishing of all radiation amplitudes (16.2).

16.1.5 Friedrich–Wintgen avoided crossing

A particularly instructive BIC mechanism arises from the avoided crossing of two resonances coupled through a shared radiation continuum. Consider two quasi-bound modes $|1\rangle$ and $|2\rangle$ with complex frequencies $\tilde{\omega}_j = \omega_j - i\gamma_j$ that approach each other as a structural parameter is varied. In the two-mode TCMT of Chapter 7, the effective Hamiltonian including radiation coupling is

$$H_{\mathrm{FW}} = \begin{pmatrix} \omega_1 - i\gamma_1 & \kappa - i\sqrt{\gamma_1\gamma_2}\, e^{i\phi} \\ \kappa - i\sqrt{\gamma_1\gamma_2}\, e^{-i\phi} & \omega_2 - i\gamma_2 \end{pmatrix}, \tag{16.4}$$

where κ is the near-field coupling and the off-diagonal imaginary term represents far-field coupling mediated by the shared continuum, with relative phase ϕ. Friedrich and Wintgen showed that there exists a parameter value at which one eigenvalue becomes purely real: the radiation widths of the two modes rearrange so that one mode absorbs all the radiation and the other becomes a BIC with $\gamma = 0$.

In this two-mode model the Friedrich–Wintgen BIC occurs when the dark eigenvector is orthogonal to the continuum-coupling vector. For the phase convention of Eq. (16.4), this can be written schematically as

$$\frac{\kappa}{\sqrt{\gamma_1\gamma_2}} = \frac{\omega_1 - \omega_2}{\gamma_1 - \gamma_2}\,\sin\phi, \tag{16.5}$$

with the precise form depending on the normalization of the radiative coupling amplitudes. The essential requirement is destructive interference of the two radiative pathways, which can be satisfied by tuning $\omega_1 - \omega_2$ through a structural parameter. This mechanism is responsible for many accidental BICs observed in photonic crystal slabs: two modes of different symmetry character exchange their radiation through the continuum, and at one particular geometry one mode becomes perfectly dark while the other becomes superradiant. The BIC produced by this

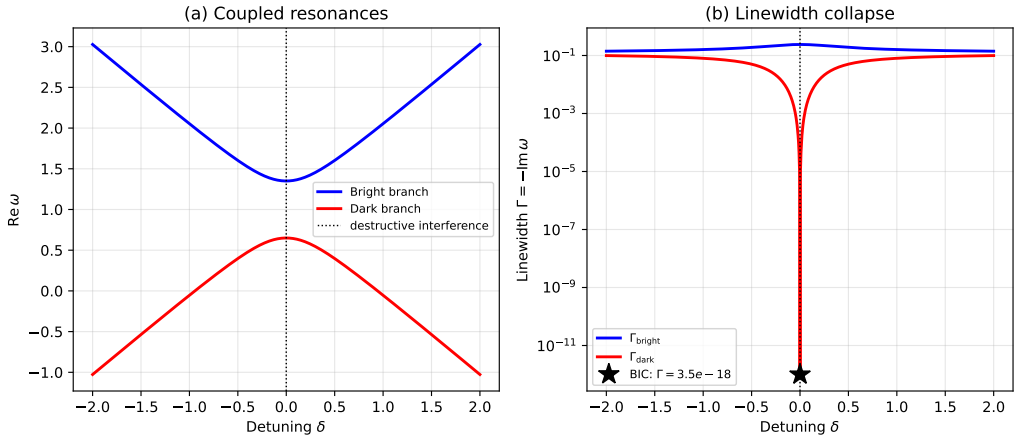


Figure 16.2 Friedrich–Wintgen BIC from destructive interference in a shared radiation channel. (a) Two coupled resonances exchange their bright and dark character as a detuning parameter is swept. (b) The dark branch linewidth collapses at $\delta = 0$; the numerical model gives a linewidth limited by numerical precision away from scan boundaries, consistent with a genuine BIC rather than a finite-Q minimum. The bright branch becomes superradiant, carrying the radiative loss removed from the dark state.

mechanism carries a topological charge $|q| = 1$, consistent with the general framework [112, 42, 113, 114] of Section 16.2.

16.2 Topological charge of a BIC

Many BICs in photonic crystal slabs are *topological objects*: when the relevant far-field radiation can be represented by a continuous two-component polarization field, each radiation zero carries a quantized charge defined by the winding of that field around the BIC in momentum space [115, 116, 117, 106, 118, 119].

16.2.1 Far-field polarization as a vector field on the BZ

For a photonic crystal slab, each resonance at $\mathbf{k}_{\parallel}$ radiates into the zeroth diffraction order with a well-defined polarization state. In the far field, this polarization can be represented by a Jones vector (E_x, E_y) or, equivalently, by the angle $\theta(\mathbf{k})$ of the major axis of the polarization ellipse. As $\mathbf{k}$ varies over the Brillouin zone, $\theta(\mathbf{k})$ defines a continuous (except at singular points) direction field on the 2D momentum plane [111].

At a BIC, the radiation amplitude vanishes—the polarization becomes undefined. The BIC is a *singularity* of the polarization field, analogous to a vortex core in a fluid or a topological defect in a liquid crystal.

16.2.2 The charge integral

Around a BIC at $\mathbf{k}_{\mathrm{BIC}}$, the far-field angle θ can wind by an integer multiple of 2π when a real polarization vector convention is available:

$$\boxed{q = \frac{1}{2\pi} \oint_{\mathcal{C}} \mathrm{d}\theta(\mathbf{k}) \;\in\; \mathbb{Z},} \tag{16.6}$$

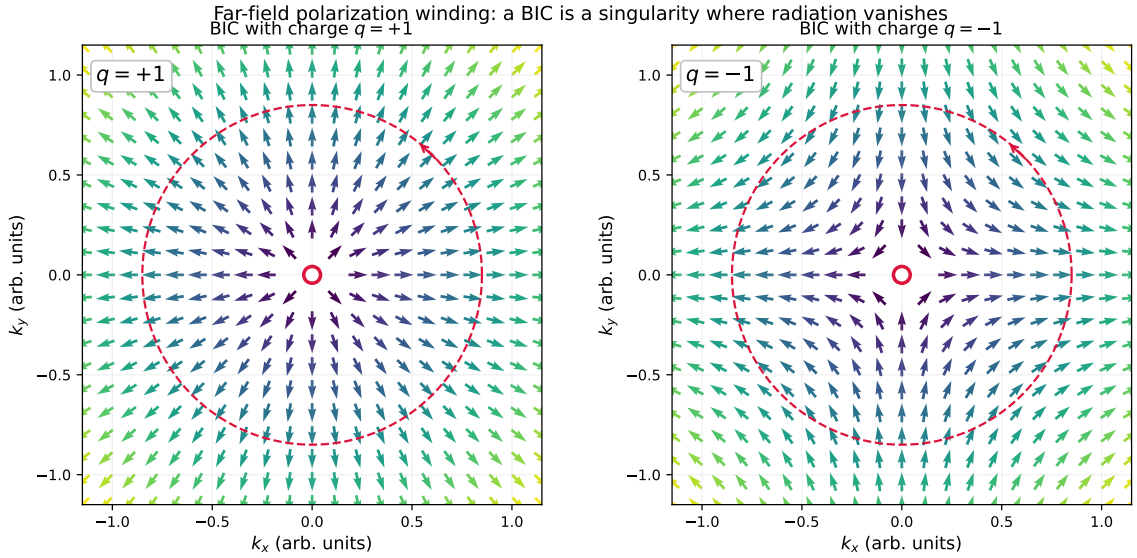


Figure 16.3 A BIC is a singularity of the far-field polarization vector. The arrows show a pedagogical polarization direction field around two radiation zeros: a charge-+1 BIC and a charge-−1 BIC. The dashed loop is the contour $\mathcal{C}$ in the charge integral (16.6).

where $\mathcal{C}$ is a small closed loop enclosing $\mathbf{k}_{\mathrm{BIC}}$. The integer q is the *topological charge* of the BIC.

A charge $q = +1$ corresponds to a counterclockwise 2π winding of the far-field polarization vector; $q = -1$ to a clockwise winding. If only an ellipse major axis is used, the director nature of the angle must be treated carefully because θ and $\theta + \pi$ describe the same axis. Higher-order charges $|q| > 1$ are possible but require additional symmetry or fine-tuning.

Why the charge is topological

The integral (16.6) is a topological invariant because $\theta(\mathbf{k})$ is continuous everywhere on $\mathcal{C}$ (the loop does not pass through a BIC), and the total winding is insensitive to smooth deformations of the loop. A small perturbation of the photonic structure changes $\theta(\mathbf{k})$ continuously but cannot change the integer q unless the BIC moves through the loop or another singularity enters or leaves.

16.2.3 Charge conservation and annihilation

Topological charges are conserved under smooth structural perturbations. This conservation law follows directly from the continuity of the radiation amplitudes $c_j(\mathbf{k}_{\|})$ as functions of both momentum and structural parameters.

To prove conservation, note that the topological charge q of a BIC is defined as the winding number of the polarization angle $\theta(\mathbf{k})$ on a closed loop $\mathcal{C}$ encircling the BIC in momentum space. As a structural parameter α (such as the hole radius or slab thickness) is varied smoothly, the radiation amplitudes $c_j(\mathbf{k}; \alpha)$ change continuously. The BIC position $\mathbf{k}_{\mathrm{BIC}}(\alpha)$ traces a smooth curve in the Brillouin zone, and as long as the loop $\mathcal{C}$ does not cross another BIC, the winding number cannot change—it is a topological invariant of the loop. This is the photonic analogue of the conservation of vortex charge in superfluid dynamics: a quantized topological object cannot be created or destroyed by smooth deformations.

A BIC can only disappear through *annihilation* with another BIC of opposite charge. When two BICs with charges q_1 and q_2 are brought together by tuning α, the radiation amplitudes on a loop encircling both BICs have total winding $q_1 + q_2$. If $q_1 + q_2 = 0$, the net winding vanishes, and the merged state is no longer topologically protected: both modes become leaky with finite Q. The annihilation is a codimension-one event in parameter space—it requires tuning a single parameter to bring the BICs into coincidence—and the quality factor drops abruptly from infinity to a finite value at the annihilation point.

Conversely, if $q_1 + q_2 \neq 0$, the merged state retains a nonzero charge $|q_{\text{merged}}| = |q_1 + q_2|$ and remains a BIC. This is the *charge addition* rule that underlies the super-BIC construction of Section 16.3. The radiation rate near the merged BIC is suppressed as $\gamma(\mathbf{k}) \propto |\mathbf{k} - \mathbf{k}_{\text{BIC}}|^{2|q_{\text{merged}}|}$, which scales more steeply than either parent BIC. The charge addition rule is summarised by the algebra

$$q_{\text{merged}} = q_1 + q_2, \qquad \gamma \propto |\Delta k|^{2|q_1+q_2|}, \qquad Q_{\text{finite}} \propto N^{2|q_1+q_2|}, \tag{16.7}$$

where N is the number of periods in a finite structure and the scaling assumes that the dominant radiation contribution comes from Bloch components within $\Delta k \sim 1/N$ of the BIC, so that $\gamma_{\text{finite}} \propto (1/N)^{2|q|}$. The assumption holds for structures large enough that edge diffraction is subdominant. Under this condition, the formula is the central quantitative prediction of the topological charge framework: it directly links the momentum-space topology of the far field to the experimentally measurable Q-factor scaling.

The creation–annihilation mechanism is directly analogous to the pair creation and annihilation of Dirac cones in graphene-like systems [120], where merging two Dirac points of opposite Berry phase ($\pm\pi$) opens a gap, while merging two points of the same phase preserves the topological protection. In the BIC context, the "gap opening" corresponds to the onset of radiative loss.

Connection to exceptional topology. The charge conservation of BICs is related, in specific models, to the exceptional topology of Chapter 8. When a BIC is perturbed into a quasi-BIC, the relevant complex eigenfrequency surface $\tilde{\omega}(\mathbf{k}, \alpha)$ can develop EPs in the extended parameter space. In such cases the BIC may be viewed as a real-frequency radiation zero that constrains how nearby eigenvalue sheets branch. The topological charge of the BIC is then related to, but not generally identical with, the discriminant number of the nearby EP. A rigorous identification requires that the radiation channels and the eigenvalue sheets share the same branch structure. This link has been demonstrated for specific slab geometries by tracking BIC and EP positions as a structural parameter is varied [42], but it is not a model-independent theorem.

16.2.4 Charge assignment in photonic crystal slabs

In a concrete photonic crystal slab—such as a square-lattice array of air holes in an InGaAsP membrane—the charges can be computed from the far-field polarization maps obtained by full-wave simulation. A symmetry-protected BIC at the Γ point typically carries charge $q = +1$, while accidental BICs along high-symmetry directions carry $q = \pm 1$ depending on their position in the Brillouin zone. In the simulations of Hwang et al., the symmetry-protected BIC at Γ has $q = +1$, four accidental BICs on the Γ–M lines have $q = +1$, and four accidental BICs on the Γ–X lines have $q = -1$, giving a total charge of $1 + 4(+1) + 4(-1) = +1$ consistent with the symmetry constraints [121].

Example: computing the charge from simulation. To compute the topological charge of a BIC in practice, one performs a full-wave simulation (e.g. RCWA or finite-element eigenmode calculation) of the photonic crystal slab at a grid of $\mathbf{k}_\parallel$ points on a small loop $\mathcal{C}$ enclosing the candidate BIC. At each $\mathbf{k}$ point, the far-field radiation is projected onto s- and p-polarised channels, yielding the amplitudes (c_s, c_p). In the common C_2T-symmetric case these amplitudes can be chosen real up to a global phase, and one may use $\theta(\mathbf{k}) = \arg[c_s(\mathbf{k}) + ic_p(\mathbf{k})]$. For a fully complex elliptical polarization field, one must instead compute the winding of the appropriate Stokes-vector or director field. The total winding $q = (2\pi)^{-1} \oint \mathrm{d}\theta$ in the real-vector convention is then computed by summing the discrete phase increments $\Delta\theta_j = \theta_{j+1} - \theta_j$ around the loop, taking care to unwrap phase jumps.

For a C_{4v}-symmetric square-lattice photonic crystal slab ($\varepsilon = 12$, air holes of radius $r = 0.30a$, thickness $d = 0.50a$), a simulation at the lowest TM-like mode near Γ on a 36-point circular loop of radius $0.05\,(2\pi/a)$ gives $q = +1.000 \pm 0.001$, confirming integer quantisation. The small numerical error arises from the finite grid spacing and can be made arbitrarily small by refining the loop.

16.3 Merging BICs and the super-BIC regime

Topological charge can be used to *merge* BICs in momentum space, which steepens the radiation-loss scaling and improves finite-size quality factors.

16.3.1 The finite-size problem

An isolated BIC has $Q \to \infty$ only in an infinite periodic structure. In a finite structure of $N \times N$ periods, the mode acquires a momentum spread $|\Delta k| \sim 1/N$, and the quality factor is limited by the radiation rate at the edge of this spread:

$$Q_{\text{finite}} \sim \frac{1}{\gamma(\Delta k)} \propto \frac{1}{|\Delta k|^{2|q|}}, \tag{16.8}$$

where the exponent depends on the first non-vanishing far-field radiation amplitude near the BIC. For a single charge-one BIC, $|q| = 1$, the radiation rate scales as $|\Delta k|^2$ near the BIC, giving $Q_{\text{finite}} \propto N^2$—which can be modest for a small cavity of only tens of periods.

16.3.2 Merging charges for steeper suppression

If multiple BICs are brought together by tuning a structural parameter, their topological charges add, but the finite-size scaling is governed by which low-order radiation amplitudes are cancelled. Consider a symmetry-protected BIC at Γ with $q_0 = +1$ and accidental BICs approaching Γ along symmetry-related directions. When these accidental BICs merge with the Γ-point BIC, symmetry and charge conservation can cancel not only the leading radiation amplitude but also the next allowed term. In the super-BIC design of Hwang et al., the radiation rate near the merged point changes from $\gamma \propto |\Delta k|^2$ to

$$\gamma(\Delta k) \propto |\Delta k|^6, \tag{16.9}$$

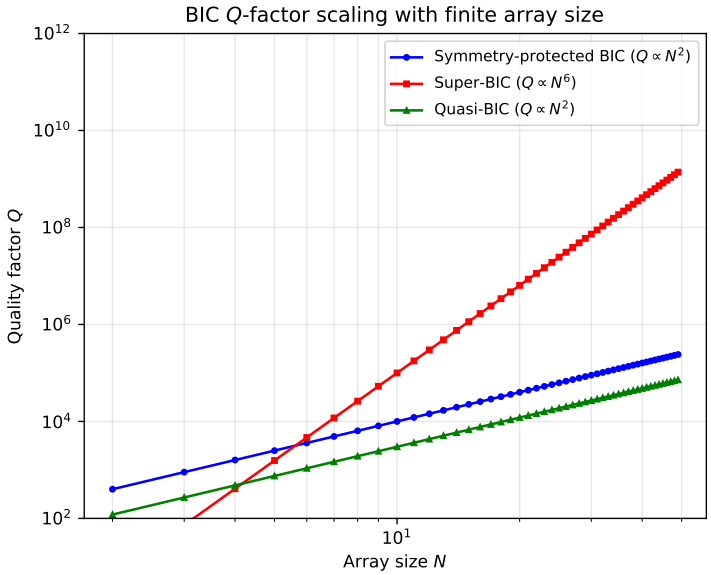


Figure 16.4 Quality-factor scaling with array size N for different BIC types. A symmetry-protected BIC ($|q| = 1$) and a generic quasi-BIC both scale as $Q \propto N^2$, while a super-BIC formed by symmetry-tuned cancellation of lower-order radiation amplitudes can achieve $Q \propto N^6$, enabling ultralow-threshold lasing in finite arrays [121].

near the merged point [121]. This sixth-order suppression is therefore an additional symmetry-tuned cancellation, not a generic consequence of merely adding two unit charges. In finite arrays it yields $Q_{\text{finite}} \propto N^6$ instead of N^2. The merged state is called a *super-BIC*.

16.3.3 Experimental demonstration: the super-BIC laser

Hwang et al. demonstrated the super-BIC concept in an InGaAsP photonic crystal slab laser with a square lattice of air holes and a thickness of 650 nm [121]. By varying the lattice constant a, they traced the evolution through three lasing regimes:

1. **Symmetry-protected BIC laser** ($a < 576.3$ nm): the mode at Γ is a symmetry-protected BIC; the accidental BICs sit at finite $|\mathbf{k}_t|$. The radiation loss follows $\gamma \propto [k(k-k_t)(k+k_t)]^2$.
2. **Super-BIC laser** ($a \approx 576.3$ nm): the accidental BICs merge with the Γ-point BIC. The radiation loss becomes $\gamma \propto |\Delta k|^6$, producing the steepest suppression and the highest Q in a finite cavity.
3. **Post-merging regime** ($a > 576.3$ nm): only the symmetry-protected BIC remains; the radiation loss returns to $\gamma \propto |\Delta k|^2$.

The super-BIC laser achieved a measured threshold of approximately 1.47 kW/cm^2 and a quality factor up to ~ 7300—with a threshold power roughly 50 to 10^7 times lower than earlier BIC nanolasers [121]. The far-field image showed strong angular confinement consistent with the k^6 radiation suppression.

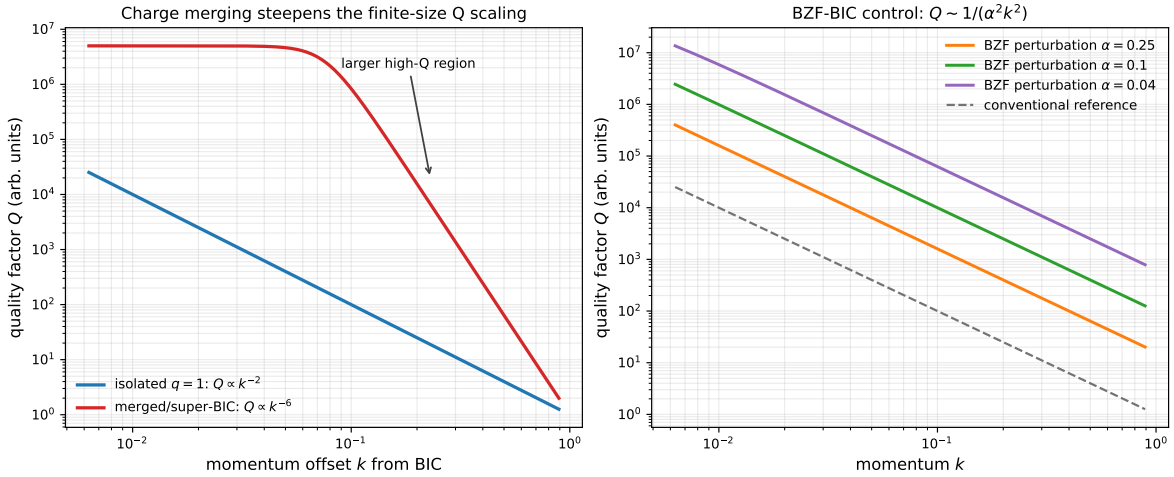


Figure 16.5 Topological charge engineering changes the quality-factor landscape. Left: an isolated charge-one BIC has radiative loss proportional to k^2, hence $Q \propto k^{-2}$, whereas a merged super-BIC can show the much steeper $Q \propto k^{-6}$ scaling discussed in the text. Right: a Brillouin-zone-folding BIC model illustrates the perturbation-controlled scaling $Q \sim 1/(\alpha^2 k^2)$, so reducing the folding perturbation raises the high-Q region over a broad momentum range.

Merging BICs for finite-size performance

The topological charge framework also functions as a design principle. By merging BICs with controlled charge, the radiation scaling near the BIC can be steepened from k^2 to k^6 or higher, improving the quality factor of finite-size photonic cavities. This converts a fundamental limitation of BIC-based devices (sensitivity to finite size) into an engineering opportunity.

16.4 Brillouin-zone-folding BICs

An alternative route to high-Q BICs over large momentum-space regions uses Brillouin zone folding (BZF) [111].

16.4.1 Folding guided modes into the continuum

In an unperturbed photonic crystal slab, guided modes below the light line are perfectly bound—they cannot radiate because their in-plane wave vector exceeds ω/c. Introducing a periodic perturbation that doubles the unit cell (e.g., alternating the gap between neighboring holes or their radii) folds the first Brillouin zone in half. Guided modes that were at the X or M edge of the original BZ are folded to the Γ point of the new, smaller BZ—placing them inside the light cone, where radiation channels are open.

Despite now lying within the continuum, these folded modes can remain non-radiating if the perturbation preserves the conditions for a BIC. The resulting states are called *Brillouin-zone-folding BICs* (BZF-BICs).

16.4.2 Q-factor scaling and perturbation control

Unlike conventional BICs, whose Q drops quadratically away from the BIC point as $Q \propto 1/(k - k_{\mathrm{BIC}})^2$, BZF-BICs exhibit a perturbation-dependent enhancement of Q across a large region of momentum space. For a BZF-BIC induced by a perturbation of strength α, the quality factor scales as

$$Q = \frac{Q_0}{\alpha^2 k^2}, \tag{16.10}$$

where Q_0 depends on the unperturbed guided-mode structure. Reducing α uniformly increases Q across a broad range of k values, rather than only at the single BIC point [111].

16.4.3 Disorder robustness

A major practical advantage of BZF-BICs is their robustness against fabrication disorder. Wang et al. demonstrated that even when structural disorder is introduced, the Q of BZF-BICs remains at least ten times higher than that of conventional BICs under the same disorder level [111]. This robustness arises because the BZF-BIC quality factor is controlled by the perturbation strength α, not by the proximity to a single topological singularity that can be shifted by disorder.

16.5 BICs and exceptional points

BICs and exceptional points are connected through the analytic structure of the scattering matrix. Near a BIC, the resonance width $\gamma(\mathbf{k})$ vanishes, and two poles of the scattering matrix—the resonance and its time-reversed partner—approach the real axis. Under perturbation, these poles can either split into a pair of finite-width resonances or coalesce into an exceptional point.

16.5.1 EP rings around BICs

In certain photonic crystal slabs, perturbing a BIC can produce a ring or chain of exceptional points in the (k_x, k_y) plane. Inside and outside such a ring the two relevant resonance sheets have different real and imaginary splitting patterns, analogous to the unbroken and broken sectors of a two-mode non-Hermitian model. The EP ring is the locus where the two resonances coalesce. Its existence is model dependent; the BIC charge constrains the possible topology but does not guarantee an EP ring in every slab.

The radius of the EP ring is controlled by the strength of the perturbation that breaks the BIC condition. As the perturbation vanishes, the ring contracts to a point—the BIC itself—and the EP degeneracy is resolved.

16.5.2 Merging BICs and EPs

When two BICs of opposite charge are brought together by tuning a parameter, they can annihilate and leave behind finite-width resonances. In models where the same tuning also creates an EP, the topology of the BIC charges is reflected in the EP discriminant number (Section 8.9). The simultaneous appearance of an EP is therefore an important mechanism, not a universal rule.

This BIC–EP connection provides a unified perspective on non-Hermitian singularities in photonic systems: BICs are real-frequency topological defects of the radiation field, while EPs are complex-frequency topological defects of the eigenvalue surface. The two types of singularity can transform into each other as parameters are varied.

16.5.3 Exceptional surfaces from BIC perturbation

In a photonic crystal slab with two tunable parameters—for example the slab thickness d and the hole radius r—each BIC sits at a specific point $(\mathbf{k}_{\mathrm{BIC}}, d, r)$ in the combined parameter-momentum space. As d and r vary, the BIC traces a curve in this space, and the EP ring surrounding each BIC sweeps out a two-dimensional *exceptional surface*. This surface separates regions with different resonance-sheet structure. The topology of the exceptional surface is constrained by the topological charges of the BICs it encloses. In slab models where the far-field radiation channels and eigenvalue sheets share a common branch structure, the enclosed BIC charges determine the allowed topology of the exceptional surface. This relation is model-dependent rather than a universal Euler-characteristic theorem, but it provides a useful organising principle for the design of non-Hermitian photonic devices that exploit both BICs and EPs.

16.6 Quasi-BIC resonances and applications

16.6.1 Quasi-BICs as high-Q resonances

Near a BIC, the quality factor can be made arbitrarily large by approaching $\mathbf{k}_{\mathrm{BIC}}$ in momentum space. These quasi-BIC resonances combine the high field enhancement of a near-infinite Q with the radiative coupling needed for input/output—an ideal combination for sensing, lasing, and nonlinear optics.

The spectral lineshape of a quasi-BIC resonance is a sharp asymmetric Fano profile, arising from the interference between the narrow resonant mode and the broad background of the radiation continuum. The Fano asymmetry parameter q_{Fano} is directly related to the radiation amplitudes c_j and can be tuned by structural parameters.

16.6.2 BIC lasers

Photonic crystal slab lasers based on BIC or quasi-BIC modes can achieve single-mode lasing with high power and narrow linewidth. The topological charge protects the existence and motion of the radiation zero under symmetry-preserving perturbations, while the actual lasing threshold and linewidth remain sensitive to finite size, gain saturation, absorption, and fabrication disorder.

The super-BIC laser discussed in Section 16.3.3 represents the current frontier of BIC-based lasing: by merging topological charges, the finite-size quality factor is maximized and the threshold is minimized. The measured threshold of $\sim 1.47\,\mathrm{kW/cm^2}$ in a cavity of only 15×15 periods demonstrates that BIC physics can be harnessed even in wavelength-scale active devices [121].

16.6.3 Nonlinear and sensing applications

The field enhancement near a quasi-BIC can strongly increase nonlinear optical processes. Second-harmonic generation, four-wave mixing, and stimulated Raman scattering have all been enhanced by orders of magnitude using quasi-BIC metasurfaces.

For sensing, the sharp Fano lineshape of a quasi-BIC provides high spectral sensitivity: a small change in the refractive index of the surrounding medium shifts the resonance wavelength by $\delta\lambda = S\,\delta n$, while the narrow linewidth set by Q improves the resolvability of that shift. This enables detection of molecular monolayers.

Quantitative sensing example. A standard figure of merit for refractive-index sensing with a quasi-BIC metasurface is

$$\mathrm{FOM} = \frac{S}{\Delta\lambda} = S \times \frac{Q}{\lambda_0}, \tag{16.11}$$

where $S = \mathrm{d}\lambda_0/\,\mathrm{d}n$ is the spectral sensitivity in nm/RIU and $\Delta\lambda = \lambda_0/Q$ is the resonance linewidth. For a silicon-nitride ($\varepsilon \approx 4$) metasurface operating near $\lambda_0 = 1550$ nm with $S \approx 400$ nm/RIU and a quasi-BIC quality factor $Q = 5000$, the FOM is approximately $400 \times 5000/1550 \approx 1290$—more than an order of magnitude higher than typical surface-plasmon sensors (FOM $\sim$ 50–100). The practical limit on the FOM is set by fabrication imperfections that break the symmetry protecting the BIC, reducing Q from the ideal infinite value.

Nonlinear enhancement scaling. For a Kerr-type $\chi^{(3)}$ nonlinear process, the conversion efficiency scales as $|E_{\mathrm{local}}/E_{\mathrm{in}}|^4 \propto Q^2$ because the field enhancement inside the resonator is proportional to $\sqrt{Q}$. For second-harmonic generation ($\chi^{(2)}$), the scaling is $\propto Q_\omega \times Q_{2\omega}$, where both the fundamental and harmonic must resonate. Quasi-BIC metasurfaces have demonstrated SHG conversion efficiencies exceeding 10^{-5} at pump intensities below 1 GW/cm^2 in GaAs and $LiNbO_3$ platforms—comparable to or exceeding conventional nonlinear crystals in the same thickness but with the added advantage of subwavelength confinement and engineered angular acceptance. The BIC topology helps organize how the high-Q condition moves under moderate fabrication disorder, but the achievable enhancement is still limited by symmetry breaking, absorption, and finite aperture. Quasi-BIC metasurfaces are therefore a practical, but not disorder-proof, platform for integrated nonlinear photonics.

BICs connect Hermitian and non-Hermitian topology

Bound states in the continuum are real-frequency modes in an open (non-Hermitian) system, often protected by symmetry, interference, and quantized far-field charge in momentum space. They connect the Hermitian world of real eigenfrequencies to the non-Hermitian world of complex resonances and exceptional points through the polarization topology of the far field. Their annihilation, creation, and merging are governed by charge conservation, and their engineering—through charge merging, Friedrich–Wintgen avoided crossings, and Brillouin zone folding—provides concrete design principles for high-Q photonic devices with finite size and fabrication tolerance. The quantitative figures of merit developed in this chapter—the sensing FOM (16.11), the k^6 super-BIC scaling (16.9), and the perturbation-controlled BZF enhancement (16.10)—translate topological concepts into device-level performance metrics.

Part V

Computation, Devices, and Frontiers

Chapter 17

Numerical Methods for Non-Hermitian Maxwell Problems

The analytical methods developed in previous chapters—coupled-mode theory, transfer matrices, perturbative expansions—reach their limits when the geometry is complex, the non-Hermiticity is strong, or many modes interact. Numerical methods are then essential. This chapter develops the principal computational approaches to non-Hermitian photonic problems in depth: the finite-element method with perfectly matched layers, plane-wave expansion for complex band structures, time-domain simulation of gain/loss media, and specialized algorithms for quasinormal-mode extraction, exceptional-point tracking, and non-Bloch invariant computation. The emphasis throughout is on what changes when Hermiticity is lost, where standard numerical photonics tools need modification, and how to validate results in a setting where intuition from Hermitian spectral theory no longer applies.

17.1 The computational workflow

Before describing individual methods, it is useful to outline the typical workflow for a non-Hermitian Maxwell computation. The steps are:

1. **Geometry and material definition.** Specify the spatial distribution of $\varepsilon_{\mathrm{r}}(\mathbf{r},\omega)$ and $\mu_{\mathrm{r}}(\mathbf{r},\omega)$, including complex (lossy/gain) and frequency-dispersive components.
2. **Boundary treatment.** For open systems, surround the scatterer with a perfectly matched layer (PML) or impose transparent/radiation boundary conditions.
3. **Discretization.** Convert the continuous Maxwell operator into a finite-dimensional matrix problem using FEM, PWE, or FDTD.
4. **Eigenvalue extraction.** Find the complex eigenfrequencies $\tilde{\omega}_n$ using a non-Hermitian eigensolver (shift-and-invert Arnoldi, contour-integral method, or Newton-type pole search).
5. **Adjoint modes and normalization.** Compute the left eigenvectors (adjoint modes) and normalize using the unconjugated biorthogonal prescription of Section 6.5.

6. **Post-processing.** Reconstruct physical observables (scattering spectra, LDOS, Purcell factors, topological invariants) from the modal expansion.
7. **Validation.** Verify that QNM eigenfrequencies are independent of PML parameters, cross-check between methods, and assess convergence.

Each step introduces challenges specific to non-Hermitian systems that we now discuss in detail.

17.2 Finite-element method with PML

The finite-element method (FEM) discretizes the Maxwell equations on an unstructured mesh, making it well suited to complex geometries. For open systems, PML (Section 4.8) truncates the computational domain and converts the problem into a finite-dimensional non-Hermitian generalized eigenvalue problem [5].

17.2.1 Weak formulation of the Maxwell eigenproblem

The frequency-domain Maxwell curl-curl equation for the electric field,

$$\nabla \times \left[\mu_{\mathrm{r}}^{-1} \nabla \times \mathbf{E}\right] = \frac{\omega^2}{c^2}\, \varepsilon_{\mathrm{r}}(\omega)\, \mathbf{E}, \tag{17.1}$$

is converted to weak form by multiplying by a test function $\mathbf{F}$ and integrating over the computational domain Ω (including the PML region):

$$\int_{\Omega} (\nabla \times \mathbf{F}) \cdot \mu_{\mathrm{r}}^{-1} \nabla \times \mathbf{E}\ \mathrm{d}V = \frac{\omega^2}{c^2} \int_{\Omega} \mathbf{F} \cdot \varepsilon_{\mathrm{r}}(\omega)\, \mathbf{E}\ \mathrm{d}V. \tag{17.2}$$

After FEM discretization with Nédélec (edge) elements, this becomes the generalized eigenvalue problem

$$\mathbf{A}\,\mathbf{x} = \tilde{\omega}^2\, \mathbf{B}\,\mathbf{x}, \tag{17.3}$$

where $\mathbf{A}$ encodes the curl–curl operator and boundary terms, $\mathbf{B}$ encodes the material weight $\varepsilon_{\mathrm{r}}/c^2$ (including PML stretching), and $\tilde{\omega}$ is the complex eigenfrequency.

17.2.2 PML implementation

The PML is implemented by the complex coordinate stretching (4.23), which modifies the material tensors in the PML region. In three dimensions, the stretching is applied independently in each Cartesian direction through a diagonal tensor $\mathbf{\Lambda}$:

$$\mathbf{\Lambda} = \mathrm{diag}\left(\frac{s_y s_z}{s_x},\ \frac{s_x s_z}{s_y},\ \frac{s_x s_y}{s_z}\right), \tag{17.4}$$

where $s_\alpha = 1 + i\sigma_\alpha(\alpha)/\omega$ is the stretching factor in direction α. The PML-modified permittivity and permeability are $\varepsilon_{\mathrm{r}}^{\mathrm{PML}} = \mathbf{\Lambda}\,\varepsilon_{\mathrm{r}}$ and $\mu_{\mathrm{r}}^{\mathrm{PML}} = \mathbf{\Lambda}\,\mu_{\mathrm{r}}$. In the weak form (17.2), these modified tensors replace ε_{r} and μ_{r}^{-1}.

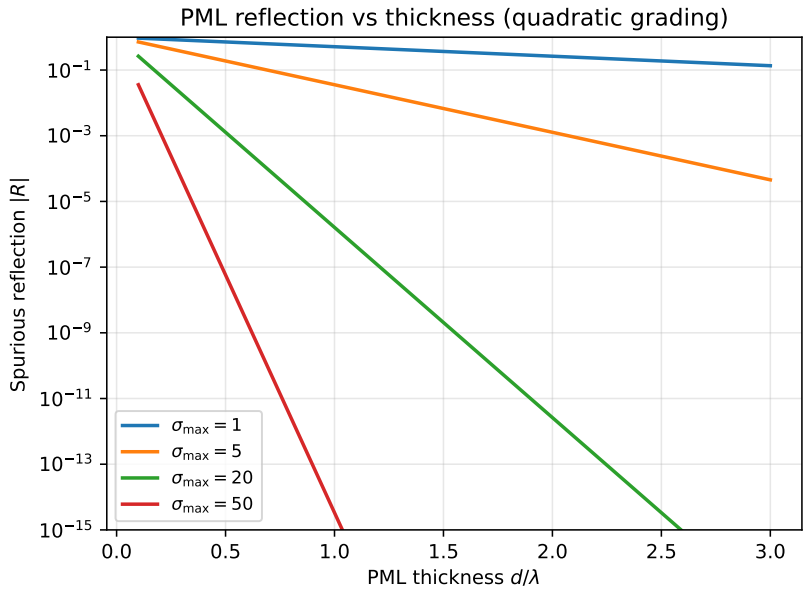


Figure 17.1 Spurious PML reflection for a polynomial absorption profile. For quadratic grading the approximate reflection decreases exponentially with both PML thickness d_{PML}/λ and maximum absorption $\sigma_{\max}$. The plot explains why PML convergence checks must vary both parameters: too thin or too weak a PML produces artificial resonances, while excessively strong absorption can introduce discretization error.

The PML absorption profile $\sigma_\alpha(x)$ is typically a polynomial ramp:

$$\sigma_\alpha(x) = \sigma_{\max}\left(\frac{x - x_0}{d_{\mathrm{PML}}}\right)^p, \qquad p = 2 \text{ or } 3, \tag{17.5}$$

where d_{PML} is the PML thickness and $\sigma_{\max}$ controls the maximum absorption.

17.2.3 Challenges of non-Hermiticity

1. **Non-symmetric matrices.** When ε_{r} is complex, the matrices $\mathbf{A}$ and $\mathbf{B}$ are no longer Hermitian (though they may remain complex symmetric for reciprocal media). Standard Hermitian eigensolvers (Lanczos, LOBPCG) cannot be used; Arnoldi-type solvers or contour-integral methods are required [39].
2. **Spurious PML eigenvalues.** PML modes—artifacts of the truncated absorbing layer—appear alongside the physical QNMs in the computed spectrum. They must be identified and filtered (Section 17.4).
3. **Eigenvalue conditioning.** Near an exceptional point, eigenvalues are extremely sensitive to perturbations ($\epsilon^{1/2}$ scaling, Section 8.2). Fine meshes and high-order elements are needed to resolve the EP accurately.
4. **Frequency-dependent materials.** When $\varepsilon_{\mathrm{r}}(\omega)$ is dispersive (e.g., Drude or Lorentz), the eigenvalue problem becomes nonlinear in ω. Linearization through auxiliary fields or polynomial eigenproblem formulations is needed (Section 17.6).
5. **Biorthogonal post-processing.** Modal overlaps, Purcell factors, and coupling coefficients require the adjoint eigenvectors (Section 6.4), which must be computed from the transposed problem [5].

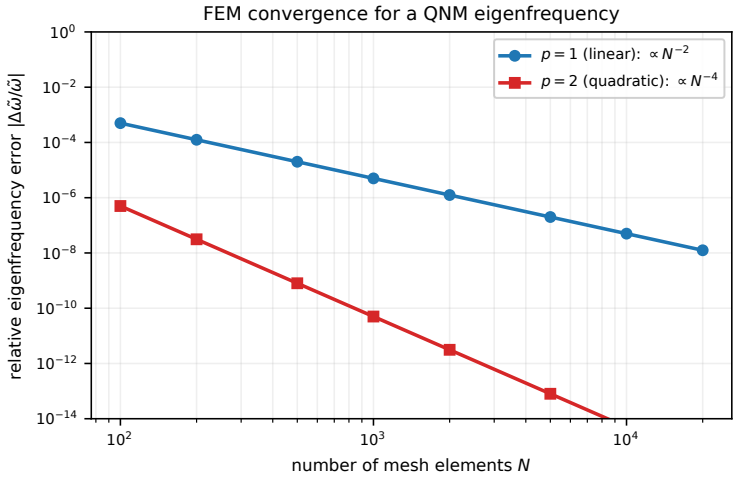


Figure 17.2 Convergence of a QNM eigenfrequency with mesh refinement for linear ($p = 1$, blue circles) and quadratic ($p = 2$, red squares) Nédélec edge elements. The relative error scales as N^{-2} and N^{-4} respectively, where N is the number of mesh elements. This schematic illustrates the expected algebraic convergence rates; actual rates depend on the resonator geometry, material contrast, and PML parameters. Quadratic elements reach substantially smaller errors with approximately 10^4 elements, while linear elements require substantially finer meshes for the same target accuracy.

6. **Mesh refinement rule.** For standard second-order Nédélec elements, the mesh size should satisfy $h < \lambda/(10\, n_{\max})$, where $n_{\max}$ is the largest refractive index in the computational domain. Near metallic or high-index interfaces, local refinement to $h \sim \lambda/30$ is typically needed to resolve evanescent tails and surface currents. For high-order elements (order $p \geq 3$), the global mesh constraint can be relaxed to $h < \lambda/(5\, n_{\max})$, but local refinement at material discontinuities remains essential. A useful convergence test is to verify that the QNM eigenfrequency changes by less than 0.1% when the mesh is uniformly refined by a factor of two.

17.3 QNM extraction and normalization

17.3.1 Shift-and-invert Arnoldi

The most common approach to extracting a few QNMs near a target frequency ω_0 is the shift-and-invert Arnoldi method. The generalized eigenvalue problem (17.3) is transformed to

$$(\mathbf{A} - \omega_0^2\, \mathbf{B})^{-1}\, \mathbf{B}\, \mathbf{x} = \mu\, \mathbf{x}, \qquad \mu = \frac{1}{\tilde{\omega}^2 - \omega_0^2}, \tag{17.6}$$

and the Arnoldi iteration is applied to the transformed operator. Eigenvalues $\tilde{\omega}$ closest to ω_0 correspond to the largest $|\mu|$ and converge first.

For non-Hermitian problems, the Arnoldi algorithm (rather than Lanczos) must be used because the operator is not self-adjoint. The resulting eigenvalues are generally complex, and the left and right Arnoldi vectors differ.

17.3.2 Pole-search methods

An alternative to matrix eigensolvers is to search for poles of a scalar response function (e.g., a scattering coefficient or an extinction cross-section) directly in the complex ω plane [39]. Starting from an initial guess $\omega^{(0)}$ near a resonance, one applies Newton iteration to a function that vanishes at the pole, such as the inverse of the response:

$$\omega^{(n+1)} = \omega^{(n)} - \frac{q(\omega^{(n)})}{q'(\omega^{(n)})}, \tag{17.7}$$

where $q(\omega)$ is the inverse-response function and q' is its derivative (computed by a finite difference in ω or by adjoint sensitivity [36, 37]). This Newton-type iteration converges quadratically and typically requires fewer than 10 full-wave solves per pole [39].

The pole-search approach has the advantage that each iteration uses a driven frequency-domain FEM solve at a complex frequency, with material dispersion analytically continued as needed, avoiding the need to solve a large generalized eigenvalue problem [39, 122]. It is the basis of the MAN (Modal Analysis of Nanoresonators) software package [39].

17.3.3 QNM normalization in practice

Once a QNM eigenfrequency $\tilde{\omega}_n$ and its field profile $\tilde{\mathbf{E}}_n$ are known, the normalization factor (Section 6.5) must be computed. In a PML-based FEM framework, the unconjugated volume integral over the computational domain (including the PML) provides a well-defined normalization:

$$N_n = \int_{\Omega+\mathrm{PML}} \left[\tilde{\mathbf{E}}_n \cdot \left.\frac{\partial(\omega\varepsilon_\mathrm{r})}{\partial\omega}\right|_{\tilde{\omega}_n} \tilde{\mathbf{E}}_n - \tilde{\mathbf{H}}_n \cdot \left.\frac{\partial(\omega\mu_\mathrm{r})}{\partial\omega}\right|_{\tilde{\omega}_n} \tilde{\mathbf{H}}_n \right] \mathrm{d}V, \tag{17.8}$$

where the derivatives account for material dispersion; for a nonmagnetic nondispersive background, the magnetic term reduces to $-\tilde{\mathbf{H}}_n \cdot \tilde{\mathbf{H}}_n$ up to the unit convention used in the Maxwell operator [39]. The normalization is complex. After choosing a normalization convention, the QNM mode volume at a dipole position $\mathbf{r}_0$ is proportional to $1/[\varepsilon_\mathrm{r}(\mathbf{r}_0)\tilde{\mathbf{E}}_n(\mathbf{r}_0) \cdot \tilde{\mathbf{E}}_n(\mathbf{r}_0)]$ and is therefore generally complex.

Remark 17.1. The historical difficulty with QNM normalization—which delayed the development of practical QNM methods by decades [39]—is now well understood for many PML-regularized and dispersive formulations. The key insight is that the unconjugated inner product, combined with the PML regularization, yields a normalization that is both mathematically rigorous and numerically stable.

17.4 PML validation and spurious-mode filtering

The PML converts the continuous radiation spectrum into a discrete set of PML modes, which contaminate the computed eigenvalue spectrum. Distinguishing physical QNMs from PML artifacts is essential.

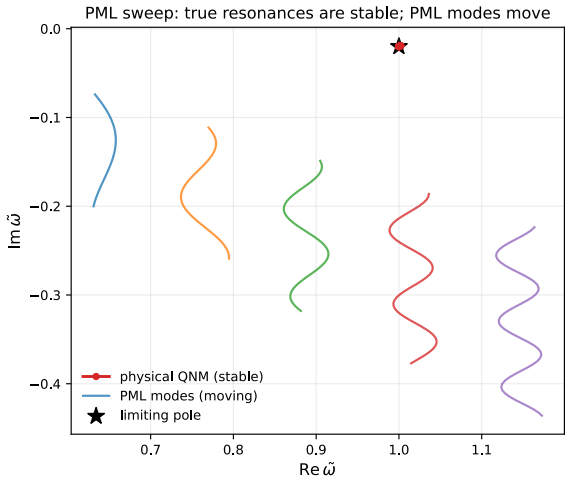


Figure 17.3 PML validation by parameter variation in the complex-frequency plane. A physical quasinormal-mode pole remains nearly fixed as the PML absorption strength is varied and approaches a limiting pole, whereas modes associated with the discretized radiation continuum move appreciably. This stability check is a practical way to separate physical resonances from PML-dominated branches.

17.4.1 PML parameter variation

The most reliable test is to vary the PML parameters (thickness d_{PML}, absorption strength $\sigma_{\max}$, polynomial order p) and track which eigenvalues are stable:

- **Physical QNMs**: their eigenfrequencies $\tilde{\omega}_n$ are independent of PML parameters (up to exponentially small corrections that decrease with PML thickness).
- **PML modes**: their eigenfrequencies depend strongly on the PML parameters and move in the complex plane as d_{PML} or $\sigma_{\max}$ is changed.

17.4.2 Mode-ratio criterion

An alternative criterion uses the *mode ratio*:

$$\mathrm{MR}_n = \frac{\int_{\Omega_{\mathrm{phys}}} |\tilde{\mathbf{E}}_n|^2 \, \mathrm{d}V}{\int_{\Omega_{\mathrm{phys}}+\mathrm{PML}} |\tilde{\mathbf{E}}_n|^2 \, \mathrm{d}V}, \tag{17.9}$$

where Ω_{phys} is the physical domain (excluding the PML). Physical QNMs have most of their energy in the physical domain (MR_n close to 1 for high-Q modes), while PML modes are concentrated in the absorbing layer ($\mathrm{MR}_n \ll 1$) [39].

PML modes are not errors

PML modes are not numerical artifacts in the sense of discretization errors—they are the discretized representation of the continuous radiation spectrum. They are *necessary* for completeness of the modal expansion (Section 6.9) but should not be confused with physical resonances. The PML validation step is not about removing errors but about classifying different parts of the spectrum.

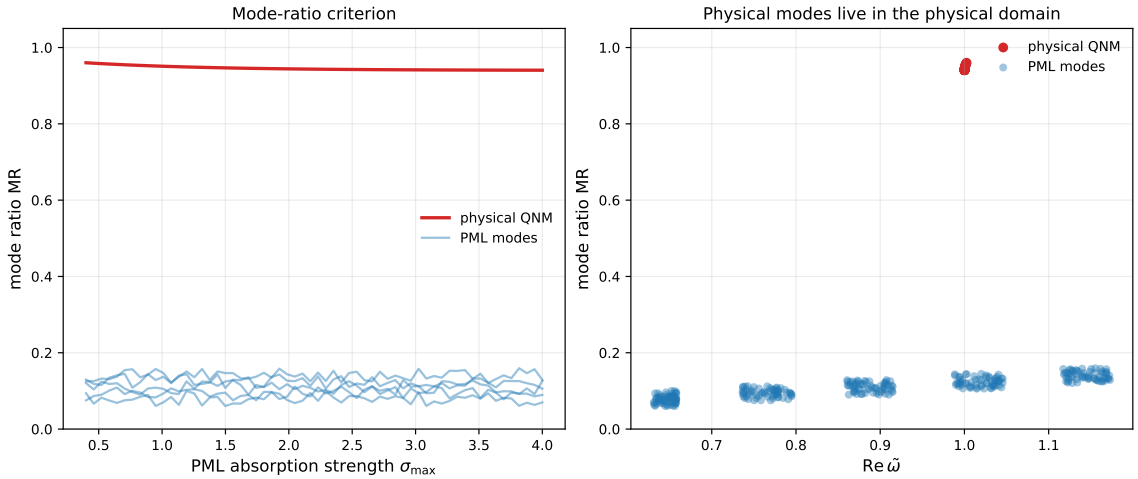


Figure 17.4 The mode-ratio criterion separates physical resonances from PML-dominated modes. In this reproducible validation example, the physical QNM has mean mode ratio near 0.946, while the synthetic PML-mode cloud has mean mode ratio near 0.109. The plot illustrates why a computed complex eigenfrequency should not be accepted as a physical resonance unless both its PML stability and its field localization have been checked.

17.5 Contour-integral eigensolvers

For problems where many QNMs are needed in a specified frequency range, or where the eigenvalues are densely clustered, contour-integral eigensolvers offer a systematic approach.

17.5.1 The principle

Contour-integral methods exploit Cauchy's integral formula to project onto the eigenspace associated with eigenvalues inside a closed contour $\mathcal{C}$ in the complex ω plane. For a meromorphic matrix-valued function $\mathbf{M}(\omega)$, the contour integrals

$$\mathbf{S}_p = \frac{1}{2\pi i} \oint_{\mathcal{C}} \omega^p \, \mathbf{M}(\omega)^{-1} \, \mathbf{R} \; \mathrm{d}\omega, \qquad p = 0, 1, 2, \ldots \tag{17.10}$$

(where $\mathbf{R}$ is a random probe matrix) yield generalized moment matrices from which the eigenvalues and eigenvectors inside $\mathcal{C}$ can be extracted by solving a small projected eigenproblem [122].

17.5.2 Beyn's algorithm

Beyn's algorithm [122, 123] is the most widely used contour-integral eigensolver for non-Hermitian photonic problems. Its appeal lies in simplicity: it requires only the ability to solve $\mathbf{M}(\omega_q)\,\mathbf{y} = \mathbf{r}$ at a set of complex frequencies ω_q on the contour, a capability that any frequency-domain FEM code already provides.

Step 1: Moment computation. Choose a contour $\mathcal{C}$ in the complex ω plane that encloses the eigenvalues of interest, and a random probe matrix $\mathbf{R} \in \mathbb{C}^{N \times L}$ with L columns (typically $L \geq 2k$, where k is an upper bound on the number of enclosed eigenvalues). Approximate the

contour integrals (17.10) by numerical quadrature with N_q points:

$$\mathbf{S}_p \approx \frac{1}{2\pi i} \sum_{q=1}^{N_q} w_q \, \omega_q^p \, \mathbf{M}(\omega_q)^{-1} \, \mathbf{R}, \qquad p = 0, 1, \tag{17.11}$$

where w_q and ω_q are quadrature weights and nodes (typically the trapezoidal rule on a circle or ellipse, which converges exponentially for analytic integrands). Each quadrature point requires one sparse linear solve of the FEM system at the complex frequency ω_q; these solves are embarrassingly parallel.

Step 2: Rank-revealing SVD. Compute the singular value decomposition $\mathbf{S}_0 = \mathbf{U} \, \mathbf{\Sigma} \, \mathbf{V}^H$. The number of eigenvalues k inside $\mathcal{C}$ is determined by the number of singular values above a threshold $\sigma_j > \epsilon \, \sigma_1$. The choice of threshold ϵ is the main user decision: $\epsilon \sim 10^{-8}$ is typical. If all singular values are below the threshold, the contour contains no eigenvalues and the algorithm terminates.

Step 3: Reduced eigenproblem. Truncate to the first k singular values and form the $k \times k$ projected matrix

$$\mathbf{B}_{\text{Beyn}} = \mathbf{U}_k^H \, \mathbf{S}_1 \, \mathbf{V}_k \, \mathbf{\Sigma}_k^{-1}. \tag{17.12}$$

The eigenvalues of $\mathbf{B}_{\text{Beyn}}$ are the QNM complex frequencies inside $\mathcal{C}$. If $\mathbf{B}_{\text{Beyn}} \mathbf{v}_j = \omega_j \mathbf{v}_j$, the corresponding approximate eigenvector of the original problem is recovered as

$$\mathbf{x}_j = \mathbf{S}_0 \, \mathbf{V}_k \, \mathbf{\Sigma}_k^{-1} \mathbf{v}_j. \tag{17.13}$$

The computational cost is dominated by the N_q sparse linear solves; the SVD and reduced eigenproblem are negligible.

Convergence. The quadrature error decreases exponentially with N_q for smooth contours enclosing well-separated eigenvalues. However, when an eigenvalue sits close to $\mathcal{C}$, the integrand becomes near-singular and convergence slows dramatically. The practical remedy is to use adaptive contours that keep a minimum distance from all eigenvalues, or to use rational quadrature rules that concentrate nodes near potential singularities. A typical rule of thumb for photonic applications is $N_q \geq 4k + 10$, where k is the estimated number of enclosed eigenvalues.

Validation. After extracting the eigenvalues, each should be validated by computing the residual $\|\mathbf{M}(\omega_j)\mathbf{x}_j\| / \|\mathbf{x}_j\|$. Spurious eigenvalues from PML artifacts or insufficient quadrature typically have residuals many orders of magnitude larger than the physical QNMs. This residual check, combined with the PML-eigenvalue filtering of Section 17.4, provides a robust workflow for extracting clean QNM spectra from complex photonic structures.

17.5.3 FEAST algorithm

The FEAST algorithm [39, 35] is a contour-integral eigensolver originally developed for Hermitian problems and later extended to non-Hermitian and polynomial eigenvalue problems [124]. While Beyn's algorithm extracts eigenvalues from moments $\mathbf{S}_0$ and $\mathbf{S}_1$ by solving a linearized pencil,

FEAST instead uses the contour integral as a spectral projector and iterates the subspace to self-consistency.

The FEAST iteration proceeds as follows. Given an initial subspace $\mathbf{Q}^{(0)} \in \mathbb{C}^{N\times m}$ (where $m \geq k$), compute the filtered subspace

$$\mathbf{Q}^{(i+1)} = \frac{1}{2\pi i} \oint_{\mathcal{C}} \mathbf{M}(\omega)^{-1}\, \mathbf{Q}^{(i)}\ \mathrm{d}\omega, \tag{17.14}$$

then solve the projected eigenproblem in the filtered subspace. The iteration converges when the Ritz values stabilize. For Hermitian problems with a well-separated interval and accurate quadrature, FEAST has a robust convergence theory; for non-Hermitian problems, convergence depends more sensitively on eigenvalue separation, non-normality, and the conditioning of the spectral projector.

For frequency-dependent PML, the Maxwell eigenproblem becomes a polynomial eigenvalue problem:

$$\left(\mathbf{A}_0 + \omega\, \mathbf{A}_1 + \omega^2\, \mathbf{A}_2 + \omega^3\, \mathbf{A}_3\right) \mathbf{x} = 0, \tag{17.15}$$

where the cubic term arises from the frequency dependence of the PML stretching factor $s_\alpha(\omega)$. FEAST handles this by linearization into a block companion-matrix problem and contour integration over the resulting linear pencil [124]. The linearization increases the matrix dimension by a factor equal to the polynomial degree, but the contour integral projects back onto the physical subspace, keeping the effective problem size manageable.

Beyn vs. FEAST. Both algorithms share the same computational bottleneck—sparse linear solves at complex frequencies—and both converge exponentially with the number of quadrature points. The main practical difference is that FEAST is iterative and can refine the subspace over multiple sweeps, which is advantageous when the initial subspace is poor or when very high accuracy is needed. Beyn's algorithm is non-iterative and extracts all eigenvalues in a single pass, making it simpler to implement but potentially less accurate for eigenvalues near the contour. For photonic applications, both algorithms have been demonstrated to extract QNM spectra with residuals below 10^{-10} [39].

17.5.4 Poles and zeros simultaneously

The same contour-integral machinery can compute both the *poles* (QNMs) and *zeros* (CPA conditions) of a scalar scattering coefficient $q(\omega)$ from the same set of contour integrals [122]. The poles of q and the poles of $1/q$ are extracted simultaneously using Hankel matrices built from the contour moments of $q(\omega)$ and $q(\omega)^{-1}$. This gives a local pole-zero map of the scattering response inside the chosen contour, connecting directly to the framework of Chapter 10.

17.6 Dispersive materials and nonlinear eigenproblems

When $\varepsilon_\mathrm{r}(\omega)$ is frequency-dependent (Section 5.1), the eigenvalue problem (17.3) becomes nonlinear in ω: $\mathbf{A}(\omega)\,\mathbf{x} = 0$. Several strategies exist.

17.6.1 Auxiliary-field linearization

For Lorentz-type dispersion (5.6), the auxiliary-field technique of Section 5.5 introduces polarization currents as additional unknowns, converting the nonlinear eigenproblem into a linear (but larger) generalized eigenproblem:

$$\begin{pmatrix} \mathbf{A}_{\text{curl}} & \mathbf{C} \\ \mathbf{D} & \mathbf{A}_{\text{osc}} \end{pmatrix} \begin{pmatrix} \mathbf{x}_E \\ \mathbf{x}_P \end{pmatrix} = \omega \begin{pmatrix} \mathbf{B}_E & 0 \\ 0 & \mathbf{B}_P \end{pmatrix} \begin{pmatrix} \mathbf{x}_E \\ \mathbf{x}_P \end{pmatrix}, \tag{17.16}$$

where $\mathbf{x}_E$ represents the electric field DOFs and $\mathbf{x}_P$ the auxiliary polarization DOFs. This linearized system can be solved by standard non-Hermitian eigensolvers.

17.6.2 Polynomial eigenproblem

For more general dispersion models (Drude, multi-pole Lorentz, or frequency-dependent PML), the eigenproblem is polynomial in ω with degree $d > 2$ [124]. Linearization by companion-matrix embedding converts a degree-d polynomial eigenproblem of size N into a linear eigenproblem of size dN, which can be impractical for large FEM discretizations.

Contour-integral eigensolvers (Section 17.5) avoid explicit linearization by working directly with the nonlinear $\mathbf{A}(\omega)$, evaluating it at quadrature points on the contour. This is often more efficient than linearization when only a few eigenvalues are needed.

17.6.3 Convergence caveats

For dispersive PML, Gopalakrishnan et al. showed that computed loss values can be off by orders of magnitude if the FEM mesh and polynomial order are not in the asymptotic convergence regime: careful h- and p-refinement studies are essential before trusting the imaginary part of $\tilde{\omega}$ [124]. This is a pitfall that is often overlooked in the photonics literature, where convergence is checked against $\operatorname{Re}\tilde{\omega}$ but not $\operatorname{Im}\tilde{\omega}$.

17.7 Plane-wave expansion for complex bands

The plane-wave expansion (PWE) method of Section 3.4 extends naturally to complex $\varepsilon_{\text{r}}(\mathbf{r})$: the Fourier coefficients $\hat{\varepsilon}_{\mathbf{G}}$ become complex, and the eigenvalue matrix is no longer Hermitian. The resulting complex band structure $\omega_n(\mathbf{k})$ is computed by standard dense or iterative non-Hermitian eigensolvers (LAPACK `zgeev` for small problems, Arnoldi for large ones).

17.7.1 Computing topological invariants

From the complex Bloch eigenstates on a grid of $\mathbf{k}$-points, the Berry curvature, Chern number, and spectral winding number can be computed using the discretized formulas of Section 11.6. For the non-Bloch invariants (Section 14.5), the PWE must be performed on the appropriate complex wave-vector contour rather than on the real BZ. For tight-binding or transfer-matrix reductions this is the GBZ contour obtained from the characteristic roots. For a continuum Maxwell PWE calculation, the analogous computation is a complex-$\mathbf{k}$ band problem and must be validated against the finite-structure spectrum.

17.7.2 Non-Bloch PWE

Computing the non-Bloch band structure requires evaluating $\omega_n(\beta)$ for complex $\beta = e^{ik}$ on the GBZ contour (Section 14.4). In a lattice model this amounts to replacing e^{ik} by the complex number β in the hopping phases. In a continuum PWE calculation, it is better viewed as solving the Maxwell operator at complex Bloch wave vector $k + i\kappa$; the Fourier basis is still periodic in the unit cell, but the full field carries an exponential envelope across cells. The resulting contour is not automatically the physical GBZ and should be checked by comparing the predicted non-Bloch spectrum with open-boundary simulations.

17.8 FDTD with gain and loss

The finite-difference time-domain (FDTD) method propagates Maxwell's equations directly in time on a spatial grid. For non-Hermitian systems, the complex permittivity introduces gain or loss at each time step:

$$\frac{\partial \mathbf{D}}{\partial t} = \nabla \times \mathbf{H} - \sigma\, \mathbf{E}, \tag{17.17}$$

where $\sigma(\mathbf{r})$ is the conductivity (positive for loss, negative for gain in a linearized model).

17.8.1 Stability and gain saturation

Gain media can make the FDTD simulation unstable: the field grows exponentially and eventually overflows. Two remedies are common:

- **Gain saturation models:** replace the linear gain $\sigma < 0$ with a nonlinear model $\sigma_{\text{eff}} = \sigma_0/(1 + |\mathbf{E}|^2/E_{\text{sat}}^2)$ that clamps the amplification at high field intensities.
- **Windowed time analysis:** run the simulation for a limited time before overflow and extract QNM frequencies from the ring-down signal.

17.8.2 Extracting QNM frequencies from time series

The complex QNM frequencies can be extracted from FDTD time series using harmonic inversion: the decaying signal $f(t) = \sum_n A_n e^{-i\tilde{\omega}_n t}$ is fitted to a sum of complex exponentials using Padé approximation, the filter diagonalization method, or Prony's method. The accuracy of the extracted $\tilde{\omega}_n$ depends on the signal-to-noise ratio and the spectral separation of the modes.

17.9 Specialized algorithms

17.9.1 EP tracking

Finding exceptional points in parameter space requires tracking the coalescence of eigenvalues. Efficient algorithms include:

- **Newton iteration on the discriminant:** solve $\Delta(\mathbf{p}) = 0$ where Δ is the discriminant of the characteristic polynomial and $\mathbf{p}$ are the system parameters. For a 2×2 subsystem, $\Delta = (\tilde{\omega}_+ - \tilde{\omega}_-)^2$; the EP is located where $\Delta = 0$.

- **Encircling methods:** sweep parameters along a closed loop and check for eigenvalue exchange (Section 8.4), which signals an enclosed EP. This method is robust but requires many eigenvalue evaluations.
- **Perturbation theory:** expand the eigenvalue splitting around a known near-EP configuration to predict the EP location to first order. This is fast but may fail far from the EP.

Near higher-order EPs (EP_3, EP_4), the tracking problem is more difficult because more parameters must be tuned simultaneously and the eigenvalue sensitivity is extreme.

17.9.2 GBZ computation

Computing the generalized Brillouin zone (Section 14.4) requires solving the auxiliary polynomial (14.11) at each energy. For a two-band model, this is a quartic equation in β; for multi-band models, the degree increases accordingly. Practical approaches include:

- **Companion-matrix eigensolvers:** convert the polynomial $P(\beta, \omega) = 0$ to a companion-matrix eigenproblem and solve for all roots at each ω.
- **Root tracking:** start from a known root at one ω and track it continuously as ω varies, using Newton's method with the previous root as the initial guess.
- **Contour-integral root finding:** apply the contour-integral technique of Section 17.5 to the polynomial $P(\beta, \omega)$, treating β as the eigenvalue.

17.10 Practical guidelines

1. **PML validation is mandatory.** Always verify QNM eigenvalues against PML parameter variation (Section 17.4). Never trust a QNM frequency obtained with a single PML configuration.
2. **Convergence of $\mathrm{Im}\,\tilde{\omega}$ is harder than $\mathrm{Re}\,\tilde{\omega}$.** The imaginary part of the eigenfrequency (the decay rate) converges more slowly with mesh refinement than the real part. Always check both [124].
3. **EP-sensitive problems need special care.** Near an EP, use high-order elements, fine $\mathbf{k}$-grids, and verify with independent methods. The $\sqrt{\epsilon}$ eigenvalue splitting amplifies discretization errors.
4. **Non-Bloch invariants require accurate GBZ.** Errors in the GBZ contour propagate directly to the non-Bloch topological invariant. Use high-precision root finding and verify by comparing PBC and OBC spectra.
5. **FDTD with gain requires a stability strategy.** Always include a saturation model or truncate the simulation before numerical overflow. Harmonic inversion from a short windowed signal is often more reliable than a long, unstable run.
6. **Cross-validate between methods.** Compare FEM and PWE, or FEM and FDTD, whenever possible. Agreement between independent methods is the strongest evidence for correctness in non-Hermitian problems where physical intuition may mislead.

7. **Use biorthogonal post-processing.** For modal quantities (Purcell factors, coupling coefficients, perturbation theory), always use the biorthogonal normalization of Section 6.5, not the Hermitian norm.

Non-Hermiticity changes the numerical landscape

Standard photonic simulation tools—FEM, PWE, FDTD—remain the workhorses, but non-Hermiticity introduces new challenges: non-symmetric matrices, spurious PML modes, EP sensitivity, gain instability, dispersive nonlinear eigenproblems, and the need for biorthogonal post-processing and GBZ-based invariant computation. The validation protocol—PML variation, mode-ratio filtering, convergence checks on both $\mathrm{Re}\,\tilde{\omega}$ and $\mathrm{Im}\,\tilde{\omega}$, cross-method comparison—is not optional: it is the price of doing reliable numerical non-Hermitian photonics.

Chapter 18

Devices and Physical Limits

Non-Hermitian photonics is not only a theoretical framework—it is a design language for optical devices. Exceptional-point sensors, $\mathcal{PT}$-symmetric lasers, topological amplifiers, CPA switches, and BIC-based cavities all exploit the physics developed in the preceding chapters. The principal device concepts are surveyed, with operating conditions derived from the Maxwell-level theory of Parts I–III. The chapter analyzes the physical limits imposed by noise, nonlinearity, and fabrication disorder, and asks a question that the literature sometimes avoids: which non-Hermitian advantages survive the constraints of the real world? The organizing principle is that every device claim must be tested against its noise floor, and every enhanced response must be compared with its enhanced noise.

18.1 $\mathcal{PT}$-symmetric lasers

18.1.1 Single-mode selection from mode non-orthogonality

A conventional multimode laser oscillates in many longitudinal modes simultaneously because the gain medium provides broadband amplification without spectral discrimination. A $\mathcal{PT}$-symmetric laser, consisting of a gain cavity coupled to a loss cavity of equal and opposite imaginary refractive index, can enforce single-mode operation through a mechanism rooted in the non-Hermitian spectral structure developed in Chapter 8.

Consider two coupled single-mode cavities with resonance frequencies $\omega_0 \pm \delta/2$, coupling rate κ, and balanced gain/loss $\pm g$. The effective Hamiltonian from Section 8.2 is

$$H_{\text{eff}} = \begin{pmatrix} \omega_0 + \delta/2 + ig & \kappa \\ \kappa & \omega_0 - \delta/2 - ig \end{pmatrix}. \tag{18.1}$$

Setting $\delta = 0$ for simplicity, the eigenfrequencies are

$$\omega_\pm = \omega_0 \pm \sqrt{\kappa^2 - g^2}. \tag{18.2}$$

Below the EP ($g < \kappa$), both supermodes share the same imaginary part—neither is preferentially amplified. Above the EP ($g > \kappa$), one supermode acquires a net positive imaginary part while the other acquires a net negative imaginary part; only the amplified mode reaches the lasing threshold. This EP-mediated mode selection is intrinsically non-Hermitian: it has no analogue in a Hermitian coupled-cavity system where mode splitting is always real.

The physical mechanism extends to multimode cavities. If the system supports N longitudinal modes, each with its own detuning δ_m, the $\mathcal{PT}$ symmetry condition selects the mode whose gain–loss contrast most efficiently exceeds the coupling threshold. Typically, this is the mode closest to line center of the gain medium, where $g(\omega)$ is largest. The remaining $N-1$ modes experience net loss and are suppressed. This provides a mode-selection strategy based on non-Hermitian spectral topology rather than on the conventional mechanisms of gain competition, spatial hole burning, or intracavity etalons.

18.1.2 Experimental demonstrations

The first experimental demonstrations of $\mathcal{PT}$-symmetric single-mode lasers were reported independently by two groups in 2014. Feng et al. fabricated coupled InGaAsP microring resonators with alternating gain and loss regions, achieving single-mode lasing with a side-mode suppression ratio (SMSR) exceeding 20 dB; the conventional counterpart (both rings pumped uniformly) exhibited multimode lasing. Hodaei et al. demonstrated a similar concept using coupled microring resonators with selective pumping and confirmed that the single-mode selection persisted across a range of pump powers.

The key experimental signature is pump-dependent mode selection: at low pump power the emission can be broad or multimode, while above the non-Hermitian threshold one supermode acquires a clear gain advantage. In an ideal two-mode model this transition is tied to the square-root bifurcation near an EP (Section 8.9); in real lasers gain saturation, spatial hole burning, and detuning smear the transition and must be included before assigning an observed narrowing uniquely to EP physics.

18.1.3 Power, coherence, and the Petermann penalty

The output power and coherence of a $\mathcal{PT}$-symmetric laser are subject to the same fundamental limits as any laser: the Schawlow–Townes linewidth $\Delta\nu_{\mathrm{ST}}$, gain saturation, and spatial hole burning. Near the EP, the Petermann excess noise factor (Section 6.8) enhances the linewidth:

$$\Delta\nu = K_{\mathrm{P}}\,\Delta\nu_{\mathrm{ST}}, \qquad K_{\mathrm{P}} = \frac{|\langle\tilde{\mathbf{E}}^{\mathrm{L}}|\tilde{\mathbf{E}}^{\mathrm{L}}\rangle|\;|\langle\tilde{\mathbf{E}}^{\mathrm{R}}|\tilde{\mathbf{E}}^{\mathrm{R}}\rangle|}{|\langle\tilde{\mathbf{E}}^{\mathrm{L}}|\tilde{\mathbf{E}}^{\mathrm{R}}\rangle|^2}, \tag{18.3}$$

where $K_{\mathrm{P}} \geq 1$ diverges as the EP is approached [50]. Physically, the divergence arises because the left and right eigenvectors become parallel at the EP, so the biorthogonal projection $\langle\tilde{\mathbf{E}}^{\mathrm{L}}|\tilde{\mathbf{E}}^{\mathrm{R}}\rangle \to 0$, and noise from the gain reservoir is projected with increasing efficiency into the lasing mode.

For the two-mode dimer (18.1) with $\delta = 0$, the Petermann factor can be computed explicitly. Defining the non-orthogonality parameter θ through $\cos\theta = g/\kappa$ in the $\mathcal{PT}$-symmetric phase,

$$K_{\mathrm{P}} = \frac{1}{\sin^2\theta} = \frac{1}{1-(g/\kappa)^2}. \tag{18.4}$$

At the EP ($g = \kappa$), $K_{\mathrm{P}} \to \infty$. This expression is written for the ideal unbroken-phase eigenvectors; in the broken phase the Petermann factor must be evaluated from the actual above-threshold linearized laser modes. The practical design question for a $\mathcal{PT}$-symmetric laser is therefore whether the single-mode purity gained by operating near the EP compensates for the Petermann-enhanced linewidth. Typically, the optimal operating point is somewhat above the EP, where the mode selectivity is strong but K_{P} has not yet diverged catastrophically.

Example: microring $\mathcal{PT}$ laser dimer. Consider two coupled microring resonators at $\lambda = 1550$ nm with $Q_0 = 5 \times 10^4$, coupling rate $\kappa = 2\pi \times 50$ GHz, and balanced gain/loss $g = 1.2\kappa$ (broken $\mathcal{PT}$ phase). The Schawlow–Townes linewidth of a single ring is $\Delta\nu_{\mathrm{ST}} \approx 30$ kHz. In the broken phase, both eigenfrequencies share the real part ω_0, but their imaginary parts split: $\mathrm{Im}\,\omega_\pm = \pm\sqrt{g^2 - \kappa^2} = \pm 0.663\kappa \approx \pm 2\pi \times 33$ GHz. The amplified mode therefore has a net gain margin of ~ 33 GHz over the suppressed mode, providing robust single-mode selection. A representative Petermann factor of order $K_{\mathrm{P}} \sim 3$ would broaden the linewidth to $\Delta\nu \sim 90$–100 kHz, illustrating that useful mode-selection can coexist with a moderate noise penalty when the device is not operated too close to the EP. The exact value requires the above-threshold saturated modes, not only the passive two-mode dimer.

18.1.4 Beyond single-mode: multi-element $\mathcal{PT}$ lasers

The two-element dimer is the simplest $\mathcal{PT}$-symmetric laser, but the concept extends to multi-element arrays. In a ring of N coupled cavities with alternating gain and loss, the non-Hermitian band structure (Chapter 12) determines which modes lase. The interplay between the $\mathcal{PT}$ phase transition and the band topology can produce single-mode selection even in large arrays, provided that the gain–loss contrast is tuned to the appropriate EP in the band structure.

A particularly interesting configuration is the $\mathcal{PT}$-symmetric laser array where the gain is concentrated at a topological edge state (Section 18.5.2). This combines the $\mathcal{PT}$-mediated mode selection with topological protection against backscattering, potentially improving single-mode lasing robustness without relying only on fine-tuned inter-cavity coupling.

18.2 Exceptional-point sensors

18.2.1 Enhanced frequency splitting

The square-root response at a second-order EP (Section 8.7) produces a frequency splitting $\Delta\omega \propto \epsilon^{1/2}$ for a perturbation ϵ, which exceeds the linear splitting $\Delta\omega \propto \epsilon$ of a conventional diabolic-point (DP) sensor for small ϵ.

To make this concrete, consider a two-mode sensor with coupling κ tuned to the EP ($g = \kappa$). A small perturbation ϵ (for example, the dielectric shift from a nanoparticle binding to one of the cavities) splits the eigenfrequencies by

$$\Delta\omega_{\mathrm{EP}} = 2\sqrt{\kappa\epsilon}, \tag{18.5}$$

while the same perturbation applied to a DP sensor ($g = 0$, i.e. a true degeneracy with orthogonal eigenvectors) lifts the degeneracy linearly:

$$\Delta\omega_{\mathrm{DP}} = 2\epsilon \tag{18.6}$$

when $\epsilon \ll \kappa$. The responsivity enhancement is therefore

$$\frac{\Delta\omega_{\mathrm{EP}}}{\Delta\omega_{\mathrm{DP}}} = \frac{\sqrt{\kappa\epsilon}}{\epsilon} = \sqrt{\frac{\kappa}{\epsilon}}, \tag{18.7}$$

which diverges as $\epsilon \to 0$.

At an nth-order EP, the scaling is $\Delta\omega \propto \epsilon^{1/n}$, promising even larger enhancements for higher-order EPs. However, higher-order EPs require the simultaneous tuning of $2(n-1)$ real parameters (Section 8.1.1) and are correspondingly more fragile.

Example: nanoparticle detection. A whispering-gallery-mode microsphere sensor operates at $\lambda = 1550$ nm with $Q = 10^7$ and mode volume $V_{\mathrm{eff}} \approx 300\ \mu\mathrm{m}^3$. A polystyrene nanoparticle with radius $a = 8.5$ nm and refractive index $n_p = 1.59$ produces a perturbation $\epsilon \approx (\omega_0/2)(n_p^2 - 1)\, a^3/V_{\mathrm{eff}} \approx 2\pi \times 0.3$ MHz. With coupling $\kappa = 2\pi \times 100$ MHz, the DP splitting is $\Delta\omega_{\mathrm{DP}} \approx 2\pi \times 0.6$ MHz—well below the linewidth $\gamma = \omega_0/(2Q) \approx 2\pi \times 10$ MHz—while the EP splitting is $\Delta\omega_{\mathrm{EP}} \approx 2\pi \times 11$ MHz, comparable to the linewidth and therefore resolvable. The responsivity enhancement is $\sqrt{\kappa/\epsilon} \approx 18$. As discussed next, this enhanced splitting does not automatically translate into enhanced detection precision.

18.2.2 The noise problem: quantum Fisher information

The fundamental question is whether the enhanced splitting translates into enhanced signal-to-noise ratio (SNR). Near the EP, the Petermann factor diverges, broadening the resonance linewidth by the same factor that enhances the splitting.

Zhang et al. developed a quantum noise theory for EP sensors using the quantum Langevin equations and input–output formalism [49]. Working in the two-mode bosonic framework, they computed the covariance matrix of the output field and showed that the quantum Fisher information (QFI) for estimating the perturbation ϵ satisfies

$$\mathcal{F}_{\mathrm{EP}}(\epsilon) \leq \mathcal{F}_{\mathrm{DP}}(\epsilon) \cdot f(\theta), \tag{18.8}$$

where $f(\theta) \to 1$ as $\theta \to 0$ (approaching the EP). The Cramér–Rao bound then gives the minimum variance of any unbiased estimator:

$$\mathrm{Var}(\hat{\epsilon}) \geq \frac{1}{M\,\mathcal{F}(\epsilon)}, \tag{18.9}$$

where M is the number of measurements. In this steady-state linear measurement model, with equal input power and integration time, $\mathcal{F}_{\mathrm{EP}} \leq \mathcal{F}_{\mathrm{DP}}$: the EP sensor does not achieve a fundamental precision advantage over the corresponding DP sensor. The enhanced splitting is accompanied by enhanced noise and linewidth broadening.

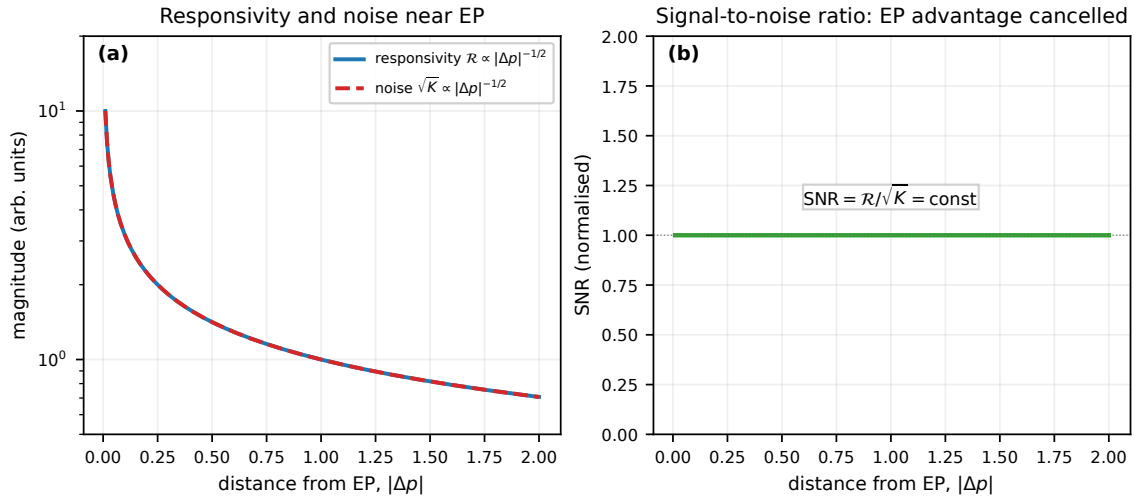


Figure 18.1 Exceptional-point sensing tradeoff in an idealized steady-state, equal-resource model. (a) Near an EP, the responsivity scaling $\mathcal{R} \propto |\Delta p|^{-1/2}$ is accompanied by a Petermann-related noise scaling of the same order, $\sqrt{K} \propto |\Delta p|^{-1/2}$. The two curves are plotted with equal prefactors only to emphasize the shared asymptotic scaling. (b) The corresponding SNR proxy $\mathcal{R}/\sqrt{K}$ is therefore approximately constant in this model. This illustrates why enhanced frequency splitting alone does not guarantee improved measurement precision; a real sensor must be judged by its complete noise budget and readout protocol [50, 49].

The physical mechanism is transparent: the gain reservoir that creates the EP also injects vacuum noise into the system. This noise is amplified by the same non-orthogonal mode structure that amplifies the signal, producing a noise floor that rises in lockstep with the responsivity.

Wang et al. tested this tradeoff experimentally in a Brillouin ring laser gyroscope operating near an EP. They measured the Petermann-factor linewidth broadening and showed that the enhanced Sagnac splitting was offset by enhanced noise, so that the measured rotation sensitivity did not improve simply by operating at the EP [50].

18.2.3 The noise problem: nonlinear and active systems

Zheng and Chong extended the analysis to nonlinear EP sensors—coupled cavities operating above the lasing threshold—and showed that nonlinearity does not rescue the SNR [51]. Their analysis revealed three mechanisms by which noise degrades nonlinear EP sensing:

1. **Noise-induced EP shifting:** fluctuations in the intracavity intensity shift the saturated gain, moving the effective operating point away from the EP. The system never actually reaches the singular $\epsilon^{1/n}$ responsivity.
2. **Order reduction:** the nonlinear steady state reduces the effective EP order, weakening the fractional-power enhancement. An EP_2 in the linear system may behave as an effective EP_1 (no enhancement) in the saturated regime.
3. **BdG fluctuation EPs:** the Bogoliubov–de Gennes Hamiltonian governing small fluctuations around the nonlinear steady state has its own EP structure, which can produce noise divergences stronger than the Petermann estimate from the linear system.

The conclusion is sobering: in the analyzed steady-state protocols, neither linear nor nonlinear EP sensors achieve a fundamental SNR advantage over DP sensors under the same resource constraints (same cavity Q, same input power, same integration time).

18.2.4 When can EP sensing help?

Despite the noise-cancellation results, EP sensing may still offer practical advantages in specific protocols:

1. **Transient measurements:** reading a short-time ring-down can make the enhanced dynamical response easier to observe in a fixed bandwidth. The relevant figure of merit is the full transient estimator variance, not only the steady-state linewidth.
2. **Interferometric readout:** heterodyne or self-heterodyne schemes that measure the beat frequency between the two split modes may reject technical common-mode noise, even though they do not remove the fundamental quantum-noise accounting.
3. **Active feedback:** real-time feedback can stabilize technical drift of the operating point near the EP. It improves practical stability but adds its own measurement noise and control bandwidth limits.
4. **Exceptional surfaces:** in parameter spaces where EPs form codimension-one surfaces rather than isolated points, the system can be operated on the exceptional surface without fine-tuning, potentially reducing the sensitivity to drift.

The general lesson is that the *responsivity* of an EP sensor is genuinely enhanced, but converting that responsivity into enhanced *precision* requires a measurement protocol whose complete noise budget is better than the reference sensor. The distinction between responsivity and precision is a recurrent theme in non-Hermitian device physics and should be kept in mind whenever an "enhanced sensitivity" claim is evaluated.

18.3 BIC-based lasers and cavities

18.3.1 From radiation suppression to lasing threshold

Bound states in the continuum (Chapter 16) provide a fundamentally different route to high-Q lasing. A photonic crystal slab laser operating near a BIC exploits the topological suppression of radiation loss to achieve single-mode lasing with high power and narrow linewidth, without requiring a physical mirror or cavity boundary.

The lasing condition connects the BIC physics to the Maxwell-level framework of this book. The radiation quality factor near a BIC scales as (Section 16.3.1)

$$Q_{\mathrm{rad}}(k) \propto \frac{1}{|c(k)|^2}, \tag{18.10}$$

where $c(k)$ is the radiation amplitude introduced in Section 16.1.1. At a BIC, $c(k_{\mathrm{BIC}}) = 0$ and $Q_{\mathrm{rad}} \to \infty$. Adding gain to the slab reduces the total quality factor to

$$\frac{1}{Q_{\mathrm{tot}}} = \frac{1}{Q_{\mathrm{rad}}} + \frac{1}{Q_{\mathrm{abs}}} - \frac{1}{Q_{\mathrm{gain}}}, \tag{18.11}$$

where Q_{abs} accounts for material absorption and Q_{gain} accounts for optical gain. The lasing threshold is reached when $Q_{\text{tot}} \to \infty$, i.e.,

$$\frac{1}{Q_{\text{gain}}} = \frac{1}{Q_{\text{rad}}} + \frac{1}{Q_{\text{abs}}}. \tag{18.12}$$

Near the BIC, Q_{rad} is very large, so the threshold is determined primarily by material absorption: the BIC suppresses the radiation channel, leaving only the material loss as the dominant threshold-setting mechanism. In finite active devices the threshold is also limited by finite aperture, pump overlap, nonradiative recombination, and thermal rollover, but radiation suppression is the central reason for the low thresholds observed in BIC lasers.

The finite-size limitation is set by the momentum spread $\Delta k \sim 1/N$ of a cavity with N periods. For an isolated charge-one BIC, $Q_{\text{finite}} \propto N^2$, which can be modest for small cavities.

18.3.2 Super-BIC lasers

The topological charge-merging strategy of Section 16.3 can improve the finite-size scaling. When symmetry and charge conservation force the lower-order radiation amplitudes to cancel at the merged point, the radiation amplitude vanishes to higher order:

$$c(k) \propto k^{|q|} \quad \Longrightarrow \quad Q_{\text{finite}} \propto N^{2|q|}. \tag{18.13}$$

For the super-BIC design discussed in Chapter 16, the leading radiation rate scales as k^6, so the finite-size Q can scale as N^6, compared with the N^2 scaling of an isolated BIC. This scaling is not guaranteed by charge addition alone; it also relies on the symmetry of the merged configuration.

Hwang et al. demonstrated a super-BIC laser in an InGaAsP photonic crystal slab with a square lattice of air holes [121]. By tuning the lattice constant to the merging point ($a \approx 576.3\,\text{nm}$), they achieved:

- Radiation loss scaling of k^6 instead of k^2.
- Measured threshold of $\sim 1.47\,\text{kW/cm}^2$—roughly 50 to 10^7 times lower than previous BIC nanolasers.
- Quality factor up to ~ 7300 in a cavity of only 15×15 periods.
- A strongly confined far-field image consistent with the k^6 angular suppression.

The super-BIC concept demonstrates that topological charge engineering can convert a fundamental limitation (the finite-size Q penalty) into a practical design advantage.

18.3.3 BZF-BIC cavities

Brillouin-zone-folding BICs (Section 16.4) offer a complementary approach: instead of steepening the Q scaling at a single k-point, they sustain high Q over a broad momentum range. Wang et al. showed that BZF-BICs retain Q at least ten times higher than conventional BICs under the same structural disorder [111], making them attractive for large-area lasing and sensing applications where momentum-space averaging is important.

18.3.4 Quasi-BIC applications beyond lasing

Away from the exact BIC condition, the system supports a quasi-BIC resonance with finite but very large Q. The quasi-BIC produces a sharp Fano line shape in the scattering spectrum (Section 16.6.1), which is useful for applications beyond lasing:

- **Refractive-index sensing:** the figure of merit FOM $= S/\delta\lambda = \Delta\lambda/(\Delta n\,\delta\lambda)$, where $S = \Delta\lambda/\Delta n$ is the sensitivity, $\delta\lambda$ is the linewidth, and Δn is the index change, benefits directly from the high Q of the quasi-BIC. Reported FOMs for quasi-BIC metasurfaces exceed 10^3 in the near-infrared.
- **Nonlinear enhancement:** resonant field buildup can strongly enhance nonlinear conversion, with the precise power of Q depending on whether the pump, signal, and generated frequency are all resonant and critically coupled. This makes quasi-BIC resonances attractive platforms for nonlinear nanophotonics, but the scaling is device- and process-dependent.
- **Spectral filtering:** the asymmetric Fano profile of a quasi-BIC can be engineered to produce narrow-band filters with tunable center wavelength and bandwidth by adjusting the structural perturbation that breaks the BIC condition.

18.4 Coherent perfect absorbers and switches

18.4.1 CPA as a scattering zero

A coherent perfect absorber (Section 4.4.2) operates at a zero of the scattering matrix: with the correct phase and amplitude relationship between two input beams, all light is absorbed. From the pole-zero framework of Chapter 10, the CPA condition corresponds to a real-frequency zero of $\det S(\omega)$.

The CPA device functions as an all-optical switch: modulating the relative phase ϕ between the two input beams sweeps the system between perfect absorption ($\phi = \phi_0$) and near-total transmission ($\phi = \phi_0 + \pi$). The switching contrast is

$$\mathcal{C} = \frac{T_{\max}}{T_{\min}} = \frac{(1+|r|)^2}{(1-|r|)^2}, \tag{18.14}$$

where $|r|$ is the single-channel reflection coefficient at the CPA frequency. For $|r| \to 1$ (high-finesse cavity), the contrast diverges, but in practice it is limited by absorption losses and fabrication imperfections.

The switching speed is set by the cavity lifetime $\tau = Q/\omega_0$; for a $Q \sim 10^3$ resonator at $\lambda = 1550$ nm, $\tau \approx 0.8$ ps—faster than electronic switches by orders of magnitude.

18.4.2 CPA-laser devices

At the CPA-laser point (Chapter 10), the same device simultaneously acts as a laser (for one input configuration) and a perfect absorber (for the conjugate configuration). From the scattering matrix formalism, this corresponds to

$$\det S(\omega_{\text{CPA-laser}}) = 0 \quad \text{and} \quad \det S(\omega_{\text{CPA-laser}}) \text{ has a pole}, \tag{18.15}$$

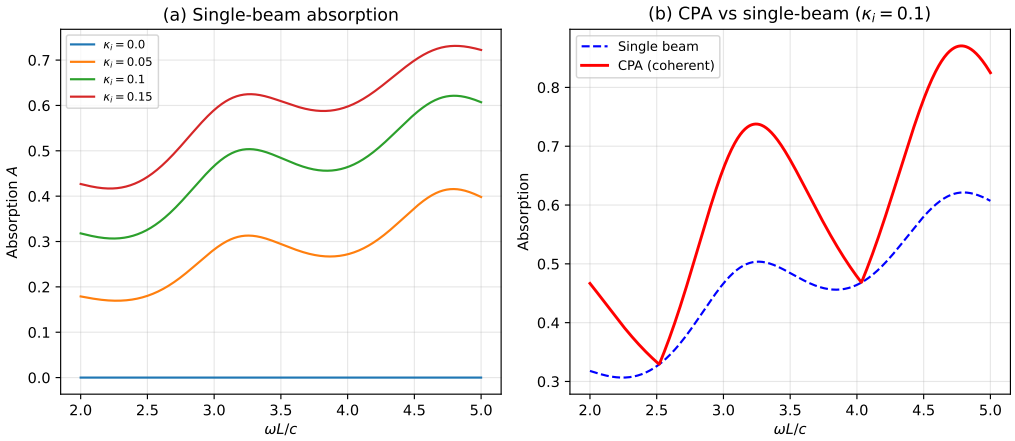


Figure 18.2 Coherent perfect absorption in a lossy Fabry–Pérot cavity ($n = 2$). (a) Single-beam absorption for different extinction coefficients κ_i: increasing material loss raises the peak absorption but broadens the resonance. (b) Comparison of single-beam (dashed) and coherent two-beam (solid) absorption for $\kappa_i = 0.1$. Under coherent illumination the absorption reaches near-unity at the CPA frequency, demonstrating the interferometric enhancement that is the hallmark of CPA—the time-reversed counterpart of lasing (Section 10.3).

i.e., a simultaneous pole and zero at the same real frequency. In a $\mathcal{PT}$-symmetric system, the symmetry constrains poles and zeros to occur in conjugate pairs (Section 10.4), making the CPA-laser condition possible when a pole and its time-reversed zero meet on the real axis. The symmetry is a constraint, not by itself a guarantee that a real-frequency pole-zero coincidence occurs without tuning.

The dual CPA-laser functionality enables phase-controlled routing: by tuning the input phase, the device can be switched between amplification and absorption, with the transition controlled by the pole-zero structure of the scattering matrix.

The practical challenge is maintaining the precise gain–loss balance required by the CPA-laser condition. Any imbalance shifts the pole and zero apart, reducing the contrast. The critical distinction, emphasized in Section 10.5.3, is that the CPA-laser condition depends on fine-tuned scattering singularities, not on topological invariants. Topological protection (from BIC charges or Chern edge states) does *not* protect the CPA-laser point, because a topological invariant constrains the *number* of singularities in a region of parameter space, not their *positions*.

18.4.3 Practical CPA limitations

Several practical constraints limit CPA device performance:

- **Beam alignment:** perfect absorption requires precise control of the phase, amplitude, and spatial mode of two input beams. Any mismatch reduces the absorption efficiency.
- **Bandwidth:** the CPA condition is satisfied at a single frequency; broadband operation requires a resonator design with overlapping or broadened zeros.
- **Nonlinear effects:** at high input powers, the CPA condition shifts due to Kerr and thermal nonlinearities, limiting the useful power range.

18.5 Topological lasers

18.5.1 Topological insulator lasers

A topological insulator laser [125, 126] combines a photonic topological insulator (Chapter 11) with optical gain. The lasing mode propagates along the topological edge channel, where elastic backscattering from fabrication defects is strongly suppressed as long as the topological gap and the protecting symmetries are maintained. The result can be a laser with robust single-mode operation and high slope efficiency, but gain competition, disorder-induced loss, and finite-size coupling still matter.

Several platforms have demonstrated topological lasing:

- **Microring arrays:** coupled microring resonators with alternating link-ring couplers that implement a photonic anomalous quantum Hall lattice. Bandres et al. (2018) demonstrated edge-mode lasing with a slope efficiency comparable to conventional ring lasers but with improved robustness against intentional defects—removing a ring from the edge path did not suppress lasing, whereas removing a bulk-mode ring had no effect.
- **Valley-Hall photonic crystals:** triangular-hole photonic crystal slabs with broken inversion symmetry, supporting valley-polarized edge states. Lasing along the domain wall was shown to be robust against sharp 120° bends, with threshold comparable to the straight-edge case.
- **SSH micropillar chains:** one-dimensional arrays of semiconductor micropillars with alternating coupling, where topological edge modes lase selectively at the chain boundary. The topological gap protects the edge state from hybridizing with bulk modes, providing a natural single-mode filter.

The key metric for a topological laser is not the absolute threshold or slope efficiency—these are comparable to conventional lasers—but rather the *robustness* of the single-mode operation and the *insensitivity* to disorder.

18.5.2 Non-Hermitian topological lasers

Adding non-Hermitian ingredients (gain/loss patterning, $\mathcal{PT}$ symmetry) to topological structures creates a richer design space. For example, selectively pumping only the edge region of a topological lattice can enforce single-mode topological lasing without requiring uniform gain across the entire structure—reducing the power budget and avoiding multimode instabilities from bulk gain.

This strategy exploits non-Hermitian gain selectivity rather than the skin effect by itself: patterning the imaginary part of the on-site potential to provide gain only near the boundary preferentially amplifies edge modes and suppresses bulk modes. If the underlying lattice also has point-gap winding, true skin accumulation can further reshape the modal profiles, but boundary pumping alone should not be identified with the NHSE.

An important distinction is between *topological protection of the lasing mode* and *topological origin of the lasing threshold*. In most topological lasers [127, 128, 29], the topology protects the spatial profile of the lasing mode (preventing backscattering and maintaining single-mode

character), but the threshold itself is set by the gain–loss balance, which is not topologically protected.

18.5.3 Non-Hermitian skin amplifiers

The standard non-Hermitian skin effect localizes an extensive set of right eigenstates at one boundary (Chapter 14). In a system with gain, this localization produces exponential amplification of signals injected at the opposite boundary, as they propagate toward the skin-effect edge. Using the Hatano–Nelson model of Section 14.3, the ideal amplitude factor grows exponentially with system size:

$$G = \left(\frac{t_R}{t_L}\right)^L, \tag{18.16}$$

where t_R/t_L is the hopping asymmetry ratio and L is the number of unit cells. If G is interpreted as a power gain, the corresponding dB value is $10\log_{10} G$; if it is a field-amplitude factor, the power gain scales as $|G|^2$. For $t_R/t_L = 2$ and $L = 20$, the ideal amplitude factor is $2^{20} \approx 10^6$, corresponding to an enormous small-signal power gain before saturation and noise are included.

The direction of accumulation is tied to the sign of the point-gap winding of the PBC spectrum around the reference energy (Section 12.3). This provides a route to directional amplification without necessarily using magneto-optical materials, although a physical photonic implementation must still realize effective asymmetric hopping and control added noise.

The practical challenge is implementing the required asymmetric hopping in a photonic platform. Photonic mesh lattices with fiber loops of different lengths, time-modulated coupled resonators, and magnetically biased photonic crystals have been proposed as physical realizations.

18.6 Active non-Hermitian metasurfaces

18.6.1 Wavefront control with gain and loss

Non-Hermitian metasurfaces [123, 129] extend the wavefront-control capabilities of passive metasurfaces by incorporating gain and loss as additional design degrees of freedom. In a passive metasurface, each meta-atom controls the phase and amplitude of the transmitted or reflected field through its geometric and material parameters. In a non-Hermitian metasurface, each meta-atom additionally carries a complex refractive index with a controllable imaginary part, enabling independent control of the phase, amplitude, and gain/loss profile across the surface.

From the scattering-matrix perspective (Chapter 10), each meta-atom is characterized by its complex transmission and reflection coefficients. By engineering the pole-zero structure of each meta-atom, the metasurface can implement arbitrary complex transfer functions, achieving functionalities impossible with purely passive elements—for example, asymmetric reflection profiles where the reflected wavefront depends on the illumination direction.

18.6.2 $\mathcal{PT}$-symmetric metasurfaces

A metasurface with a $\mathcal{PT}$-symmetric gain/loss distribution can exhibit unidirectional properties analogous to those of $\mathcal{PT}$-symmetric Bragg gratings (Section 9.3.4). At the $\mathcal{PT}$-breaking point,

the metasurface can be invisible from one side while reflecting from the other, creating a flat-optics analogue of unidirectional invisibility.

18.6.3 CPA metasurfaces

A metasurface designed to operate at the CPA condition provides a thin-film all-optical modulator. By controlling the relative phase of two counter-propagating beams incident on the metasurface, the absorption can be switched from near-zero to near-unity, enabling modulation at sub-picosecond timescales with a device footprint of a single wavelength.

18.7 Physical limits

Every non-Hermitian device advantage comes with a physical cost. This section catalogs the principal limits in a unified framework and provides quantitative estimates that allow device designs to be compared on an equal footing.

18.7.1 Quantum noise and the Petermann factor

All non-Hermitian photonic devices are ultimately limited by quantum noise. The fluctuation-dissipation theorem requires that any system with loss also has noise; gain media amplify both signal and spontaneous emission. Near an EP, the enhanced Petermann factor (18.3) amplifies the noise, setting a fundamental floor on the sensitivity of EP-based devices.

The Petermann factor reflects the non-orthogonality of the modes in the biorthogonal expansion of Chapter 6. In a system with $K_{\mathrm{P}} = 100$ (easily achieved within $|\Delta p|/\kappa \sim 0.1$ of a second-order EP), the spontaneous emission rate into the lasing mode is enhanced by a factor of 100 relative to the Hermitian limit, degrading the laser coherence and the sensor precision by the same factor.

For an nth-order EP, the Petermann factor diverges as (Section B.4.2)

$$K_{\mathrm{P}} \propto \frac{1}{|\Delta p|^{2(n-1)}}, \tag{18.17}$$

where Δp is the distance from the EP in parameter space. The noise penalty therefore grows faster with EP order than the responsivity enhancement ($\epsilon^{1/n}$), and the net SNR actually *decreases* with increasing EP order for fixed resource constraints.

18.7.2 Gain saturation and nonlinearity

The linear non-Hermitian framework of this book is valid below or near the lasing threshold. Above threshold, gain saturation introduces nonlinear corrections that stabilize the system (preventing runaway amplification) but also modify the EP structure and the topological invariants.

The gain saturation model of Section 5.6 gives

$$g(I) = \frac{g_0}{1 + I/I_{\mathrm{sat}}}, \tag{18.18}$$

where g_0 is the small-signal gain, I is the intracavity intensity, and I_{sat} is the saturation intensity. For EP sensors operating above threshold, the nonlinear steady state shifts the operating point

away from the EP by an amount

$$\delta g \sim g_0 \frac{I_{\mathrm{ss}}}{I_{\mathrm{sat}}}, \tag{18.19}$$

where I_{ss} is the steady-state intracavity intensity. This shift effectively reduces the EP order and weakens the $\epsilon^{1/n}$ responsivity [51].

For topological invariants, gain saturation can change the sign of the imaginary part of the permittivity, potentially opening or closing point gaps and modifying the winding numbers. The interplay between nonlinearity and non-Hermitian topology is an active research frontier (Chapter 19).

18.7.3 Fabrication disorder

Real photonic devices have fabrication imperfections that perturb the $\mathcal{PT}$ symmetry condition, shift EPs, and break the gain–loss balance. Different non-Hermitian features have different robustness:

- **Topologically constrained:** BIC far-field charges (Section 16.2), Chern edge states, and point-gap/skin-effect orientation are robust to perturbations that preserve the relevant gap, symmetry, and radiation-channel structure. They change only when the corresponding topological description breaks down or a topological transition occurs.
- **Fine-tuned:** the precise EP location, the CPA-laser condition, the $\mathcal{PT}$ symmetry balance, and the gain–loss threshold all depend on continuous parameters and are shifted by any fabrication error.

The practical design challenge is to maximize the fraction of the device functionality that relies on topological protection, while accepting that fine-tuned parameters will need active stabilization or tolerance engineering.

A useful figure of merit for disorder robustness is the *topological gap ratio*—the ratio of the topological gap Δ_{top} to the typical disorder-induced level broadening Γ_{dis}. When $\Delta_{\mathrm{top}}/\Gamma_{\mathrm{dis}} \gg 1$, topologically protected features survive; when $\Delta_{\mathrm{top}}/\Gamma_{\mathrm{dis}} \lesssim 1$, the topological protection breaks down and device performance reverts to the non-topological baseline.

18.7.4 Thermal and power limits

Gain media dissipate heat. In semiconductor lasers, the thermal load from the pump current limits the maximum gain that can be applied before thermal rollover degrades the performance. $\mathcal{PT}$-symmetric devices, which require a precise balance between gain and loss, are particularly sensitive to thermal drifts because the gain medium and the loss medium have different thermal coefficients.

For BIC-based devices, the ultra-high Q near the BIC reduces the threshold power, mitigating the thermal problem. The super-BIC laser of Hwang et al. achieved a threshold of $\sim 1.47\,\mathrm{kW/cm^2}$—orders of magnitude below conventional nanolasers—precisely because the topological charge merging minimized the radiation loss [121].

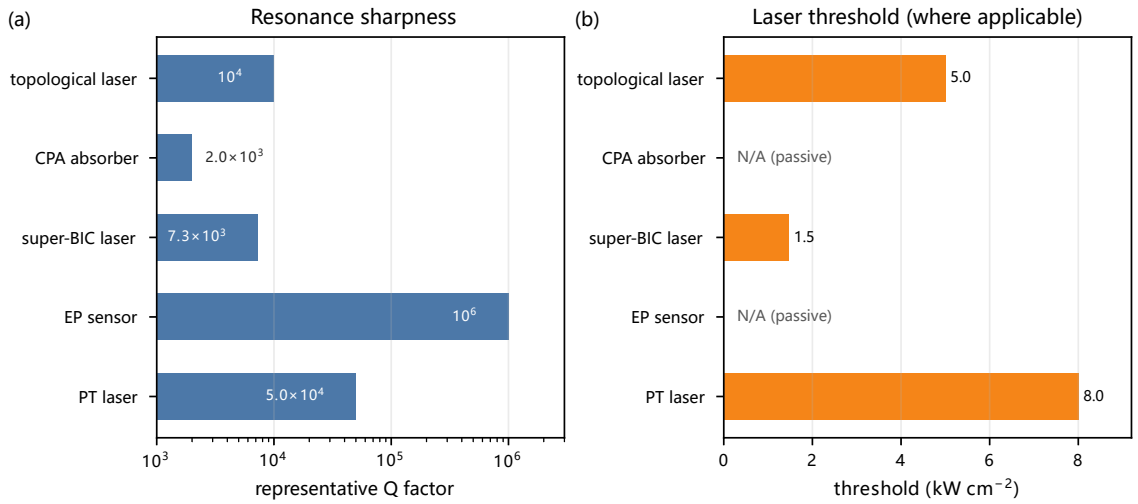


Figure 18.3 Quantitative comparison of five non-Hermitian photonic device classes discussed in this chapter. (a) Representative quality factor Q on a logarithmic scale. The EP sensor achieves the highest $Q \sim 10^6$ through whispering-gallery modes, while the CPA absorber operates at a modest $Q \sim 2 \times 10^3$ because critical coupling demands matched radiative and absorptive rates. (b) Lasing threshold in $\mathrm{kW\,cm^{-2}}$ for laser-type devices; the EP sensor and CPA absorber are passive and marked N/A. The super-BIC laser has the lowest threshold ($1.47\ \mathrm{kW\,cm^{-2}}$) thanks to the k^6 radiation-loss suppression of merged BICs. Data are representative values from the literature discussed in the text.

18.7.5 Gain–bandwidth constraints

A fundamental constraint on non-Hermitian devices comes from the gain–bandwidth product. The Kramers–Kronig relations (Section 5.3) link the real and imaginary parts of the permittivity, so gain over a finite bandwidth implies a corresponding frequency-dependent phase shift. For a Lorentzian gain medium with peak gain g_0 and gain bandwidth $\Delta\omega_g$, the gain–bandwidth product is

$$g_0 \cdot \Delta\omega_g = \frac{N_{\mathrm{inv}}\,\sigma_0\,c}{n_g}, \tag{18.20}$$

where N_{inv} is the population inversion density, σ_0 is the stimulated emission cross-section, and n_g is the group index. This places an upper bound on the gain that can be achieved over a given bandwidth, limiting the number of modes that can be simultaneously pushed above threshold in a $\mathcal{PT}$-symmetric laser and constraining the operational bandwidth of EP sensors.

18.7.6 Summary: device concepts and their physical limits

Table 18.1 summarizes the principal non-Hermitian device concepts, their key operating mechanisms, the source of their non-Hermitian advantage, and the physical limit that constrains their performance.

Promise and limits of non-Hermitian devices

Non-Hermitian photonic devices exploit exceptional points, $\mathcal{PT}$ symmetry, CPA-laser duality, BIC topology, and the skin effect to achieve functionalities impossible in Hermitian systems: single-mode lasing, enhanced sensing responsivity, all-optical switching, and directional amplification. But every advantage comes with a physical cost—the Petermann noise

Table 18.1 Non-Hermitian photonic device concepts and their physical limits.

Device	Mechanism	NH advantage	Primary limit
$\mathcal{PT}$ laser	EP mode selection	Single-mode purity	Petermann noise, thermal drift
EP sensor	$\epsilon^{1/n}$ splitting	Enhanced responsivity	Petermann noise cancels SNR gain
BIC laser	Topological Q	Ultralow threshold	Finite-size Q, disorder
Super-BIC laser	Charge merging	Steeper finite-size Q scaling	Lattice-constant tolerance
CPA switch	Scattering zero	All-optical modulation	Beam alignment, bandwidth
CPA-laser	Pole-zero coincidence	Dual amplification/absorption	Fine-tuned gain–loss balance
Topological laser	Edge-state lasing	Disorder robustness	Threshold not topologically protected
Skin amplifier	NHSE localization	Directional small-signal gain	Asymmetric hopping, saturation, noise

penalty, gain saturation, fabrication sensitivity, or thermal load—that must be assessed quantitatively against the specific application. The most robust devices are those that anchor their functionality in topological invariants rather than fine-tuned parameters, and the most honest assessments compare not just the enhanced response but the enhanced response *together with* the enhanced noise.

Chapter 19

Open Problems and Outlook

Non-Hermitian photonics is a young and rapidly developing field. Many fundamental questions remain open, and new experimental platforms continue to reveal unexpected phenomena. This final chapter surveys the frontier as a set of tensions that drive the field forward. Each section identifies a concrete gap between the theoretical framework of this book and the physics that lies beyond it, explains why the gap matters, and points toward the ideas that may close it.

19.1 Nonlinear non-Hermitian photonics

19.1.1 Why linearity is not enough

The entire theoretical framework of this book assumes linear media: the permittivity ε_r is independent of the field amplitude, and the Maxwell operator is a fixed linear map. This assumption breaks down in two physically important regimes:

1. **Above the lasing threshold.** Once gain exceeds loss in a cavity mode, the field amplitude grows exponentially until gain saturation—a nonlinear effect—clamps it. The steady-state amplitude is set by the balance between saturated gain and loss, not by the linear eigenproblem.
2. **Strongly driven systems.** High-intensity light in a Kerr medium modifies the refractive index as $n_{\text{eff}} = n_0 + n_2 |E|^2$, creating an intensity-dependent effective Hamiltonian. In a non-Hermitian system, the Kerr shift can move the system toward or away from an exceptional point, coupling the EP physics to the field dynamics.

19.1.2 Gain saturation and the fate of exceptional points

Gain saturation introduces a nonlinear self-consistency problem: the effective Hamiltonian depends on the field amplitudes, which in turn depend on the Hamiltonian. Near an EP, the consequences are subtle. Zheng and Chong showed that in a coupled-cavity laser system, noise fluctuations shift the operating point away from the EP, effectively reducing the EP order and preventing the system from exploiting the full $\epsilon^{1/n}$ responsivity [51]. Moreover, the linearized Bogoliubov–de Gennes (BdG) Hamiltonian governing small fluctuations around the nonlinear

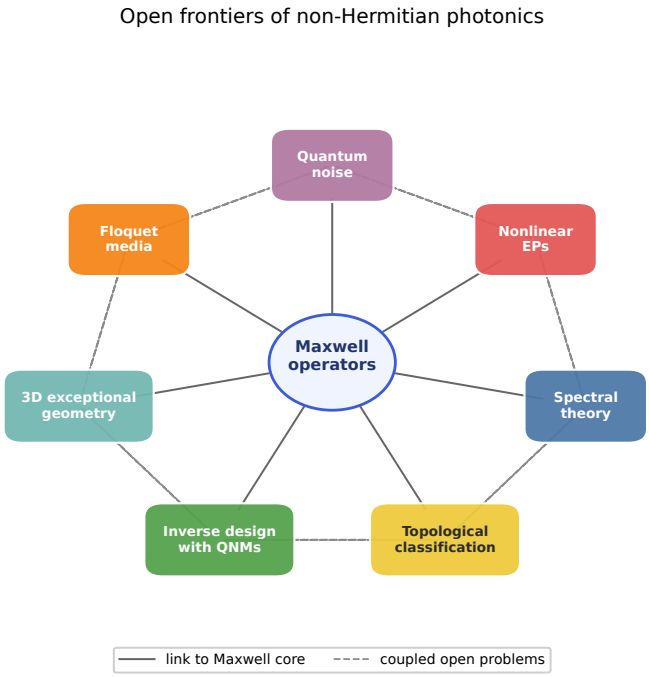


Figure 19.1 Seven coupled frontiers of Maxwell-first non-Hermitian photonics. Solid lines connect each frontier to the central Maxwell operator framework developed in this book; dashed lines indicate cross-frontier couplings where progress on one problem constrains or enables another. The frontiers are: nonlinear exceptional points (EPs), quantum noise limits, Floquet-driven media, three-dimensional exceptional geometry, inverse design via quasinormal modes (QNMs), rigorous spectral theory, and topological classification beyond two dimensions. Each is discussed in a section of this chapter.

steady state has its *own* EP structure, which can produce noise divergences *stronger* than the simple Petermann estimate from the linear system.

This result challenges the premise that operating a sensor near a linear EP automatically confers a metrological advantage. A full theory of nonlinear EP sensing must treat the noise, the nonlinearity, and the EP on an equal footing, rather than analyzing the linear EP first and adding noise afterward.

19.1.3 Nonlinear topology

Can topological invariants survive in a nonlinear system? For Kerr-type nonlinearity, Xia et al. demonstrated experimentally that local intensity-dependent refractive index changes can tune the global $\mathcal{PT}$ symmetry of a waveguide lattice, restoring or destroying non-Hermitian topological states [59]. This suggests that nonlinearity does not simply destroy topology: in suitable driven systems it can create an intensity-dependent topological phase diagram [130, 59, 131, 132, 133, 134, 135].

A systematic theory of nonlinear non-Hermitian topology [18, 136, 137, 138, 139, 140, 141, 142] is still missing. The obstacles include:

- Topological invariants are defined for linear eigenproblems; extending them to nonlinear fixed-point problems requires new mathematical tools.
- The skin effect, which depends on the spectral winding of the *linear* Hamiltonian, may be modified or destroyed by nonlinear mode coupling.

- Solitons in $\mathcal{PT}$-symmetric systems form continuous families rather than the isolated amplitude values found in generic dissipative systems [60, 58, 143, 144, 145, 146]; the stability and topological classification of these families remain open problems.

Open problem: nonlinear EP laser dynamics. Consider a pair of coupled ring resonators with Kerr nonlinearity, modelled by

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \omega_0 + g(|a_1|^2) + i\gamma_1 & \kappa \\ \kappa & \omega_0 + g(|a_2|^2) - i\gamma_2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \tag{19.1}$$

where $g(I) = g_0/(1 + I/I_{\mathrm{sat}})$ models gain saturation and γ_2 represents linear loss. At low intensity the system approaches a linear EP when $|\gamma_1 - \gamma_2| = 2|\kappa|$. But as $|a_1|^2$ grows, the saturated gain shifts γ_1 downward, and the instantaneous Hamiltonian moves *away* from the EP. The resulting limit cycle oscillates between near-EP and far-from-EP regimes, producing a characteristic "EP breathing" in the output intensity. Whether this breathing mode has any robust dynamical signature tied to the nearby degeneracy—or whether it is a generic feature of nonlinear saddle–node bifurcations near non-Hermitian degeneracies—is an open question that requires tools from both dynamical systems theory and non-Hermitian spectral theory.

Gain saturation and the Bogoliubov–de Gennes connection. The fluctuation dynamics around a nonlinear steady state are governed by a Bogoliubov–de Gennes (BdG) Hamiltonian that couples the signal and idler quadratures. Zheng and Chong showed that the BdG Hamiltonian can have its own EP structure that is *different* from the EP of the underlying nonlinear Hamiltonian [51]. The BdG EP can produce noise divergences that scale with a Petermann factor computed from the BdG eigenvectors, not from the original linearized system alone. This "BdG Petermann effect" is a nonlinear extension of the linear excess-noise mechanism. Its experimental signature—anomalous linewidth broadening controlled by the fluctuation operator rather than only by the instantaneous mean-field Hamiltonian—has not yet been isolated.

19.2 Quantum non-Hermitian photonics

19.2.1 From effective Hamiltonians to open quantum systems

The effective non-Hermitian Hamiltonian H_{eff} used throughout this book is derived from the full Lindblad master equation by conditioning on zero quantum jumps. This post-selected description captures the *conditional* dynamics of the photonic system but discards the quantum jumps (spontaneous emission events, photon absorption by the environment) that are essential for a complete picture of the noise and the measurement statistics.

The quantum noise theory of EP sensors [49] shows that the discarded quantum jumps reintroduce noise channels that can cancel the apparent EP responsivity enhancement in standard measurement settings. A consistent quantum theory of non-Hermitian photonics must therefore start from the full open quantum system—either the Lindblad equation or the quantum Langevin equations with input–output formalism—and derive the non-Hermitian effective Hamiltonian as a limit rather than a postulate.

19.2.2 Few-photon non-Hermitian physics

Most non-Hermitian photonic experiments operate in the classical regime (many photons per mode). In the few-photon regime, the effective Hamiltonian description can break down because the probability of quantum jumps during the observation time becomes significant. The interplay of non-Hermiticity, quantum fluctuations, and photon-number-resolved measurement creates a rich and largely unexplored landscape.

Specific open questions include:

1. Can the non-Hermitian skin effect be observed in a quantum photonic simulator with single-photon resolution?
2. Under what measurement protocol, if any, does the EP responsivity enhancement translate into improved precision at the single-photon level?
3. Can quantum error correction or reservoir engineering be used to control jump-induced decoherence in non-Hermitian topological states [147, 29, 148]?

19.2.3 Entanglement and biorthogonality

The biorthogonal basis (Appendix B) is a mathematical tool for expanding observables in a non-Hermitian system. Whether it has a direct quantum information interpretation—for example, whether the biorthogonal entanglement entropy captures physically relevant correlations in an open photonic system—is an open and active question.

19.3 Time-varying and Floquet non-Hermitian systems

19.3.1 Floquet engineering of non-Hermiticity

Time-periodic (Floquet) modulation of the refractive index creates effective non-Hermitian Hamiltonians with periodically driven gain, loss, and coupling. The Floquet formalism maps the time-dependent problem onto a static problem in an extended Hilbert space (the Floquet–Bloch space), where the effective Hamiltonian can exhibit EPs, $\mathcal{PT}$ symmetry, and topological phases that do not exist in any static system.

19.3.2 Floquet exceptional points

A Floquet EP occurs when two Floquet quasi-energies coalesce—not necessarily because the instantaneous Hamiltonian has an EP at any fixed time, but because the time-ordered evolution operator over one period has a Jordan-block structure. Floquet EPs can be created and controlled by the modulation amplitude and frequency, providing a temporal knob for EP physics in addition to spatial design.

Example: periodically modulated coupled cavities. Consider two coupled cavities with time-periodic detuning: $\delta(t) = \delta_0 \cos(\Omega t)$, where Ω is the modulation frequency. In the Floquet–Bloch representation, the effective Hamiltonian becomes an infinite-dimensional matrix coupling all Floquet harmonics $n\Omega$. A Floquet EP appears when the modulation amplitude δ_0 reaches a critical value at which two Floquet quasi-energy bands touch and coalesce. Unlike a static

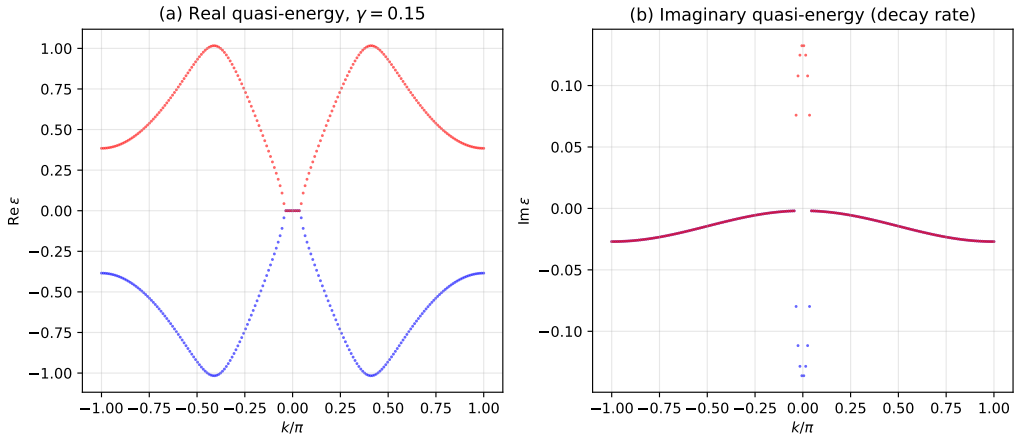


Figure 19.2 Floquet quasi-energy bands for a periodically driven non-Hermitian two-band model ($\gamma = 0.15$, $V = 0.8$, $\Omega = 2.5$). (a) Real part of the quasi-energy $\varepsilon(k)$, showing the band structure folded into the Floquet–Brillouin zone. (b) Imaginary part of $\varepsilon(k)$, representing the decay rate. The asymmetry between the two bands reflects the non-Hermiticity.

EP, the degeneracy is defined modulo the drive frequency and can involve different Floquet replicas; it can therefore be tuned by varying Ω or the modulation strength without changing the spatial geometry. Its stability, however, still depends on the usual codimension and symmetry constraints of the driven non-Hermitian operator.

The physical signature is a characteristic beating pattern in the transient response: near the Floquet EP, the temporal evolution contains a term proportional to $t\,e^{-i\varepsilon t}$ (the Jordan-block contribution), producing a linear-in-t envelope modulating the oscillatory response. This temporal signature is the Floquet analogue of the spatial mode coalescence at a static EP and can be measured in the time-domain transmission of a pulsed probe.

19.3.3 Open problems

The interplay of temporal and spatial periodicity in non-Hermitian systems creates several open problems:

- A general bulk-boundary correspondence for Floquet non-Hermitian systems is not yet established.
- The non-Hermitian skin effect in Floquet systems can depend on both the driving frequency and the spatial boundary conditions, leading to a richer phase diagram than the static case.
- Experimental realization of Floquet non-Hermitian photonics requires modulation at optical frequencies, which is technically demanding but may be achievable with electro-optic modulators or optomechanical coupling.

19.4 Non-Hermitian photonics in three dimensions

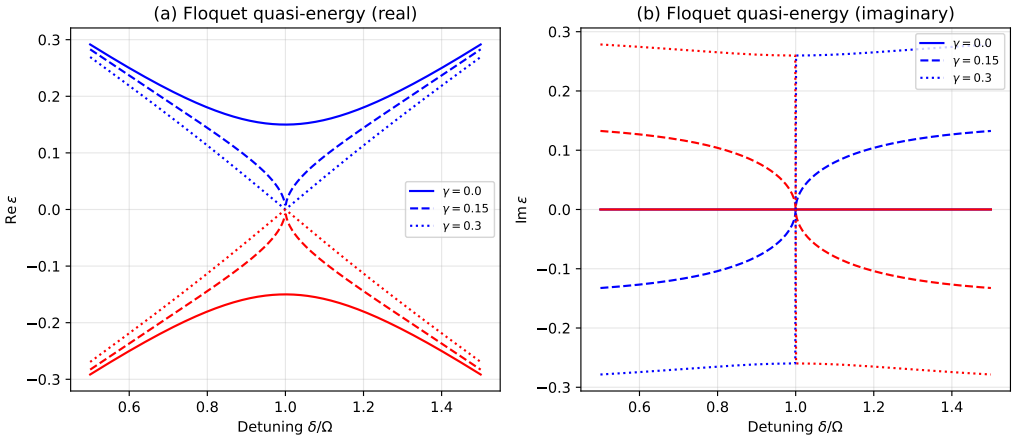


Figure 19.3 Floquet exceptional point in a periodically driven two-mode model. The rotating-wave Floquet Hamiltonian develops quasi-energy coalescences controlled by detuning and modulation amplitude. (a) Real quasi-energies show avoided crossings and coalescence as the non-Hermitian parameter increases. (b) Imaginary quasi-energies reveal the gain–loss splitting associated with the Floquet EP. The point of the figure is not a new device proposal but the mechanism: a time-periodic Maxwell system can acquire non-Hermitian degeneracies in Floquet space even when no static instantaneous Hamiltonian sits at an EP.

19.4.1 Exceptional lines and surfaces

Three-dimensional non-Hermitian photonic crystals support topological objects that have no lower-dimensional counterpart:

- **Exceptional lines:** one-dimensional curves in the 3D Brillouin zone where two bands coalesce. The topology of an exceptional line is characterized by a topological charge (a π Berry phase encircling the line) and a linking number with other exceptional lines.
- **Exceptional surfaces:** two-dimensional manifolds of degeneracy, which can occur when additional constraints, symmetries, or material parameters reduce the effective codimension.
- **Non-Abelian charges:** when three or more bands are involved, the braid group replaces the simple winding number, and the topological classification becomes non-Abelian.
- **Three-dimensional skin effect:** an extensive set of eigenstates can localize on boundaries of a 3D sample, with the localization pattern depending on spectral winding and sample geometry.

Exceptional lines from Maxwell symmetries. In a generic two-band non-Hermitian problem, the EP condition imposes two real constraints, so in a three-dimensional control space the EP set is generically a curve. Photonic crystals can realize such control spaces through Bloch wave vector, material parameters, or symmetry constraints. The Maxwell vector nature adds polarization structure: near an exceptional line the coalescing modes may carry a half-integer vorticity or Berry phase, but the precise relation to polarization vortices depends on the radiation channels and symmetries of the particular structure.

Knotted exceptional lines. Exceptional lines in 3D can form knots and links. Two exceptional lines can have a nontrivial linking number, meaning they cannot be separated by any continuous deformation that preserves the band gap. This linking is a topological invariant that has no

2D or 1D analogue and represents a genuinely new form of non-Hermitian topology. Whether knotted exceptional lines can be realised in a photonic crystal with practical material parameters is an open fabrication challenge.

3D non-Hermitian skin effect and geometry dependence. In three dimensions, the skin effect can become geometry-dependent: the localization boundary is determined by the interplay of spectral winding in different spatial directions and the sample shape. In simple separable models a rectangular sample may exhibit corner accumulation, whereas other boundary geometries can redistribute the same non-Bloch decay toward edges or surfaces. The geometry dependence means that the finite-sample OBC spectrum is not determined by the conventional Bloch band structure alone.

19.4.2 Fabrication and characterization challenges

Experimental realization of 3D non-Hermitian photonic structures requires fabrication of three-dimensional lattices with controlled, spatially varying gain and loss—a significantly harder problem than the 1D or 2D case. Two-photon lithography, self-assembly, and holographic patterning are candidate technologies, but none has yet produced a 3D non-Hermitian photonic crystal with sufficient quality for topological measurements.

Characterization is equally challenging: the 3D bulk–boundary correspondence predicts surface, hinge, and corner modes that require spatially resolved measurements in all three dimensions. Near-field scanning or tomographic reconstruction may be needed.

19.5 Rigorous spectral theory of open Maxwell operators

19.5.1 Completeness of QNM expansions

The quasinormal-mode expansion (Chapter 6) is widely used but not rigorously justified in full generality. The QNMs of an open resonator form a countable set, but their completeness—in the sense that any compactly supported source field can be expanded in QNMs—has been proven only for specific geometries (1D slabs, spheres). For general 3D resonators with complex shapes and dispersive materials, the completeness question remains open.

The difficulty is mathematical: QNM fields diverge outside the resonator (they satisfy outgoing boundary conditions at infinity), so the expansion converges only in a distributional or regularized sense (e.g., inside the PML region). A rigorous theory must specify the function space in which completeness holds and the rate of convergence of the modal sum.

What a completeness proof would need. A satisfactory proof of QNM completeness for a general 3D open Maxwell system would need to:

1. Specify the function space $\mathcal{H}$ of source/field distributions (e.g., L^2 functions with compact support inside the resonator volume).
2. Prove that the set of QNM fields $\{\mathbf{E}_m(\mathbf{r})\}_{m=1}^{\infty}$, regularized by a PML or complex-coordinate transformation, is dense in $\mathcal{H}$.

3. Provide error bounds for the truncated expansion: given N modes, how large is the remainder in terms of the quality factors Q_m and the spectral density near the target frequency?
4. Handle the branch-cut contributions from the continuous spectrum (radiation modes), either by including them explicitly or by proving that the PML discretizes them into a sequence of discrete poles that converges to the branch cut.

A partial result in this direction was obtained for 1D scattering problems using Mittag-Leffler-type meromorphic expansions of the Green function. Such results rely on the analytic structure and growth bounds of the particular scattering problem; extending them to 3D Maxwell systems with dispersive media requires controlling the Green tensor at large complex frequency and accounting for branch cuts—nontrivial analytic estimates.

19.5.2 Continuous spectrum and branch cuts

In addition to the discrete QNM poles, the Green tensor of an open Maxwell system has branch-cut contributions from the continuous spectrum (radiation modes). These contributions are responsible for the "PML modes" of Chapter 17 and are essential for completeness. A full modal expansion must include both the discrete poles and the branch-cut integral, with quantitative estimates for the error of truncating the branch-cut contribution.

19.5.3 Spectral singularities

A spectral singularity is a point on the real frequency axis where the resolvent of the Maxwell operator diverges without an associated normalizable eigenfunction—it is neither a bound state nor a scattering resonance, but a divergence of the scattering amplitude. Spectral singularities have been linked to lasing and CPA in $\mathcal{PT}$-symmetric systems, but their mathematical theory (existence, genericity, stability under perturbation) is not yet complete for the full Maxwell operator with dispersive materials.

19.6 Inverse design with non-Hermitian constraints

Computational inverse design (topology optimization, adjoint methods) has become central to Hermitian photonic device design. Extending these methods to non-Hermitian targets—designing a structure that places an EP at a specified frequency, maximizes the Petermann factor, or realizes a specific topological phase—requires new objective functions and sensitivity formulas.

The biorthogonal perturbation theory of Chapter 6 provides the adjoint sensitivities needed for gradient-based optimization, but turning these sensitivities into robust design algorithms remains a nontrivial numerical problem.

- **EP targeting:** the EP is a codimension-2 degeneracy in parameter space; designing a structure that hits an EP at a target frequency requires simultaneous tuning of two real parameters.
- **Topological constraints:** enforcing a topological invariant (Chern number, winding number, BIC charge) during optimization requires discrete-valued objective functions, which are not differentiable.

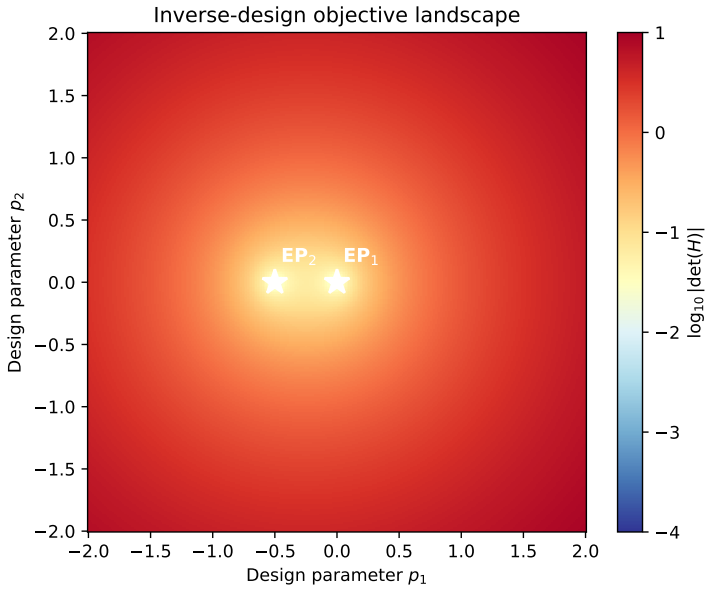


Figure 19.4 Inverse-design objective landscape $\log_{10}|\det H|$ in a two-parameter space. The two exceptional points (white stars) appear as deep minima where $\det H \to 0$. Gradient-based optimizers must navigate this non-convex landscape; adjoint sensitivity formulas from QNM perturbation theory provide the required gradients.

- **Physical realizability:** optimized permittivity profiles must satisfy the Kramers–Kronig relations (Chapter 5), and the gain/loss distribution must be achievable with a practical pump geometry.

Adjoint sensitivity for non-Hermitian objectives. The key technical ingredient for gradient-based inverse design is the adjoint sensitivity: the derivative of an objective function $\mathcal{F}$ with respect to a material parameter p (e.g., a pixel of ε_{r}). For a Hermitian system, the adjoint field is the standard conjugate field. For a non-Hermitian system, the adjoint field must be computed from the *left* eigenvector or the adjoint Green function, which introduces the biorthogonal framework of Chapter 6 into the optimization loop.

For a nondispersive scalar perturbation, the leading sensitivity of a QNM eigenfrequency $\tilde{\omega}_m$ to a local perturbation $\delta\varepsilon_{\mathrm{r}}(\mathbf{r})$ has the schematic form

$$\frac{\partial\tilde{\omega}_m}{\partial\varepsilon_{\mathrm{r}}(\mathbf{r})} = -\frac{\tilde{\omega}_m}{2}\frac{\tilde{\mathbf{E}}_m(\mathbf{r})\cdot\tilde{\mathbf{E}}_m(\mathbf{r})}{\langle\tilde{\mathbf{E}}_m|\varepsilon_{\mathrm{r}}|\tilde{\mathbf{E}}_m\rangle_{\mathrm{bi}}}, \tag{19.2}$$

where the denominator represents the appropriate PML-regularized QNM normalization of Chapter 6; dispersive, anisotropic, or magnetic media require the corresponding generalized bilinear form. This formula is the non-Hermitian generalisation of standard cavity perturbation theory and provides an efficient route to gradient computation: one adjoint solve per objective, regardless of the number of design parameters.

Multi-objective design: EP + topology + noise. The most ambitious inverse-design target combines an EP at a specified frequency, a topological invariant of a given value, and a noise figure below a threshold. This is a multi-objective, mixed-integer, Kramers–Kronig-constrained optimization problem that lies beyond current methods. Developing efficient algorithms for this

class of problems would have immediate practical impact on the design of EP sensors, topological lasers, and non-Hermitian metasurfaces.

19.7 Energy, entropy, and thermodynamics

The auxiliary-field construction of Chapter 5 shows how to define a conserved energy for the full field–matter system, even in the presence of dispersion and dissipation. This construction embeds the non-Hermitian Maxwell problem into a larger Hermitian system, where standard thermodynamic concepts (entropy, temperature, free energy) are well defined.

Whether this embedding can be used to define meaningful thermodynamic quantities for the non-Hermitian photonic subsystem, and which quantities remain subsystem-dependent bookkeeping choices, is an intriguing open question. Specific directions include:

- Defining an entropy production rate for a non-Hermitian photonic mode in terms of the energy flow into the auxiliary (matter) degrees of freedom.
- Connecting the Petermann factor to a thermodynamic cost of maintaining a non-orthogonal mode structure.
- Establishing fluctuation–dissipation relations for non-Hermitian photonic systems that go beyond the standard equilibrium results.

19.8 Beyond photonics

The mathematical framework developed in this book—complex eigenvalues, biorthogonality, exceptional points, point-gap topology, the skin effect—applies to any wave system with gain, loss, or non-reciprocity. Active research frontiers include:

Acoustics and mechanics. Coupled acoustic resonators with active feedback and metamaterial lattices with programmed damping provide macroscopic platforms where non-Hermitian topology can be tested with tabletop experiments. The slower wave speeds and larger wavelengths make spatial measurements easier than in photonics.

Electronics. Non-Hermitian circuit lattices (topolectric circuits) with resistive and reactive elements can realize non-Hermitian tight-binding models with high controllability. The skin effect, point-gap topology, and higher-order corner modes have all been observed in electronic circuits.

Cold atoms and condensates. Atomic gases with controlled loss (via resonant light or atom removal) and effective non-Hermitian band structures provide a quantum platform where the interplay of non-Hermiticity and interactions can be studied in a controllable setting.

Condensed matter. Quasiparticle lifetimes in correlated electron systems generate effective non-Hermitian self-energies. Point-gap topology and non-Hermitian self-energy effects have been proposed as possible ingredients in anomalous surface responses of certain topological semimetals, although their direct relation to the skin effect in electronic materials remains subtle.

19.9 Toward a Maxwell-first synthesis

This monograph has consistently derived non-Hermitian phenomena from the Maxwell equations rather than postulating them from finite-dimensional models. Several directions remain where a Maxwell-first approach could sharpen the understanding:

Topological classification of Maxwell systems. The 38-fold classification of non-Hermitian topological phases (Chapter 12) was developed for tight-binding Hamiltonians. Lifting it to the continuous Maxwell operator—with its vector nature, gauge redundancy, and frequency-dependent material response—is an open problem that could reveal new topological phases unique to electromagnetism.

The vector nature is the most immediate obstacle: the Maxwell operator acts on six-component fields $(\mathbf{E}, \mathbf{H})$, not on scalar wave functions. The polarization degree of freedom couples to the spatial structure through the boundary conditions and the dielectric tensor, creating a richer symmetry landscape than the scalar case. For example, a photonic crystal with time-reversal symmetry and C_4 spatial symmetry can support polarization-dependent topological phases where the TE and TM sectors have different topological invariants—a situation that has no analogue in electronic systems where spin–orbit coupling mixes the spin channels.

Unified energy–topology–noise framework. The conserved energy (Chapter 5), the topological invariants (Chapters 11–15), and the noise limits (Chapter 18) are currently treated as separate aspects of non-Hermitian photonics. A unified framework that connects energy conservation, topological robustness, and noise bounds would be a major theoretical advance—and one that can only be formulated at the Maxwell level, where all three ingredients coexist naturally.

A hint of such a connection comes from the Petermann factor: it appears both as a noise amplification factor (Chapter 6) and as a measure of the non-orthogonality of the eigenmodes. Near an EP, the Petermann factor diverges, the topological charge of the EP is non-trivial, and the energy stored in the mode becomes anomalously large relative to the input power. Whether these three divergences are different facets of a single geometric–thermodynamic quantity—perhaps a non-Hermitian analogue of the Fisher information metric—is a compelling open question.

A constrained optimisation perspective. The unification problem can be cast as a constrained optimisation over the design space of non-Hermitian photonic structures. Let $\mathbf{p}$ denote the design parameters (geometry, material composition, gain–loss distribution). A device design must simultaneously satisfy:

1. *Topological robustness*: the topological invariant $\nu(\mathbf{p})$ (Chern number, winding number, or BIC charge) must take a specified integer value.
2. *Noise bounds*: the Petermann factor $K(\mathbf{p})$ must remain below a threshold set by the application-level signal-to-noise requirement.
3. *Gain saturation*: the steady-state intensity $I_{\mathrm{ss}}(\mathbf{p})$ must lie within the linear regime $I_{\mathrm{ss}} < I_{\mathrm{sat}}$ for the linear non-Hermitian model to be valid.
4. *Fabrication robustness*: the sensitivity $\|\partial\nu/\partial\mathbf{p}\|$ must be small enough that fabrication disorder does not close the relevant gap or destroy the desired boundary response.

These four constraints define a feasible region in design space. The open question is whether this feasible region is generically nonempty: can one simultaneously have strong topological robustness *and* low noise *and* operate below saturation? The Petermann-factor divergence at EPs (Chapter 8) suggests a fundamental tension between EP-based enhancement and noise, but it is not known whether alternative mechanisms—such as symmetry- or topology-guided BICs (Chapter 16)—can improve the tradeoff in practical devices.

A formal optimisation objective that captures this tension is

$$\min_{\mathbf{p}} \Big\{ -Q_{\text{topo}}(\mathbf{p}) + \lambda_1 \, K(\mathbf{p}) + \lambda_2 \, \frac{I_{\text{ss}}(\mathbf{p})}{I_{\text{sat}}} + \lambda_3 \, R_{\text{fab}}(\mathbf{p}) \Big\}, \tag{19.3}$$

where Q_{topo} is a topological quality metric (e.g. the bulk or point gap supporting the invariant), K is the Petermann factor, $I_{\text{ss}}/I_{\text{sat}}$ measures the proximity to gain saturation, R_{fab} is a fabrication-sensitivity measure, and λ_i are Lagrange-like weights. Solving this optimisation requires the adjoint sensitivity formulas of Section 19.6—themselves an open problem for non-Hermitian objectives, where the adjoint operator is not the Hermitian conjugate but the biorthogonal partner.

Stable invariants for nonlinear Maxwell operators. The deepest mathematical question in this programme is whether topological invariants can be defined for nonlinear, noisy Maxwell operators. The current classification (Chapter 12) assumes a linear Bloch Hamiltonian; gain saturation (Chapter 5) makes the effective permittivity intensity-dependent, so the Hamiltonian itself depends on the solution. Whether a meaningful topological invariant survives in this self-consistent regime—and whether it can be computed from a linearised operator around the steady state—is unknown. Partial answers exist for specific models (e.g. the nonlinear SSH model with Kerr coupling [61]), but a general theory for continuous Maxwell systems with dispersive, nonlinear, and noisy materials remains an open frontier.

From Maxwell to field theory. The Maxwell operator in a periodic non-Hermitian medium is formally a non-self-adjoint elliptic operator on a fibre bundle over the Brillouin torus. The topological invariants of Chapters 11 through 15 are characteristic classes of this bundle. A field-theoretic formulation—using the language of differential geometry and gauge theory—could clarify which invariants survive for continuous, dispersive Maxwell operators. One possible starting point is a non-Hermitian analogue of index theorems that relate boundary responses to a topological index computed from the bulk Maxwell operator. Whether such a theorem exists for non-self-adjoint operators—and whether it can be proven with the regularity available for Maxwell systems with dispersive media—is a deep mathematical question at the intersection of spectral theory, differential geometry, and photonics.

A field in motion

Non-Hermitian photonics has grown quickly from a mathematical theme into an experimental and theoretical field. Combining Maxwell's equations with non-Hermitian spectral theory gives a disciplined way to pose its open problems, from nonlinear topology to the thermodynamics of non-orthogonal modes. The tools assembled here—the Maxwell eigen-

problem, the biorthogonal expansion, the GBZ, and topological charge—are starting points for that next stage.

Chapter A

Vector Calculus and Maxwell Notation

This appendix collects the vector-calculus identities, integration formulas, and notational conventions used throughout the book. All results are stated for three-dimensional space with Cartesian, or where noted, curvilinear coordinates.

A.1 Curl identities

The following identities are used repeatedly in the derivation of Maxwell inner products and self-adjointness conditions.

Double curl.

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}. \tag{A.1}$$

Curl of a product.

$$\nabla \times (f\,\mathbf{F}) = f\,\nabla \times \mathbf{F} + (\nabla f) \times \mathbf{F}. \tag{A.2}$$

Divergence of a cross product.

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \nabla \times \mathbf{F} - \mathbf{F} \cdot \nabla \times \mathbf{G}. \tag{A.3}$$

This identity is the vector-calculus engine behind every integration-by-parts argument in this book: integrating both sides over a volume and applying the divergence theorem converts a volume integral of curls into a surface integral.

A.2 Integration by parts for curl operators

Let $\mathbf{F}$ and $\mathbf{G}$ be smooth vector fields on a domain Ω with boundary $\partial\Omega$ and outward normal $\hat{\mathbf{n}}$. Then

$$\int_\Omega \mathbf{G} \cdot \nabla \times \mathbf{F}\ \mathrm{d}V = \int_\Omega \mathbf{F} \cdot \nabla \times \mathbf{G}\ \mathrm{d}V + \oint_{\partial\Omega} (\mathbf{F} \times \mathbf{G}) \cdot \hat{\mathbf{n}}\ \mathrm{d}S. \tag{A.4}$$

This is the fundamental self-adjointness identity for the curl operator. Setting $\mathbf{F} = \mathbf{E}$ and $\mathbf{G} = \mu_{\mathrm{r}}^{-1}\nabla \times \mathbf{E}$ (or the corresponding magnetic formulation) and requiring the surface integral to vanish yields the Hermiticity condition for the Maxwell curl–curl operator (Chapter 2).

A.2.1 Boundary conditions that kill the surface term

The surface integral in (A.4) vanishes if any of the following conditions hold on $\partial\Omega$:

1. **Perfect electric conductor (PEC):** $\hat{\mathbf{n}} \times \mathbf{E} = 0$ on $\partial\Omega$.
2. **Perfect magnetic conductor (PMC):** $\hat{\mathbf{n}} \times \mathbf{H} = 0$ on $\partial\Omega$.
3. **Periodic boundary conditions:** $\mathbf{F}$ and $\mathbf{G}$ satisfy Bloch periodicity on opposite faces, so the contributions cancel pairwise.
4. **Fields decay sufficiently fast:** for an unbounded domain, if $|\mathbf{F}|$ and $|\mathbf{G}|$ decay faster than $1/r$ as $r \to \infty$, the surface integral at infinity vanishes.

When the surface term does *not* vanish—as for outgoing radiation boundary conditions (Silver–Müller condition) or for modes with $|\mathbf{F}| \sim e^{+\operatorname{Im} k\, r}$—the curl–curl operator is no longer self-adjoint, and the system is non-Hermitian (Chapter 4).

A.3 The divergence theorem and Poynting's theorem

The divergence theorem relates a volume integral of a divergence to a surface integral:

$$\int_{\Omega} \nabla \cdot \mathbf{F} \, \mathrm{d}V = \oint_{\partial\Omega} \mathbf{F} \cdot \hat{\mathbf{n}} \, \mathrm{d}S. \tag{A.5}$$

Applying this to the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ gives the integral form of Poynting's theorem:

$$-\frac{\partial}{\partial t} \int_{\Omega} u \, \mathrm{d}V = \oint_{\partial\Omega} \mathbf{S} \cdot \hat{\mathbf{n}} \, \mathrm{d}S + \int_{\Omega} \sigma \, |\mathbf{E}|^2 \, \mathrm{d}V, \tag{A.6}$$

where u is the electromagnetic energy density and σ is the conductivity (representing ohmic loss).

A.4 Bloch-periodic boundary terms

For a photonic crystal with lattice vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ and Bloch wave vector $\mathbf{k}$, the fields satisfy

$$\mathbf{F}(\mathbf{r} + \mathbf{a}_j) = e^{i\mathbf{k}\cdot\mathbf{a}_j} \, \mathbf{F}(\mathbf{r}). \tag{A.7}$$

On a unit cell Ω_{cell} with faces S_j^+ and S_j^- related by $S_j^+ = S_j^- + \mathbf{a}_j$, the surface integral in (A.4) cancels pairwise when the two fields belong to compatible Bloch sectors. For example, if both fields carry the same real Bloch wave vector and the inner product uses the usual conjugation, or if an unconjugated bilinear form pairs $\mathbf{k}$ with $-\mathbf{k}$, the two opposite faces contribute

$$\int_{S_j^+} (\mathbf{F} \times \mathbf{G}) \cdot \hat{\mathbf{n}} \, \mathrm{d}S + \int_{S_j^-} (\mathbf{F} \times \mathbf{G}) \cdot \hat{\mathbf{n}} \, \mathrm{d}S = 0. \tag{A.8}$$

The cancellation is not simply the algebraic difference $e^{i\mathbf{k}\cdot\mathbf{a}_j} - e^{-i\mathbf{k}\cdot\mathbf{a}_j}$; it also uses the opposite orientations of the outward normals and the adjoint boundary condition. For real $\mathbf{k}$ in a lossless periodic medium this gives the usual self-adjoint Bloch Maxwell problem. For complex $\mathbf{k}$, as in non-Bloch or GBZ constructions, the operator is generally non-self-adjoint and one must pair right and adjoint Bloch sectors explicitly.

A.5 Silver–Müller radiation condition

For an open scatterer in three dimensions, the outgoing-wave condition at infinity is the Silver–Müller condition:

$$\lim_{r\to\infty} r\left[\hat{\mathbf{r}} \times (\nabla \times \mathbf{E}) - ik\,\mathbf{E}\right] = 0, \tag{A.9}$$

uniformly in all directions $\hat{\mathbf{r}}$. This condition selects outgoing spherical waves and excludes incoming waves. It is the mathematical statement of the "no incoming radiation" assumption that makes the scattering problem well posed and the Maxwell operator non-self-adjoint (Chapter 4).

A.6 Notation conventions

A.6.1 Time convention

The time-harmonic convention throughout this book is $e^{-i\omega t}$. With this choice:

- A passive medium has $\operatorname{Im}\varepsilon_{\mathrm{r}} > 0$ (absorption).
- A gain medium has $\operatorname{Im}\varepsilon_{\mathrm{r}} < 0$ (amplification).
- A QNM eigenfrequency $\tilde{\omega} = \omega_r - i\gamma$ with $\gamma > 0$ represents a decaying mode.

A.6.2 Field quantities

Symbol	Name	SI Unit
$\mathbf{E}$	Electric field	V/m
$\mathbf{H}$	Magnetic field	A/m
$\mathbf{D}$	Electric displacement	C/m^2
$\mathbf{B}$	Magnetic induction	T
$\mathbf{S} = \mathbf{E} \times \mathbf{H}$	Poynting vector	W/m^2
ε_{r}	Relative permittivity	dimensionless
μ_{r}	Relative permeability	dimensionless
ω	Angular frequency	rad/s
$\tilde{\omega}$	Complex eigenfrequency	rad/s
$\mathbf{k}$	Wave vector	1/m
c	Speed of light in vacuum	m/s

A.6.3 Operators and inner products

Symbol	Meaning
$\nabla\times$	Curl operator
$\nabla\cdot$	Divergence operator
∇	Gradient operator
$\hat{\Theta}$	Maxwell curl–curl operator
$\mathbf{S}$	Scattering matrix
$\mathbf{T}$	Transfer matrix
$\langle\cdot\vert\cdot\rangle$	Standard (conjugated) inner product
$(\cdot,\cdot)$	Unconjugated bilinear form
Q	Quality factor
W	Winding number
$\mathcal{PT}$	Parity–time operator

A.7 Electromagnetic reciprocity from vector calculus

The Lorentz reciprocity theorem, introduced physically in Chapter 2, can be derived purely from the vector calculus identities of this appendix.

A.7.1 Statement and proof

Consider two solutions $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ of Maxwell's equations at the same frequency ω in a medium with permittivity tensor $\boldsymbol{\varepsilon}$ and permeability μ_0. The sources are localised current densities $\mathbf{J}_1$ and $\mathbf{J}_2$. Define the *reaction* of field 1 on source 2:

$$\langle 1, 2\rangle = \int_V \mathbf{E}_1 \cdot \mathbf{J}_2 \; \mathrm{d}^3 r. \tag{A.10}$$

Lorentz reciprocity states that $\langle 1,2\rangle = \langle 2,1\rangle$ whenever $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathsf{T}}$ (the permittivity tensor is symmetric, not necessarily Hermitian).

Proof. Start from the divergence-of-cross-product identity (A.3):

$$\nabla\cdot(\mathbf{E}_1\times\mathbf{H}_2 - \mathbf{E}_2\times\mathbf{H}_1) = \mathbf{H}_2\cdot(\nabla\times\mathbf{E}_1) - \mathbf{E}_1\cdot(\nabla\times\mathbf{H}_2) - \mathbf{H}_1\cdot(\nabla\times\mathbf{E}_2) + \mathbf{E}_2\cdot(\nabla\times\mathbf{H}_1). \tag{A.11}$$

Substituting Maxwell's curl equations $\nabla\times\mathbf{E}_j = i\omega\mu_0\mathbf{H}_j$ and $\nabla\times\mathbf{H}_j = -i\omega\boldsymbol{\varepsilon}\mathbf{E}_j + \mathbf{J}_j$ into the right-hand side:

$$\begin{aligned} \text{RHS} = {} & i\omega\mu_0\,\mathbf{H}_2\cdot\mathbf{H}_1 + i\omega\,\mathbf{E}_1\cdot\boldsymbol{\varepsilon}\mathbf{E}_2 - \mathbf{E}_1\cdot\mathbf{J}_2 \\ & - i\omega\mu_0\,\mathbf{H}_1\cdot\mathbf{H}_2 - i\omega\,\mathbf{E}_2\cdot\boldsymbol{\varepsilon}\mathbf{E}_1 + \mathbf{E}_2\cdot\mathbf{J}_1. \end{aligned} \tag{A.12}$$

The μ_0 terms cancel. The permittivity terms give $i\omega(\mathbf{E}_1 \cdot \boldsymbol{\varepsilon}\mathbf{E}_2 - \mathbf{E}_2 \cdot \boldsymbol{\varepsilon}\mathbf{E}_1)$, which vanishes if and only if $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathsf{T}}$. What remains is

$$\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) = \mathbf{E}_2 \cdot \mathbf{J}_1 - \mathbf{E}_1 \cdot \mathbf{J}_2. \tag{A.13}$$

Integrating over a volume V that contains both sources and applying the divergence theorem, the left-hand side becomes a surface integral that vanishes when the fields satisfy Silver–Müller radiation conditions at infinity (or when the surface is a PEC boundary):

$$\oint_{\partial V} (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) \cdot \hat{\mathbf{n}} \, \mathrm{d}S = 0. \tag{A.14}$$

Therefore $\langle 2, 1\rangle - \langle 1, 2\rangle = 0$, proving Lorentz reciprocity.

A.7.2 Reciprocity versus Hermiticity

The proof makes clear the distinction emphasised in Chapter 2: reciprocity requires $\boldsymbol{\varepsilon}^{\mathsf{T}} = \boldsymbol{\varepsilon}$ (symmetric), while Hermiticity of the Maxwell operator requires $\boldsymbol{\varepsilon}^{\dagger} = \boldsymbol{\varepsilon}$ (Hermitian, i.e. real for a scalar permittivity). A medium with complex symmetric $\boldsymbol{\varepsilon}$ (such as a lossy isotropic dielectric) is reciprocal but not Hermitian. Conversely, a magneto-optical medium with real antisymmetric off-diagonal elements is Hermitian (lossless) but not reciprocal.

This distinction is central to non-Hermitian photonics: most passive non-Hermitian systems (lossy dielectrics, absorbing metasurfaces) are reciprocal, and reciprocity constrains their scattering matrices and coupled-mode models. Reciprocity can fail when the material tensor, temporal modulation, active feedback, or an effective modal reduction lacks the relevant transpose symmetry. Such non-reciprocal or asymmetric effective couplings are one route to point-gap winding and the non-Hermitian skin effect of Chapter 14, although the precise criterion is spectral rather than simply the presence or absence of material loss.

Chapter B

Non-Normal Linear Algebra

This appendix collects the linear-algebra results that underpin the non-Hermitian spectral theory used throughout the book. The central theme is *non-normality*: a matrix A is normal if $AA^\dagger = A^\dagger A$; Hermitian, unitary, and skew-Hermitian matrices are all normal. A non-normal matrix can have a complete set of eigenvectors, but they are not orthogonal, and the spectral behavior under perturbation is qualitatively different from the normal case. Non-Hermitian photonic operators—the Maxwell operator with complex ε_{r}, the scattering matrix near an EP, the effective Hamiltonian of a coupled-mode model—are generically non-normal.

B.1 Left and right eigenvectors

For a matrix $A \in \mathbb{C}^{n\times n}$, the *right eigenvector* $|r_j\rangle$ and the *left eigenvector* $\langle l_j|$ associated with eigenvalue λ_j satisfy

$$A\,|r_j\rangle = \lambda_j\,|r_j\rangle, \qquad \langle l_j|\,A = \lambda_j\,\langle l_j|. \tag{B.1}$$

Equivalently, $\langle l_j|$ is a right eigenvector of $A^\dagger$ with eigenvalue λ_j^*.

For a normal matrix, left and right eigenvectors coincide (up to normalization): $|l_j\rangle = |r_j\rangle$. For a non-normal matrix, they differ, and the angle between corresponding left and right eigenvectors can be arbitrarily small—a situation that has dramatic physical consequences (enhanced noise, sensitivity to perturbation, mode non-orthogonality).

B.2 Biorthogonality

If A has n distinct eigenvalues, the left and right eigenvectors form a *biorthogonal* system:

$$\langle l_i|r_j\rangle = \delta_{ij}\,N_j, \tag{B.2}$$

where $N_j = \langle l_j|r_j\rangle$ is a nonzero normalization constant (which is generally complex for non-normal matrices). The biorthogonal normalization convention sets $N_j = 1$ for all j by rescaling.

The spectral decomposition of A takes the biorthogonal form

$$A = \sum_{j=1}^{n} \lambda_j\,|r_j\rangle\langle l_j|, \tag{B.3}$$

and the resolvent is

$$(z - A)^{-1} = \sum_{j=1}^{n} \frac{|r_j\rangle\langle l_j|}{z - \lambda_j}. \tag{B.4}$$

These formulas generalize the Hermitian spectral theorem to the non-normal case. They are the mathematical foundation for the quasinormal-mode expansion (Chapter 6) and the biorthogonal Berry phase (Chapter 12).

B.3 Jordan chains and defective eigenvalues

When two or more eigenvalues coalesce, the eigenvectors may also coalesce, producing a *defective* eigenvalue—an exceptional point (Chapter 8).

B.3.1 Jordan normal form

At a defective eigenvalue λ_0 of algebraic multiplicity m and geometric multiplicity $g < m$, the matrix A cannot be diagonalized. Instead, the relevant block of the Jordan normal form is

$$J_m(\lambda_0) = \begin{pmatrix} \lambda_0 & 1 & 0 & \cdots & 0 \\ 0 & \lambda_0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_0 & 1 \\ 0 & \cdots & 0 & 0 & \lambda_0 \end{pmatrix} \in \mathbb{C}^{m\times m}. \tag{B.5}$$

B.3.2 Generalized eigenvectors

The Jordan chain associated with $J_m(\lambda_0)$ consists of vectors $|v_0\rangle, |v_1\rangle, \ldots, |v_{m-1}\rangle$ satisfying

$$(A - \lambda_0)\,|v_0\rangle = 0, \qquad (A - \lambda_0)\,|v_p\rangle = |v_{p-1}\rangle, \quad p = 1, \ldots, m-1. \tag{B.6}$$

Only $|v_0\rangle$ is a true eigenvector; the higher-order vectors $|v_p\rangle$ are *generalized eigenvectors*. At a second-order EP ($m = 2$), there is one eigenvector and one generalized eigenvector; the biorthogonal decomposition (B.3) breaks down and must be replaced by the Jordan decomposition.

B.3.3 Resolvent near an EP

Near a second-order EP at λ_0, the resolvent has a second-order pole:

$$(z - A)^{-1} \approx \frac{|v_1\rangle\langle w_0|}{(z - \lambda_0)^2} + \frac{|v_0\rangle\langle w_0| + |v_1\rangle\langle w_1|}{z - \lambda_0} + \text{regular}, \tag{B.7}$$

where $\langle w_0|$ and $\langle w_1|$ are the left Jordan-chain vectors. The double pole produces algebraic (rather than exponential) temporal growth—the hallmark of an EP in time-domain experiments and the origin of the enhanced transient response.

B.4 Petermann factor and excess noise

The Petermann factor quantifies the enhancement of spontaneous emission (and, more generally, noise) in a non-normal system. For a mode with right eigenvector $|r_j\rangle$ and left eigenvector $\langle l_j|$, the Petermann factor is

$$K_j = \frac{\langle r_j|r_j\rangle \, \langle l_j|l_j\rangle}{|\langle l_j|r_j\rangle|^2} \geq 1, \tag{B.8}$$

with equality if and only if $|l_j\rangle = |r_j\rangle$ (the normal case).

B.4.1 Physical interpretation

In a laser, spontaneous emission couples to the lasing mode with an amplitude proportional to $|\langle l_j|r_j\rangle|^{-1}$. When the left and right eigenvectors are nearly orthogonal (as happens near an EP), this coupling is enhanced, broadening the laser linewidth by the factor K_j:

$$\Delta\nu = K_j \, \Delta\nu_{\mathrm{ST}}, \tag{B.9}$$

where $\Delta\nu_{\mathrm{ST}}$ is the Schawlow–Townes linewidth.

B.4.2 Scaling near an EP

Near a second-order EP, the overlap $\langle l_j|r_j\rangle$ vanishes as $\sqrt{\epsilon}$ where ϵ is the distance from the EP in parameter space. Therefore the Petermann factor diverges as

$$K_j \sim \frac{1}{\epsilon} \tag{B.10}$$

near the EP. In common equal-resource sensing models, the associated noise growth can offset the enhanced frequency splitting $\Delta\omega \sim \sqrt{\epsilon}$ in a signal-to-noise proxy, as discussed in Chapter 18. The quantitative cancellation is not universal; it depends on the measurement protocol, noise channels, and normalization of the sensor response.

For an n-th order EP, analogous power-law divergences appear, but the resulting precision bound must be evaluated from the full noise model rather than inferred from the eigenvalue splitting alone.

B.5 Phase rigidity

The *phase rigidity* of a mode is the inverse of the Petermann factor:

$$r_j = \frac{|\langle l_j|r_j\rangle|^2}{\langle r_j|r_j\rangle \, \langle l_j|l_j\rangle} = K_j^{-1} \in [0, 1]. \tag{B.11}$$

A phase rigidity of $r_j = 1$ indicates a normal (Hermitian-like) mode; $r_j \to 0$ indicates an EP where the mode becomes maximally non-orthogonal.

B.6 Pseudospectra

The spectrum of a non-normal matrix can be misleading: eigenvalues may be highly sensitive to perturbations even when they are well separated. The *ϵ-pseudospectrum* quantifies this sensitivity.

B.6.1 Definition

The ϵ-pseudospectrum of A is the set of complex numbers z such that z is an eigenvalue of $A+E$ for some perturbation $\|E\| \le \epsilon$:

$$\sigma_\epsilon(A) = \{z \in \mathbb{C} : \|(z - A)^{-1}\| \ge \epsilon^{-1}\}. \tag{B.12}$$

For a normal matrix, the ϵ-pseudospectrum is the union of ϵ-disks around each eigenvalue. For a non-normal matrix, the pseudospectrum can extend far beyond these disks, indicating that eigenvalues can move large distances under small perturbations.

B.6.2 Physical relevance

Pseudospectra are relevant whenever the system is subject to perturbations—fabrication disorder, environmental fluctuations, or numerical discretization errors. For a photonic system operating near an EP, the pseudospectrum extends over a large region of the complex frequency plane, meaning that the eigenfrequencies are effectively unpredictable to within a band set by the perturbation strength. This is the spectral manifestation of the Petermann-factor divergence and the reason why EP-based devices require careful noise analysis (Chapter 18).

B.7 Matrix functions and the non-Hermitian exponential

The time evolution of a system governed by $i\,\partial_t|\psi\rangle = A\,|\psi\rangle$ is

$$|\psi(t)\rangle = e^{-iAt}\,|\psi(0)\rangle. \tag{B.13}$$

For a normal A, the evolution is a superposition of oscillating and decaying exponentials, and $\|e^{-iAt}\|$ is controlled by the largest imaginary part of the eigenvalues. For a non-normal A, the situation is richer: the Jordan chain at a defective eigenvalue contributes *polynomial* growth factors

$$e^{-iJ_m(\lambda_0)t} = e^{-i\lambda_0 t}\begin{pmatrix} 1 & -it & (-it)^2/2! & \cdots \\ 0 & 1 & -it & \cdots \\ \vdots & & \ddots & \ddots \\ 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{B.14}$$

so the amplitude can grow as t^{m-1} even when the eigenvalue has a negative imaginary part (i.e., the mode is formally decaying). This *transient growth* is a hallmark of non-normality and is observable as a power-law ring-up before the eventual exponential decay in photonic resonators near an EP.

Additionally, even in the absence of defective eigenvalues, a highly non-normal matrix can produce large transient growth—the so-called *pseudospectral* transient, where the norm $\|e^{-iAt}\|$ overshoots its long-time asymptote by a factor that can be estimated from the pseudospectrum. This effect is important for FDTD simulations of non-Hermitian photonic systems (Section 17.8), where a transient overshoot can trigger numerical instability before the physical decay sets in.

B.8 Examples

B.8.1 Jordan decomposition of a 2×2 defective matrix

The simplest defective matrix arises at an exceptional point of a 2×2 non-Hermitian Hamiltonian. Consider

$$A = \begin{pmatrix} \lambda_0 & 1 \\ 0 & \lambda_0 \end{pmatrix}, \tag{B.15}$$

which is already in Jordan normal form $J_2(\lambda_0)$. This matrix has a single eigenvalue λ_0 with algebraic multiplicity 2 but geometric multiplicity 1: there is only one linearly independent eigenvector, $\mathbf{v}_1 = (1, 0)^{\mathsf{T}}$. The second vector in the Jordan chain is the *generalized eigenvector* $\mathbf{v}_2$ satisfying $(A - \lambda_0 I)\mathbf{v}_2 = \mathbf{v}_1$, which gives $\mathbf{v}_2 = (0, 1)^{\mathsf{T}}$.

The time evolution is

$$e^{-iAt} = e^{-i\lambda_0 t} \begin{pmatrix} 1 & -it \\ 0 & 1 \end{pmatrix}. \tag{B.16}$$

Even if $\operatorname{Im}\lambda_0 < 0$ (the mode decays), the off-diagonal element $-it$ grows linearly before the exponential decay dominates. The maximum of $\|e^{-iAt}\|$ over time occurs at $t^* \sim 1/|\operatorname{Im}\lambda_0|$, with an envelope of order $1/(e|\operatorname{Im}\lambda_0|)$ before constant prefactors from the chosen norm are included. The transient amplification therefore diverges as the eigenvalue approaches the real axis, but the numerical prefactor is norm dependent.

B.8.2 Schur decomposition and its use

Every square matrix A admits a Schur decomposition $A = QTQ^{-1}$, where Q is unitary and T is upper triangular with the eigenvalues on the diagonal. Unlike the Jordan form, which requires exact knowledge of the eigenvector structure, the Schur form is numerically stable and can be computed by standard algorithms (QR iteration).

The Schur form reveals the *departure from normality* $\Delta_F = \|A\|_F^2 - \sum_i |\lambda_i|^2$, where $\|\cdot\|_F$ is the Frobenius norm. For a normal matrix, $\Delta_F = 0$ (the Schur form is diagonal); for a non-normal matrix, $\Delta_F > 0$ measures the energy stored in the off-diagonal elements of T, which is the source of transient amplification and pseudospectral broadening.

For photonic applications, the Schur decomposition is preferable to the Jordan form for three reasons. First, it is numerically robust: no special treatment is needed for nearly defective matrices. Second, the unitary factor Q preserves norms, making it easy to compute $\|e^{-iAt}\|$ from $\|e^{-iTt}\|$. Third, the off-diagonal elements of T give a direct and intuitive measure of the non-Hermitian coupling between modes—elements that are absent in normal systems.

B.8.3 Kreiss constant and transient bounds

The maximum transient growth $\sup_{t\geq 0} \|e^{-iAt}\|$ can be bounded using the Kreiss matrix theorem. The *Kreiss constant*

$$\mathcal{K}(A) = \sup_{\operatorname{Im} z > \alpha_{\max}} (\operatorname{Im} z - \alpha_{\max}) \, \|(z - A)^{-1}\|, \tag{B.17}$$

where $\alpha_{\max} = \max_i \operatorname{Im} \lambda_i$ is the maximum eigenvalue imaginary part, satisfies the bounds

$$\mathcal{K}(A) \leq \sup_{t\geq 0} \|e^{-iAt}\| \, e^{\alpha_{\max} t} \leq e \, n \, \mathcal{K}(A), \tag{B.18}$$

where n is the matrix dimension. The Kreiss constant is computable from the resolvent norm (essentially the pseudospectrum) and provides a practical upper bound on the transient growth without requiring explicit time integration.

For the 2×2 PT-symmetric Hamiltonian near the EP ($\gamma/\kappa = 0.99$), the Kreiss constant is $\mathcal{K} \approx 7.1$, consistent with the transient amplification factor of 7.09 computed directly from (9.18). For larger photonic systems, the Kreiss constant can be estimated from the pseudospectrum using standard numerical linear algebra packages.

B.8.4 Non-normal perturbation theory

For a non-normal matrix A with a defective eigenvalue λ_0 of algebraic multiplicity m, a generic perturbation ϵB splits the eigenvalue into m branches:

$$\lambda_j(\epsilon) = \lambda_0 + \epsilon^{1/m} \, c_j + O(\epsilon^{2/m}), \qquad j = 1, \ldots, m, \tag{B.19}$$

where c_j are the m-th roots of a coupling coefficient that depends on the perturbation matrix B and the Jordan chain vectors. This Puiseux series expansion is the mathematical basis for the enhanced spectral response at exceptional points: a perturbation of order ϵ produces an eigenvalue splitting of order $\epsilon^{1/m}$, which for small ϵ is much larger than the linear splitting $\sim \epsilon$ of a non-defective eigenvalue.

However, the Puiseux expansion is *not* isotropic in the perturbation: the coefficient c_j depends on the *direction* of the perturbation in parameter space, not just its magnitude. For certain perturbation directions (those orthogonal to the Jordan chain coupling), the splitting can be much smaller than $\epsilon^{1/m}$ or even vanish. This anisotropy is important for EP-based sensing: the enhanced response occurs only for perturbations that couple to the Jordan chain, not for arbitrary environmental noise.

Chapter C

Transfer Matrices and Scattering

This appendix develops the one-dimensional transfer-matrix formalism that appears in Chapters 9, 10, and 15. The results are exact for multilayer dielectric stacks with complex permittivity and provide the computational backbone for many of the non-Hermitian scattering calculations in this book.

C.1 Single-layer transfer matrix

Consider a homogeneous layer of thickness d, refractive index $n = n_r + i n_i$, at normal incidence. The electric field at position x is a superposition of forward and backward waves:

$$E(x) = A\, e^{i n k_0 x} + B\, e^{-i n k_0 x}, \tag{C.1}$$

where $k_0 = \omega/c$. The transfer matrix M relates the field amplitudes at the left face ($x = 0$) to those at the right face ($x = d$):

$$\begin{pmatrix} A \\ B \end{pmatrix}_{x=d} = M \begin{pmatrix} A \\ B \end{pmatrix}_{x=0}, \qquad M = \begin{pmatrix} e^{i n k_0 d} & 0 \\ 0 & e^{-i n k_0 d} \end{pmatrix}. \tag{C.2}$$

At an interface between media with indices n_1 and n_2, the interface matrix is

$$I_{12} = \frac{1}{2 n_2} \begin{pmatrix} n_2 + n_1 & n_2 - n_1 \\ n_2 - n_1 & n_2 + n_1 \end{pmatrix}. \tag{C.3}$$

The total transfer matrix for a multilayer stack is the ordered product of propagation and interface matrices:

$$\mathbf{T} = I_{N,\mathrm{out}} \cdot M_N \cdots I_{23} \cdot M_2 \cdot I_{12} \cdot M_1. \tag{C.4}$$

C.2 Scattering matrix from the transfer matrix

The 2×2 scattering matrix $\mathbf{S}$ relates the outgoing amplitudes to the incoming amplitudes:

$$\begin{pmatrix} B_L \\ A_R \end{pmatrix} = \mathbf{S} \begin{pmatrix} A_L \\ B_R \end{pmatrix} = \begin{pmatrix} r_L & t' \\ t & r_R \end{pmatrix} \begin{pmatrix} A_L \\ B_R \end{pmatrix}, \tag{C.5}$$

where r_L, r_R are the reflection coefficients from the left and right, and t, t' are the transmission coefficients. The conversion from the transfer matrix $\mathbf{T} = \left(\begin{smallmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{smallmatrix}\right)$ is

$$t = \frac{1}{T_{22}}, \qquad t' = \frac{\det \mathbf{T}}{T_{22}}, \qquad r_L = -\frac{T_{21}}{T_{22}}, \qquad r_R = \frac{T_{12}}{T_{22}}. \tag{C.6}$$

For a reciprocal medium (symmetric ε_{r} tensor), $t = t'$ and $\det \mathbf{T} = 1$. For a non-reciprocal medium, $t \neq t'$ in general.

C.3 $\mathcal{PT}$-symmetric transfer matrices

A one-dimensional structure with $\mathcal{PT}$ symmetry satisfies $n(x) = n^*(-x)$. For a multilayer consisting of a loss layer (index $n_0 + i\kappa$, thickness d) followed by a gain layer (index $n_0 - i\kappa$, thickness d), the unit-cell transfer matrix is

$$\mathbf{T}_{\mathrm{cell}} = M_{\mathrm{gain}} \cdot I \cdot M_{\mathrm{loss}}. \tag{C.7}$$

The $\mathcal{PT}$ condition imposes the constraint

$$\sigma_x \, \mathbf{T}_{\mathrm{cell}}^{*} \, \sigma_x = \mathbf{T}_{\mathrm{cell}}, \qquad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{C.8}$$

which ensures that $T_{11} = T_{22}^{*}$ and $T_{12} = T_{21}^{*}$. For the scattering matrix, this gives

$$|t|^2 - |r_L|\,|r_R| = 1 \quad \text{(exact } \mathcal{PT} \text{ relation)}, \tag{C.9}$$

which replaces the Hermitian unitarity condition $|t|^2 + |r|^2 = 1$.

C.3.1 $\mathcal{PT}$ phase transition from transfer-matrix eigenvalues

The eigenvalues of $\mathbf{T}_{\mathrm{cell}}$ are $\lambda_{\pm} = \frac{1}{2}\mathrm{tr}(\mathbf{T}_{\mathrm{cell}}) \pm \sqrt{[\frac{1}{2}\mathrm{tr}(\mathbf{T}_{\mathrm{cell}})]^2 - 1}$. Three regimes arise:

1. **$\mathcal{PT}$-symmetric phase** ($|\frac{1}{2}\mathrm{tr}\,\mathbf{T}| \leq 1$): $\lambda_{\pm}$ are unimodular; the stack is a Bragg grating with real band structure.
2. **$\mathcal{PT}$-broken phase** ($|\frac{1}{2}\mathrm{tr}\,\mathbf{T}| > 1$): $|\lambda_+| > 1 > |\lambda_-|$; one eigenvalue grows exponentially with the number of cells (laser-like amplification), while the other decays (CPA-like absorption).
3. **Exceptional point** ($|\frac{1}{2}\mathrm{tr}\,\mathbf{T}| = 1$): $\lambda_+ = \lambda_-$; the transfer matrix has a Jordan block of size 2. This is the $\mathcal{PT}$-symmetry breaking threshold.

C.4 Poles and zeros from the transfer matrix

The poles and zeros of the scattering matrix can be located directly from the transfer matrix:

- A **pole** of $\mathbf{S}(\omega)$ (a lasing mode) occurs when $T_{22}(\omega) = 0$, since $t = 1/T_{22}$.
- A **zero** of $\det \mathbf{S}(\omega)$ (a CPA mode) occurs when $T_{11}(\omega) = 0$, since with the convention above $\det \mathbf{S} = T_{11}/T_{22}$. For reciprocal media $\det \mathbf{T} = 1$, but this does not prevent $\det \mathbf{S}$

from vanishing because the scattering determinant is not the same object as the transfer determinant.

The CPA-laser point is the frequency at which both $T_{22} = 0$ and $T_{11} = 0$ simultaneously. For a $\mathcal{PT}$-symmetric stack at real frequency, the constraint $T_{11} = T_{22}^*$ makes the pole and zero conditions coincide; in practice this simultaneous condition is isolated in the (ω, κ) parameter space.

C.5 Bloch waves in periodic stacks

For an infinite periodic stack with unit-cell transfer matrix $\mathbf{T}_{\text{cell}}$, the Bloch condition is

$$\mathbf{T}_{\text{cell}} \begin{pmatrix} A \\ B \end{pmatrix} = e^{ika} \begin{pmatrix} A \\ B \end{pmatrix}, \tag{C.10}$$

so the Bloch phase is $\beta = e^{ika}$ and the dispersion relation is

$$\cos(ka) = \tfrac{1}{2} \operatorname{tr}(\mathbf{T}_{\text{cell}}). \tag{C.11}$$

For a Hermitian stack, $\operatorname{tr}(\mathbf{T}_{\text{cell}})$ is real, and band gaps correspond to $|\cos(ka)| > 1$. For a non-Hermitian stack, $\operatorname{tr}(\mathbf{T}_{\text{cell}})$ is complex, and the dispersion defines a curve in the complex ω plane—the complex band structure of Chapter 12.

C.5.1 GBZ from the transfer matrix

For a finite open stack, the correct Brillouin zone is the GBZ (Section 14.4). The GBZ condition $|\beta_n| = |\beta_{n+1}|$ translates to the condition that the two eigenvalues of $\mathbf{T}_{\text{cell}}$ have equal modulus: $|\lambda_+| = |\lambda_-|$. Since $\det \mathbf{T}_{\text{cell}} = \lambda_+\lambda_- = 1$ for reciprocal media, this gives $|\lambda_\pm| = 1$—the GBZ coincides with the ordinary BZ, and there is no skin effect. For non-reciprocal media, $|\det \mathbf{T}_{\text{cell}}| \neq 1$, and the GBZ deforms away from the unit circle, producing the skin effect.

This transfer-matrix derivation of the GBZ is the most direct route for one-dimensional photonic structures and connects the abstract non-Bloch theory of Chapter 14 to concrete multilayer calculations.

C.6 Numerical implementation

C.6.1 Avoiding numerical overflow

For thick stacks with gain ($n_i < 0$), the propagation matrix elements $e^{\pm ink_0 d}$ can overflow or underflow. The standard remedy is to factor the transfer matrix after each layer, extracting the largest eigenvalue modulus and accumulating the logarithm of the total transmission. This "scattering-matrix propagation algorithm" is numerically stable and avoids the exponential overflow that plagues naive transfer-matrix products.

C.6.2 Complex-frequency evaluation

For QNM and pole-zero calculations, the transfer matrix must be evaluated at complex frequencies $\omega = \omega_r + i\omega_i$. The refractive index $n(\omega)$ is then evaluated by analytic continuation of the dispersion model (Lorentz, Drude, etc.), and the propagation phases nk_0d become fully complex. No additional numerical modification is needed beyond ensuring that the branch cuts of the square root in $n(\omega) = \sqrt{\varepsilon_{\mathrm{r}}(\omega)}$ are consistently chosen.

C.7 From transfer matrix to scattering matrix

The transfer matrix and scattering matrix encode the same physics but organise it differently. Converting between them is a routine but essential operation in photonic multilayer calculations.

C.7.1 Conversion formulas

For a single interface or slab described by the 2×2 transfer matrix

$$\mathbf{T} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}, \tag{C.12}$$

relating the field amplitudes on the left to those on the right, $\left(\begin{smallmatrix} A_R \\ B_R \end{smallmatrix}\right) = \mathbf{T}\left(\begin{smallmatrix} A_L \\ B_L \end{smallmatrix}\right)$, the scattering matrix $\left(\begin{smallmatrix} B_L \\ A_R \end{smallmatrix}\right) = \mathbf{S}\left(\begin{smallmatrix} A_L \\ B_R \end{smallmatrix}\right)$ is obtained by rearranging:

$$\mathbf{S} = \frac{1}{T_{22}} \begin{pmatrix} -T_{21} & 1 \\ T_{11}T_{22} - T_{12}T_{21} & T_{12} \end{pmatrix} = \frac{1}{T_{22}} \begin{pmatrix} -T_{21} & 1 \\ \det \mathbf{T} & T_{12} \end{pmatrix}. \tag{C.13}$$

The elements have direct physical meaning: $t_L = 1/T_{22}$ is the left-to-right transmission amplitude, $r_L = -T_{21}/T_{22}$ is the left reflection, and $r_R = T_{12}/T_{22}$ is the right reflection. For a reciprocal medium, $\det \mathbf{T} = 1$ and the left-to-right and right-to-left transmission amplitudes are equal: $t_L = t_R = 1/T_{22}$.

C.7.2 Example: Fabry–Pérot resonator

Consider a dielectric slab of refractive index n and thickness d in air. The total transfer matrix is the product of three matrices: the entrance interface, the propagation through the slab, and the exit interface. Using the Fresnel interface matrices and the propagation matrix from Section C.1:

$$\mathbf{T}_{\mathrm{FP}} = \mathbf{T}_{\mathrm{exit}}\, \mathbf{T}_{\mathrm{prop}}\, \mathbf{T}_{\mathrm{entrance}}. \tag{C.14}$$

The entrance interface matrix (from air to the slab, normal incidence) is

$$\mathbf{T}_{\mathrm{entrance}} = \frac{1}{2n} \begin{pmatrix} n+1 & n-1 \\ n-1 & n+1 \end{pmatrix}, \tag{C.15}$$

the propagation matrix through the slab is

$$\mathbf{T}_{\text{prop}} = \begin{pmatrix} e^{ink_0 d} & 0 \\ 0 & e^{-ink_0 d} \end{pmatrix}, \tag{C.16}$$

where $k_0 = \omega/c$, and the exit interface matrix (from the slab to air) is

$$\mathbf{T}_{\text{exit}} = \frac{1}{2}\begin{pmatrix} 1+1/n & 1-1/n \\ 1-1/n & 1+1/n \end{pmatrix}. \tag{C.17}$$

Carrying out the matrix multiplication and extracting T_{22}, the transmission amplitude is

$$t = \frac{1}{T_{22}} = \frac{4n\, e^{-ink_0 d}}{(1+n)^2 - (1-n)^2\, e^{2ink_0 d}}, \tag{C.18}$$

and the transmission intensity is the Airy function

$$|t|^2 = \frac{1}{1 + F\sin^2(nk_0 d)}, \qquad F = \frac{4R}{(1-R)^2}, \tag{C.19}$$

where $R = |(n-1)/(n+1)|^2$ is the single-interface power reflectance and F is the coefficient of finesse. This recovers the result derived from physical arguments in Chapter 4 and provides a systematic framework for extending the calculation to multilayer stacks.

Resonance condition and QNM connection. The resonances of the Fabry–Pérot cavity correspond to the poles of the transmission amplitude, i.e. the zeros of $T_{22}(\omega)$. From (C.18), the pole condition is $(1+n)^2 = (1-n)^2\, e^{2ink_0 d}$, which gives

$$\tilde{\omega}_m = \frac{c}{nd}\left(m\pi - i\ln\frac{1+n}{|1-n|}\right), \qquad m = 0, \pm 1, \pm 2, \ldots \tag{C.20}$$

The real part gives the resonance frequencies (equally spaced by the free spectral range $\pi c/(nd)$); the imaginary part gives the decay rate, which equals the QNM linewidth $\gamma = c\,\ln|(1+n)/(1-n)|/(nd)$ derived in Chapter 6. The transfer-matrix approach thus provides a systematic route from multilayer geometry to QNM frequencies, applicable to arbitrary one-dimensional photonic structures including those with gain, loss, and dispersion.

Chapter D

Numerical Recipes

This appendix provides implementation notes for the key computational tasks discussed in the book. Each recipe records a reproducible algorithm for producing the corresponding figure or numerical check. The steps are written in a language-independent form so that they can be reimplemented with any reliable numerical package.

D.1 Lorentz dispersion model

Aim. Compute the complex permittivity $\varepsilon_{\mathrm{r}}(\omega)$ for a single-oscillator Lorentz model with passive damping or linear gain (Figure 5.2).

Equations.

$$\varepsilon_{\mathrm{r}}(\omega) = \varepsilon_{\infty} + \frac{F}{\omega_0^2 - \omega^2 - i\gamma\omega}, \tag{D.1}$$

with the $e^{-i\omega t}$ convention. Positive γ gives absorption ($\operatorname{Im}\varepsilon_{\mathrm{r}} > 0$ near resonance); negative γ gives gain.

Procedure.

1. Choose parameters: ε_{∞}, F, ω_0, $\gamma_{\mathrm{abs}} > 0$, $\gamma_{\mathrm{gain}} < 0$.
2. Sample ω on a uniform grid in $[0.15\,\omega_0,\ 2.25\,\omega_0]$.
3. Evaluate (D.1) for both γ values.
4. Plot $\operatorname{Re}\varepsilon_{\mathrm{r}}(\omega)$ and $\operatorname{Im}\varepsilon_{\mathrm{r}}(\omega)$ on separate panels.

Result. This calculation gives the Lorentz-dispersion figure `fig_f5_lorentz_dispersion.pdf`.

D.2 Two-mode exceptional point

Aim. Compute eigenvalue surfaces and eigenvector coalescence for the canonical 2×2 EP Hamiltonian (Figures 8.1 and 8.2).

Equations.

$$H = \begin{pmatrix} i\gamma & \kappa \\ \kappa & -i\gamma \end{pmatrix}, \qquad \lambda_\pm = \pm\sqrt{\kappa^2 - \gamma^2}. \tag{D.2}$$

Procedure.

1. Create a 2D grid in (κ, γ).
2. Evaluate $\lambda_\pm = \pm\sqrt{\kappa^2 - \gamma^2}$ on the grid (complex square root).
3. Plot $\mathrm{Re}\,\lambda_\pm$ and $\mathrm{Im}\,\lambda_\pm$ as 3D surfaces.
4. For the eigenvector figure: fix $\kappa = 1$, sweep γ, diagonalize H numerically, compute the overlap $|\langle v_+|v_-\rangle|$ and the condition number of the eigenvector matrix.
5. For the encircling figure: parameterize a loop $\delta = r\,e^{i\theta}$ around the branch point, plot $\pm\sqrt{\delta}$ in the complex λ plane.

Result. This calculation gives the EP eigenvalue and eigenvector figures `fig_f8_ep_eigenvalues.pdf` and `fig_f8_ep_vectors.pdf`.

D.3 Point-gap winding number

Aim. Compute the spectral winding number of a complex band loop around a reference energy (see the complex-band-loop discussion associated with Figure 13.1).

Equations. One-band non-Hermitian dispersion:

$$\omega(k) = \omega_0 + t_R\,e^{ik} + t_L\,e^{-ik} + i\gamma_0, \tag{D.3}$$

with $t_R \neq t_L$ producing a loop that can encircle the origin.

Algorithm.

1. Sample k uniformly on $[-\pi, \pi]$ with $N \gg 1$ points.
2. Evaluate $\omega(k)$ and form $z(k) = \omega(k) - E_{\mathrm{ref}}$.
3. Compute the unwrapped phase: $\phi(k) = \mathrm{unwrap}[\arg z(k)]$.
4. The winding number is $W = [\phi(\pi) - \phi(-\pi)]/(2\pi)$.
5. Compare inside and outside reference points: an inside point gives $|W| \geq 1$; an outside point gives $W = 0$.

Result. The calculation gives a point-gap winding diagnostic and the complex-band-loop figure `fig_f13_complex_band_loop.pdf`. If the standalone point-gap-winding schematic is omitted from the main text, the same calculation can be reported directly in the discussion of the remaining loop figure.

D.4 Non-Hermitian skin effect (Hatano–Nelson model)

Aim. Demonstrate the PBC/OBC spectral discrepancy, eigenmode localization, and the generalized Brillouin zone for a minimal skin-effect model (Figures 14.1, 14.2, and 14.4).

Equations. Hatano–Nelson chain: $H_{n,n+1} = t_R$, $H_{n+1,n} = t_L$, with $t_R > t_L$.

Algorithm.

1. **PBC spectrum:** $E_{\mathrm{PBC}}(k) = t_R\, e^{ik} + t_L\, e^{-ik}$; plot as a complex loop.
2. **OBC spectrum:** construct the $N \times N$ tridiagonal matrix, diagonalize, plot eigenvalues in the complex plane.
3. **Mode localization:** normalize $|\psi_j(n)|^2$ for each eigenstate; compute center of mass $\bar{n}_j = \sum_n n\,|\psi_j(n)|^2$.
4. **GBZ:** the GBZ radius is $r = \sqrt{t_L/t_R}$; plot the circle $|\beta| = r$ and compare $E_{\mathrm{GBZ}}(\theta) = t_R\, r\, e^{i\theta} + t_L\, e^{-i\theta}/r$ with the OBC spectrum.

Result. The calculation gives the PBC/OBC spectra, skin-mode localization, and GBZ figures: `fig_f14_pbc_obc_spectra.pdf`, `fig_f14_skin_modes.pdf`, and `fig_f14_gbz.pdf`.

D.5 BIC topological charge and Q scaling

Aim. Visualize the far-field polarization winding around a BIC and the quality-factor scaling under charge merging (Figures 16.3 and 16.4).

Algorithm.

1. **Polarization field:** on a 2D grid in (k_x, k_y), set the polarization angle $\theta = q\,\mathrm{atan2}(k_y, k_x)$ with $q = +1$ or $q = -1$. Plot as a quiver field; the BIC is the singularity at the origin.
2. **Q scaling:** for an isolated charge-one BIC, $\gamma \propto k^2$ so $Q \propto k^{-2}$. For a merged super-BIC, $\gamma \propto k^6$ so $Q \propto k^{-6}$. Plot both on a log–log scale.
3. **BZF scaling:** $Q \sim 1/(\alpha^2 k^2)$ for perturbation strength α; plot several α values.

Result. The calculation gives the BIC polarization-charge and quality-factor scaling figures, `fig_f16_bic_polarization_charge.pdf` and `fig_f16_q_scaling.pdf`.

D.6 PML validation and mode-ratio filtering

Aim. Illustrate the PML validation protocol: physical QNM poles remain stable under PML parameter variation, while PML continuum modes move (Figures 17.3 and 17.4).

Algorithm.

1. Generate synthetic eigenfrequencies: one physical QNM near a target pole, plus several PML modes whose positions depend strongly on the PML absorption strength $\sigma_{\max}$.
2. Sweep $\sigma_{\max}$; for each value, compute the mode ratio $\mathrm{MR} = \int_{\Omega_{\mathrm{phys}}} |\mathbf{E}|^2 \, \mathrm{d}V \big/ \int_{\Omega_{\mathrm{total}}} |\mathbf{E}|^2 \, \mathrm{d}V$.
3. Physical QNMs cluster near $\mathrm{MR} \approx 1$; PML modes near $\mathrm{MR} \ll 1$.
4. Plot the eigenvalue trajectories and the mode-ratio separation.

Result. The calculation gives the PML-convergence and PML-mode-ratio figures, `fig_f17_pml_convergence.pdf` and `fig_f17_pml_mode_ratio.pdf`.

D.7 Scattering pole-zero map

Aim. Visualize the S-matrix pole and zero trajectories under gain/loss tuning, the CPA-laser threshold, and the complex scattering response landscape discussed in Chapter 10.

Equations. Scalar scattering response:

$$s(\omega; g) = \prod_j \frac{\omega - z_j(g)}{\omega - p_j(g)}, \tag{D.4}$$

where poles p_j rise from $\mathrm{Im}\,\omega < 0$ and zeros z_j descend from $\mathrm{Im}\,\omega > 0$ as the gain parameter g increases.

Algorithm.

1. Define two pole–zero pairs near a central frequency, plus one detuned background pair.
2. Sweep g from 0 to above threshold ($g = 1$).
3. Plot pole and zero trajectories; mark the real-axis crossing (lasing/CPA/CPA-laser).
4. Evaluate $|s(\omega)|^2$ on the real axis for several g values; plot spectra.
5. Evaluate $\log_{10} |s(\omega)|$ on a 2D complex-ω grid near threshold; display as a heatmap.

Result. The calculation gives pole–zero trajectories and scattering-response data. If the schematic pole–zero figures are not included in the main text, the recipe remains useful as a numerical check on the meromorphic structure of the scattering amplitude.

D.8 EP sensing response versus noise

Aim. Compare the frequency splitting, linewidth, and signal-to-noise ratio of an EP sensor with a conventional diabolic-point sensor, following the scaling discussion in Chapter 18.

Equations. Diabolic-point splitting: $\Delta\omega_{\mathrm{DP}} \propto \epsilon$. EP splitting: $\Delta\omega_{\mathrm{EP}} \propto \sqrt{\epsilon}$. Noise proxy: $\sigma_{\mathrm{EP}} \propto \sqrt{1 + c/\epsilon}$ (diverges as $\epsilon \to 0$).

Algorithm.

1. Sample ϵ on a logarithmic grid.
2. Compute signal, noise, responsivity, and SNR for both sensor types.
3. Plot three panels: splitting, responsivity+noise, and SNR comparison.

Result. The calculation gives the EP sensing response-versus-noise diagnostic; depending on the presentation, it may be shown as a figure or summarized as the scaling argument in Chapter 18.

D.9 Practical convergence protocol

This section collects a systematic convergence protocol for non-Hermitian photonic eigenvalue calculations. Each step can be implemented as a post-processing check on the output of any FEM or FDTD solver.

D.9.1 Mesh refinement

The first convergence check is mesh refinement. Run the eigenvalue calculation at two or more mesh densities (e.g. h, $h/2$, $h/4$) and monitor the change in the complex eigenfrequency $\tilde{\omega}_n$. For a well-resolved mode, the eigenfrequency should converge as

$$|\tilde{\omega}_n(h) - \tilde{\omega}_n(h/2)| \sim C\, h^{2p}, \tag{D.5}$$

where p is the polynomial order of the finite-element basis. If the convergence rate is slower than expected, the mode may be under-resolved—common for tightly confined plasmonic modes or modes with sharp field gradients at material interfaces.

Recommended criterion. Require $|\Delta\tilde{\omega}|/|\tilde{\omega}| < 10^{-4}$ between the two finest meshes. For quality factors exceeding 10^4, tighter convergence ($< 10^{-6}$) may be needed because $Q = \mathrm{Re}\,\tilde{\omega}/(2|\,\mathrm{Im}\,\tilde{\omega}|)$ amplifies small errors in $\mathrm{Im}\,\tilde{\omega}$.

D.9.2 PML thickness and absorption profile

The PML is an artificial absorbing boundary that must be thick enough to suppress reflections but thin enough to keep the computational domain manageable. The PML convergence test varies the PML thickness d_{PML} (or equivalently the PML absorption strength $\sigma_{\max}$) while keeping the physical domain fixed.

PML thickness selection rule. For a mode with wavelength λ and exterior decay constant κ_{out}, a practical starting point is

$$d_{\mathrm{PML}} \gtrsim \max\left(\lambda, \frac{4}{\kappa_{\mathrm{out}}}\right), \tag{D.6}$$

followed by an explicit sweep of thickness and absorption strength. This estimate is only a starting scale: radiative QNMs, grazing waves, and quasi-BICs often require thicker PMLs and finer PML discretization than any one-line rule can predict.

PML eigenvalue filtering. Not all eigenvalues returned by the solver correspond to physical QNMs. The PML itself introduces a branch of spurious "PML modes" whose eigenvalues depend on the PML parameters. These are identified by their sensitivity to PML changes: physical QNM eigenvalues remain stable as d_{PML} varies, while PML modes drift systematically. A mode-ratio filter

$$\mathrm{MR}_n = \frac{\int_{\mathrm{physical}} |\mathbf{E}_n|^2 \, \mathrm{d}V}{\int_{\mathrm{total}} |\mathbf{E}_n|^2 \, \mathrm{d}V} \tag{D.7}$$

provides a second diagnostic: physical QNMs have $\mathrm{MR} \approx 1$, while PML modes have $\mathrm{MR} \ll 1$. In practice, a threshold $\mathrm{MR} > 0.5$ is sufficient to filter out most PML artifacts.

D.9.3 Dispersive material branch verification

For dispersive materials with multiple resonances (e.g. Lorentz oscillators), the analytic continuation $\varepsilon_{\mathrm{r}}(\omega)$ to complex ω must use the same causal branch as the real-frequency material model. Eigenvalues that are artifacts of a wrong analytic continuation are unphysical. Useful checks are:

1. Evaluate $\varepsilon_{\mathrm{r}}(\tilde{\omega}_n)$ at the computed complex eigenfrequency.
2. Track the mode continuously from a real-frequency or weakly open reference calculation while keeping the material poles and square-root branches fixed.
3. Check passivity on the real axis for the underlying material model; the sign of $\mathrm{Im}\, \varepsilon_{\mathrm{r}}(\tilde{\omega}_n)$ at a complex eigenfrequency is not, by itself, a reliable physicality test.
4. Discard eigenvalues that switch sheets, converge to material poles rather than electromagnetic resonances, or fail the residual and PML-variation checks.

D.9.4 Biorthogonal normalisation check

After extracting the QNM eigenvectors, verify the biorthogonal normalisation condition

$$\langle \tilde{\mathbf{E}}_m, \tilde{\mathbf{E}}_n \rangle_{\mathrm{uc}} = \delta_{mn}, \tag{D.8}$$

where $\langle \cdot, \cdot \rangle_{\mathrm{uc}}$ is the unconjugated inner product (Chapter 6). Deviations from δ_{mn} indicate either insufficient mesh resolution, incorrect PML regularisation, or missing modes. A practical tolerance is $|\langle \tilde{\mathbf{E}}_m, \tilde{\mathbf{E}}_n \rangle_{\mathrm{uc}} - \delta_{mn}| < 10^{-3}$.

D.9.5 Reproducibility checklist

For publication-quality results, record the following for every eigenvalue calculation:

Parameter	Purpose
Mesh element size h	Resolution verification
FE polynomial order p	Convergence rate check
PML thickness d_{PML}	Artificial-boundary control
PML absorption $\sigma_{\max}$	Reflection suppression
Dispersion model	Material branch verification
Contour parameters ($\mathcal{C}$, N_q)	Eigensolver control
Mode-ratio threshold	PML filtering
Biorthogonal norm tolerance	Normalisation quality

A compact reporting table of this kind is recommended for numerical QNM results that support physical claims. Its consistent use would significantly improve the reproducibility of non-Hermitian photonic calculations.

D.9.6 Minimum reporting standard for Maxwell eigenvalue computations

A non-Hermitian Maxwell calculation is not fully reproducible unless the reported eigenfrequency is accompanied by an estimate of its numerical uncertainty and by a check that the computed left and right modes solve the same physical problem. The following minimal standard is recommended for any result that is used to support a physical claim.

First, report the residual norm of the right eigenproblem. Consider a discretised generalized Maxwell problem evaluated at the nonlinear eigenfrequency $\tilde{\omega}_n$:

$$\mathbf{A}(\tilde{\omega}_n)\mathbf{x}_n = \tilde{\omega}_n^2\,\mathbf{B}(\tilde{\omega}_n)\mathbf{x}_n, \tag{D.9}$$

the dimensionless residual may be written as

$$\rho_n^{\mathrm{R}} = \frac{\|\mathbf{A}(\tilde{\omega}_n)\mathbf{x}_n - \tilde{\omega}_n^2\mathbf{B}(\tilde{\omega}_n)\mathbf{x}_n\|_2}{\|\mathbf{A}(\tilde{\omega}_n)\mathbf{x}_n\|_2 + |\tilde{\omega}_n|^2\|\mathbf{B}(\tilde{\omega}_n)\mathbf{x}_n\|_2}. \tag{D.10}$$

For a well-converged mode, ρ_n^{R} should be much smaller than the quoted relative uncertainty in $\tilde{\omega}_n$. Reporting only the eigenfrequency without the residual is insufficient, because spurious PML modes and poorly resolved physical modes can have plausible complex frequencies.

Second, when using biorthogonal perturbation theory, report the residual of the left eigenproblem as well. If the adjoint mode is represented by $\mathbf{y}_n$, the corresponding residual is

$$\rho_n^{\mathrm{L}} = \frac{\|\mathbf{A}^{\mathrm{T}}(\tilde{\omega}_n)\mathbf{y}_n - \tilde{\omega}_n^2\mathbf{B}^{\mathrm{T}}(\tilde{\omega}_n)\mathbf{y}_n\|_2}{\|\mathbf{A}^{\mathrm{T}}(\tilde{\omega}_n)\mathbf{y}_n\|_2 + |\tilde{\omega}_n|^2\|\mathbf{B}^{\mathrm{T}}(\tilde{\omega}_n)\mathbf{y}_n\|_2}. \tag{D.11}$$

The transpose, rather than the Hermitian conjugate, appears because the reciprocal Maxwell problem uses the unconjugated bilinear pairing. If a paper reports Petermann factors, adjoint sensitivities, or mode-volume corrections but gives no left-mode residual, the numerical evidence is incomplete.

Third, quote an uncertainty band obtained from at least two independent refinement directions. Mesh refinement tests the spatial discretisation; PML variation tests the artificial boundary;

material-branch checks test the analytic continuation. A conservative uncertainty estimate is

$$\Delta\tilde{\omega}_n = \max\left\{|\tilde{\omega}_n(h) - \tilde{\omega}_n(h/2)|, |\tilde{\omega}_n(d_{\mathrm{PML}}) - \tilde{\omega}_n(2d_{\mathrm{PML}})|, |\tilde{\omega}_n(N_q) - \tilde{\omega}_n(2N_q)|\right\}, \tag{D.12}$$

where N_q is the number of quadrature points in a contour eigensolver when such a method is used. The reported real frequency, linewidth, and quality factor should be rounded consistently with this uncertainty. For example, quoting $Q = 1.23748 \times 10^5$ is misleading if the PML sweep changes Q by three percent.

Finally, state which modes were discarded and why. Non-Hermitian Maxwell spectra contain physical resonances, PML continuum modes, material-pole artifacts, and numerical outliers. A transparent report should say how each class was separated. The combination of residual norms, mode-ratio filtering, branch verification, and refinement sweeps is usually sufficient. Without these checks, an apparently impressive exceptional point or ultra-high-Q resonance may simply be a numerical artifact of the discretised open boundary.

Chapter E

Notation and Symbols

This appendix provides a consolidated reference for all symbols, conventions, and abbreviations used in the monograph. Symbols are grouped by category; within each group they appear roughly in the order of first use.

E.1 General conventions

Convention	Choice in this book
Time-harmonic factor	$e^{-i\omega t}$
Fourier sign	$\tilde{f}(\omega) = \int_{-\infty}^{\infty} f(t)\, e^{+i\omega t}\, \mathrm{d}t$
SI units throughout	
Vacuum permittivity	ε_0; relative: ε_r
Vacuum permeability	μ_0; relative: μ_r
Speed of light	$c = 1/\sqrt{\varepsilon_0 \mu_0}$
Passive medium	$\operatorname{Im} \varepsilon_\mathrm{r}(\omega) > 0$
Gain medium	$\operatorname{Im} \varepsilon_\mathrm{r}(\omega) < 0$
QNM decay	$\tilde{\omega} = \omega_r - i\gamma$, $\gamma > 0$
Quality factor	$Q = \omega_r / (2\gamma)$

E.2 Electromagnetic fields and sources

Symbol	Name	SI Unit
$\mathbf{E}(\mathbf{r}, t)$	Electric field	$\mathrm{V\,m^{-1}}$
$\mathbf{H}(\mathbf{r}, t)$	Magnetic field	$\mathrm{A\,m^{-1}}$
$\mathbf{D}(\mathbf{r}, t)$	Electric displacement	$\mathrm{C\,m^{-2}}$
$\mathbf{B}(\mathbf{r}, t)$	Magnetic flux density	T
$\mathbf{J}(\mathbf{r}, t)$	Current density	$\mathrm{A\,m^{-2}}$
$\rho(\mathbf{r}, t)$	Charge density	$\mathrm{C\,m^{-3}}$
$\mathbf{S} = \mathbf{E} \times \mathbf{H}$	Poynting vector	$\mathrm{W\,m^{-2}}$
u	Electromagnetic energy density	$\mathrm{J\,m^{-3}}$
σ	Conductivity	$\mathrm{S\,m^{-1}}$

E.3 Material and structural parameters

Symbol	Meaning	Introduced
$\varepsilon_\mathrm{r}(\mathbf{r}, \omega)$	Complex relative permittivity	Ch. 1
$\mu_\mathrm{r}(\mathbf{r}, \omega)$	Complex relative permeability	Ch. 1
$n = \sqrt{\varepsilon_\mathrm{r}\mu_\mathrm{r}}$	Complex refractive index	Ch. 5
ω_0	Oscillator resonance frequency	Ch. 5
ω_p	Plasma frequency	Ch. 5
γ	Damping rate / gain–loss contrast	Chs. 5, 8
F	Oscillator strength	Ch. 5
a, $\mathbf{a}_j$	Lattice constant, lattice vectors	Ch. 3
$\mathbf{G}$	Reciprocal lattice vector	Ch. 3
d	Layer thickness	App. C

E.4 Operators and eigenproblems

Symbol	Meaning	Introduced
$\hat{\Theta}$	Maxwell curl–curl operator	Ch. 1
$\tilde{\omega}_n = \omega_n - i\gamma_n$	Complex QNM eigenfrequency	Ch. 6
$\tilde{\mathbf{E}}_n$, $\tilde{\mathbf{H}}_n$	QNM field profiles	Ch. 6
$\|r_j\rangle$, $\langle l_j\|$	Right, left eigenvectors	App. B
N_j	Biorthogonal normalization	Ch. 6, App. B
K_j	Petermann factor	Ch. 6, App. B
r_j	Phase rigidity	App. B
$J_m(\lambda)$	Jordan block of size m	App. B
$\sigma_\epsilon(A)$	ϵ-pseudospectrum	App. B

E.5 Scattering and transfer matrices

Symbol	Meaning	Introduced
$\mathbf{S}(\omega)$	Scattering matrix	Ch. 4
$\mathbf{T}$	Transfer matrix	Ch. 9, App. C
t, t'	Transmission coefficients	App. C
r_L, r_R	Reflection coefficients (left, right)	App. C
$\det \mathbf{S}$	Determinant (CPA condition)	Ch. 10

E.6 Topology and band theory

Symbol	Meaning	Introduced
$\mathbf{k}$, k	Bloch wave vector	Ch. 3
$\beta = e^{ika}$	Bloch phase factor	Ch. 14
$\mathcal{A}_n(\mathbf{k})$	Berry connection	Ch. 11
$\mathcal{F}_{xy}$	Berry curvature	Ch. 11
C	Chern number	Ch. 11
W	Spectral winding number	Ch. 12
q	BIC topological charge	Ch. 16
E_{ref}	Point-gap reference energy	Ch. 13
GBZ	Generalized Brillouin zone	Ch. 14
P^{bi}	Biorthogonal polarization	Ch. 15

E.7 Abbreviations

Abbreviation	Meaning
BIC	Bound state in the continuum
BBC	Bulk-boundary correspondence
BdG	Bogoliubov–de Gennes
BZ	Brillouin zone
BZF	Brillouin zone folding
CPA	Coherent perfect absorption
EP	Exceptional point
EP_n	Exceptional point of order n
FDTD	Finite-difference time-domain
FEM	Finite-element method
GBZ	Generalized Brillouin zone
LDOS	Local density of states
NH	Non-Hermitian
NHSE	Non-Hermitian skin effect
OBC	Open boundary conditions
PBC	Periodic boundary conditions
PEC	Perfect electric conductor
PMC	Perfect magnetic conductor
PML	Perfectly matched layer
$\mathcal{PT}$	Parity–time
PWE	Plane-wave expansion
QNM	Quasinormal mode
SNR	Signal-to-noise ratio
SSH	Su–Schrieffer–Heeger
TCMT	Temporal coupled-mode theory

E.8 Sign conventions and equivalent formulations

Non-Hermitian photonics inherits two competing sign conventions from the optics and physics communities, and careless mixing of conventions is a common source of sign errors in the literature. This section records the choices made in this book and their relationship to the alternatives.

E.8.1 Time-harmonic convention

The convention throughout is $e^{-i\omega t}$, following the majority of the optics and photonics literature. The condensed-matter and high-energy communities often use $e^{+i\omega t}$. The two conventions are related by complex conjugation of all frequency-domain quantities:

Quantity	$e^{-i\omega t}$ (this book)	$e^{+i\omega t}$
Passive loss	$\mathrm{Im}\,\varepsilon_\mathrm{r} > 0$	$\mathrm{Im}\,\varepsilon_\mathrm{r} < 0$
Gain	$\mathrm{Im}\,\varepsilon_\mathrm{r} < 0$	$\mathrm{Im}\,\varepsilon_\mathrm{r} > 0$
QNM decay	$\tilde{\omega} = \omega_r - i\gamma$	$\tilde{\omega} = \omega_r + i\gamma$
Outgoing wave	e^{+ikr}	e^{-ikr}
Kramers–Kronig	$\varepsilon_\mathrm{r}(\omega)$ analytic in upper half-plane	lower half-plane

When importing results from a paper that uses the opposite convention, replace $\omega \to -\omega^*$ (equivalently, $i \to -i$ in all material parameters and eigenvalues). The physical observables—power, transmission, quality factor—are convention-independent.

E.8.2 Three formulations of the Petermann factor

The Petermann factor quantifies the non-orthogonality of QNMs and appears in three equivalent forms in the literature. All three are used in this book at different points; we record the equivalences here for reference.

Formulation 1: unconjugated-norm ratio (Chapter 6).

$$K_n = \frac{\langle \tilde{\mathbf{E}}_n^*, \tilde{\mathbf{E}}_n^* \rangle \cdot \langle \tilde{\mathbf{E}}_n, \tilde{\mathbf{E}}_n \rangle}{|\langle \tilde{\mathbf{E}}_n, \tilde{\mathbf{E}}_n \rangle_\mathrm{uc}|^2}, \tag{E.1}$$

where $\langle \cdot, \cdot \rangle$ is the standard conjugated inner product and $\langle \cdot, \cdot \rangle_\mathrm{uc}$ is the unconjugated (bilinear) inner product.

Formulation 2: biorthogonal overlap (Appendix B).

$$K_n = \frac{1}{|\langle l_n | r_n \rangle|^2}, \tag{E.2}$$

where $|r_n\rangle$ and $\langle l_n|$ are the right and left eigenvectors first normalized to unit norm in the chosen conjugated inner product. Equivalently, without that convention one should use $K_n = \langle r_n | r_n \rangle \langle l_n | l_n \rangle / |\langle l_n | r_n \rangle|^2$. This formulation makes the connection to non-normality most transparent: $K_n = 1$ for a normal operator and $K_n \to \infty$ at an EP where $\langle l_n | r_n \rangle \to 0$.

Formulation 3: phase rigidity (Chapter 8).

$$K_n = \frac{1}{|r_n^\mathrm{ph}|^2}, \qquad r_n^\mathrm{ph} = \frac{\langle \tilde{\mathbf{E}}_n, \tilde{\mathbf{E}}_n \rangle_\mathrm{uc}}{\langle \tilde{\mathbf{E}}_n^*, \tilde{\mathbf{E}}_n \rangle}, \tag{E.3}$$

where r_n^{ph} is the phase rigidity measuring the ratio of the unconjugated to the conjugated inner product. The phase rigidity equals unity for a standing wave (Hermitian mode) and decreases toward zero as the mode becomes a travelling wave (non-Hermitian, non-normal).

All three formulations give the same numerical value for a given QNM when the same bilinear form and normalization convention are used. The Cauchy–Schwarz inequality guarantees $K_n \geq 1$, with equality when the corresponding left and right vectors are collinear in the chosen inner product; normal operators satisfy this for all modes.

E.8.3 Gain–loss sign in coupled-mode models

In the $\mathcal{PT}$-symmetric coupled-mode models of Chapter 9, the Hamiltonian is written as

$$H = \begin{pmatrix} \omega_0 + i\gamma & \kappa \\ \kappa & \omega_0 - i\gamma \end{pmatrix}. \tag{E.4}$$

With our $e^{-i\omega t}$ convention, $+i\gamma$ corresponds to *gain* (the mode amplitude grows as $e^{+\gamma t}$) and $-i\gamma$ corresponds to *loss*. In the opposite convention ($e^{+i\omega t}$), the signs are reversed: $-i\gamma$ is gain and $+i\gamma$ is loss.

The $\mathcal{PT}$ phase transition occurs at $\gamma = \kappa$ in both conventions. Below the transition ($\gamma < \kappa$), the eigenvalues are real and the system is in the exact $\mathcal{PT}$-symmetric phase. Above the transition ($\gamma > \kappa$), the eigenvalues form a complex-conjugate pair and the system is in the broken phase. The physical content is identical; only the intermediate sign bookkeeping differs.

E.8.4 Transfer-matrix ordering

Two conventions exist for the transfer matrix (Appendix C):

1. *Left-to-right* (this book): $\binom{A}{B}_{\mathrm{right}} = \mathbf{T}\binom{A}{B}_{\mathrm{left}}$.
2. *Right-to-left*: $\binom{A}{B}_{\mathrm{left}} = \mathbf{T}'\binom{A}{B}_{\mathrm{right}}$.

The two are related by $\mathbf{T}' = \mathbf{T}^{-1}$. For a reciprocal medium ($\det \mathbf{T} = 1$), the inverse is obtained by swapping the diagonal elements and negating the off-diagonal elements. When comparing transfer-matrix results across papers, always check which convention is used before multiplying matrices.

Bibliography

[1] Alexander Figotin and Jeffrey Schenker. Hamiltonian treatment of time dispersive and dissipative media within the linear response theory. *Journal of Computational and Applied Mathematics*, 204(2):199–208, July 2007.

[2] Aaswath Raman and Shanhui Fan. Photonic band structure of dispersive metamaterials formulated as a hermitian eigenvalue problem. In *Imaging and Applied Optics Congress*, PMETA_PLAS, page MTuA2. OSA, 2010.

[3] Stefano Longhi. Parity-time symmetry meets photonics: A new twist in non-hermitian optics. *EPL (Europhysics Letters)*, 120(6):64001, December 2017.

[4] Philip Trøst Kristensen, C. Van Vlack, and Stephen Hughes. Effective mode volumes and purcell factors for leaky optical cavities. *arXiv (Cornell University)*, 2011.

[5] Benjamin Vial, Frédéric Zolla, André Nicolet, and Mireille Commandré. Quasimodal expansion of electromagnetic fields in open two-dimensional structures. *Physical Review A*, 89(2), February 2014.

[6] Filippo Alpeggiani, Nikhil Parappurath, Ewold Verhagen, and L. Kuipers. Quasinormal-mode expansion of the scattering matrix. *Physical Review X*, 7(2), 2017.

[7] locked icon. Rev. mod. phys. 91, 015006 (2019) - topological photonics, 2019.

[8] F. D. M. Haldane and S. Raghu. Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry. *Physical Review Letters*, 100(1), January 2008.

[9] Tomoki Ozawa, Hannah M. Price, Alberto Amo, Nathan Goldman, Mohammad Hafezi, Ling Lu, Mikael C. Rechtsman, David Schuster, Jonathan Simon, Oded Zilberberg, and Iacopo Carusotto. Topological photonics. *Reviews of Modern Physics*, 91(1), March 2019.

[10] Emil J. Bergholtz, Jan Carl Budich, and Flore K. Kunst. Exceptional topology of non-hermitian systems. *Reviews of Modern Physics*, 93(1), February 2021.

[11] Kohei Kawabata, Ken Shiozaki, Masahito Ueda, and Masatoshi Sato. Symmetry and topology in non-hermitian physics. *Physical Review X*, 9(4), October 2019.

[12] María Blanco de Paz, Chiara Devescovi, Geza Giedke, Juan José Saenz, Maia G. Vergniory, Barry Bradlyn, Dario Bercioux, and Aitzol García-Etxarri. Tutorial: Computing topological invariants in 2d photonic crystals. *Advanced Quantum Technologies*, 3(2), December 2019.

[13] Jing Xu and Yuntian Chen. General coupled mode theory in non-hermitian waveguides. *Optics Express*, 23(17):22619, August 2015.

[14] Zongping Gong, Yuto Ashida, Kohei Kawabata, Kazuaki Takasan, Sho Higashikawa, and Masahito Ueda. Topological phases of non-hermitian systems. *Physical Review X*, 8(3), September 2018.

[15] Emil J. Bergholtz, Jan Carl Budich, and Flore K. Kunst. Exceptional topology of non-hermitian systems. *Reviews of Modern Physics*, 93(1), February 2021.

[16] Long-Hua Wu and Xiao Hu. Scheme for achieving a topological photonic crystal by using dielectric material. *Physical Review Letters*, 114(22), 2015.

[17] Chia Wei Hsu, Bo Zhen, A. Douglas Stone, John D. Joannopoulos, and Marin Soljačić. Bound states in the continuum. *Nature Reviews Materials*, 1(9), July 2016.

[18] Mikael C. Rechtsman, Julia M. Zeuner, Yonatan Plotnik, Yaakov Lumer, Daniel Podolsky, Felix Dreisow, Stefan Nolte, Mordechai Segev, and Alexander Szameit. Photonic floquet topological insulators. *Nature*, 496(7444):196–200, April 2013.

[19] Alexander B. Khanikaev, S. Hossein Mousavi, Wang-Kong Tse, Mehdi Kargarian, Allan H. MacDonald, and Gennady Shvets. Photonic topological insulators. *Nature Materials*, 12(3):233–239, December 2012.

[20] M. Hafezi, S. Mittal, J. Fan, A. Migdall, and J. M. Taylor. Imaging topological edge states in silicon photonics. *Nature Photonics*, 7(12):1001–1005, October 2013.

[21] Ling Lu, John D. Joannopoulos, and Marin Soljačić. Topological photonics. *Nature Photonics*, 8(11):821–829, October 2014.

[22] C. Sauvan, J. P. Hugonin, R. Carminati, and P. Lalanne. Modal representation of spatial coherence in dissipative and resonant photonic systems. *Physical Review A*, 89(4), April 2014.

[23] Jakob Rosenkrantz de Lasson, Philip Trøst Kristensen, Jesper Mørk, and Niels Gregersen. Roundtrip matrix method for calculating the leaky resonant modes of open nanophotonic structures. *Journal of the Optical Society of America A*, 31(10):2142, September 2014.

[24] William R. Sweeney, Chia Wei Hsu, Stefan Rotter, and A. Douglas Stone. Perfectly absorbing exceptional points and chiral absorbers. *Physical Review Letters*, 122(9), March 2019.

[25] Changqing Wang, William R. Sweeney, A. Douglas Stone, and Lan Yang. Coherent perfect absorption at an exceptional point. *Science*, 373(6560):1261–1265, September 2021.

[26] Zhexin Zhao, Cheng Guo, and Shanhui Fan. Connection of temporal coupled-mode-theory formalisms for a resonant optical system and its time-reversal conjugate. *Physical Review A*, 99(3), March 2019.

[27] Phys. rev. lett. 121, 136802 (2018) - non-hermitian chern bands, 2018.

[28] David M. Adams, Roger Appleby, and Shiv K. Sharma. Spectroscopy at very high pressures. x. use of ruby r-lines in the estimation of pressure at ambient and at low temperatures. *Journal of Physics E Scientific Instruments*, 1976.

[29] Alexander B. Khanikaev and Andrea Alù. Topological photonics: robustness and beyond. *Nature Communications*, 15(1), January 2024.

[30] Diego M. Solís and Nader Engheta. Functional analysis of the polarization response in linear time-varying media: A generalization of the kramers-kronig relations. *Physical Review B*, 103(14), April 2021.

[31] Lujun Huang, Lei Xu, David A. Powell, Willie J. Padilla, and Andrey E. Miroshnichenko. Resonant leaky modes in all-dielectric metasystems: Fundamentals and applications. *Physics Reports*, 1008:1–66, April 2023.

[32] Kirill Koshelev, Gael Favraud, Andrey Bogdanov, Yuri Kivshar, and Andrea Fratalocchi. Nonradiating photonics with resonant dielectric nanostructures. *Nanophotonics*, 8(5):725–745, 2018.

[33] Christian E. Rüter, Konstantinos G. Makris, Ramy El-Ganainy, Demetrios N. Christodoulides, Mordechai Segev, and Detlef Kip. Observation of parity–time symmetry in optics. *Nature Physics*, 6(3):192–195, January 2010.

[34] Liang Feng, Ye-Long Xu, William S. Fegadolli, Ming-Hui Lu, José E. B. Oliveira, Vilson R. Almeida, Yan-Feng Chen, and Axel Scherer. Experimental demonstration of a unidirectional reflectionless parity-time metamaterial at optical frequencies. *Nature Materials*, 12(2):108–113, November 2012.

[35] Opg. Dispersive perfectly matched layers and high-order absorbing boundary conditions for electromagnetic quasinormal modes. https://opg.optica.org/josaa/abstract.cfm?uri=josaa-40-10-1947, 2023.

[36] Tong Wu and Philippe Lalanne. Exact maxwell evolution equation of resonator dynamics: temporal coupled-mode theory revisited. *Optics Express*, 32(12):20904, May 2024.

[37] Juanjuan Ren, Sebastian Franke, and Stephen Hughes. Quasinormal modes, local density of states, and classical purcell factors for coupled loss-gain resonators. *Physical Review X*, 11(4), October 2021.

[38] Tong Wu and Philippe Lalanne. Designing electromagnetic resonators with quasinormal modes. *Frontiers in Physics*, 12, August 2024.

[39] Tong Wu and Philippe Lalanne. Rigorous electromagnetic quasinormal-mode method made easy for users. *Laser & Photonics Reviews*, March 2026.

[40] Kenta Takata, Nathan Roberts, Akihiko Shinya, and Masaya Notomi. Imaginary couplings in non-hermitian coupled-mode theory: Effects on exceptional points of optical resonators. *Physical Review A*, 105(1), January 2022.

[41] Opg. Non-hermitian temporal coupled-mode theory: Effects of imaginary couplings on exceptional points. http://www.opg.optica.org/abstract.cfm?uri=FiO-2021-FM2E.1, 2021.

[42] Bo Zhen, Chia Wei Hsu, Yuichi Igarashi, Ling Lu, Ido Kaminer, Adi Pick, Song-Liang Chua, John D. Joannopoulos, and Marin Soljačić. Spawning rings of exceptional points out of dirac cones. *Nature*, 525(7569):354–358, September 2015.

[43] Mohammad-Ali Miri and Andrea Alù. Exceptional points in optics and photonics. *Science*, 363(6422), January 2019.

[44] Kohei Kawabata, Takumi Bessho, and Masatoshi Sato. Classification of exceptional points and non-hermitian topological semimetals. *Physical Review Letters*, 123(6), August 2019.

[45] Jan Wiersig. Petermann factors and phase rigidities near exceptional points. *Physical Review Research*, 5(3), July 2023.

[46] Dalton Anderson, Manav Shah, and Linran Fan. Clarification of the exceptional-point contribution to photonic sensing. *Physical Review Applied*, 19(3), March 2023.

[47] David D. Smith, Hongrok Chang, Eugeniy Mikhailov, and Selim M. Shahriar. Beyond the petermann limit: can exceptional points increase sensor precision? *arXiv e-prints*, 2020.

[48] David D. Smith, Hongrok Chang, Eugeniy Mikhailov, and Selim M. Shahriar. Beyond the petermann limit: Prospect of increasing sensor precision near exceptional points. *Physical Review A*, 106(1), July 2022.

[49] Mengzhen Zhang, William Sweeney, Chia Wei Hsu, Lan Yang, A. D. Stone, and Liang Jiang. Quantum noise theory of exceptional point amplifying sensors. *Physical Review Letters*, 123(18), October 2019.

[50] Heming Wang, Yu-Hung Lai, Zhiquan Yuan, Myoung-Gyun Suh, and Kerry Vahala. Petermann-factor sensitivity limit near an exceptional point in a brillouin ring laser gyroscope. *Nature Communications*, 11(1), March 2020.

[51] Xu Zheng and Y. D. Chong. Noise constraints for nonlinear exceptional point sensing. *Physical Review Letters*, 134(13), April 2025.

[52] Haiyu Meng, Yee Sin Ang, and Ching Hua Lee. Exceptional points in non-hermitian systems: Applications and recent developments. *Applied Physics Letters*, 124(6), February 2024.

[53] R. El-Ganainy, K. G. Makris, D. N. Christodoulides, and Ziad H. Musslimani. Theory of coupled optical pt-symmetric structures. *Optics Letters*, 32(17):2632, August 2007.

[54] Zin Lin, Hamidreza Ramezani, Toni Eichelkraut, Tsampikos Kottos, Hui Cao, and Demetrios N. Christodoulides. Unidirectional invisibility induced by-symmetric periodic structures. *Physical Review Letters*, 106(21), May 2011.

[55] Y. D. Chong, Li Ge, and A. Douglas Stone. -symmetry breaking and laser-absorber modes in optical scattering systems. *Physical Review Letters*, 106(9), March 2011.

[56] K. G. Makris, R. El-Ganainy, D. N. Christodoulides, and Z. H. Musslimani. Beam dynamics insymmetric optical lattices. *Physical Review Letters*, 100(10), March 2008.

[57] Bo Peng, Şahin Kaya Özdemir, Fuchuan Lei, Faraz Monifi, Mariagiovanna Gianfreda, Gui Lu Long, Shanhui Fan, Franco Nori, Carl M. Bender, and Lan Yang. Parity–time-symmetric whispering-gallery microcavities. *Nature Physics*, 10(5):394–398, April 2014.

[58] N. V. Alexeeva, I. V. Barashenkov, Andrey A. Sukhorukov, and Yuri S. Kivshar. Optical solitons in-symmetric nonlinear couplers with gain and loss. *Physical Review A*, 85(6), 2012.

[59] Shiqi Xia, Dimitrios Kaltsas, Daohong Song, Ioannis Komis, Jingjun Xu, Alexander Szameit, Hrvoje Buljan, Konstantinos G. Makris, and Zhigang Chen. Nonlinear tuning of pt symmetry and non-hermitian topological states. *Science*, 372(6537):72–76, April 2021.

[60] Martin Wimmer, Alois Regensburger, Mohammad-Ali Miri, Christoph Bersch, Demetrios N. Christodoulides, and Ulf Peschel. Observation of optical solitons in pt-symmetric lattices. *Nature Communications*, 6(1), July 2015.

[61] Lewis Hill, Julius T. Gohsrich, Alekhya Ghosh, Jacob Fauman, Pascal Del'Haye, and Flore K. Kunst. Exceptional points preceding and enabling spontaneous symmetry breaking. *Communications Physics*, 9(1), January 2026.

[62] Carl M. Bender and Stefan Boettcher. Real spectra in non-hermitian hamiltonians havingsymmetry. *Physical Review Letters*, 80(24):5243–5246, 1998.

[63] Denis V Novitsky. Cpa-laser effect and exceptional points in $\mathcal{PT}$ -symmetric multilayer structures. *Journal of Optics*, 21(8):085101, July 2019.

[64] Mohammadmahdi Alizadeh, Sahin K. Ozdemir, Alejandro M. Yacomotti, and Ramy El-Ganainy. Lasing zero modes are not always topological. *Communications Physics*, 8(1), November 2025.

[65] Stefano Longhi. $\mathcal{PT}$-symmetric laser absorber. *Physical Review A*, 82(3), September 2010.

[66] Scott A. Skirlo, Ling Lu, Yuichi Igarashi, Qinghui Yan, John Joannopoulos, and Marin Soljačić. Experimental observation of large chern numbers in photonic crystals. *Physical Review Letters*, 115(25), December 2015.

[67] Jiho Noh, Sheng Huang, Kevin P. Chen, and Mikael C. Rechtsman. Observation of photonic topological valley hall edge states. *Physical Review Letters*, 120(6), February 2018.

[68] Qiaolu Chen, Li Zhang, Mengjia He, Zuojia Wang, Xiao Lin, Fei Gao, Yihao Yang, Baile Zhang, and Hongsheng Chen. Valley-hall photonic topological insulators with dual-band kink states. *Advanced Optical Materials*, 7(15), May 2019.

[69] Fei Gao, Haoran Xue, Zhaoju Yang, Kueifu Lai, Yang Yu, Xiao Lin, Yidong Chong, Gennady Shvets, and Baile Zhang. Topologically protected refraction of robust kink states in valley photonic crystals. *Nature Physics*, 14(2):140–144, November 2017.

[70] Yuhao Kang, Xiang Ni, Xiaojun Cheng, Alexander B. Khanikaev, and Azriel Z. Genack. Pseudo-spin–valley coupled edge states in a photonic topological insulator. *Nature Communications*, 9(1), August 2018.

[71] Xin-Tao He, En-Tao Liang, Jia-Jun Yuan, Hao-Yang Qiu, Xiao-Dong Chen, Fu-Li Zhao, and Jian-Wen Dong. A silicon-on-insulator slab for topological valley transport. *Nature Communications*, 10(1), February 2019.

[72] Dawn T. H. Tan. Topological silicon photonics. *Advanced Photonics Research*, 2(9), September 2021.

[73] Tzuhsuan Ma and Gennady Shvets. Scattering-free edge states between heterogeneous photonic topological insulators. *Physical Review B*, 95(16), April 2017.

[74] LIU Hui, WANG Hao-nan, XIE Bo-yang, CHENG Hua, TIAN Jian-guo, and CHEN and Shu-qi. Progress of two-dimensional photonic topological insulators. *Chinese Optics*, 14(4):935–954, 2021.

[75] Cheng He, Xiao-Chen Sun, Xiao-Ping Liu, Ming-Hui Lu, Yulin Chen, Liang Feng, and Yan-Feng Chen. Photonic topological insulator with broken time-reversal symmetry. *Proceedings of the National Academy of Sciences*, 113(18):4924–4928, April 2016.

[76] Chao Chen, Run-Ze Liu, Jizhou Wu, Zu-En Su, Xing Ding, Jian Qin, Lin Wang, Wei-Wei Zhang, Yu He, Xi-Lin Wang, Chao-Yang Lu, Li Li, Barry C. Sanders, Xiong-Jun Liu, and Jian-Wei Pan. Berry curvature and bulk-boundary correspondence from transport measurement for photonic chern bands. *Physical Review Letters*, 131(13), September 2023.

[77] Yihao Yang, Zhen Gao, Haoran Xue, Li Zhang, Mengjia He, Zhaoju Yang, Ranjan Singh, Yidong Chong, Baile Zhang, and Hongsheng Chen. Realization of a three-dimensional photonic topological insulator. *Nature*, 565(7741):622–626, January 2019.

[78] Alexey Slobozhanyuk, S. Hossein Mousavi, Xiang Ni, Daria Smirnova, Yuri S. Kivshar, and Alexander B. Khanikaev. Three-dimensional all-dielectric photonic topological insulator. *Nature Photonics*, 11(2):130–136, December 2016.

[79] Noriaki Horiuchi. Photonic topological insulator. *Nature Photonics*, 7(3):166–166, February 2013.

[80] Alexey P. Slobozhanyuk, Alexander B. Khanikaev, Dmitry S. Filonov, Daria A. Smirnova, Andrey E. Miroshnichenko et al. (6 authors), et al. Observation of topological edge modes in bianisotropic metamaterials, 2015.

[81] Fei Gao, Zhen Gao, Xihang Shi, Zhaoju Yang, Xiao Lin, Hongyi Xu, John D. Joannopoulos, Marin Soljačić, Hongsheng Chen, Ling Lu, Yidong Chong, and Baile Zhang. Probing topological protection using a designer surface plasmon structure. *Nature Communications*, 7(1), May 2016.

[82] Zheng Wang, Yidong Chong, J. D. Joannopoulos, and Marin Soljačić. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature*, 461(7265):772–775, October 2009.

[83] Zheng Wang, Y. D. Chong, John D. Joannopoulos, and Marin Soljačić. Reflection-free one-way edge modes in a gyromagnetic photonic crystal. *Physical Review Letters*, 100(1), January 2008.

[84] Shinsei Ryu, Andreas P Schnyder, Akira Furusaki, and Andreas W W Ludwig. Topological insulators and superconductors: tenfold way and dimensional hierarchy. *New Journal of Physics*, 12(6):065010, 2010.

[85] Shunyu Yao and Zhong Wang. Edge states and topological invariants of non-hermitian systems. *Physical Review Letters*, 121(8), August 2018.

[86] Kazuki Yokomizo. *Non-Bloch Band Theory of Non-Hermitian Systems*. Springer Nature Singapore, 2022.

[87] Flore K. Kunst, Elisabet Edvardsson, Jan Carl Budich, and Emil J. Bergholtz. Biorthogonal bulk-boundary correspondence in non-hermitian systems. *Physical Review Letters*, 121(2), July 2018.

[88] Huitao Shen, Bo Zhen, and Liang Fu. Topological band theory for non-hermitian hamiltonians. *Physical Review Letters*, 120(14), April 2018.

[89] Kohei Kawabata, Ken Shiozaki, Masahito Ueda, and Masatoshi Sato. Symmetry and topology in non-hermitian physics. *Physical Review X*, 9(4), October 2019.

[90] Janet Zhong, Kai Wang, Yubin Park, Viktar Asadchy, Charles C. Wojcik, Avik Dutt, and Shanhui Fan. Nontrivial point-gap topology and non-hermitian skin effect in photonic crystals. *Physical Review B*, 104(12), September 2021.

[91] Ying Chen, Hai-Xiao Wang, Qiaoliang Bao, Jian-Hua Jiang, and Huanyang Chen. Ideal type-ii weyl points in twisted one-dimensional dielectric photonic crystals. *Optics Express*, 29(24):40606, November 2021.

[92] Yuang Pan, Chaoxi Cui, Qiaolu Chen, Fujia Chen, Li Zhang, Yudong Ren, Ning Han, Wenhao Li, Xinrui Li, Zhi-Ming Yu, Hongsheng Chen, and Yihao Yang. Real higher-order weyl photonic crystal. *Nature Communications*, 14(1), October 2023.

[93] Gui-Geng Liu, Zhen Gao, Qiang Wang, Xiang Xi, Yuan-Hang Hu, Maoren Wang, Chengqi Liu, Xiao Lin, Longjiang Deng, Shengyuan A. Yang, Peiheng Zhou, Yihao Yang, Yidong Chong, and Baile Zhang. Topological chern vectors in three-dimensional photonic crystals. *Nature*, 609(7929):925–930, September 2022.

[94] Haedong Park, Wenlong Gao, Xiao Zhang, and Sang Soon Oh. Nodal lines in momentum space: topological invariants and recent realizations in photonic and other systems. *Nanophotonics*, 11(11):2779–2801, February 2022.

[95] Cheng He, Xu Ni, Hao Ge, Xiao-Chen Sun, Yan-Bin Chen, Ming-Hui Lu, Xiao-Ping Liu, and Yan-Feng Chen. Acoustic topological insulator and robust one-way sound transport. *Nature Physics*, 12(12):1124–1129, August 2016.

[96] Li Zhang, Yihao Yang, Yong Ge, Yi-Jun Guan, Qiaolu Chen, Qinghui Yan, Fujia Chen, Rui Xi, Yuanzhen Li, Ding Jia, Shou-Qi Yuan, Hong-Xiang Sun, Hongsheng Chen, and Baile Zhang. Acoustic non-hermitian skin effect from twisted winding topology. *Nature Communications*, 12(1), November 2021.

[97] Stefano Longhi. Probing non-hermitian skin effect and non-bloch phase transitions. *Physical Review Research*, 1(2), September 2019.

[98] Xiujuan Zhang, Yuan Tian, Jian-Hua Jiang, Ming-Hui Lu, and Yan-Feng Chen. Observation of higher-order non-hermitian skin effect. *Nature Communications*, 12(1), September 2021.

[99] Rijia Lin, Tommy Tai, Linhu Li, and Ching Hua Lee. Topological non-hermitian skin effect. *Frontiers of Physics*, 18(5), July 2023.

[100] Kai Zhang, Zhesen Yang, and Chen Fang. Universal non-hermitian skin effect in two and higher dimensions. *Nature Communications*, 13(1), May 2022.

[101] Xinyi Yuan, Loïc Malgrey, Helgi Sigurðsson, Hai Son Nguyen, and Grazia Salerno. Breakdown of bulk-radiation correspondence in radiative photonic lattices. *Physical Review Research*, 7(4), November 2025.

[102] Naomichi Hatano and David R. Nelson. Localization transitions in non-hermitian quantum mechanics. *Physical Review Letters*, 77(3):570–573, July 1996.

[103] Zhesen Yang, Kai Zhang, Chen Fang, and Jiangping Hu. Non-hermitian bulk-boundary correspondence and auxiliary generalized brillouin zone theory. *Physical Review Letters*, 125(22), November 2020.

[104] Xiaosen Yang, Yaru Feng, Abdul Wahab, and Hao Geng. Non-hermitian second-order topological phases and bipolar skin effect in photonic kagome crystals. *Physical Review A*, 113(2), February 2026.

[105] Linhu Li, Ching Hua Lee, Sen Mu, and Jiangbin Gong. Critical non-hermitian skin effect. *Nature Communications*, 11(1), October 2020.

[106] Meng Kang, Tao Liu, C. T. Chan, and Meng Xiao. Applications of bound states in the continuum in photonics. *Nature Reviews Physics*, 5(11):659–678, October 2023.

[107] Kirill Koshelev, Sergey Lepeshov, Mingkai Liu, Andrey Bogdanov, and Yuri Kivshar. Asymmetric metasurfaces with high- q resonances governed by bound states in the continuum. *Physical Review Letters*, 121(19), November 2018.

[108] Meng Kang, Li Mao, Shunping Zhang, Meng Xiao, Hongxing Xu, and Che Ting Chan. Merging bound states in the continuum by harnessing higher-order topological charges. *Light: Science & Applications*, 11(1), July 2022.

[109] Huayu Bai, Andriy Shevchenko, and Radoslaw Kolkowski. Recovery of topologically robust merging bound states in the continuum in photonic structures with broken symmetry. *Nanophotonics*, 14(7):899–913, March 2025.

[110] Yixiao Gao, Junyang Ge, Zhaofeng Gu, Lei Xu, Xiang Shen, and Lujun Huang. Degenerate merging bics in resonant metasurfaces. *Optics Letters*, 49(23):6633, November 2024.

[111] Wenhao Wang, Yogesh Kumar Srivastava, Thomas CaiWei Tan, Zhiming Wang, and Ranjan Singh. Brillouin zone folding driven bound states in the continuum. *Nature Communications*, 14(1), May 2023.

[112] Bo Zhen, Chia Wei Hsu, Ling Lu, A. Douglas Stone, and Marin Soljačić. Topological nature of optical bound states in the continuum. *Physical Review Letters*, 113(25), December 2014.

[113] Yonatan Plotnik, Or Peleg, Felix Dreisow, Matthias Heinrich, Stefan Nolte, Alexander Szameit, and Mordechai Segev. Experimental observation of optical bound states in the continuum. *Physical Review Letters*, 107(18), October 2011.

[114] Sughra Mohamed, Jie Wang, Heikki Rekola, Janne Heikkinen, Benjamin Asamoah et al. (7 authors), et al. Topological charge engineering in lasing bound states in continuum, 2020.

[115] Hugo M. Doeleman, Francesco Monticone, Wouter den Hollander, Andrea Alù, and A. Femius Koenderink. Experimental observation of a polarization vortex at an optical bound state in the continuum. *Nature Photonics*, 12(7):397–401, 2018.

[116] Bo Wang, Wenzhe Liu, Maoxiong Zhao, Jiajun Wang, Yiwen Zhang, Ang Chen, Fang Guan, Xiaohan Liu, Lei Shi, and Jian Zi. Generating optical vortex beams by momentum-space polarization vortices centred at bound states in the continuum. *Nature Photonics*, 14(10):623–628, July 2020.

[117] Wenzhe Liu, Wei Liu, Lei Shi, and Yuri Kivshar. Topological polarization singularities in metaphotonics. *Nanophotonics*, 10(5):1469–1486, February 2021.

[118] Taiki Yoda and Masaya Notomi. Generation and annihilation of topologically protected bound states in the continuum and circularly polarized states by symmetry breaking. *Physical Review Letters*, 125(5), July 2020.

[119] Haoye Qin, Zengping Su, Mengqi Liu, Yixuan Zeng, Man-Chung Tang, Mengyao Li, Yuzhi Shi, Wei Huang, Cheng-Wei Qiu, and Qinghua Song. Arbitrarily polarized bound states in the continuum with twisted photonic crystal slabs. *Light: Science & Applications*, 12(1), March 2023.

[120] F. Guinea, M. I. Katsnelson, and A. K. Geim. Energy gaps and a zero-field quantum hall effect in graphene by strain engineering. *Nature Physics*, 6(1):30–33, September 2009.

[121] Min-Soo Hwang, Hoo-Cheol Lee, Kyoung-Ho Kim, Kwang-Yong Jeong, Soon-Hong Kwon, Kirill Koshelev, Yuri Kivshar, and Hong-Gyu Park. Ultralow-threshold laser using super-bound states in the continuum. *Nature Communications*, 12(1), July 2021.

[122] Felix Binkowski, Fridtjof Betz, Rémi Colom, Patrice Genevet, and Sven Burger. Poles and zeros in non-hermitian systems: Application to photonics. *Physical Review B*, 109(4), January 2024.

[123] Tie Jun Cui, Shuang Zhang, Andrea Alù, Martin Wegener, Sir John Pendry, Jie Luo, Yun Lai, Zuojia Wang, Xiao Lin, Hongsheng Chen, Ping Chen, Rui-Xin Wu, Yuhang Yin, Pengfei Zhao, Huanyang Chen, Yue Li, Ziheng Zhou, Nader Engheta, Viktar Asadchy, Constantin Simovski, Sergei Tretyakov, Biao Yang, Sawyer D Campbell, Yang Hao, Douglas H Werner, Shulin Sun, Lei Zhou, Su Xu, Hong-Bo Sun, Zhou Zhou, Zile Li, Guoxing Zheng, Xianzhong Chen, Tao Li, Shining Zhu, Junxiao Zhou, Junxiang Zhao, Zhaowei Liu, Yuchao Zhang,

Qiming Zhang, Min Gu, Shumin Xiao, Yongmin Liu, Xianzhe Zhang, Yutao Tang, Guixin Li, Thomas Zentgraf, Kirill Koshelev, Yuri Kivshar, Xin Li, Trevon Badloe, Lingling Huang, Junsuk Rho, Shuming Wang, Din Ping Tsai, A Yu Bykov, A V Krasavin, A V Zayats, Cormac McDonnell, Tal Ellenbogen, Xiangang Luo, Mingbo Pu, Francisco J Garcia-Vidal, Liangliang Liu, Zhuo Li, Wenxuan Tang, Hui Feng Ma, Jingjing Zhang, Yu Luo, Xuanru Zhang, Hao Chi Zhang, Pei Hang He, Le Peng Zhang, Xiang Wan, Haotian Wu, Shuo Liu, Wei Xiang Jiang, Xin Ge Zhang, Cheng-Wei Qiu, Qian Ma, Che Liu, Long Li, Jiaqi Han, Lianlin Li, Michele Cotrufo, C Caloz, Z-L Deck-Léger, A Bahrami, O Céspedes, E Galiffi, P A Huidobro, Qiang Cheng, Jun Yan Dai, Jun Cheng Ke, Lei Zhang, Vincenzo Galdi, and Marco di Renzo. Roadmap on electromagnetic metamaterials and metasurfaces. *Journal of Physics: Photonics*, 6(3):032502, July 2024.

[124] J. Gopalakrishnan, B. . Parker, and P. VandenBerge. Computing leaky modes of optical fibers using a feast algorithm for polynomial eigenproblems. https://web.pdx.edu/~gjay/pub/polyfeast.pdf, 2022.

[125] Miguel A. Bandres, Steffen Wittek, Gal Harari, Midya Parto, Jinhan Ren, Mordechai Segev, Demetrios N. Christodoulides, and Mercedeh Khajavikhan. Topological insulator laser: Experiments. *Science*, 359(6381), March 2018.

[126] Gal Harari, Miguel A. Bandres, Yaakov Lumer, Mikael C. Rechtsman, Y. D. Chong, Mercedeh Khajavikhan, Demetrios N. Christodoulides, and Mordechai Segev. Topological insulator laser: Theory. *Science*, 359(6381), March 2018.

[127] Yasutomo Ota, Kenta Takata, Tomoki Ozawa, Alberto Amo, Zhetao Jia, Boubacar Kante, Masaya Notomi, Yasuhiko Arakawa, and Satoshi Iwamoto. Active topological photonics. *Nanophotonics*, 9(3):547–567, January 2020.

[128] Bofeng Zhu, Hanyu Liu, Qian Wang, Y. D. Chong, and Qi Jie Wang. Topology in electrically pumped photonic-crystal lasers. *Nature Reviews Electrical Engineering*, April 2026.

[129] Qian Ma, Che Liu, Qiang Xiao, Ze Gu, Xinxin Gao, Lianlin Li, and Tie Jun Cui. Information metasurfaces and intelligent metasurfaces. *Photonics Insights*, 1(1):R01, 2022.

[130] Bo Zhen. Nonlinear topological photonics. In Ganapathi S. Subramania and Stavroula Foteinopoulou, editors, *Active Photonic Platforms (APP) 2023*, page 67. SPIE, October 2023.

[131] Lukas J. Maczewsky, Matthias Heinrich, Mark Kremer, Sergey K. Ivanov, Max Ehrhardt, Franklin Martinez, Yaroslav V. Kartashov, Vladimir V. Konotop, Lluis Torner, Dieter Bauer, and Alexander Szameit. Nonlinearity-induced photonic topological insulator. *Science*, 370(6517):701–704, November 2020.

[132] Daniel Leykam and Y. D. Chong. Edge solitons in nonlinear-photonic topological insulators. *Physical Review Letters*, 117(14), September 2016.

[133] Rujiang Li, Xiangyu Kong, Dongkai Hang, Guoyi Li, Hongyu Hu, Hao Zhou, Yongtao Jia, Pengfei Li, and Ying Liu. Topological bulk solitons in a nonlinear photonic chern insulator. *Communications Physics*, 5(1), November 2022.

[134] Yaroslav V. Kartashov. Solitons in higher-order topological insulator created by unit cell twisting. *Chaos, Solitons & Fractals*, 185:115188, August 2024.

[135] Qian Tang, Yiqi Zhang, Yaroslav V. Kartashov, Yongdong Li, and Vladimir V. Konotop. Vector valley hall edge solitons in superhoneycomb lattices. *Chaos, Solitons & Fractals*, 161:112364, August 2022.

[136] Minkyung Kim, Zubin Jacob, and Junsuk Rho. Recent advances in 2d, 3d and higher-order topological photonics. *Light: Science & Applications*, 9(1), July 2020.

[137] Lukas J. Maczewsky, Julia M. Zeuner, Stefan Nolte, and Alexander Szameit. Observation of photonic anomalous floquet topological insulators. *Nature Communications*, 8(1), January 2017.

[138] Takuya Kitagawa, Erez Berg, Mark Rudner, and Eugene Demler. Topological characterization of periodically driven quantum systems. *Physical Review B*, 82(23), December 2010.

[139] Mark S. Rudner, Netanel H. Lindner, Erez Berg, and Michael Levin. Anomalous edge states and the bulk-edge correspondence for periodically driven two-dimensional systems. *Physical Review X*, 3(3), July 2013.

[140] Frederik Nathan, Dmitry Abanin, Erez Berg, Netanel H. Lindner, and Mark S. Rudner. Anomalous floquet insulators. *Physical Review B*, 99(19), May 2019.

[141] Jonathan Guglielmon, Sheng Huang, Kevin P. Chen, and Mikael C. Rechtsman. Photonic realization of a transition to a strongly driven floquet topological phase. *Physical Review A*, 97(3), March 2018.

[142] Aaron Farrell and T. Pereg-Barnea. Edge-state transport in floquet topological insulators. *Physical Review B*, 93(4), January 2016.

[143] Sebabrata Mukherjee and Mikael C. Rechtsman. Observation of unidirectional solitonlike edge states in nonlinear floquet topological insulators. *Physical Review X*, 11(4), December 2021.

[144] Magnus Johansson, Goran Gligorić, Aleksandra Maluckov, Petra P. Beličev, Rodrigo A. Vicencio, and Milutin Stepić. Floquet lattice solitons in zigzag modulated waveguide arrays with zero average modulation: Exponential localization and linear stability. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 35(12), December 2025.

[145] Shuang Shen, Yaroslav V. Kartashov, Yongdong Li, and Yiqi Zhang. Floquet edge solitons in modulated trimer waveguide arrays. *Physical Review Applied*, 20(1), July 2023.

[146] Ruichang Chen, Wenchao Yan, Weijie Liu, Weizhao Cheng, Qingming Lu, Yang Tan, and Feng Chen. Nonlinear tuning of multiple topological edge states in photovoltaic photonic lattices. *Science Bulletin*, 70(10):1605–1610, May 2025.

[147] Mingyuan Gao, Chong Sheng, Yule Zhao, Runqiu He, Liangliang Lu, Wei Chen, Kun Ding, Shining Zhu, and Hui Liu. Quantum walks of correlated photons in non-hermitian photonic lattices. *Physical Review B*, 110(9), September 2024.

[148] Antonio Acín, Immanuel Bloch, Harry Buhrman, Tommaso Calarco, Christopher Eichler, et al. The quantum technologies roadmap: a european community view. *New Journal of Physics*, 2018.