
Topological Photonics from Maxwell's Equations

A Maxwell-First Foundation for Topological Phases of Light

Wenhao Li and Yihao Yang*

*Corresponding author: yangyihao@zju.edu.cn

Second Edition Draft

A Maxwell-first monograph on topological photonics: Bloch bands, energy inner products, Berry curvature, Chern numbers, edge states, synthetic gauge fields, Floquet driving, non-Hermitian topology, bound states in the continuum, valley photonics, three-dimensional topological phases, and nonlinear topological effects, all organized from the frequency-domain Maxwell operator rather than by analogy alone.

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.

Why this book.

Topological photonics has grown from the proposal of one-way photonic crystal waveguides to a broad framework for controlling light with band geometry, symmetry, and topology. Many introductions begin from the language of electronic topological insulators and then translate the results to optics. This book instead keeps Maxwell’s equations at the center. The six-component electromagnetic field, the material-dependent energy inner product, bosonic time reversal, dispersion, bianisotropy, loss, gain, and radiation are treated as native features of the subject, not as afterthoughts to a condensed-matter analogy.

The intended reader is a graduate student or researcher with background in electrodynamics and some familiarity with quantum mechanics or solid-state physics. No prior command of differential geometry or topological band theory is assumed; Berry phases, Chern numbers, bulk–edge correspondence, and related tools are introduced when the Maxwell problems require them.

Organization of the manuscript. Part I develops the Maxwell operator, electromagnetic inner products, Bloch bands, Berry geometry, Chern numbers, and the basic two-dimensional photonic Chern-insulator picture. Part II turns to synthetic gauge fields, coupled resonators, Floquet modulation, non-Hermitian effects, and bound states in the continuum. Part III covers

valley photonics, three-dimensional topological phases, Weyl physics, and nonlinear frontiers. The appendices collect differential-form notation, symmetry background, numerical algorithms, and spectral-theoretic material used throughout the text.

The manuscript uses SI units and the time dependence $e^{-i\omega t}$. Its emphasis is pedagogical but Maxwell-first: electronic analogies are used for orientation, while the derivations are written in the language of electromagnetic modes and material response.

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

Book page: <https://oxelra-yhy.github.io/Lumin/books/topological-photonics/>

Book review repository: <https://github.com/Oxelra-yhy/Topological-Photonics>

Contents

I	Maxwell Foundations and Two-Dimensional Topology	1
1	Introduction	2
	Foundational Clarifications	2
1.1	Why Do We Need “Yet Another” Topological Photonics Textbook?	2
1.2	A Brief History of Photonic Topological States	3
1.2.1	Intellectual Origins: From the Quantum Hall Effect to Photonic Analogies	3
1.2.2	Foundational Papers: Raghu–Haldane and the Wang Experiment	4
1.2.3	From Broken TRS to Preserved TRS: The Photonic Spin Hall Effect	4
1.2.4	Three-Dimensional Topological Photonics: Weyl Points and Nodal Lines .	5
1.2.5	Non-Hermitian Topology and Topological Lasers	5
1.2.6	Floquet Topology and Synthetic Dimensions	5
1.3	Overview of the Book	5
1.4	Comparison of Representative Schemes	8
2	Maxwell Operator	11
	Why a Maxwell-first Foundation is Necessary?	11
2.1	Time–Frequency Domain and Complex Amplitudes	12
2.1.1	Source-Free Maxwell Equations	12
2.1.2	Monochromatic Steady State and Time-Factor Convention	12
2.2	Linear Relations and Material Matrix	12
2.2.1	The Most General Linear Local Constitutive Relations	12
2.2.2	The Lossless Condition	13
2.3	The Six-Component Field Vector and the Maxwell Operator	16
2.3.1	Compact Form	16
2.3.2	The Generalized Eigenvalue Problem Under the Non-Dispersive Approximation	16
2.4	The Energy Inner Product and Positive Definiteness	16
2.4.1	Why Six Components Rather Than Three?	17
2.4.2	The Material-Weighted (Energy) Inner Product	20
2.5	The Self-Adjointness Theorem	20

2.6	Influence of Boundary Conditions	24
2.6.1	Closed Boundaries	24
2.6.2	Periodic Boundaries and Bloch Parameterization	25
2.6.3	Open Boundaries and Complex-Frequency Spectra	25
2.7	Reciprocity: The Lorentz Reciprocity Theorem	25
2.7.1	Microscopic Origin: Onsager Reciprocal Relations	26
2.8	Time-Reversal Symmetry	27
2.8.1	The Anti-Unitary Operator $\mathcal{T}$	27
2.8.2	Conditions for TRS to Hold	27
2.8.3	Distinguishing TRS from Reciprocity	28
2.8.4	Photonic $\mathcal{T}^2 = +1$ and the Absence of Kramers' Theorem	30
2.9	Standard Eigenvalue Problems	32
2.9.1	Numerical Standardization	32
2.10	Energy Inner Product for Dispersive Media	33
	Effective Hermitianization of Dispersive Media	33
	Chapter Summary	34
3	Bloch's Theorem and Photonic Band Theory	35
	Why Periodicity Changes the Spectral Problem?	35
3.1	Photonic Crystals and Lattice Structure	36
3.1.1	Mathematical Formulation of Spatial Periodicity	36
3.1.2	Reciprocal Lattice and Brillouin Zone	36
3.2	Translation Operators and Bloch's Theorem	37
3.2.1	Definition and Properties of Translation Operators	37
3.2.2	Bloch's Theorem	37
3.2.3	Why Does Periodicity Necessarily Lead to Energy Bands?	38
3.3	The Cell-Periodic Maxwell Operator $\hat{N}(\mathbf{k})$	42
3.3.1	Substituting the Bloch Form into Maxwell's Equations	42
3.3.2	Cell-Periodic Energy Inner Product	42
3.3.3	Discrete Spectrum of the Cell-Periodic Problem	43
3.4	Photonic Band Structure	44
3.4.1	Dispersion Relations $\omega_n(\mathbf{k})$	44
3.4.2	Band Diagrams Along High-Symmetry Lines	44
3.5	Photonic Bandgaps	44
3.5.1	Definition of Bandgaps	44
3.5.2	Physical Factors Affecting the Bandgap	45
3.5.3	Two-Dimensional Scalar Equations	45
3.6	Variational Principle and Bandgap Estimation	46
3.6.1	Rayleigh Quotient	46
3.6.2	Physical Intuition for the Bandgap	46

3.6.3	Redundancy of $\mathbf{k}$	46
3.6.4	Periodic Gauge	47
3.7	Overview of Numerical Methods	47
3.7.1	Plane-Wave Expansion Method (PWE)	47
3.7.2	Finite Element Method (FEM)	47
3.7.3	FDTD and Time-Domain Methods	48
3.8	$\mathbf{k} \cdot \mathbf{p}$ Perturbation Theory	48
3.8.1	Expansion of the Band Near $\mathbf{k}_0$	48
3.9	Degenerate Points and Dirac Cones	49
3.9.1	$\mathbf{k} \cdot \mathbf{p}$ Theory for Twofold Degeneracies	49
3.9.2	Conditions for a Dirac Cone	49
3.10	The Light Cone and Radiative Modes	52
	Chapter Summary	52
4	Berry Geometry and Topological Invariants	54
	Why Band Theory Is Not Yet Topology?	54
4.1	Adiabatic Evolution and the Berry Phase	54
4.1.1	Parameter Space and the Adiabatic Assumption	54
4.2	Berry Connection and Berry Curvature	55
4.2.1	Berry Connection	55
4.2.2	Berry Curvature	55
4.2.3	Sum Rule for the Berry Curvature	56
4.2.4	Geometric Intuition and a Two-Level Model	56
4.3	Kubo Formula for the Berry Curvature	57
4.4	The Chern Number	58
4.4.1	Definition and Integrality	58
4.4.2	Deformation, Gap Closing, and $\mathcal{T}$	59
4.5	The Bulk–Edge Correspondence Principle	59
4.5.1	Statement	59
4.5.2	Jackiw–Rebbi Mechanism	59
4.6	Multi-Band Chern Numbers and the Wilson Loop	60
4.6.1	Non-Abelian Berry Geometry	60
4.6.2	Wilson Loop	61
4.7	Discrete Numerical Algorithms	61
4.7.1	Link Variables	61
4.7.2	The Fukui–Hatsugai–Suzuki Method	62
4.7.3	Discrete Wilson Loop	62
4.8	The $\mathbb{Z}_2$ Invariant	63
4.9	Introduction to HOTI	63
	Chapter Summary	64

5	Photonic Chern Insulators and the Quantum Hall Effect	65
	Why Topology Must Become Chiral Transport?	65
5.1	Gyrotropic Media and Magneto-Optical Response	65
5.1.1	Gyrotropic Tensors	66
5.1.2	Microscopic Origin and Frequency Limitation	67
5.2	Band Structure of the 2D Photonic Crystal	67
5.2.1	Wang Model and Polarization Reduction	67
5.2.2	Bloch Form and the Wang Bandgap	68
5.2.3	Plane-Wave Matrix Construction	68
5.3	Computation and Physical Meaning of the Chern Number	69
5.3.1	FHS Computation	69
5.3.2	$\mathcal{T}$, Berry Curvature, and Gap Closing	70
5.4	Unidirectional Edge States and Bulk–Edge Correspondence	70
5.4.1	Edge Count	70
5.4.2	Supercell Dispersion and Localization	71
5.4.3	Robustness of One-Way Propagation	71
5.5	Wang Experiment and Physical Interpretation	71
5.6	Beyond the Wang Model	72
	Chapter Summary	72
6	Photonic QSHE and $\mathbb{Z}_2$ Topology	74
	Why Topology Must Be Rebuilt Without Kramers Theorem?	74
6.1	Bosonic $\mathcal{T}$ and the Absence of Kramers’ Theorem	74
6.1.1	Time Reversal for Fermions and Bosons	74
6.2	Crystalline Symmetry and Pseudo-Spin Construction	75
6.2.1	C_{6v} Degeneracies at Γ	75
6.2.2	Shrinkage, Expansion, and Band Inversion	75
6.2.3	Pseudo-Spin Basis	76
6.3	Pseudo-Time-Reversal Symmetry and the Effective BHZ Model	76
6.3.1	$\mathbf{k} \cdot \mathbf{p}$ Hamiltonian Near Γ	76
6.3.2	Pseudo- $\mathcal{T}$	77
6.4	The Pseudo- $\mathbb{Z}_2$ Invariant	78
6.4.1	Pseudo-Spin Chern Number	78
6.5	Helical Edge States	78
6.5.1	Bulk–Edge Correspondence	78
6.5.2	Analytical Edge-State Solution	79
6.6	Backscattering Protection and Its Limitations	80
6.6.1	Selection Rule	80
6.6.2	Quantitative Scaling	80
6.7	Experimental Realization of the Wu–Hu Model	81

Chapter Summary	81
II Synthetic, Driven, and Open Photonic Topology	82
7 Coupled Resonators	83
Why Topology Can Be Engineered Through Coupling?	83
7.1 From Continuous Media to the Tight-Binding Model	84
7.1.1 Physical Motivation for Discretization	84
7.1.2 Peierls Substitution and Synthetic Flux	84
7.1.3 Coupled-Mode Derivation from Maxwell's Equations	85
7.2 The Hofstadter Model and Landau-Gauge Flux	85
7.2.1 Uniform Flux on a Square Lattice	86
7.2.2 Hofstadter Butterfly and TKNN Labels	86
7.3 Microring Resonator Arrays	87
7.3.1 Ring Resonators as Lattice Sites	87
7.3.2 The Hafezi Scheme	87
7.3.3 Effective Hopping from Link Rings	88
7.3.4 CW/CCW Pseudo-Spin	89
7.4 Synthetic Dimensions	89
7.4.1 Internal Degrees of Freedom as Lattice Sites	89
7.4.2 Frequency-Space Coupled-Mode Equations	90
7.4.3 Thouless Pumping	90
7.4.4 Beyond Frequency Space	91
Chapter Summary	91
8 Floquet Topology	93
Why Topology Need Not Be Static?	93
8.1 Foundations of Floquet Theory	94
8.1.1 The Floquet Theorem	94
8.1.2 Periodicity of Quasi-Energy and New Bandgaps	95
8.2 The Effective Floquet Hamiltonian and Winding Number	95
8.2.1 Floquet–Bloch Correspondence and Physical Intuition	95
8.2.2 The Effective Hamiltonian	96
8.2.3 Chern Numbers Are Not Enough: The Discovery of the Anomalous Phase	96
8.2.4 Definition of the Winding Number	96
8.3 Minimal Driven Models and the Anomalous Phase	97
8.4 Paraxial Optics as Floquet Evolution	98
8.5 Helical Waveguide Floquet Topological Insulators	99
8.5.1 Straight Waveguide Arrays and Graphene	99
8.5.2 Helical Waveguides and the Effective Vector Potential	100

8.5.3	Effective Peierls Phase and Topological Gap	101
8.6	Floquet Topological Phase Diagram	103
8.7	Other Floquet Schemes in Optics	104
	Chapter Summary	105
9	Non-Hermitian Topological Photonics	106
	Why Topology Must Survive Openness?	106
9.1	Open Maxwell Systems	107
9.1.1	Biorthogonal Basis and Left/Right Eigenvectors	108
9.1.2	Non-Hermitian Effective Inner Product	108
9.2	Exceptional Points and Their Topological Charge	109
9.2.1	Canonical Model of a Second-Order EP	109
9.2.2	Topological Charge of the EP	110
9.3	PT Symmetry	110
9.3.1	Definition of Photonic PT Symmetry	110
9.3.2	PT-Symmetric Photonic Implementations	111
9.4	Non-Hermitian Topological Classification	111
9.4.1	Redefinition of Gaps	112
9.4.2	Non-Hermitian Symmetry Classification	112
9.4.3	Non-Hermitian Chern Number	112
9.5	NHSE, GBZ, and Non-Bloch Bulk–Edge Correspondence	112
9.5.1	From Bloch Bulk States to Skin Modes	113
9.5.2	Generalized Brillouin Zone	113
9.5.3	The Hatano–Nelson Model	113
9.5.4	Non-Bloch Bulk–Edge Correspondence	114
9.6	Non-Hermitian Photonic Systems	115
9.7	Quasinormal Modes and Non-Hermitian Topology	115
	Chapter Summary	115
10	BICs and Topological Radiation	117
	Why Topology Must Confront the Radiation Continuum?	117
10.1	Basic Concepts of BICs	117
10.1.1	Statement of the Problem	117
10.1.2	History of BICs	118
10.2	Classification and Protection Mechanisms of BICs	118
10.2.1	Symmetry-Protected BICs	118
10.2.2	Topologically Protected BICs: Vortices of the Radiation Polarization Field	119
10.2.3	Explicit Computation of the BIC Topological Charge	120
10.2.4	Friedrich–Wintgen BIC	121
10.3	Quasi-BICs and High- Q Resonances	122
10.3.1	Experimental Characterization of BICs	123

10.4 Connection of BICs to Topological Photonics	125
10.4.1 Polarization Vortices and Berry Phase	125
III Valley, Three-Dimensional, and Nonlinear Topology	126
11 Valley Topological Photonic Crystals	127
Why Topology Can Be Local in Momentum Space?	127
11.1 Dirac Cones in Honeycomb Photonic Crystals	128
11.1.1 Geometry and Symmetry of the Honeycomb Lattice	128
11.1.2 C_{3v} Little Group and Dirac Degeneracy	128
11.1.3 Valley Hall Effect and Valley Hall Conductivity	129
11.1.4 Frequency-Domain Eigenequation for TM Modes	130
11.1.5 Physical Picture of the Dirac Cone Dispersion	130
11.2 Honeycomb Tight-Binding Model	131
11.3 $\mathbf{k} \cdot \mathbf{p}$ Theory and Effective Dirac Hamiltonian	131
11.3.1 Degenerate Perturbation Expansion	132
11.3.2 Symmetry Constraints and the Dirac Form	132
11.3.3 Second-Order $\mathbf{k} \cdot \mathbf{p}$ Corrections and Trigonal Warping	133
11.4 Valley Chern Number	133
11.4.1 Sublattice Inequivalence and the Mass Term	133
11.4.2 Berry Curvature of the Massive Dirac Fermion	134
11.4.3 Definition and Computation of the Valley Chern Number	135
11.5 Rigorous Theory of Domain-Wall Interface States	136
11.5.1 The Jackiw–Rebbi Mechanism	136
11.5.2 Interface States from the K and K' Valleys	137
11.5.3 Exact Solution for a Continuous Mass Profile	137
11.5.4 Domain-Wall Types and Differences in Topological Protection	137
11.6 Numerical Verification: The Supercell Method	138
11.7 120° Bend Transmission and Experimental Verification	138
11.8 Analysis of the Limitations of Valley Topological Protection	138
12 Three-Dimensional Topological Photonics	140
Why Topology Changes in Three Dimensions?	140
12.1 The Maxwell Eigenvalue Problem in 3D Photonic Crystals	140
12.1.1 Three-Dimensional Bloch Eigenequation	140
12.1.2 Polarization Degrees of Freedom	141
12.1.3 Symmetry and Classification of Band Degeneracies	141
12.2 Weyl Photonic Crystals	141
12.2.1 $\mathbf{k} \cdot \mathbf{p}$ Theory of Weyl Points	141
12.2.2 Topological Charge and Chirality	142

12.2.3	Realizing Weyl Points in Photonic Crystals	143
12.3	Fermi Arcs and Surface States	144
12.3.1	Bulk–Edge Correspondence and the Chirality of Weyl Points	144
12.3.2	Proof of Surface State Existence	145
12.4	Dirac Points and Nodal-Line Photonic Crystals	145
12.4.1	Symmetry-Protected Dirac Points	145
12.4.2	Nodal-Line Photonic Crystals	146
12.5	Numerical and Experimental Signatures	146
13	Nonlinear Topological Photonics	148
	Why Topology Must Go Beyond Linearity?	148
13.1	General Framework of the Nonlinear Maxwell Equations	148
13.1.1	Maxwell’s Equations with Nonlinear Polarization	148
13.1.2	Frequency-Domain Equations and Harmonic Balance	149
13.2	Weak Nonlinear Expansion	150
13.2.1	Slowly Varying Envelope Approximation	150
13.2.2	Nonlinear Corrections to Berry Curvature	150
13.3	Self-Consistent Theory of Nonlinear Bloch Waves	151
13.3.1	Definition of the Nonlinear Eigenvalue Problem	151
13.3.2	Power-Dependent Band Structure	152
13.3.3	Nonlinear Berry Phase	152
13.4	Mathematical Structure of Nonlinear Topological Theory	152
13.4.1	Energy Functional of the Nonlinear System	153
13.4.2	Bogoliubov–de Gennes Stability Analysis	153
13.4.3	Symmetry of the BdG Equation	153
13.4.4	Topological Classification of Nonlinear Systems	153
13.5	Topological Solitons and Nonlinear Edge States	154
13.6	Nonlinearity-Induced Topological Phase Transitions	155
13.6.1	Power-Driven Bandgap Closure and Chern Number Jump	155
13.7	Extension Directions	156
A	Differential Forms and Electromagnetism	158
	Why Differential Forms Are Worth Using Here?	158
A.1	Foundations of Exterior Algebra	159
A.1.1	Definition of Differential Forms	159
A.1.2	Wedge Product and Exterior Derivative	159
A.1.3	Hodge Duality	160
A.2	Maxwell’s Equations in Exterior Form	162
A.2.1	Electromagnetic Field as a Differential Form	162
A.2.2	Maxwell’s Equations in Differential Forms	162
A.2.3	Constitutive Relations	163

A.3	Connections and Curvature in Fiber-Bundle Language	163
A.3.1	Principal Bundles and Connections	163
A.3.2	The Chern Number as a Characteristic Class	164
A.3.3	Non-Abelian Generalization	164
A.4	Stokes' Theorem and Topological Quantization	164
A.4.1	First Chern Number on a Magnetic Monopole Bundle: Step-by-Step Computation	165
A.5	Further Applications	166
A.5.1	Maxwell's Equations in Differential-Form Notation	167
A.5.2	Hodge Duality and Maxwell's Equations in Media	167
A.5.3	Chern–Simons Forms and Topological Invariants	167
A.5.4	Fiber-Bundle Theory and Photonic Bands	168
B	Point and Space Groups	169
	Why Symmetry Must Be Made Explicit?	169
B.1	Basic Concepts of Groups	170
B.2	Group Representation Theory	170
B.2.1	Basic Definitions	170
B.2.2	Characters	170
B.3	Point Groups Commonly Used in Photonic Crystals	171
B.3.1	C_{nv} Groups	171
B.3.2	Time Reversal and Magnetic Point Groups	172
B.4	Space Groups and the Symmetry of Bloch States	172
B.4.1	Space Groups	174
B.4.2	Compatibility Relations	174
B.5	Selection Rules and Radiative Coupling	175
B.6	Character Tables and Applications	177
B.6.1	Character Table of the C_{6v} Point Group	177
B.6.2	C_{3v} Point Group and K -Point Symmetry	177
B.6.3	C_{4v} Point Group and the Square Lattice	178
B.6.4	Selection Rules and Coupling Matrix Elements	178
B.6.5	Time-Reversal Symmetry and Kramers' Theorem	178
C	Numerical Algorithm Implementation	180
	Why Topological Photonics Needs Reproducible Algorithms?	180
C.1	Plane-Wave Expansion Method (PWE)	181
C.1.1	Problem Setup	181
C.1.2	Fourier Expansion	181
C.1.3	PWE Algorithm Pseudocode	181
C.1.4	Complete Numerical Workflow for 1D Photonic Crystal PWE	182
C.2	FHS Lattice Berry Curvature	184

C.2.1	FHS Algorithm Pseudocode	184
C.2.2	FHS Algorithm Debugging Checklist	184
C.3	Wilson Loop Method	186
C.3.1	Wilson Loop Algorithm Pseudocode	186
C.4	Supercell Method and Edge States	187
C.4.1	Algorithm Concept	187
C.4.2	Supercell Edge State Algorithm Pseudocode	187
C.5	Transfer Matrix Method (1D)	188
C.5.1	Single-Layer Transfer Matrix	188
C.5.2	Transfer Matrix of One Period	188
C.5.3	Bloch Condition and Dispersion Relation	188
C.5.4	Zak Phase via the Transfer Matrix	188
C.6	Programming Recommendations	188
C.6.1	Recommended Tools	188
C.6.2	Debugging and Verification Checklist	189
C.7	Numerical Methods	189
C.7.1	Self-Consistent Iteration Algorithm for Nonlinear Bloch Waves	190
C.7.2	Numerical Solution of the Bogoliubov–de Gennes Equation	190
C.7.3	Implementation Details of the Split-Step Fourier Method	191
C.7.4	Imaginary-Time Evolution for Topological Edge Solitons	191
C.7.5	Supercell Method for 3D Topological Photonic Crystals	192
C.7.6	Efficient Numerical Computation of Berry Curvature	192
C.8	Implementation Details	193
C.8.1	Stabilization of the Discrete Berry Curvature Algorithm	193
C.8.2	Subspace Projection for Nearly Degenerate Bands	193
C.8.3	Numerical Errors and Grid Convergence	194
C.8.4	Pseudocode Framework	194
C.9	Unified Implementation Framework	194
C.10	Error Estimation and Convergence	197
C.10.1	Triple Decomposition of Discretization Error	197
C.10.2	Residuals and A Posteriori Error Estimates for Eigenvalue Problems	198
C.10.3	Why Preconditioners Are Especially Important in Maxwell Problems	198
C.10.4	Polar Decomposition Error Control in Wilson Loops	199
C.10.5	Spurious Edge State Criteria in Supercell Calculations	200
C.10.6	Comparison of PWE, FEM, and FDTD Error Sources in Topological Calculations	200
C.10.7	Executable Unified Diagnostic Workflow	201
C.11	Numerical Tools for 3D Topology	202
C.11.1	From Block-Tridiagonal Discrete Maxwell Operator to Surface Green’s Function	202

C.11.2 Schur Complement and Self-Energy Fixed-Point Equation	203
C.11.3 Layer Decimation and Sancho–Rubio-Type Doubling Iteration	203
C.11.4 Discrete Flux Integration of Weyl Point Topological Charge	204
C.12 Wilson-Loop Error Control	205
C.12.1 Discrete Wilson Loop and the Continuous Limit	205
C.12.2 Nearly Degenerate Subspaces and Parallel Transport Gauge	206
C.12.3 Comparison of Polar Decomposition, QR, and SVD Correction	207
C.12.4 Credible Surface State Criteria in Supercell Calculations	208
C.12.5 From “Program Output” to “Textbook Figures”: Selection Criteria	209
D Supplement on Maxwell Spectral Theory	210
D.1 Green’s Function of the Maxwell Operator	210
D.2 Local Density of States and the Purcell Effect	210
D.3 Variational Principle and the Rayleigh Quotient	211
D.4 The Courant–Fischer Principle and Bandgap Sensitivity	211

Part I

Maxwell Foundations and Two-Dimensional Topology

Chapter 1

Introduction

Foundational Clarifications

The sole purpose of this chapter is not to “finish the history,” but to establish an **actionable coordinate system for learning**: Where must you trust Maxwell’s equations rather than analogies? Where is “topology” merely an elegant metaphor, and where does it become a verifiable theorem?

To prevent the reader from developing incorrect intuitions from the outset, this book deliberately places the following three warnings in the present chapter:

1. **The inner product is not a decoration.** If the inner product is chosen incorrectly (e.g., using the unweighted L^2 inner product), the gauge structure of the Berry connection is destroyed, and the Chern number may no longer even be an integer.
2. **TRS and reciprocity are two distinct propositions.** Reciprocity constrains the scattering matrix (port-exchange symmetry), whereas time-reversal symmetry (TRS) constrains the field transformation under time reversal; the two become separable in the presence of magneto-optical or time-varying media.
3. **Photonic devices are inherently open systems.** Radiative leakage means that eigenfrequencies can become complex, causing Hermitian spectral theory to break down directly; topological protection does not “automatically hold”—one must specify its domain of validity.

1.1 Why Do We Need “Yet Another” Topological Photonics Textbook?

Since Haldane and Raghu proposed the theoretical prediction of topological edge states in the context of photonic crystals in 2005 [1, 2], topological photonics has evolved from a purely theoretical concept into an experimental discipline spanning the microwave, terahertz, infrared, and visible frequency bands. As of 2026, the field has produced at least three landmark reviews [3, 4, 5], hundreds of original research articles, and several excellent conference tutorials. However, these resources share a common limitation in their narrative strategy: they almost universally introduce topological concepts starting from **tight-binding models of condensed**

matter physics, relying on the “photon–electron analogy” metaphor, and only subsequently discuss the peculiarities of photonic systems.

This “analogy first, correction later” narrative is efficient in research reviews, but it leaves three systematic deficiencies in a pedagogical context:

1. **The inner product problem is concealed.** In electronic systems, $\langle\psi|\phi\rangle = \int \psi^* \phi \, d^3r$ is the natural L^2 inner product, and the definition of the Berry connection requires no further discussion. In photonic systems, the correct inner product is the material-weighted energy inner product $\langle\mathbf{f}_1, \mathbf{f}_2\rangle_M$, whose positive definiteness depends on the lossless condition and whose form depends on the constitutive tensor $\bar{M}$. Ignoring this point leaves all subsequent definitions of topological quantities suspended in mid-air.
2. **Symmetry constraints are misplaced.** In electronic systems, TRS guarantees spin degeneracy via Kramers’ theorem, and the $\mathbb{Z}_2$ classification follows naturally. In photonic systems, TRS is an anti-unitary operator whose specific form depends on the magneto-optical response of the material; meanwhile, reciprocity (Lorentz reciprocity) is a symmetry proposition independent of TRS that constrains the scattering matrix rather than the Berry curvature. Conflating the two is one of the most common errors in the literature.
3. **Open systems are neglected.** The condensed matter analogy implicitly assumes closed systems (Hermitian Hamiltonian, real eigenvalues), but real photonic devices almost always involve radiative leakage, material absorption, or gain. Non-Hermitian topology is not an “advanced extension”—it is the *everyday reality* of photonics experiments.

The guiding axiom of this book is **Maxwell-first**: starting from the generalized eigenvalue problem of the frequency-domain Maxwell equations, $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$, establishing the self-adjointness of the operator under the material-weighted inner product, and then introducing Bloch parameterization, Berry geometry, and topological invariants. The tight-binding model from condensed matter is not the logical starting point, but rather an *effective approximation* valid under specific conditions.

1.2 A Brief History of Photonic Topological States

1.2.1 Intellectual Origins: From the Quantum Hall Effect to Photonic Analogies

In 1980, von Klitzing discovered the quantum Hall effect (QHE) in a two-dimensional electron gas, where the Hall conductance exhibits precisely quantized plateaus. In 1982, Thouless, Kohmoto, Nightingale, and den Nijs (TKNN) proved that the quantized Hall conductance equals the sum of the Chern numbers of the occupied energy bands—a topological invariant that is robust against perturbations and disorder. This discovery profoundly revealed the hidden geometric structure in band theory: the Berry phase and Berry curvature.

The core insight enabling the generalization of topological concepts from electronic systems to bosonic systems (photons, phonons) is that the definition of topological invariants requires only a *band structure* and an *inner product*; it does not depend on particle statistics (fermions or bosons) or on the existence of a Fermi level. The key physical prerequisites are:

- The system possesses well-defined frequency bands with **complete bandgaps** between them;
- A positive-definite inner product exists on the eigenstate space, so that the Berry connection and curvature are well defined.

1.2.2 Foundational Papers: Raghu–Haldane and the Wang Experiment

Between 2005 and 2008, Raghu and Haldane explicitly proposed that in two-dimensional photonic crystals, if TRS is broken through the magneto-optical effect (gyro-anisotropy under an applied static magnetic field), the photonic bands can acquire nonzero Chern numbers, and the boundaries of the photonic crystal will support **unidirectionally propagating** chiral edge states [1, 2]. The key steps of this prediction are:

1. Write Maxwell’s equations as a generalized eigenvalue problem $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$;
2. Verify the self-adjointness of the operator under the material-weighted inner product (lossless condition);
3. Use Bloch’s theorem to parameterize the problem as a family of eigenvalue problems in $\mathbf{k}$ -space;
4. Compute the Chern number $\mathcal{C}_n$ of each band over the Brillouin zone.

In 2009, Wang et al. provided the first experimental verification of unidirectional propagation of photonic topological edge states in a two-dimensional photonic crystal made of YIG (yttrium iron garnet) at microwave frequencies (GHz), demonstrating the absence of backscattering even in the presence of metallic scatterers [6]. This experiment is regarded as the classic verification of the “photonic quantum Hall effect.”

1.2.3 From Broken TRS to Preserved TRS: The Photonic Spin Hall Effect

Although the microwave magneto-optical scheme is conceptually clean, it is difficult to extend to optical frequencies—the magneto-optical response is extremely weak at optical frequencies. Between 2013 and 2015, two complementary approaches were proposed:

- **Bianisotropic metamaterials** (Khanikaev et al., 2013 [7]): Through carefully designed metamaterial unit cells, an effective spin–orbit coupling is constructed to realize a photonic analogue of the quantum spin Hall effect (QSHE).
- **Crystalline symmetry protection** (Wu & Hu, 2015 [8]): In a dielectric photonic crystal with C_{6v} symmetry, the Dirac point is moved to the Γ point via “zone folding,” and a bandgap is opened by expanding or shrinking the pillar diameter, realizing pseudo-spin topological states without any magnetic materials.

The common feature of both approaches is that under preserved TRS, the total Chern number is zero, but one can define a *pseudo-spin Chern number* or a $\mathbb{Z}_2$ invariant to characterize the topological phase. The edge states are not strictly unidirectional but rather “spin-locked” (pseudo-spin up/down propagate in opposite directions).

1.2.4 Three-Dimensional Topological Photonics: Weyl Points and Nodal Lines

The band structure of three-dimensional photonic crystals can support Weyl points—linear crossings of two bands in the Brillouin zone with topological charges (monopole charges) of ± 1 . Lu et al. (predicted in 2013 [9], experimentally realized in 2015 [10]) observed Weyl points and their hallmark surface states—Fermi arcs—in a double-helix photonic crystal. The richness of three-dimensional topological photonic crystals lies in the fact that, compared to two dimensions, three-dimensional systems support a greater variety of band crossings (Weyl, Dirac, nodal lines/surfaces) and more complex symmetry-protection mechanisms.

1.2.5 Non-Hermitian Topology and Topological Lasers

The gain and loss inherent to real photonic systems introduce essential non-Hermiticity. In 2018, Bandres, Harari, et al. realized a topological laser in a microring resonator array [11, 12]—the lasing mode is localized on the edge state and inherits the robustness of topological protection. The theoretical framework of non-Hermitian topology requires a re-examination of the relationships among the inner product, self-adjointness, and the spectrum, introducing new concepts such as exceptional points, the non-Hermitian skin effect, and others [13].

1.2.6 Floquet Topology and Synthetic Dimensions

Through periodic temporal modulation (Floquet engineering), effective topological bandgaps can be induced without altering the static material [14]. Synthetic dimensions exploit the frequency, orbital angular momentum, or spatial mode index of light as additional “dimensions,” enabling the simulation of higher-dimensional topological physics in physically low-dimensional systems [15].

1.3 Overview of the Book

Below we outline the role of each chapter, providing the reader with a bird’s-eye-view knowledge map.

Chapter 2: The Maxwell Operator and Inner Product

The book begins from the six-component Maxwell generalized eigenvalue problem $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$. This chapter defines the material-weighted energy inner product, proves the self-adjointness theorem in the lossless closed setting, and separates losslessness, reciprocity, and time-reversal symmetry as distinct material conditions. It also explains why dispersive media require a modified energy metric or an auxiliary-field enlargement.

Chapter 3: Bloch Band Theory

Periodic materials turn the Maxwell eigenproblem into a family of cell-periodic problems indexed by $\mathbf{k}$ in the Brillouin zone. This chapter derives the photonic Bloch theorem, the cell-periodic Maxwell operator $\hat{N}(\mathbf{k})$, the meaning of photonic bands and bandgaps, and the basic $\mathbf{k} \cdot \mathbf{p}$ theory used to describe Dirac cones and band inversions.

Chapter 4: Berry Geometry and Topological Invariants

Berry connection, Berry curvature, Chern number, Wilson loops, and discrete FHS algorithms are introduced using the material-weighted inner product. The chapter also explains why Chern numbers are integer-valued, how gap closings change topology, and how bulk–edge correspondence emerges from the same geometric data.

Chapter 5: Photonic Chern Insulators—Magneto-Optical Effects and the Quantum Hall Analogy

This chapter treats the most direct photonic analogue of the quantum Hall effect: a gyrotropic material breaks TRS, opens a topological bandgap, and supports chiral edge states. The focus is on the Maxwell/PWE formulation of magneto-optical photonic crystals, the role of the gyrotropic tensor, and the microwave experiments that established photonic Chern transport.

Chapter 6: Photonic Quantum Spin Hall Effect—Pseudo-Spin and Pseudo-Kramers Pairs

Ordinary photons have $\mathcal{T}^2 = +1$, so photonic QSHE must be engineered rather than inherited from Kramers theorem. This chapter explains pseudo-spin construction, band inversion in crystalline photonic structures, effective BHZ models, helical edge states, and the symmetry conditions under which pseudo-spin protection is meaningful.

Chapter 7: Coupled Resonators and Synthetic Gauge Fields

Coupled-mode theory connects Maxwell modes to tight-binding-like resonator models. This chapter develops synthetic Peierls phases, effective magnetic flux, Hofstadter physics in resonator arrays, frequency synthetic dimensions, and Thouless pumping.

Chapter 8: Floquet Topological Photonics

Periodic driving produces quasi-energy bands and topology that has no static analogue. This chapter introduces Floquet operators, 0- and π -gaps, anomalous winding phases, and the paraxial waveguide realization of driven topological transport.

Chapter 9: Non-Hermitian Topological Photonics

Gain, loss, radiation, and asymmetric coupling force photonic topology beyond Hermitian band theory. This chapter introduces biorthogonal eigenvectors, exceptional points, non-Bloch topology, skin effects, topological lasers, and the distinction between point-gap and line-gap classifications.

Chapter 10: Bound States in the Continuum and Topological Radiation

This chapter turns from closed photonic band theory to open photonic systems, where modes above the light cone normally radiate into the continuum and acquire a finite linewidth. It introduces bound states in the continuum as exceptional non-radiating modes embedded inside the radiation continuum, and classifies their protection mechanisms into symmetry-protected BICs, topologically protected BICs, and Friedrich–Wintgen BICs arising from destructive interference between leaky modes. The key topological object is the far-field polarization vortex in momentum space, whose winding number defines the BIC topological charge and controls the robustness,

motion, and pair annihilation of BICs. The chapter then explains how quasi-BICs produce experimentally observable high- Q Fano resonances, how the scaling law $Q \propto (\delta k)^{-2|q|}$ links topological charge to device performance, and why BIC physics provides a natural bridge between topological band theory, non-Hermitian radiation loss, sensing, lasing, and nonlinear photonics.

Chapter 11: Valley Topological Photonic Crystals

This chapter develops valley topological photonic crystals as a symmetry-based and experimentally accessible route to topological waveguiding. Starting from honeycomb photonic crystals, it explains how Dirac cones arise at the inequivalent valleys K and K' , and how breaking spatial inversion symmetry opens a valley bandgap described by an effective massive Dirac Hamiltonian. The Berry curvature, valley Chern number, and Jackiw–Rebbi domain-wall mechanism are then introduced to explain the emergence of valley-polarized interface states. A central message of the chapter is that valley protection is not an absolute topological protection: it relies on the suppression of intervalley scattering and therefore depends on interface orientation, disorder length scale, and symmetry preservation.

Chapter 12: Three-Dimensional Topological Photonics

This chapter extends the band-topological framework from two-dimensional photonic crystals to fully three-dimensional Maxwell systems. It first formulates the three-dimensional Bloch eigenvalue problem for vector electromagnetic fields, emphasizing the additional role of polarization, gauge choice, and crystalline symmetries. The main focus is on Weyl points, Dirac points, nodal lines, and nodal surfaces, whose topology is characterized by Berry flux, topological charge, Wilson loops, and symmetry constraints. The chapter also explains the bulk–surface correspondence in three dimensions through photonic surface states and Fermi arcs, highlighting how these objects are observed in frequency-resolved and surface-sensitive electromagnetic measurements rather than through an electronic Fermi level.

Chapter 13: Nonlinear Topological Photonics

This chapter discusses what happens when the linear eigenvalue structure underlying most topological band theory is modified by optical nonlinearity. Beginning with Maxwell’s equations including nonlinear polarization, it introduces perturbative corrections, harmonic-balance formulations, and self-consistent nonlinear Bloch modes. The chapter then develops the physics of power-dependent band structures, nonlinear Berry phases, topological solitons, nonlinear edge states, and nonlinearity-induced topological phase transitions. Special care is taken to distinguish different meanings of topology in nonlinear systems: topology may be assigned to stationary nonlinear solutions, to linearized Bogoliubov–de Gennes spectra, or to self-consistent effective bands, but it cannot be naively imported from linear Hermitian band theory without specifying the underlying state and stability problem.

Four Appendices

Appendix A restates Maxwell’s equations and Chern classes in the language of differential forms and fiber bundles, providing a solid differential-geometric foundation for Berry geometry (step-by-step computation of the Chern number $C_1 = n$ for the magnetic monopole bundle on S^2). Appendix B provides tools for point group and space group analysis of photonic crystals—derivation of character tables from scratch (complete demonstration for C_{4v}), compatibility

relations, and Wigner–Eckart selection rules. Appendix C provides directly reproducible numerical algorithm pseudocode: a complete one-dimensional PWE Python workflow (including the inverse rule and convergence tests), the FHS discrete Berry curvature algorithm (with an 8-item debugging checklist), and the Wilson loop algorithm (with Wannier band plotting). Appendix D gives a concise introduction to spectral theory for the Maxwell operator, covering essential functional-analytic background on self-adjoint extensions, spectral decomposition, and the variational characterization of eigenvalues.

1.4 Comparison of Five Representative Topological Photonic Schemes

Before delving into the technical chapters, it is useful to compare several representative schemes from a macroscopic perspective. Table 1.1 does not exhaust the book: BIC topology, valley topology, three-dimensional topological photonics, and nonlinear topology are treated as additional main threads in later chapters.

Complementarity and Convergence Trends Among Schemes

As can be seen from Table 1.1, no single scheme dominates across all evaluation dimensions. The **magneto-optical scheme** (Chapter 5) provides truly unidirectional transport and the strongest robustness, but is limited by the weakness of magneto-optical response at optical frequencies; its clearest realizations remain in the microwave band. The **pseudo-spin scheme** (Chapter 6) elegantly circumvents the absence of Kramers’ theorem, realizing helical edge states on dielectric platforms, but its protection relies on the integrity of C_{6v} symmetry—any defect that breaks this symmetry (e.g., a 90° bend rather than a 120° bend) leads to significant backscattering. The **coupled resonator scheme** (Chapter 7) synthesizes effective magnetic flux through path phase differences, offering the greatest design flexibility and pioneering the synthetic dimension paradigm—exploring higher-dimensional topology with only a one-dimensional physical structure. The **Floquet scheme** (Chapter 8) reveals the anomalous topological phase unique to dynamical systems ($C_{\text{eff}} = 0$ yet edge states exist), and in waveguide arrays one can image the formation of topological states cross-section by cross-section along z . The **non-Hermitian scheme** (Chapter 9) opens entirely new physical dimensions—exceptional points, the non-Hermitian skin effect, and the 38-fold non-Hermitian symmetry classification, phenomena that simply do not exist in Hermitian systems.

Current research frontiers are actively exploring *scheme convergence*:

- **Non-Hermitian + topology**: Introducing gain/loss into topological edge states to realize topological lasers (Chapter 9).
- **Floquet + synthetic dimensions**: Combining Floquet modulation with frequency synthetic dimensions in time-multiplexed systems to explore the 4D quantum Hall effect.
- **Nonlinear + topology**: Exploiting the Kerr effect to induce self-induced topological phase transitions and topological solitons in topological edge states (Chapter 13).
- **Pseudo-spin + BIC**: Using C_{6v} symmetry-protected BICs to realize high- Q topological resonance states, enhancing nonlinear optical effects.

Table 1.1 Comparison of five representative topological photonic schemes.

Property	Magneto-optical Chern insulator (Ch. 5)	Pseudo-spin QSHE (Ch. 6)	Coupled resonators (Ch. 7)	Floquet topology (Ch. 8)	Non-Hermitian topology (Ch. 9)
TRS status	Explicitly broken (external field)	Emergent pseudo-TRS	Design dependent; often synthetic gauge	Effectively broken (periodic driving)	Hermiticity broken; TRS depends on implementation
Topological invariant	Chern number $C \in \mathbb{Z}$	Pseudo-spin Chern number $C_{\pm}$	TKNN/Chern	Winding number W_{ϵ}	Non-Bloch Chern number; winding number W
Protection mechanism	Band topology (robust)	C_{6v} symmetry (conditional)	Synthetic gauge invariance	Quasi-energy band topology	Point gap / line gap
Edge-state type	Unidirectional (chiral)	Helical (pseudo-spin locked)	Unidirectional or helical	Chiral (incl. anomalous)	Skin states; chiral
Bandgap mechanism	Magneto-optical splitting, TRS breaking	Band inversion $M_0 B < 0$	Hofstadter flux	Quasi-energy 0-/ π -gap	Point/line gap in complex spectrum
Operating frequency	Most mature in microwave; optical response weak	Microwave $\rightarrow$ near-IR	Microwave $\rightarrow$ optical	Near-IR $\rightarrow$ visible	Microwave $\rightarrow$ optical
Experimental platform	YIG ferrite rod arrays	Dielectric PhC slabs	Microrings / microwave cavities	Femtosecond-written waveguide arrays	Gain/loss microrings; fiber loops
Core advantage	Truly unidirectional; robust	No magnetic field; CMOS compatible	Flexible; synthetic dimensions	Room temperature; z -evolution imaging	New physics (EP, NHSE)
Core limitation	Optical magneto-optical response is weak	Pseudo-TRS fragile; C_3 sensitive	Quality factor limited	Paraxial approximation	Biorthogonal inner product not positive definite
Representative experiment	Wang <i>et al.</i> , 2009	Wu/Hu, 2015	Hafezi <i>et al.</i> , 2013	Rechtsman <i>et al.</i> , 2013	Weidemann <i>et al.</i> , 2020

Remark (Engineering Guidelines for Scheme Selection). For practical applications, scheme selection should follow these priorities: (1) First determine the operating frequency band—this strongly constrains the magneto-optical scheme and often favors dielectric, resonator, or Floquet platforms at optical frequencies; (2) Next determine the required edge-state type—unidirectional transport requires TRS breaking (magneto-optical or Floquet), while helical transport can use the pseudo-spin or resonator scheme; (3) Finally consider fabrication compatibility—for CMOS compatibility, prefer the dielectric pseudo-spin or microring scheme. In all cases, one should verify the *quantitative* strength of topological protection at the target frequency and geometry (transmission loss, backscattering rate), rather than relying solely on qualitative topological arguments.

Table 1.2 Principal notation conventions used throughout the book.

Symbol	Meaning	Convention or definition
$\mathbf{E}, \mathbf{H}$	Electric and magnetic field complex amplitudes	$e^{-i\omega t}$ convention
$\mathbf{D}, \mathbf{B}$	Electric displacement and magnetic induction	Frequency-domain complex amplitudes
$\mathbf{f} = (\mathbf{E}, \mathbf{H})^\top$	Six-component field vector	Electromagnetic state vector
$\bar{M}(\mathbf{r}, \omega)$	6×6 material matrix	Contains $\boldsymbol{\varepsilon}, \boldsymbol{\mu}, \bar{\xi}, \bar{\zeta}$
$\hat{N}$	Maxwell curl operator	$\begin{pmatrix} 0 & i\nabla \times \\ -i\nabla \times & 0 \end{pmatrix}$
$\langle \cdot, \cdot \rangle_M$	Material-weighted inner product	$\int \mathbf{f}_1^\dagger \bar{M} \mathbf{f}_2 d^3r$
$\mathbf{k}$	Bloch wave vector	$\mathbf{k} \in \text{BZ}$
$u_{n\mathbf{k}}(\mathbf{r})$	Bloch-periodic part	$\mathbf{f}_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}$
$\mathcal{A}_n(\mathbf{k})$	Berry connection	$-i\langle u_{n\mathbf{k}}, \nabla_{\mathbf{k}} u_{n\mathbf{k}} \rangle_M$
$\mathcal{F}_n(\mathbf{k})$	Berry curvature	$\nabla_{\mathbf{k}} \times \mathcal{A}_n$
$\mathcal{C}_n$	Chern number	$(2\pi)^{-1} \int_{\text{BZ}} \mathcal{F}_n d^2k$
$\mathcal{W}$	Wilson loop operator	Path-ordered exponential
$\mathcal{T}$	Time-reversal operator	Anti-unitary
$\mathcal{P}$	Spatial inversion operator	Unitary

Remark. Throughout the book, SI units are used exclusively. The time-factor convention is $e^{-i\omega t}$ (engineering convention), which differs from the $e^{+i\omega t}$ convention used in some physics literature by a complex conjugation. The reader should be aware of this difference when cross-referencing other sources, as it affects the sign of the Berry connection.

Chapter 2

The Maxwell Operator and Energy Inner Product

Why a Maxwell-first Foundation is Necessary?

The logical foundation of topological photonics is the generalized eigenvalue problem of frequency-domain Maxwell equations, not an imported tight-binding Hamiltonian. This chapter establishes the three pieces that all later topological constructions use:

$$\hat{N}\mathbf{f} = \omega \bar{M}\mathbf{f}, \quad \langle \mathbf{f}, \mathbf{g} \rangle_M = \int \mathbf{f}^\dagger \bar{M} \mathbf{g} d^3r, \quad \text{symmetry constraints on } \bar{M}.$$

The first two determine the correct operator and inner product; the third separates reciprocity from time-reversal symmetry and identifies when Chern or pseudo-spin topology can appear.

Before writing down the frequency-domain Maxwell equations, it is useful to fix what the frequency representation assumes. Linearity makes different frequencies independent, while causality ties dispersion and loss through the Kramers–Kronig relations:

$$\text{Re}[\varepsilon(\omega)] - \varepsilon_\infty = \frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\omega' \text{Im}[\varepsilon(\omega')]}{\omega'^2 - \omega^2} d\omega'. \quad (2.1)$$

Thus a frequency-independent, lossless material matrix is an approximation. When dispersion or gain/loss is important, the linear Hermitian generalized eigenproblem must be modified; the last section gives the auxiliary-field construction for a Drude medium, and Chapter 9 treats genuinely non-Hermitian effects.

Remark (Global impact of the time-factor convention). This book adopts the $e^{-i\omega t}$ convention. If the $e^{+i\omega t}$ convention is used instead (as in some engineering references), then: (i) causality requires $\chi(\omega)$ to be analytic in the *lower* half-plane; (ii) loss corresponds to $\text{Im}[\varepsilon] < 0$ rather than > 0 ; (iii) the form of the TRS operator $\hat{\Theta}$ is unchanged, but the sign convention of the Berry connection is flipped. When reading different references, always confirm the time-factor convention first—this is one of the most common sources of sign confusion (see Pitfall 6 at the end of this chapter).

With the above physical picture in hand, we now formally write down the frequency-domain Maxwell equations.

2.1 From the Time Domain to the Frequency Domain: Complex Amplitude Conventions

2.1.1 Source-Free Maxwell Equations

In regions free of free charges and free currents, the time-domain Maxwell equations are

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}(\mathbf{r}, t)}{\partial t}, \quad (2.2a)$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t}, \quad (2.2b)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = 0, \quad (2.2c)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0. \quad (2.2d)$$

2.1.2 Monochromatic Steady State and Time-Factor Convention

For a monochromatic (single-frequency) steady state, we adopt the complex amplitude representation

$$\mathbf{E}(\mathbf{r}, t) = \text{Re}\{\mathbf{E}(\mathbf{r}) e^{-i\omega t}\}, \quad (2.3)$$

where $\omega > 0$ is the angular frequency and $\mathbf{E}(\mathbf{r})$ is the complex amplitude vector field. The same convention is adopted for $\mathbf{H}$, $\mathbf{D}$, and $\mathbf{B}$.

Remark. Throughout this book, we use the engineering time convention $e^{-i\omega t}$. Some physics literature instead adopts the $e^{+i\omega t}$ convention. Whenever this choice affects the sign of a formula, the corresponding change will be stated explicitly. With the $e^{-i\omega t}$ convention, a positive imaginary part of the permittivity, $\text{Im } \varepsilon > 0$, corresponds to loss.

Substituting eq. (2.3) into eq. (2.2) yields the frequency-domain equations

$$\nabla \times \mathbf{E}(\mathbf{r}) = i\omega \mathbf{B}(\mathbf{r}), \quad (2.4a)$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = -i\omega \mathbf{D}(\mathbf{r}), \quad (2.4b)$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = 0, \quad (2.4c)$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0. \quad (2.4d)$$

2.2 Linear Constitutive Relations and the Material Matrix

2.2.1 The Most General Linear Local Constitutive Relations

In the most general linear, local, possibly anisotropic medium that allows magneto-electric coupling (bianisotropy), the frequency-domain constitutive relations are

$$\begin{pmatrix} \mathbf{D} \\ \mathbf{B} \end{pmatrix} = \underbrace{\begin{pmatrix} \varepsilon(\mathbf{r}, \omega) & \bar{\xi}(\mathbf{r}, \omega) \\ \bar{\zeta}(\mathbf{r}, \omega) & \mu(\mathbf{r}, \omega) \end{pmatrix}}_{\bar{\mathbf{M}}(\mathbf{r}, \omega)} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}. \quad (2.5)$$

Here:

- ϵ : 3×3 permittivity tensor (SI units: F/m);
- μ : 3×3 permeability tensor (SI units: H/m);
- $\bar{\xi}, \bar{\zeta}$: 3×3 bianisotropy coupling tensors (SI units: s/m);
- $\bar{M}$: 6×6 material matrix, completely characterizing the linear electromagnetic response of the medium.

Definition 2.1 (Symmetry classification of the material matrix). Based on the matrix structure of $\bar{M}$, the following important classes of material symmetry can be distinguished:

1. **Isotropic**: $\epsilon = \epsilon \mathbf{I}_3$, $\mu = \mu \mathbf{I}_3$, $\bar{\xi} = \bar{\zeta} = 0$.
2. **Anisotropic (no magneto-electric coupling)**: $\bar{\xi} = \bar{\zeta} = 0$, but ϵ or μ is not proportional to $\mathbf{I}_3$.
3. **Gyro-anisotropic (magneto-optical)**: ϵ has antisymmetric imaginary components (gyromagnetic parameter g), breaking TRS.
4. **Bianisotropic**: $\bar{\xi}, \bar{\zeta} \neq 0$. Chiral media ($\bar{\xi} = i\kappa \mathbf{I}_3$, $\bar{\zeta} = -i\kappa \mathbf{I}_3$) are the most common subclass.

2.2.2 The Lossless Condition

Definition 2.2 (Lossless medium). If and only if $\bar{M} = \bar{M}^\dagger$ (in the sense of frequency-domain complex amplitudes), the interaction between the electromagnetic field and the medium produces no net energy dissipation or gain. The medium is then called **lossless**.

Specifically, the lossless condition implies

$$\epsilon = \epsilon^\dagger, \quad \mu = \mu^\dagger, \quad \bar{\xi} = \bar{\zeta}^\dagger. \quad (2.6)$$

Remark. For real materials, the lossless condition holds only approximately in frequency ranges far from resonance where intrinsic losses can be neglected. In dispersive media, $\bar{M}(\omega)$ satisfies the Kramers–Kronig relations, and strict losslessness over the entire frequency range is impossible. This book first adopts the lossless, non-dispersive limit when establishing the formal framework, and then discusses the consequences of relaxing these conditions in later chapters (Chapter 8, Appendix A).

Physical discussion: Why is the material matrix the “metadata” of the entire theory? In condensed matter physics, the Hamiltonian H itself contains all the information about the system; in classical electromagnetism, the Maxwell curl operator $\hat{N}$ is the same for all materials—the information that distinguishes different systems is entirely encoded in the material matrix $\bar{M}$. This means that the values of topological invariants (such as the Chern number), symmetry protections (such as TRS), and qualitative spectral features (presence or absence of bandgaps) are determined by the Maxwell operator together with the matrix structure of $\bar{M}$. One may say that $\bar{M}$ plays the role of a metric on the electromagnetic state space.

This analogy is not merely rhetorical. In subsequent chapters we will see that the definition of the Berry connection *requires* an inner product, and the inner product is precisely given by $\bar{M}$. Therefore, if you write $\bar{M}$ incorrectly—for instance, by omitting the bianisotropic coupling terms $\bar{\xi}, \bar{\zeta}$, or by erroneously using a diagonal ϵ for a magneto-optical material—then the Berry

curvature and Chern numbers computed subsequently will be **physically meaningless**, even if the mathematical procedure itself reports no errors.

To help the reader build concrete physical intuition about the structure of $\bar{M}$, we present two important worked examples.

Example 2.3 (The Polder gyromagnetic model). Ferrites under an applied DC magnetic field $\mathbf{B}_0 = B_0 \hat{z}$ exhibit strong magneto-optical effects and are the key material for realizing photonic Chern insulators (Chapter 4). Their permeability tensor is given by the **Polder model**:

$$\boldsymbol{\mu} = \mu_0 \begin{pmatrix} \mu_r & -i\kappa & 0 \\ i\kappa & \mu_r & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.7)$$

where

$$\mu_r = 1 + \frac{\omega_m \omega_0}{\omega_0^2 - \omega^2}, \quad \kappa = \frac{\omega_m \omega}{\omega_0^2 - \omega^2}. \quad (2.8)$$

Here $\omega_0 = \gamma B_0$ is the precession frequency (γ is the gyromagnetic ratio) and $\omega_m = \gamma \mu_0 M_s$ is proportional to the saturation magnetization M_s .

Physical interpretation.

- The diagonal element μ_r describes the “ordinary” permeability in directions parallel and perpendicular to the applied field.
- The antisymmetric imaginary parts $\pm i\kappa$ (gyromagnetic parameters) describe the different responses of left- and right-circularly polarized components caused by magnetization precession. Physically, this is because the applied magnetic field breaks the degeneracy between left and right circular polarizations—analogueous to the Zeeman effect lifting spin degeneracy in electronic systems.
- When $\omega \rightarrow \omega_0$, both μ_r and κ diverge—this is ferromagnetic resonance (FMR), at which point losses cannot be neglected and the lossless approximation breaks down.

Symmetry verification.

1. **Losslessness:** $\boldsymbol{\mu}^\dagger = \boldsymbol{\mu}$ (since μ_r, κ are real numbers, and $(-i\kappa)^* = i\kappa$ is exactly the corresponding element after transposition). ✓
2. **Reciprocity:** $\boldsymbol{\mu}^\top \neq \boldsymbol{\mu}$ (since $(-i\kappa)^\top = -i\kappa \neq i\kappa$), so magneto-optical materials are **nonreciprocal**. ✗
3. **TRS:** $\boldsymbol{\mu}^* \neq \boldsymbol{\mu}$ (since $(-i\kappa)^* = i\kappa \neq -i\kappa$), so TRS is broken. ✗

Therefore, Polder ferrites simultaneously break reciprocity and TRS, which is precisely the material condition required for realizing nonzero Chern numbers. This is the core material used by Wang et al. in their landmark 2009 experiment [6].

Positive-definiteness condition. Positive definiteness of $\boldsymbol{\mu}$ requires all three eigenvalues to be positive. The eigenvalue in the z direction is μ_0 (always positive). The two eigenvalues in the

xy plane are $\mu_0(\mu_r \pm \kappa)$, so the positive-definiteness condition is

$$\mu_r + \kappa > 0 \quad \text{and} \quad \mu_r - \kappa > 0, \quad \Longleftrightarrow \quad |\kappa| < \mu_r. \quad (2.9)$$

Using eq. (2.8), one can verify that this condition is typically satisfied far from the FMR frequency ω_0 .

Example 2.4 (Chiral medium). Chiral media (such as solutions of helical molecules, chiral metamaterials) possess nonzero magneto-electric coupling, with constitutive relations

$$\mathbf{D} = \varepsilon \mathbf{E} + i\kappa_c \mathbf{H}, \quad \mathbf{B} = \mu \mathbf{H} - i\kappa_c \mathbf{E}, \quad (2.10)$$

where κ_c is the (real) chirality parameter. The corresponding 6×6 material matrix is

$$\bar{M} = \begin{pmatrix} \varepsilon \mathbf{I}_3 & i\kappa_c \mathbf{I}_3 \\ -i\kappa_c \mathbf{I}_3 & \mu \mathbf{I}_3 \end{pmatrix}. \quad (2.11)$$

Symmetry analysis.

1. **Losslessness:** $\bar{M}^\dagger = \begin{pmatrix} \varepsilon \mathbf{I}_3 & i\kappa_c \mathbf{I}_3 \\ -i\kappa_c \mathbf{I}_3 & \mu \mathbf{I}_3 \end{pmatrix} = \bar{M}$. ✓ (Key: $(i\kappa_c)^* = -i\kappa_c$ is exactly the conjugate transpose of the $\bar{\zeta}$ block.)
2. **Reciprocity:** By definition eq. (2.46), we need $\bar{\xi} = -\bar{\zeta}^\top$. Here $\bar{\xi} = i\kappa_c \mathbf{I}_3$, $\bar{\zeta} = -i\kappa_c \mathbf{I}_3$, so $-\bar{\zeta}^\top = i\kappa_c \mathbf{I}_3 = \bar{\xi}$. ✓ That is, chiral media are **reciprocal**.
3. **TRS:** By eq. (2.51), we need $\bar{\xi}^* = -\bar{\xi}$. $(i\kappa_c)^* = -i\kappa_c = -(i\kappa_c)$. ✓ Chiral media **preserve TRS**.

Physical consequences. Chiral media assign different refractive indices to the two circular polarization components of a plane wave (optical activity), but since TRS is preserved, the total Chern number of any TRS-invariant band subspace is zero. Chiral media can produce polarization rotation but cannot realize unidirectionally propagating topological edge states. This stands in sharp contrast to magneto-optical materials, which achieve nonzero Chern numbers and unidirectional edge states by breaking TRS.

Positive-definiteness condition. $\bar{M}$ is positive definite if $\varepsilon > 0$, $\mu > 0$, and the Schur complement $\varepsilon - \kappa_c^2/\mu > 0$, i.e.,

$$\kappa_c^2 < \varepsilon\mu. \quad (2.12)$$

This is the physical upper bound on the chirality parameter: excessively strong chirality ($\kappa_c^2 \geq \varepsilon\mu$) causes the material matrix to lose positive definiteness, corresponding to an unstable material response.

The two examples above demonstrate how the matrix structure of $\bar{M}$ encodes physical symmetries, and how symmetries constrain the possible values of topological invariants. In the next section, we integrate the field variables and $\bar{M}$ into a unified six-component operator formalism, which serves as the starting point for all subsequent rigorous derivations.

2.3 The Six-Component Field Vector and the Maxwell Operator

2.3.1 Compact Form

The electric and magnetic field complex amplitudes are combined into a six-component field vector

$$\mathbf{f}(\mathbf{r}) \equiv \begin{pmatrix} \mathbf{E}(\mathbf{r}) \\ \mathbf{H}(\mathbf{r}) \end{pmatrix} \in \mathbb{C}^6, \quad (2.13)$$

and the Maxwell curl operator is defined as

$$\hat{N} \equiv \begin{pmatrix} 0 & i\nabla \times \\ -i\nabla \times & 0 \end{pmatrix}. \quad (2.14)$$

The frequency-domain Maxwell equations eq. (2.4) are then equivalent to

$$\boxed{\hat{N} \mathbf{f}(\mathbf{r}) = \omega \bar{M}(\mathbf{r}, \omega) \mathbf{f}(\mathbf{r})}. \quad (2.15)$$

2.3.2 The Generalized Eigenvalue Problem Under the Non-Dispersive Approximation

When $\bar{M}$ is independent of ω (or when dispersion is negligible in the frequency range of interest), eq. (2.15) becomes a standard **generalized eigenvalue problem**:

$$\hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n, \quad (2.16)$$

where ω_n is the eigenfrequency and $\mathbf{f}_n$ is the corresponding eigenmode.

Remark. Unlike the *standard* eigenvalue problem $H\psi = E\psi$ of the Schrödinger equation, eq. (2.16) has the material matrix $\bar{M}$ on the right-hand side—this is a **generalized** eigenvalue problem. The role of $\bar{M}$ is analogous to a metric: it defines the geometric structure of the state space.

Example 2.5 (One-dimensional layered medium). Consider an isotropic, lossless medium layered along z and uniform in the xy plane, $\boldsymbol{\varepsilon} = \varepsilon(z)\mathbf{I}_3$, $\boldsymbol{\mu} = \mu_0\mathbf{I}_3$, $\bar{\xi} = \bar{\zeta} = 0$. For TE modes (E_x, H_y, H_z nonzero), eq. (2.16) reduces to the scalar problem

$$-\frac{d}{dz} \left(\frac{1}{\mu_0} \frac{dE_x}{dz} \right) = \omega^2 \varepsilon(z) E_x, \quad (2.17)$$

which is precisely the starting point of the Joannopoulos textbook [16]. However, we emphasize: scalarization is valid only under specific symmetry and polarization conditions; in general three-dimensional or bianisotropic situations, the six-component vector form must be retained.

2.4 The Energy Inner Product and Positive Definiteness

2.4.1 Why Six Components Rather Than Three?

Beginners often ask: why not follow the Joannopoulos textbook and retain only the $\mathbf{H}$ (or $\mathbf{E}$) component, reducing the problem to a three-component equation such as $\hat{\Theta}_H \mathbf{H} = (\omega^2/c^2)\mathbf{H}$? The answer involves three levels:

Mathematical level: The symmetry of the generalized eigenvalue problem is more natural. For non-bianisotropic media, eliminating $\mathbf{E}$ gives $\nabla \times \varepsilon^{-1} \nabla \times \mathbf{H} = \omega^2 \mu_t \mathbf{H}$; in the common nonmagnetic case this reduces to the Joannopoulos operator $\hat{\Theta}_H = \nabla \times \varepsilon^{-1} \nabla \times$ with eigenvalue ω^2/c^2 . Thus the eigenvalue of the three-component form is ω^2 (or ω^2/c^2 after restoring units) rather than ω . This has two consequences: (1) the correspondence between positive and negative frequencies is obscured by the ω^2 mapping; (2) when defining the Berry connection, the natural parameterization should be a family of projection operators with ω as the continuous parameter; ω^2 parameterization introduces an unnecessary singularity at $\omega = 0$. The six-component form $\hat{N}\mathbf{f} = \omega \bar{M}\mathbf{f}$ is linear in ω , so the symmetry between positive- and negative-frequency solutions $(\omega_n, \mathbf{f}_n) \leftrightarrow (-\omega_n, \sigma_z \mathbf{f}_n^*)$ —where $\sigma_z = \text{diag}(\mathbf{I}_3, -\mathbf{I}_3)$ —can be clearly expressed.

Physical level: Bianisotropic materials cannot be reduced to three components. When $\bar{\xi} \neq 0$ or $\bar{\zeta} \neq 0$ (e.g., chiral media or the magneto-electric effect on the surface of a topological insulator), the electric and magnetic fields are coupled in the constitutive relations, and no closed equation involving only $\mathbf{H}$ or only $\mathbf{E}$ exists. Such materials are increasingly important in topological photonics—for instance, the axion electrodynamics on the surface of a three-dimensional topological insulator is described precisely by $\bar{\xi} = \bar{\zeta} = \alpha(\theta/\pi)\mathbf{I}_3$.

Topological level: The gauge freedom of the Berry connection requires a complete basis. When defining the Berry connection of Bloch eigenmodes $\mathcal{A}_n(\mathbf{k}) = -i\langle \mathbf{f}_{n\mathbf{k}}, \nabla_{\mathbf{k}} \mathbf{f}_{n\mathbf{k}} \rangle_M$, the inner product $\langle \cdot, \cdot \rangle_M$ is naturally defined on the six-component space. If truncated to three components, the form of the material-weighted inner product becomes dependent on the specific elimination procedure, losing uniformity across different materials. The six-component inner product is the *only* definition that applies simultaneously to all material types, all polarization states, and all dimensions.

Summary. The three-component reduction $\hat{\Theta}_H$ is a special case of the six-component form under specific conditions—it is highly efficient for two-dimensional problems in isotropic or simple anisotropic media, and this book will also use it in numerical examples (see theorem 2.6). However, as a *theoretical framework*, the six-component form is the logical starting point.

Example 2.6 (2D TM/TE scalarization: From six components to the Joannopoulos form). Consider a two-dimensional system that is translationally invariant along z ($\partial_z = 0$), with $\varepsilon = \varepsilon(x, y)\mathbf{I}_3$, $\mu = \mu(x, y)\mathbf{I}_3$, $\bar{\xi} = \bar{\zeta} = 0$. Starting from the six-component form $\hat{N}\mathbf{f} = \omega \bar{M}\mathbf{f}$, we systematically reduce to the scalar equations of the Joannopoulos textbook.

Step 1: Write out the component form. The $i\nabla \times$ in $\hat{N}$ decomposes under $\partial_z = 0$ as:

$$i\nabla \times = i \begin{pmatrix} 0 & -\partial_z & \partial_y \\ \partial_z & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix} \xrightarrow{\partial_z=0} i \begin{pmatrix} 0 & 0 & \partial_y \\ 0 & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix}. \quad (2.18)$$

The six-component equation $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$ expands into six scalar equations.

Step 2: TM polarization (E_z, H_x, H_y nonzero). For TM polarization we take

$$\mathbf{E} = (0, 0, E_z), \quad \mathbf{H} = (H_x, H_y, 0).$$

The relevant component equations are

$$i\partial_y H_z = \omega\varepsilon E_x \rightarrow E_x = 0, \quad (2.19)$$

$$-i\partial_x H_z = \omega\varepsilon E_y \rightarrow E_y = 0, \quad (2.20)$$

$$i(\partial_x H_y - \partial_y H_x) = \omega\varepsilon E_z, \quad (2.21)$$

$$-i(\partial_y E_z - \partial_z E_y) = \omega\mu H_x, \quad (2.22)$$

$$-i(\partial_z E_x - \partial_x E_z) = \omega\mu H_y. \quad (2.23)$$

Using $\partial_z = 0$ and $E_x = E_y = 0$, the last two equations give

$$H_x = -\frac{i}{\omega\mu} \partial_y E_z, \quad H_y = +\frac{i}{\omega\mu} \partial_x E_z.$$

Substituting these into eq. (2.21) gives

$$\boxed{-\nabla_{xy} \cdot \left(\frac{1}{\mu(\mathbf{r})} \nabla_{xy} E_z \right) = \omega^2 \varepsilon(\mathbf{r}) E_z.} \quad (2.24)$$

Step 3: TE polarization (H_z, E_x, E_y nonzero). For TE polarization we take

$$\mathbf{H} = (0, 0, H_z), \quad \mathbf{E} = (E_x, E_y, 0).$$

The first two equations from $i\nabla \times \mathbf{H} = \omega\varepsilon\mathbf{E}$ give

$$E_x = \frac{i}{\omega\varepsilon} \partial_y H_z, \quad E_y = -\frac{i}{\omega\varepsilon} \partial_x H_z.$$

Substituting these into

$$-i(\partial_x E_y - \partial_y E_x) = \omega\mu H_z$$

gives

$$\boxed{-\nabla_{xy} \cdot \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla_{xy} H_z \right) = \omega^2 \mu(\mathbf{r}) H_z.} \quad (2.25)$$

Key observations. eq. (2.24) and eq. (2.25) are precisely the starting point of the Joannopoulos textbook [16]. However, note several restrictions:

- They are valid **only for isotropic** (or at least z -isotropic), **non-bianisotropic** materials.
- They hold **only in two dimensions** ($\partial_z = 0$). In three dimensions, TM/TE no longer decouple.
- TM and TE are **completely independent**: each possesses its own independent band structure with no mutual coupling. This means that every band of a two-dimensional isotropic photonic crystal is either purely TM or purely TE— but this property is broken when magneto-optical effects or C_6 point-group coupling is introduced (see Chapter 6).

Example 2.7 (Errors caused by the L^2 inner product: A concrete counterexample). The danger of using the wrong inner product can be seen already in a one-dimensional periodic medium. Let $\mathbf{f}_1, \mathbf{f}_2$ be two Bloch-TE modes with different eigenfrequencies. Self-adjointness gives the physically correct orthogonality relation

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \int_0^a (\varepsilon E_{z,1}^* E_{z,2} + \mu H_{x,1}^* H_{x,2} + \mu H_{y,1}^* H_{y,2}) dz = 0. \quad (2.26)$$

By contrast, the naive L^2 overlap $\int_0^a (E_{z,1}^* E_{z,2} + H_{x,1}^* H_{x,2} + H_{y,1}^* H_{y,2}) dz$ is generally nonzero in an inhomogeneous medium. For example, for $\varepsilon(z) = 1 + 3 \cos^2(\pi z/a)$, $\mu = 1$, two first-band modes at $k_1 = 0.2\pi/a$ and $k_2 = 0.4\pi/a$ can have normalized L^2 overlap of order 10^{-1} while their material-weighted overlap is zero to numerical precision. The same mistake propagates directly into Berry geometry:

$$\mathcal{A}_n^{(L^2)}(k) = i \langle \mathbf{f}_{nk}, \partial_k \mathbf{f}_{nk} \rangle_{L^2} \neq i \langle \mathbf{f}_{nk}, \partial_k \mathbf{f}_{nk} \rangle_M = \mathcal{A}_n(k), \quad (2.27)$$

which is why the material-weighted inner product is fixed before any topological invariant is introduced.

The three examples above demonstrate from different perspectives the necessity of the six-component form and the material-weighted inner product. Having fully understood the structure of $\bar{M}$ and the operator properties of $\hat{N}$, we can now formally establish the complete theoretical framework of the energy inner product and prove that it endows $\hat{N}$ with self-adjointness—the most central theorem of this chapter.

In quantum mechanics, the inner product on the state space is the standard L^2 inner product: $\langle \psi | \phi \rangle = \int \psi^*(\mathbf{r}) \phi(\mathbf{r}) d^3r$. For photonic systems, one might similarly define a “naive inner product”

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2} = \int d^3r (\mathbf{E}_1^* \cdot \mathbf{E}_2 + \mathbf{H}_1^* \cdot \mathbf{H}_2). \quad (2.28)$$

However, this inner product has serious deficiencies:

1. Under $\langle \cdot, \cdot \rangle_{L^2}$, $\hat{N}$ is **not** self-adjoint (unless $\bar{M}$ happens to be a constant multiple of the identity matrix, i.e., vacuum).
2. The orthogonality of eigenmodes under $\langle \cdot, \cdot \rangle_{L^2}$ does not correspond to energy orthogonality.
3. When the Berry connection is defined under $\langle \cdot, \cdot \rangle_{L^2}$, the correspondence between its curvature and physical observables (such as the anomalous velocity) is destroyed.

2.4.2 The Material-Weighted (Energy) Inner Product

Definition 2.8 (Energy inner product). For six-component field vectors $\mathbf{f}_1, \mathbf{f}_2$, the material-weighted inner product is defined as

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M \equiv \int d^3r \mathbf{f}_1^\dagger(\mathbf{r}) \bar{M}(\mathbf{r}) \mathbf{f}_2(\mathbf{r}). \quad (2.29)$$

Expanded, this reads

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \int d^3r (\mathbf{E}_1^* \cdot \boldsymbol{\epsilon} \mathbf{E}_2 + \mathbf{E}_1^* \cdot \bar{\boldsymbol{\xi}} \mathbf{H}_2 + \mathbf{H}_1^* \cdot \bar{\boldsymbol{\zeta}} \mathbf{E}_2 + \mathbf{H}_1^* \cdot \boldsymbol{\mu} \mathbf{H}_2). \quad (2.30)$$

When $\bar{\boldsymbol{\xi}} = \bar{\boldsymbol{\zeta}} = 0$ (no bianisotropy), this simplifies to

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \int d^3r (\mathbf{E}_1^* \cdot \boldsymbol{\epsilon} \mathbf{E}_2 + \mathbf{H}_1^* \cdot \boldsymbol{\mu} \mathbf{H}_2), \quad (2.31)$$

in which case $\frac{1}{4} \langle \mathbf{f}, \mathbf{f} \rangle_M$ is exactly the volume integral of the time-averaged electromagnetic energy density.

Proposition 2.9 (Positive definiteness). *If $\bar{M}(\mathbf{r})$ is a positive-definite Hermitian matrix for all $\mathbf{r}$ in the domain Ω , then $\langle \cdot, \cdot \rangle_M$ is a positive-definite inner product on $L^2(\Omega; \mathbb{C}^6)$.*

Proof. The positive-definite Hermitian matrix $\bar{M}$ guarantees that for any nonzero $\mathbf{f}(\mathbf{r})$, $\mathbf{f}^\dagger(\mathbf{r}) \bar{M}(\mathbf{r}) \mathbf{f}(\mathbf{r}) > 0$. After integration, $\langle \mathbf{f}, \mathbf{f} \rangle_M = \int \mathbf{f}^\dagger \bar{M} \mathbf{f} d^3r > 0$ (for nonzero $\mathbf{f}$). Conjugate symmetry $\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \overline{\langle \mathbf{f}_2, \mathbf{f}_1 \rangle_M}$ is directly guaranteed by $\bar{M} = \bar{M}^\dagger$. Linearity and non-degeneracy are obvious. $\square$

Remark. Positive definiteness requires all eigenvalues of $\bar{M}$ to be positive. For vacuum ($\boldsymbol{\epsilon} = \epsilon_0 \mathbf{I}_3$, $\boldsymbol{\mu} = \mu_0 \mathbf{I}_3$), $\bar{M}$ is obviously positive definite. For general lossless media, positive definiteness is equivalent to the positive definiteness of $\boldsymbol{\epsilon}$ and $\boldsymbol{\mu}$ plus appropriate constraints on $\bar{\boldsymbol{\xi}}$ and $\bar{\boldsymbol{\zeta}}$ (Schur complement condition). Media with gain or loss ($\bar{M} \neq \bar{M}^\dagger$) will be discussed in Chapter 9.

2.5 The Self-Adjointness Theorem

With the energy inner product in hand, we can rigorously state and prove the self-adjointness of the metric-weighted Maxwell operator $\hat{H}_M = \bar{M}^{-1} \hat{N}$ —one of the most important theorems in this entire book, because it guarantees a real eigenvalue spectrum and orthogonal eigenmodes, thereby providing Berry geometry with a sound mathematical foundation.

Self-adjointness has three consequences that will be used repeatedly: real eigenfrequencies, energy orthogonality, and modal completeness. In particular, for distinct eigenfrequencies,

$$\omega_m \neq \omega_n \implies \langle \mathbf{f}_m, \mathbf{f}_n \rangle_M = \int \mathbf{f}_m^\dagger \bar{M} \mathbf{f}_n d^3r = 0. \quad (2.32)$$

Remark (Comparison with L^2 orthogonality). Under the ordinary L^2 inner product $\langle \mathbf{f}_m | \mathbf{f}_n \rangle_{L^2} = \int \mathbf{f}_m^\dagger \mathbf{f}_n d^3r$, $\hat{N}$ is self-adjoint under the same boundary conditions, but the generalized eigenmodes solve $\hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n$, not $\hat{N} \mathbf{f}_n = \omega_n \mathbf{f}_n$. Therefore different-frequency Maxwell modes are not

generally orthogonal in the unweighted L^2 sense. Material-weighted orthogonality is the physically correct choice.

Remark (Breakdown and repair in lossy systems). When the material has losses, the imaginary part of $\bar{M}$ causes the metric-weighted Maxwell operator to lose self-adjointness, and the threefold guarantee fails one by one: eigenfrequencies become complex (modes decay), left and right eigenvectors are no longer equal (biorthogonal bases replace orthogonal bases), and completeness must be corrected to an overcomplete quasinormal mode (QNM) expansion. Chapter 9 will systematically address these breakdowns.

Theorem 2.10 (Self-adjointness of the Maxwell operator). *Let $\bar{M}(\mathbf{r})$ be lossless, Hermitian, and positive definite,*

$$\bar{M}(\mathbf{r}) = \bar{M}^\dagger(\mathbf{r}), \quad \bar{M}(\mathbf{r}) > 0.$$

Assume that the fields satisfy boundary conditions for which the surface terms generated by integration by parts vanish, such as closed perfectly conducting boundaries or periodic/Bloch-periodic boundaries. Define the Maxwell operator

$$\hat{H}_M = \bar{M}^{-1} \hat{N},$$

where

$$\hat{N} = \begin{pmatrix} 0 & i\nabla \times \\ -i\nabla \times & 0 \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}.$$

Then $\hat{H}_M$ is self-adjoint under the energy inner product

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \int_{\Omega} d^3r \mathbf{f}_1^\dagger(\mathbf{r}) \bar{M}(\mathbf{r}) \mathbf{f}_2(\mathbf{r}).$$

That is,

$$\boxed{\langle \mathbf{f}_1, \hat{H}_M \mathbf{f}_2 \rangle_M = \langle \hat{H}_M \mathbf{f}_1, \mathbf{f}_2 \rangle_M.} \quad (2.33)$$

Equivalently, the generalized Maxwell eigenvalue problem

$$\hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n$$

is Hermitian with respect to the positive-definite metric $\bar{M}$.

Proof. We first establish the integration-by-parts identity for $\hat{N}$ under the standard L^2 inner product

$$\langle \mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2} = \int_{\Omega} d^3r \mathbf{f}_1^\dagger(\mathbf{r}) \mathbf{f}_2(\mathbf{r}).$$

For two fields

$$\mathbf{f}_j = \begin{pmatrix} \mathbf{E}_j \\ \mathbf{H}_j \end{pmatrix}, \quad j = 1, 2,$$

we have

$$\begin{aligned}\langle \mathbf{f}_1, \hat{N}\mathbf{f}_2 \rangle_{L^2} &= \int_{\Omega} d^3r [\mathbf{E}_1^* \cdot (i\nabla \times \mathbf{H}_2) + \mathbf{H}_1^* \cdot (-i\nabla \times \mathbf{E}_2)] \\ &= i \int_{\Omega} d^3r \mathbf{E}_1^* \cdot (\nabla \times \mathbf{H}_2) - i \int_{\Omega} d^3r \mathbf{H}_1^* \cdot (\nabla \times \mathbf{E}_2).\end{aligned}\quad (2.34)$$

Using the vector identity

$$\mathbf{A} \cdot (\nabla \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \nabla \cdot (\mathbf{B} \times \mathbf{A}),$$

with $\mathbf{A} = \mathbf{E}_1^*$, $\mathbf{B} = \mathbf{H}_2$ in the first term and $\mathbf{A} = \mathbf{H}_1^*$, $\mathbf{B} = \mathbf{E}_2$ in the second term, we obtain

$$\begin{aligned}\langle \mathbf{f}_1, \hat{N}\mathbf{f}_2 \rangle_{L^2} &= i \int_{\Omega} d^3r \mathbf{H}_2 \cdot (\nabla \times \mathbf{E}_1^*) - i \int_{\Omega} d^3r \mathbf{E}_2 \cdot (\nabla \times \mathbf{H}_1^*) \\ &\quad - i \oint_{\partial\Omega} dS \mathbf{n} \cdot (\mathbf{H}_2 \times \mathbf{E}_1^*) + i \oint_{\partial\Omega} dS \mathbf{n} \cdot (\mathbf{E}_2 \times \mathbf{H}_1^*).\end{aligned}\quad (2.35)$$

Since

$$(\hat{N}\mathbf{f}_1)^\dagger \mathbf{f}_2 = (-i\nabla \times \mathbf{H}_1^*) \cdot \mathbf{E}_2 + (i\nabla \times \mathbf{E}_1^*) \cdot \mathbf{H}_2,$$

the volume terms are precisely

$$\langle \hat{N}\mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2}.$$

Therefore,

$$\langle \mathbf{f}_1, \hat{N}\mathbf{f}_2 \rangle_{L^2} = \langle \hat{N}\mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2} + i \oint_{\partial\Omega} dS \mathbf{n} \cdot (\mathbf{E}_2 \times \mathbf{H}_1^* - \mathbf{H}_2 \times \mathbf{E}_1^*). \quad (2.36)$$

Under the assumed boundary conditions, the surface integral vanishes. For example, for a perfectly conducting boundary,

$$\mathbf{n} \times \mathbf{E}_j|_{\partial\Omega} = 0, \quad j = 1, 2,$$

and hence both surface terms vanish. For periodic or Bloch-periodic boundary conditions, the contributions from opposite faces cancel. Thus,

$$\langle \mathbf{f}_1, \hat{N}\mathbf{f}_2 \rangle_{L^2} = \langle \hat{N}\mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2}. \quad (2.37)$$

In this sense, $\hat{N}$ is formally self-adjoint under the standard L^2 inner product.

We now prove the self-adjointness of the Maxwell operator

$$\hat{H}_M = \bar{M}^{-1} \hat{N}$$

under the energy inner product. By definition,

$$\begin{aligned}\langle \mathbf{f}_1, \hat{H}_M \mathbf{f}_2 \rangle_M &= \int_{\Omega} d^3r \mathbf{f}_1^\dagger \bar{M} (\bar{M}^{-1} \hat{N} \mathbf{f}_2) \\ &= \int_{\Omega} d^3r \mathbf{f}_1^\dagger \hat{N} \mathbf{f}_2 \\ &= \langle \mathbf{f}_1, \hat{N} \mathbf{f}_2 \rangle_{L^2}.\end{aligned}\quad (2.38)$$

Using (2.37), this becomes

$$\langle \mathbf{f}_1, \hat{H}_M \mathbf{f}_2 \rangle_M = \langle \hat{N} \mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2}. \quad (2.39)$$

On the other hand,

$$\langle \hat{H}_M \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \int_{\Omega} d^3r \left(\bar{M}^{-1} \hat{N} \mathbf{f}_1 \right)^{\dagger} \bar{M} \mathbf{f}_2. \quad (2.40)$$

Because $\bar{M} = \bar{M}^{\dagger}$ and $\bar{M} > 0$, its inverse is also Hermitian,

$$(\bar{M}^{-1})^{\dagger} = \bar{M}^{-1}.$$

Therefore,

$$\begin{aligned} \langle \hat{H}_M \mathbf{f}_1, \mathbf{f}_2 \rangle_M &= \int_{\Omega} d^3r (\hat{N} \mathbf{f}_1)^{\dagger} (\bar{M}^{-1})^{\dagger} \bar{M} \mathbf{f}_2 \\ &= \int_{\Omega} d^3r (\hat{N} \mathbf{f}_1)^{\dagger} \mathbf{f}_2 \\ &= \langle \hat{N} \mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2}. \end{aligned} \quad (2.41)$$

Hence,

$$\langle \mathbf{f}_1, \hat{H}_M \mathbf{f}_2 \rangle_M = \langle \hat{H}_M \mathbf{f}_1, \mathbf{f}_2 \rangle_M.$$

This proves that $\hat{H}_M = \bar{M}^{-1} \hat{N}$ is self-adjoint under the energy inner product. $\square$

Corollary 2.11 (Reality of the spectrum and energy orthogonality). *Under the same assumptions as in theorem 2.10, the generalized Maxwell eigenvalue problem*

$$\hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n \quad (2.42)$$

has the following properties:

1. *the eigenfrequencies ω_n are real;*
2. *eigenmodes with different eigenfrequencies are orthogonal under the energy inner product;*
3. *in a closed finite domain, the spectrum is discrete, whereas for periodic systems the eigenfrequencies form bands.*

Proof. The generalized eigenvalue problem

$$\hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n$$

is equivalent to the ordinary eigenvalue problem

$$\hat{H}_M \mathbf{f}_n = \omega_n \mathbf{f}_n, \quad \hat{H}_M = \bar{M}^{-1} \hat{N}.$$

Since $\hat{H}_M$ is self-adjoint under the positive-definite energy inner product, its eigenvalues are real and its eigenmodes can be chosen to be orthogonal under this inner product.

Equivalently, one may verify this directly. Let $\mathbf{f}_m$ and $\mathbf{f}_n$ satisfy

$$\hat{N} \mathbf{f}_m = \omega_m \bar{M} \mathbf{f}_m, \quad \hat{N} \mathbf{f}_n = \omega_n \bar{M} \mathbf{f}_n.$$

Using the L^2 self-adjointness of $\hat{N}$, we have

$$\langle \mathbf{f}_m, \hat{N} \mathbf{f}_n \rangle_{L^2} = \langle \hat{N} \mathbf{f}_m, \mathbf{f}_n \rangle_{L^2}.$$

The left-hand side gives

$$\langle \mathbf{f}_m, \hat{N} \mathbf{f}_n \rangle_{L^2} = \omega_n \langle \mathbf{f}_m, \bar{M} \mathbf{f}_n \rangle_{L^2} = \omega_n \langle \mathbf{f}_m, \mathbf{f}_n \rangle_M.$$

The right-hand side gives

$$\langle \hat{N} \mathbf{f}_m, \mathbf{f}_n \rangle_{L^2} = \langle \omega_m \bar{M} \mathbf{f}_m, \mathbf{f}_n \rangle_{L^2} = \omega_m^* \langle \bar{M} \mathbf{f}_m, \mathbf{f}_n \rangle_{L^2}.$$

Since $\bar{M} = \bar{M}^\dagger$,

$$\langle \bar{M} \mathbf{f}_m, \mathbf{f}_n \rangle_{L^2} = \langle \mathbf{f}_m, \bar{M} \mathbf{f}_n \rangle_{L^2} = \langle \mathbf{f}_m, \mathbf{f}_n \rangle_M.$$

Therefore,

$$(\omega_n - \omega_m^*) \langle \mathbf{f}_m, \mathbf{f}_n \rangle_M = 0. \quad (2.43)$$

Taking $m = n$ gives

$$(\omega_n - \omega_n^*) \langle \mathbf{f}_n, \mathbf{f}_n \rangle_M = 0.$$

Because $\bar{M}$ is positive definite,

$$\langle \mathbf{f}_n, \mathbf{f}_n \rangle_M > 0$$

for any nonzero eigenmode. Hence

$$\omega_n = \omega_n^*,$$

so the eigenfrequencies are real. For $m \neq n$, if

$$\omega_m \neq \omega_n,$$

then (2.43) gives

$$\langle \mathbf{f}_m, \mathbf{f}_n \rangle_M = 0.$$

Thus eigenmodes with distinct eigenfrequencies are orthogonal under the energy inner product. Within degenerate eigenspaces, the eigenmodes may be orthonormalized using the same positive-definite energy inner product. $\square$

2.6 The Influence of Boundary Conditions on Spectral Structure

The spectrum of eq. (2.16) depends not only on the spatial distribution of $\bar{M}$, but also critically on the boundary conditions.

2.6.1 Closed Boundaries

Consider a finite domain Ω enclosed by a perfect electric conductor (PEC), with boundary condition

$$\hat{\mathbf{n}} \times \mathbf{E}|_{\partial\Omega} = 0. \quad (2.44)$$

In this case the spectrum is discrete: $0 \leq \omega_1 \leq \omega_2 \leq \dots$, and the eigenmodes form a countable orthonormal set. This is the simplest “textbook” setting.

2.6.2 Periodic Boundaries and Bloch Parameterization

When $\bar{M}(\mathbf{r})$ has lattice periodicity $\bar{M}(\mathbf{r} + \mathbf{R}) = \bar{M}(\mathbf{r})$ ($\mathbf{R}$ being a lattice vector), Bloch’s theorem can be used to reduce the eigenvalue problem of the infinite periodic system to a parameterized problem on a single unit cell. This is the core content of Chapter 3; here we only give a preview:

The periodic part $\mathbf{u}_{n\mathbf{k}}$ of the Bloch field $\mathbf{f}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ satisfies a $\mathbf{k}$ -dependent eigenequation

$$\hat{N}(\mathbf{k}) \mathbf{u}_{n\mathbf{k}} = \omega_n(\mathbf{k}) \bar{M} \mathbf{u}_{n\mathbf{k}}, \quad (2.45)$$

where $\hat{N}(\mathbf{k})$ is the curl operator after the substitution $\nabla \rightarrow \nabla + i\mathbf{k}$. The dispersion relations $\omega_n(\mathbf{k})$ constitute the **photonic bands**, and topological invariants are defined on the Brillouin zone.

2.6.3 Open Boundaries and Complex-Frequency Spectra

Real photonic devices are rarely fully closed or infinitely periodic. Radiative leakage causes eigenfrequencies to become complex, $\tilde{\omega}_n = \omega'_n - i\gamma_n/2$ ($\gamma_n > 0$ corresponds to decay), and the eigenmodes become **quasinormal modes** (QNM). In this case, the lossless condition on $\bar{M}$ is insufficient to guarantee a real spectrum—the radiative boundary conditions themselves introduce effective “non-Hermiticity.”

Remark. There are two routes for defining topological invariants in open systems: (1) based on the complex-frequency band structure and generalized Berry phase of QNMs; (2) based on the phase topology of the scattering matrix (S matrix) or reflection matrix. Each has its own applicable scenarios, to be discussed in detail in Chapter 8. This book makes it explicit: **whenever open systems are involved, one must specify which topological definition is being used**; the Chern number formula for closed systems cannot be applied unmodified to QNMs.

2.7 Reciprocity: The Lorentz Reciprocity Theorem

Reciprocity is one of the most fundamental symmetries in classical electromagnetism, but it is frequently confused with time-reversal symmetry. This section provides a rigorous distinction.

Theorem 2.12 (Lorentz reciprocity theorem). *Let $(\mathbf{E}_1, \mathbf{H}_1)$ and $(\mathbf{E}_2, \mathbf{H}_2)$ be frequency-domain fields excited by sources $\mathbf{J}_1$ and $\mathbf{J}_2$, respectively, in the same linear medium. If the material tensors satisfy*

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top, \quad \boldsymbol{\mu} = \boldsymbol{\mu}^\top, \quad \bar{\boldsymbol{\xi}} = -\bar{\boldsymbol{\zeta}}^\top, \quad (2.46)$$

then over all space

$$\int d^3r \mathbf{J}_1 \cdot \mathbf{E}_2 = \int d^3r \mathbf{J}_2 \cdot \mathbf{E}_1. \quad (2.47)$$

Proof. Starting from Maxwell’s equations, using the vector identity $\nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1) = \mathbf{H}_2 \cdot \nabla \times \mathbf{E}_1 - \mathbf{E}_1 \cdot \nabla \times \mathbf{H}_2 - \mathbf{H}_1 \cdot \nabla \times \mathbf{E}_2 + \mathbf{E}_2 \cdot \nabla \times \mathbf{H}_1$, substituting the frequency-domain

Maxwell equations eq. (2.4) and using the constitutive relations eq. (2.5), after rearrangement the divergence term integrates to zero over the surface at infinity (radiation condition), yielding

$$\int d^3r (\mathbf{J}_1 \cdot \mathbf{E}_2 - \mathbf{J}_2 \cdot \mathbf{E}_1) = -i\omega \int d^3r [\mathbf{E}_1 \cdot (\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^\top) \mathbf{E}_2 + \mathbf{H}_1 \cdot (\boldsymbol{\mu} - \boldsymbol{\mu}^\top) \mathbf{H}_2 + \dots]. \quad (2.48)$$

When eq. (2.46) holds, the right-hand side vanishes, giving eq. (2.47). $\square$

Remark. In the reciprocity condition eq. (2.46), $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\top$ is the *transpose*, not the conjugate transpose. For magneto-optical materials (gyro-anisotropic), $\boldsymbol{\varepsilon}$ has an antisymmetric imaginary part: $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}^\top$, so magneto-optical materials are **nonreciprocal**. In the common lossless local gyrotropic setting, the same bias also breaks TRS and makes nonzero photonic Chern numbers possible [1].

Example 2.13 (Scattering matrix symmetry in reciprocal systems). In a reciprocal system, the multi-port scattering matrix satisfies $S = S^\top$ (note: transpose, not conjugate transpose). This means $S_{ij} = S_{ji}$: the transmission coefficient from port i to port j equals that from j to i . Nonreciprocal devices (such as isolators and circulators) use the magneto-optical effect to break this symmetry.

2.7.1 Microscopic Origin: Onsager Reciprocal Relations

The Lorentz reciprocity theorem is a *macroscopic* electromagnetic theorem equivalent to the symmetry of the material tensor $\boldsymbol{\varepsilon}^\top = \boldsymbol{\varepsilon}$. But this symmetry is not accidental—it is rooted in the **Onsager reciprocal relations** (1931) from statistical mechanics, which in turn originate from the time-reversal invariance of microscopic equations of motion.

Precise distinction among three reciprocity/symmetry concepts

1. **Onsager reciprocity** (microscopic, always holds): $\varepsilon_{ij}(\omega, \mathbf{B}_0) = \varepsilon_{ji}(\omega, -\mathbf{B}_0)$. This is a direct corollary of microscopic reversibility for linear response near equilibrium—under an external magnetic field $\mathbf{B}_0$, time-reversing the carrier equations of motion while simultaneously reversing $\mathbf{B}_0$ returns the system to its original state. Onsager reciprocity holds for *any* linear material, including magneto-optical media.
2. **Lorentz reciprocity** (macroscopic, conditional): $\varepsilon_{ij}(\omega) = \varepsilon_{ji}(\omega)$, i.e., $\boldsymbol{\varepsilon}^\top = \boldsymbol{\varepsilon}$. When all time-odd bias fields vanish (or their influence is negligible), the Onsager relation reduces to Lorentz reciprocity. Magneto-optical media ($\mathbf{B}_0 \neq 0$, $\boldsymbol{\varepsilon}$ contains the antisymmetric component *ig*) violate this condition.
3. **Time-reversal symmetry** (TRS, detailed in §2.8): $\sigma_T \bar{M}^* \sigma_T = \bar{M}$; in simple non-bianisotropic media this reduces to $\varepsilon^* = \varepsilon$ and $\mu^* = \mu$. This is an *independent* symmetry from reciprocity (see Table 2.1).

The logical relationship among the three can be summarized as:

$$\text{Onsager (always holds)} \xrightarrow{\mathbf{B}_0=0} \text{Lorentz reciprocity} \xleftarrow{\text{independent}} \text{TRS } (\varepsilon^* = \varepsilon)$$

A reciprocal but lossy system (such as absorption in ordinary glass) satisfies Lorentz reciprocity but is not a closed time-reversal-invariant Hermitian system; a biased magneto-optical but lossless

system (such as ideal YIG at microwave frequencies) breaks both TRS and Lorentz reciprocity. Only a lossless medium without an external magnetic field satisfies all three simultaneously.

Remark (Nonlinear generalization). The Lorentz reciprocity theorem can be generalized to weakly nonlinear media. In parametric processes, Miller's rule guarantees that $\chi_{ijk}^{(2)}$ has symmetry under the joint permutation of its indices and frequencies, which can be viewed as a corollary of the nonlinear Onsager relation. However, strong nonlinear effects (such as optical bistability) can break reciprocity—this is one of the theoretical foundations for magnet-free nonreciprocal devices. See Chapter 13 for a discussion of nonlinear topological photonics.

2.8 Time-Reversal Symmetry

2.8.1 The Anti-Unitary Operator $\mathcal{T}$

Definition 2.14 (Time reversal of electromagnetic fields). The time-reversal operator $\mathcal{T}$ is an **anti-unitary** (antilinear, antiunitary) operator whose action on frequency-domain complex amplitude fields is

$$\mathcal{T}: \quad \mathbf{E}(\mathbf{r}) \mapsto \mathbf{E}^*(\mathbf{r}), \quad \mathbf{H}(\mathbf{r}) \mapsto -\mathbf{H}^*(\mathbf{r}). \quad (2.49)$$

Equivalently, for the six-component field vector

$$\mathcal{T}\mathbf{f} = \begin{pmatrix} \mathbf{I}_3 & 0 \\ 0 & -\mathbf{I}_3 \end{pmatrix} \mathbf{f}^* \equiv \sigma_T \mathbf{f}^*, \quad (2.50)$$

where $\sigma_T = \text{diag}(1, 1, 1, -1, -1, -1)$.

Remark. The anti-unitarity of $\mathcal{T}$ ($\mathcal{T}(\alpha\mathbf{f}) = \alpha^*\mathcal{T}\mathbf{f}$) means that it cannot be represented by a unitary matrix. This is a feature shared by photonic TRS and electronic TRS, but in photonic systems there is no Kramers degeneracy (because $\mathcal{T}^2 = +1$ rather than -1). This difference profoundly affects the topological classification.

2.8.2 Conditions for TRS to Hold

Proposition 2.15 (TRS condition). *The generalized eigenvalue problem eq. (2.16) is invariant under $\mathcal{T}$ if and only if the material matrix satisfies*

$$\sigma_T \bar{M}^*(\mathbf{r}) \sigma_T = \bar{M}(\mathbf{r}). \quad (2.51)$$

Expanded in components: $\boldsymbol{\varepsilon}^* = \boldsymbol{\varepsilon}$ (real), $\boldsymbol{\mu}^* = \boldsymbol{\mu}$ (real), $\bar{\xi}^* = -\bar{\xi}$, $\bar{\zeta}^* = -\bar{\zeta}$.

Proof. Apply $\mathcal{T}$ to both sides of $\hat{N}\mathbf{f}_n = \omega_n \bar{M}\mathbf{f}_n$: $\mathcal{T}\hat{N}\mathbf{f}_n = \omega_n^* \sigma_T \bar{M}^* \mathbf{f}_n^*$.

On the other hand, using the reality of $\nabla \times$ and the antilinearity of $\mathcal{T}$, one can verify that $\mathcal{T}\hat{N} = \hat{N}\mathcal{T}$ (the structure of $\hat{N}$ guarantees this: i changes sign under complex conjugation, and the sign arrangement of σ_T exactly compensates). Therefore $\hat{N}(\mathcal{T}\mathbf{f}_n) = \omega_n^* \sigma_T \bar{M}^* \sigma_T (\mathcal{T}\mathbf{f}_n)$.

For $\mathcal{T}\mathbf{f}_n$ to also be an eigenmode with eigenfrequency ω_n (real, as guaranteed by self-adjointness), we need $\sigma_T \bar{M}^* \sigma_T = \bar{M}$. $\square$

2.8.3 Distinguishing TRS from Reciprocity

Table 2.1 Comparison between reciprocity and time-reversal symmetry.

Feature	Reciprocity	Time-reversal symmetry
Mathematical nature	Lorentz reciprocity theorem; a bilinear identity between two solutions.	Anti-unitary symmetry of the wave equation or effective Hamiltonian.
Material condition	For local linear media, $\epsilon = \epsilon^\top$, $\mu = \mu^\top$, and $\xi = -\xi^\top$.	In a lossless frequency-domain formulation, $\epsilon = \epsilon^*$ and $\mu = \mu^*$.
Broken by	Magneto-optical bias, moving media, or other nonreciprocal couplings.	Magneto-optical bias; gain or loss can make the effective eigenproblem non-Hermitian.
Effect on S matrix	With reciprocal port normalization, $S = S^\top$.	Time reversal relates the scattering process to its reversed process in a lossless system.
Effect on Berry curvature	No direct constraint.	For a TRS-invariant isolated band or band subspace, $\Omega(\mathbf{k}) = -\Omega(-\mathbf{k})$.
Effect on Chern number	No direct constraint.	The total Chern number of a TRS-invariant band subspace vanishes.

Remark. table 2.1 clearly shows that a system can simultaneously possess both reciprocity and TRS (e.g., ordinary dielectrics), can simultaneously break both (magneto-optical materials), but intermediate cases also exist (for example, lossy but non-magnetic materials preserve Lorentz reciprocity but are not Hermitian TRS systems in the closed-system sense). A frequently occurring error in the literature is to equate “nonreciprocal” with “TRS-broken”—the two are equivalent in lossless, local media, but are not equivalent in lossy or nonlocal cases.

The comparison table above distinguishes reciprocity from TRS using abstract matrix conditions. To give the reader *actionable physical intuition*, we use two classic device-level examples to “ground” these conditions.

Example 2.16 (Nonreciprocal scattering matrix of a three-port microwave circulator). Consider a lossless three-port Y-junction with C_3 symmetry and a ferrite bias $\mathbf{B}_0 = B_0 \hat{z}$. In the port basis, C_3 symmetry forces the scattering matrix to be circulant:

$$S = \begin{pmatrix} \alpha & \gamma & \beta \\ \beta & \alpha & \gamma \\ \gamma & \beta & \alpha \end{pmatrix}. \quad (2.52)$$

Reciprocity requires $S = S^\top$, which here is equivalent to

$$S = S^\top \iff \beta = \gamma. \quad (2.53)$$

Perfect clockwise circulation corresponds to

$$\alpha = 0, \quad \beta = e^{i\phi}, \quad \gamma = 0, \quad (2.54)$$

and hence

$$S = e^{i\phi} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = e^{i\phi} P. \quad (2.55)$$

The microscopic origin is the gyromagnetic permeability of the biased ferrite,

$$\bar{\mu} = \begin{pmatrix} \mu & i\kappa & 0 \\ -i\kappa & \mu & 0 \\ 0 & 0 & \mu_z \end{pmatrix}, \quad (2.56)$$

which is Hermitian but not symmetric for $\kappa \neq 0$. Thus Lorentz reciprocity is broken. Time reversal sends $\mathbf{B}_0 \rightarrow -\mathbf{B}_0$ and $\kappa \rightarrow -\kappa$, reversing the preferred circulation direction.

Example 2.17 (Faraday rotation and time-reversal symmetry breaking). A linearly polarized plane wave propagates along the z direction in a uniform magneto-optical medium whose permittivity tensor is

$$\varepsilon = \begin{pmatrix} \varepsilon_d & -ig & 0 \\ ig & \varepsilon_d & 0 \\ 0 & 0 & \varepsilon_z \end{pmatrix}, \quad \boldsymbol{\mu} = \mu_0 \mathbb{I}, \quad (2.57)$$

where $g > 0$ is proportional to the strength of the applied DC magnetic field B_0 . For a plane wave propagating along z , Maxwell's equations give

$$k^2 \mathbf{e} = \frac{\omega^2}{c^2} \boldsymbol{\varepsilon}_\perp \mathbf{e}, \quad \text{where } \boldsymbol{\varepsilon}_\perp = \begin{pmatrix} \varepsilon_d & -ig \\ ig & \varepsilon_d \end{pmatrix}. \quad (2.58)$$

The eigenvalues of $\boldsymbol{\varepsilon}_\perp$ are $\varepsilon_\pm = \varepsilon_d \pm g$, with corresponding eigenvectors

$$\begin{aligned} \hat{\mathbf{e}}_+ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, & \text{left circular polarization (LCP),} \\ \hat{\mathbf{e}}_- &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, & \text{right circular polarization (RCP).} \end{aligned} \quad (2.59)$$

The corresponding propagation constants are $k_\pm = (\omega/c)\sqrt{\varepsilon_\pm\mu_0}$.

The incident linearly polarized wave (x direction) is $\mathbf{E}(0) = E_0 \hat{x} = (E_0/\sqrt{2})(\hat{\mathbf{e}}_+ + \hat{\mathbf{e}}_-)$. After propagating a distance L , the two circular components acquire different phases, so the polarization plane rotates by

$$\theta_F = \frac{(k_+ - k_-)L}{2} = \frac{\omega L}{2c} (\sqrt{\varepsilon_+ \mu_0} - \sqrt{\varepsilon_- \mu_0}) \approx \frac{\omega L g}{2c \sqrt{\varepsilon_d \mu_0}} \quad (2.60)$$

where the last step is the linear approximation for $g \ll \varepsilon_d$.

The crucial point is non-reciprocal accumulation. On a round trip through a Faraday medium, the rotation is $2\theta_F$, not zero. In operator language, $\boldsymbol{\varepsilon}^* \neq \boldsymbol{\varepsilon}$, so eq. (2.51) is violated. This differs

from natural optical activity in reciprocal chiral media, where forward and backward rotations cancel:

Table 2.2 Comparison between Faraday rotation and natural optical activity.

	Faraday rotation	Natural optical activity
Rotation direction	Determined by the direction of $\mathbf{B}_0$; unchanged when the propagation direction is reversed.	Determined by the propagation direction; forward and backward rotations have opposite signs.
Round-trip rotation	$2\theta_F$; cumulative.	0; cancels after a round trip.
TRS	Broken.	Preserved.
Reciprocity	Broken; for example, $\boldsymbol{\varepsilon} \neq \boldsymbol{\varepsilon}^\top$ due to imaginary off-diagonal components.	Preserved; reciprocal bianisotropic media satisfy $\bar{\boldsymbol{\xi}} = -\bar{\boldsymbol{\zeta}}^\top$.
Topological consequence	Can support bands with nonzero Chern number.	The total Chern number of a TRS-invariant band subspace must vanish when time-reversal symmetry is preserved.

Remark. The irreversibility of Faraday rotation is the physical basis for constructing optical isolators. Combining a Faraday rotator with a polarizer achieves unidirectional transmission, which is used in laser systems to protect the source from back-reflections. In topological photonics, the Faraday effect is the core mechanism for realizing photonic Chern insulators (chapter 5)—a nonzero g leads to nontrivial Chern numbers in the band structure.

The two examples above translate the operator conditions for reciprocity/TRS into observable experimental phenomena: the unidirectional circulation of the circulator directly corresponds to $S \neq S^\top$, and the irreversible polarization rotation of Faraday effect directly corresponds to the absence of $\hat{K}$ symmetry. After understanding these physical pictures, when the reader encounters “break TRS to obtain a nonzero Chern number” in subsequent chapters, they should immediately be able to associate it with the specific material (magneto-optical ferrite) and phenomenon (Faraday rotation).

2.8.4 Photonic $\mathcal{T}^2 = +1$ and the Absence of Kramers’ Theorem

The two preceding examples demonstrated the direct physical consequences of TRS breaking. But even when TRS *is preserved* (no magneto-optical effects, $\boldsymbol{\varepsilon}^* = \boldsymbol{\varepsilon}$), photonic systems still have a **fundamental difference** from electronic systems— the sign of the square of the TRS operator.

Square of the six-component TRS operator. From eq. (2.51), the photonic TRS operator is $\mathcal{T} = \sigma_T \hat{K}$, where $\hat{K}$ is complex conjugation and $\sigma_T = \begin{pmatrix} I_3 & 0 \\ 0 & -I_3 \end{pmatrix}$. Direct verification:

$$\mathcal{T}^2 = \sigma_T \hat{K} \sigma_T \hat{K} = \sigma_T (\sigma_T^*) \hat{K}^2 = \sigma_T \sigma_T \cdot I = I, \quad \implies \quad \boxed{\mathcal{T}^2 = +1 \quad (\text{photons}).} \quad (2.61)$$

Here we used the fact that σ_T is a real matrix ($\sigma_T^* = \sigma_T$) and $\sigma_T^2 = I$ (the squares of the diagonal blocks $\pm I_3$ are all I_3).

Square for spin- $\frac{1}{2}$ electronic systems. For comparison, the electronic TRS operator is $\Theta_e = i\sigma_y \hat{K}$ (spin $\frac{1}{2}$), $\Theta_e^2 = (i\sigma_y)(i\sigma_y)^* \hat{K}^2 = (i\sigma_y)(-i\sigma_y) = \sigma_y^2 = -I$, i.e., $\Theta_e^2 = -1$. This sign difference determines whether Kramers' degeneracy theorem holds:

Kramers' Theorem (holds only when $\Theta^2 = -1$)

If $\Theta^2 = -1$, then for any eigenstate $|\psi\rangle$, $\Theta|\psi\rangle$ is linearly independent of $|\psi\rangle$ and has the same eigenvalue; therefore *every energy level is at least doubly degenerate*.

The photonic $\mathcal{T}^2 = +1$ **does not satisfy** the prerequisite of Kramers' theorem; therefore TRS alone *cannot guarantee* degeneracy of photonic bands. This is one of the most important physical facts in the entire book.

Comparison	Electrons (spin-1/2)	Photons (six-component)	Photons with pseudo-TRS
Θ^2	-1	$+1$	-1 emergent
Kramers degeneracy	✓ protected	× not guaranteed	✓ emergent
AZ symmetry class	AII	AI	AII-like
Topological invariant	$\mathbb{Z}_2$	trivial in 2D stable class AI	emergent $\mathbb{Z}_2$
Protection strength	Strong; requires TRS only	Conditional; ordinary photonic TRS alone does not give Kramers protection	Conditional; requires an additional pseudo-spin symmetry

Systematic comparison.

Remark (Altland–Zirnbauer classification in photonic systems). The AZ tenfold classification is based on three discrete symmetries—TRS (Θ), particle–hole symmetry (Ξ), and chiral symmetry (Π)—applied to the Hamiltonian. When TRS is preserved, photonic systems belong to **class AI** ($\Theta^2 = +1$, no Ξ or Π), whose stable two-dimensional classification is topologically trivial. After TRS is broken, the system belongs to **class A** (no discrete symmetries), which in two dimensions is classified by the integer Chern number. This explains why the magneto-optical scheme in Chapter 5 directly yields a nonzero Chern number, while Chapter 6 must *construct* a pseudo-TRS operator with $\hat{\Theta}^2 = -1$ to obtain $\mathbb{Z}_2$ protection. See the review by Chiu et al. [17] for the complete AZ periodic table.

This discussion reveals the core tension in photonic topology: **nature does not provide a photonic version of Kramers degeneracy**, so every realization of a photonic topological phase must make a unique design at the material or geometric level— this is precisely the subject of the subsequent chapters.

2.9 From Generalized to Standard Eigenvalue Problems: The $\bar{M}^{-1}\hat{N}$ Form

In numerical practice, it is sometimes more convenient to convert eq. (2.16) into a standard eigenvalue problem. If $\bar{M}$ is invertible (which is guaranteed when it is positive definite), define

$$\hat{H}_M \equiv \bar{M}^{-1}\hat{N}, \quad (2.62)$$

so that eq. (2.16) becomes

$$\hat{H}_M \mathbf{f}_n = \omega_n \mathbf{f}_n. \quad (2.63)$$

Proposition 2.18. $\hat{H}_M$ is self-adjoint under $\langle \cdot, \cdot \rangle_M$.

Proof. $\langle \mathbf{f}_1, \hat{H}_M \mathbf{f}_2 \rangle_M = \int \mathbf{f}_1^\dagger \bar{M} (\bar{M}^{-1} \hat{N} \mathbf{f}_2) d^3r = \int \mathbf{f}_1^\dagger \hat{N} \mathbf{f}_2 d^3r = \langle \mathbf{f}_1, \hat{N} \mathbf{f}_2 \rangle_{L^2}$.

Similarly, $\langle \hat{H}_M \mathbf{f}_1, \mathbf{f}_2 \rangle_M = \langle \hat{N} \mathbf{f}_1, \mathbf{f}_2 \rangle_{L^2}$.

By the L^2 self-adjointness of $\hat{N}$ (eq. (2.37)), the two are equal. $\square$

Remark. The famous operator $\hat{\Theta}_H = \nabla \times \frac{1}{\varepsilon(\mathbf{r})} \nabla \times$ in the Joannopoulos textbook [16] is the nonmagnetic special case obtained when only the $\mathbf{H}$ component is retained and there is no bianisotropy. This book uses the six-component form to maintain generality.

2.9.1 Numerical Standardization

In practical numerical computations, directly constructing $\hat{H}_M = \bar{M}^{-1}\hat{N}$ may introduce numerical instability (condition number issues with $\bar{M}^{-1}$). A better strategy exploits the positive definiteness of $\bar{M}$ to perform a **Cholesky decomposition**.

By the Hermitian positive definiteness of $\bar{M}$ (theorem 2.9), there exists a unique lower-triangular matrix L such that

$$\bar{M} = LL^\dagger. \quad (2.64)$$

Left-multiplying both sides of eq. (2.16) by L^{-1} and inserting $L^{-\dagger}L^\dagger$:

$$\underbrace{L^{-1}\hat{N}L^{-\dagger}}_{\hat{H}_{\text{std}}} \underbrace{L^\dagger \mathbf{f}_n}_{\tilde{\mathbf{f}}_n} = \omega_n \tilde{\mathbf{f}}_n. \quad (2.65)$$

The new operator $\hat{H}_{\text{std}}$ is a standard Hermitian matrix, which can be solved by standard Hermitian eigensolvers.

Remark. In practice, generalized Hermitian eigensolver routines usually perform this standardization internally. Manual construction is mainly useful for custom iterative solvers, GPU implementations, or preconditioning studies.

2.10 Energy Inner Product for Dispersive Media: A Preview

All results in this chapter are based on the non-dispersive approximation $\bar{M} \approx \text{const.}$ When material dispersion cannot be neglected, the correct energy density becomes the Brillouin form

$$U = \frac{1}{4} \int d^3r \mathbf{f}^\dagger \frac{\partial(\omega \bar{M})}{\partial \omega} \mathbf{f}, \quad (2.66)$$

and the energy inner product is modified accordingly. A systematic approach is to introduce auxiliary fields, lifting (linearizing) the dispersive system into a higher-dimensional non-dispersive system [18]. See Appendix A for details.

Effective Hermitianization of Dispersive Media: The Auxiliary Field Method

Example 2.19 (Auxiliary field Hermitianization of a Drude dispersive medium). Consider a plasma described by the Drude model: $\varepsilon(\omega) = \varepsilon_\infty - \omega_p^2/(\omega^2 + i\gamma\omega)$. In the frequency domain, ε depends explicitly on ω , causing the eigenvalue problem $\hat{N}\mathbf{f} = \omega \bar{M}(\omega)\mathbf{f}$ to become a **nonlinear eigenvalue problem**— ω appears on both sides of the equation, and the self-adjointness and orthogonality framework established in this chapter no longer directly applies.

Step 1: Introduce an auxiliary polarization field. The physical origin of Drude dispersion is the inertial response of free electrons. In the time domain, the polarization current $\mathbf{J}_p$ satisfies the equation of motion

$$\frac{\partial \mathbf{J}_p}{\partial t} + \gamma \mathbf{J}_p = \varepsilon_0 \omega_p^2 \mathbf{E}. \quad (2.67)$$

Treating $\mathbf{J}_p$ as an independent dynamical degree of freedom (auxiliary field), the state vector of the system is extended from the six-component $\mathbf{f} = (\mathbf{E}, \mathbf{H})^T$ to the nine-component $\tilde{\mathbf{f}} = (\mathbf{E}, \mathbf{H}, \mathbf{J}_p)^T$.

Step 2: Construct a linear GEP in the extended state space. In the frequency domain (lossless limit $\gamma \rightarrow 0$), eq. (2.67) gives $\mathbf{J}_p = \varepsilon_0 \omega_p^2 \mathbf{E}/(-i\omega)$. Define the extended material matrix

$$\tilde{M} = \begin{pmatrix} \varepsilon_\infty \varepsilon_0 & 0 & 0 \\ 0 & \mu_0 & 0 \\ 0 & 0 & \varepsilon_0^{-1} \omega_p^{-2} \end{pmatrix}, \quad \tilde{N} = \begin{pmatrix} 0 & i\nabla \times & -i \\ -i\nabla \times & 0 & 0 \\ i & 0 & 0 \end{pmatrix}. \quad (2.68)$$

The extended GEP is

$$\boxed{\tilde{N} \tilde{\mathbf{f}} = \omega \tilde{M} \tilde{\mathbf{f}}}, \quad (2.69)$$

which is a **linear** eigenvalue problem in ω !

Step 3: Self-adjointness is restored. In the lossless limit $\gamma = 0$, $\tilde{M}$ is a real, diagonal, positive-definite matrix ($\varepsilon_\infty > 0$, $\mu_0 > 0$, $\omega_p^{-2} > 0$), and $\tilde{N}$ is self-adjoint under the standard L^2 inner product. Equivalently, $\tilde{M}^{-1} \tilde{N}$ is self-adjoint under the extended energy inner product $\langle \tilde{\mathbf{f}}_1 | \tilde{\mathbf{f}}_2 \rangle_{\tilde{M}} = \int \tilde{\mathbf{f}}_1^\dagger \tilde{M} \tilde{\mathbf{f}}_2 d^3r$: $\tilde{N}^\dagger = \tilde{N}$ (can be verified directly). Therefore, **all theorems of this chapter (real spectrum, orthogonality, completeness) are fully restored in the extended state space.**

Step 4: Physical meaning and generalizations.

- The introduction of the auxiliary field $\mathbf{J}_p$ converts dispersive information from “frequency-dependent material parameters” into “additional dynamical degrees of freedom”—mathematically, this is **linearization**.
- For a Lorentz oscillator $\varepsilon(\omega) = \varepsilon_\infty + \omega_p^2/(\omega_0^2 - \omega^2 - i\gamma\omega)$, one needs to introduce the auxiliary polarization $\mathbf{P}$ and polarization velocity $\dot{\mathbf{P}}$ as two fields, extending the state space to twelve components $(\mathbf{E}, \mathbf{H}, \mathbf{P}, \dot{\mathbf{P}})^T$.
- Multiple resonances $\sum_j$ correspond to multiple auxiliary fields—the state space dimension $= 6 + 3N_{\text{poles}}$.
- In the lossy case $\gamma > 0$, $\tilde{N}$ acquires an anti-Hermitian correction $\propto i\gamma$ —self-adjointness is broken, which is precisely the starting point of non-Hermitian topological theory in Chapter 9.

Remark (Silveirinha’s auxiliary field theory). The above construction was first systematized by Silveirinha [18], who proved that as long as the dispersion function $\varepsilon(\omega)$ is a rational function (i.e., can be expanded in a finite number of poles), the auxiliary field method can exactly linearize the nonlinear EVP into a standard GEP, and the topological invariants (Chern number, Zak phase, etc.) of the extended system are rigorously consistent with physical observables.

Chapter Summary

Starting from the time-domain Maxwell equations, this chapter has progressively established a rigorous frequency-domain operator theory framework. The central result is the lossless generalized eigenvalue problem

$$\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}, \quad \hat{N} = \begin{pmatrix} 0 & i\nabla \times \\ -i\nabla \times & 0 \end{pmatrix}, \quad \mathbf{f} = (\mathbf{E}, \mathbf{H})^T.$$

When $\bar{M} = \bar{M}^\dagger > 0$ and the boundary terms vanish, $\hat{H}_M = \bar{M}^{-1}\hat{N}$ is self-adjoint under the material-weighted inner product $\langle \mathbf{f}, \mathbf{g} \rangle_M = \int \mathbf{f}^\dagger \bar{M} \mathbf{g} d^3r$. This gives real frequencies, energy orthogonality, and a stable foundation for Berry geometry.

The material matrix also determines symmetry. Losslessness uses the Hermitian condition $\bar{M} = \bar{M}^\dagger$; Lorentz reciprocity uses transpose conditions on $\boldsymbol{\varepsilon}, \boldsymbol{\mu}, \bar{\boldsymbol{\xi}}, \bar{\boldsymbol{\zeta}}$; time reversal uses the anti-unitary action $\mathcal{T}\mathbf{f} = \sigma_T \mathbf{f}^*$ and the condition $\sigma_T \bar{M}^* \sigma_T = \bar{M}$. These are distinct notions, and confusing reciprocity with TRS leads to wrong conclusions about Chern topology.

Finally, the six-component formalism is not decorative. It is what makes the inner product, bianisotropic materials, positive/negative-frequency pairing, and dispersive auxiliary-field constructions fit into a single operator framework. Bloch theory, Berry curvature, Chern numbers, pseudo-spin topology, and non-Hermitian extensions all build on this structure.

Chapter 3

Bloch's Theorem and Photonic Band Theory

Why Periodicity Changes the Spectral Problem?

In Chapter 2, we established the complete framework of the Maxwell generalized eigenvalue problem $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$: self-adjointness guarantees real eigenfrequencies, the energy inner product guarantees orthogonal completeness, and reciprocity and TRS provide symmetry constraints on the spectrum. However, that chapter dealt with **general materials**—no assumptions were made about the spatial structure of $\bar{M}(\mathbf{r})$.

This chapter introduces a key assumption: $\bar{M}(\mathbf{r})$ **has spatial periodicity**. This assumption has extremely far-reaching consequences: it decomposes a continuous-spectrum problem on an infinite-dimensional function space into a family of **discrete-spectrum problems on a compact domain**, parameterized by the wave vector $\mathbf{k}$. This decomposition—Bloch's theorem—is the prerequisite for all band theory and topological classification.

- **From continuous spectrum to discrete spectrum:** The eigenfrequencies of the original problem form a continuous set (imagine the free photon $\omega = c|\mathbf{k}|/\sqrt{\varepsilon}$ in a homogeneous medium); discussing topological invariants directly on all of space is meaningless. The Bloch decomposition reduces the problem to a discrete spectrum on the unit cell Ω , making “the n -th energy band” a well-defined concept—this is the prerequisite for defining the Berry connection and Chern number.
- **From real space to momentum space:** Periodicity introduces the Brillouin zone BZ as a parameter space. BZ itself is a compact manifold (a torus), and the vector bundle defined on it—the Bloch bundle— carries topological invariants. Without BZ, there is no Berry geometry.
- **Origin of gauge freedom:** The phase of the cell function $\mathbf{u}_{n\mathbf{k}}$ can be chosen independently at each $\mathbf{k}$ — this “redundant freedom” is precisely the origin of the Berry connection (i.e., the “virtual vector potential”). When the “magnetic flux” of this vector potential (the Chern number) is nonzero, no globally smooth gauge can be found—this is the mathematical root of topological nontriviality.

Remark (**Three common misconceptions**).

Misconception one: “Bloch’s theorem is just borrowed from solid-state physics.”

The translation-group argument is shared with electronic band theory, but the operator is different: photonics uses the six-component generalized Maxwell problem $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$, whose self-adjointness comes from the curl operator and the material-weighted inner product.

Misconception two: “The bandgap depends only on dielectric contrast.” Contrast matters, but so do connectivity, lattice symmetry, filling fraction, and polarization. In 2D, for example, TM and TE gaps need not overlap.

Misconception three: “The band diagram contains all topological information.”

High-symmetry-line plots are only one-dimensional slices through BZ. Topology depends on the phase evolution of $\mathbf{u}_{n\mathbf{k}}$ over the whole Brillouin zone, not only on the eigenfrequencies.

We now build this reduction in four steps: define the lattice and its reciprocal space, prove the photonic Bloch theorem, derive the cell-periodic Maxwell operator $\hat{N}(\mathbf{k})$, and use it to describe bands, gaps, gauges, and low-energy Dirac models.

3.1 Photonic Crystals and Lattice Structure

3.1.1 Mathematical Formulation of Spatial Periodicity

Let the material matrix $\bar{M}(\mathbf{r})$ of the medium satisfy discrete translational symmetry: there exist d linearly independent basis vectors $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d$ ($d = 1, 2$, or 3) such that

$$\bar{M}(\mathbf{r} + \mathbf{R}) = \bar{M}(\mathbf{r}), \quad \forall \mathbf{R} = \sum_{i=1}^d n_i \mathbf{a}_i, \quad n_i \in \mathbb{Z}. \quad (3.1)$$

The set $\{\mathbf{R}\}$ constitutes the **Bravais lattice** Λ . Due to the periodicity of the material matrix, the curl operator $\hat{N}$ itself is position-independent, while the material weight $\bar{M}(\mathbf{r})$ in the generalized eigenvalue problem eq. (2.16) (Chapter 2) is a periodic function.

Definition 3.1 (Unit cell). The **Wigner–Seitz unit cell** Ω is defined as the set of points closer to the origin than to any other lattice point (or equidistant):

$$\Omega = \{\mathbf{r} \mid |\mathbf{r}| \leq |\mathbf{r} - \mathbf{R}|, \forall \mathbf{R} \in \Lambda \setminus \{0\}\}. \quad (3.2)$$

In numerical computations, the equivalent parallelepiped unit cell $\Omega_{\text{para}} = \{s_1 \mathbf{a}_1 + s_2 \mathbf{a}_2 + s_3 \mathbf{a}_3 \mid 0 \leq s_i < 1\}$ is commonly used.

3.1.2 Reciprocal Lattice and Brillouin Zone

Define the **reciprocal lattice basis vectors** $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_d$ satisfying

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij}. \quad (3.3)$$

The reciprocal lattice vectors $\mathbf{G} = \sum_i m_i \mathbf{b}_i$ ($m_i \in \mathbb{Z}$) constitute the **reciprocal lattice** Λ^* .

Definition 3.2 (First Brillouin zone). The first Brillouin zone BZ is the Wigner–Seitz cell of the reciprocal lattice:

$$\text{BZ} = \{\mathbf{k} \mid |\mathbf{k}| \leq |\mathbf{k} - \mathbf{G}|, \forall \mathbf{G} \in \Lambda^* \setminus \{0\}\}. \quad (3.4)$$

Example 3.3 (Two-dimensional triangular lattice). Let the lattice constant be a , with basis vectors $\mathbf{a}_1 = a(1, 0)$, $\mathbf{a}_2 = a(1/2, \sqrt{3}/2)$. From eq. (3.3),

$$\mathbf{b}_1 = \frac{2\pi}{a}(1, -1/\sqrt{3}), \quad \mathbf{b}_2 = \frac{2\pi}{a}(0, 2/\sqrt{3}). \quad (3.5)$$

The first Brillouin zone is a regular hexagon, with high-symmetry points $\Gamma = (0, 0)$, $K = (4\pi/3a, 0)$, $M = (\pi/a, \pi/\sqrt{3}a)$.

Example 3.4 (Two-dimensional square lattice). Basis vectors $\mathbf{a}_1 = a(1, 0)$, $\mathbf{a}_2 = a(0, 1)$. Reciprocal lattice basis vectors $\mathbf{b}_1 = (2\pi/a)(1, 0)$, $\mathbf{b}_2 = (2\pi/a)(0, 1)$. The Brillouin zone is the square $[-\pi/a, \pi/a]^2$, with high-symmetry points $\Gamma = (0, 0)$, $X = (\pi/a, 0)$, $M = (\pi/a, \pi/a)$.

3.2 Translation Operators and Bloch's Theorem

3.2.1 Definition and Properties of Translation Operators

For any lattice vector $\mathbf{R} \in \Lambda$, define the **translation operator** $\hat{T}_{\mathbf{R}}$:

$$(\hat{T}_{\mathbf{R}} \mathbf{f})(\mathbf{r}) = \mathbf{f}(\mathbf{r} - \mathbf{R}). \quad (3.6)$$

Proposition 3.5. *The translation operators have the following properties:*

- (i) **Group structure:** $\hat{T}_{\mathbf{R}_1} \hat{T}_{\mathbf{R}_2} = \hat{T}_{\mathbf{R}_1 + \mathbf{R}_2}$, $\hat{T}_0 = \hat{I}$, $\hat{T}_{\mathbf{R}}^{-1} = \hat{T}_{-\mathbf{R}}$.
- (ii) **Unitarity:** $\hat{T}_{\mathbf{R}}$ is unitary under the energy inner product $\langle \cdot, \cdot \rangle_M$.
- (iii) **Commutation with the Maxwell operator:** $[\hat{T}_{\mathbf{R}}, \hat{N}] = 0$ and $\hat{T}_{\mathbf{R}} \bar{M}(\mathbf{r}) \hat{T}_{\mathbf{R}}^{-1} = \bar{M}(\mathbf{r})$.

Proof. (i) Follows directly from the definition.

(ii) We need to show $\langle \hat{T}_{\mathbf{R}} \mathbf{f}_1, \hat{T}_{\mathbf{R}} \mathbf{f}_2 \rangle_M = \langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M$.

$$\begin{aligned} \langle \hat{T}_{\mathbf{R}} \mathbf{f}_1, \hat{T}_{\mathbf{R}} \mathbf{f}_2 \rangle_M &= \int_{\mathbb{R}^d} \mathbf{f}_1^\dagger(\mathbf{r} - \mathbf{R}) \bar{M}(\mathbf{r}) \mathbf{f}_2(\mathbf{r} - \mathbf{R}) d^d r \\ &= \int_{\mathbb{R}^d} \mathbf{f}_1^\dagger(\mathbf{r}') \bar{M}(\mathbf{r}' + \mathbf{R}) \mathbf{f}_2(\mathbf{r}') d^d r' \quad (\mathbf{r}' = \mathbf{r} - \mathbf{R}) \\ &= \int_{\mathbb{R}^d} \mathbf{f}_1^\dagger(\mathbf{r}') \bar{M}(\mathbf{r}') \mathbf{f}_2(\mathbf{r}') d^d r' = \langle \mathbf{f}_1, \mathbf{f}_2 \rangle_M, \end{aligned} \quad (3.7)$$

where the third equality uses the periodicity of $\bar{M}$ (eq. (3.1)).

(iii) $\hat{N}$ contains only $\nabla \times$; acting on $\mathbf{f}(\mathbf{r} - \mathbf{R})$ is equivalent to $(\hat{N} \mathbf{f})(\mathbf{r} - \mathbf{R})$, so $\hat{T}_{\mathbf{R}} \hat{N} = \hat{N} \hat{T}_{\mathbf{R}}$. Also, $\hat{T}_{\mathbf{R}} \bar{M}(\mathbf{r}) \hat{T}_{\mathbf{R}}^{-1} \mathbf{f}(\mathbf{r}) = \bar{M}(\mathbf{r} - \mathbf{R}) \mathbf{f}(\mathbf{r}) = \bar{M}(\mathbf{r}) \mathbf{f}(\mathbf{r})$. $\square$

3.2.2 Bloch's Theorem

By property (iii), the generalized eigenvalue problem $\hat{N} \mathbf{f} = \omega \bar{M} \mathbf{f}$ commutes with all translation operators $\{\hat{T}_{\mathbf{R}}\}$. Since $\{\hat{T}_{\mathbf{R}}\}$ forms an Abelian group, they can be simultaneously diagonalized.

Theorem 3.6 (Photonic Bloch theorem). *Let $\bar{M}(\mathbf{r} + \mathbf{R}) = \bar{M}(\mathbf{r})$ hold for all $\mathbf{R} \in \Lambda$. Then the eigenfields $\mathbf{f}(\mathbf{r})$ of the generalized eigenvalue problem*

$$\hat{N} \mathbf{f}(\mathbf{r}) = \omega \bar{M}(\mathbf{r}) \mathbf{f}(\mathbf{r}) \quad (3.8)$$

can be chosen to simultaneously satisfy the Bloch condition:

$$\mathbf{f}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_{n\mathbf{k}}(\mathbf{r}), \quad (3.9)$$

*where $\mathbf{k} \in \text{BZ}$ is the **Bloch wave vector**, n is the **band index**, and $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ is a cell-periodic function with lattice periodicity:*

$$\mathbf{u}_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = \mathbf{u}_{n\mathbf{k}}(\mathbf{r}), \quad \forall \mathbf{R} \in \Lambda. \quad (3.10)$$

Proof. Step 1: Since $\hat{T}_{\mathbf{R}}$ is unitary and all translation operators commute (Abelian group), by the spectral theorem they have a common eigenbasis. Let $\mathbf{f}$ be simultaneously a generalized eigenfunction of $\hat{N}$ and an eigenfunction of all $\hat{T}_{\mathbf{R}}$:

$$\hat{T}_{\mathbf{R}} \mathbf{f} = \lambda(\mathbf{R}) \mathbf{f}. \quad (3.11)$$

Step 2: By the group homomorphism property $\hat{T}_{\mathbf{R}_1 + \mathbf{R}_2} = \hat{T}_{\mathbf{R}_1} \hat{T}_{\mathbf{R}_2}$, the eigenvalues satisfy

$$\lambda(\mathbf{R}_1 + \mathbf{R}_2) = \lambda(\mathbf{R}_1) \lambda(\mathbf{R}_2). \quad (3.12)$$

Furthermore, by unitarity $|\lambda(\mathbf{R})| = 1$. Therefore λ is a group homomorphism from Λ to $U(1)$. All $U(1)$ characters of $\Lambda \cong \mathbb{Z}^d$ are given by

$$\lambda(\mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}} \quad (3.13)$$

where $\mathbf{k}$ is uniquely determined modulo Λ^* (i.e., $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ give the same character).

Step 3: From eq. (3.11) and eq. (3.13), $\mathbf{f}(\mathbf{r} - \mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}} \mathbf{f}(\mathbf{r})$, i.e., $\mathbf{f}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \mathbf{f}(\mathbf{r})$. Defining $\mathbf{u}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{f}(\mathbf{r})$, we get

$$\mathbf{u}(\mathbf{r} + \mathbf{R}) = e^{-i\mathbf{k}\cdot(\mathbf{r} + \mathbf{R})} \mathbf{f}(\mathbf{r} + \mathbf{R}) = e^{-i\mathbf{k}\cdot(\mathbf{r} + \mathbf{R})} e^{i\mathbf{k}\cdot\mathbf{R}} \mathbf{f}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{f}(\mathbf{r}) = \mathbf{u}(\mathbf{r}). \quad (3.14)$$

Therefore $\mathbf{u}$ is a periodic function, and $\mathbf{f} = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}$ is in Bloch form. $\square$

Remark. The above proof is carried out entirely within the framework of the Maxwell generalized eigenvalue problem, without resorting to the Schrödinger equation or tight-binding models. The only key inputs are: (1) the periodicity of $\bar{M}$; (2) the unitarity and Abelian group structure of $\hat{T}_{\mathbf{R}}$.

Remark. Unlike the electronic Bloch theorem, the Bloch states $\mathbf{f}_{n\mathbf{k}}$ here are six-component vector fields; each component $(E_x, E_y, E_z, H_x, H_y, H_z)$ satisfies the phase factor of eq. (3.9).

3.2.3 Why Does Periodicity Necessarily Lead to Energy Bands?

Before entering the general framework of the cell-periodic operator, it is worth pausing to understand from a physical perspective the core question of “why periodicity leads to energy

bands.” Beginners often view energy bands as a purely mathematical result—a corollary of Bloch's theorem—while overlooking the simple yet profound physical picture behind them.

Homogeneous medium: Free photons. In a homogeneous isotropic medium ($\varepsilon(\mathbf{r}) \equiv \varepsilon_0$, $\mu(\mathbf{r}) \equiv \mu_0$), the eigenmodes of Maxwell's equations are plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}$, with dispersion relation

$$\omega(\mathbf{k}) = \frac{c}{n}|\mathbf{k}|, \quad n = \sqrt{\varepsilon_0\mu_0}, \quad (3.15)$$

which is a light cone with its apex at the origin—continuous, isotropic, and without bandgaps. “Folding” this dispersion relation into the first Brillouin zone (i.e., replacing $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{G}$ with the smallest $|\mathbf{G}|$) produces multiple “folded bands” that cross at the Brillouin zone boundary. These crossing points are **accidental degeneracies**—in the absence of periodic modulation, they will not be opened.

Introducing periodicity: Bragg scattering opens bandgaps. Once $\varepsilon(\mathbf{r})$ acquires spatial periodicity $\varepsilon(\mathbf{r} + \mathbf{R}) = \varepsilon(\mathbf{r})$, the cell function $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ of the Bloch states is no longer constant but varies with $\mathbf{r}$ within the unit cell. At the Brillouin zone boundary, the wave vectors $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ are degenerate (due to folding), and the Fourier component $\hat{\varepsilon}_{\mathbf{G}}$ of the periodic dielectric constant couples these two originally degenerate plane waves. This is equivalent to a 2×2 perturbation:

$$\begin{pmatrix} \omega_{\mathbf{k}}^{(0)} & V_{\mathbf{G}} \\ V_{\mathbf{G}}^* & \omega_{\mathbf{k}+\mathbf{G}}^{(0)} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \omega \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad (3.16)$$

where $\omega_{\mathbf{k}}^{(0)} = \omega_{\mathbf{k}+\mathbf{G}}^{(0)}$ (degeneracy condition) and the coupling matrix element $V_{\mathbf{G}}$ is proportional to $\hat{\varepsilon}_{\mathbf{G}}$. As long as $V_{\mathbf{G}} \neq 0$, the degeneracy is split into

$$\omega_{\pm} = \omega_{\mathbf{k}}^{(0)} \pm |V_{\mathbf{G}}|, \quad (3.17)$$

which is the **Bragg bandgap**.

This picture yields three key insights:

1. **The bandgap width is proportional to the Fourier component $\hat{\varepsilon}_{\mathbf{G}}$:** The larger the dielectric contrast, the larger $\hat{\varepsilon}_{\mathbf{G}}$, and the wider the bandgap. This explains why high-contrast photonic crystals (such as silicon/air systems) readily open complete bandgaps.
2. **The bandgap position is determined by the Brillouin zone geometry:** The position of the degeneracy points depends on the geometric relationship of the crossing between $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$, i.e., the shape of the Brillouin zone boundary. This explains why different lattice types (square, triangular, honeycomb) differ in the ease of opening bandgaps.
3. **Not all degeneracies will be opened:** If $\hat{\varepsilon}_{\mathbf{G}} = 0$ (i.e., that Fourier component happens to vanish), or if symmetry protects the degeneracy (the dimension of the irreducible representation is greater than 1), then the bandgap will not open—this is the physical origin of **symmetry-protected Dirac cones** (see section 3.9 for details).

Remark. This Bragg picture is structurally completely parallel to the “Nearly Free Electron” (NFE) model in condensed matter physics. However, two differences must be noted: (1) the “free dispersion” of photons is linear $\omega \propto |\mathbf{k}|$ (rather than the electronic parabola $E \propto k^2$), so

the behavior of the photonic band structure near the Γ point is fundamentally different from that of electrons; (2) photons are vector fields, and the polarization degree of freedom leads to additional degeneracies and splittings—this is particularly evident in the bandgap differences between TM and TE modes (the 1D benchmark below isolates the simpler Bragg mechanism).

Example 3.7 (Photonic Kronig–Penney model). The one-dimensional layered structure is the simplest and most complete model for understanding photonic energy bands. Consider a periodically layered medium along the z direction:

$$\varepsilon(z) = \begin{cases} \varepsilon_1, & 0 < z < d_1 \\ \varepsilon_2, & d_1 < z < a \end{cases}, \quad \mu(z) \equiv 1, \quad a = d_1 + d_2. \quad (3.18)$$

Starting from the six-component framework of Chapter 2, we systematically reduce to the transfer matrix method.

Step 1: From six components to the TM scalar equation. In Chapter 2, we already proved that for a one-dimensional structure with $\partial_x = \partial_y = 0$ and $\mu = 1$, TM polarization (E_z nonzero, $H_z = 0$) satisfies the scalar equation

$$-\frac{d^2 E_z}{dz^2} = \frac{\omega^2}{c^2} \varepsilon(z) E_z. \quad (3.19)$$

This is a one-dimensional Sturm–Liouville problem, with $\varepsilon(z)$ playing the role of a “potential function” (note: not a “potential well”—there is no sign reversal in the photonic equation).

Step 2: General solution in each layer. In layer j where $\varepsilon = \varepsilon_j$, the equation has constant coefficients:

$$E_z''(z) + k_j^2 E_z(z) = 0, \quad k_j \equiv \frac{\omega}{c} \sqrt{\varepsilon_j}. \quad (3.20)$$

The general solution is $E_z^{(j)}(z) = A_j e^{ik_j z} + B_j e^{-ik_j z}$. The corresponding magnetic field component is given by $H_y = -\frac{1}{i\omega\mu_0} \frac{dE_z}{dz}$:

$$H_y^{(j)}(z) = \frac{k_j}{\omega\mu_0} (A_j e^{ik_j z} - B_j e^{-ik_j z}). \quad (3.21)$$

Step 3: Transfer matrix. E_z and H_y are continuous at interfaces (Maxwell boundary conditions). Define the state vector $\mathbf{s}(z) = (E_z(z), H_y(z))^T$; then propagation through layer j of thickness d_j corresponds to the transfer matrix

$$T_j = \begin{pmatrix} \cos(k_j d_j) & -\frac{i\omega\mu_0}{k_j} \sin(k_j d_j) \\ -\frac{ik_j}{\omega\mu_0} \sin(k_j d_j) & \cos(k_j d_j) \end{pmatrix}. \quad (3.22)$$

Note that $\det T_j = 1$ (energy flux conservation), and $T_j \in \text{SL}(2, \mathbb{R})$ for real frequencies. The transfer matrix for one complete period is

$$T_{\text{cell}} = T_2 T_1. \quad (3.23)$$

Step 4: Bloch condition and dispersion relation. Bloch's theorem requires $\mathbf{s}(z+a) = e^{iKa} \mathbf{s}(z)$, where K is the Bloch wave vector. Therefore e^{iKa} is an eigenvalue of T_{cell} . Since

$\det T_{\text{cell}} = 1$, the two eigenvalues are $e^{\pm iKa}$, so $\text{tr}(T_{\text{cell}}) = 2 \cos(Ka)$. Expanding gives

$$\cos(Ka) = \cos(k_1 d_1) \cos(k_2 d_2) - \frac{1}{2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \sin(k_1 d_1) \sin(k_2 d_2). \quad (3.24)$$

This is the Kronig–Penney dispersion relation for photonic crystals (compare with the δ -potential Kronig–Penney model for electrons: the photonic version has no δ potential but rather a square-wave dielectric function).

Physical analysis:

- When $|\text{RHS}| \leq 1$, Ka is real—the Bloch wave propagates, corresponding to an **energy band**.
- When $|\text{RHS}| > 1$, Ka becomes complex—the Bloch wave is evanescent, corresponding to a **bandgap**.
- Bandgaps appear near $k_1 d_1 + k_2 d_2 = m\pi$ (m a positive integer), i.e., when the round-trip optical path in one period is an integer multiple of half a wavelength—this is the Bragg condition.
- When $\varepsilon_1 = \varepsilon_2$, $k_1 = k_2$, and the RHS simplifies to $\cos(k_1 a)$, yielding no bandgap—the free-photon dispersion of a homogeneous medium.
- The bandgap width is proportional to $|k_1/k_2 - k_2/k_1| = |\sqrt{\varepsilon_1/\varepsilon_2} - \sqrt{\varepsilon_2/\varepsilon_1}|$; hence the stronger the dielectric contrast, the stronger the Bragg scattering, and typically the wider the bandgap.

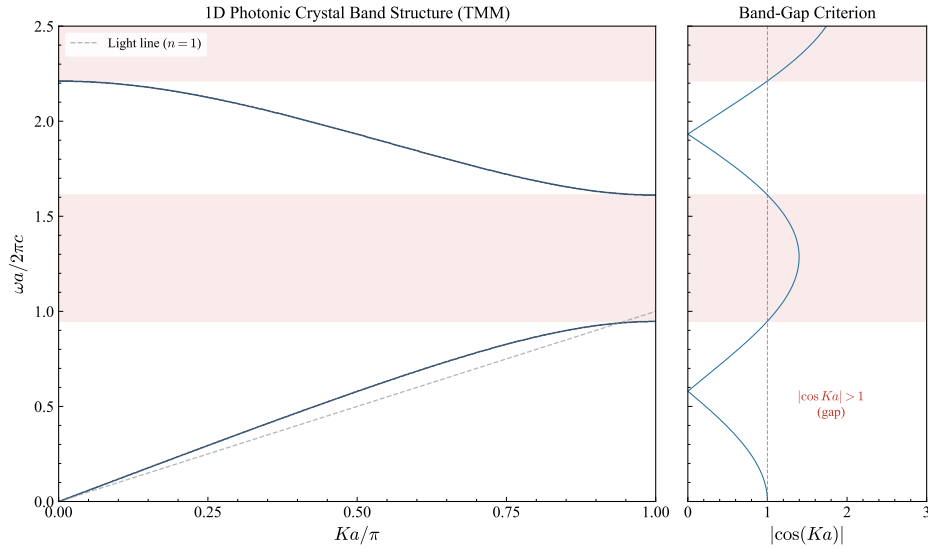


Figure 3.1 Kronig–Penney dispersion relation of a 1D layered photonic crystal. The left panel shows the pass bands corresponding to allowed Bloch wave vectors; shaded regions correspond to forbidden bands where $|\cos(Ka)| > 1$. The right panel directly plots the dispersion criterion $\cos(Ka)$, linking the “pass band/bandgap” classification to the eigenvalue structure of the transfer matrix. Unlike the parabolic folding in electronic systems, the free dispersion of photonic systems is approximately linear; the crossings formed after folding are opened into bandgaps by Bragg coupling.

Remark. The power of the transfer matrix method lies in its ability to handle **any** one-dimensional periodic structure— one simply multiplies the transfer matrices of each layer in sequence. For

continuously varying $\varepsilon(z)$, the profile can be discretized into a multilayer step function and then taken to the limit (equivalent to numerically solving the ODE). However, the transfer matrix method **cannot** be generalized to two and three dimensions—this is precisely why general methods such as the plane-wave expansion (PWE) or finite element method (FEM) are needed (see Appendix C).

From a concrete example to the general framework. The one-dimensional model above demonstrates all the physics of photonic band formation: periodicity $\rightarrow$ Bloch condition $\rightarrow$ folding $\rightarrow$ Bragg coupling $\rightarrow$ bandgap. In more general two- and three-dimensional cases, although there are no longer analytical transfer matrix formulas, the physical picture is exactly the same—the only differences are that the Brillouin zone geometry is more complex and the polarization degree of freedom introduces more mode coupling and symmetry constraints. With this physical intuition in hand, we now enter the general mathematical framework of the cell-periodic Maxwell operator.

3.3 The Cell-Periodic Maxwell Operator $\hat{N}(\mathbf{k})$

3.3.1 Substituting the Bloch Form into Maxwell's Equations

Substituting eq. (3.9) into $\hat{N}\mathbf{f} = \omega\bar{M}\mathbf{f}$. Note that when ∇ acts on $e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}(\mathbf{r})$,

$$\nabla \times (e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{u}) = e^{i\mathbf{k}\cdot\mathbf{r}}(\nabla + i\mathbf{k}) \times \mathbf{u}. \quad (3.25)$$

Therefore, the generalized eigenvalue problem becomes

$$\boxed{\hat{N}(\mathbf{k}) \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) = \omega_{n\mathbf{k}} \bar{M}(\mathbf{r}) \mathbf{u}_{n\mathbf{k}}(\mathbf{r})}, \quad (3.26)$$

where $\hat{N}(\mathbf{k})$ is the **cell-periodic Maxwell operator**:

$$\hat{N}(\mathbf{k}) = \begin{pmatrix} 0 & (\nabla + i\mathbf{k}) \times \\ -(\nabla + i\mathbf{k}) \times & 0 \end{pmatrix}. \quad (3.27)$$

Remark. The domain of $\hat{N}(\mathbf{k})$ is the space of **periodic functions** on the unit cell Ω , $L^2_{\text{per}}(\Omega, \mathbb{C}^6)$, rather than the full space $L^2(\mathbb{R}^d, \mathbb{C}^6)$. This is a crucial distinction—the original problem on the infinite-dimensional full space has a continuous spectrum, while the cell-periodic problem on the compact domain has a discrete spectrum.

3.3.2 Cell-Periodic Energy Inner Product

On the cell-periodic space $L^2_{\text{per}}(\Omega, \mathbb{C}^6)$, define the **cell-periodic energy inner product**:

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle_{M\Omega} = \int_{\Omega} \mathbf{u}_1^\dagger(\mathbf{r}) \bar{M}(\mathbf{r}) \mathbf{u}_2(\mathbf{r}) d^d r. \quad (3.28)$$

Proposition 3.8 (Self-adjointness of the cell-periodic operator). *Under $\langle \cdot, \cdot \rangle_{M\Omega}$, $\hat{N}(\mathbf{k})$ is self-adjoint for each $\mathbf{k} \in \text{BZ}$ (when $\bar{M} = \bar{M}^\dagger$).*

Proof. It remains to prove the symmetry of $\bar{M}^{-1}\hat{N}(\mathbf{k})$ with respect to the material-weighted inner product:

$$\langle \mathbf{u}_1, \bar{M}^{-1}\hat{N}(\mathbf{k})\mathbf{u}_2 \rangle_{M\Omega} = \langle \bar{M}^{-1}\hat{N}(\mathbf{k})\mathbf{u}_1, \mathbf{u}_2 \rangle_{M\Omega}. \quad (3.29)$$

Equivalently, since the factor $\bar{M}$ in the inner product cancels the $\bar{M}^{-1}$ in the operator, this is the $L^2(\Omega)$ symmetry of $\hat{N}(\mathbf{k})$:

$$\langle \mathbf{u}_1, \hat{N}(\mathbf{k})\mathbf{u}_2 \rangle_{L^2(\Omega)} = \langle \hat{N}(\mathbf{k})\mathbf{u}_1, \mathbf{u}_2 \rangle_{L^2(\Omega)}. \quad (3.30)$$

Consider the upper off-diagonal block $(\nabla + i\mathbf{k}) \times$ of $\hat{N}(\mathbf{k})$:

$$\int_{\Omega} \mathbf{u}_1^\dagger \cdot [(\nabla + i\mathbf{k}) \times \mathbf{u}_2] \, d^d r = \int_{\Omega} [(\nabla - i\mathbf{k}) \times \mathbf{u}_1]^\dagger \cdot \mathbf{u}_2 \, d^d r + \oint_{\partial\Omega} (\cdots) \, dS. \quad (3.31)$$

Periodic boundary conditions ensure that the surface integral over $\partial\Omega$ cancels (the values of $\mathbf{u}$ are the same on opposite faces, but the outward normals are opposite). Moreover, $(\nabla - i\mathbf{k}) \times = [(\nabla + i\mathbf{k}) \times]^\dagger$ holds in L^2_{per} (when $\mathbf{k}$ is a real vector, the sign change from conjugating $i\mathbf{k}$ exactly compensates), so $\hat{N}(\mathbf{k})$ is self-adjoint. $\square$

3.3.3 Discrete Spectrum of the Cell-Periodic Problem

Theorem 3.9 (Discrete eigenspectrum). *For each fixed $\mathbf{k} \in \text{BZ}$, the cell-periodic generalized eigenvalue problem eq. (3.26) has countably infinitely many real eigenfrequencies*

$$\cdots \leq \omega_{-2,\mathbf{k}} \leq \omega_{-1,\mathbf{k}} < 0 < \omega_{1,\mathbf{k}} \leq \omega_{2,\mathbf{k}} \leq \cdots, \quad (3.32)$$

and the corresponding eigenfunctions $\{\mathbf{u}_{n\mathbf{k}}\}$ form a complete orthonormal set in $L^2_{\text{per}}(\Omega, \mathbb{C}^6)$ under the inner product $\langle \cdot, \cdot \rangle_{M\Omega}$.

Proof sketch. Ω is a compact domain, and $(\nabla + i\mathbf{k}) \times$ under periodic boundary conditions has its domain in a subspace of the Sobolev space $H^1_{\text{per}}(\Omega)$. The standard spectral theory of self-adjoint operators with compact resolvent (see [19] and PDE textbooks) guarantees a discrete real spectrum and complete orthogonality. The positive definiteness of $\bar{M}$ ensures the eigenfrequencies are real and can be divided into positive and negative branches. $\square$

Remark. The negative-frequency branch $\omega_{n,\mathbf{k}} < 0$ is not physically independent; it is the “pairing” of the positive-frequency solutions under the $e^{-i\omega t}$ convention: if $(\omega, \mathbf{f})$ is a solution, then $(-\omega, \mathbf{f}^*)$ is also a solution. In subsequent discussions of topological invariants in this book, we usually focus only on the positive-frequency branch. However, in non-Hermitian systems the division into positive and negative branches may become anomalous—this is one of the topics of Chapter 9.

Quasi-Momentum Conservation

Discrete translation symmetry makes $\mathbf{k}$ a good quantum number, but only modulo a reciprocal lattice vector. Scattering in a periodic medium therefore obeys

$$\mathbf{k}_{\text{incident}} = \mathbf{k}_{\text{scattered}} + \mathbf{G}, \quad \mathbf{G} \in \Lambda^*. \quad (3.33)$$

Processes with $\mathbf{G} \neq 0$ are Umklapp processes. The reduced value of $\mathbf{k}$ inside the first Brillouin zone is still well defined, which is what makes $\omega_n(\mathbf{k})$ a band function and makes the Berry connection in Chapter 4 a function on BZ.

Remark. Photonic bands differ from electronic bands in two practical ways. First, the free-photon dispersion is linear rather than parabolic. Second, Maxwell equations naturally contain positive- and negative-frequency partners. These features matter for symmetry classification, but the organizing parameter of the periodic problem is still the Bloch quasi-momentum $\mathbf{k}$.

3.4 Photonic Band Structure

3.4.1 Dispersion Relations $\omega_n(\mathbf{k})$

As $\mathbf{k}$ varies continuously within BZ, each eigenfrequency $\omega_{n,\mathbf{k}}$ as a function of $\mathbf{k}$ constitutes the n -th **photonic energy band** $\omega_n(\mathbf{k})$. The collection of all bands $\{\omega_n(\mathbf{k})\}_{n \geq 1}$ is called the **photonic band structure**.

Definition 3.10 (Analyticity of bands). For a simple (non-degenerate) eigenvalue $\omega_n(\mathbf{k}_0)$, perturbation theory guarantees that $\omega_n(\mathbf{k})$ is an analytic function of $\mathbf{k}$ in a neighborhood of $\mathbf{k}_0$. At degenerate points, band crossings or anti-crossings may occur, whose classification is determined by higher-order terms in perturbation theory.

3.4.2 Band Diagrams Along High-Symmetry Lines

In practice, $\omega_n(\mathbf{k})$ is typically plotted along high-symmetry paths on the boundary of the first Brillouin zone, for example:

- Square lattice: $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$
- Triangular lattice: $\Gamma \rightarrow M \rightarrow K \rightarrow \Gamma$
- Honeycomb lattice: $\Gamma \rightarrow K \rightarrow M \rightarrow \Gamma$ (same as the triangular lattice, but different folding due to the different unit cell)

Remark. High-symmetry lines are merely a slice on a set of measure zero in BZ. The complete information of the band structure is contained in $\omega_n(\mathbf{k})$ over the entire BZ. Topological invariants are integral quantities defined over the entire BZ and cannot be inferred from high-symmetry line slices alone (unless symmetry-protected parameters provide constraints).

3.5 Photonic Bandgaps

3.5.1 Definition of Bandgaps

Definition 3.11 (Complete photonic bandgap). If there exists a frequency interval $(\omega_{\text{low}}, \omega_{\text{high}})$ such that

$$\omega_{\text{low}} = \max_{\mathbf{k} \in \text{BZ}} \omega_n(\mathbf{k}), \quad \omega_{\text{high}} = \min_{\mathbf{k} \in \text{BZ}} \omega_{n+1}(\mathbf{k}), \quad \omega_{\text{low}} < \omega_{\text{high}}, \quad (3.34)$$

then a **complete photonic bandgap** exists between the n -th and $(n+1)$ -th bands.

Definition 3.12 (Directional bandgap). If the above inequality holds only for $\mathbf{k}$ along specific directions, it is called a **directional** (or partial) **bandgap**. A directional bandgap does not block light in all propagation directions and therefore does not possess the “omnidirectional prohibition” property of a complete bandgap.

3.5.2 Physical Factors Affecting the Bandgap

The main factors affecting whether a photonic bandgap exists and its width include:

1. **Dielectric contrast:** The larger $\varepsilon_{\max}/\varepsilon_{\min}$, the wider the bandgap. This is directly related to the Rayleigh–Ritz variational principle: field concentration in high- ε regions lowers the frequency, field concentration in low- ε regions raises the frequency; the greater the contrast, the larger the inter-band frequency splitting.
2. **Connectivity:** Whether the dielectric material forms a connected network. Empirically, a connected high- ε structure favors TM bandgaps (2D), while an air background with high- ε cylinders favors TE bandgaps (2D). Three-dimensional complete bandgaps typically require “diamond” or “gyroid” type structures [16].
3. **Lattice symmetry:** Higher rotational symmetry (e.g., sixfold) tends to make the frequency ranges of bandgaps in different directions of BZ more uniform, promoting the “merging” of directional bandgaps into a complete bandgap.
4. **Polarization** (2D only): In 2D photonic crystals, TM modes ($\mathbf{E}$ along z) and TE modes ($\mathbf{H}$ along z) each satisfy a scalar equation (see next section), and their bandgap conditions differ. A complete 2D bandgap requires the TM and TE bandgaps to overlap in frequency.

3.5.3 Two-Dimensional Scalar Equations

In a 2D photonic crystal (ε and μ depend only on (x, y)), Maxwell’s equations decompose into two independent scalar problems. Setting $k_z = 0$ (in-plane propagation),

TM polarization (E_z mode)

$$-(\nabla_{xy} + i\mathbf{k}) \cdot \frac{1}{\mu(\mathbf{r})} (\nabla_{xy} + i\mathbf{k}) E_z(\mathbf{r}) = \frac{\omega^2}{c^2} \varepsilon(\mathbf{r}) E_z(\mathbf{r}), \quad (3.35)$$

TE polarization (H_z mode)

$$-(\nabla_{xy} + i\mathbf{k}) \cdot \frac{1}{\varepsilon(\mathbf{r})} (\nabla_{xy} + i\mathbf{k}) H_z(\mathbf{r}) = \frac{\omega^2}{c^2} \mu(\mathbf{r}) H_z(\mathbf{r}), \quad (3.36)$$

where $\mathbf{r} = (x, y)$ and $\mathbf{k} = (k_x, k_y)$ is the in-plane Bloch wave vector.

Remark. eqs. (3.35) and (3.36) are standard generalized eigenvalue problems for second-order self-adjoint elliptic operators, with ε and μ as the weights on the right-hand side, respectively. This is precisely the equivalent form of the $\hat{\Theta}_H$ operator from the Joannopoulos textbook [16], Chapter 2, in the special case of 2D with $k_z = 0$. We proved its self-adjointness in Chapter 2.

3.6 Variational Principle and Bandgap Estimation

3.6.1 Rayleigh Quotient

For the cell-periodic generalized eigenvalue problem eq. (3.26), define the **Rayleigh quotient**:

$$R[\mathbf{u}; \mathbf{k}] = \frac{\langle \mathbf{u}, \hat{N}(\mathbf{k})\mathbf{u} \rangle_{L^2(\Omega)}}{\langle \mathbf{u}, \bar{M}\mathbf{u} \rangle_{L^2(\Omega)}} = \frac{\langle \mathbf{u}, \hat{N}(\mathbf{k})\mathbf{u} \rangle_{L^2(\Omega)}}{\langle \mathbf{u}, \mathbf{u} \rangle_{M\Omega}}. \quad (3.37)$$

Remark. Since the spectrum of $\hat{N}(\mathbf{k})$ includes both positive and negative branches, the extremal principle for the Rayleigh quotient requires restriction to positive frequencies. For the n -th band of the positive-frequency branch,

$$\omega_n(\mathbf{k}) = \min_{\substack{V_n \subset L^2_{\text{per}}(\Omega) \\ \dim V_n = n}} \max_{\mathbf{u} \in V_n \setminus \{0\}} R[\mathbf{u}; \mathbf{k}], \quad (3.38)$$

which is the min-max theorem adapted to this generalized eigenvalue problem.

3.6.2 Physical Intuition for the Bandgap

Using the Rayleigh quotient, one can give an intuitive explanation for the opening of a gap. Near a Bragg plane, the two counter-propagating waves form two orthogonal standing-wave patterns. The lower band-edge mode places more field weight in high- ε regions and is pulled downward in frequency; the upper band-edge mode is forced, by orthogonality, to place more field weight in low- ε regions and is pushed upward. This is the variational version of the Bragg splitting in eq. (3.16) and of the Kronig–Penney condition eq. (3.24).

The frequency difference between the two standing waves is the bandgap width $\Delta\omega$. Qualitatively,

$$\boxed{\frac{\Delta\omega}{\omega_{\text{mid}}} \sim \frac{\varepsilon_{\text{max}} - \varepsilon_{\text{min}}}{\varepsilon_{\text{max}} + \varepsilon_{\text{min}}} = \frac{\Delta\varepsilon}{\bar{\varepsilon}}}. \quad (3.39)$$

In 2D, a complete gap requires the TM and TE gaps to overlap. Connected high- ε networks often favor TM gaps, while isolated high- ε inclusions can favor TE gaps. In 3D the vector nature of Maxwell fields makes this design problem more restrictive; diamond-like connectivity is the canonical example where a complete gap can open at moderate contrast [20, 16].

Remark (From bandgaps to topology). The existence of a bandgap is the **prerequisite** for topological classification: only after a bandgap separates the bands into “occupied” and “empty” bands do topological invariants such as the Chern number become meaningful (see chapter 4). Therefore, the standing-wave physical picture not only explains the origin of bandgaps but also indirectly answers “why certain frequency ranges can support topological edge states.”

3.6.3 Redundancy of $\mathbf{k}$

By eq. (3.13), $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ ($\mathbf{G} \in \Lambda^*$) give the same Bloch state. Therefore $\mathbf{k}$ is meaningful only modulo Λ^* , and BZ is the fundamental domain of $\mathbf{k}$ space modulo Λ^* .

3.6.4 Periodic Gauge

On the boundary of BZ, $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ represent the same physical state. However, the **phase** of the cell function $\mathbf{u}_{n\mathbf{k}}$ can be freely chosen— this phase freedom is called **gauge freedom**.

Definition 3.13 (Periodic gauge). If on each pair of equivalent boundaries $\mathbf{k}_\partial$ and $\mathbf{k}_\partial + \mathbf{G}$ of BZ, the cell functions satisfy

$$\mathbf{u}_{n,\mathbf{k}_\partial+\mathbf{G}}(\mathbf{r}) = e^{-i\mathbf{G}\cdot\mathbf{r}} \mathbf{u}_{n,\mathbf{k}_\partial}(\mathbf{r}), \quad (3.40)$$

this is called the **periodic gauge**.

Remark. The periodic gauge makes the Bloch wave $\mathbf{f}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ continuous on the BZ boundary. However, $\mathbf{u}_{n\mathbf{k}}$ itself may be discontinuous on the BZ boundary. When the Chern number $\mathcal{C}_n \neq 0$, it is impossible to find a $\mathbf{u}_{n\mathbf{k}}$ that is both smooth and periodic over the entire BZ— this is the core manifestation of the topological obstruction, to be discussed in detail in Chapter 4.

3.7 Overview of Numerical Methods

3.7.1 Plane-Wave Expansion Method (PWE)

The most direct numerical method is to expand the cell function $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ in reciprocal lattice vectors:

$$\mathbf{u}_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G} \in \Lambda^*} \hat{\mathbf{u}}_{n\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}. \quad (3.41)$$

Substituting into eq. (3.26) yields the matrix eigenvalue problem

$$\sum_{\mathbf{G}'} N_{\mathbf{G}\mathbf{G}'}(\mathbf{k}) \hat{\mathbf{u}}(\mathbf{G}') = \omega \sum_{\mathbf{G}'} M_{\mathbf{G}\mathbf{G}'} \hat{\mathbf{u}}(\mathbf{G}'), \quad (3.42)$$

where

$$N_{\mathbf{G}\mathbf{G}'}(\mathbf{k}) = \delta_{\mathbf{G}\mathbf{G}'} \begin{pmatrix} 0 & -i(\mathbf{k} + \mathbf{G}) \times \\ i(\mathbf{k} + \mathbf{G}) \times & 0 \end{pmatrix}, \quad (3.43)$$

$$M_{\mathbf{G}\mathbf{G}'} = \hat{M}(\mathbf{G} - \mathbf{G}') = \frac{1}{|\Omega|} \int_{\Omega} \bar{M}(\mathbf{r}) e^{-i(\mathbf{G}-\mathbf{G}')\cdot\mathbf{r}} d^d r. \quad (3.44)$$

Remark. In practice, $\mathbf{G}$ must be truncated to a finite set $|\mathbf{G}| \leq G_{\max}$. The $\hat{N}$ block is simple because it is diagonal in $\mathbf{G}$, while material inhomogeneity enters through the convolution matrix $M_{\mathbf{G}\mathbf{G}'}$. Convergence depends on the decay of the Fourier coefficients of $\varepsilon(\mathbf{r})$. Discontinuous interfaces converge slowly and are usually treated with improved Fourier factorization rules; the implementation details are deferred to Appendix C.

3.7.2 Finite Element Method (FEM)

The finite element method divides the unit cell into triangular or tetrahedral elements and represents $\mathbf{u}_{n\mathbf{k}}$ by local basis functions. It is better than PWE for curved interfaces, complex

unit cells, adaptive meshes, and edge-element treatments of Maxwell constraints. Its cost is the loss of the simple reciprocal-space convolution structure, so sparse eigensolvers are needed.

3.7.3 FDTD and Time-Domain Methods

The finite-difference time-domain method solves Maxwell equations in time and extracts frequencies by Fourier analysis. A single broadband pulse can probe many modes, which is useful for transmission spectra and edge-state dynamics. For high-precision band diagrams, however, the frequency resolution is limited by the simulation time and by boundary treatment.

3.8 $\mathbf{k} \cdot \mathbf{p}$ Perturbation Theory

3.8.1 Expansion of the Band Near $\mathbf{k}_0$

Suppose that at $\mathbf{k}_0$, the n -th band is non-degenerate, with eigenfrequency $\omega_n(\mathbf{k}_0)$ and cell function $\mathbf{u}_{n\mathbf{k}_0}$. When $\mathbf{k} = \mathbf{k}_0 + \delta\mathbf{k}$, $\hat{N}(\mathbf{k}) = \hat{N}(\mathbf{k}_0) + \delta\hat{N}$, where

$$\delta\hat{N} = \begin{pmatrix} 0 & i\delta\mathbf{k} \times \\ -i\delta\mathbf{k} \times & 0 \end{pmatrix}. \quad (3.45)$$

This is a perturbation with $\delta\mathbf{k}$ as the small parameter.

Theorem 3.14 ($\mathbf{k} \cdot \mathbf{p}$ perturbation theory). *To first order,*

$$\omega_n(\mathbf{k}_0 + \delta\mathbf{k}) = \omega_n(\mathbf{k}_0) + \delta\mathbf{k} \cdot \mathbf{v}_n(\mathbf{k}_0) + O(|\delta\mathbf{k}|^2), \quad (3.46)$$

where the **group velocity** $\mathbf{v}_n(\mathbf{k}) = \nabla_{\mathbf{k}}\omega_n(\mathbf{k})$ is given by

$$\mathbf{v}_n(\mathbf{k}_0) = \frac{1}{\langle \mathbf{u}_{n\mathbf{k}_0}, \mathbf{u}_{n\mathbf{k}_0} \rangle_{M_\Omega}} \langle \mathbf{u}_{n\mathbf{k}_0}, \bar{M}^{-1} \frac{\partial \hat{N}}{\partial \mathbf{k}} \Big|_{\mathbf{k}_0} \mathbf{u}_{n\mathbf{k}_0} \rangle_{M_\Omega}. \quad (3.47)$$

To second order in the non-degenerate case, the band-curvature tensor is

$$\frac{1}{m_{ij}^*} \equiv \frac{\partial^2 \omega_n}{\partial k_i \partial k_j} \Big|_{\mathbf{k}_0} = \frac{2}{\langle \mathbf{u}_{n\mathbf{k}_0}, \mathbf{u}_{n\mathbf{k}_0} \rangle_{M_\Omega}} \sum_{m \neq n} \frac{\text{Re} [V_{nm}^{(i)} V_{mn}^{(j)}]}{\omega_n(\mathbf{k}_0) - \omega_m(\mathbf{k}_0)}, \quad (3.48)$$

where

$$V_{nm}^{(i)} = \langle \mathbf{u}_{n\mathbf{k}_0}, \bar{M}^{-1} (\partial_{k_i} \hat{N}) \mathbf{u}_{m\mathbf{k}_0} \rangle_{M_\Omega}. \quad (3.49)$$

Proof sketch. Expand both $\omega_n(\mathbf{k}_0 + \delta\mathbf{k})$ and $\mathbf{u}_{n,\mathbf{k}_0+\delta\mathbf{k}}$ in powers of $\delta\mathbf{k}$, substitute into eq. (3.26), and compare order by order in $\delta\mathbf{k}$. The zeroth order is the original problem. Taking the inner product $\langle \cdot, \cdot \rangle_{M_\Omega}$ of the first-order equation with $\mathbf{u}_{n\mathbf{k}_0}$ and using self-adjointness to eliminate the first-order correction to $\mathbf{u}$ yields eq. (3.47). The second order is similar but requires the sum over $m \neq n$. See the exercises for details. $\square$

Remark. The $\partial \hat{N} / \partial \mathbf{k}$ in eq. (3.47) is given by eq. (3.45); its physical meaning is “the cross-coupling between the curl and $\delta\mathbf{k}$ ”—this is completely analogous to the role of the momentum

operator $\hat{p}$ in $\mathbf{k} \cdot \mathbf{p}$ perturbation theory in solid-state physics. This is also why we call this section “ $\mathbf{k} \cdot \mathbf{p}$ perturbation theory.”

3.9 Degenerate Points and Dirac Cones

3.9.1 $\mathbf{k} \cdot \mathbf{p}$ Theory for Twofold Degeneracies

When the n -th and $(n+1)$ -th bands are degenerate at $\mathbf{k}_0$ ($\omega_n(\mathbf{k}_0) = \omega_{n+1}(\mathbf{k}_0)$), **degenerate perturbation theory** must be used. Let the degenerate subspace be spanned by $\{\mathbf{u}_1, \mathbf{u}_2\}$ (already orthonormalized); then the effective Hamiltonian is a 2×2 matrix

$$H_{ab}^{\text{eff}}(\delta\mathbf{k}) = \omega_0 \delta_{ab} + \frac{1}{\langle \mathbf{u}_a, \mathbf{u}_a \rangle_{M\Omega}} \langle \mathbf{u}_a, \bar{M}^{-1} \delta \hat{N} \mathbf{u}_b \rangle_{M\Omega}, \quad a, b \in \{1, 2\}. \quad (3.50)$$

3.9.2 Conditions for a Dirac Cone

Proposition 3.15 (Linear degeneracy). *If $H^{\text{eff}}(\delta\mathbf{k})$ can be written as*

$$H^{\text{eff}} = \omega_0 \sigma_0 + v_x \delta k_x \sigma_x + v_y \delta k_y \sigma_y, \quad (3.51)$$

where σ_0 is the identity matrix, σ_x, σ_y are Pauli matrices, and $v_x, v_y \neq 0$, then the dispersion relation is $\omega_{\pm} = \omega_0 \pm \sqrt{v_x^2 \delta k_x^2 + v_y^2 \delta k_y^2}$, i.e., a **Dirac cone**.

Remark. Dirac cones appear at the K point of hexagonal lattices (C_3 symmetry-protected), or at the Γ point of square lattices (requiring accidental degeneracy of specific irreducible representations of C_{4v}). When a σ_z term (“mass term”) appears in H^{eff} , the Dirac cone is opened into a bandgap—this is the key mechanism for realizing topological phase transitions (Chapter 6). Wu and Hu [8], by shrinking/expanding the unit cell in a C_{6v} honeycomb lattice, continuously tuned the sign of the “mass” to realize the photonic QSHE state.

Example 3.16 (Dirac cone and valley mass in a honeycomb photonic crystal). A honeycomb photonic crystal provides the canonical setting for Dirac-cone physics. Near the inequivalent valleys K and K' , two bands may touch linearly, and the low-energy physics is described by an effective Dirac Hamiltonian. In this example we derive the symmetry-allowed form of the two-band Hamiltonian near the K point.

Step 1: Lattice structure and valleys. The honeycomb lattice is not a Bravais lattice, but consists of two triangular sublattices, denoted by A and B . It may be described by the triangular Bravais lattice with primitive vectors

$$\mathbf{a}_1 = a \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = a \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}. \quad (3.52)$$

The corresponding reciprocal lattice vectors are

$$\mathbf{b}_1 = \frac{2\pi}{a} \begin{pmatrix} 1 \\ -1/\sqrt{3} \end{pmatrix}, \quad \mathbf{b}_2 = \frac{2\pi}{a} \begin{pmatrix} 0 \\ 2/\sqrt{3} \end{pmatrix}. \quad (3.53)$$

The first Brillouin zone is a hexagon with two inequivalent corners, usually denoted by K and K' . We expand the Bloch wave vector as

$$\mathbf{k} = \mathbf{K} + \delta\mathbf{k}, \quad (3.54)$$

where $|\delta\mathbf{k}|$ is small.

Step 2: Two-mode subspace at the K point. Assume that two Bloch modes are degenerate at the K point:

$$\omega_1(\mathbf{K}) = \omega_2(\mathbf{K}) = \omega_0. \quad (3.55)$$

This degeneracy is enforced in the ideal honeycomb structure by the symmetry of the little group at K together with the equivalence of the two sublattices. Let

$$\{\mathbf{u}_1, \mathbf{u}_2\}$$

be an orthonormal basis for this degenerate subspace. Under a C_3 rotation about the center of the unit cell, we may choose this basis such that

$$D(C_3) = \begin{pmatrix} e^{i2\pi/3} & 0 \\ 0 & e^{-i2\pi/3} \end{pmatrix}. \quad (3.56)$$

Thus the two basis modes carry angular momenta $\ell = +1$ and $\ell = -1$ modulo 3.

Step 3: Symmetry constraints on the effective Hamiltonian. Projecting the full Maxwell operator onto the two-mode subspace gives the effective Hamiltonian

$$H^{\text{eff}}(\delta\mathbf{k}) = h_0(\delta\mathbf{k})\sigma_0 + h_x(\delta\mathbf{k})\sigma_x + h_y(\delta\mathbf{k})\sigma_y + h_z(\delta\mathbf{k})\sigma_z. \quad (3.57)$$

The C_3 covariance condition is

$$D(C_3)H^{\text{eff}}(C_3^{-1}\delta\mathbf{k})D(C_3)^{-1} = H^{\text{eff}}(\delta\mathbf{k}). \quad (3.58)$$

At linear order in $\delta\mathbf{k}$, this condition allows the off-diagonal Dirac coupling

$$h_x + ih_y \propto \delta k_x - i\delta k_y, \quad (3.59)$$

up to basis-dependent phase conventions. It forbids a linear-in-momentum contribution to h_z , but it does not by itself forbid a constant mass term. Thus the most general leading-order two-band Hamiltonian near K is

$$H_K^{\text{eff}}(\delta\mathbf{k}) = \omega_0\sigma_0 + v(\delta k_x\sigma_x - \delta k_y\sigma_y) + m\sigma_z. \quad (3.60)$$

The precise signs of the σ_x and σ_y terms depend on the choice of basis and on the convention for the valley.

If the two sublattices are equivalent, an additional symmetry exchanges the two components of the two-mode basis and forces

$$m = 0. \quad (3.61)$$

Then one obtains the massless Dirac Hamiltonian

$$H_K^{\text{eff}}(\delta\mathbf{k}) = \omega_0\sigma_0 + v(\delta k_x\sigma_x - \delta k_y\sigma_y), \quad (3.62)$$

with dispersion

$$\omega_{\pm}(\delta\mathbf{k}) = \omega_0 \pm v|\delta\mathbf{k}|. \quad (3.63)$$

Step 4: Physical origin of the mass term. A constant mass term

$$m\sigma_z$$

appears when the symmetry that makes the two components equivalent is broken. For example, in a honeycomb lattice this can be achieved by making the two sublattices inequivalent. In a photonic crystal, this may correspond to using different scatterer radii or different dielectric constants on the A and B sublattices. The spectrum then becomes

$$\omega_{\pm}(\delta\mathbf{k}) = \omega_0 \pm \sqrt{v^2|\delta\mathbf{k}|^2 + m^2}, \quad (3.64)$$

and the Dirac point is gapped by

$$\Delta\omega = 2|m|. \quad (3.65)$$

The sign of m distinguishes two valley-mass configurations. Across an interface where m changes sign, the valley Chern number changes, and valley-polarized interface states may appear, provided intervalley scattering is sufficiently weak. The time-reversed valley K' is described by a related Dirac Hamiltonian with the valley-dependent signs reversed.

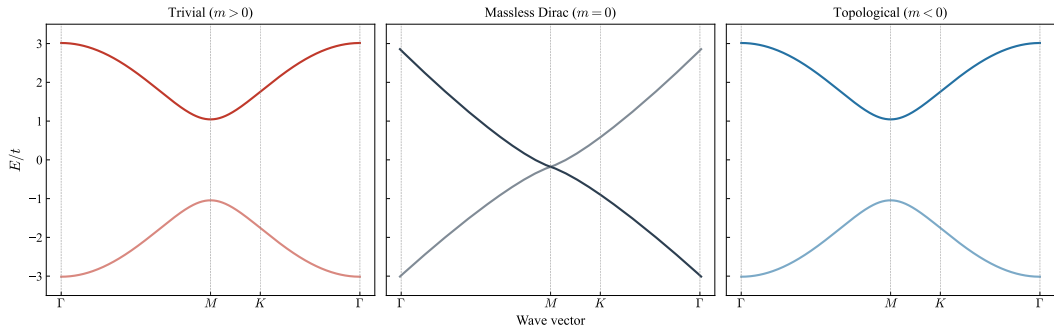


Figure 3.2 Dirac cone and mass-term-driven topological phase transition at the K point of a honeycomb lattice photonic crystal. The figure compares three cases along the high-symmetry path: $m > 0$, $m = 0$, and $m < 0$. At $m = 0$ a massless Dirac cone appears; when structural deformation shifts the effective mass term away from zero, the degeneracy is opened into a bandgap, and the sign reversal of the mass term corresponds to band inversion and a change in topological index. The two-band effective theory shown in this figure is precisely the spectral manifestation of eq. (3.51).

This is an exemplar of the “Maxwell-first” standpoint of this book: we do not “transplant” the electronic Dirac cone from graphene to the photonic system, but start from the projection of the Maxwell eigenproblem onto the nearly degenerate Bloch mode subspace to obtain an isomorphic two-band effective description. Correspondingly, Berry curvature, mass sign reversal, and the emergence of boundary states are first and foremost consequences of electromagnetic mode structural reorganization, and only then may be written in Dirac language.

For the lower band, the Berry curvature can be written as

$$\Omega_{-}(\delta\mathbf{k}) = \frac{m v^2}{2(v^2|\delta\mathbf{k}|^2 + m^2)^{3/2}}. \quad (3.66)$$

The curvature is concentrated near the gapped Dirac point and changes sign with m . Integrating the continuum Dirac theory gives the local contribution

$$\frac{1}{2\pi} \int_{\mathbb{R}^2} \Omega_{xy} d^2(\delta k) = \frac{1}{2} \text{sgn}(m). \quad (3.67)$$

A single Dirac cone therefore carries a half-integer local Berry flux, not a stand-alone lattice Chern number. The full integer invariant requires all valleys and the ultraviolet regularization of the lattice; this is developed systematically in Chapters 4 and 11.

Remark. eq. (3.66) is kept here only as a bridge from band theory to topology. The detailed gauge-invariant construction of Berry curvature and Chern number belongs to the next chapter.

3.10 The Light Cone and Radiative Modes

In slab photonic crystals, periodicity exists only in the in-plane (x, y) directions; along z the system has finite thickness, with air or substrate outside.

Definition 3.17 (Light cone). For an external homogeneous medium (refractive index n_{cl}), the **light cone** is defined as the region in frequency–wave-vector space where $\omega > c|\mathbf{k}_{\parallel}|/n_{\text{cl}}$. Modes within the light cone can couple to the radiative continuum, i.e., “leak” into free space.

Remark. Modes below the light cone are guided by total internal reflection and decay evanescently in the cladding. Modes inside the light cone generally have finite radiative lifetimes and should be treated as resonances. This brief distinction is enough for band theory; BICs and radiative topology are treated later in Chapter 10.

Chapter Summary

Periodic material parameters reduce the Maxwell eigenproblem on all space to a family of cell-periodic generalized eigenproblems indexed by the Bloch vector $\mathbf{k} \in \text{BZ}$. The translation operators form an Abelian group, so eigenfields can be written as

$$\mathbf{f}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_{n\mathbf{k}}(\mathbf{r}), \quad \mathbf{u}_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = \mathbf{u}_{n\mathbf{k}}(\mathbf{r}).$$

The resulting operator $\hat{N}(\mathbf{k})$ is self-adjoint under the cell energy inner product and has a discrete real spectrum for each fixed $\mathbf{k}$.

Bands $\omega_n(\mathbf{k})$, bandgaps, and high-symmetry-line diagrams are therefore all consequences of this Bloch decomposition. Bandgaps arise from Bragg coupling and can be understood through the variational tendency of fields to concentrate in high- or low-dielectric regions. The 1D Kronig–Penney model is the analytical benchmark; PWE, FEM, and FDTD are the main numerical routes in higher dimensions.

The same Bloch framework also introduces the gauge freedom of $\mathbf{u}_{n\mathbf{k}}$ over the Brillouin zone, preparing the Berry geometry of Chapter 4. Near isolated or degenerate points, $\mathbf{k} \cdot \mathbf{p}$ perturbation theory gives group velocities, effective masses, and two-band Dirac Hamiltonians. In honeycomb photonic crystals, a Dirac mass opens the K/K' valley degeneracy and provides the local band-theory origin of the topological phase transitions studied in later chapters.

Chapter 4

Berry Geometry and Topological Invariants

Why Band Theory Is Not Yet Topology?

The previous chapters established the Maxwell operator and the photonic Bloch eigenproblem. At that stage the band functions $\omega_n(\mathbf{k})$ are known, but the topology of a band is not contained only in its eigenvalues. It is contained in how the eigenvectors $\mathbf{u}_{n\mathbf{k}}$ vary over the Brillouin zone.

This chapter supplies the missing geometric layer. The Berry connection measures the local phase transport of Bloch eigenstates, the Berry curvature is the associated gauge-invariant local curvature, and the Chern number is the global topological invariant obtained by integrating that curvature over the Brillouin zone. For Maxwell systems all of these quantities must be defined using the material-weighted inner product inherited from the energy structure of the frequency-domain problem.

Remark (**Common misconception warnings**).

1. Berry phase is not merely an arbitrary phase convention. Its curvature controls physical response, and its integral controls edge mode count.
2. A nonzero Chern number is meaningful only for an isolated band or an isolated band group. Gap closings are precisely where the invariant can change.
3. Photonic systems do not naturally have fermionic Kramers degeneracy. A pseudo- $\mathbb{Z}_2$ structure requires extra symmetry, as in the Wu–Hu mechanism of Chapter 6.

The logical chain of the chapter is

$$\mathcal{A} \longrightarrow \mathcal{F} \longrightarrow \mathcal{C} \longrightarrow \text{bulk-edge correspondence}.$$

4.1 Adiabatic Evolution and the Berry Phase

4.1.1 Parameter Space and the Adiabatic Assumption

The cell-periodic Maxwell eigenproblem is

$$\hat{N}(\mathbf{k})\mathbf{u}_{n\mathbf{k}}(\mathbf{r}) = \omega_n(\mathbf{k})\bar{M}(\mathbf{r})\mathbf{u}_{n\mathbf{k}}(\mathbf{r}). \quad (4.1)$$

Let $\mathbf{k}$ vary slowly along a path in the Brillouin zone. The adiabatic assumption is that the band remains separated from the others:

$$\min_{m \neq n} |\omega_n(\mathbf{k}) - \omega_m(\mathbf{k})| = \Delta > 0 \quad \text{along the path.} \quad (4.2)$$

The eigenstate then follows the same band, but it can acquire a geometric phase.

With the normalization $\langle \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} = 1$, the overlap after a small displacement is

$$\langle \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n, \mathbf{k}+d\mathbf{k}} \rangle_{M_\Omega} = 1 + \langle \mathbf{u}_{n\mathbf{k}}, \partial_{k_i} \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} dk_i + O(|d\mathbf{k}|^2).$$

The derivative of the normalization condition implies that the linear term is purely imaginary. This imaginary part is the Berry connection.

4.2 Berry Connection and Berry Curvature

4.2.1 Berry Connection

Definition 4.1 (Berry connection). For an isolated band, the Berry connection is

$$\mathcal{A}_{n,i}(\mathbf{k}) = -i \langle \mathbf{u}_{n\mathbf{k}}, \partial_{k_i} \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} = -i \int_{\Omega} \mathbf{u}_{n\mathbf{k}}^\dagger(\mathbf{r}) \bar{M}(\mathbf{r}) \partial_{k_i} \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) d^d r. \quad (4.3)$$

Proposition 4.2. The Berry connection $\mathcal{A}_{n,i}(\mathbf{k})$ is real.

Proof. Differentiating $\langle \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} = 1$ with respect to k_i gives

$$\langle \partial_{k_i} \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} + \langle \mathbf{u}_{n\mathbf{k}}, \partial_{k_i} \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} = 0.$$

Since $\bar{M}$ is Hermitian, the two terms are complex conjugates. Thus $\langle \mathbf{u}_{n\mathbf{k}}, \partial_{k_i} \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega}$ is purely imaginary, and (4.3) is real. $\square$

Proposition 4.3 (Gauge transformation of the Berry connection). Under the gauge transformation $\mathbf{u}_{n\mathbf{k}} \rightarrow e^{i\phi_n(\mathbf{k})} \mathbf{u}_{n\mathbf{k}}$,

$$\mathcal{A}_{n,i}(\mathbf{k}) \rightarrow \mathcal{A}_{n,i}(\mathbf{k}) + \partial_{k_i} \phi_n(\mathbf{k}). \quad (4.4)$$

Proof. Substitute $e^{i\phi_n(\mathbf{k})} \mathbf{u}_{n\mathbf{k}}$ into (4.3). The derivative acts both on the state and on the phase factor, producing the additional term $\partial_{k_i} \phi_n$. $\square$

The Berry connection is therefore gauge-dependent, like the electromagnetic vector potential. The curvature is the gauge-invariant quantity.

4.2.2 Berry Curvature

Definition 4.4 (Berry curvature). In two dimensions,

$$\mathcal{F}_n(\mathbf{k}) = \partial_{k_x} \mathcal{A}_{n,y} - \partial_{k_y} \mathcal{A}_{n,x} = 2 \operatorname{Im} \langle \partial_{k_x} \mathbf{u}_{n\mathbf{k}}, \partial_{k_y} \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega}. \quad (4.5)$$

In d dimensions, $\mathcal{F}_{n,ij} = \partial_{k_i} \mathcal{A}_{n,j} - \partial_{k_j} \mathcal{A}_{n,i}$.

The last equality in (4.5) is often the most useful working form. It follows directly from the definition of $\mathcal{A}_i$:

$$\begin{aligned} \partial_{k_x} \mathcal{A}_y - \partial_{k_y} \mathcal{A}_x = -i \Big[& \langle \partial_{k_x} \mathbf{u}, \partial_{k_y} \mathbf{u} \rangle_{M_\Omega} + \langle \mathbf{u}, \partial_{k_x} \partial_{k_y} \mathbf{u} \rangle_{M_\Omega} \\ & - \langle \partial_{k_y} \mathbf{u}, \partial_{k_x} \mathbf{u} \rangle_{M_\Omega} - \langle \mathbf{u}, \partial_{k_y} \partial_{k_x} \mathbf{u} \rangle_{M_\Omega} \Big]. \end{aligned} \quad (4.6)$$

For a smooth gauge, the mixed second derivatives cancel. The remaining two terms are complex conjugates, so their antisymmetric combination gives $2 \operatorname{Im} \langle \partial_{k_x} \mathbf{u}, \partial_{k_y} \mathbf{u} \rangle_{M_\Omega}$. This derivation also makes clear why curvature is local and gauge-invariant, while the connection itself is not.

Proposition 4.5 (Gauge invariance of the Berry curvature). *$\mathcal{F}_n$ is invariant under $U(1)$ gauge transformations.*

Proof. Using (4.4), the extra term in the curvature is $\partial_{k_i} \partial_{k_j} \phi - \partial_{k_j} \partial_{k_i} \phi = 0$. $\square$

4.2.3 Sum Rule for the Berry Curvature

Proposition 4.6 (Curvature sum rule). *For a complete band basis,*

$$\sum_n \mathcal{F}_n(\mathbf{k}) = 0. \quad (4.7)$$

Sketch. Differentiate the completeness relation $\sum_n |\mathbf{u}_{n\mathbf{k}}\rangle \langle \mathbf{u}_{n\mathbf{k}}| \bar{M} = \hat{I}$ twice with respect to $\mathbf{k}$ and take the antisymmetric part. Since the identity operator has no curvature, the total curvature vanishes. $\square$

4.2.4 Geometric Intuition and a Two-Level Model

The Berry connection is a connection on the Bloch-state line bundle over the Brillouin zone. The Berry phase around a closed loop is the holonomy of this connection, and the Berry curvature is its local curvature. Unlike the holonomy of a vector on an open patch of a sphere, the integral over the whole Brillouin zone is quantized because the Brillouin zone is a compact closed manifold.

A useful physical response formula is the semiclassical anomalous velocity

$$\dot{\mathbf{r}} = \nabla_{\mathbf{k}} \omega_n(\mathbf{k}) + \dot{\mathbf{k}} \times \hat{\mathbf{z}} \mathcal{F}_n(\mathbf{k}), \quad (4.8)$$

which shows that Berry curvature can produce transverse wave-packet drift.

Example 4.7 (Berry curvature of a two-level system). Consider

$$H(\mathbf{k}) = d_0(\mathbf{k})\sigma_0 + \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma}. \quad (4.9)$$

The scalar term d_0 does not affect the eigenstates. For the lower band, with $\hat{\mathbf{d}} = \mathbf{d}/|\mathbf{d}|$, write

$$\hat{\mathbf{d}} = (\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta).$$

A convenient lower-band eigenstate is

$$|u_{-}\rangle = \begin{pmatrix} -\sin(\theta/2)e^{-i\varphi} \\ \cos(\theta/2) \end{pmatrix}.$$

Then

$$\mathcal{A}_i = -i\langle u_{-} | \partial_{k_i} u_{-} \rangle = -\frac{1 - \cos \theta}{2} \partial_{k_i} \varphi.$$

Taking the curl gives

$$\mathcal{F}_{-} = -\frac{1}{2} \sin \theta (\partial_{k_x} \theta \partial_{k_y} \varphi - \partial_{k_y} \theta \partial_{k_x} \varphi),$$

which is equivalently

$$\mathcal{F}_{-}(\mathbf{k}) = -\frac{1}{2} \hat{\mathbf{d}} \cdot (\partial_{k_x} \hat{\mathbf{d}} \times \partial_{k_y} \hat{\mathbf{d}}), \quad (4.10)$$

while the upper band has the opposite sign. The lower-band Chern number is

$$\mathcal{C}_{-} = -\frac{1}{4\pi} \int_{\text{BZ}} \hat{\mathbf{d}} \cdot (\partial_{k_x} \hat{\mathbf{d}} \times \partial_{k_y} \hat{\mathbf{d}}) d^2 k. \quad (4.11)$$

Thus the Chern number is the degree of the map $\hat{\mathbf{d}} : \text{BZ} \rightarrow S^2$, up to the sign convention of the band.

For a continuum massive Dirac cone $\mathbf{d} = (vk_x, vk_y, m)$,

$$\mathcal{F}_{-}(\mathbf{k}) = -\frac{1}{2} \frac{mv^2}{(v^2 k^2 + m^2)^{3/2}}. \quad (4.12)$$

The integral can be evaluated explicitly:

$$\mathcal{C}_{-}^{\text{Dirac}} = \frac{1}{2\pi} \int_0^\infty 2\pi k dk \left[-\frac{1}{2} \frac{mv^2}{(v^2 k^2 + m^2)^{3/2}} \right] = -\frac{1}{2} \text{sgn}(m).$$

This half-integer is not the Chern number of a complete lattice band; it is the local Berry-flux contribution of one Dirac cone. In a lattice model, all valleys and ultraviolet regularization must be included, so the total Chern number is an integer.

4.3 Kubo Formula for the Berry Curvature

The Kubo formula avoids direct numerical differentiation of eigenstates.

Theorem 4.8 (Kubo formula).

$$\mathcal{F}_n(\mathbf{k}) = 2 \sum_{m \neq n} \frac{\text{Im} \left[\langle \mathbf{u}_{n\mathbf{k}}, \bar{M}^{-1}(\partial_{k_x} \hat{N}) \mathbf{u}_{m\mathbf{k}} \rangle_{M_\Omega} \langle \mathbf{u}_{m\mathbf{k}}, \bar{M}^{-1}(\partial_{k_y} \hat{N}) \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega} \right]}{[\omega_n(\mathbf{k}) - \omega_m(\mathbf{k})]^2}. \quad (4.13)$$

Proof. Non-degenerate perturbation theory gives

$$\partial_{k_i} \mathbf{u}_{n\mathbf{k}} = \sum_{m \neq n} \frac{\langle \mathbf{u}_{m\mathbf{k}}, \bar{M}^{-1}(\partial_{k_i} \hat{N}) \mathbf{u}_{n\mathbf{k}} \rangle_{M_\Omega}}{\omega_n - \omega_m} \mathbf{u}_{m\mathbf{k}}. \quad (4.14)$$

This expression is obtained by differentiating $\hat{N}\mathbf{u}_{n\mathbf{k}} = \omega_n \bar{M}\mathbf{u}_{n\mathbf{k}}$ and projecting onto $\mathbf{u}_{m\mathbf{k}}$ with $m \neq n$:

$$\langle \mathbf{u}_{m\mathbf{k}}, \bar{M}^{-1}(\partial_{k_i} \hat{N})\mathbf{u}_{n\mathbf{k}} \rangle_{M\Omega} = (\omega_n - \omega_m) \langle \mathbf{u}_{m\mathbf{k}}, \partial_{k_i} \mathbf{u}_{n\mathbf{k}} \rangle_{M\Omega}.$$

Substituting the expansion for both $\partial_{k_x} \mathbf{u}_{n\mathbf{k}}$ and $\partial_{k_y} \mathbf{u}_{n\mathbf{k}}$ into $2 \operatorname{Im} \langle \partial_{k_x} \mathbf{u}_{n\mathbf{k}}, \partial_{k_y} \mathbf{u}_{n\mathbf{k}} \rangle_{M\Omega}$, orthonormality leaves only equal intermediate-band indices:

$$\langle \partial_{k_x} \mathbf{u}_{n\mathbf{k}}, \partial_{k_y} \mathbf{u}_{n\mathbf{k}} \rangle_{M\Omega} = \sum_{m \neq n} \frac{\langle \mathbf{u}_{n\mathbf{k}}, \bar{M}^{-1}(\partial_{k_x} \hat{N})\mathbf{u}_{m\mathbf{k}} \rangle_{M\Omega} \langle \mathbf{u}_{m\mathbf{k}}, \bar{M}^{-1}(\partial_{k_y} \hat{N})\mathbf{u}_{n\mathbf{k}} \rangle_{M\Omega}}{(\omega_n - \omega_m)^2}.$$

Taking twice the imaginary part gives (4.13). $\square$

The denominator explains why Berry curvature is concentrated near small gaps and why it becomes singular at topological phase transitions.

4.4 The Chern Number

4.4.1 Definition and Integrality

Definition 4.9 (Chern number). For an isolated band on a two-dimensional Brillouin zone,

$$\mathcal{C}_n = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_n(\mathbf{k}) d^2k \in \mathbb{Z}. \quad (4.15)$$

Theorem 4.10 (Integrality of the Chern number). $\mathcal{C}_n$ is an integer.

Physical proof. Cover the Brillouin zone by two patches U_1, U_2 on which smooth gauges exist. On the overlap boundary γ , the two gauges are related by $\mathbf{u}^{(2)} = e^{ig(\mathbf{k})}\mathbf{u}^{(1)}$. The corresponding Berry connections satisfy

$$\mathcal{A}^{(2)} = \mathcal{A}^{(1)} + \nabla_{\mathbf{k}} g.$$

Using Stokes' theorem,

$$2\pi\mathcal{C}_n = \int_{U_1} \mathcal{F} + \int_{U_2} \mathcal{F} = \oint_{\gamma} (\mathcal{A}^{(1)} - \mathcal{A}^{(2)}) \cdot d\mathbf{k} = - \oint_{\gamma} \nabla_{\mathbf{k}} g \cdot d\mathbf{k}.$$

The relative minus sign comes from the opposite orientations with which the same overlap boundary is seen from the two patches. Since e^{ig} is single-valued on the closed curve, $\Delta g = 2\pi\ell$ with $\ell \in \mathbb{Z}$. The integer ℓ is the winding number of the transition function $e^{ig} : \gamma \rightarrow U(1)$. Hence $\mathcal{C}_n \in \mathbb{Z}$. $\square$

Proposition 4.11. The integer Chern number does not depend on replacing the material-weighted inner product by another smooth positive-definite Hermitian inner product.

Sketch. Changing the positive-definite inner product changes the Berry connection by a globally defined smooth one-form on the Brillouin zone. The difference of the corresponding Berry curvatures is exact. Its integral over the boundaryless torus BZ vanishes. $\square$

4.4.2 Deformation, Gap Closing, and $\mathcal{T}$

Proposition 4.12 (Invariance of the Chern number under adiabatic deformations). *If an isolated band remains separated from all other bands by a nonzero gap during a continuous parameter deformation, its Chern number is unchanged.*

Proof. The Chern number varies continuously under smooth deformation of an isolated bundle, but it is integer-valued. It must therefore remain constant. $\square$

Thus a topological phase transition requires a bandgap closing. In Dirac language this is the point where the mass term changes sign and the Berry curvature flux can be transferred between bands.

Theorem 4.13. *If an isolated nondegenerate band is mapped to itself by bosonic time-reversal symmetry, then its Chern number is zero. More generally, the total Chern number of a time-reversal-invariant band subspace is zero.*

Proof. Time reversal maps $\mathbf{k}$ to $-\mathbf{k}$ and reverses Berry curvature: $\mathcal{F}(\mathbf{k}) = -\mathcal{F}(-\mathbf{k})$ for a self-conjugate isolated band. The Brillouin-zone integral cancels pairwise. For an invariant band group, the trace of the non-Abelian curvature obeys the same cancellation. $\square$

Corollary 4.14. *A nonzero total Chern number for a photonic band subspace requires breaking time-reversal symmetry.*

4.5 The Bulk–Edge Correspondence Principle

4.5.1 Statement

At an interface, the topological data of the bulk bandgap can change. The consequence is spectral flow through the gap.

Theorem 4.15 (Bulk–edge correspondence). *Let the material on the left have gap Chern number $\mathcal{C}_L$ and the material on the right have gap Chern number $\mathcal{C}_R$. The net number of chiral edge branches crossing the common gap is*

$$n_+ - n_- = \mathcal{C}_L - \mathcal{C}_R, \quad (4.16)$$

where n_+ and n_- count positive- and negative-velocity edge branches.

The edge-state spectrum in a strip geometry gives the most visible form of the theorem. Bulk bands form projected continua, while edge branches traverse the gap. The net crossing number cannot change unless the bulk gap closes.

4.5.2 Jackiw–Rebbi Mechanism

The simplest microscopic picture is a Dirac mass domain wall.

Example 4.16 (Jackiw–Rebbi zero mode: Bound state on a domain wall). Consider

$$\hat{H}\psi(x) = [-iv\sigma_x\partial_x + m(x)\sigma_z]\psi(x) = E\psi(x), \quad (4.17)$$

where $m(x \rightarrow +\infty) = +m_0$ and $m(x \rightarrow -\infty) = -m_0$. Set $E = 0$. Multiplying the zero-energy equation by σ_x gives

$$-iv\partial_x\psi(x) - im(x)\sigma_y\psi(x) = 0,$$

or

$$\partial_x\psi(x) = -\frac{m(x)}{v}\sigma_y\psi(x).$$

Let $\sigma_y\chi_s = s\chi_s$ with $s = \pm 1$ and set $\psi(x) = f_s(x)\chi_s$. Then

$$\partial_x f_s(x) = -s \frac{m(x)}{v} f_s(x).$$

For the mass profile specified above, the normalizable choice is $s = +1$. The normalizable bound state is

$$\boxed{\psi_0(x) = \mathcal{N} \exp\left[-\frac{1}{v} \int_0^x m(x') dx'\right] \chi, \quad E_0 = 0.} \quad (4.18)$$

It exists because the exponent decays both as $x \rightarrow +\infty$ and as $x \rightarrow -\infty$ only when the mass changes sign. Restoring the momentum parallel to the interface turns this bound state into a chiral dispersing edge branch.

This model is not a substitute for the full Maxwell problem. It is the low-energy envelope theory obtained after projecting near a band inversion point. Its value is that it makes the bulk–edge principle concrete: a spatial change of topological mass forces an interface mode.

4.6 Multi-Band Chern Numbers and the Wilson Loop

4.6.1 Non-Abelian Berry Geometry

When a group of N bands is isolated from all other bands but internally degenerate or entangled, Berry geometry becomes non-Abelian.

Definition 4.17 (Non-Abelian Berry connection). For an isolated band group,

$$[\mathcal{A}_i(\mathbf{k})]_{ab} = -i\langle \mathbf{u}_{a\mathbf{k}}, \partial_{k_i} \mathbf{u}_{b\mathbf{k}} \rangle_{M_\Omega}, \quad a, b = 1, \dots, N. \quad (4.19)$$

The curvature is

$$\mathcal{F}_{ij} = \partial_{k_i} \mathcal{A}_j - \partial_{k_j} \mathcal{A}_i + i[\mathcal{A}_i, \mathcal{A}_j]. \quad (4.20)$$

The total Chern number of the group is obtained from $\text{Tr } \mathcal{F}$.

4.6.2 Wilson Loop

Definition 4.18 (Wilson loop operator). Along a closed path ℓ in the Brillouin zone,

$$\mathcal{W}_\ell = \mathcal{P} \exp \left(i \oint_\ell \mathcal{A}_i(\mathbf{k}) dk_i \right), \quad (4.21)$$

where $\mathcal{P}$ denotes path ordering.

Proposition 4.19 (Gauge covariance of the Wilson loop). *Under a smooth $U(N)$ gauge transformation, a Wilson loop based at $\mathbf{k}_0$ transforms as*

$$\mathcal{W}_\ell \rightarrow V^\dagger(\mathbf{k}_0) \mathcal{W}_\ell V(\mathbf{k}_0).$$

Therefore its eigenvalues are gauge-invariant.

The covariance follows from the transformation law

$$\mathcal{A}_i \rightarrow V^\dagger \mathcal{A}_i V - i V^\dagger \partial_{k_i} V.$$

Along a discretized path, neighboring factors contain endpoint matrices that cancel telescopically. On a closed loop only the base-point matrices remain, which gives the similarity transformation above. Thus the Wilson-loop phases, not the individual matrix elements, are the physical quantities.

For a loop along k_x at fixed k_y ,

$$\mathcal{W}(k_y) = \mathcal{P} \exp \left(i \int_{-\pi/a}^{\pi/a} \mathcal{A}_x(k_x, k_y) dk_x \right). \quad (4.22)$$

The phases of its eigenvalues are hybrid Wannier centers. Their winding as k_y varies diagnoses Chern topology and, for TRS-protected band groups, the $\mathbb{Z}_2$ index.

The detailed relation to Wannier localizability and symmetry indicators is important but belongs more naturally to Appendix B and the numerical appendix. Here the point is simply that Wilson loops are the multi-band generalization of Berry phase.

4.7 Discrete Numerical Algorithms

4.7.1 Link Variables

On a discrete Brillouin-zone grid, define

Definition 4.20 (Link variable).

$$U_{n,i}(\mathbf{k}) = \frac{\langle \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n,\mathbf{k}+\hat{e}_i \Delta k_i} \rangle_{M_\Omega}}{|\langle \mathbf{u}_{n\mathbf{k}}, \mathbf{u}_{n,\mathbf{k}+\hat{e}_i \Delta k_i} \rangle_{M_\Omega}|}. \quad (4.23)$$

Endpoint gauge phases cancel in products around closed plaquettes.

4.7.2 The Fukui–Hatsugai–Suzuki Method

Definition 4.21 (Lattice Berry curvature (FHS)). For a plaquette based at $\mathbf{k}$,

$$\tilde{\mathcal{F}}_n(\mathbf{k}) = \frac{1}{i} \ln [U_{n,x}(\mathbf{k})U_{n,y}(\mathbf{k} + \hat{e}_x\Delta k_x)U_{n,x}^*(\mathbf{k} + \hat{e}_y\Delta k_y)U_{n,y}^*(\mathbf{k})], \quad (4.24)$$

with the logarithm taken on the principal branch $(-\pi, \pi]$.

Each link variable has unit modulus and changes under a gauge transformation only by endpoint phases:

$$U_i(\mathbf{k}) \rightarrow e^{-i\phi(\mathbf{k})}U_i(\mathbf{k})e^{i\phi(\mathbf{k}+\hat{e}_i\Delta k_i)}.$$

The four endpoint phases cancel around a plaquette. The FHS flux is therefore gauge-invariant on each plaquette, even before taking the continuum limit.

Theorem 4.22 (FHS Chern number).

$$\mathcal{C}_n^{\text{lat}} = \frac{1}{2\pi} \sum_{\mathbf{k} \in \text{grid}} \tilde{\mathcal{F}}_n(\mathbf{k}). \quad (4.25)$$

For an isolated band and a mesh that resolves the curvature, this integer converges to the continuum Chern number.

Example 4.23 (Explicit FHS computation: the Qi–Wu–Zhang model). For

$$H(\mathbf{k}) = \sin k_x \sigma_x + \sin k_y \sigma_y + (m + \cos k_x + \cos k_y) \sigma_z,$$

diagonalize $H(\mathbf{k}_{ij})$ on a periodic mesh and choose the lower-band eigenvector at every grid point. The eigenvector phases may be arbitrary; the plaquette product removes them. The plaquette Wilson loop

$$W_{xy}(\mathbf{k}_{ij}) = U_x(\mathbf{k}_{ij})U_y(\mathbf{k}_{i+1,j})U_x(\mathbf{k}_{i,j+1})^*U_y(\mathbf{k}_{ij})^*$$

gives the lattice flux $\tilde{\mathcal{F}}(\mathbf{k}_{ij}) = \text{Im} \log W_{xy}$. The numerical procedure is therefore:

$$\{|u_-(\mathbf{k}_{ij})\rangle\} \longrightarrow U_x, U_y \longrightarrow W_{xy} \longrightarrow \frac{1}{2\pi} \sum_{ij} \text{Im} \log W_{xy}.$$

For example, with the convention of this chapter, the lower band has $\mathcal{C}_- = -1$ for $-2 < m < 0$ and $\mathcal{C}_- = +1$ for $0 < m < 2$. At $m = -2, 0, 2$ the gap closes, so the Chern number is not defined there and can change across those values.

4.7.3 Discrete Wilson Loop

Definition 4.24 (Discrete Wilson loop). For an isolated N -band group, define the overlap matrix

$$[F(\mathbf{k}_j)]_{ab} = \langle \mathbf{u}_{a,\mathbf{k}_j}, \mathbf{u}_{b,\mathbf{k}_{j+1}} \rangle_{M_\Omega}. \quad (4.26)$$

The discrete Wilson loop along the k_x direction is

$$\mathcal{W}(k_y) = \prod_{j=0}^{N_1-1} F(\mathbf{k}_j), \quad (4.27)$$

with periodic closure $\mathbf{k}_{N_1} = \mathbf{k}_0 + \mathbf{G}_x$.

In practice one usually unitarizes each overlap matrix by polar decomposition before forming the product. This improves numerical stability without changing the continuum limit.

4.8 The $\mathbb{Z}_2$ Invariant

For fermionic $\mathcal{T}^2 = -1$ systems, the total Chern number vanishes but a $\mathbb{Z}_2$ invariant can remain nontrivial. The invariant can be diagnosed by the partner-switching pattern of Wilson-loop phases between time-reversal-invariant momenta.

For ordinary photons, physical time reversal satisfies $\mathcal{T}^2 = +1$, so the electronic class-AII $\mathbb{Z}_2$ invariant is not automatically available. Photonic pseudo- $\mathbb{Z}_2$ phases require an additional symmetry that constructs an effective pseudo- $\mathcal{T}$ with square -1 in a restricted subspace. This is the starting point of Chapter 6.

The robust practical message is:

- Chern topology requires $\mathcal{T}$ breaking and is integer-valued.
- Fermionic or pseudo-fermionic $\mathcal{T}$ can instead produce $\mathbb{Z}_2$ topology.
- In photonics, the validity of pseudo- $\mathbb{Z}_2$ protection must be checked from the actual symmetry of the platform.

4.9 Introduction to Higher-Dimensional Topological Invariants

In three-dimensional systems, Berry curvature can be integrated over closed two-dimensional surfaces. This is the natural language of Weyl points.

Definition 4.25 (Weyl point). A Weyl point is an isolated linear crossing in a three-dimensional Brillouin zone described locally by

$$H_{\text{eff}} = \omega_W \sigma_0 + \sum_{i=x,y,z} v_i \delta k_i \sigma_i.$$

Its chirality is the Chern number on a small sphere enclosing the crossing, equal to $\text{sgn}(v_x v_y v_z)$ in this diagonal convention.

Proposition 4.26 (Nielsen–Ninomiya theorem). *In a periodic lattice, the total Weyl chirality over the Brillouin zone is zero. Weyl points therefore appear in opposite-chirality sets.*

Higher-order topological invariants, such as nested Wilson loops and multipole moments, diagnose corner and hinge states. They are mentioned here only to locate them within the Berry-geometry family; the detailed photonic physics belongs to later chapters.

Chapter Summary

Berry geometry turns Bloch eigenvectors into topology. The Berry connection is gauge-dependent, the Berry curvature is gauge-invariant, and the Chern number is the integral of curvature over the Brillouin zone. For photonic bands, these quantities must be defined using the material-weighted Maxwell inner product.

The Chern number is integer-valued and stable under any deformation that keeps the band isolated. It changes only through gap closing. Time-reversal symmetry forces the total Chern number of a self-conjugate band or band group to vanish, which is why photonic Chern insulators require gyrotropic or otherwise $\mathcal{T}$ -breaking physics.

Bulk–edge correspondence states that the net number of chiral edge branches is the Chern-number difference across an interface. The Jackiw–Rebbi domain-wall solution gives the simplest analytical picture: a sign-changing Dirac mass binds an interface mode.

For numerical work, the FHS method gives a gauge-invariant lattice Chern number, and Wilson loops generalize Berry phase to multi-band subspaces. These tools are the common language used by the following chapters.

Chapter 5

Photonic Chern Insulators and the Quantum Hall Effect

Why Topology Must Become Chiral Transport?

The first three chapters built the framework: Maxwell operators, Bloch bands, Berry curvature, and Chern numbers. This chapter applies that framework to the first experimentally verified topological photonic system: the magneto-optical photonic Chern insulator. The physical chain is direct:

$$\text{gyrotropic material} \longrightarrow \mathcal{T} \text{ breaking} \longrightarrow \mathcal{C} \neq 0 \longrightarrow \text{one-way edge transport.}$$

Raghu and Haldane [1] proposed that a gyromagnetic photonic crystal can acquire nonzero band Chern numbers. Wang et al. [6, 21] then observed microwave unidirectional edge propagation in a YIG photonic crystal. This makes the photonic Chern insulator the cleanest first example of the bulk–edge correspondence in topological photonics.

Remark (Common misconception warnings).

1. A small gyrotropic parameter does not by itself make the phase topologically trivial. The Chern number changes only when a gap closes. What becomes small is the observable bandgap.
2. Opening a bandgap is not the same as producing a topological phase. The diagnostic is the Chern number, not the mere existence of a gap.
3. Microwave YIG experiments cannot be scaled directly to optical frequencies. The magneto-optical figure of merit drops by orders of magnitude in the near-infrared and visible regimes.

The main line of this chapter is therefore deliberately narrow: construct the gyrotropic material tensor, derive the 2D eigenproblem, compute the Chern number, and connect it to chiral edge transport.

5.1 Gyrotropic Media and Magneto-Optical Response

5.1.1 Gyrotropic Tensors

When a ferrite or YIG material is biased by a static magnetic field $\mathbf{B}_0 = B_0 \hat{z}$, the transverse electromagnetic response becomes gyrotropic. With the time convention $e^{-i\omega t}$, a common lossless form of the permittivity tensor is

Definition 5.1 (Gyromagnetic permittivity tensor).

$$\boldsymbol{\varepsilon}(\omega) = \varepsilon_0 \begin{pmatrix} \varepsilon_{\perp}(\omega) & -ig(\omega) & 0 \\ ig(\omega) & \varepsilon_{\perp}(\omega) & 0 \\ 0 & 0 & \varepsilon_z(\omega) \end{pmatrix}. \quad (5.1)$$

Here g is the gyrotropic parameter.

The tensor is Hermitian in the lossless limit, but it is not real when $g \neq 0$. Thus the material can remain energy-conserving while breaking time-reversal symmetry. For ferrites near ferromagnetic resonance, the permeability tensor takes the same gyrotropic form,

$$\boldsymbol{\mu}(\omega) = \mu_0 \begin{pmatrix} \mu_{\perp}(\omega) & -i\kappa(\omega) & 0 \\ i\kappa(\omega) & \mu_{\perp}(\omega) & 0 \\ 0 & 0 & \mu_z(\omega) \end{pmatrix}. \quad (5.2)$$

The Polder model gives

$$\mu_{\perp} = 1 + \frac{\omega_0 \omega_m}{\omega_0^2 - \omega^2}, \quad \kappa = \frac{\omega \omega_m}{\omega_0^2 - \omega^2}, \quad (5.3)$$

where $\omega_0 = \gamma B_0$ and $\omega_m = \gamma \mu_0 M_s$.

Assembling the electric and magnetic tensors into the material matrix

$$\bar{M} = \begin{pmatrix} \boldsymbol{\varepsilon} & 0 \\ 0 & \boldsymbol{\mu} \end{pmatrix},$$

one obtains the basic criterion for the photonic Chern phase.

Proposition 5.2 (Sufficient condition for $\mathcal{T}$ breaking). *If $g \neq 0$ or $\kappa \neq 0$, then $\bar{M}^* \neq \bar{M}$ and physical time-reversal symmetry is broken. If the tensors remain Hermitian and positive definite in the frequency window of interest, the closed lossless Maxwell eigenproblem still has real eigenfrequencies and a well-defined band topology.*

Proof. The material condition for ordinary $\mathcal{T}$ in the frequency-domain Maxwell problem requires the constitutive tensors to be real in the appropriate basis. The imaginary antisymmetric off-diagonal entries violate this condition. Hermiticity, however, is a separate statement: the off-diagonal entries are complex conjugates of each other, so the energy inner product and real spectrum remain meaningful in the lossless closed setting. $\square$

5.1.2 Microscopic Origin and Frequency Limitation

The gyrotropic tensor can be understood from cyclotron motion. For free carriers of effective mass m^* in $\mathbf{B}_0 = B_0 \hat{z}$, the frequency-domain Lorentz equation gives

$$-m^* \omega^2 \mathbf{r} = -e \mathbf{E} + i \omega e \mathbf{r} \times B_0 \hat{z} - i m^* \omega \gamma \mathbf{r}. \quad (5.4)$$

With

$$\omega_c = \frac{e B_0}{m^*}, \quad \omega_p = \sqrt{\frac{N e^2}{m^* \varepsilon_0}},$$

the resulting Drude gyrotropic response has the form

$$\bar{\varepsilon}(\omega) = \begin{pmatrix} \varepsilon_{\perp}(\omega) & -i g(\omega) & 0 \\ i g(\omega) & \varepsilon_{\perp}(\omega) & 0 \\ 0 & 0 & \varepsilon_z(\omega) \end{pmatrix}, \quad (5.5)$$

where

$$\varepsilon_{\perp}(\omega) = \varepsilon_{\infty} - \frac{\omega_p^2 (\omega + i\gamma)}{\omega [(\omega + i\gamma)^2 - \omega_c^2]}, \quad (5.6)$$

$$g(\omega) = \frac{\omega_p^2 \omega_c}{\omega [(\omega + i\gamma)^2 - \omega_c^2]}. \quad (5.7)$$

In the lossless high-frequency limit,

$$g(\omega) \simeq \frac{\omega_p^2 \omega_c}{\omega^3}, \quad \omega \gg \omega_c. \quad (5.8)$$

This scaling is the physical reason why magneto-optical Chern insulators work cleanly in the microwave regime but become difficult at optical frequencies.

In the Wang experiment, YIG operates near a ferromagnetic resonance at GHz frequencies, and $\kappa/\mu_{\perp}$ can be of order 0.1–1. In near-infrared garnets, by contrast, the effective off-diagonal-to-diagonal ratio is typically 10^{-3} or smaller, shrinking the topological bandgap below typical fabrication and loss scales.

5.2 Band Structure of the 2D Gyromagnetic Photonic Crystal

5.2.1 Wang Model and Polarization Reduction

The classic model is a two-dimensional square lattice of YIG cylinders in air, with the bias field along $\hat{z}$. Assume invariance along z . Maxwell's equations then separate into two closed polarizations: TM modes (E_z, H_x, H_y) and TE modes (H_z, E_x, E_y) .

For TM modes, the x, y components of $\nabla \times \mathbf{E} = i\omega \boldsymbol{\mu} \mathbf{H}$ give

$$i\omega \begin{pmatrix} \mu_{\perp} & -i\kappa \\ i\kappa & \mu_{\perp} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix} = \begin{pmatrix} -\partial_y E_z \\ \partial_x E_z \end{pmatrix}. \quad (5.9)$$

Together with the z component of $\nabla \times \mathbf{H} = -i\omega\epsilon\mathbf{E}$, this yields

$$\nabla \cdot [\boldsymbol{\mu}_{\perp}^{-1}(\mathbf{r})\nabla E_z(\mathbf{r})] = -\omega^2\epsilon_z(\mathbf{r})E_z(\mathbf{r}), \quad (5.10)$$

where

$$\boldsymbol{\mu}_{\perp}^{-1} = \frac{1}{\mu_{\perp}^2 - \kappa^2} \begin{pmatrix} \mu_{\perp} & i\kappa \\ -i\kappa & \mu_{\perp} \end{pmatrix}. \quad (5.11)$$

The TE equation has the analogous form

$$\nabla \cdot [\boldsymbol{\epsilon}_{\perp}^{-1}(\mathbf{r})\nabla H_z(\mathbf{r})] = -\omega^2\mu_z(\mathbf{r})H_z(\mathbf{r}). \quad (5.12)$$

Remark. The off-diagonal gyrotropic term is what distinguishes the Chern-insulator problem from an ordinary scalar photonic crystal. When $\kappa = 0$, (5.10) reduces to a standard scalar Helmholtz eigenproblem. When $\kappa \neq 0$, $\mathcal{T}$ is broken and degeneracies can open into topological gaps.

5.2.2 Bloch Form and the Wang Bandgap

For the TM problem, write $E_z(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$. The cell-periodic equation is

$$(\nabla + i\mathbf{k}) \cdot [\boldsymbol{\mu}_{\perp}^{-1}(\mathbf{r})(\nabla + i\mathbf{k})u_{n\mathbf{k}}(\mathbf{r})] = -\omega_n^2(\mathbf{k})\epsilon_z(\mathbf{r})u_{n\mathbf{k}}(\mathbf{r}). \quad (5.13)$$

It can be solved by plane-wave expansion, finite elements, or other Hermitian eigensolvers adapted to the material-weighted inner product.

Example 5.3 (Band computation for the Wang model). For a square lattice with YIG cylinders of radius $r/a \simeq 0.11$ and microwave parameters near 4.5 GHz, turning on the gyromagnetic permeability κ lifts the relevant degeneracy and opens a complete TM bandgap. The two bands adjacent to the gap acquire opposite Chern numbers; in the standard convention the lower and upper bands carry $\mathcal{C} = +1$ and $\mathcal{C} = -1$, respectively.

5.2.3 Plane-Wave Matrix Construction

For gyromagnetic media, the main numerical subtlety is the tensor inverse rule. One should not Fourier-expand each component of $\boldsymbol{\mu}_{\perp}^{-1}(\mathbf{r})$ independently across a discontinuous interface. Instead, build the Fourier matrix of the full tensor and invert it as a block matrix.

Example 5.4 (Explicit construction of the PWE matrix equation). Expand the cell-periodic field as

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}}.$$

For a periodic tensor component $f(\mathbf{r})$, let

$$\hat{f}_{\mathbf{G}-\mathbf{G}'} = \frac{1}{\Omega} \int_{\text{cell}} f(\mathbf{r}) e^{-i(\mathbf{G}-\mathbf{G}')\cdot\mathbf{r}} d^2r.$$

Construct the $2N_G \times 2N_G$ block Fourier matrix

$$\mathcal{M} = \begin{pmatrix} [\hat{\mu}_\perp]_{\mathbf{G},\mathbf{G}'} & -i[\hat{\kappa}]_{\mathbf{G},\mathbf{G}'} \\ i[\hat{\kappa}]_{\mathbf{G},\mathbf{G}'} & [\hat{\mu}_\perp]_{\mathbf{G},\mathbf{G}'} \end{pmatrix}. \quad (5.14)$$

Then invert the whole matrix:

$$\mathcal{M}^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}.$$

The PWE generalized eigenproblem has the form

$$\boxed{\sum_{\mathbf{G}'} H_{\mathbf{G}\mathbf{G}'}(\mathbf{k}) c_{\mathbf{G}'} = \omega^2 \sum_{\mathbf{G}'} S_{\mathbf{G}\mathbf{G}'} c_{\mathbf{G}'},} \quad (5.15)$$

with

$$H_{\mathbf{G}\mathbf{G}'}(\mathbf{k}) = (\mathbf{k} + \mathbf{G})^T \begin{pmatrix} \mathbf{A}_{\mathbf{G},\mathbf{G}'} & \mathbf{B}_{\mathbf{G},\mathbf{G}'} \\ \mathbf{C}_{\mathbf{G},\mathbf{G}'} & \mathbf{D}_{\mathbf{G},\mathbf{G}'} \end{pmatrix} (\mathbf{k} + \mathbf{G}'), \quad (5.16)$$

and

$$S_{\mathbf{G}\mathbf{G}'} = \hat{\varepsilon}_{\mathbf{G}-\mathbf{G}'}. \quad (5.17)$$

This is the implementable matrix form used to compute the gyromagnetic TM bands.

Remark. The full tensor inversion is a numerical necessity, not just a formal nicety. Directly expanding the inverse components of a discontinuous tensor produces slow convergence and can distort small topological gaps.

5.3 Computation and Physical Meaning of the Chern Number

5.3.1 FHS Computation

For numerical Bloch eigenstates, the Chern number is most robustly computed by the Fukui–Hatsugai–Suzuki method. On a discrete Brillouin-zone mesh, define the link variables

$$U_\mu(\mathbf{k}_\ell) = \frac{\langle u_{n,\ell}, u_{n,\ell+\hat{\mu}} \rangle_M}{|\langle u_{n,\ell}, u_{n,\ell+\hat{\mu}} \rangle_M|}, \quad \mu = 1, 2. \quad (5.18)$$

The material-weighted inner product is essential for photonic modes. The lattice Berry flux is

$$\tilde{F}_{12}(\mathbf{k}_\ell) = \text{Im} \log [U_1(\mathbf{k}_\ell) U_2(\mathbf{k}_\ell + \hat{1}) U_1^*(\mathbf{k}_\ell + \hat{2}) U_2^*(\mathbf{k}_\ell)], \quad (5.19)$$

and

$$\mathcal{C}_n = \frac{1}{2\pi} \sum_{\ell} \tilde{F}_{12}(\mathbf{k}_\ell). \quad (5.20)$$

For an isolated band and a sufficiently fine mesh, this returns an integer equal to the continuum Chern number.

5.3.2 $\mathcal{T}$, Berry Curvature, and Gap Closing

In a $\mathcal{T}$ -symmetric two-dimensional system, the Berry curvature is odd in momentum and the Chern number vanishes. Thus a nonzero Chern number requires $\mathcal{T}$ breaking. In the Wang model, $\kappa \neq 0$ supplies precisely this ingredient.

Theorem 5.5 (Necessary condition for a topological phase transition). *During a continuous parameter variation, the Chern number can change only if the relevant bandgap closes and reopens.*

Proof. The Chern number is integer-valued and continuous under smooth deformations of an isolated band. It is therefore constant as long as the band remains isolated. It can jump only when the band ceases to be isolated, i.e. at a band touching. $\square$

The local mechanism is the same as in a two-band Dirac model:

$$H(\mathbf{k}) = d_x(\mathbf{k})\sigma_x + d_y(\mathbf{k})\sigma_y + d_z(\mathbf{k})\sigma_z.$$

The Berry curvature is controlled by the wrapping of $\hat{\mathbf{d}} = \mathbf{d}/|\mathbf{d}|$ over the sphere. When the mass term changes sign at a Dirac point, the Berry flux attached to that point changes and the Chern number jumps. The full Haldane phase diagram is not needed here; the only needed lesson is that magneto-optical $\mathcal{T}$ breaking supplies a Dirac mass whose sign and gap-closing events determine the band Chern numbers.

5.4 Unidirectional Edge States and Bulk–Edge Correspondence

5.4.1 Edge Count

To compute edge states, use a strip or supercell geometry: finite in one direction and Bloch-periodic in the other. For an interface between a photonic Chern insulator and a trivial material or metallic wall, the number of chiral edge branches is determined by the Chern number difference across the interface.

Corollary 5.6 (Number of edge states in a photonic Chern insulator). *Inside a common bandgap, the net number of unidirectional edge branches equals*

$$\Delta\mathcal{C} = \mathcal{C}_{\text{left}} - \mathcal{C}_{\text{right}}.$$

For the Wang interface with $\Delta\mathcal{C} = 1$, one edge branch crosses the gap.

The edge dispersion connects the lower projected bulk band to the upper projected bulk band. It cannot terminate inside the gap without either meeting another counter-propagating branch or closing the bulk gap. This is the spectral-flow form of the bulk–edge correspondence.

5.4.2 Supercell Dispersion and Localization

In a strip calculation, the generalized eigenproblem can be written schematically as

$$\sum_{G'_x} \sum_{j'} H_{(G_x, j), (G'_x, j')}(k_x) c_{n, (G'_x, j')} = \omega_n^2(k_x) \sum_{G'_x} \sum_{j'} S_{(G_x, j), (G'_x, j')}(k_x) c_{n, (G'_x, j')}. \quad (5.21)$$

Here j labels the cell across the strip and G_x labels reciprocal vectors along the periodic direction. The same tensor-PWE construction as in section 5.2.3 is used locally, but one direction is kept in real space.

An edge state is identified by its localization near the interface. Away from the boundary, its field decays approximately as

$$|E_z(j)| \propto e^{-|j-j_{\text{edge}}|/\xi}. \quad (5.22)$$

Near a Dirac gap with velocity v and mass m , the decay length is

$$\xi \simeq \frac{v}{\sqrt{m^2 - (\omega - \omega_D)^2}}, \quad (5.23)$$

and is shortest near the middle of the gap.

5.4.3 Robustness of One-Way Propagation

Theorem 5.7 (Topologically protected unidirectional propagation). *At a frequency inside a complete topological bandgap, if only one edge channel exists and it has group velocity $v_g > 0$, then elastic backscattering into an edge channel with $v_g < 0$ is forbidden because no such channel exists in the gap.*

Physical argument. Bulk states are absent by the bandgap. A counter-propagating edge state is absent because the Chern number difference permits only one chiral branch. Thus a local defect can change the phase and spatial profile of the edge wave, but it has no backward propagating channel into which it can scatter. $\square$

This is stronger than the conditional protection of pseudo-spin or valley schemes. It does not rely on suppressing intervalley or pseudo-spin mixing. Its practical limits instead come from loss, finite-size leakage, disorder that closes the gap, and imperfect excitation into the edge channel.

5.5 Wang Experiment and Physical Interpretation

The Wang experiment used a square-lattice YIG cylinder array at microwave frequencies with a metallic boundary. The applied magnetic field biased the YIG cylinders, giving a large gyromagnetic permeability and a topological TM bandgap near 4.5 GHz.

The observed signatures match the theory:

- forward transmission along the edge remained high across the topological bandgap;
- backward transmission was strongly suppressed;
- a metallic obstacle placed along the edge did not produce appreciable reflection;
- reversing the magnetic bias reverses the propagation direction.

The experiment is the photonic analogue of the quantum Hall effect, but with important differences. There is no filled Fermi sea and no electronic Hall conductance. The topological invariant is a photonic band Chern number computed using the Maxwell energy inner product, and the observable is one-way electromagnetic transport in a bandgap.

5.6 Frequency Limitation and Other Chern Routes

The Wang model succeeds because microwave ferrites provide strong gyromagnetic response. A useful figure of merit is

$$\text{FOM} = \frac{|g(\omega)|}{\varepsilon_1(\omega)} \quad \text{or} \quad \frac{|\kappa(\omega)|}{\mu_\perp(\omega)}, \quad (5.24)$$

depending on whether electric or magnetic gyrotropy is used. The topological gap is typically proportional to this ratio.

At microwave frequencies, ferrites near ferromagnetic resonance can have an order-one magnetic gyrotropic ratio. At telecom or visible frequencies, the magneto-optical response of garnets is orders of magnitude smaller, so the gap is often narrower than fabrication disorder, absorption, and radiation leakage. This limitation motivates the following chapters: pseudo-spin topology avoids magnetic bias by using crystalline symmetry, coupled resonator arrays use geometric phases, and Floquet platforms use periodic driving.

Other Chern-insulator geometries, such as triangular and honeycomb lattices, realize the same logic with Dirac points at different high-symmetry momenta. The triangular and honeycomb cases connect most directly to the Haldane-model picture: a $\mathcal{T}$ -breaking gyrotropic perturbation opens a Dirac gap whose mass sign determines the Chern number. For this chapter, these variants are secondary; the Wang model already contains the full material–band–Chern–edge chain.

Chapter Summary

The photonic Chern insulator is the cleanest realization of chiral topological transport in Maxwell systems. Gyrotropic tensors such as (5.1) and (5.2) break $\mathcal{T}$ while remaining Hermitian in the lossless regime. This allows photonic bands to carry nonzero Chern numbers.

For the two-dimensional gyromagnetic photonic crystal, TM and TE polarizations reduce to scalar eigenproblems with tensor-valued inverse material coefficients. The PWE implementation must use the full tensor inverse rule; otherwise small magneto-optical gaps can be numerically unreliable. The Chern number is then computed from material-weighted Bloch eigenstates, for example by the FHS method.

The bulk–edge correspondence gives the observable consequence: an interface with $\Delta\mathcal{C} = 1$ supports one unidirectional edge branch. Because there is no backward channel inside the gap, elastic backscattering is forbidden in the ideal lossless setting. The Wang microwave experiment verified this prediction using YIG cylinders and observed robust one-way transmission around obstacles.

The main limitation is frequency. Strong gyromagnetic response is available in microwave ferrites, but the magneto-optical figure of merit becomes very small at optical frequencies. That

bottleneck is the reason later chapters develop pseudo-spin, resonator, Floquet, valley, and other routes to topological photonics.

Chapter 6

The Photonic Quantum Spin Hall Effect and $\mathbb{Z}_2$ Topological Insulators

Why Topology Must Be Rebuilt Without Kramers Theorem?

The photonic Chern insulator of Chapter 5 realizes unidirectional edge states by explicitly breaking $\mathcal{T}$. This is conceptually clean, but magneto-optical effects become weak at optical frequencies, so the topological bandgap can be too small for robust integrated devices. The natural question is whether a QSHE-like photonic phase can be obtained without relying on a magnetic bias.

For electrons, the quantum spin Hall effect is protected by fermionic time-reversal symmetry, $\hat{\Theta}^2 = -1$, and by Kramers degeneracy. For photons, ordinary time reversal is bosonic, $\hat{\Theta}^2 = +1$, so this route is unavailable. The Wu–Hu idea is to replace intrinsic spin with crystalline pseudo-spin: C_{6v} symmetry creates the relevant degeneracies, a shrinkage–expansion deformation inverts p - and d -like modes, and the resulting four-band effective theory takes the form of the BHZ model.

Remark (Common misconception warnings).

1. The photonic pseudo-QSHE is not identical to the electronic QSHE. Its pseudo- $\mathcal{T}$ is an emergent crystalline symmetry, not fundamental time reversal.
2. The p and d orbitals are symmetry labels of Bloch modes at Γ , not free-space orbital angular momentum of light.
3. The relevant protection is conditional. Defects, interfaces, or bends that mix pseudo-spin sectors can produce backscattering.

The chapter therefore follows one main line:

bosonic $\mathcal{T} \longrightarrow C_{6v}$ pseudo-spin $\longrightarrow$ BHZ effective model $\longrightarrow$ helical edge states and their limitations.

6.1 Bosonic $\mathcal{T}$ and the Absence of Kramers' Theorem

6.1.1 Time Reversal for Fermions and Bosons

The time-reversal operator $\hat{\Theta}$ is anti-unitary. The crucial distinction is the square of this operator.

Definition 6.1 (Fermionic and bosonic $\mathcal{T}$).

- Fermionic systems with half-integer spin satisfy $\hat{\Theta}^2 = -1$. Kramers' theorem applies.
- Bosonic systems with integer spin satisfy $\hat{\Theta}^2 = +1$. Kramers' theorem does not apply.

Proposition 6.2 ($\mathcal{T}$ type of photons). *For Maxwell fields in the usual lossless reciprocal setting, the physical time-reversal operation satisfies $\hat{\Theta}^2 = +1$. Photons therefore have bosonic $\mathcal{T}$.*

Proof. In the convention used in Chapter 2, time reversal complex conjugates the frequency-domain field and applies the appropriate sign changes to the magnetic components. Applying this operation twice returns the original electromagnetic field. Hence $\hat{\Theta}^2 = +1$. $\square$

Theorem 6.3 (Absence of Kramers' theorem). *In a system with $\hat{\Theta}^2 = +1$, time-reversal symmetry does not force eigenstates to appear in degenerate pairs.*

Proof. Let $|u\rangle$ be a non-degenerate eigenstate. For a fermionic anti-unitary symmetry with $\hat{\Theta}^2 = -1$, $|u\rangle$ and $\hat{\Theta}|u\rangle$ must be linearly independent, producing a Kramers pair. For $\hat{\Theta}^2 = +1$, the relation $\hat{\Theta}|u\rangle = e^{i\phi}|u\rangle$ is self-consistent, so no degeneracy is required. $\square$

Consequently, a two-dimensional photonic system protected only by ordinary bosonic $\mathcal{T}$ is not an electronic class-AII $\mathbb{Z}_2$ topological insulator. To build a QSHE analogue one must construct an effective pseudo-spin and an effective pseudo-time-reversal symmetry inside a restricted photonic subspace.

6.2 Crystalline Symmetry and Pseudo-Spin Construction

6.2.1 C_{6v} Degeneracies at Γ

Consider a two-dimensional photonic crystal with C_{6v} point-group symmetry. At the Γ point, the full point group acts as the little group on Bloch eigenmodes. The two-dimensional irreducible representations E_1 and E_2 therefore enforce twofold degeneracies.

Proposition 6.4 (Symmetry-protected degeneracy at the Γ point). *In a C_{6v} photonic crystal, Bloch modes at Γ belonging to the two-dimensional representations E_1 or E_2 are necessarily twofold degenerate.*

The modes in the E_1 representation are conventionally labeled p_x, p_y , while those in E_2 are labeled $d_{x^2-y^2}, d_{xy}$. These names express transformation properties under the point group; they are not claims about free-space orbital angular momentum.

6.2.2 Shrinkage, Expansion, and Band Inversion

Wu and Hu [8] used a honeycomb-like cluster geometry in which six dielectric cylinders sit around a lattice point. Changing the cluster radius R while preserving C_{6v} symmetry tunes the ordering of the p - and d -like doublets.

Definition 6.5 (Shrunk and expanded configurations). Let a be the lattice constant and R the distance from each cylinder center to the cluster center.

- Shrunk configuration: $R < a/3$.
- Critical configuration: $R = a/3$ in the idealized model.
- Expanded configuration: $R > a/3$.

Theorem 6.6 (Wu–Hu band inversion mechanism). *In the shrunk configuration, the p doublet lies below the d doublet near Γ . In the expanded configuration, the d doublet drops below the p doublet. The transition between the two closes the gap at Γ and produces a fourfold Dirac point in the ideal symmetric model.*

Physical argument. The p and d modes belong to different two-dimensional representations of C_{6v} . As the cluster is expanded, inter-cell coupling changes the relative energies of these representations. At the inversion point the two doublets meet; a $\mathbf{k} \cdot \mathbf{p}$ expansion around Γ gives two coupled Dirac cones. Changing the ordering of the doublets is therefore equivalent to changing the sign of the Dirac mass in the effective theory. $\square$

6.2.3 Pseudo-Spin Basis

Definition 6.7 (Photonic pseudo-spin). Define

$$|p_+\rangle = \frac{1}{\sqrt{2}}(|p_x\rangle + i|p_y\rangle), \quad |p_-\rangle = \frac{1}{\sqrt{2}}(|p_x\rangle - i|p_y\rangle), \quad (6.1)$$

$$|d_+\rangle = \frac{1}{\sqrt{2}}(|d_{x^2-y^2}\rangle + i|d_{xy}\rangle), \quad |d_-\rangle = \frac{1}{\sqrt{2}}(|d_{x^2-y^2}\rangle - i|d_{xy}\rangle). \quad (6.2)$$

The subspace $\{|p_+\rangle, |d_+\rangle\}$ is called pseudo-spin up, and $\{|p_-\rangle, |d_-\rangle\}$ is called pseudo-spin down.

The point of this construction is not merely notation. In this basis, symmetry forces the low-energy Hamiltonian to split into two time-reversed blocks, closely paralleling the BHZ model for the electronic QSHE [22].

6.3 Pseudo-Time-Reversal Symmetry and the Effective BHZ Model

6.3.1 $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian Near Γ

Use the basis

$$\{|p_+\rangle, |d_+\rangle, |p_-\rangle, |d_-\rangle\}.$$

To second order in $\mathbf{k}$, the C_{6v} constraints give the following four-band effective Hamiltonian.

Theorem 6.8 (Photonic BHZ effective Hamiltonian). *Near Γ , the Wu–Hu photonic crystal is described by*

$$\mathcal{H}_{\text{eff}}(\mathbf{k}) = \begin{pmatrix} h(\mathbf{k}) & 0 \\ 0 & h^*(-\mathbf{k}) \end{pmatrix}, \quad (6.3)$$

where

$$h(\mathbf{k}) = \epsilon(\mathbf{k})\mathbb{I}_2 + \begin{pmatrix} M(\mathbf{k}) & A(k_x + ik_y) \\ A(k_x - ik_y) & -M(\mathbf{k}) \end{pmatrix}. \quad (6.4)$$

Here

$$\epsilon(\mathbf{k}) = C + D(k_x^2 + k_y^2), \quad M(\mathbf{k}) = M_0 + B(k_x^2 + k_y^2).$$

The upper block acts on pseudo-spin up and the lower block on pseudo-spin down.

Symmetry derivation. Under a C_6 rotation,

$$|p_{\pm}\rangle \mapsto e^{\mp i\pi/3} |p_{\pm}\rangle, \quad |d_{\pm}\rangle \mapsto e^{\mp 2i\pi/3} |d_{\pm}\rangle,$$

while $k_{\pm} = k_x \pm ik_y$ transforms as $k_{\pm} \mapsto e^{\pm i\pi/3} k_{\pm}$. In the upper block, the angular momentum difference between p_+ and d_+ is one unit, so the lowest allowed off-diagonal terms are proportional to k_+ and k_- . The diagonal terms must have zero angular momentum and therefore start from constants plus k^2 .

Mirror symmetry maps $p_+ \leftrightarrow p_-$, $d_+ \leftrightarrow d_-$, and $k_y \rightarrow -k_y$. This fixes the lower block to be $h^*(-\mathbf{k})$ up to basis conventions. Combining these constraints gives (6.3)–(6.4). $\square$

The parameters have direct physical meanings. A is the Dirac velocity; M_0 is the band-inversion mass, essentially the relative ordering of p and d doublets; B regularizes the mass away from Γ ; C and D shift and tilt the bands without changing the topology. The band-inverted regime is detected by the relative sign of the mass at $\mathbf{k} = 0$ and at the regularization scale. In the convention above,

$$\boxed{M_0/B < 0 \text{ topological,} \quad M_0/B > 0 \text{ trivial.}} \quad (6.5)$$

6.3.2 Pseudo- $\mathcal{T}$

The block structure suggests an effective anti-unitary operation

$$\hat{\mathcal{T}} = \begin{pmatrix} 0 & -\mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{pmatrix} \hat{K}, \quad (6.6)$$

where $\hat{K}$ denotes complex conjugation. This operator satisfies

$$\hat{\mathcal{T}}^2 = -1, \quad \hat{\mathcal{T}} \mathcal{H}_{\text{eff}}(\mathbf{k}) \hat{\mathcal{T}}^{-1} = \mathcal{H}_{\text{eff}}(-\mathbf{k}).$$

Proposition 6.9 (Pseudo-Kramers degeneracy). *At momenta where the four-band effective model and pseudo- $\mathcal{T}$ are valid, the operator $\hat{\mathcal{T}}$ produces Kramers-like pairs because $\hat{\mathcal{T}}^2 = -1$.*

Proof. The matrix part of (6.6) squares to $-\mathbb{I}_4$, and it is real, so complex conjugation does not change it. Thus $\hat{\mathcal{T}}^2 = -\mathbb{I}_4$. The standard Kramers argument then applies inside this effective subspace: $|u\rangle$ and $\hat{\mathcal{T}}|u\rangle$ are orthogonal and degenerate whenever they belong to the same momentum sector. $\square$

Remark (Fragility of pseudo- $\mathcal{T}$). Pseudo- $\mathcal{T}$ is produced by crystalline symmetry and by the separation of the two pseudo-spin blocks. It can fail when disorder, boundary orientation, or higher-order terms mix pseudo-spin sectors. This is the essential difference between the photonic pseudo-QSHE and the electronic QSHE: the former is symmetry-engineered and conditional, while the latter is protected by fundamental fermionic time reversal.

6.4 The Pseudo- $\mathbb{Z}_2$ Invariant

6.4.1 Pseudo-Spin Chern Number

Because (6.3) is block diagonal, one can define Chern numbers for the two pseudo-spin blocks separately.

Definition 6.10 (Pseudo-spin Chern number).

$$\mathcal{C}_\uparrow = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_\uparrow(\mathbf{k}) d^2k, \quad \mathcal{C}_\downarrow = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_\downarrow(\mathbf{k}) d^2k. \quad (6.7)$$

True $\mathcal{T}$ gives $\mathcal{C}_\uparrow + \mathcal{C}_\downarrow = 0$. The pseudo-spin Chern number is

$$\mathcal{C}_s = \frac{\mathcal{C}_\uparrow - \mathcal{C}_\downarrow}{2}. \quad (6.8)$$

Theorem 6.11 (Pseudo-spin Chern number of the BHZ model). *For (6.3)–(6.4),*

$$M_0/B < 0 \Rightarrow |\mathcal{C}_s| = 1, \quad M_0/B > 0 \Rightarrow \mathcal{C}_s = 0.$$

Proof. For the upper block,

$$\mathbf{d}(\mathbf{k}) = (Ak_x, -Ak_y, M_0 + Bk^2)$$

maps momentum space to a two-level Bloch sphere after continuum compactification. If M_0 and the regularizing mass Bk^2 have opposite signs, the normalized vector $\hat{\mathbf{d}}$ covers the sphere once and the block has Chern number ± 1 . If they have the same sign, the mapping is topologically trivial. The lower block is the time-reversed partner and has the opposite Chern number. $\square$

When pseudo- $\mathcal{T}$ is maintained, the parity of $\mathcal{C}_s$ plays the role of a pseudo- $\mathbb{Z}_2$ index:

$$\nu = \mathcal{C}_s \mod 2. \quad (6.9)$$

This should be read as a crystalline symmetry-protected index, not as the fundamental class-AII invariant of electronic topological insulators.

6.5 Helical Edge States

6.5.1 Bulk–Edge Correspondence

Theorem 6.12 (Pseudo-spin bulk–edge correspondence). *At an interface between the expanded phase ($|\mathcal{C}_s| = 1$) and the shrunk phase ($\mathcal{C}_s = 0$), the ideal pseudo-spin-conserving theory supports a pair of helical edge states: pseudo-spin up propagates in one direction and pseudo-spin down propagates in the opposite direction.*

Argument. The pseudo-spin-up block has a Chern-number difference of one across the interface, so it contributes one chiral edge mode. The pseudo-spin-down block has the opposite Chern-

number difference and contributes the counter-propagating partner. The two modes cross at the projected Γ point when pseudo- $\mathcal{T}$ is preserved. $\square$

This produces pseudo-spin-momentum locking:

Proposition 6.13 (Pseudo-spin-momentum locking). *In the helical edge-state subspace, propagation direction is correlated with pseudo-spin. A source that overlaps predominantly with one pseudo-spin preferentially excites one propagation direction.*

6.5.2 Analytical Edge-State Solution

The edge dispersion can be derived directly from the BHZ Hamiltonian. Consider an interface at $y = 0$, with translation symmetry along x . Let the mass change sign:

$$M_0(y) = \begin{cases} M_T < 0, & y > 0 \quad (\text{topological}), \\ M_P > 0, & y < 0 \quad (\text{trivial}). \end{cases} \quad (6.10)$$

Neglecting C, D and retaining the leading Dirac terms, the upper block becomes

$$h_{\uparrow}(k_x, -i\partial_y) = Ak_x\sigma_x - iA\sigma_y\partial_y + M_0(y)\sigma_z. \quad (6.11)$$

Example 6.14 (Helical edge states of the semi-infinite BHZ model). At $k_x = 0$, the bound-state equation is

$$[-iA\sigma_y\partial_y + M_0(y)\sigma_z]\phi(y) = 0. \quad (6.12)$$

Multiplying by σ_y and using $\sigma_y\sigma_z = i\sigma_x$ gives

$$\partial_y\phi(y) = \frac{M_0(y)}{A}\sigma_x\phi(y).$$

Choose the σ_x eigenstate $\chi_+ = (1, 1)^T/\sqrt{2}$. Then

$$\boxed{\phi_0(y) = f_0 \exp\left[\frac{1}{A} \int_0^y M_0(y') dy'\right] \chi_+}. \quad (6.13)$$

This state is normalizable exactly because $M_0(y)$ changes sign. The penetration lengths on the two sides are

$$\xi_T = \frac{|A|}{|M_T|}, \quad \xi_P = \frac{|A|}{|M_P|}. \quad (6.14)$$

Treating $Ak_x\sigma_x$ as a perturbation on this bound state gives

$$\boxed{E_{\uparrow}(k_x) = Ak_x + O(k_x^3)}. \quad (6.15)$$

The lower block is related by pseudo- $\mathcal{T}$, so

$$E_{\downarrow}(k_x) = -Ak_x + O(k_x^3). \quad (6.16)$$

The two branches cross at $k_x = 0$ and form the helical edge pair.

If the quadratic mass term is restored, the effective decay length varies as

$$\xi(k_x) \simeq \frac{|A|}{|M_0 + Bk_x^2|}.$$

It diverges as the edge dispersion merges into the bulk continuum. This gives a practical estimate for the required supercell width in numerical calculations and for the minimum device width needed to suppress hybridization between opposite edges.

6.6 Backscattering Protection and Its Limitations

6.6.1 Selection Rule

In the ideal effective model, pseudo-spin up and pseudo-spin down do not mix. Since the two propagation directions are tied to opposite pseudo-spins, backscattering requires a perturbation that couples the two pseudo-spin sectors. Forward scattering within the same pseudo-spin channel is allowed, but it does not reverse propagation.

Let $\hat{V}$ be a local perturbation. In the projected edge subspace, the backscattering matrix element is

$$T_{\downarrow\leftarrow\uparrow} = \langle -k_x, \downarrow | \hat{V} | k_x, \uparrow \rangle.$$

If $\hat{V}$ preserves the crystalline selection rules that define pseudo-spin, this matrix element vanishes to leading order. If $\hat{V}$ contains symmetry-breaking components, it can become nonzero.

6.6.2 Quantitative Scaling

Decompose the perturbation into C_{6v} symmetry channels,

$$\hat{V} = V_{A_1} \hat{P}_{A_1} + V_{A_2} \hat{P}_{A_2} + V_{E_1} \hat{P}_{E_1} + V_{E_2} \hat{P}_{E_2} + \dots \quad (6.17)$$

The fully symmetric A_1 component does not mix pseudo-spins. Components that break the relevant rotational or mirror selection rules can couple the helical partners. In the Born approximation,

$$|T_{\downarrow\leftarrow\uparrow}|^2 \propto |V_E|^2 |\langle \phi_{\downarrow} | \hat{P}_E | \phi_{\uparrow} \rangle|^2, \quad (6.18)$$

where E denotes a pseudo-spin-mixing symmetry channel. The corresponding reflectance scales as

$$R \propto \left| \frac{V_E}{\Delta\omega_{\text{gap}}} \right|^2. \quad (6.19)$$

This formula captures the engineering lesson. The gap suppresses coupling to bulk modes, while crystalline selection rules suppress backscattering between edge modes. Neither suppression is absolute. Strong defects, sharp non-compatible bends, or disorder with large pseudo-spin-mixing components require full-wave simulation rather than relying only on the effective model.

Remark (Difference from electronic QSHE). In an electronic QSHE edge, non-magnetic elastic backscattering of a single helical pair is forbidden by fundamental $\mathcal{T}$ and Kramers theorem. In a photonic pseudo-QSHE edge, backscattering is only conditionally suppressed by crystalline symmetry. This is why lattice-compatible bends can be robust while arbitrary bends may reflect.

6.7 Experimental Realization of the Wu–Hu Model

The original Wu–Hu proposal [8] uses an all-dielectric photonic crystal with shrunken and expanded clusters. Microwave-scale implementations often use effectively two-dimensional parallel-plate or rod structures, where the relevant modes are well described by a scalar TM-like model. The main observations are:

- a band inversion between p - and d -like doublets at Γ ;
- helical edge bands at an expanded–shrunken interface;
- pseudo-spin-selective directional excitation;
- high transmission around symmetry-compatible bends;
- finite reflection when defects break the protecting crystalline selection rules.

At optical frequencies, related designs have been explored in photonic crystal slabs and waveguide platforms. The same symmetry logic applies, but radiation loss, finite slab thickness, polarization mixing, and fabrication disorder must be included in the design. For the purpose of this chapter, these implementations are evidence for the same main point: pseudo-QSHE physics is experimentally useful, but its protection must always be stated with the symmetry assumptions attached.

The Khanikaev bianisotropic scheme [7] is historically important, but it follows a different route: pseudo-spin is constructed from electromagnetic duality and bianisotropic $\mathbf{E}$ – $\mathbf{H}$ coupling rather than from p/d crystalline band inversion. For this chapter it is best treated as a related platform rather than a second main derivation.

Chapter Summary

Photons have bosonic time reversal, $\hat{\Theta}^2 = +1$, so ordinary photonic $\mathcal{T}$ does not provide Kramers degeneracy or the electronic class-AII $\mathbb{Z}_2$ invariant. The Wu–Hu mechanism rebuilds a QSHE-like structure by using C_{6v} crystalline symmetry to define $p_{\pm}/d_{\pm}$ pseudo-spins and by tuning a shrinkage–expansion band inversion at Γ .

The effective four-band theory is the photonic BHZ Hamiltonian

$$\mathcal{H}_{\text{eff}}(\mathbf{k}) = \begin{pmatrix} h(\mathbf{k}) & 0 \\ 0 & h^*(-\mathbf{k}) \end{pmatrix},$$

with $M_0/B < 0$ labeling the band-inverted phase. Separate Chern numbers can be assigned to the two pseudo-spin blocks, and their parity defines a pseudo- $\mathbb{Z}_2$ index when the crystalline symmetry and block separation remain valid.

At an expanded–shrunken interface, the mass changes sign and produces a Jackiw–Rebbi edge state. The two pseudo-spin blocks give opposite dispersions, $E_{\uparrow}(k_x) = Ak_x$ and $E_{\downarrow}(k_x) = -Ak_x$, forming a helical pair. The protection is useful but conditional: backscattering is suppressed when defects preserve the pseudo-spin selection rules, and finite when symmetry-breaking components mix the two sectors.

Part II

Synthetic, Driven, and Open Photonic Topology

Chapter 7

Coupled Resonators and Synthetic Gauge Fields

Why Topology Can Be Engineered Through Coupling?

The two preceding application chapters were based on continuous-medium photonic crystals: topological properties were determined by material tensors, lattice symmetry, and the spectrum of the Maxwell operator. Coupled resonator arrays introduce a different route. Instead of engineering topology primarily through a continuous dielectric landscape, one first isolates a small set of resonant modes and then engineers their couplings.

This shift has three consequences. First, the Maxwell problem can be projected to a tight-binding model when the resonators are high- Q , the modes are spectrally isolated, and the inter-site coupling is weak. Second, hopping phases can be controlled by optical path length, so a synthetic gauge field can be produced without using magneto-optical materials. Third, internal degrees of freedom such as frequency modes can be reinterpreted as synthetic lattice coordinates, making higher-dimensional topological models accessible in lower-dimensional physical structures.

Remark (Common misconception warnings).

1. The tight-binding model is not an uncontrolled simplification of Maxwell's equations. It is a controlled projection under high- Q , isolated-mode, weak-coupling conditions.
2. A synthetic magnetic flux is not a real magnetic field. It reproduces the Peierls phases of a magnetic field in the effective lattice Hamiltonian, but it originates from optical propagation phases.
3. The pseudo- $\mathcal{T}$ of microring arrays is not fundamental electronic time-reversal symmetry. It relies on the near-decoupling of clockwise and counterclockwise circulation sectors and is sensitive to backscattering.

The organizing chain of this chapter is therefore

Maxwell modes $\longrightarrow$ coupled-mode equations $\longrightarrow$ tight-binding lattice
 $\longrightarrow$ Peierls phases and topology.

The main physical platform will be microring resonator arrays, where this chain is especially transparent.

7.1 From Continuous Media to the Tight-Binding Model

7.1.1 Physical Motivation for Discretization

In Chapter 3, the photonic band structure was obtained from a continuous Maxwell operator $\hat{N}(\mathbf{k})$. For coupled resonator arrays, the relevant low-energy degrees of freedom are localized resonant modes. If each resonator supports one dominant mode near ω_0 , the field can be expanded in this localized-mode basis and the Maxwell eigenproblem reduces to a matrix eigenproblem.

Definition 7.1 (Tight-binding Hamiltonian). Let each lattice site $\mathbf{R}_i$ support one mode $|i\rangle$ at frequency ω_i . The nearest-neighbor coupling coefficient is t_{ij} , generally complex. The tight-binding Hamiltonian is

$$\hat{H}_{\text{TB}} = \sum_i \omega_i |i\rangle \langle i| + \sum_{\langle i,j \rangle} t_{ij} |i\rangle \langle j|, \quad (7.1)$$

where $\langle i, j \rangle$ denotes nearest-neighbor pairs. Hermiticity requires $t_{ji} = t_{ij}^*$.

The approximation is meaningful only within a hierarchy of scales. The modal spacing or free spectral range must be much larger than the coupling strength and the linewidth, so that a single-mode description is valid. For coherent band transport, the coupling strength must also exceed the relevant loss rate. Nearest-neighbor truncation further requires next-nearest-neighbor overlap to be small compared with the dominant coupling.

7.1.2 Peierls Substitution and Synthetic Flux

In electronic tight-binding models, a magnetic field enters through the Peierls substitution

$$t_{ij} \rightarrow t_{ij} e^{i\phi_{ij}}, \quad \phi_{ij} = \frac{e}{\hbar} \int_{\mathbf{R}_j}^{\mathbf{R}_i} \mathbf{A} \cdot d\mathbf{l}.$$

In resonator arrays the same mathematical structure can be engineered by optical path phases. The relevant observable is not an individual link phase, which depends on the choice of site phases, but the accumulated phase around a closed loop.

Definition 7.2 (Synthetic magnetic flux). For a closed path C enclosing a plaquette, the synthetic magnetic flux is

$$\Phi_C = \sum_{(i,j) \in C} \phi_{ij} \pmod{2\pi}, \quad (7.2)$$

where the sum is taken counterclockwise. The condition $\Phi_C \neq 0 \pmod{2\pi}$ means that the effective tight-binding model feels a nonzero gauge flux.

Proposition 7.3 (Gauge invariance of the synthetic flux). Under the site-dependent phase redefinition $|i\rangle \rightarrow e^{i\chi_i} |i\rangle$, the link phase transforms as

$$\phi_{ij} \rightarrow \phi_{ij} + \chi_i - \chi_j.$$

The sum around any closed loop is unchanged because all site phases cancel. Thus Φ_C is gauge-invariant.

7.1.3 Coupled-Mode Derivation from Maxwell's Equations

The tight-binding model in (7.1) should be viewed as a projection of Maxwell's equations, not as a phenomenological assumption. Consider two high- Q dielectric resonators separated by a distance d , each supporting isolated modes $\mathbf{E}_a(\mathbf{r})$ and $\mathbf{E}_b(\mathbf{r})$ at frequencies ω_a and ω_b .

The isolated resonator a satisfies

$$\nabla \times \frac{1}{\mu_a(\mathbf{r})} \nabla \times \mathbf{E}_a = \omega_a^2 \varepsilon_a(\mathbf{r}) \mathbf{E}_a, \quad (7.3)$$

with the mode normalized as $\int \varepsilon_a |\mathbf{E}_a|^2 d^3r = 1$. When the resonators are brought close together, the total field is expanded as

$$\mathbf{E}(\mathbf{r}, t) = c_a(t) \mathbf{E}_a(\mathbf{r}) + c_b(t) \mathbf{E}_b(\mathbf{r}).$$

Substitution into the full Maxwell equation and projection onto the two-mode subspace gives the coupled-mode eigenproblem

$$\begin{pmatrix} \omega_a & \kappa_{ab} \\ \kappa_{ba} & \omega_b \end{pmatrix} \begin{pmatrix} c_a \\ c_b \end{pmatrix} = \omega \begin{pmatrix} c_a \\ c_b \end{pmatrix}. \quad (7.4)$$

Up to convention-dependent normalization factors, the coupling coefficient has the schematic overlap form

$$\kappa_{ab} = \int (\varepsilon - \varepsilon_a) \mathbf{E}_a^* \cdot \mathbf{E}_b d^3r + \int \left(\frac{1}{\mu} - \frac{1}{\mu_a} \right) (\nabla \times \mathbf{E}_a^*) \cdot (\nabla \times \mathbf{E}_b) d^3r. \quad (7.5)$$

The coupling is controlled by the evanescent tail of one resonator mode inside the perturbation region of the other resonator. For dielectric resonators this overlap typically decays exponentially with distance, $|\kappa_{ab}| \sim e^{-d/l_c}$, giving a physical justification for nearest-neighbor truncation.

For a reciprocal lossless two-resonator system, $\kappa_{ba} = \kappa_{ab}^*$, and (7.4) is exactly the two-site tight-binding Hamiltonian

$$\boxed{\hat{H}_{\text{TB}} = \omega_a |a\rangle \langle a| + \omega_b |b\rangle \langle b| + \kappa_{ab} |a\rangle \langle b| + \kappa_{ab}^* |b\rangle \langle a|}. \quad (7.6)$$

Generalizing to many resonators gives Definition 7.1.

Remark (Validity of the tight-binding projection). The useful dimensionless checks are $|\kappa|/\Delta_{\text{mode}} \ll 1$ for single-mode projection, $\gamma/|\kappa| \ll 1$ for coherent transport, and $|\kappa_{\text{nnn}}/\kappa_{\text{nn}}| \ll 1$ for nearest-neighbor truncation. These inequalities are the practical boundary between a resonator lattice and a full multimode Maxwell problem.

7.2 The Hofstadter Model and Landau-Gauge Flux

7.2.1 Uniform Flux on a Square Lattice

The canonical topological model for a resonator array is the Hofstadter model: a square lattice with uniform flux Φ through each plaquette. In the Landau gauge,

$$\phi_{ij}^{(x)} = 0, \quad \phi_{ij}^{(y)} = 2\pi\alpha m,$$

where m is the x -coordinate of the site and $\alpha = \Phi/(2\pi)$ is the reduced flux.

Theorem 7.4 (Harper equation). *For the square-lattice Hofstadter model in the Landau gauge, the Bloch eigenvalue equation reduces to*

$$t u_{m+1} + t u_{m-1} + 2t \cos(2\pi\alpha m + k_y a) u_m = E u_m, \quad (7.7)$$

where u_m is the wave amplitude at the m -th site in the x direction and k_y is the Bloch wave vector along y .

Proof. Take the hopping Hamiltonian

$$\hat{H} = t \sum_{m,n} \left[|m+1, n\rangle \langle m, n| + e^{i2\pi\alpha m} |m, n+1\rangle \langle m, n| + \text{h.c.} \right]. \quad (7.8)$$

Using $\psi_{m,n} = e^{ik_y n a} u_m$, hopping in the y direction contributes $2t \cos(2\pi\alpha m + k_y a) u_m$. The remaining two terms are the nearest-neighbor difference operator in m , which gives (7.7). $\square$

7.2.2 Hofstadter Butterfly and TKNN Labels

When $\alpha = p/q$ is rational, the Harper equation has period q and the original band splits into q subbands. The spectrum as a function of α forms the Hofstadter butterfly. For this chapter, the essential fact is not the fractal itself, but the way each open gap is labeled by an integer Chern number.

Theorem 7.5 (Topological structure of the Hofstadter butterfly). *For $\alpha = p/q$, the Chern numbers $\{C_n\}$ of the Hofstadter subbands satisfy the TKNN Diophantine equation*

$$\sum_{n=1}^r C_n = t_r, \quad r = q s_r + p t_r, \quad (7.9)$$

where $s_r, t_r \in \mathbb{Z}$ and the representative is chosen with $|t_r| \leq q/2$.

For the experimentally common value $\alpha = 1/4$, the Harper equation becomes a 4×4 Bloch problem. In a folded Brillouin zone, one may write

$$\begin{pmatrix} 2 \cos \theta_0 & 1 & 0 & e^{-ik_x \cdot 4a} \\ 1 & 2 \cos \theta_1 & 1 & 0 \\ 0 & 1 & 2 \cos \theta_2 & 1 \\ e^{ik_x \cdot 4a} & 0 & 1 & 2 \cos \theta_3 \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix} = \frac{E}{t} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad (7.10)$$

with $\theta_m = \pi m/2 + k_y a$. The TKNN equation gives the common Chern assignment

$$(C_1, C_2, C_3, C_4) = (1, 1, -3, 1),$$

with the central gap depending on the perturbation that opens it in the nearest-neighbor square-lattice model. The robust experimental message is that the relevant gaps carry chiral edge channels whose number is fixed by the gap Chern label.

Remark. The number-theoretic structure of the full butterfly, including irrational flux and Cantor-set spectra, is conceptually beautiful but not central to the photonic implementation discussed here. For resonator arrays, the main role of the Hofstadter model is to translate engineered link phases into measurable topological bands and edge states.

7.3 Microring Resonator Arrays

7.3.1 Ring Resonators as Lattice Sites

Definition 7.6 (Microring resonator lattice site). A single microring resonator supports two nearly degenerate propagating modes: clockwise (CW) and counterclockwise (CCW). The angular resonance frequency is

$$\omega_0 = \frac{2\pi mc}{n_{\text{eff}} L},$$

where m is the longitudinal mode order, L is the ring circumference, and n_{eff} is the effective refractive index. If CW–CCW backscattering is negligible, each ring behaves as a two-state site.

Adjacent rings are coupled by direct evanescent coupling or by link rings. The propagation phase accumulated by light along a connecting path,

$$\phi = n_{\text{eff}} k_0 L_{\text{path}},$$

becomes the Peierls phase of the effective hopping coefficient.

Proposition 7.7 (Hopping phase and path length). *In a microring resonator array,*

$$t_{ij} = |t_{ij}| e^{i\phi_{ij}}, \quad \phi_{ij} = n_{\text{eff}} k_0 L_{ij} + \phi_{\text{coupler}}. \quad (7.11)$$

By designing L_{ij} to vary with position, one can realize $\Phi_C = \sum_C \phi_{ij} \neq 0$ and hence a synthetic magnetic flux.

7.3.2 The Hafezi Scheme

Hafezi et al. proposed and experimentally realized a photonic Hofstadter model in microring resonator arrays [23, 24]. The site rings form a square lattice, while auxiliary link rings mediate hopping. The x -direction hopping phases are taken uniform, and the y -direction link ring lengths vary linearly with the x coordinate. This implements the Landau gauge,

$$\phi_y(m) = 2\pi\alpha m + \phi_0.$$

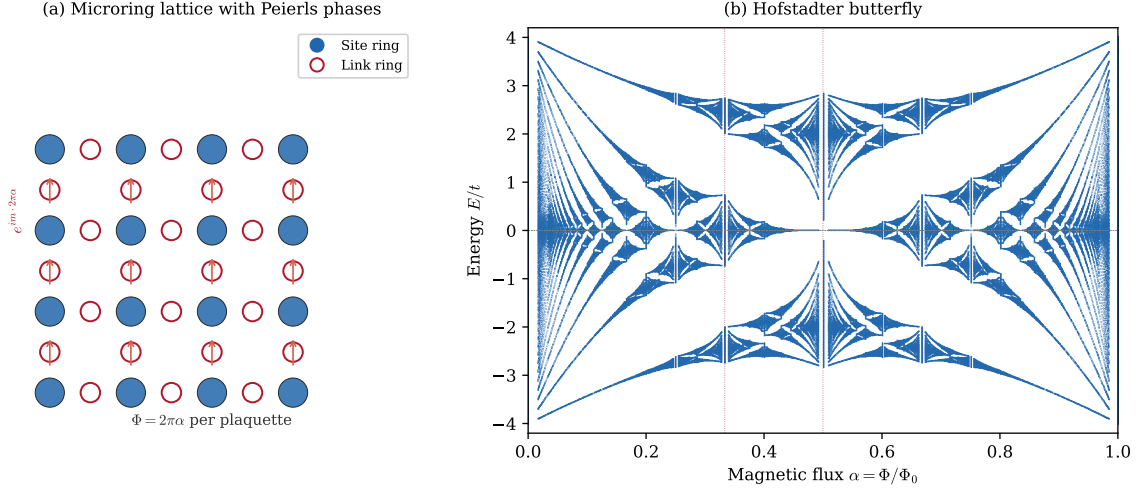


Figure 7.1 Peierls-phase engineering in microring resonator arrays. (a) Site rings and link rings implement position-dependent hopping phases. (b) The corresponding Harper–Hofstadter spectrum splits into flux-dependent topological subbands.

Taking $\alpha = 1/4$, the spectrum splits into four subbands, and edge states appear in the gaps. The 2013 InP experiment observed one-way edge propagation within a selected circulation sector [24]. This is the key point: the synthetic gauge field is produced by lithographic path-length engineering rather than by a magnetic optical response.

7.3.3 Effective Hopping from Link Rings

The quantitative connection between geometry and flux follows from a simple coupled-mode/scattering picture. Let site rings of perimeter L_s be coupled pairwise by link rings of perimeter L_l . At each directional coupler, use the unitary scattering matrix

$$S_{\text{coupler}} = \begin{pmatrix} \tau & i\kappa \\ i\kappa & \tau \end{pmatrix}, \quad |\tau|^2 + |\kappa|^2 = 1. \quad (7.12)$$

A photon hopping from site ring A to site ring B couples out with amplitude $i\kappa$, propagates through half of the link ring with phase $e^{i\phi_l/2}$, and couples back with another factor $i\kappa$. Thus

$$t_{AB} = -\kappa^2 e^{i\phi_l/2}, \quad (7.13)$$

where $\phi_l = n_{\text{eff}} k_0 L_l$. The hopping strength is set by $|\kappa|^2$, while the hopping phase is set by the link-ring propagation phase.

To realize the Landau gauge, choose link-ring lengths along the y direction as

$$L_l(m_x) = L_0 + m_x \delta L. \quad (7.14)$$

Then

$$\phi(m_x) = \frac{n_{\text{eff}} k_0}{2} (L_0 + m_x \delta L) = \phi_0 + m_x \underbrace{\frac{n_{\text{eff}} k_0 \delta L}{2}}_{2\pi\alpha}. \quad (7.15)$$

The flux per plaquette is therefore

$$\boxed{\Phi = 2\pi\alpha = \frac{\pi n_{\text{eff}}\delta L}{\lambda}}. \quad (7.16)$$

For $\alpha = 1/4$, this gives $\delta L = \lambda/(2n_{\text{eff}})$. At telecom wavelengths with $n_{\text{eff}} \approx 2.8$, the required length gradient is on the order of hundreds of nanometers, within lithographic control.

Remark (Physical essence of the link-ring gradient). Equation (7.16) is the central design formula of the Hafezi scheme. A geometric length gradient plays the role of a magnetic vector potential in the projected tight-binding model. The phase is nevertheless dispersive through $n_{\text{eff}}(\lambda)$ and k_0L , so the flux is designed for an operating wavelength rather than being a frequency-independent material constant.

7.3.4 CW/CCW Pseudo-Spin

Microring resonators naturally support CW and CCW propagation directions. For the CW mode, propagation through a link ring contributes a phase $+\phi_l/2$; for the CCW mode, the reversed path contributes $-\phi_l/2$. The two circulation sectors therefore experience opposite synthetic magnetic fields.

Proposition 7.8 (Intrinsic pseudo- $\mathcal{T}$ of microring arrays). *In a microring array without CW-CCW coupling, the effective Hamiltonians satisfy*

$$H_{CCW}(\mathbf{k}) = H_{CW}^*(-\mathbf{k}). \quad (7.17)$$

This has the same block structure as time-reversed quantum-spin-Hall models, but the symmetry is pseudo- $\mathcal{T}$ rather than fundamental $\mathcal{T}$.

Backscattering from sidewall roughness, asymmetric bending loss, or fabrication imperfections can mix CW and CCW sectors. Thus the microring scheme can realize robust transport within a circulation sector, but its protection must be assessed through the actual backscattering rate and loss, not only through the ideal topological model.

7.4 Synthetic Dimensions

7.4.1 Internal Degrees of Freedom as Lattice Sites

Synthetic dimensions reinterpret a discrete internal degree of freedom as a lattice coordinate. The most important example for resonator arrays is the frequency ladder of a microring.

Definition 7.9 (Synthetic dimension). Let a physical system have discrete internal states $\{m\}$ and controllable couplings between them. Treating m as a lattice coordinate produces an effective extra dimension. A one-dimensional physical array plus one synthetic dimension can therefore simulate a two-dimensional lattice.

A single microring supports resonances

$$\omega_m = \omega_0 + m\Omega_{\text{FSR}}.$$

If an electro-optic modulator is driven at $\Omega = \Omega_{\text{FSR}}$, it couples neighboring frequency modes $m \leftrightarrow m \pm 1$. The modulation phase becomes a controllable hopping phase in frequency space.

7.4.2 Frequency-Space Coupled-Mode Equations

For a microring of circumference L , the resonances are approximately

$$\omega_m = \omega_0 + m\Omega_{\text{FSR}}, \quad \Omega_{\text{FSR}} = \frac{2\pi c}{n_{\text{eff}}L}.$$

An electro-optic drive $V(t) = V_0 \cos(\Omega t + \phi)$ produces the refractive-index modulation

$$\delta n(t) = \frac{1}{2} n_{\text{eff}}^3 r_{33} \frac{V_0}{d} \cos(\Omega t + \phi), \quad (7.18)$$

where r_{33} is the Pockels coefficient and d is the electrode gap. Expanding the intracavity field as

$$\mathbf{E}(t) = \sum_m a_m(t) e^{-i\omega_m t} \mathbf{e}_m(\mathbf{r})$$

and applying the rotating-wave approximation gives

$$i \frac{da_m}{dt} = \omega_m a_m + g(e^{i\phi} a_{m-1} + e^{-i\phi} a_{m+1}) - i \frac{\gamma_m}{2} a_m, \quad (7.19)$$

with an effective modulation-induced hopping

$$g = \frac{\omega_0}{2} \cdot \frac{1}{2} n_{\text{eff}}^2 r_{33} \frac{V_0}{d}.$$

The modulation phase ϕ is therefore a synthetic gauge parameter.

For a one-dimensional chain of such rings, with spatial coupling J and a ring-dependent modulation phase $\phi_n = \phi_0 + 2\pi\alpha n$, the effective two-dimensional lattice equation is

$$E a_{n,m} = \omega_m a_{n,m} + J(a_{n+1,m} + a_{n-1,m}) + g(e^{i\phi_n} a_{n,m-1} + e^{-i\phi_n} a_{n,m+1}). \quad (7.20)$$

Here n is a real-space coordinate and m is the frequency synthetic coordinate. Equation (7.20) is the Harper model with flux $\Phi = 2\pi\alpha$.

Remark (Advantages and limitations of frequency synthetic dimensions). The advantage is that a one-dimensional physical structure can emulate a two-dimensional topological lattice, with tunable flux controlled by electronic modulation phases. The limitations are equally important: the synthetic dimension has finite size, mode dispersion detunes high-order frequency sites, losses are mode-dependent, and the boundary is usually soft rather than a hard geometric termination.

7.4.3 Thouless Pumping

Synthetic dimensions naturally connect to topological pumping. Consider a one-dimensional periodic system with a cyclic parameter θ , such as a slowly varied modulation phase. In the adiabatic limit, the net displacement per cycle is fixed by the Chern number over the (k, θ) parameter space.

Theorem 7.10 (Thouless pumping and the Chern number). *For a one-dimensional band that remains gapped during one adiabatic cycle of θ , the center-of-mass displacement is*

$$\Delta x = C_1 a, \quad (7.21)$$

where C_1 is the first Chern number on the two-dimensional parameter space (k, θ) .

A standard representative is the Rice–Mele model with

$$t_1(\theta) = t + \delta t \cos \theta, \quad t_2(\theta) = t - \delta t \cos \theta, \quad \Delta(\theta) = \Delta_0 \sin \theta.$$

When θ winds once around the gap-closing point in parameter space, the Wannier center shifts by an integer number of lattice constants. In frequency synthetic dimensions, the same mechanism appears as a quantized shift of photon occupation along the frequency ladder.

7.4.4 Beyond Frequency Space

Other photonic degrees of freedom can also serve as synthetic coordinates: orbital angular momentum, polarization, time bins, and spatial modes of a multimode waveguide. For the present chapter they should be viewed as extensions of the same idea, not as separate main lines. Frequency modes are the closest continuation of coupled-resonator physics because resonators naturally provide an ordered mode ladder and electro-optic modulation supplies tunable nearest-neighbor hopping.

The main caution is that locality in a synthetic coordinate is not identical to locality in real space. An “edge” in frequency space is a boundary of accessible resonant modes, usually produced by finite bandwidth, dispersion, and loss. Therefore a spectral accumulation near the end of a frequency window should be interpreted as topological edge transport only when it is accompanied by the predicted directional and robust response under modulation-phase variation.

The same design logic applies beyond InP or silicon microrings, including superconducting, acoustic, and mechanical resonator arrays. The platform details differ, but the checks are the same: the resonator Q must be high enough to define bands, the coupling must exceed loss while remaining within the single-mode regime, the phase control must realize the desired flux, and disorder or backscattering must not mix the protected sectors too strongly. This portability is the strength of coupled-resonator topology; the price is that the usable topological bandwidth is limited by coupling strength, disorder, and loss.

Chapter Summary

Coupled resonator arrays provide a controlled bridge from Maxwell’s equations to tight-binding topological models. The bridge is coupled-mode theory: localized high- Q modes become lattice sites, evanescent or waveguide-mediated overlap becomes hopping, and optical path phases become Peierls phases.

The central implementation is the microring Hofstadter array. A link-ring length gradient produces the flux

$$\Phi = \frac{\pi n_{\text{eff}} \delta L}{\lambda},$$

which realizes the Landau-gauge Harper model and its Chern-labeled gaps. The CW and CCW circulation sectors experience opposite fluxes, giving a pseudo-spin structure whose protection depends on suppressing backscattering.

Synthetic dimensions extend the same philosophy. Frequency modes of a modulated ring form an effective lattice coordinate, and a modulation-phase gradient produces the Harper equation in the combined real-space–frequency-space lattice. This makes resonator networks one of the most flexible platforms for synthetic gauge fields, while also making their limitations clear: finite Q , dispersion, loss, disorder, and soft synthetic boundaries must be checked quantitatively.

Chapter 8

Floquet Topological Insulators and Periodically Driven Systems

Why Topology Need Not Be Static?

This chapter introduces three conceptual leaps not touched upon in previous chapters:

- **From static to periodically driven: energy to quasi-energy.** In all preceding chapters, the topological classification of the system was built on the eigenvalues and eigenstates of the Bloch Hamiltonian $\hat{H}(\mathbf{k})$. When system parameters vary with period T , energy is no longer conserved; it is replaced by the quasi-energy ε , defined on the compact space $(-\pi/T, \pi/T]$. The direct consequence of this compactification is that the band structure simultaneously possesses bandgaps at both $\varepsilon = 0$ and $\varepsilon = \pi/T$, rather than just the single ground-state bandgap of the static case.
- **From Chern numbers to winding numbers: the incompleteness of static topology.** For Floquet systems, knowing only the Chern number of the effective Hamiltonian H_{eff} is *insufficient*. The key discovery of Rudner et al. in 2013 [4] was: there exist phases where $\mathcal{C}_{\text{eff}} = 0$ yet chiral edge states exist in every bandgap (the anomalous Floquet phase). A complete topological classification requires the **winding number** W defined on (k, t) space, which contains the microstructural information of $U(\mathbf{k}, t)$ throughout the entire driving period, not merely a snapshot at $t = T$.
- **From material response to geometric design: breaking the magneto-optical frequency limitation.** The core difficulty of Chapter 5 was the rapid decay of the magneto-optical effect at optical frequencies. The Floquet scheme provides a completely different $\mathcal{T}$ -breaking mechanism: *time-periodic driving* itself breaks $\mathcal{T}$, without requiring magneto-optical materials. The helical waveguide realization is particularly elegant: the propagation direction z plays the role of “time,” and the helical geometry plays the role of “periodic driving”; the entire $\mathcal{T}$ breaking is controlled purely by fabrication geometry and can be realized at any wavelength.

Remark (Common misconception warnings). Three most common misconceptions when first learning Floquet topology:

1. **Quasi-energy is not energy:** Quasi-energy ε is defined only modulo $2\pi/T$ and cannot be directly associated with the system's energy absorption/emission. This point is especially critical in photonic systems, because “quasi-frequency” is not the actual spectral frequency.
2. **The effective Hamiltonian H_{eff} loses microstructure:** The branch choice in $H_{\text{eff}} = (i/T) \log U(T)$ discards π -gap information, so the topological invariants of H_{eff} are *insufficient* for a complete classification of Floquet phases.
3. **The paraxial approximation is not a small-parameter expansion:** In photonic Floquet systems, “time” is actually the z coordinate; the paraxial approximation reduces the vector Maxwell equations to a scalar Schrödinger form, which requires the transverse structure to vary much more slowly than the wavelength—it fails for subwavelength structures.

The topological photonic systems discussed in previous chapters—whether magneto-optical Chern insulators (Chapter 5) or crystalline symmetry-protected pseudo-QSHE (Chapter 6)—were all **static** systems: material parameters and geometric structure do not vary with time, and topological properties are entirely determined by the time-independent Bloch Hamiltonian $\hat{H}(\mathbf{k})$.

This chapter introduces an entirely new dimension: **time-periodic driving** (Floquet driving). When system parameters vary with period T , energy is no longer a good quantum number; it is replaced by the **quasi-energy** ε , defined on the Brillouin zone $(-\pi/T, \pi/T]$. This additional periodicity—the compactification of quasi-energy—brings entirely new topological classifications, including **anomalous topological phases** that do not exist in static systems.

8.1 Foundations of Floquet Theory

8.1.1 The Floquet Theorem

Theorem 8.1 (Floquet theorem). *Let the evolution equation of a linear system be*

$$i \frac{\partial}{\partial t} |\psi(t)\rangle = \hat{H}(t) |\psi(t)\rangle, \quad (8.1)$$

where $\hat{H}(t+T) = \hat{H}(t)$ is a periodic Hamiltonian. Then there exists a set of solutions (Floquet states) of the form:

$$|\psi_n(t)\rangle = e^{-i\varepsilon_n t} |\phi_n(t)\rangle, \quad (8.2)$$

where ε_n is the **quasi-energy** and $|\phi_n(t+T)\rangle = |\phi_n(t)\rangle$ is the **Floquet mode** (a periodic function).

Proof. Construct the single-period evolution operator

$$U(T) = \mathcal{T} \exp \left[-i \int_0^T \hat{H}(t') dt' \right], \quad (8.3)$$

where $\mathcal{T}$ is the time-ordering operator. $U(T)$ is a unitary operator whose eigenvalues can be written as $e^{-i\varepsilon_n T}$; the corresponding eigenstates $|u_n\rangle$ satisfy $U(T) |u_n\rangle = e^{-i\varepsilon_n T} |u_n\rangle$. Defining $|\phi_n(t)\rangle = e^{i\varepsilon_n t} U(t, 0) |u_n\rangle$, one directly verifies $|\phi_n(t+T)\rangle = |\phi_n(t)\rangle$. $\square$

Remark. The Floquet theorem is the temporal counterpart of Bloch's theorem: Bloch's theorem exploits spatial periodicity, while the Floquet theorem exploits temporal periodicity. The mathematical structures are completely parallel—the spatial Bloch wave vector $\mathbf{k}$ corresponds to the quasi-energy ε , and the spatial Brillouin zone $(-\pi/a, \pi/a]$ corresponds to the quasi-energy Brillouin zone $(-\pi/T, \pi/T]$.

The decomposition in Theorem 8.1 has two pieces with different physical meanings. The quasi-energy controls the phase accumulated after an integer number of periods, while the periodic Floquet mode records the intra-period micromotion. In a static system these two roles collapse into one Hamiltonian eigenproblem. In a driven system they remain distinct, and this distinction is precisely what allows anomalous Floquet topology.

8.1.2 Periodicity of Quasi-Energy and New Bandgaps

Definition 8.2 (Quasi-energy Brillouin zone). The quasi-energy ε_n is uniquely determined only modulo $2\pi/T$: ε_n and $\varepsilon_n + 2\pi m/T$ ($m \in \mathbb{Z}$) describe the same physical state. Therefore the Brillouin zone for quasi-energy is

$$\varepsilon \in \left(-\frac{\pi}{T}, \frac{\pi}{T}\right]. \quad (8.4)$$

Proposition 8.3 (Two types of bandgaps). *In the quasi-energy spectrum, there are two topologically inequivalent bandgap positions:*

- $\varepsilon = 0$ **bandgap**: *Similar to the bandgap in static systems.*
- $\varepsilon = \pi/T$ **bandgap**: *A bandgap at the boundary of the quasi-energy Brillouin zone—unique to Floquet systems and absent in static systems.*

Edge states in the two types of bandgaps can exist independently, labeled by different topological invariants.

8.2 The Effective Floquet Hamiltonian and Winding Number

8.2.1 Floquet–Bloch Correspondence and Physical Intuition

Floquet theory is the time-periodic analogue of Bloch theory. In a spatially periodic system, translations are diagonalized by Bloch wave vectors $\mathbf{k}$; in a time-periodic system, one-period evolution is diagonalized by quasi-energies ε through

$$U(T) |u_n\rangle = e^{-i\varepsilon_n T} |u_n\rangle. \quad (8.5)$$

The crucial difference is that quasi-energy is compact: ε and $\varepsilon + 2\pi/T$ describe the same Floquet eigenvalue. This compactness simultaneously creates the additional π -gap and makes the topology of the full time-evolution operator relevant. Thus Floquet systems require more than the Chern numbers of a static-looking effective Hamiltonian.

Concept	Bloch	Floquet
Periodicity	spatial lattice translation	time-periodic evolution
Label	wave vector $\mathbf{k}$	quasi-energy ε
Fundamental operator	translation or Bloch Hamiltonian	one-period unitary $U(T)$
Compact variable	Brillouin zone	quasi-energy Brillouin zone
New gap	ordinary band gap	0-gap and π -gap

This table is not merely an analogy. It explains why Floquet topology has an additional layer: the topology of $U(\mathbf{k}, t)$ during a period can be nontrivial even when the stroboscopic effective Hamiltonian looks topologically trivial.

8.2.2 The Effective Hamiltonian

Definition 8.4 (Effective Floquet Hamiltonian). Define the effective Hamiltonian as

$$\hat{H}_{\text{eff}} = \frac{i}{T} \ln U(T), \quad (8.6)$$

whose eigenvalues are the quasi-energies ε_n .

Remark (Branch choice of the matrix logarithm). The definition of $\ln U(T)$ depends on the branch choice: different branches correspond to different “starting points” for the quasi-energy. This ambiguity reflects the core feature of Floquet topological classification—the Chern number of $\hat{H}_{\text{eff}}$ is **insufficient** for a complete characterization of topological phases.

Concretely, choosing a branch cut in the logarithm is equivalent to choosing which quasi-energy gap is treated as the reference gap. A cut through the 0-gap and a cut through the π -gap can produce different effective Hamiltonians, even though both come from the same $U(T)$. Thus H_{eff} should always be read together with its chosen quasi-energy branch.

8.2.3 Chern Numbers Are Not Enough: The Discovery of the Anomalous Phase

Theorem 8.5 (Topological classification of Floquet systems). *The topological classification of a two-dimensional Floquet system ($\hat{H}(\mathbf{k}, t)$ with both spatial and temporal periodicity) is determined by **two** independent integer invariants:*

1. The Chern number C_n of the effective Hamiltonian $\hat{H}_{\text{eff}}(\mathbf{k})$ (labeling the number of edge states in the $\varepsilon = 0$ bandgap).
2. The **Floquet winding number** $W[\varepsilon]$ (labeling the number of edge states in the $\varepsilon = \pi/T$ bandgap).

8.2.4 Definition of the Winding Number

Definition 8.6 (Floquet winding number). For a two-band system, define the winding number of the time-evolution operator $U(\mathbf{k}, t)$ over one driving period as

$$W = \frac{1}{8\pi^2} \int_0^T dt \int_{\text{BZ}} d^2k \, \text{tr} [U^{-1} \partial_t U, U^{-1} \partial_{k_x} U, U^{-1} \partial_{k_y} U], \quad (8.7)$$

where the brackets denote the combination of three matrices $[A, B, C] = ABC - ACB + BCA - BAC + CAB - CBA$ (complete antisymmetrization). W is an integer-valued topological invariant defined on the three-dimensional space $(k_x, k_y, t) \in \text{BZ} \times S^1$.

Proposition 8.7 (Existence of the anomalous phase). *When all band Chern numbers are zero ($C_n = 0$) but $W \neq 0$, the system is in an **anomalous Floquet topological insulator** phase—topologically protected chiral edge states exist, but they cannot be understood from the classification of static systems.*

For a general multi-band Floquet system, one should specify the quasi-energy gap used as the branch cut of the logarithm. Denote the winding associated with a gap at quasi-energy ε_* by W_{ε_*} . The edge-state count is then a gap-resolved statement: the number of chiral modes crossing a given gap is determined by the winding of the evolution with that gap chosen as the reference cut. This is the precise sense in which the π -gap is an independent topological object rather than a high-energy copy of the 0-gap.

This gap-resolved viewpoint also clarifies the role of band Chern numbers. In static systems, edge-state counting is usually organized by the Chern number of bands below a single spectral gap. In Floquet systems there is no absolute “below” because quasi-energy is periodic. One may choose a branch cut at the 0-gap or at the π -gap, and these two choices can lead to different effective Hamiltonians. The difference between the corresponding edge counts is precisely the dynamical information carried by the winding number.

8.3 Minimal Driven Models and the Anomalous Phase

A minimal way to see the two-gap structure is a split-step evolution. Let one period consist of two noncommuting steps,

$$U(k, T) = e^{-iH_2 T/2} e^{-iH_1 T/2}, \quad (8.8)$$

with

$$H_1 = J_1(\cos k \sigma_x + \sin k \sigma_y), \quad (8.9)$$

$$H_2 = J_2 \sigma_z. \quad (8.10)$$

Using the Pauli-matrix exponential formula, the quasi-energy dispersion is

$$\cos(\varepsilon T) = \cos(J_1 T/2) \cos(J_2 T/2) - \sin(J_1 T/2) \sin(J_2 T/2) \cos k. \quad (8.11)$$

The 0-gap closes when the quasi-energy reaches $\varepsilon = 0$, while the π -gap closes when $\varepsilon = \pi/T$. These two closures occur on different parameter lines, so edge states in the two gaps can be labeled independently.

For the two-step model above, the gap-closing conditions are particularly transparent. The 0-gap is controlled by the relative size of the two steps, while the π -gap is controlled by their sum over one period:

$$\text{0-gap closure:} \quad (J_1 - J_2)T/2 = 0 \mod \pi, \quad (8.12)$$

$$\pi\text{-gap closure:} \quad (J_1 + J_2)T/2 = \pi \mod \pi. \quad (8.13)$$

Thus a parameter change can close only one of the two gaps. This is the minimal algebraic reason that Floquet systems can support phases with edge modes in one gap but not the other.

The same logic underlies the Rudner anomalous Floquet phase. A driving protocol may return to a topologically trivial stroboscopic operator $U(T)$, so that the effective Hamiltonian has zero Chern number, while the intra-period evolution $U(\mathbf{k}, t)$ winds nontrivially in (k_x, k_y, t) space. This gives $W \neq 0$ even when $C_{\text{eff}} = 0$, and therefore produces chiral edge states that cannot be predicted from the final-time band Chern number alone.

In the canonical two-dimensional five-step drive, the system can even satisfy $U(\mathbf{k}, T) = \mathbf{I}$ for all $\mathbf{k}$ at a symmetric driving point. The effective Hamiltonian then appears trivial, but the intermediate-time map $U(\mathbf{k}, t)$ sweeps a nontrivial cycle in the unitary group. This is the cleanest mathematical picture of an anomalous Floquet insulator: topology is stored in the motion, not in the final snapshot.

A compact version of the Rudner construction makes this explicit. Divide one period into several hopping steps around a plaquette. During each step only one directed coupling is active, so a wave packet is moved sequentially around the unit cell. At the end of the period the bulk may return to its starting point, giving a trivial-looking $U(\mathbf{k}, T)$, while an edge wave packet cannot complete the same closed loop and is transported chirally along the boundary. The bulk evolution is therefore topologically nontrivial even though the final bulk Floquet operator can be homotopic to the identity.

This picture also explains why the anomalous phase is not captured by a static Chern number. A Chern number only sees the eigenvectors of a Hamiltonian at one quasi-energy branch; the anomalous phase is instead a pumping process distributed over the whole period. The boundary mode is the trace left by this pump when the closed bulk orbit is cut open by an edge.

Remark (Micromotion and anomalous edge states). The object carrying anomalous Floquet topology is not only $U(T)$ but the entire micromotion $U(\mathbf{k}, t)$ during one period. The effective Hamiltonian describes stroboscopic dynamics after choosing a logarithm branch; the winding number describes the homotopy class of the full evolution. This is why $C_{\text{eff}} = 0$ can still coexist with protected chiral edge states in the 0 or π quasi-energy gaps.

8.4 Paraxial Optics as Floquet Evolution

The optical platform most directly connected to Floquet topology is the paraxial waveguide array. For a slowly varying refractive-index profile, the field envelope satisfies

$$i \frac{\partial \psi}{\partial z} = -\frac{1}{2k_0} \nabla_{\perp}^2 \psi - \frac{k_0 \Delta n(\mathbf{r}_{\perp}, z)}{n_0} \psi. \quad (8.14)$$

This has the form of a Schrödinger equation with the propagation coordinate z playing the role of time. Therefore a structure periodic in z realizes a Floquet drive, with one spatial period of the waveguide array corresponding to one driving period.

The analogy is powerful but not literal. In optics, $|\psi|^2$ is a directly imageable classical intensity, and the sample can be inspected at different propagation distances. At the same time, the paraxial mapping is an approximation: for strong bending, large index contrast, or polarization-sensitive effects, one must return to the full Maxwell problem. This is why the Floquet description is a controlled effective theory rather than a replacement for Maxwell optics.

The practical value of the mapping is threefold. First, the propagation coordinate z provides direct access to micromotion: by imaging different cross sections one can observe how an edge state forms during the drive. Second, the optical intensity is a classical field intensity rather than a quantum probability distribution, so no state tomography is needed to visualize transport. Third, periodic geometric modulation of waveguides can implement a drive without fast temporal modulation of material parameters.

The main limitation is equally important. The paraxial equation assumes slowly varying envelopes, weak index contrast, and small bending angles. When the helix radius, modulation frequency, or index contrast becomes too large, the effective Schrödinger picture may miss polarization mixing, radiation loss, and higher-order Maxwell corrections. In that regime the Floquet language may still be useful, but the starting eigenproblem must again be Maxwell's equations.

Remark (High-frequency averaging and its limits). When the driving period is short compared with the intrinsic coupling time scale, the periodically driven system can often be organized through a high-frequency or Floquet–Magnus expansion. Schematically,

$$U(T) = \mathcal{T} \exp \left[-i \int_0^T H(t) dt \right] = e^{-iH_F T}, \quad (8.15)$$

where H_F begins with the time-averaged Hamiltonian and then receives commutator corrections from different moments within the period. This approximation is useful for helical waveguides when the pitch is short enough that coupling phases accumulated in one period remain controlled.

However, the high-frequency Hamiltonian is not the full Floquet topology. Near resonant driving, or when micromotion within the period becomes large, the averaged Hamiltonian can miss the anomalous phase. The reliable diagnostic is then the full evolution $U(\mathbf{k}, t)$ and its winding, not only a truncated effective Hamiltonian.

8.5 Helical Waveguide Floquet Topological Insulators

8.5.1 Straight Waveguide Arrays and Graphene

Arranging waveguides on a honeycomb lattice (graphene structure), with adjacent waveguides coupled through evanescent fields. In the tight-binding approximation:

$$i \frac{da_i}{dz} = \beta_i a_i + \sum_{\langle j \rangle} C_{ij} a_j, \quad (8.16)$$

where a_i is the mode amplitude of the i -th waveguide, β_i is the propagation constant, and C_{ij} is the coupling coefficient.

For **straight** waveguides (Δn independent of z), the system is equivalent to the static graphene Hamiltonian—Dirac cones exist at the K and K' points, but there is no topological bandgap ($\mathcal{T}$ preserved, $C = 0$).

8.5.2 Helical Waveguides and the Effective Vector Potential

One of the most experimentally successful Floquet optical platforms is waveguide arrays with periodic helical modulation along the propagation direction. If the center trajectory of the i -th waveguide is written as

$$(x_i(z), y_i(z)) = (X_i + R \cos \Omega z, Y_i + R \sin \Omega z), \quad (8.17)$$

where (X_i, Y_i) is the equilibrium position (honeycomb lattice site), R is the helix radius, $\Omega = 2\pi/\Lambda$, and Λ is the helix period, then the propagation direction z naturally plays the role of the “time” variable in the Floquet problem. Echoing the earlier discussion of anomalous Floquet edge modes, the key advantage of this platform is that the periodic driving is not abstractly written into Hamiltonian parameters but is directly encoded in the sample geometry; therefore micromotion, quasi-gaps, and boundary transport can be spatially resolved along the propagation direction. It is precisely in this sense that helical waveguides became the representative platform for bringing Floquet topology from effective models to imageable experiments.

Theorem 8.8 (Effective vector potential of helical waveguides). *Under the paraxial approximation, the propagation coordinate z in a helical waveguide array plays the role of time in Floquet periodic evolution. After a unitary transformation to the co-rotating frame, the original paraxial equation (8.14) can be rewritten as a two-dimensional Schrödinger-type equation with an effective gauge potential:*

$$i\partial_z \psi = \frac{1}{2k_0 n_0} [-i\nabla_\perp - \mathbf{A}(z)]^2 \psi + V(\mathbf{r}_\perp) \psi, \quad (8.18)$$

where the periodicity of $\mathbf{A}(z)$ comes from the helical modulation of the waveguide center trajectories, so that one propagation period corresponds to one Floquet driving period. When the helix radius and pitch satisfy the slowly varying approximation, the system can be described in the high-frequency limit by an effective Hamiltonian in the averaged sense, exhibiting equivalent Peierls phases and chiral edge transport.

The theorem has two physical ingredients. The first is the coordinate transformation into the frame co-moving with the waveguides, which turns the transverse velocity of the helix into a gauge potential. The second is the tight-binding projection onto guided modes, where this gauge potential appears as a direction-dependent Peierls phase in the evanescent couplings. Together these steps convert a purely geometric modulation of waveguide positions into an effective time-periodic gauge field.

Experimental observations:

1. When light is injected at a zigzag boundary of the honeycomb lattice, it propagates **unidirectionally** along the boundary—without diffusing into the bulk.
2. When a defect (missing waveguide) is introduced in the propagation path, light circumvents the defect and continues along the boundary—**backscattering immunity**.
3. Non-helical waveguides (straight waveguides): light diffuses from the injection point into the bulk; no edge states exist.

This was the **first experimental realization of a Floquet topological insulator in any physical system**.

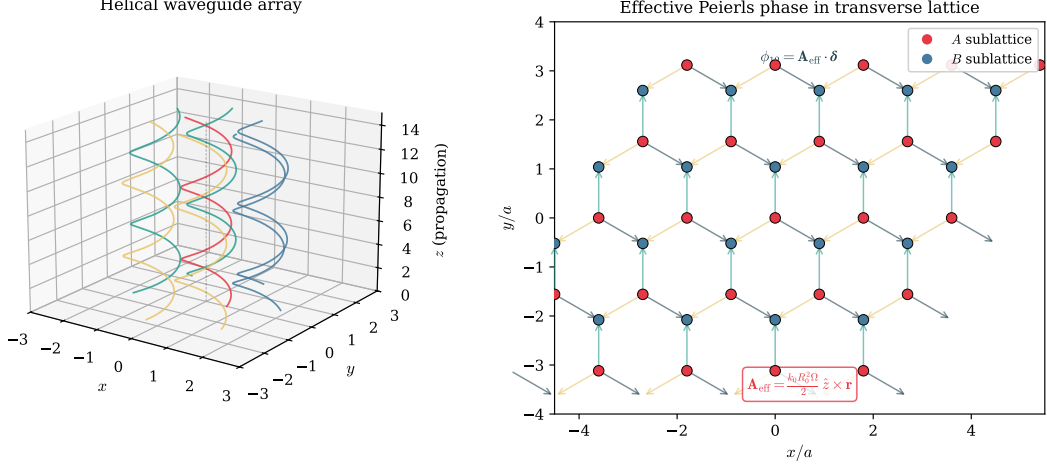


Figure 8.1 Effective gauge field in a helical waveguide array. (Left) Three-dimensional view: waveguides arranged on a honeycomb lattice undergo helical motion along the propagation direction z , with helix radius R_0 and period Z . (Right) Cross-sectional effective lattice: in the rotating frame, the coupling between waveguides acquires direction-dependent Peierls phases $\phi_{ij} = \frac{n_{\text{eff}}\omega}{c} \frac{2\pi R_0}{Z} \hat{z} \times \mathbf{d}_{ij}$, equivalent to a uniform effective flux threading each hexagonal cell.

The experiment is also pedagogically important because it directly visualizes Floquet micromotion. The field profile at a given output facet is not merely a band-structure measurement; it is a snapshot of the driven evolution after a finite propagation length. This makes helical waveguides unusually transparent compared with many electronic Floquet proposals, where micromotion is often harder to observe directly.

Three observables are especially useful for identifying the Floquet origin of the edge transport:

- **Boundary confinement:** injected light remains localized near the edge instead of spreading through the bulk.
- **Defect bypassing:** the edge wave packet circulates around missing or detuned waveguides without strong reflection.
- **Drive dependence:** removing the helical modulation, or tuning it outside the topological window, destroys the chiral edge channel.

8.5.3 Effective Peierls Phase and Topological Gap

The helical motion generates a time-periodic effective vector potential. For a waveguide trajectory

$$\mathbf{r}_j(z) = \mathbf{R}_j + R[\cos(\Omega z + \phi_j)\hat{x} + \sin(\Omega z + \phi_j)\hat{y}], \quad (8.19)$$

the co-rotating frame gives, schematically,

$$\mathbf{A}_{\text{eff}} = k_0 n_0 R \Omega (-\sin \Omega z, \cos \Omega z, 0), \quad (8.20)$$

so nearest-neighbor hopping acquires a Peierls phase,

$$C_j \rightarrow C_j \exp \left[-i \int_{\mathbf{R}_A}^{\mathbf{R}_B} \mathbf{A}_{\text{eff}} \cdot d\boldsymbol{\ell} \right]. \quad (8.21)$$

After averaging over one helical period, this phase plays the same role as a Haldane-type mass term in the honeycomb lattice. A useful estimate is

$$M_{\text{eff}} \sim 3\sqrt{3} C_2 \sin \eta, \quad \eta = k_0 n_0 R \Omega |\delta|, \quad (8.22)$$

with topological gap

$$\Delta\omega_{\text{gap}} \sim 2|M_{\text{eff}}|. \quad (8.23)$$

Thus the dimensionless helix parameter η controls both the strength of time-reversal-symmetry breaking and the size of the Floquet gap.

The Rechtsman et al. helical waveguide experiment [14] used this mechanism to realize chiral edge transport in a femtosecond-laser-written honeycomb array. The important point for the present chapter is not the detailed parameter table, but the design logic: choose η large enough to open a visible topological gap while remaining within the paraxial and finite-length propagation window.

The same parameter also affects the effective nearest-neighbor coupling. In a tight-binding description the helical drive can renormalize the bandwidth through Bessel-function factors such as $J_0(\eta)$. Therefore increasing η is not always beneficial: a larger Peierls phase opens a larger gap, but too large a drive can flatten the bands, reduce group velocity, and make finite-length propagation less effective. The useful design window is roughly the regime in which the topological gap is visible while the bandwidth and paraxial approximation remain reliable.

A common high-frequency estimate writes the renormalized nearest-neighbor coupling as

$$C_{\text{eff}} \simeq C_0 J_0(\eta). \quad (8.24)$$

This formula should not be overinterpreted as an exact universal result, but it captures the main design tension: the same drive that opens the topological gap can also reduce the bandwidth. Near the first zero of J_0 , the bands become very flat, which can enhance localization but also slows propagation and makes finite-size effects more visible.

For the Rechtsman platform, the relevant scale is set by

$$\eta = k_0 n_0 R \Omega |\delta|, \quad (8.25)$$

where R is the helix radius, $\Omega = 2\pi/\Lambda$ is set by the pitch, and $|\delta|$ is the nearest-neighbor separation. The experimental parameters put η at order unity: large enough to generate a measurable gap, but small enough that the paraxial approximation and finite sample length remain useful. This order-of-magnitude check is more important for the chapter's logic than the detailed numerical table.

For example, using visible light in glass, a helix radius of several microns, a centimeter-scale pitch, and a tens-of-microns lattice spacing gives $\eta \sim 1$. This places the experiment in the useful middle regime: the effective gauge phase is no longer perturbatively tiny, but the waveguides are still slowly varying on the optical wavelength scale. This is the practical reason the helical scheme can realize sizable optical-frequency topological gaps without relying on weak magneto-optical effects.

This tradeoff can be summarized as follows:

- **Small η :** the gap is too small and the system is close to the undriven honeycomb lattice.
- **Intermediate η :** the Peierls phase is large enough to open a measurable topological gap while the bandwidth remains usable.
- **Large η :** coupling renormalization and paraxial corrections become important; the driven-band picture may remain qualitative but loses simple quantitative control.

Compared with magneto-optical topological photonics, the helical scheme has a major advantage at optical frequencies: the effective gauge phase is set by geometry and can be made order unity. Its cost is fabrication complexity and the fact that the propagation distance is finite, so the observed edge transport is always a finite-time Floquet evolution rather than an infinite-time eigenstate measurement.

8.6 Floquet Topological Phase Diagram

The topology of a driven photonic lattice changes when either the 0-gap or the π -gap closes and reopens. In helical waveguide arrays this can be organized by the driving strength $\eta = k_0 R \Omega |\delta|$: at $\eta = 0$ the honeycomb array has graphene-like Dirac cones and no topological gap; for finite η the effective Peierls phase opens a Floquet gap; at stronger driving the π -gap can close and reopen, allowing an anomalous phase with $C_{\text{eff}} = 0$ but $W_\pi \neq 0$.

A schematic one-parameter sequence is:

- **No drive:** the honeycomb lattice has Dirac cones and no topological gap.
- **Moderate drive:** a gap opens at quasi-energy 0, and the system behaves like a Floquet Chern insulator.
- **Gap closing:** a topological transition occurs when either the 0-gap or π -gap closes.
- **Anomalous regime:** both gaps reopen, but the edge-state count is controlled by the winding of the full evolution rather than by C_{eff} alone.

In multi-parameter split-step models, these transitions appear as phase-boundary lines such as $\alpha = \beta$ for a 0-gap closure and $\alpha + \beta = \pi$ for a π -gap closure. The helical waveguide platform collapses this richer phase diagram onto a smaller set of geometric control parameters, chiefly the helix radius and pitch.

Regime	Gap behavior	Topological interpretation
Undriven	Dirac point, no gap	static honeycomb lattice, trivial for edge transport
Floquet Chern	0-gap opens	effective Haldane-like mass and chiral edge state
Transition	0 or π gap closes	winding can change only through gap closing
Anomalous	both gaps reopen	edge modes controlled by W , not only by C_{eff}

Theorem 8.9 (Bulk–edge correspondence of the anomalous phase). *In a two-dimensional Floquet system, chiral edge states are counted separately in the 0 and π quasi-energy gaps. A*

phase with $C_{\text{eff}} = 0$ but nonzero Floquet winding has protected edge states even though no static Chern number predicts them.

In practice, the gap-resolved edge count should be read as a statement about spectral flow. An edge branch can enter the quasi-energy zone through the π boundary and reappear at $-\pi/T$, so a plot restricted to a single branch of ε may look discontinuous. The discontinuity is not physical; it is a consequence of drawing a periodic quasi-energy axis on a line.

For this reason, the most reliable diagnostic is to plot the spectrum on the quasi-energy circle and mark separately which edge branches cross the 0-gap and which cross the π -gap. This avoids the common mistake of treating the π -gap as merely the top of an ordinary energy band.

The anomalous phase is therefore not a small modification of a Chern insulator. It is a genuinely dynamical phase: its edge states are tied to the winding of the full evolution over one period and to the compactness of quasi-energy.

This is the reason that experiments and simulations should inspect both quasi-energy gaps. If only the effective Hamiltonian spectrum near one branch cut is studied, the anomalous edge states can be missed.

It is useful to contrast the three common cases:

- **Static Chern phase:** a single band gap is characterized by the Chern number of occupied bands.
- **Ordinary Floquet Chern phase:** the drive opens a gap and the effective Hamiltonian has nonzero Chern number, so the phase resembles a driven version of a static Chern insulator.
- **Anomalous Floquet phase:** the effective Chern number can vanish, but the winding of $U(\mathbf{k}, t)$ is nonzero and edge modes still cross a quasi-energy gap.

The third case is the genuinely new one. It cannot be reproduced by simply finding a static Hamiltonian with the same stroboscopic spectrum, because the topology is carried by the path through unitary space during the period.

8.7 Other Floquet Schemes in Optics

Helical waveguides are the cleanest spatial-propagation realization, but they are not the only optical route to Floquet topology. Time-varying media can create sidebands at $\omega \pm \Omega$, which may be interpreted as hopping along a synthetic frequency dimension. Spatiotemporal modulation,

$$n(\mathbf{r}, t) = n_0 + \delta n \cos(\mathbf{q} \cdot \mathbf{r} - \Omega t), \quad (8.26)$$

selects a propagation direction and can break reciprocity without relying on magneto-optical materials. Periodically bent waveguides and alternating-coupling waveguide arrays provide geometric variants of the same idea: a periodically varying optical structure realizes a unitary evolution whose topology must be read from the full period, not only from an averaged Hamiltonian.

These alternatives are useful to mention, but they should not distract from the main line of this chapter. They are different implementations of the same Floquet principle: periodic

modulation creates a compact quasi-energy spectrum, and the topology of the drive is encoded in the one-period evolution.

The practical distinction is how the modulation is applied. Helical waveguides modulate geometry along the propagation direction; time-varying media modulate material parameters in real time; synthetic-frequency implementations modulate resonators to couple different frequency sidebands; and spatiotemporal modulation combines both space and time dependence to engineer direction-selective scattering. The mathematical Floquet structure is shared, but the experimental constraints and loss mechanisms differ substantially.

- **Temporal modulation:** couples optical frequencies separated by integer multiples of the modulation frequency and naturally realizes synthetic frequency lattices.
- **Spatiotemporal modulation:** carries a modulation wave vector $\mathbf{q}$ and can therefore distinguish forward and backward scattering channels.
- **Geometric modulation:** bends or twists waveguides along z , turning spatial propagation into a directly imageable Floquet evolution.

These platforms should be viewed as complementary rather than competing: each implements periodic evolution, but each has different limits on modulation strength, loss, bandwidth, and fabrication complexity.

Chapter Summary

This chapter established the Floquet route to topological photonics. Temporal periodicity leads to quasi-energy, defined modulo $2\pi/T$, so a driven system has both a conventional 0-gap and a Floquet-specific π -gap. The effective Hamiltonian $H_{\text{eff}} = (i/T) \log U(T)$ is useful, but its Chern numbers are not complete because the logarithm branch and the intra-period micromotion both matter.

The essential new invariant is the Floquet winding number of the full evolution $U(\mathbf{k}, t)$. It explains the anomalous Floquet phase, where $C_{\text{eff}} = 0$ but protected chiral edge states still appear in a quasi-energy gap. This is the central conceptual difference between static and periodically driven topology.

In photonics, paraxial waveguide arrays realize Floquet evolution because the propagation coordinate z plays the role of time. Helical waveguides implement a periodic effective vector potential, generate Peierls phases and a Haldane-type gap, and provide the canonical experimental realization of Floquet topological edge transport. Other schemes, including temporal modulation, spatiotemporal modulation, and periodically bent waveguides, fit the same basic principle: topology must be assigned to the full one-period evolution rather than only to an averaged static band structure.

Core conclusion: Floquet topology is topology of evolution, not merely topology of an effective static Hamiltonian. In photonic systems this idea is especially concrete because periodic propagation structures allow the evolution itself to be engineered and observed.

Chapter 9

Non-Hermitian Topological Photonics

Why Topology Must Survive Openness?

The framework of the previous seven chapters was built on the core assumption $\bar{M} = \bar{M}^\dagger$ (lossless, gain-free systems). This chapter will systematically dismantle that assumption, revealing three **fundamental transformations**—not merely perturbative corrections:

- **Transformation 1: Real spectrum $\rightarrow$ complex spectrum.** Once $\bar{M} \neq \bar{M}^\dagger$, the eigenfrequency ω_n acquires an imaginary part, describing temporal growth or decay of the mode. More profoundly, at special points in parameter space—exceptional points—two eigenvalues *merge* into one (algebraic degeneracy rather than geometric degeneracy), forming a Jordan block structure. This type of *defective degeneracy* is impossible in Hermitian systems, and is the most essential new phenomenon of non-Hermitian physics.
- **Transformation 2: Orthogonality $\rightarrow$ biorthogonality.** The eigenvectors of a Hermitian operator are orthogonal under the material-weighted inner product. The right eigenvectors $\{|\psi_n^R\rangle\}$ of a non-Hermitian operator are *not orthogonal* to one another; one must introduce the left eigenvectors $\{\langle\psi_n^L|\}$ to form a biorthogonal basis. All topological tools such as the Berry connection, Chern number, and bulk–edge correspondence must be *redefined* using the biorthogonal basis.
- **Transformation 3: Bloch $\rightarrow$ non-Bloch.** The non-Hermitian skin effect (NHSE) shows that under open boundary conditions, a macroscopic set of bulk states can become exponentially localized at the boundary—the Brillouin zone $\mathbf{k} \in \mathbb{R}$ is no longer the correct framework for describing bulk states, and must be generalized to the **generalized Brillouin zone** (GBZ) of $\beta = e^{ik} \in \mathbb{C}$.

More importantly, in optics, “non-Hermitian” does not mean first writing down a complex tight-binding Hamiltonian and then seeking its topological meaning, but rather starting from the **physical fact of open electromagnetic systems**: radiative boundary conditions allow energy to leak to infinity, complex permittivity and permeability encode material absorption and gain, and resonant modes coupled to the continuum naturally appear as quasinormal modes (QNMs). In other words, the starting point of non-Hermiticity in photonics is that the Maxwell operator is no longer self-adjoint, not the formal addition of a complex potential to a Hermitian tight-binding model. Therefore, although this chapter will frequently use simplified languages such as effective Hamiltonians, PT-symmetric dimers, and Hatano–Nelson chains, these models

should all be understood as projections of the open Maxwell problem onto finite-mode subspaces, not as a priori assumptions detached from the electromagnetic eigenproblem.

Remark (Common misconception warnings). Three intuitions that must be abandoned before studying this chapter:

1. **Non-Hermitian $\neq$ Hermitian plus a perturbation**: the Jordan-block behavior caused by eigenstate coalescence at an EP cannot be handled by ordinary Taylor perturbation theory; it requires a Puiseux expansion with fractional powers such as $\lambda^{1/2}$.
2. **EP $\neq$ Dirac point**: a Dirac point is a linear crossing of two *independent* eigenstates (integer topological charge ± 1), whereas an EP is where two eigenstates *merge* into one (half-integer topological charge $\pm 1/2$).
3. **PT symmetry $\neq$ Hermiticity**: a real spectrum in the PT-unbroken phase does not imply the operator is self-adjoint. Modes are still biorthogonal rather than orthogonal, and the definition of Berry geometry is fundamentally different.

9.1 From Open Maxwell Systems to Non-Hermitian Effective Operators

Chapter 2 established the self-adjointness of the Maxwell operator in lossless closed systems (theorem 2.10)—this was the foundation of real spectra, orthogonal completeness, and Berry geometry. This section systematically removes the lossless assumption $\bar{M} = \bar{M}^\dagger$, showing how the non-Hermitian material matrix leads to complex spectra, biorthogonal bases, and exceptional points.

When the material contains loss or gain, the permittivity and permeability tensors acquire anti-Hermitian parts:

$$\boldsymbol{\varepsilon}(\omega) = \boldsymbol{\varepsilon}'(\omega) + i\boldsymbol{\varepsilon}''(\omega), \quad \boldsymbol{\mu}(\omega) = \boldsymbol{\mu}'(\omega) + i\boldsymbol{\mu}''(\omega), \quad (9.1)$$

where $\boldsymbol{\varepsilon}' = \frac{1}{2}(\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}^\dagger)$ is the Hermitian part, $\boldsymbol{\varepsilon}'' = \frac{1}{2i}(\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^\dagger)$ is the anti-Hermitian part. The lossless condition $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\dagger$ is equivalent to $\boldsymbol{\varepsilon}'' = 0$.

Definition 9.1 (Non-Hermitian material matrix). Define the non-Hermitian material matrix

$$\bar{M}_{\text{NH}} = \begin{pmatrix} \boldsymbol{\varepsilon} & \boldsymbol{\xi} \\ \boldsymbol{\zeta} & \boldsymbol{\mu} \end{pmatrix}, \quad \bar{M}_{\text{NH}} \neq \bar{M}_{\text{NH}}^\dagger. \quad (9.2)$$

The system's generalized eigenvalue problem remains $\hat{N}\mathbf{f} = \omega \bar{M}_{\text{NH}}\mathbf{f}$, but $\bar{M}_{\text{NH}}$ is no longer Hermitian.

Proposition 9.2 (Complexification of eigenfrequencies). *When $\bar{M}_{\text{NH}} \neq \bar{M}_{\text{NH}}^\dagger$, the eigenfrequency ω_n is generally complex: $\omega_n = \omega'_n + i\omega''_n$. $\omega''_n > 0$ corresponds to a gain mode (exponentially growing amplitude), while $\omega''_n < 0$ corresponds to a dissipative mode (exponentially decaying amplitude).*

Proof. Consider $\hat{N}\mathbf{f}_n = \omega_n \bar{M}_{\text{NH}}\mathbf{f}_n$. In the Hermitian case, the reality of ω_n is guaranteed by $\langle \mathbf{f}_m, \hat{N}\mathbf{f}_n \rangle_M = \langle \hat{N}\mathbf{f}_m, \mathbf{f}_n \rangle_M$. When $\bar{M}_{\text{NH}} \neq \bar{M}_{\text{NH}}^\dagger$, the inner product $\langle \mathbf{f}, \mathbf{g} \rangle_M = \int \mathbf{f}^\dagger \bar{M}_{\text{NH}} \mathbf{g} d^3\mathbf{r}$ no

longer satisfies $\langle \mathbf{f}, \mathbf{g} \rangle_M = \overline{\langle \mathbf{g}, \mathbf{f} \rangle_M}$. In closed source-free systems we could rely on theorem 2.10 for a real spectrum and orthogonal expansion; after open boundaries, radiation, gain, and loss are introduced, this framework must be replaced by the more general biorthogonal spectral theory. Since $\int \mathbf{f}_m^\dagger (\bar{M}_{\text{NH}} - \bar{M}_{\text{NH}}^\dagger) \mathbf{f}_n d^3\mathbf{r} \neq 0$, ω_n is no longer constrained to be real. In the time evolution $e^{-i\omega_n t} = e^{-i\omega'_n t} e^{\omega''_n t}$, the sign of ω''_n determines growth or decay of the mode. $\square$

9.1.1 Biorthogonal Basis and Left/Right Eigenvectors

Definition 9.3 (Left and right generalized eigenproblems). Define the **right generalized eigenproblem**

$$\hat{N} |\mathbf{f}_n^R\rangle = \omega_n \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle, \quad (9.3)$$

and the **left generalized eigenproblem**

$$\langle \mathbf{f}_n^L | \hat{N} = \omega_n \langle \mathbf{f}_n^L | \bar{M}_{\text{NH}}, \quad (9.4)$$

or equivalently, $\hat{N}^\dagger |\mathbf{f}_n^L\rangle = \omega_n^* \bar{M}_{\text{NH}}^\dagger |\mathbf{f}_n^L\rangle$.

Theorem 9.4 (Biorthogonality). *In the non-degenerate case, the left and right eigenvectors satisfy the biorthogonality relation:*

$$\langle \mathbf{f}_m^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle = \delta_{mn} \mathcal{N}_n, \quad (9.5)$$

where $\mathcal{N}_n = \langle \mathbf{f}_n^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle$ is the normalization constant.

Proof. From the right eigenproblem $\hat{N} |\mathbf{f}_n^R\rangle = \omega_n \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle$ and the left eigenproblem $\langle \mathbf{f}_m^L | \hat{N} = \omega_m \langle \mathbf{f}_m^L | \bar{M}_{\text{NH}}$. The biorthogonality relation follows from these self-consistent left and right generalized eigenproblems and does not require $\hat{N}$ itself to be Hermitian or anti-Hermitian. Specifically, multiplying the left equation on the right by $|\mathbf{f}_n^R\rangle$:

$$\langle \mathbf{f}_m^L | \hat{N} |\mathbf{f}_n^R\rangle = \omega_m \langle \mathbf{f}_m^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle.$$

At the same time, from the right eigenproblem:

$$\langle \mathbf{f}_m^L | \hat{N} |\mathbf{f}_n^R\rangle = \omega_n \langle \mathbf{f}_m^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle.$$

Subtracting gives $(\omega_m - \omega_n) \langle \mathbf{f}_m^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle = 0$. When $\omega_m \neq \omega_n$, $\langle \mathbf{f}_m^L | \bar{M}_{\text{NH}} |\mathbf{f}_n^R\rangle = 0$. $\square$

Remark. The normalization constant $\mathcal{N}_n$ can be zero! When $\mathcal{N}_n = 0$, the biorthogonal basis collapses—this is precisely the signature of an exceptional point.

9.1.2 Non-Hermitian Effective Inner Product

In Hermitian topology, the Berry connection is built from a positive-definite material-weighted inner product. In the non-Hermitian case, the object needed for spectral decomposition and Berry geometry is instead the **biorthogonal pairing**

$$\langle \mathbf{f}_n^L | \bar{M}_{\text{NH}} |\mathbf{f}_m^R\rangle. \quad (9.6)$$

This pairing is algebraic rather than an ordinary stored-energy norm: it restores projection formulas and mode residues for non-self-adjoint operators, but it is generally not positive definite. Consequently, the non-Hermitian Berry connection may be complex-valued, and its topological meaning must be tied to a line gap, point gap, or isolated band subspace rather than to the usual Hermitian energy norm.

Remark (Biorthogonal pairing is not the same as the energy inner product). In passive Hermitian systems, orthogonality, energy conservation, and spectral expansion are often carried by the same material-weighted inner product. In open or gain–loss systems, these roles separate: the energy-type quadratic form still describes stored energy, dissipation, and power balance, while the biorthogonal pairing is the correct tool for spectral expansion and Berry geometry.

9.2 Exceptional Points and Their Topological Charge

Definition 9.5 (Exceptional point). A point $\mathbf{k}_{\text{EP}}$ in parameter space is called an **exceptional point** (EP) if, at that point, two (or more) eigenvalues of the non-Hermitian operator coalesce, and the corresponding eigenvectors also coalesce (i.e., a Jordan block appears):

$$\omega_1(\mathbf{k}_{\text{EP}}) = \omega_2(\mathbf{k}_{\text{EP}}), \quad |\mathbf{f}_1^R(\mathbf{k}_{\text{EP}})\rangle \propto |\mathbf{f}_2^R(\mathbf{k}_{\text{EP}})\rangle. \quad (9.7)$$

Remark. An EP is fundamentally different from a degeneracy point (Dirac point) in a Hermitian system: at a Dirac point, eigenvalues coalesce but eigenvectors remain linearly independent (geometric multiplicity equals algebraic multiplicity); at an EP, the geometric multiplicity is *strictly smaller* than the algebraic multiplicity, and the operator cannot be diagonalized.

9.2.1 Canonical Model of a Second-Order EP

Consider the two-level non-Hermitian effective Hamiltonian:

$$H_{\text{eff}}(\mathbf{k}) = \begin{pmatrix} \omega_0 + \delta(\mathbf{k}) + i\gamma_1 & \kappa \\ \kappa & \omega_0 - \delta(\mathbf{k}) + i\gamma_2 \end{pmatrix}, \quad (9.8)$$

where $\delta(\mathbf{k})$ is the detuning, $\gamma_{1,2}$ are gain/loss rates, and κ is the coupling strength.

Proposition 9.6 (EP condition). *For the real-parameter model (9.8), the second-order EP condition is*

$$\kappa^2 + \left[\delta + \frac{i}{2}(\gamma_1 - \gamma_2) \right]^2 = 0. \quad (9.9)$$

If κ is real and $\delta, \gamma_1, \gamma_2$ are real controls, this requires

$$\delta = 0, \quad \frac{\gamma_1 - \gamma_2}{2} = \pm \kappa. \quad (9.10)$$

*For the **PT**-symmetric case $\gamma_1 = -\gamma_2 \equiv \gamma$, the above simplifies to*

$$\delta = 0, \quad |\gamma| = \kappa, \quad (9.11)$$

and the coalescence point on the usual $\delta = 0$ slice occurs at $|\gamma| = \kappa$.

Proof. $\omega_{\pm} = \omega_0 + \frac{i}{2}(\gamma_1 + \gamma_2) \pm \sqrt{\kappa^2 + [\delta + \frac{i}{2}(\gamma_1 - \gamma_2)]^2}$. The EP requires the discriminant to vanish: $\kappa^2 + [\delta + \frac{i}{2}(\gamma_1 - \gamma_2)]^2 = 0$. For real parameters this separates into real and imaginary parts, giving $\delta = 0$ and $(\gamma_1 - \gamma_2)/2 = \pm\kappa$. For the **PT-symmetric** case $\gamma_1 = -\gamma_2 \equiv \gamma$, the real and imaginary parts of the discriminant give $\delta = 0$ and $|\gamma| = \kappa$, and the EP appears on the $\delta = 0$ slice at $|\gamma| = \kappa$. $\square$

9.2.2 Topological Charge of the EP

Definition 9.7 (Winding number of the EP discriminant). Define the discriminant $\Delta(\mathbf{k}) = (\omega_+ - \omega_-)^2$. In a two-dimensional parameter space, the EP is an isolated zero of $\Delta = 0$. Along a closed path $\mathcal{C}$ encircling the EP, the phase winding number of the discriminant is

$$\nu_{\text{EP}} = \frac{1}{2\pi} \oint_{\mathcal{C}} d \arg \Delta(\mathbf{k}) \in \mathbb{Z}. \quad (9.12)$$

A second-order EP has $|\nu_{\text{EP}}| = 1$.

Theorem 9.8 (Eigenvalue exchange of the EP). *After one loop around the EP, the two eigenvalues exchange: $\omega_+ \rightarrow \omega_-$, $\omega_- \rightarrow \omega_+$. This is the direct consequence of the EP winding number $\nu_{\text{EP}} = \pm 1$.*

Proof. $\omega_{\pm}$ are the two branches of $\sqrt{\Delta(\mathbf{k})}$. Δ has a first-order zero at the EP ($\nu_{\text{EP}} = \pm 1$), so $\sqrt{\Delta}$ has a branch point there. Upon encircling $\mathcal{C}$ once, $\arg \Delta$ changes by 2π , therefore $\arg \sqrt{\Delta}$ changes by π , i.e., $\sqrt{\Delta} \rightarrow -\sqrt{\Delta}$, and the two eigenvalues exchange. Returning to the original arrangement requires two loops—this is the Riemann-sheet structure of the EP. $\square$

A second-order EP therefore has two equivalent topological signatures: the discriminant has an integer winding, while the eigenvalue branch $\sqrt{\Delta}$ has a half-integer vorticity. One loop exchanges the two eigenvalues; two loops are required to return to the original sheet. This square-root Riemann-sheet structure is the essential distinction between an EP and an ordinary Hermitian degeneracy.

9.3 PT Symmetry

9.3.1 Definition of Photonic PT Symmetry

Definition 9.9 (Photonic PT symmetry). A photonic system has **parity–time-reversal** (PT) symmetry, if there exist a parity operator $\mathcal{P}$ and a time-reversal operator $\mathcal{T}$ such that

$$[\mathcal{PT}, \bar{M}_{\text{NH}}^{-1} \hat{N}] = 0. \quad (9.13)$$

For a one-dimensional system, $\mathcal{P} : x \rightarrow -x$, and $\mathcal{T}$ is complex conjugation. Physical implementation: spatially symmetric distributions of gain ($\varepsilon'' > 0$) and loss ($\varepsilon'' < 0$), satisfying $\varepsilon(x) = \varepsilon^*(-x)$.

Theorem 9.10 (PT-symmetric phases). *A finite-dimensional PT-symmetric system with a real characteristic polynomial typically has two phases:*

1. **PT-unbroken phase** ($\gamma < \gamma_{\text{EP}}$ in the two-mode model): eigenvalues in the considered sector are real, and eigenstates are eigenstates of the PT operator (spontaneous PT symmetry preserved).
2. **PT-broken phase** ($\gamma > \gamma_{\text{EP}}$): the corresponding eigenvalues appear as complex conjugate pairs, and PT symmetry is spontaneously broken.

The transition point is an exceptional point in the generic two-level scenario.

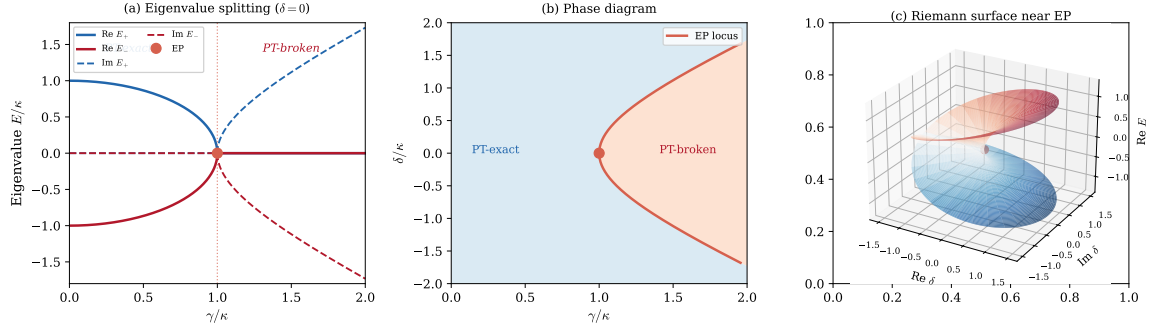


Figure 9.1 Exceptional points and phase diagram of a PT-symmetric coupled model. (Left) Real parts and imaginary parts of the eigenvalues in parameter space $(\gamma/\kappa, \delta/\kappa)$, with the PT-unbroken and PT-broken phases lying on the $\delta = 0$ slice: for $\gamma < \kappa$ the two eigenvalues are real, whereas for $\gamma > \kappa$ they split into a complex-conjugate pair. (Right) Real-part and imaginary-part Riemann surfaces for the $\delta = 0$ cross-section, where the exceptional point $\gamma_{\text{EP}} = \kappa$ is the self-intersection point where the two energy sheets merge.

Proof. Taking the PT-symmetric special case of Eq. (9.8) ($\gamma_1 = -\gamma_2 = \gamma$, $\delta = 0$) as an example, the eigenvalues are $\omega_{\pm} = \omega_0 \pm \sqrt{\kappa^2 - \gamma^2}$. When $\gamma < \kappa$, $\sqrt{\kappa^2 - \gamma^2}$ is real, so both $\omega_{\pm}$ are real. When $\gamma > \kappa$, $\sqrt{\kappa^2 - \gamma^2} = i\sqrt{\gamma^2 - \kappa^2}$, and $\omega_{\pm} = \omega_0 \pm i\sqrt{\gamma^2 - \kappa^2}$ form a complex-conjugate pair. The transition point $\gamma = \kappa$ is exactly the EP. In larger PT-symmetric systems, the same logic applies pairwise to the sector undergoing PT breaking, although different mode pairs may reach EPs at different thresholds. $\square$

9.3.2 PT-Symmetric Photonic Implementations

Implementing PT symmetry in a photonic structure requires

$$\varepsilon(\mathbf{r}) = \varepsilon^*(-\mathbf{r}), \quad (9.14)$$

so the real part of the dielectric profile is even and the imaginary part is odd. One-dimensional gain-loss photonic crystals and coupled microrings are common realizations. In both cases, increasing the gain-loss contrast drives the same basic scenario: a real-spectrum PT-unbroken sector reaches an EP and then enters a PT-broken phase with complex-conjugate eigenfrequencies. These examples are useful experimentally, but their main conceptual role in this chapter is to illustrate how EP physics appears in open Maxwell systems.

9.4 Non-Hermitian Topological Classification

9.4.1 Redefinition of Gaps

Definition 9.11 (Line gap and point gap). In the complex-energy plane, the concept of a “gap” in a non-Hermitian system must be generalized:

1. **Line gap:** there exists a line ℓ in the complex plane such that all eigenvalues lie on the same side of ℓ . In particular, if ℓ is the real axis, this requires all eigenvalues to have imaginary parts of the same sign.
2. **Point gap:** there exists a reference point E_{ref} in the complex plane such that $H_{\text{eff}}(\mathbf{k}) - E_{\text{ref}}\mathbf{I}$ is invertible for all $\mathbf{k}$ (i.e., E_{ref} is not on the spectrum).

Remark. A line gap is a special case of a point gap (any point on ℓ can serve as E_{ref}). The bandgap of a Hermitian system is equivalent to a line gap (taking ℓ as the vertical line through a point in the bandgap on the real axis). But a non-Hermitian system may have a point gap without having a line gap—this is an entirely new topological possibility.

9.4.2 Non-Hermitian Symmetry Classification

Non-Hermitian symmetry classification is richer than the Hermitian tenfold way because one must specify both the gap type and the dagger/non-dagger versions of time-reversal, particle-hole, and chiral constraints [13]. For the present photonic discussion, the key point is simpler: reciprocity, PT symmetry, and crystalline symmetries constrain which line-gap or point-gap invariants are meaningful, but the classification must always be stated together with the relevant gap.

9.4.3 Non-Hermitian Chern Number

Definition 9.12 (Non-Hermitian Chern number). For a two-dimensional non-Hermitian system with an isolated band (or isolated band subspace) separated by a line gap, one can define the Chern number based on the biorthogonal basis:

$$C_n^{\text{NH}} = \frac{1}{2\pi} \int_{\text{BZ}} \mathcal{F}_n^{\text{NH}} d^2k, \quad (9.15)$$

where $\mathcal{F}_n^{\text{NH}} = \nabla_{\mathbf{k}} \times \mathcal{A}_n^{\text{NH}}$ uses the non-Hermitian Berry connection defined above.

Proposition 9.13 (Integrality of the non-Hermitian Chern number). *Although $\mathcal{F}_n^{\text{NH}}$ is generally complex-valued, C_n^{NH} remains an integer provided the relevant band or band subspace is isolated and no exceptional degeneracy is crossed.*

Proof sketch. Using the completeness of the biorthogonal basis and the line-gap condition, one can express $\mathcal{F}_n^{\text{NH}}$ as the Jacobian of a mapping, whose topological quantization is guaranteed by an argument analogous to the Hermitian case. The key step is to prove that although $\mathcal{A}_n^{\text{NH}}$ is complex-valued, its integral over the Brillouin zone still equals 2π times an integer—this is essentially because the homotopy class of the mapping from the base space (T^2) to the classifying space is still determined by π_2 . A rigorous proof can be found in [13]. $\square$

9.5 NHSE, GBZ, and Non-Bloch Bulk–Edge Correspondence

9.5.1 From Bloch Bulk States to Skin Modes

The **non-Hermitian skin effect** (NHSE) is the accumulation of a macroscopic set of nominal bulk eigenstates near an open boundary. It is not a small boundary correction to ordinary Bloch theory: in many non-Hermitian systems, periodic and open boundary conditions lead to different spectra even in the thermodynamic limit.

Theorem 9.14 (NHSE and the point-gap winding number). *For a one-dimensional non-Hermitian system $H(k)$, define*

$$W(E_{\text{ref}}) = \frac{1}{2\pi i} \oint \frac{d}{dk} \ln \det[H(k) - E_{\text{ref}}] dk. \quad (9.16)$$

When $W(E_{\text{ref}}) \neq 0$ for a reference point inside a point gap, a generic finite-range 1D system exhibits the NHSE under open boundary conditions.

9.5.2 Generalized Brillouin Zone

Definition 9.15 (Generalized Brillouin zone [25]). The generalized Brillouin zone (GBZ) is the closed trajectory of the complex Bloch factor $\beta = e^{ik}$ selected by the open-boundary eigenvalue equation. Unlike the ordinary Brillouin zone, it need not satisfy $|\beta| = 1$.

The GBZ is therefore not merely a small imaginary shift of k . It is a replacement of the Bloch parameterization by the complex trajectory compatible with open boundaries. Topological invariants relevant to OBC spectra must be computed on this trajectory rather than on the ordinary unit circle.

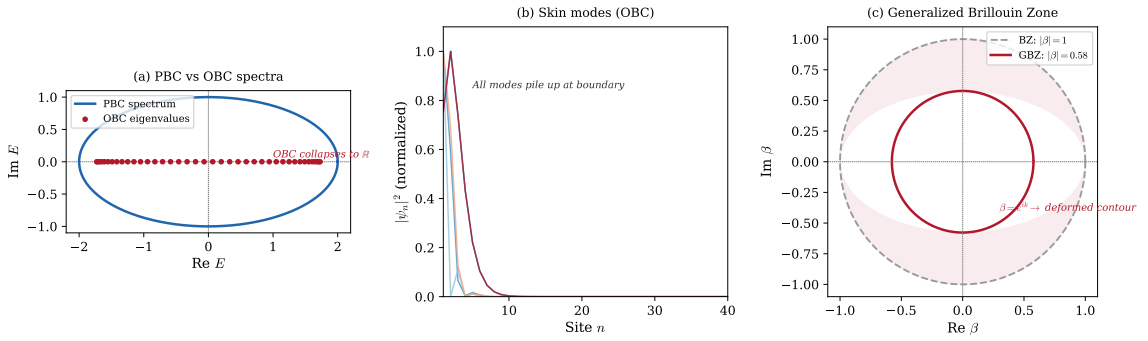


Figure 9.2 Non-Hermitian skin effect (NHSE): Hatano–Nelson model. (Left) Complex-energy spectra under periodic boundary conditions (PBC, blue loop) and open boundary conditions (OBC, red real segment). (Center) OBC eigenstates localized at the same boundary. (Right) The generalized Brillouin zone (GBZ, red) deviates from the unit circle.

9.5.3 The Hatano–Nelson Model

The minimal model is the nonreciprocal chain

$$H = \sum_n (t_R |n+1\rangle\langle n| + t_L |n\rangle\langle n+1|). \quad (9.17)$$

Under PBC, $\beta = e^{ik}$ and

$$E(k) = t_R e^{ik} + t_L e^{-ik}, \quad (9.18)$$

which traces an ellipse in the complex-energy plane when $t_R \neq t_L$. Under OBC, the relevant roots have equal modulus and the thermodynamic spectrum becomes

$$E(\phi) = 2\sqrt{t_R t_L} \cos \phi. \quad (9.19)$$

The corresponding GBZ is

$$\mathcal{C}_\beta = \{\beta \in \mathbb{C} : |\beta| = (t_L/t_R)^{1/2}\}, \quad (9.20)$$

and the skin depth is set by

$$\xi^{-1} = \frac{1}{2} \left| \ln \frac{t_R}{t_L} \right|. \quad (9.21)$$

Thus the bulk states themselves become non-Bloch exponential waves. In photonics, such nonreciprocal effective hopping can arise from magneto-optical effects, time modulation, direction-selective scattering, or synthetic dimensions.

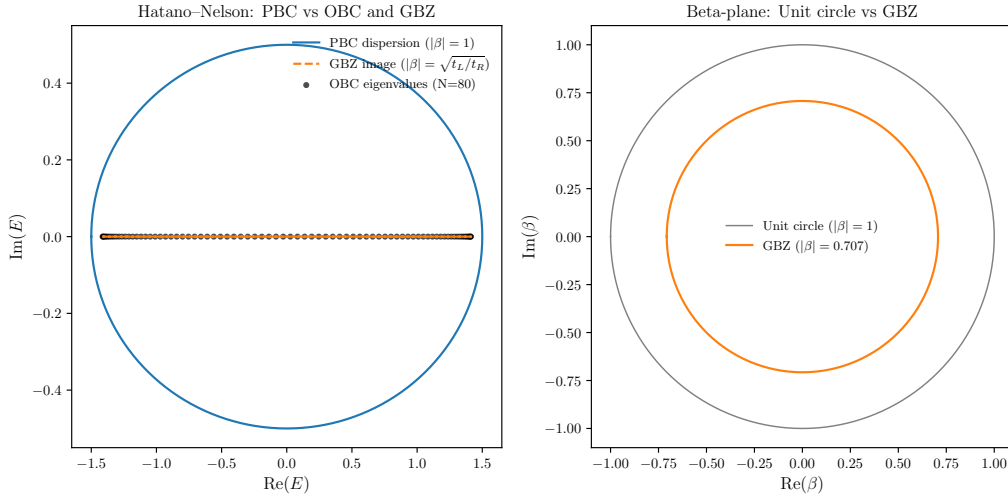


Figure 9.3 PBC spectrum, OBC spectrum, and GBZ of the Hatano–Nelson model. The figure illustrates why the OBC spectrum cannot be read from the ordinary Bloch loop.

9.5.4 Non-Bloch Bulk–Edge Correspondence

Proposition 9.16 (Breakdown of conventional bulk–edge correspondence). *In systems with NHSE, topological invariants computed from ordinary Bloch bands under PBC may fail to predict the OBC boundary spectrum, because the open-boundary bulk states are parameterized by the GBZ rather than by $|\beta| = 1$.*

Theorem 9.17 (Non-Hermitian bulk–edge correspondence (GBZ version)). *For one-dimensional non-Hermitian lattice models with NHSE, winding numbers defined on the GBZ can restore the correspondence with OBC boundary states:*

$$N_{\text{edge}} = |W^{\text{GBZ}}| \quad (9.22)$$

for the relevant 1D chiral or winding problem. Higher-dimensional analogues require model-specific non-Bloch invariants and are not captured by a single universal formula.

9.6 Non-Hermitian Topological Phenomena in Photonic Systems

Gain, loss, radiation leakage, and nonreciprocal coupling make non-Hermitian topology directly relevant to photonics. Three representative consequences are especially important.

First, gain–loss photonic crystals can realize PT symmetry and EP physics. A PT-symmetric SSH chain or coupled-resonator array illustrates how increasing gain–loss contrast can move a system from a line-gapped topological regime to an EP-mediated transition. The detailed invariant depends on whether one uses a line gap or a point gap.

Second, weak material loss does not automatically destroy Chern topology. If a gyromagnetic or otherwise topological photonic band remains separated by a line gap, the non-Hermitian Chern number is continuously connected to the Hermitian-limit Chern number:

$$C_n^{\text{NH}} = C_n \quad (9.23)$$

under gap-preserving loss or gain perturbations.

Third, topological lasers are non-Hermitian photonic devices because gain and loss determine the lasing threshold of edge and bulk modes. However, above threshold they involve gain saturation and nonlinear mode competition, so they should not be treated as a direct consequence of linear non-Hermitian band topology alone.

9.7 Quasinormal Modes and Non-Hermitian Topology

Quasinormal modes (QNMs) are the eigenmodes of open photonic systems satisfying outgoing radiation conditions:

$$\hat{N}\mathbf{f}_n^{\text{QNM}} = \tilde{\omega}_n \bar{M}\mathbf{f}_n^{\text{QNM}}, \quad \tilde{\omega}_n = \omega_n - i\frac{\Gamma_n}{2}. \quad (9.24)$$

Their complex frequency is caused by radiation leakage even when the material itself is lossless. This makes QNMs the natural Maxwell-level origin of many effective non-Hermitian Hamiltonians used in photonics.

BICs and Fano resonances can be viewed within the same open-system spectral language: a BIC is a limiting case where the radiative linewidth tends to zero, while a Fano resonance is the interference response between a resonant QNM and a continuum channel. These topics are important, but in this chapter they serve mainly to emphasize that non-Hermiticity in photonics begins with open electromagnetic boundary conditions, not merely with adding complex onsite potentials.

Chapter Summary

This chapter extended the Maxwell-topology framework from closed Hermitian systems to open, gain–loss, and radiative photonic systems. The first structural change is spectral: eigenfrequencies become complex, right eigenvectors are no longer mutually orthogonal, and Berry geometry must

be formulated using left–right biorthogonal states. Exceptional points are the local spectral singularities of this theory: eigenvalues and eigenvectors coalesce, producing a square-root Riemann surface and eigenvalue exchange under encircling.

The second structural change is topological. Non-Hermitian systems require a distinction between line gaps and point gaps, and the Chern number or winding number must be defined relative to the relevant gap. Point gaps also underlie the non-Hermitian skin effect: under open boundary conditions, bulk states can become exponentially localized, and the ordinary Brillouin zone must be replaced by the generalized Brillouin zone to recover bulk–edge correspondence.

For photonics, the physical source of non-Hermiticity is not only material gain and loss but also openness itself. Quasinormal modes, BIC limits, Fano resonances, PT-symmetric devices, non-Hermitian Chern bands, and topological lasers are all different manifestations of this open Maxwell viewpoint. The central lesson is that non-Hermitian topology changes the spectrum, the inner product, and the meaning of bulk–edge correspondence at the same time.

Bound States in the Continuum and Topological Radiation

Why Topology Must Confront the Radiation Continuum?

The previous eight chapters established a complete theoretical framework of topological photonics—from the self-adjointness of the Maxwell operator to Bloch bands, from Berry geometry to various topological invariants, and from magneto-optical Chern insulators to Floquet topology. But all these theories implicitly assumed a key condition: a *closed system*. Real photonic structures—especially photonic crystal slabs—inevitably couple to the radiation continuum, and are therefore *open systems*. This introduces three fundamental conceptual leaps:

- **Closed $\rightarrow$ open.** When a photonic structure couples to the radiation continuum, the mode acquires a finite linewidth $\Gamma > 0$, and the eigenfrequency becomes complex, $\tilde{\omega} = \omega_0 - i\Gamma/2$. The BIC is the strange exception to this picture—embedded in the continuum yet non-radiating ($\Gamma = 0$), and its protecting mechanism is often of *topological* origin. This tightly connects topological photonics with non-Hermitian physics and photonic crystal slab experiments.

Remark (Common misconception warnings).

1. **BIC $\neq$ dark mode:** a dark mode (e.g., a symmetry-forbidden transition) is decoupled from the radiation continuum, but is not necessarily embedded within the continuum. The uniqueness of a BIC is that its frequency lies *inside* the radiation continuum.
2. **The Q factor of a quasi-BIC is not infinite:** any perturbation that breaks the exact BIC condition (finite size, material absorption, fabrication error) reduces Q from ∞ to a finite value. The topological charge $|q|$ determines the *sensitivity* of Q to perturbations (the exponent in the scaling law), but does not guarantee arbitrarily large Q .

10.1 Basic Concepts of BICs

10.1.1 Statement of the Problem

Consider a finite-thickness photonic crystal slab. Its modes fall into two classes:

1. **Guided modes:** frequencies below the light cone ($\omega < c|\mathbf{k}_{\parallel}|$), with fields decaying exponentially outside the slab and zero linewidth;
2. **Radiative/leaky modes:** frequencies above the light cone, with fields propagating to the far field as radiation waves; these are quasinormal modes (QNMs) with finite linewidth $\Gamma > 0$.

A BIC is a special exception: its frequency lies above the light cone (embedded in the radiation continuum), yet its linewidth is $\Gamma = 0$ —it is a true bound state.

Definition 10.1 (Bound state in the continuum). Let $\{\omega_c(\mathbf{k}_{\parallel})\}$ denote the frequency range of the radiation continuum. A mode $\mathbf{f}_{\text{BIC}}$ is a BIC if its eigenfrequency ω_{BIC} satisfies:

1. $\omega_{\text{BIC}} \in \{\omega_c\}$ (embedded in the continuum);
2. $\mathbf{f}_{\text{BIC}}$ decays exponentially in the far field ($|\mathbf{f}_{\text{BIC}}| \rightarrow 0$ as $|z| \rightarrow \infty$);
3. the linewidth $\Gamma_{\text{BIC}} = 0$ (no radiative leakage).

10.1.2 History of BICs

The concept of the BIC was first proposed by von Neumann and Wigner in 1929 (in the quantum-mechanical context, where they constructed a potential supporting a bound state embedded in the continuum), but it was not until the 2000s that BICs were widely studied and experimentally verified in photonic systems. The advantages of photonic BICs are:

- the designability of photonic systems makes BICs easier to realize and control;
- the high- Q resonances (quasi-BICs) near BICs are important for sensing, lasing, and nonlinear optics.

10.2 Classification and Protection Mechanisms of BICs

10.2.1 Symmetry-Protected BICs

Theorem 10.2 (Symmetry-protected BIC). *Let a photonic crystal slab have spatial symmetry group G , and let the radiation channel belong to an irreducible representation Γ_{rad} of G . If a mode belongs to an irreducible representation Γ_{mode} different from Γ_{rad} , then the coupling matrix element between the mode and the radiation channel vanishes:*

$$\langle \Gamma_{\text{rad}} | V | \Gamma_{\text{mode}} \rangle = 0 \quad (\Gamma_{\text{rad}} \neq \Gamma_{\text{mode}}), \quad (10.1)$$

where V is the coupling operator between the slab and the radiation field. The mode is therefore a BIC.

Proof. This is a direct consequence of the Wigner–Eckart theorem. The coupling operator V transforms as the identity representation of G (because Maxwell’s equations preserve crystalline symmetry), so the matrix element $\langle \Gamma_{\text{rad}} | V | \Gamma_{\text{mode}} \rangle$ can be nonzero only when $\Gamma_{\text{rad}} = \Gamma_{\text{mode}}$. $\square$

Example 10.3 (Γ -point symmetry-protected BIC). In a square-lattice photonic crystal slab, the point group at the Γ point ($\mathbf{k}_{\parallel} = 0$) is C_{4v} . The normally outgoing plane-wave radiation

channel is described by the transverse electric-field components (E_x, E_y) and therefore transforms as the two-dimensional vector representation E of C_{4v} . Modes belonging to one-dimensional representations such as A_1, A_2, B_1 , or B_2 are symmetry-incompatible with this vertical radiation channel and can therefore form symmetry-protected BICs.

Remark. A symmetry-protected BIC generally acquires a finite linewidth when moving away from the high-symmetry point (because the symmetry group at a general $\mathbf{k}_{\parallel}$ point is lower, with lower-dimensional irreducible representations). However, the linewidth growth can be further suppressed by a topological mechanism—this leads to topological BICs.

10.2.2 Topologically Protected BICs: Vortices of the Radiation Polarization Field

Definition 10.4 (Radiation polarization vector). For a leaky mode $\mathbf{f}_n(\mathbf{k}_{\parallel})$ above the light cone in a photonic crystal slab, define the upward-radiating far-field polarization vector as

$$\mathbf{c}_n(\mathbf{k}_{\parallel}) = \begin{pmatrix} c_x(\mathbf{k}_{\parallel}) \\ c_y(\mathbf{k}_{\parallel}) \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{x}} \cdot \mathbf{E}_n^{\text{far}} \\ \hat{\mathbf{y}} \cdot \mathbf{E}_n^{\text{far}} \end{pmatrix}, \quad (10.2)$$

i.e., the projection of the far-field electric field onto the xy plane. $\mathbf{c}_n$ is defined in momentum space $\mathbf{k}_{\parallel}$.

Theorem 10.5 (BICs as polarization vortices [26]). A BIC corresponds to a **vortex** in the far-field polarization vector field $\mathbf{c}_n(\mathbf{k}_{\parallel})$: at the BIC momentum $\mathbf{k}_{\text{BIC}}$, $\mathbf{c}_n$ vanishes (radiation is zero), and the polarization direction has a nonzero winding number around that point. In the common case where $C_2\mathcal{T}$ symmetry allows the radiating Jones vector to be chosen real, define $\phi(\mathbf{k}_{\parallel}) = \arg[c_x(\mathbf{k}_{\parallel}) + ic_y(\mathbf{k}_{\parallel})]$. The BIC charge is

$$q = \frac{1}{2\pi} \oint_C d\phi(\mathbf{k}_{\parallel}) \in \mathbb{Z}. \quad (10.3)$$

This winding number q is called the **topological charge** of the BIC.

Proof sketch. $\mathbf{c}_n(\mathbf{k}_{\parallel})$ is a Jones vector defined on the two-dimensional $\mathbf{k}_{\parallel}$ space. Under the real gauge available in many time-reversal-invariant slab geometries, the complex scalar $c_x + ic_y$ records the far-field polarization direction. Its zeros (the BICs) are isolated, and the winding number of the angle ϕ around the zero is a topological quantity—it can only change when BICs annihilate in pairs (charge conservation: $q_1 + q_2 = 0$) or when they detach from the continuum. For elliptically polarized far fields, the equivalent formulation uses the winding of the polarization ellipse reconstructed from the Stokes parameters. $\square$

Proposition 10.6 (Robustness of topological BICs). A single topological BIC cannot be removed by a continuous parameter perturbation that preserves the symmetry—it can only move in $\mathbf{k}_{\parallel}$ space, or annihilate with another BIC of opposite charge. This is because the winding number q is an integer and cannot vary continuously.

Example 10.7 (Charge conservation and BIC pair annihilation). In a photonic crystal slab with C_{2v} symmetry, there is a BIC with $q = +1$ at the Γ point. By breaking C_{2v} down to C_2

(e.g., introducing anisotropy), this BIC can split into two $q = +1$ BICs and one $q = -1$ BIC (total charge conserved: $+1 = +1 + 1 - 1$), located at different $\mathbf{k}_{\parallel}$ positions. Further tuning of parameters can bring the $+1$ and -1 BICs together to annihilate, leaving one $q = +1$ BIC—the total topological charge remains conserved throughout the process [26].

10.2.3 Explicit Computation of the BIC Topological Charge

Example 10.8 (Explicit computation of the BIC topological charge: C_{2v} photonic crystal slab). Consider a two-dimensional photonic crystal slab with C_{2v} symmetry (square lattice of dielectric rods in air), and study the far-field radiation characteristics of a TM-like mode near the Γ point ($\mathbf{k}_{\parallel} = \mathbf{0}$).

Step 1: Symmetry constraints on the far-field polarization vector. The far-field radiation can be described by the Jones vector $\mathbf{c}(\mathbf{k}_{\parallel}) = (c_x, c_y)^T$, where c_x and c_y are the complex amplitudes of the x - and y -polarized components, respectively. The group C_{2v} contains two mirror reflections σ_x ($y \rightarrow -y$) and σ_y ($x \rightarrow -x$).

For a mode belonging to the irreducible representation B_2 , the mirror constraints are:

$$\sigma_x : c_x(k_x, -k_y) = c_x(k_x, k_y), \quad c_y(k_x, -k_y) = -c_y(k_x, k_y) \quad (10.4)$$

$$\sigma_y : c_x(-k_x, k_y) = -c_x(k_x, k_y), \quad c_y(-k_x, k_y) = c_y(k_x, k_y) \quad (10.5)$$

This means c_x is odd in k_x and even in k_y , while c_y is even in k_x and odd in k_y .

Step 2: Lowest-order expansion and the BIC. Performing a Taylor expansion near the Γ point, the lowest nonzero terms are:

$$\boxed{c_x \approx \alpha k_x, \quad c_y \approx \beta k_y} \quad (10.6)$$

where α, β are complex coefficients. At $\mathbf{k}_{\parallel} = \mathbf{0}$, $c_x = c_y = 0$ —the **radiation amplitude is strictly zero**, which is exactly the symmetry-protected BIC. The physical origin is that the Γ -point mode belongs to the B_2 representation, which is symmetry-incompatible with the far-field radiation channel (plane waves belonging to the E representation), so the coupling is strictly forbidden by symmetry.

Step 3: Explicit computation of the topological charge. Define the phase $\phi(\mathbf{k}_{\parallel}) = \arg(c_x + ic_y)$ (for linearly polarized far fields, this is equivalent to the polarization direction angle). Parametrize a circular path of radius δk centered at Γ : $k_x = \delta k \cos \theta$, $k_y = \delta k \sin \theta$, $\theta \in [0, 2\pi)$. Substituting into Eq. (10.6):

$$\phi(\theta) = \arg(\alpha \cos \theta + i\beta \sin \theta) \quad (10.7)$$

When α, β are real and have the same sign, $\phi(\theta)$ is strictly equal to θ , and as θ winds from 0 to 2π , ϕ also increases by 2π . The topological charge is therefore:

$$\boxed{q = \frac{1}{2\pi} \oint_{\Gamma} d\phi = +1} \quad (10.8)$$

This is the most common topological charge at the Γ point of C_{2v} (or higher C_{4v} , C_{6v}) slabs.

Step 4: Generalization to Stokes parameters. For elliptically polarized far fields (α, β complex), one should reconstruct the polarization ellipse from the Stokes parameters. The linear-polarization orientation is $\phi_{\text{pol}} = \frac{1}{2} \arg(S_1 + iS_2)$, with the understanding that the ellipse axis is defined modulo π . In the linearly polarized limit this reduces to the phase winding above; in the general case the charge is obtained from the winding of the ellipse orientation around the polarization singularity.

Remark. Higher symmetry (such as C_{6v}) can produce BICs with $|q| > 1$ at the Γ point: the lowest-order allowed terms are $c_x \sim k^{|q|} \cos(|q|\theta)$, $c_y \sim k^{|q|} \sin(|q|\theta)$. But such high-charge BICs are extremely fragile—small perturbations split them into $|q|$ BICs with $q = \pm 1$ (charge conservation) [26].

10.2.4 Friedrich–Wintgen BIC

Example 10.9 (Complete diagonalization of the Friedrich–Wintgen BIC and the Fano line shape). Consider two leaky modes (frequencies ω_1, ω_2 , linewidths γ_1, γ_2) coupled through a common radiation continuum channel.

Step 1: Effective Hamiltonian and the BIC condition. Within the CMT framework, the two-mode effective Hamiltonian is:

$$H_{\text{eff}} = \begin{pmatrix} \omega_1 - i\gamma_1/2 & V_{12} - i\sqrt{\gamma_1\gamma_2}/2 \\ V_{12} - i\sqrt{\gamma_1\gamma_2}/2 & \omega_2 - i\gamma_2/2 \end{pmatrix} \quad (10.9)$$

where V_{12} is the direct near-field coupling and $-i\sqrt{\gamma_1\gamma_2}/2$ is the indirect coupling mediated by the far-field continuum (from the Fano–Anderson model). The eigenvalues of H_{eff} are analytically:

$$\Omega_{\pm} = \frac{(\omega_1 + \omega_2) - i(\gamma_1 + \gamma_2)/2}{2} \pm \sqrt{\left(\frac{\Delta\omega - i\Delta\gamma/2}{2}\right)^2 + \left(V_{12} - i\frac{\sqrt{\gamma_1\gamma_2}}{2}\right)^2} \quad (10.10)$$

where $\Delta\omega = \omega_1 - \omega_2$, $\Delta\gamma = \gamma_1 - \gamma_2$. The BIC condition requires the imaginary part of one eigenvalue to vanish—i.e., linewidth $\Gamma_- = 0$:

$$\boxed{\Delta\omega\sqrt{\gamma_1\gamma_2} = V_{12}(\gamma_1 - \gamma_2)} \quad (\text{Friedrich–Wintgen condition}) \quad (10.11)$$

up to a sign fixed by the relative phase convention of the two radiation amplitudes. Equivalently, the non-radiating eigenvector must be orthogonal to the radiation vector $(\sqrt{\gamma_1}, \sqrt{\gamma_2})^T$ and must simultaneously diagonalize the Hermitian part of H_{eff} . Physically, exact destructive interference of the two radiation pathways turns one hybridized mode into a BIC.

Step 2: quasi-BIC scaling law. Suppose the FW-BIC is realized at parameter $\alpha = \alpha_0$ (e.g., hole size or angle); when deviating from the FW condition by $\delta\alpha = \alpha - \alpha_0$, the relation between linewidth and detuning is:

$$\boxed{\Gamma(\delta\alpha) \propto (\delta\alpha)^2, \quad Q(\delta\alpha) \propto (\delta\alpha)^{-2}} \quad (10.12)$$

This **quadratic scaling law** is the universal feature of FW-BICs, directly resulting from the expansion of the eigenvalues of H_{eff} near the BIC point.

Step 3: Fano resonance line shape. Embedding H_{eff} into the scattering-matrix framework, the transmission spectrum of the quasi-BIC exhibits an asymmetric Fano line shape:

$$T(\omega) = \left| \frac{(\omega - \omega_{\text{res}})/\gamma_a + q_F}{(\omega - \omega_{\text{res}})/\gamma_a + i} \right|^2 \cdot T_{\text{bg}} \quad (10.13)$$

where q_F is the Fano asymmetry parameter ($q_F \rightarrow \infty$ reduces to a Lorentzian line shape, $q_F = 0$ gives an antiresonance), and γ_a is the radiative linewidth of the quasi-BIC. In the BIC limit $\delta\alpha \rightarrow 0$, $\gamma_a \rightarrow 0$ and $q_F \rightarrow \pm\infty$ —the resonance disappears (becomes a dark mode).

Remark. There are practical upper limits on the Q factor of quasi-BICs: (1) finite sample size limits $Q \lesssim (\omega/c)(L/2)^2/\lambda$; (2) material absorption limits $Q \lesssim \text{Re}(\varepsilon)/\text{Im}(\varepsilon)$; (3) fabrication precision limits symmetry—for symmetry-protected BICs, $Q \propto (\delta\text{shape})^{-2}$. Therefore ultra-high Q ($> 10^6$) requires careful engineering optimization rather than relying only on topology.

10.3 Quasi-BICs and High- Q Resonances

The practical application value of BICs is mainly realized through quasi-BICs in their vicinity.

Definition 10.10 (Quasi-BIC). Applying a perturbation to a BIC (such as moving away from a high-symmetry point or introducing a slight asymmetry), the BIC acquires a finite but very small linewidth Γ and becomes a high- Q leaky resonant mode. With the convention $\tilde{\omega} = \omega_0 - i\Gamma/2$, the quality factor is $Q = \omega_0/\Gamma$ and can become extremely large.

Proposition 10.11 (Q -factor scaling law of quasi-BICs). *Near a symmetry-protected BIC, when deviating from the high-symmetry point by $\delta k = |\mathbf{k}_{\parallel} - \mathbf{k}_{\text{BIC}}|$, the generic single-radiation-channel linewidth scaling is determined by the lowest nonzero order of the far-field amplitude. For a simple polarization vortex with charge q this gives*

$$\Gamma(\delta k) \propto (\delta k)^{2|q|}. \quad (10.14)$$

Additional radiation channels, nongeneric higher-order zeros, or symmetry-forbidden lower-order terms can modify this exponent.

Proof sketch. The far-field radiation amplitude satisfies $|c_n| \propto (\delta k)^{|q|}$ (from the winding-number definition of the polarization vortex), the radiative power satisfies $P_{\text{rad}} \propto |c_n|^2 \propto (\delta k)^{2|q|}$, and the linewidth satisfies $\Gamma \propto P_{\text{rad}} \propto (\delta k)^{2|q|}$. $\square$

Example 10.12 (High- Q metasurfaces). Dielectric metasurfaces based on quasi-BICs have been widely used for:

- high-sensitivity refractive-index sensing ($Q > 10^4$ resonances enabling enhanced detection limits);
- ultra-low-threshold nonlinear optics (BIC-enhanced field localization increasing second-/third-harmonic efficiency by several orders of magnitude);
- lasers (BIC-provided surface-emitting high- Q modes).

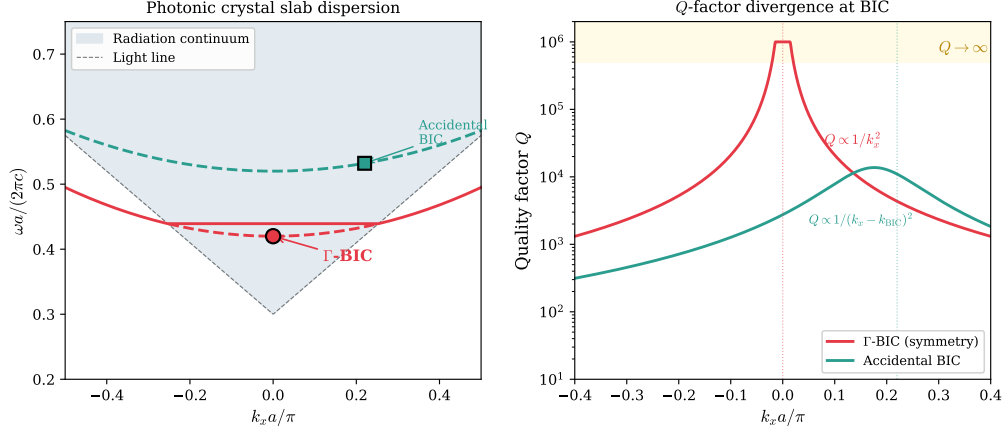


Figure 10.1 Dispersion and quality-factor divergence of a quasi-BIC. The left panel schematically shows a symmetry-protected BIC in a photonic crystal slab and the quasi-BIC dispersion in its vicinity; when the in-plane wave vector deviates from the BIC point, radiative leakage reopens and the linewidth grows with increasing δk . The right panel shows the typical scaling $Q \propto \delta k^{-2}$, illustrating that the ultra-high- Q resonances near a BIC are not accidental spikes, but are controlled by the systematic cancellation of coupling to the radiation channel at the symmetry-restored point.

10.3.1 Experimental Characterization of BICs

The theoretical prediction of a BIC must be experimentally verified, and the core challenge is: the BIC itself does not radiate, so it cannot be directly detected by far-field spectroscopy—experimentally one can only observe the quasi-BIC resonance *near* the BIC and infer the existence of the BIC from the way its linewidth tends to zero.

Angle-resolved spectroscopy For a photonic crystal slab, the incidence angle θ (or in-plane wave vector $\mathbf{k}_{\parallel}$) is the most natural tuning parameter. In angle-resolved reflection/transmission spectra, leaky modes appear as Fano resonances:

$$R(\omega) = R_0 \frac{(F\gamma + \omega - \omega_0)^2}{(\omega - \omega_0)^2 + \gamma^2}, \quad (10.15)$$

where $\gamma(\mathbf{k}_{\parallel})$ is the radiative linewidth and F is the Fano asymmetry parameter. As $\mathbf{k}_{\parallel} \rightarrow \mathbf{k}_{\text{BIC}}$, $\gamma \rightarrow 0$, and the resonance line shape disappears—this *dark point* is the experimental fingerprint of the BIC.

Measurement of the far-field polarization via Stokes parameters Extracting the topological charge q requires measuring the complete information of the far-field polarization. Measure the four Stokes parameters S_0, S_1, S_2, S_3 at each $\mathbf{k}_{\parallel}$, and map $(S_1, S_2, S_3)/S_0$ onto the Poincare sphere; then the BIC position corresponds to a *vortex singularity* on the sphere, and its winding number is the topological charge:

$$q = \frac{1}{2\pi} \oint_{\mathcal{C}} d\phi_{\text{pol}}, \quad (10.16)$$

where $\phi_{\text{pol}} = \frac{1}{2} \arg(S_1 + iS_2)$ is the linear-polarization orientation angle, defined modulo π , and $\mathcal{C}$ is a closed path in $\mathbf{k}_{\parallel}$ space enclosing the BIC.

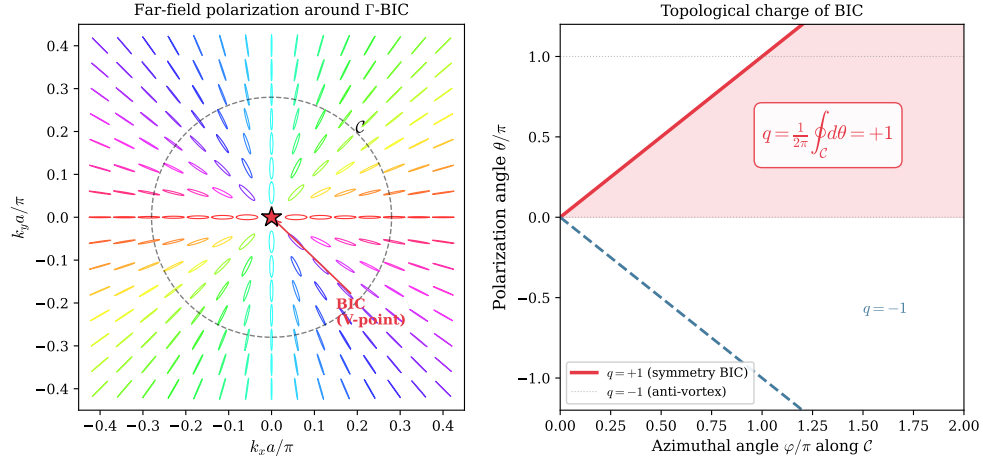


Figure 10.2 Far-field polarization vortex and topological charge of a BIC. The left panel shows the winding texture of the principal axis direction of the far-field polarization ellipse in momentum space; when encircling the BIC point once, the polarization orientation angle winds by an integer multiple of 2π , thereby defining the topological charge $q = \frac{1}{2\pi} \oint d\phi_{\text{pol}}$. The right panel schematically shows the polarization singularity structure reconstructed from Stokes parameters, illustrating that the BIC is not merely the limiting point of a high- Q resonance, but a topological defect in the far-field radiation texture.

Table 10.1 Systematic comparison of three BIC types.

Feature	Symmetry-protected BIC	Topologically protected BIC	Friedrich–Wintgen BIC
Protecting mechanism	Point-group symmetry forbids radiation.	Conservation of the far-field polarization-vortex charge.	Destructive interference between two leaky resonances.
Robustness	Stable as long as the protecting symmetry is preserved.	Charge $ q $ is conserved; the BIC position can move in momentum space.	Fine-tuned and generally not robust.
Q scaling law	$Q \propto k_{\parallel}^{-2}$	$Q \propto k_{\parallel}^{-2 q }$	Finite- Q resonance peak near the cancellation condition.
Typical $ q $	Fixed by symmetry.	≥ 1 and designable.	Not applicable; non-topological.
Representative platform	Photonic-crystal slab at Γ .	Photonic-crystal slab away from Γ .	Waveguide–resonator or two-resonance coupling system.

Q -factor engineering: three layers of optimization In practical applications, the Q factor of a quasi-BIC is constrained by three layers of factors:

1. **Topological-charge layer:** higher $|q|$ yields a steeper scaling law $Q(\delta k) \propto \delta k^{-2|q|}$ —a BIC with $|q| = 2$ has an order-of-magnitude higher Q than a $|q| = 1$ BIC at the same δk .
2. **Finite-size layer:** the period number N limits the minimum resolvable $\delta k \sim 2\pi/(Na) \rightarrow Q_{\max} \propto N^{2|q|} \rightarrow$ for $|q| = 1$, one needs $N > 100$ to reach $Q \sim 10^4$.
3. **Material-absorption layer:** $Q_{\text{abs}} = \omega \text{Re}(\epsilon)/(2 \text{Im}(\epsilon)) \rightarrow$ for all-dielectric platforms (Si_3N_4 , TiO_2), $Q_{\text{abs}} \sim 10^5$ – 10^6 , whereas for metallic plasmonics $Q_{\text{abs}} \sim 10$ – 10^2 .

10.4 Connection of BICs to Topological Photonics

10.4.1 Polarization Vortices and Berry Phase

The vortices of the far-field polarization field $\mathbf{c}_n(\mathbf{k}_{\parallel})$ in $\mathbf{k}_{\parallel}$ space (Definition 10.4 and Theorem 10.5) have a deep connection to the Berry geometry of Chapter 4.

Proposition 10.13 (BIC topological charge and Berry phase of the band). *For a time-reversal invariant system, the topological charge q of the BIC can be related to the Berry phase accumulated by the resonant band along a loop surrounding $\mathbf{k}_{\text{BIC}}$, together with the gauge and symmetry selection rules of the radiation channel. It is therefore a property of the far-field polarization texture on a loop in momentum space, not simply the value of the Berry curvature at the BIC point [27].*

Chapter Summary

1. A BIC is a true bound state embedded in the radiation continuum—its frequency lies above the light cone, yet its radiative linewidth is exactly zero. Its protection mechanisms include symmetry protection (irreducible-representation mismatch) and topological protection (winding-number conservation of the far-field polarization vortex).
2. The topological charge q of a BIC (the phase winding number of the far-field polarization vector) is an integer, and can change only through creation or annihilation of BIC pairs—topological conservation gives BICs robustness to parameter perturbations. The most common topological charge at the Γ point of a C_{2v} slab is $|q| = 1$.
3. The Friedrich–Wintgen BIC arises from destructive interference between two modes; the BIC condition is $\Delta\omega\sqrt{\gamma_1\gamma_2} = V_{12}(\gamma_1 - \gamma_2)$ up to the relative radiation-phase convention, requiring no symmetry protection but requiring parameter fine-tuning.
4. The quasi-BIC Q -factor scaling law $Q \propto (\delta k)^{-2|q|}$ directly links topological charge to device performance—for $|q| = 1$, $Q \propto (\delta k)^{-2}$, which is the key mechanism behind Fano-resonance-enhanced sensing and nonlinear optics.
5. A BIC can be viewed as a singular limiting point in the QNM spectrum ($\text{Im}\omega \rightarrow 0$), rather than an independent mode theory completely separate from QNMs. Comparisons between the two must always specify the roles of boundary conditions, normalization, and experimental observables.

Part III

Valley, Three-Dimensional, and Nonlinear Topology

Valley Topological Photonic Crystals

Why Topology Can Be Local in Momentum Space?

Chapters 5 and 6 respectively demonstrated two paradigms for realizing topologically protected edge states: breaking time-reversal symmetry (magneto-optical effect) and constructing pseudo-spin via C_{6v} symmetry. However, both schemes face their own experimental challenges: the magneto-optical scheme requires strong magnetic fields and special materials, while the pseudo-spin scheme demands precise C_{6v} symmetry.

Valley topology provides a third path—it requires only **breaking spatial inversion symmetry** (rather than time-reversal symmetry), and can realize topological domain-wall states in all-dielectric systems through simple geometric perturbations. This makes valley topological photonic crystals one of the experimentally most accessible and application-promising topological photonic systems to date.

However, valley topology also raises a profound conceptual issue: the valley Chern number is a half-integer rather than an integer, and strictly speaking is **not a topological invariant**—it depends on the choice of integration region, and is not protected when intervalley scattering is present. Understanding the nature and limitations of this “quasi-topological” protection is the core intellectual goal of this chapter.

Remark (Common misconception warnings).

1. **“The valley Chern number is a topological invariant”**—not rigorous. The valley Chern number $C_K = \pm 1/2$ depends on the choice of integration region (how to partition the areas near K and K'), and there is no strict bulk–edge correspondence analogous to that of the integer Chern number.
2. **“Valley topological edge states are immune to all defects”**—wrong. Valley topological protection is effective only against defects that do not induce intervalley scattering. Defects that break C_3 symmetry or scattering at armchair-type interfaces are not protected.
3. **“Valley topology is equivalent to a Chern insulator”**—fundamentally different. The edge states of a Chern insulator are guaranteed by an integer topological invariant, whereas the existence of valley topological domain-wall states depends on the continuum approximation and the suppression of intervalley scattering.

11.1 Dirac Cones in Honeycomb Photonic Crystals

11.1.1 Geometry and Symmetry of the Honeycomb Lattice

Consider a two-dimensional photonic crystal: two inequivalent dielectric rods (sublattices A and B) are placed at each site of a triangular lattice, forming a honeycomb arrangement. Let the lattice constant be a , with basis vectors

$$\mathbf{a}_1 = a \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{a}_2 = a \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}. \quad (11.1)$$

The corresponding reciprocal lattice vectors are

$$\mathbf{b}_1 = \frac{2\pi}{a} \begin{pmatrix} 1 \\ -1/\sqrt{3} \end{pmatrix}, \quad \mathbf{b}_2 = \frac{2\pi}{a} \begin{pmatrix} 0 \\ 2/\sqrt{3} \end{pmatrix}. \quad (11.2)$$

The first Brillouin zone of the honeycomb lattice is a regular hexagon, whose six vertices fall into two groups of inequivalent high-symmetry points:

$$\mathbf{K} = \frac{2\pi}{3a} \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad \mathbf{K}' = \frac{2\pi}{3a} \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}. \quad (11.3)$$

K and K' are related by time reversal $\mathcal{T}$: $\mathcal{T} : \mathbf{K} \leftrightarrow -\mathbf{K} \equiv \mathbf{K}'$ (modulo a reciprocal lattice vector).

When the two sublattices are identical ($\varepsilon_A = \varepsilon_B$, $r_A = r_B$), the system has C_{6v} point-group symmetry, including C_6 rotation, σ_v mirror reflection, and spatial inversion $\mathcal{P}$.

11.1.2 C_{3v} Little Group and Dirac Degeneracy

At the K point, the little group is C_{3v} (among the C_6 rotations, only C_3 maps K back to an equivalent K point). The character table of the C_{3v} group is:

Table 11.1 Character table of the C_{3v} group

C_{3v}	E	$2C_3$	$3\sigma_v$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0

The concept of valley topological photonic crystals was first proposed by Dong et al. [28], and subsequently Chen et al. systematically studied valley-contrast physics from the perspective of all-dielectric photonic crystals [29]. Experimentally, Gao et al. first observed topologically protected refraction in valley photonic crystals [30], and Lu et al. realized valley topological transport in acoustic systems [31]. The review article by Tang et al. provides a comprehensive summary of the physics, design, and applications of topological photonic crystals [32].

Figure 11.1 shows the band structure of the honeycomb photonic crystal. When the two sublattices are perfectly equivalent, the system has C_6 symmetry, and Dirac cones appear at

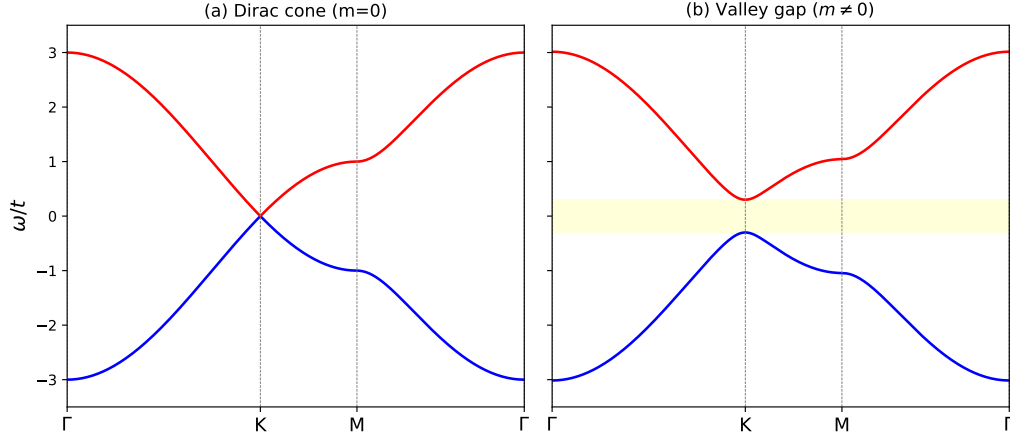


Figure 11.1 Band structure of the honeycomb tight-binding model. (a) When the two sublattices are equivalent ($m = 0$), Dirac cones appear at the K and K' points, with bands crossing linearly at the Dirac point. (b) When sublattice asymmetry is introduced ($m \neq 0$), the Dirac cone opens a bandgap (yellow-shaded region), forming a valley bandgap. The Berry curvatures at the two valleys have opposite signs, endowing the system with a nonzero valley Chern number $C_v = C_K - C_{K'} = \pm 1$. This figure was generated by numerical computation of the tight-binding model.

the Brillouin zone corners K and K' . Breaking spatial inversion symmetry (making the two sublattices inequivalent) opens the Dirac cone into a bandgap, forming a valley bandgap. The Berry curvatures at the two valleys have opposite signs—this is the physical origin of the valley topological effect.

11.1.3 Valley Hall Effect and Valley Hall Conductivity

By analogy with the Hall conductivity in the quantum Hall effect, one can define a valley Hall conductivity to quantify the valley topological effect:

$$\sigma_{xy}^{\text{valley}} = \frac{e^2}{h} C_v, \quad (11.4)$$

where $C_v = C_K - C_{K'}$ is the valley Chern number. In photonic systems, the analogue of the valley Hall conductivity is the deflection angle of photons in the transverse direction, which is proportional to the valley Chern number.

It must be emphasized that the valley Chern number C_v is not a strict topological invariant: it depends on the specific way the Brillouin zone is divided into the K valley and K' valley. However, as long as intervalley scattering is negligible (i.e., the spatial variation scale of perturbations is much larger than the lattice constant), C_v is a well-defined quantity and valley topological protection is effective.

Proposition 11.1 (Group-theoretic origin of the K -point Dirac degeneracy). *Let a honeycomb photonic crystal have C_{6v} symmetry. At the K point, the two bands belonging to the two-dimensional irreducible representation E of C_{3v} are necessarily degenerate. This symmetry-protected twofold degeneracy forms the **Dirac cone**.*

Proof. The little group C_{3v} at the K point contains three conjugacy classes: $\{E\}$, $\{C_3, C_3^{-1}\}$, $\{\sigma_v, \sigma'_v, \sigma''_v\}$. Its two-dimensional irreducible representation E has dimension 2, so Bloch states belonging to the E representation are necessarily twofold degenerate at the K point.

More specifically, let the two degenerate Bloch states at the K point be $|\phi_1\rangle$ and $|\phi_2\rangle$; they transform under C_3 rotation as

$$\hat{C}_3|\phi_1\rangle = e^{i2\pi/3}|\phi_1\rangle, \quad \hat{C}_3|\phi_2\rangle = e^{-i2\pi/3}|\phi_2\rangle. \quad (11.5)$$

Since the C_3 eigenvalues are different ($e^{\pm i2\pi/3}$), these two states cannot be mixed by any perturbation that preserves C_3 symmetry—the degeneracy is symmetry-protected. $\square$

11.1.4 Frequency-Domain Eigenequation for TM Modes

For TM modes (E_z polarization), the frequency-domain Maxwell eigenequation is

$$-\nabla \cdot \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla H_z \right) = \frac{\omega^2}{c^2} H_z, \quad (11.6)$$

where $\varepsilon(\mathbf{r})$ has the periodicity of the honeycomb lattice. By Bloch's theorem, the eigenmodes are parameterized as

$$H_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}), \quad (11.7)$$

where $u_{n\mathbf{k}}$ has lattice periodicity. Substituting into the eigenequation gives the $\mathbf{k}$ -parameterized cell equation:

$$-(\nabla + i\mathbf{k}) \cdot \left[\frac{1}{\varepsilon(\mathbf{r})} (\nabla + i\mathbf{k}) u_{n\mathbf{k}} \right] = \frac{\omega_{n\mathbf{k}}^2}{c^2} u_{n\mathbf{k}}. \quad (11.8)$$

Near the K point, the dispersion of the two bands belonging to the E representation is linear:

$$\omega_{\pm}(\mathbf{k}) = \omega_D \pm v_D |\delta\mathbf{k}| + O(|\delta\mathbf{k}|^2), \quad (11.9)$$

where ω_D is the Dirac frequency, v_D is the group velocity, and $\delta\mathbf{k} = \mathbf{k} - \mathbf{K}$.

11.1.5 Physical Picture of the Dirac Cone Dispersion

The formation of Dirac cones in honeycomb photonic crystals can be understood from two complementary perspectives:

Band-folding picture When the Brillouin zone of a triangular lattice is folded into the Brillouin zone of the honeycomb lattice (two-atom basis), two bands that were originally near the Γ point cross at the new K point. Such an “artificial” crossing can usually be lifted by perturbation, but the E representation of C_{3v} symmetry precisely prevents this lifting.

Tight-binding picture Treating the localized modes on each sublattice as “atomic orbitals,” the coupling between sublattices produces bonding and antibonding modes. At the K point, the Bloch phase factors $e^{i\mathbf{K} \cdot \boldsymbol{\delta}}$ ($\boldsymbol{\delta}$ being the nearest-neighbor vectors) sum to exactly zero, and the bonding–antibonding splitting vanishes, forming the Dirac degeneracy.

Both pictures are intuitive but not rigorous—a rigorous derivation must return to $\mathbf{k} \cdot \mathbf{p}$ perturbation theory, which is the content of the next section.

11.2 Honeycomb Tight-Binding Model and Band Structure

The tight-binding model is useful here mainly as a compact way to expose the sublattice structure. Keeping only nearest-neighbor coupling between the A and B sublattices, the three bond vectors can be written as

$$\boldsymbol{\delta}_1 = \frac{a}{\sqrt{3}} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{\delta}_2 = \frac{a}{\sqrt{3}} \begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix}, \quad \boldsymbol{\delta}_3 = \frac{a}{\sqrt{3}} \begin{pmatrix} -\sqrt{3}/2 \\ -1/2 \end{pmatrix}. \quad (11.10)$$

The corresponding two-band Hamiltonian is

$$\mathcal{H}(\mathbf{k}) = \begin{pmatrix} \varepsilon_A & t f(\mathbf{k}) \\ t f^*(\mathbf{k}) & \varepsilon_B \end{pmatrix}, \quad (11.11)$$

where t is the nearest-neighbor hopping integral and the structure factor is

$$f(\mathbf{k}) = \sum_{j=1}^3 e^{i\mathbf{k} \cdot \boldsymbol{\delta}_j} = e^{ik_y a / \sqrt{3}} + 2 \cos\left(\frac{k_x a}{2}\right) e^{-ik_y a / (2\sqrt{3})}. \quad (11.12)$$

Writing $\bar{\varepsilon} = (\varepsilon_A + \varepsilon_B)/2$ and $m = (\varepsilon_A - \varepsilon_B)/2$, Eq. (11.11) becomes

$$\mathcal{H}(\mathbf{k}) = \bar{\varepsilon} \sigma_0 + \text{Re}[t f(\mathbf{k})] \sigma_x - \text{Im}[t f(\mathbf{k})] \sigma_y + m \sigma_z, \quad (11.13)$$

with band frequencies

$$\omega_{\pm}^2(\mathbf{k}) = \bar{\varepsilon} \pm \sqrt{m^2 + t^2 |f(\mathbf{k})|^2}, \quad (11.14)$$

so the sublattice imbalance m is already visible as the Dirac mass.

The Dirac cone appears where $f(\mathbf{k}) = 0$. Direct computation:

$$f(\mathbf{K}) = e^{i \cdot 0} + 2 \cos\left(\frac{2\pi}{3}\right) e^{-i \cdot 0} = 1 + 2 \cdot (-1/2) = 0. \quad (11.15)$$

Similarly $f(\mathbf{K}') = 0$. Therefore, when $m = 0$ ($\varepsilon_A = \varepsilon_B$), the two bands are degenerate at the K and K' points—these are the Dirac cones.

Expanding $f(\mathbf{k})$ near the K point to first order:

$$f(\mathbf{K} + \delta\mathbf{k}) \approx \frac{\sqrt{3}a}{2} (\delta k_x - i \delta k_y) + O(|\delta\mathbf{k}|^2). \quad (11.16)$$

Substituting into the Hamiltonian recovers the effective Dirac form of (11.23), with $v_D = \sqrt{3}ta/2$.

11.3 $\mathbf{k} \cdot \mathbf{p}$ Perturbation Theory and the Effective Dirac Hamiltonian

11.3.1 Degenerate Perturbation Expansion

To rigorously derive the Dirac cone, we perform a $\mathbf{k} \cdot \mathbf{p}$ expansion of the Bloch eigenequation near the K point. The general framework was established in Chapter 3; here we apply it to the special case of the honeycomb photonic crystal.

Let the two degenerate states at the K point be $|u_1\rangle$ and $|u_2\rangle$, orthonormal under the material-weighted inner product

$$\langle u|v\rangle_\varepsilon = \int_{\text{cell}} \varepsilon(\mathbf{r}) u^*(\mathbf{r}) v(\mathbf{r}) d^2r. \quad (11.17)$$

Substitute $\mathbf{k} = \mathbf{K} + \delta\mathbf{k}$ into the cell equation (11.8) and treat $\delta\mathbf{k}$ as a perturbation parameter.

In the degenerate subspace, first-order perturbation gives the effective 2×2 Hamiltonian:

$$[\mathcal{H}_{\text{eff}}]_{ij} = \omega_D^2 \delta_{ij} + \sum_{\alpha=x,y} \delta k_\alpha \langle u_i | \hat{V}_\alpha | u_j \rangle_\varepsilon, \quad (11.18)$$

where the velocity operator is

$$\hat{V}_\alpha = \frac{c^2}{\omega_D} \left[\frac{1}{\varepsilon(\mathbf{r})} (\nabla + i\mathbf{K})_\alpha + (\nabla + i\mathbf{K})_\alpha \frac{1}{\varepsilon(\mathbf{r})} \right]. \quad (11.19)$$

11.3.2 Symmetry Constraints and the Dirac Form

C_{3v} symmetry imposes strong constraints on the matrix elements. Let the transformation of $|u_1\rangle$ and $|u_2\rangle$ under C_3 be

$$\hat{C}_3 |u_1\rangle = e^{i2\pi/3} |u_1\rangle, \quad \hat{C}_3 |u_2\rangle = e^{-i2\pi/3} |u_2\rangle; \quad (11.20)$$

then the diagonal matrix elements $\langle u_i | \hat{V}_\alpha | u_i \rangle$ are forced to zero by C_3 symmetry.

Proof that the diagonal matrix elements vanish. Consider $\langle u_1 | \hat{V}_x | u_1 \rangle$. Under C_3 rotation, $\hat{V}_x \rightarrow \cos(2\pi/3) \hat{V}_x - \sin(2\pi/3) \hat{V}_y$, while $|u_1\rangle \rightarrow e^{i2\pi/3} |u_1\rangle$. Therefore

$$\begin{aligned} \langle u_1 | \hat{V}_x | u_1 \rangle &= \langle u_1 | \hat{C}_3^\dagger \hat{C}_3 \hat{V}_x \hat{C}_3^\dagger \hat{C}_3 | u_1 \rangle \\ &= e^{-i2\pi/3} e^{i2\pi/3} [\cos(2\pi/3) \langle u_1 | \hat{V}_x | u_1 \rangle - \sin(2\pi/3) \langle u_1 | \hat{V}_y | u_1 \rangle]. \end{aligned} \quad (11.21)$$

Setting up a similar equation for the $\hat{V}_y$ component yields a homogeneous system whose only solution is $\langle u_1 | \hat{V}_x | u_1 \rangle = \langle u_1 | \hat{V}_y | u_1 \rangle = 0$. $\square$

The off-diagonal matrix elements are determined by the C_3 selection rules:

$$\langle u_1 | \hat{V}_x | u_2 \rangle = v_D, \quad \langle u_1 | \hat{V}_y | u_2 \rangle = i v_D, \quad (11.22)$$

where v_D is a real number (the group velocity). Therefore the effective Hamiltonian is

$$\boxed{\mathcal{H}_{\text{eff}}(\delta\mathbf{k}) = \omega_D \sigma_0 + v_D (\delta k_x \sigma_x + \delta k_y \sigma_y)}. \quad (11.23)$$

This is precisely the photonic version of the **massless Dirac equation**.

Remark (Analogy and differences with electronic graphene). The effective Hamiltonian above is formally identical to the Dirac cone of electrons in graphene. But the key differences are:

1. The photonic “mass” arises from symmetry breaking rather than spin–orbit coupling;
2. The inner product includes the material weight $\varepsilon(\mathbf{r})$, and the velocity v_D is determined by material-weighted matrix elements;
3. The photonic system has no Fermi surface; topological effects manifest in a frequency bandgap rather than an energy bandgap;
4. The effective Hamiltonian at the K' point is $\mathcal{H}_{K'} = \omega_D \sigma_0 + v_D(-\delta k_x \sigma_x + \delta k_y \sigma_y)$, i.e., the σ_x term has the opposite sign—a direct consequence of time-reversal symmetry.

11.3.3 Second-Order $\mathbf{k} \cdot \mathbf{p}$ Corrections and Trigonal Warping

When a more accurate dispersion relation is needed, second-order $\mathbf{k} \cdot \mathbf{p}$ corrections must be included. The second-order terms contain two types of contributions:

Remote-band corrections (Löwdin partitioning) Bands outside the degenerate subspace contribute effective-mass terms through second-order perturbation:

$$\mathcal{H}_{ij}^{(2)} = \sum_{\alpha, \beta} \delta k_\alpha \delta k_\beta \sum_{n \notin \{1,2\}} \frac{\langle u_i | \hat{V}_\alpha | u_n \rangle \langle u_n | \hat{V}_\beta | u_j \rangle}{\omega_D^2 - \omega_n^2}. \quad (11.24)$$

Trigonal warping The lowest-order anisotropic term allowed by C_3 symmetry is

$$\mathcal{H}_{\text{warp}} = \lambda_w (\delta k_+^3 \sigma_- + \delta k_-^3 \sigma_+), \quad (11.25)$$

where $\delta k_\pm = \delta k_x \pm i \delta k_y$, $\sigma_\pm = (\sigma_x \pm i \sigma_y)/2$, and λ_w is the warping coefficient. This term deforms the constant-frequency contours from circular to triangular, and becomes important at larger $|\delta \mathbf{k}|$.

11.4 Inversion Breaking and Valley Chern Number

11.4.1 Sublattice Inequivalence and the Mass Term

When the dielectric rods on the two sublattices differ (e.g., radii $r_A \neq r_B$ or permittivities $\varepsilon_A \neq \varepsilon_B$), spatial inversion symmetry $\mathcal{P}$ is broken while C_3 rotational symmetry is preserved. The breaking of $\mathcal{P}$ lifts the degeneracy protection of the E representation at the K point—the Dirac cone opens a bandgap.

At the level of the effective Hamiltonian, $\mathcal{P}$ breaking introduces a **mass term** $m \sigma_z$:

$$\boxed{\mathcal{H}_{\text{eff}}(\delta \mathbf{k}) = \omega_D \sigma_0 + v_D (\delta k_x \sigma_x + \delta k_y \sigma_y) + m \sigma_z.} \quad (11.26)$$

Proposition 11.2 (Microscopic expression for the mass term). *In $\mathbf{k} \cdot \mathbf{p}$ perturbation theory, the mass term m is determined by the sublattice inequivalence:*

$$m = \frac{c^2}{2\omega_D} \int_{cell} \Delta\left(\frac{1}{\varepsilon}\right) (|\nabla u_1 + i\mathbf{K}u_1|^2 - |\nabla u_2 + i\mathbf{K}u_2|^2) d^2r, \quad (11.27)$$

where $\Delta(1/\varepsilon) = 1/\varepsilon_{perturbed} - 1/\varepsilon_{unperturbed}$ characterizes the degree of inversion breaking.

Proof. Substitute $\varepsilon(\mathbf{r}) = \varepsilon_0(\mathbf{r}) + \delta\varepsilon(\mathbf{r})$ into the eigenequation, where ε_0 has $\mathcal{P}$ symmetry and $\delta\varepsilon$ breaks $\mathcal{P}$. In degenerate perturbation theory, the coefficient of the σ_z term is

$$m = \frac{1}{2} \left(\langle u_1 | \hat{V}_{\text{pert}} | u_1 \rangle - \langle u_2 | \hat{V}_{\text{pert}} | u_2 \rangle \right), \quad (11.28)$$

where $\hat{V}_{\text{pert}}$ is the perturbation operator induced by $\delta\varepsilon$. Since $\mathcal{P}$ exchanges $|u_1\rangle$ and $|u_2\rangle$ (corresponding to exchanging sublattices A and B), when $\mathcal{P}$ is broken the two diagonal matrix elements are no longer equal, and the mass term is nonzero. $\square$

The key physical point is not that the mass term changes sign between the two valleys. For a time-reversal-invariant valley-Hall photonic crystal with inversion symmetry breaking, the Dirac mass has the same sign at K and K' :

$$m_{K'} = m_K. \quad (11.29)$$

Time-reversal symmetry maps K to K' and reverses the valley chirality of the Dirac kinetic term, rather than forcing the sublattice mass σ_z to change sign. Consequently, the Berry curvature and the valley Chern numbers at the two valleys have opposite signs:

$$\Omega_K(\mathbf{q}) = -\Omega_{K'}(\mathbf{q}), \quad C_K = -C_{K'}. \quad (11.30)$$

11.4.2 Berry Curvature of the Massive Dirac Fermion

The Berry curvature of the lower band of the massive Dirac Hamiltonian eq. (11.26) can be computed analytically. Ignoring the constant term $\omega_D\sigma_0$, we write

$$\mathcal{H}(\delta\mathbf{k}) = \mathbf{d}(\delta\mathbf{k}) \cdot \boldsymbol{\sigma}, \quad \mathbf{d} = (v_D\delta k_x, v_D\delta k_y, m).$$

For the lower band, the Berry curvature is

$$\Omega_-(\delta\mathbf{k}) = -\frac{1}{2} \hat{\mathbf{d}} \cdot \left(\frac{\partial \hat{\mathbf{d}}}{\partial(\delta k_x)} \times \frac{\partial \hat{\mathbf{d}}}{\partial(\delta k_y)} \right) = -\frac{mv_D^2}{2(v_D^2|\delta\mathbf{k}|^2 + m^2)^{3/2}}. \quad (11.31)$$

Derivation of the Berry curvature formula. Let

$$D = |\mathbf{d}| = \sqrt{v_D^2|\delta\mathbf{k}|^2 + m^2}, \quad \hat{\mathbf{d}} = \frac{\mathbf{d}}{D}.$$

A useful identity is

$$\hat{\mathbf{d}} \cdot \left(\partial_{k_x} \hat{\mathbf{d}} \times \partial_{k_y} \hat{\mathbf{d}} \right) = \frac{\mathbf{d} \cdot (\partial_{k_x} \mathbf{d} \times \partial_{k_y} \mathbf{d})}{|\mathbf{d}|^3}. \quad (11.32)$$

For the massive Dirac model,

$$\partial_{\delta k_x} \mathbf{d} = (v_D, 0, 0), \quad \partial_{\delta k_y} \mathbf{d} = (0, v_D, 0).$$

Therefore

$$\partial_{\delta k_x} \mathbf{d} \times \partial_{\delta k_y} \mathbf{d} = (0, 0, v_D^2),$$

and hence

$$\mathbf{d} \cdot (\partial_{\delta k_x} \mathbf{d} \times \partial_{\delta k_y} \mathbf{d}) = m v_D^2.$$

Substituting this into (11.32) gives

$$\hat{\mathbf{d}} \cdot \left(\frac{\partial \hat{\mathbf{d}}}{\partial(\delta k_x)} \times \frac{\partial \hat{\mathbf{d}}}{\partial(\delta k_y)} \right) = \frac{m v_D^2}{(v_D^2 |\delta \mathbf{k}|^2 + m^2)^{3/2}}.$$

Thus the lower-band Berry curvature is

$$\Omega_-(\delta \mathbf{k}) = -\frac{m v_D^2}{2 (v_D^2 |\delta \mathbf{k}|^2 + m^2)^{3/2}}.$$

□

The Berry curvature is maximal in magnitude at the valley center $\delta \mathbf{k} = 0$:

$$\Omega_-(0) = -\frac{\text{sgn}(m) v_D^2}{2|m|^2}. \quad (11.33)$$

It decays as $|\delta \mathbf{k}|^{-3}$ away from the valley. Thus the Berry curvature is strongly concentrated near the K or K' point, which is the geometric origin of valley topology.

For the time-reversed valley, the orientation of the Dirac Hamiltonian is reversed, so the sign of the Berry curvature changes accordingly. This is why the two valleys usually carry opposite valley Berry curvature in a time-reversal-symmetric honeycomb system.

11.4.3 Definition and Computation of the Valley Chern Number

Definition 11.3 (Valley Chern number). The valley Chern number of the lower band near the K valley is defined as

$$C_K = \frac{1}{2\pi} \int_{\text{near } K} \Omega_-(\delta \mathbf{k}) d^2(\delta \mathbf{k}). \quad (11.34)$$

Integrating the Berry curvature (11.31) over the entire $\delta \mathbf{k}$ plane:

$$\begin{aligned} C_K &= \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} \left(-\frac{m v_D^2}{2(v_D^2 q^2 + m^2)^{3/2}} \right) q dq d\theta \\ &= -\frac{m v_D^2}{2} \int_0^\infty \frac{q dq}{(v_D^2 q^2 + m^2)^{3/2}} \\ &= -\frac{m v_D^2}{2} \cdot \left[-\frac{1}{v_D^2 (v_D^2 q^2 + m^2)^{1/2}} \right]_0^\infty = -\frac{1}{2} \text{sgn}(m). \end{aligned} \quad (11.35)$$

Similarly, $C_{K'} = +\frac{1}{2}\text{sgn}(m)$. The valley Chern number difference is

$$\Delta C_v = C_K - C_{K'} = -\text{sgn}(m). \quad (11.36)$$

Theorem 11.4 (Half-integer quantization of the valley Chern number). *In the massive Dirac approximation, the Chern number of a single valley is strictly $\pm 1/2$. The sum of the Chern numbers of the two valleys is zero ($C_K + C_{K'} = 0$), consistent with the total Chern number being zero over the full Brillouin zone (a necessary consequence of $\mathcal{T}$ symmetry).*

Remark (“Quasi-topological” nature of the valley Chern number). It must be emphasized: the valley Chern number $C_K = \pm 1/2$ is **not** a strict topological invariant. Its value depends on how the integration region “near K ” is defined. In the continuum approximation, the integration region can be extended to the entire plane, giving exactly $\pm 1/2$. But on a real discrete lattice, the Berry curvature distributions of the K and K' valleys overlap, and the deviation depends on $m/(v_D\pi/a)$.

11.5 Rigorous Theory of Domain-Wall Interface States

11.5.1 The Jackiw–Rebbi Mechanism

Consider two semi-infinite photonic crystal regions joined at $y = 0$: the left side ($y < 0$) has mass $m_L > 0$, the right side ($y > 0$) has mass $m_R < 0$. The mass undergoes a sign reversal at the interface.

In 1976, Jackiw and Rebbi proved that a sign reversal of the mass in the one-dimensional Dirac equation necessarily leads to a zero-energy bound state. This mechanism directly generalizes to our photonic system.

Theorem 11.5 (Existence of photonic domain-wall states). *Let the mass $m(y)$ in the massive Dirac equation $[v_D(-i\partial_y\sigma_y + k_x\sigma_x) + m(y)\sigma_z]\phi(y) = (\omega - \omega_D)\phi(y)$ change sign at $y = 0$. Then there exists a state bound to the interface, with dispersion relation*

$$\omega(k_x) = \omega_D + v_D k_x \text{sgn}(m_L). \quad (11.37)$$

Proof. Let the interface-state wave function be $\phi(y) = \begin{pmatrix} \phi_1(y) \\ \phi_2(y) \end{pmatrix} e^{ik_x x}$. Substituting into the Dirac equation gives the coupled equations:

$$v_D k_x \phi_2 + v_D \partial_y \phi_2 + m(y) \phi_1 = (\omega - \omega_D) \phi_1, \quad (11.38a)$$

$$v_D k_x \phi_1 - v_D \partial_y \phi_1 - m(y) \phi_2 = (\omega - \omega_D) \phi_2. \quad (11.38b)$$

Seeking the zero mode at $k_x = 0$: set $\omega = \omega_D$, then

$$v_D \partial_y \phi_2 + m(y) \phi_1 = 0, \quad -v_D \partial_y \phi_1 - m(y) \phi_2 = 0. \quad (11.39)$$

Taking $\phi_2 = i\phi_1$ gives

$$\phi_1(y) = N \exp\left(-\frac{1}{v_D} \int_0^y |m(y')| dy'\right), \quad (11.40)$$

where the fact that $m_L > 0$ and $m_R < 0$ ensures that ϕ_1 decays exponentially as $y \rightarrow \pm\infty$.

For finite k_x , first-order perturbation theory gives the dispersion correction $\omega - \omega_D = v_D k_x \text{sgn}(m_L)$. $\square$

11.5.2 Interface States from the K and K' Valleys

At the K' valley, since $m_{K'} = -m_K$, the group velocity of the interface state is reversed:

$$\omega_K(k_x) = \omega_D + v_D k_x, \quad \omega_{K'}(k_x) = \omega_D - v_D k_x. \quad (11.41)$$

The interface states from the two valleys form a pair of **counter-propagating** modes.

11.5.3 Exact Solution for a Continuous Mass Profile

For a smooth mass distribution $m(y) = m_0 \tanh(y/\xi)$ (where ξ is the domain-wall width), the Dirac equation can be solved exactly. The zero-mode wave function is

$$\phi_1(y) = N [\cosh(y/\xi)]^{-m_0 \xi / v_D}. \quad (11.42)$$

The penetration depth is $\delta \sim \xi + v_D/|m_0|$. In the step-function limit $\xi \rightarrow 0$, $\delta \rightarrow v_D/|m_0|$.

11.5.4 Domain-Wall Types and Differences in Topological Protection

The honeycomb photonic crystal has two basic domain-wall configurations:

Definition 11.6 (Zigzag and armchair domain walls).

- **Zigzag domain wall:** the interface runs along the zigzag direction. In the interface Brillouin zone, the projections of the K and K' valleys are **separated**.
- **Armchair domain wall:** the interface runs along the armchair direction. The projections of the K and K' valleys **coincide**.

Proposition 11.7 (Suppression of intervalley scattering at zigzag domain walls). *At a zigzag domain wall, scattering from the K valley to the K' valley requires a large momentum transfer $|\Delta k_{\parallel}| \sim 4\pi/(3a)$. For slowly varying defects, this large momentum transfer is strongly suppressed.*

Proposition 11.8 (Lack of protection at armchair domain walls). *At an armchair domain wall, the K and K' valleys project onto the same $k_{\parallel}$, so **any** perturbation can potentially cause intervalley scattering.*

Remark (Essential difference from Chern insulator edge states). This dependence on domain-wall type profoundly reveals the essential difference between valley topology and integer-Chern topology:

- **Chern insulator:** the existence of edge states is guaranteed by an integer topological invariant, immune to **all** perturbations that do not close the bulk bandgap, regardless of boundary orientation.
- **Valley topology:** the protection of domain-wall states depends on the suppression of intervalley scattering, which is a **dynamical** condition rather than a topological one.

11.6 Numerical Verification: The Supercell Method

The continuum theory can be checked by a supercell calculation. Construct a strip containing two domains with opposite masses: the first N periods have the $m > 0$ configuration, the last N periods have the $m < 0$ configuration, and Bloch periodicity is maintained along the interface direction. In the supercell, the frequency-domain eigenequation is

$$-(\nabla + ik_x \hat{x}) \cdot \left[\frac{1}{\varepsilon_{\text{sc}}(\mathbf{r})} (\nabla + ik_x \hat{x}) u_{n,k_x} \right] = \frac{\omega_n^2(k_x)}{c^2} u_{n,k_x}. \quad (11.43)$$

Solving by the plane-wave expansion or finite-element method gives the numerical signatures expected from the Jackiw–Rebbi analysis:

1. A pair of domain-wall bands appears within the bulk gap and connects the projected K and K' valleys.
2. The two branches have opposite group velocities and are associated with opposite valleys.
3. The field distribution decays exponentially away from the interface, with penetration depth $\xi \sim 1/|\Delta m|$.
4. Zigzag domain walls show much weaker intervalley mixing than armchair domain walls.

In a finite supercell, the two interfaces may weakly couple through bulk-state tunneling and produce a small avoided crossing. This splitting decays exponentially with the strip width and is a finite-size effect rather than a breakdown of the valley-domain-wall mechanism.

11.7 120° Bend Transmission and Experimental Verification

Proposition 11.9 (120° bend valley conservation). *At a zigzag domain wall, a 120° bend rotates the propagation direction from $\hat{x}$ to $C_3 \hat{x}$. Since C_3 rotation maps the K point back to the K point (modulo a reciprocal lattice vector), scattering at the bend remains within the same valley. Therefore the transmittance is close to 1.*

This is the basic reason why valley domain-wall waveguides can show low loss at sharp 120° bends. By contrast, a 60° bend maps K to K' and therefore tends to induce intervalley scattering. Valley photonic waveguides are thus experimentally attractive in microwave, terahertz, and integrated-photonic platforms, but their bend robustness remains tied to preservation of the valley index.

11.8 Analysis of the Limitations of Valley Topological Protection

The protection of valley domain-wall states is conditional. It relies on suppressing intervalley scattering, so it is strongest for smooth disorder, zigzag-type interfaces, and bends compatible with the C_3 structure of the valley points. It is weakened by short-range defects, armchair-type interface segments, 60° bends, and finite structures where opposite interfaces can hybridize.

This should be contrasted with Chern-insulator edge states, whose existence follows from an integer Chern number and does not depend on valley separation. Valley photonic crystals are

therefore best understood as a practically useful and experimentally convenient quasi-topological platform, not as a system with the same unconditional protection as a Chern phase.

Chapter Summary

Starting from the C_{6v} symmetry of the honeycomb photonic crystal, this chapter derived the formation of Dirac cones at the K/K' points and the mass term induced by spatial inversion breaking via $\mathbf{k} \cdot \mathbf{p}$ perturbation theory, defined the valley Chern number and analyzed its half-integer nature. Through the Jackiw–Rebbi mechanism, the analytical theory of domain-wall interface states was established, and the differences in topological protection between zigzag and armchair domain walls were discussed.

Core conclusion: Valley topology provides an experimentally convenient “quasi-topological” protection mechanism, but its protection is inherently **approximate** and **conditional**.

Chapter 12

Three-Dimensional Topological Photonics

Why Topology Changes in Three Dimensions?

The preceding chapters primarily discussed topological phenomena in two-dimensional photonic systems—whether Chern insulators (Chapter 5), quantum spin Hall analogues (Chapter 6), or valley topology (Chapter 11), their topological invariants were all defined on a two-dimensional Brillouin zone. However, the real world is three-dimensional, and extending topological concepts to three-dimensional photonic crystals not only has fundamental theoretical value but also opens up intrinsically three-dimensional phenomena—**Weyl points**, **Fermi arc** surface states, **nodal lines**, and **nodal surfaces**.

The core physics of three-dimensional topological photonics is: in the three-dimensional Brillouin zone, band crossing points (degeneracies) can carry topological charges, and their existence and stability are jointly determined by symmetry and topological protection. This chapter starts from the three-dimensional Maxwell eigenvalue problem, uses $\mathbf{k} \cdot \mathbf{p}$ theory to analyze the effective Hamiltonians near band crossing points, defines the corresponding topological invariants, and discusses the bulk–surface correspondence.

12.1 The Maxwell Eigenvalue Problem in 3D Photonic Crystals

12.1.1 Three-Dimensional Bloch Eigenequation

In a three-dimensional nonmagnetic periodic medium with scalar $\varepsilon(\mathbf{r})$, the frequency-domain Maxwell eigenequation for the magnetic field is

$$\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H} \right) = \frac{\omega^2}{c^2} \mathbf{H}, \quad (12.1)$$

where $\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{R})$ for all lattice vectors $\mathbf{R}$. By Bloch's theorem, the eigenmodes are

$$\mathbf{H}_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{u}_{n\mathbf{k}}(\mathbf{r}), \quad (12.2)$$

where $\mathbf{u}_{n\mathbf{k}}$ has lattice periodicity and $\mathbf{k}$ ranges over the three-dimensional Brillouin zone (BZ).

The key difference from the two-dimensional case is: in three-dimensional systems, $\mathbf{H}$ is a vector field, and the transversality condition $\nabla \cdot \mathbf{H} = 0$ (or, more generally, $\nabla \cdot \mathbf{B} = 0$) provides an additional constraint, leaving two physical polarization degrees of freedom at each $\mathbf{k}$ point. This vector nature plays a fundamental role in band degeneracies and topological classification.

Substituting the Bloch form gives the $\mathbf{k}$ -parameterized cell equation:

$$(\nabla + i\mathbf{k}) \times \left[\frac{1}{\varepsilon(\mathbf{r})} (\nabla + i\mathbf{k}) \times \mathbf{u}_{n\mathbf{k}} \right] = \frac{\omega_{n\mathbf{k}}^2}{c^2} \mathbf{u}_{n\mathbf{k}}. \quad (12.3)$$

12.1.2 Polarization Degrees of Freedom

For the nonmagnetic $\mathbf{H}$ formulation above, the Bloch states $\mathbf{u}_{n\mathbf{k}}(\mathbf{r})$ in three-dimensional photonic crystals are transverse vector fields:

$$(\nabla + i\mathbf{k}) \cdot \mathbf{u}_{n\mathbf{k}}(\mathbf{r}) = 0. \quad (12.4)$$

Thus the physical state space at each $\mathbf{k}$ is not a scalar Bloch space but the transverse subspace of vector fields. For the purposes of this chapter, the important point is not the numerical gauge choice, but the fact that 3D photonic bands carry vector-polarization structure while their band crossings live in a three-dimensional momentum space.

12.1.3 Symmetry and Classification of Band Degeneracies

The space-group symmetry of a three-dimensional photonic crystal determines which degeneracies can appear at high-symmetry points, lines, or planes of the Brillouin zone. The main possibilities are:

- **Weyl points:** isolated twofold degeneracies carrying integer Chern charge.
- **Dirac points:** fourfold degeneracies, often understood as two opposite-chirality Weyl points pinned together by crystalline symmetry.
- **Nodal lines:** one-dimensional band-crossing curves protected by mirror, glide, inversion–time-reversal, or related symmetries.
- **Nodal surfaces:** two-dimensional degeneracy manifolds, usually tied to nonsymmorphic space-group constraints.

The rest of this chapter focuses on the most important case, Weyl points and their Fermi-arc surface states, and then briefly treats Dirac points and nodal lines as related 3D degeneracies.

12.2 Weyl Photonic Crystals

12.2.1 $\mathbf{k} \cdot \mathbf{p}$ Theory of Weyl Points

Near the crossing point $\mathbf{k}_W$ of two bands, the effective Hamiltonian can be expanded as

$$\mathcal{H}_{\text{eff}}(\delta\mathbf{k}) = E_0 \sigma_0 + \sum_{i,j=x,y,z} v_{ij} \delta k_i \sigma_j + O(|\delta\mathbf{k}|^2), \quad (12.5)$$

where $\delta\mathbf{k} = \mathbf{k} - \mathbf{k}_W$, σ_j are the Pauli matrices, v_{ij} is the 3×3 velocity tensor, and E_0 is the crossing-point energy.

Theorem 12.1 (Topological stability of Weyl points). *When $\det(v_{ij}) \neq 0$, the Weyl point is **topologically stable**—any small perturbation that preserves translational symmetry can only shift the Weyl point’s position in $\mathbf{k}$ space but cannot eliminate it. A Weyl point can only disappear by annihilating with another Weyl point of opposite chirality.*

Proof. The eigenvalues of the effective Hamiltonian $\mathcal{H} = \sum_{ij} v_{ij} \delta k_i \sigma_j$ (ignoring $E_0 \sigma_0$) are

$$E_{\pm}(\delta\mathbf{k}) = \pm \sqrt{\sum_j \left(\sum_i v_{ij} \delta k_i \right)^2}. \quad (12.6)$$

The zero-energy condition $E = 0$ requires three independent equations

$$\sum_i v_{ij} \delta k_i = 0, \quad j = x, y, z \quad (12.7)$$

to hold simultaneously. These are three linear equations in the three-dimensional $\delta\mathbf{k}$ space. When $\det(v_{ij}) \neq 0$, the unique solution is $\delta\mathbf{k} = 0$ —an **isolated point** in the 3D BZ.

Now consider a small perturbation $\delta\mathcal{H} = \sum_j \epsilon_j \sigma_j$. The zero-energy condition after perturbation becomes

$$\sum_i v_{ij} \delta k_i + \epsilon_j = 0, \quad j = x, y, z. \quad (12.8)$$

The solution is $\delta k_i = -\sum_j (v^{-1})_{ij} \epsilon_j$ —the Weyl point moves to a new position but is **not eliminated**.

The fundamental reason for this stability is **dimension counting**: a generic two-band degeneracy requires simultaneously satisfying three real conditions (the coefficients of the three Pauli matrices being zero), and the 3D $\mathbf{k}$ space provides exactly three tunable coordinates. Therefore a Weyl point is a stable codimension-three degeneracy in three momentum dimensions. $\square$

Remark (Comparison with 2D Dirac points). In two-dimensional systems, a stable Dirac point usually requires a symmetry that forbids the σ_z mass term; then only two conditions (σ_x and σ_y coefficients being zero) remain, matching the two momentum coordinates. If that symmetry is broken, a σ_z mass opens the gap—this is precisely the origin of valley topology in Chapter 11. Three-dimensional Weyl points lack this particular weakness because the σ_z coefficient is one of the three momentum-linear coordinates rather than an independent mass.

12.2.2 Topological Charge and Chirality

The topological charge (chirality) of a Weyl point is defined as the Chern number on a closed surface enclosing that point:

$$\chi = \frac{1}{2\pi} \oint_S \mathbf{\Omega}(\mathbf{k}) \cdot d\mathbf{S} = \text{sgn}[\det(v_{ij})] = \pm 1, \quad (12.9)$$

where $\mathbf{\Omega}(\mathbf{k}) = \nabla_{\mathbf{k}} \times \mathbf{A}(\mathbf{k})$ is the Berry curvature vector and, following the convention of Chapter 4, $\mathbf{A}(\mathbf{k}) = -i\langle u_{\mathbf{k}} | \nabla_{\mathbf{k}} | u_{\mathbf{k}} \rangle$ is the Berry connection. For the two bands of a minimal Weyl Hamiltonian the Chern numbers are opposite; the sign attached to “upper” or “lower” band depends on this band convention, while the Weyl chirality itself is $\text{sgn}[\det(v_{ij})]$.

A Weyl point is a “magnetic monopole” of Berry curvature in $\mathbf{k}$ space—Berry curvature radiates outward from a $\chi = +1$ Weyl point and converges into a $\chi = -1$ Weyl point. This analogy is not merely formal: the divergence of the Berry curvature $\nabla_{\mathbf{k}} \cdot \mathbf{\Omega} = 2\pi \sum_{\alpha} \chi_{\alpha} \delta^3(\mathbf{k} - \mathbf{k}_{W,\alpha})$ is the exact transcription of the magnetic monopole equation.

Theorem 12.2 (Photonic version of the Nielsen–Ninomiya theorem). *In the compact Brillouin zone (periodic boundary conditions), Weyl points always appear in pairs, and the total topological charge is zero:*

$$\sum_{\alpha} \chi_{\alpha} = 0. \quad (12.10)$$

Proof. By Gauss’s theorem, the integral of the divergence of Berry curvature over the entire BZ equals the flux of Berry curvature through the BZ boundary. But the BZ is a compact manifold without boundary (opposite faces are identified), so the boundary flux is zero:

$$0 = \int_{\text{BZ}} \nabla_{\mathbf{k}} \cdot \mathbf{\Omega} d^3k = 2\pi \sum_{\alpha} \chi_{\alpha}. \quad (12.11)$$

Therefore $\sum_{\alpha} \chi_{\alpha} = 0$. □

12.2.3 Realizing Weyl Points in Photonic Crystals

Two main routes to realizing Weyl points in photonic crystals:

Breaking time-reversal symmetry Introducing magneto-optical materials (such as gyroelectric media) into a three-dimensional photonic crystal gives the ε tensor an antisymmetric imaginary part:

$$\varepsilon = \begin{pmatrix} \varepsilon_{\perp} & i\kappa & 0 \\ -i\kappa & \varepsilon_{\perp} & 0 \\ 0 & 0 & \varepsilon_z \end{pmatrix}. \quad (12.12)$$

This is the direct analogue of magnetic Weyl semimetals in condensed matter. When $\mathcal{T}$ is broken but $\mathcal{P}$ is preserved, inversion relates Weyl points at $\mathbf{k}$ and $-\mathbf{k}$ with opposite chirality, so the minimum number can be two. They need not sit at time-reversal-invariant momenta.

Breaking spatial inversion symmetry Preserving $\mathcal{T}$ but breaking $\mathcal{P}$. In this case, due to $\mathcal{T}$ symmetry, Weyl points at $\mathbf{k}_W$ and $-\mathbf{k}_W$ have the same chirality. Combined with the Nielsen–Ninomiya theorem, a minimum of 4 Weyl points is required (two pairs, each pair related by $\mathcal{T}$).

Remark (Tilted Weyl cones). A term proportional to $\mathbf{v}_0 \cdot \delta\mathbf{k} \sigma_0$ tilts the Weyl cone without changing its chirality. Strong tilting can produce the so-called type-II Weyl point, but this distinction affects the iso-frequency geometry more than the basic monopole charge and Fermi-arc bulk–surface correspondence emphasized here.

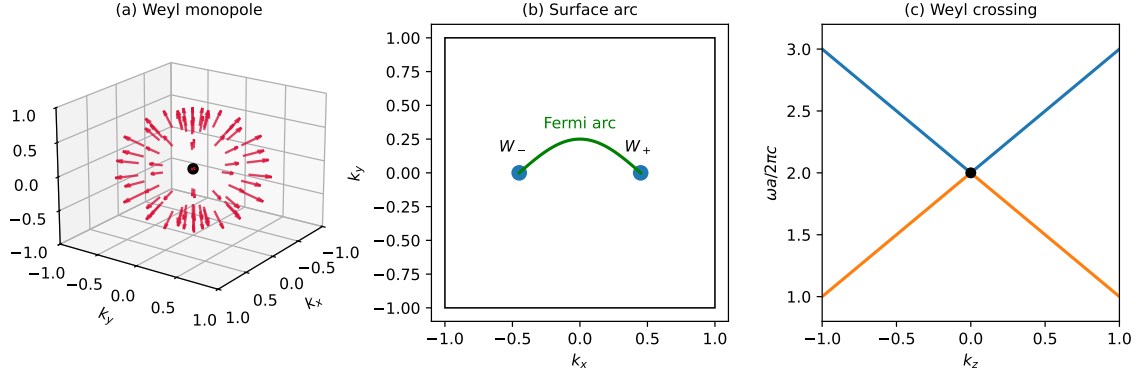


Figure 12.1 Schematic of the topological structure of a Weyl photonic crystal: Berry curvature magnetic monopoles, surface Fermi arcs, and the linear dispersion through the Weyl point.

12.3 Fermi Arcs and Surface States

The most striking physical consequence of Weyl points is their surface states—Fermi arcs. Unlike the edge states of two-dimensional topological insulators, Fermi arcs are open iso-frequency line segments connecting the projections of two Weyl points of opposite chirality onto the surface Brillouin zone. This section rigorously derives the existence and dispersion relation of Fermi arcs from the bulk–edge correspondence.

12.3.1 Bulk–Edge Correspondence and the Chirality of Weyl Points

Consider a three-dimensional photonic crystal containing one pair of Weyl points, located at $\mathbf{k}_+$ and $\mathbf{k}_-$ in momentum space, with chiralities $\chi_+ = +1$ and $\chi_- = -1$ respectively.

For any two-dimensional cross-section S (at fixed k_z) that does not pass through a Weyl point, one can define the Chern number on that cross-section:

$$C(k_z) = \frac{1}{2\pi} \int_{S(k_z)} \Omega_{xy}(\mathbf{k}) dk_x dk_y. \quad (12.13)$$

Theorem 12.3 (Chern number jump and Weyl point chirality). *When k_z sweeps from one side of the Weyl point $\mathbf{k}_+$ to the other, the Chern number jumps:*

$$\Delta C = C(k_z^+) - C(k_z^-) = \chi_+, \quad (12.14)$$

where $k_z^\pm$ are above and below the k_z coordinate of the Weyl point.

Proof. A Weyl point is a magnetic monopole source of Berry curvature. Near the Weyl point, the divergence of the Berry curvature is

$$\nabla_{\mathbf{k}} \cdot \boldsymbol{\Omega}(\mathbf{k}) = 2\pi\chi \delta^3(\mathbf{k} - \mathbf{k}_W), \quad (12.15)$$

where χ is the chirality.

Applying Gauss's theorem to an infinitesimal region enclosing the Weyl point:

$$\oint_{\partial V} \boldsymbol{\Omega} \cdot d\mathbf{S} = \int_V \nabla_{\mathbf{k}} \cdot \boldsymbol{\Omega} d^3k = 2\pi\chi. \quad (12.16)$$

Decomposing the closed surface ∂V into upper and lower k_z cross-sections, the upper cross-section contributes $+2\pi C(k_z^+)$, the lower contributes $-2\pi C(k_z^-)$, and the lateral contribution vanishes in the thin-slab limit used to compare two infinitesimally separated slices. Therefore $2\pi[C(k_z^+) - C(k_z^-)] = 2\pi\chi$, i.e., $\Delta C = \chi$. $\square$

The physical meaning of this theorem is: Weyl points are “sources” and “sinks” of Chern number in momentum space. A positive-chirality Weyl point generates one unit of Chern number; a negative-chirality one absorbs one unit.

12.3.2 Proof of Surface State Existence

Theorem 12.4 (Existence of Fermi arcs). *Let a three-dimensional photonic crystal contain a pair of Weyl points of opposite chirality, $\mathbf{k}_+$ ($\chi = +1$) and $\mathbf{k}_-$ ($\chi = -1$). On a surface whose projection separates the two Weyl points in the surface Brillouin zone, surface states form open iso-frequency contours whose endpoints lie at the projected Weyl points. The detailed connection pattern depends on the surface termination and on other Weyl points at the same frequency.*

Proof. Consider a semi-infinite system with surface normal $\hat{z}$. Equivalently, choose a family of two-dimensional bulk slices parallel to the surface and parameterized by the remaining momentum component.

Between the two Weyl points, the slice Chern number is nonzero (by the previous theorem, it jumps when the slice crosses a Weyl point).

By the bulk–edge correspondence principle, each two-dimensional slice with nonzero Chern number has $|C|$ chiral edge states at the boundary. Therefore, over the interval of slice momenta between the projected Weyl points, topologically protected surface states exist.

These surface states form an open curve in the surface Brillouin zone connecting $\bar{\mathbf{k}}_+$ and $\bar{\mathbf{k}}_-$ —the Fermi arc. $\square$

A notable feature of the Fermi arc is that it is an **open** iso-frequency line: it has two endpoints (the Weyl point projections), unlike the closed Fermi surfaces of ordinary metals. This topological feature can be directly observed through angle-resolved reflection spectroscopy.

Numerically, the same conclusion is usually checked by computing Berry-flux charge on a small enclosing surface around each Weyl point, or by tracking the Chern number of two-dimensional momentum slices across the Weyl point. The detailed connection pattern of Fermi arcs can depend on surface termination, but their endpoints and net connectivity are fixed by the projected Weyl charges.

12.4 Dirac Points and Nodal-Line Photonic Crystals

Weyl points are the generic stable twofold degeneracies in three dimensions. Other 3D degeneracies usually require additional crystalline constraints.

12.4.1 Symmetry-Protected Dirac Points

In electronic spin-1/2 systems, time-reversal and inversion symmetries can enforce Kramers degeneracy. Ordinary photonic time reversal satisfies $\Theta^2 = +1$, so fourfold photonic Dirac points

instead require suitable crystalline representations or accidental tuning. They can be viewed as two Weyl points of opposite chirality pinned together by symmetry:

$$\mathcal{H}_{\text{Dirac}}(\delta\mathbf{k}) = \begin{pmatrix} \mathcal{H}_+(\delta\mathbf{k}) & \Delta(\delta\mathbf{k}) \\ \Delta^\dagger(\delta\mathbf{k}) & \mathcal{H}_-(\delta\mathbf{k}) \end{pmatrix}. \quad (12.17)$$

When symmetry forces the coupling Δ to vanish, the two opposite Weyl charges remain superposed. If that symmetry is broken, the Dirac point can either split into Weyl points or gap out.

Proposition 12.5 ($\mathcal{T} + \mathcal{P}$ forbids isolated Weyl points). *In a nonmagnetic, inversion-symmetric photonic crystal with ordinary bosonic $\mathcal{T}$ and $\mathcal{P}$, the Berry curvature of a nondegenerate isolated band vanishes away from degeneracies. Therefore an isolated twofold Weyl point with nonzero Chern charge is forbidden unless one of these symmetries is broken or the relevant subspace realizes an additional pseudo-spin structure.*

12.4.2 Nodal-Line Photonic Crystals

When two bands are degenerate along a closed curve in the Brillouin zone, a **nodal line** is formed. Its basic topological invariant is the Berry phase around a loop encircling the line:

$$\gamma = \oint_{\ell} \mathcal{A}(\mathbf{k}) \cdot d\mathbf{k} = 0 \text{ or } \pi \pmod{2\pi}, \quad (12.18)$$

where the quantization requires a protecting symmetry such as mirror, inversion–time-reversal, glide, or another appropriate crystalline constraint. A nontrivial nodal line can support drumhead surface states inside the projected nodal loop. Compared with Fermi arcs, which are one-dimensional open contours connecting projected Weyl points, drumhead states fill a two-dimensional region in the surface Brillouin zone.

12.5 Numerical and Experimental Signatures

In calculations, Weyl points are usually located by first finding small gaps between target bands and then verifying the Berry-flux charge through a small enclosing surface. Equivalently, one tracks the Chern number of two-dimensional momentum slices and checks that it jumps when the slice crosses a Weyl point.

Experimentally, the key photonic signatures are the bulk Weyl degeneracies and the surface Fermi arcs. The first experimental verification in 3D photonic crystals was achieved in the microwave regime using a double-gyroid structure with broken inversion symmetry [10]. Moving the same physics to optical frequencies is harder because it requires precise three-dimensional nanofabrication and careful control of material loss, but the topological mechanism is the same: Weyl points act as Berry-curvature monopoles, and Fermi arcs appear as their surface manifestation.

Chapter Summary

This chapter extended topological concepts from two dimensions to three-dimensional photonic crystals, focusing on Weyl points, Fermi arcs, Dirac points, and nodal lines as representative three-dimensional topological structures. Starting from the three-dimensional Maxwell eigenequation, $\mathbf{k} \cdot \mathbf{p}$ theory was used to derive the effective Hamiltonians and topological invariants of band crossings. The bulk–surface correspondence was established through slice Chern numbers, explaining the topological origin of Fermi arc surface states.

Core conclusion: The richness of three-dimensional topological photonics stems from the high dimensionality of the 3D $\mathbf{k}$ space. In this setting, isolated Weyl points can carry integer monopole charges, their projections are connected by Fermi arcs, and symmetry can further stabilize related Dirac points and nodal lines.

Nonlinear Topological Photonics: From Maxwell's Equations to Topo- logical Invariants

Why Topology Must Go Beyond Linearity?

The topological photonics theory established in all preceding chapters was based on **linear Maxwell's equations**—the dielectric constant $\varepsilon(\mathbf{r})$ does not depend on the field strength. However, real photonic systems inevitably exhibit nonlinear responses under strong-field conditions: the Kerr effect, second-harmonic generation, parametric amplification, etc. A fundamental question therefore arises: **Does topological protection still hold in nonlinear systems? If so, what is its mathematical basis?**

This chapter starts from the nonlinear Maxwell equations and systematically constructs a unified theoretical framework for nonlinear topological photonics [33]. The core difficulty of this task lies in the fact that the cornerstones of linear topological theory—Bloch's theorem, band structure, Berry phase—all rely on the linear superposition principle of the eigenvalue problem, and nonlinear systems fundamentally break this structure. We will show how to recover topological concepts at different physical levels through **weak nonlinear expansion**, **self-consistent construction of nonlinear Bloch waves**, and **variational theory of topological solitons**.

13.1 General Framework of the Nonlinear Maxwell Equations

13.1.1 Maxwell's Equations with Nonlinear Polarization

In non-magnetic media, Maxwell's equations including the nonlinear response are

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \quad (13.1a)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \quad (13.1b)$$

where the electric displacement vector contains both linear and nonlinear contributions:

$$\mathbf{D} = \varepsilon_0 \varepsilon(\mathbf{r}) \mathbf{E} + \mathbf{P}_{\text{NL}}(\mathbf{E}). \quad (13.2)$$

For third-order (Kerr-type) nonlinearity—the most prevalent nonlinear mechanism in photonic systems—the nonlinear polarization is

$$\mathbf{P}_{\text{NL}} = \varepsilon_0 \chi^{(3)}(\mathbf{r}) |\mathbf{E}|^2 \mathbf{E}, \quad (13.3)$$

where $\chi^{(3)}(\mathbf{r})$ is the third-order nonlinear susceptibility. In centrosymmetric media (such as silicon, silicate glass), the second-order nonlinearity $\chi^{(2)}$ vanishes, and the Kerr effect is the dominant nonlinearity.

For second-order nonlinearity (such as in non-centrosymmetric crystals like LiNbO_3),

$$P_i^{\text{NL}} = \varepsilon_0 \sum_{jk} \chi_{ijk}^{(2)}(\mathbf{r}) E_j E_k. \quad (13.4)$$

Remark (The fundamental break between nonlinear and linear theory). In linear theory, one can solve for the eigenmodes $\{\mathbf{E}_{n\mathbf{k}}\}$, and any field is a linear superposition of eigenmodes. In nonlinear systems, the superposition principle fails: the sum of two solutions is generally no longer a solution. Therefore, Bloch's theorem, band structure, Berry phase, and other concepts all need to be reexamined. This is the core theoretical challenge of nonlinear topological photonics.

13.1.2 Frequency-Domain Equations and Harmonic Balance

Expanding the field in harmonic components:

$$\mathbf{E}(\mathbf{r}, t) = \sum_m \mathbf{E}_m(\mathbf{r}) e^{-im\omega t} + \text{c.c.}, \quad (13.5)$$

substituting into the nonlinear Maxwell equations and collecting terms at the same frequency yields the **harmonic balance equations**:

$$\nabla \times (\nabla \times \mathbf{E}_m) - \frac{(m\omega)^2}{c^2} \varepsilon(\mathbf{r}) \mathbf{E}_m = \frac{(m\omega)^2}{\varepsilon_0 c^2} \mathbf{P}_m^{\text{NL}}[\{\mathbf{E}_n\}], \quad (13.6)$$

where $\mathbf{P}_m^{\text{NL}}$ is the nonlinear coupling term involving all harmonic components. This is a system of **nonlinearly coupled partial differential equations** that no longer has the structure of a linear eigenvalue problem.

For Kerr nonlinearity, the self-phase modulation term at the fundamental frequency ($m = 1$) is

$$\mathbf{P}_1^{\text{NL}} \approx \varepsilon_0 \chi^{(3)}(\mathbf{r}) (2|\mathbf{E}_1|^2 + |\mathbf{E}_{-1}|^2) \mathbf{E}_1 = 3\varepsilon_0 \chi^{(3)} |\mathbf{E}_1|^2 \mathbf{E}_1, \quad (13.7)$$

where the last step uses $\mathbf{E}_{-1} = \mathbf{E}_1^*$.

In anisotropic media, the nonlinear susceptibility should be treated as a tensor, but the scalar Kerr form above is sufficient for the topological mechanisms emphasized in this chapter.

13.2 Weak Nonlinear Expansion and Perturbative Corrections to Topological Invariants

13.2.1 Slowly Varying Envelope Approximation

When nonlinear effects are weak relative to the linear bandgap, the **slowly varying envelope approximation** (SVEA) can be employed. Let the field be a slowly modulated linear Bloch mode:

$$\mathbf{E}(\mathbf{r}, t) = \sum_n A_n(\mathbf{R}, T) \mathbf{u}_{n\mathbf{k}_0}(\mathbf{r}) e^{i(\mathbf{k}_0 \cdot \mathbf{r} - \omega_n t)} + \text{c.c.}, \quad (13.8)$$

where $A_n(\mathbf{R}, T)$ is the slowly varying envelope, and $\mathbf{R} = \epsilon \mathbf{r}$, $T = \epsilon t$ are slow variables ($\epsilon \ll 1$ characterizes the nonlinear strength).

Substituting this ansatz into the nonlinear Maxwell equations and collecting terms at each order of ϵ , the linear Bloch equation is recovered at leading order. At the next order, the solvability condition gives the **topological generalization of the nonlinear Schrödinger equation** (NLSE):

$$i \frac{\partial A_n}{\partial T} + \sum_m D_{nm} (-i \nabla_{\mathbf{R}}) A_m + \gamma_n |A_n|^2 A_n = 0, \quad (13.9)$$

where D_{nm} is the dispersion matrix (determined by the linear band structure) and γ_n is the effective nonlinear coefficient:

$$\gamma_n = \frac{3\omega_n}{4} \frac{\int_{\text{cell}} \chi^{(3)}(\mathbf{r}) |\mathbf{u}_{n\mathbf{k}_0}(\mathbf{r})|^4 d^3r}{\int_{\text{cell}} \varepsilon(\mathbf{r}) |\mathbf{u}_{n\mathbf{k}_0}(\mathbf{r})|^2 d^3r}. \quad (13.10)$$

The detailed form of D_{nm} is determined by the local band dispersion and inter-band couplings. For the present chapter, its role is to connect the linear topological band structure to the nonlinear envelope dynamics.

13.2.2 Nonlinear Corrections to Berry Curvature

Nonlinear effects influence Berry geometry by modifying the effective Hamiltonian. In the weakly nonlinear limit, the effective Bloch Hamiltonian acquires an intensity-dependent correction:

$$\mathcal{H}_{\text{eff}}(\mathbf{k}, |A|^2) = \mathcal{H}_0(\mathbf{k}) + \gamma |A|^2 \mathcal{V}(\mathbf{k}), \quad (13.11)$$

where $\mathcal{H}_0(\mathbf{k})$ is the linear Hamiltonian and $\mathcal{V}(\mathbf{k})$ is the nonlinear correction matrix.

For a fixed-amplitude perturbative branch, the Berry curvature correspondingly acquires a correction:

$$\Omega_n(\mathbf{k}, |A|^2) = \Omega_n^{(0)}(\mathbf{k}) + \gamma |A|^2 \Omega_n^{(1)}(\mathbf{k}) + O(\gamma^2), \quad (13.12)$$

where $\Omega_n^{(0)}$ is the linear Berry curvature and $\Omega_n^{(1)}$ is the first-order nonlinear correction.

The explicit expression for the first-order correction $\Omega_n^{(1)}$ can be obtained from standard perturbation theory:

$$\Omega_n^{(1)}(\mathbf{k}) = -2 \text{Im} \sum_{m \neq n} \frac{\langle \partial_{k_x} u_n | u_m \rangle \langle u_m | \mathcal{V} | \partial_{k_y} u_n \rangle - (x \leftrightarrow y)}{(\omega_n - \omega_m)^2}. \quad (13.13)$$

Theorem 13.1 (Conditional stability of the Chern number under weak nonlinearity). *Let the n -th linear band have a nonzero bandgap Δ from adjacent bands. Assume that, at fixed normalization or power, there exists a single-valued smooth branch of nonlinear Bloch waves over the whole BZ and that the relevant spectral gap separating this branch from neighboring branches remains open. If the nonlinear strength satisfies $\gamma|A|^2 \ll \Delta$, then the Chern number assigned to this branch,*

$$C_n(|A|^2) = \frac{1}{2\pi} \int_{BZ} \Omega_n(\mathbf{k}, |A|^2) d^2k \quad (13.14)$$

remains an integer and equals the linear value, $C_n(|A|^2) = C_n^{(0)}$, as long as this continuous branch and its gap persist.

Proof. Under the assumptions above, the nonlinear branch can be represented perturbatively by a continuous family of effective linearized eigenproblems. Define

$$\mathcal{H}(\lambda) = \mathcal{H}_0 + \lambda \mathcal{V}, \quad \lambda = \gamma|A|^2. \quad (13.15)$$

The bandgap between the n -th and $(n+1)$ -th bands is

$$\Delta(\lambda) = \min_{\mathbf{k}} [\omega_{n+1}(\mathbf{k}, \lambda) - \omega_n(\mathbf{k}, \lambda)]. \quad (13.16)$$

By perturbation theory, $|\Delta(\lambda) - \Delta(0)| \leq \lambda \|\mathcal{V}\|_{\text{op}}$, where $\|\mathcal{V}\|_{\text{op}}$ is the operator norm of $\mathcal{V}$. Therefore when $\lambda < \Delta(0)/\|\mathcal{V}\|_{\text{op}}$, $\Delta(\lambda) > 0$.

In the interval where $\Delta(\lambda) > 0$, $C_n(\lambda)$ is a continuous integer-valued function of λ . Since the continuity of an integer-valued function implies constancy, $C_n(\lambda) = C_n(0) = C_n^{(0)}$. $\square$

The physical implication of Theorem 13.1 is that weak nonlinearity cannot by itself change the integer invariant of a well-defined, gapped nonlinear branch. It may shift the edge-state frequency and modify the group velocity, but the branch can be continuously connected to the linear topological state as long as the relevant gap remains open. This statement does **not** guarantee dynamical stability: modulational or parametric instabilities must be checked through the Bogoliubov–de Gennes analysis in §13.4.

13.3 Self-Consistent Theory of Nonlinear Bloch Waves

13.3.1 Definition of the Nonlinear Eigenvalue Problem

Going beyond the weak nonlinear approximation, one can directly seek **nonlinear Bloch wave** solutions of the nonlinear Maxwell equations—nonlinear steady-state solutions with Bloch periodicity:

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_{\mathbf{k}}(\mathbf{r}) e^{-i\omega t}, \quad \mathbf{E}_{\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k} \cdot \mathbf{R}} \mathbf{E}_{\mathbf{k}}(\mathbf{r}). \quad (13.17)$$

Substituting into the Kerr-nonlinear Maxwell equations yields the **nonlinear eigenvalue problem**:

$$\nabla \times (\nabla \times \mathbf{E}_{\mathbf{k}}) = \frac{\omega^2}{c^2} [\varepsilon(\mathbf{r}) + \chi^{(3)}(\mathbf{r}) |\mathbf{E}_{\mathbf{k}}|^2] \mathbf{E}_{\mathbf{k}}. \quad (13.18)$$

This is no longer a standard eigenvalue problem, because the “effective dielectric constant” $\varepsilon_{\text{eff}} = \varepsilon + \chi^{(3)}|\mathbf{E}|^2$ depends on the eigenfunction itself.

In practice, nonlinear Bloch waves are obtained by continuation from a linear Bloch mode or by self-consistent iteration at fixed power. The numerical details are secondary here; the important point is that a nonlinear band is meaningful only after a particular smooth branch and normalization convention have been selected.

13.3.2 Power-Dependent Band Structure

The frequency $\omega(\mathbf{k}, P)$ of nonlinear Bloch waves depends not only on the wave vector $\mathbf{k}$ but also on the power $P \propto \int_{\text{cell}} |\mathbf{E}_{\mathbf{k}}|^2 d^3r$. This leads to a **power-dependent band structure**—nonlinear bands.

For self-focusing nonlinearity ($\chi^{(3)} > 0$), the effective refractive index increases with power, and the bands shift downward in frequency. The first-order approximation for the frequency shift is

$$\delta\omega_n(\mathbf{k}) \approx -\frac{\omega_n}{2} \frac{\int_{\text{cell}} \chi^{(3)} |\mathbf{u}_{n\mathbf{k}}|^4 d^3r}{\int_{\text{cell}} \varepsilon |\mathbf{u}_{n\mathbf{k}}|^2 d^3r} \cdot P. \quad (13.19)$$

When the shift becomes comparable to the bandgap, the gap may close—this is the precursor of a nonlinearity-induced topological phase transition.

If several self-consistent solutions coexist at the same $\mathbf{k}$ and P , the nonlinear band becomes multivalued. Such bistability is useful for switching, but it also shows why nonlinear topology must always specify the branch whose invariant is being discussed.

13.3.3 Nonlinear Berry Phase

For nonlinear Bloch waves, one can formally define a Berry connection:

$$\mathcal{A}_i^{\text{NL}}(\mathbf{k}) = -i \frac{\langle \mathbf{E}_{\mathbf{k}} | \partial_{k_i} | \mathbf{E}_{\mathbf{k}} \rangle_{\varepsilon}}{\langle \mathbf{E}_{\mathbf{k}} | \mathbf{E}_{\mathbf{k}} \rangle_{\varepsilon}}, \quad (13.20)$$

where the inner product $\langle \cdot | \cdot \rangle_{\varepsilon}$ is the material-weighted inner product and the sign follows Chapter 4. However, since the nonlinear eigenvalue problem is state-dependent rather than a fixed Hermitian eigenproblem, the definitions of Berry curvature and Chern number require additional care.

Remark (Gauge invariance of the nonlinear Berry phase). The gauge transformation $|\mathbf{E}_{\mathbf{k}}\rangle \rightarrow e^{i\phi(\mathbf{k})} |\mathbf{E}_{\mathbf{k}}\rangle$ still preserves the Berry curvature, but integer quantization requires additional conditions—namely, a fixed normalization convention, a globally continuous nonlinear Bloch branch over the entire BZ, and an open gap to neighboring branches or to the relevant linearized spectrum.

13.4 Mathematical Structure of Nonlinear Topological Theory

13.4.1 Energy Functional of the Nonlinear System

The nonlinear Maxwell equations can be derived from a variational principle applied to an energy functional. For a Kerr-nonlinear system, the energy functional is

$$\mathcal{E}[\mathbf{E}] = \frac{1}{2} \int \left[|\nabla \times \mathbf{E}|^2 - \frac{\omega^2}{c^2} \varepsilon |\mathbf{E}|^2 - \frac{\omega^2}{2c^2} \chi^{(3)} |\mathbf{E}|^4 \right] d^3r. \quad (13.21)$$

Nonlinear Bloch waves are critical points of $\mathcal{E}$ under the constraint of Bloch periodicity: $\delta\mathcal{E}/\delta\mathbf{E}^* = 0$.

The Hessian matrix $\delta^2\mathcal{E}/\delta\mathbf{E}^*\delta\mathbf{E}$ of the energy functional is precisely the linearized operator $\hat{L}$; its positive-definiteness determines whether the critical point is a minimum (stable) or a saddle point (unstable).

13.4.2 Bogoliubov–de Gennes Stability Analysis

The stability of a nonlinear Bloch wave $\mathbf{E}_0$ is determined by the spectrum of its linearized operator. Let the perturbation be $\delta\mathbf{E} = \mathbf{u}e^{\lambda t} + \mathbf{v}^*e^{\lambda^*t}$; substituting into the nonlinear Maxwell equations and linearizing gives the **Bogoliubov–de Gennes** (BdG) equation:

$$\begin{pmatrix} \hat{L} & \hat{M} \\ -\hat{M}^* & -\hat{L}^* \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}, \quad (13.22)$$

where

$$\hat{L} = \nabla \times (\nabla \times) - \frac{\omega^2}{c^2} [\varepsilon + 2\chi^{(3)}|\mathbf{E}_0|^2], \quad (13.23)$$

$$\hat{M} = -\frac{\omega^2}{c^2} \chi^{(3)} \mathbf{E}_0^2. \quad (13.24)$$

13.4.3 Symmetry of the BdG Equation

The BdG equation possesses **particle–hole symmetry**: if $(\mathbf{u}, \mathbf{v})^T$ is an eigenstate with eigenvalue λ , then $(\mathbf{v}^*, \mathbf{u}^*)^T$ is an eigenstate with eigenvalue $-\lambda^*$. This means the BdG spectrum is symmetric about the imaginary axis, and instabilities ($\text{Re}(\lambda) > 0$) always appear in pairs.

Stability criterion:

- All λ have $\text{Re}(\lambda) \leq 0$: dynamically stable.
- There exists $\text{Re}(\lambda) > 0$: unstable (modulational instability or parametric instability).

13.4.4 Topological Classification of Nonlinear Systems

The topological classification of nonlinear systems is necessarily conditional. Nonlinear Bloch waves may fail to exist globally, may be multivalued, and cannot be superposed. As a result, topological invariants are well defined only after one specifies a smooth branch, a normalization convention, and a spectral gap in either the self-consistent or linearized problem. In the weak nonlinear limit this reduces to Theorem 13.1; for solitons and strongly nonlinear branches, the relevant topological statement must be supplemented by existence and BdG stability.

13.5 Topological Solitons and Nonlinear Edge States

When a linear topological edge state propagates in a Kerr medium, self-phase modulation can balance edge-state dispersion and form a **topological edge soliton**: a nonlinear wave packet localized both by topology in the transverse direction and by nonlinearity along the propagation direction.

Let the edge-state dispersion be expanded near k_0 as

$$\omega(k_x) \approx \omega_0 + v_g(k_x - k_0) + \frac{1}{2}D(k_x - k_0)^2, \quad (13.25)$$

where v_g is the group velocity and D is the group-velocity dispersion. Under the slowly varying envelope approximation, the edge envelope satisfies

$$i\frac{\partial\psi}{\partial t} + \frac{D}{2}\frac{\partial^2\psi}{\partial x^2} + \gamma_{\text{edge}}|\psi|^2\psi = 0, \quad (13.26)$$

with

$$\gamma_{\text{edge}} = \frac{3\omega_0}{4} \frac{\int \chi^{(3)}(\mathbf{r})|\mathbf{u}_{\text{edge}}(\mathbf{r})|^4 d^2r}{\int \varepsilon(\mathbf{r})|\mathbf{u}_{\text{edge}}(\mathbf{r})|^2 d^2r}. \quad (13.27)$$

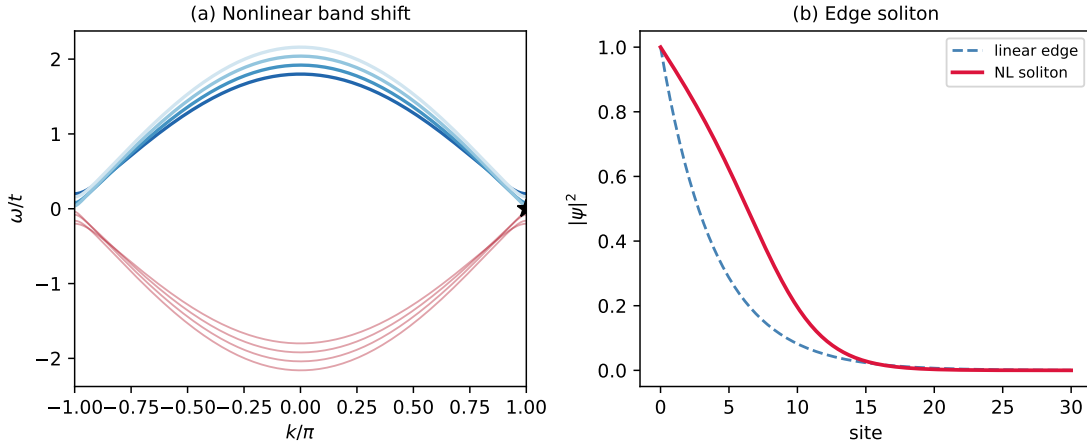


Figure 13.1 Schematic of power-dependent band shift and edge solitons in nonlinear topological photonics.

When $D\gamma_{\text{edge}} > 0$, Eq. (13.26) supports a bright-soliton envelope,

$$\psi(x, t) = \psi_0 \operatorname{sech}\left(\frac{x - v_g t}{L_s}\right) e^{i(\kappa x - \Omega t)}, \quad (13.28)$$

where $L_s = \sqrt{|D|/(\gamma_{\text{edge}}|\psi_0|^2)}$ and $\Omega = \gamma_{\text{edge}}|\psi_0|^2/2$. Other nonlinear excitations may occur in specific models, but the bright edge soliton is the cleanest example of a nonlinear continuation of a topological edge mode.

For a one-dimensional topological photonic model with Kerr nonlinearity,

$$i\frac{\partial\psi}{\partial z} = -\frac{1}{2k_0}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi + g|\psi|^2\psi, \quad (13.29)$$

stationary nonlinear edge states $\psi(x, z) = u(x)e^{i\beta z}$ satisfy

$$-\frac{1}{2k_0}u'' + V(x)u + g|u|^2u = \beta u. \quad (13.30)$$

Theorem 13.2 (Existence of topological edge solitons). *Let $V(x)$ support an isolated topological edge state $\phi_0(x)$ with eigenvalue β_0 lying in a bulk bandgap. For sufficiently weak Kerr nonlinearity, there exists a nonlinear edge-soliton branch $u(x; g)$ such that*

$$u(x; 0) = \phi_0(x), \quad \beta(g) = \beta_0 + g \int |\phi_0|^4 dx + O(g^2), \quad (13.31)$$

and the solution remains exponentially localized near the boundary as long as $\beta(g)$ stays inside the gap.

The theorem is a continuation statement: it follows from the isolated nature of the linear edge eigenvalue and the implicit-function or Lyapunov–Schmidt argument. It does not imply unconditional robustness at arbitrary power.

Proposition 13.3 (Continuation of topological protection under weak nonlinearity). *In the weakly nonlinear limit, $|g||u|^2 \ll \Delta_{\text{gap}}$, a topological edge soliton can inherit the gap-protected character of its parent linear edge state, provided the nonlinear branch remains smooth, separated from bulk modes, and dynamically stable.*

Strong nonlinearity can invalidate this continuation in three main ways: the nonlinear frequency shift can push the branch into a bulk band, modulational instability can destabilize the soliton, or the self-consistent spectrum can undergo a nonlinearity-induced topological transition. Full stability therefore requires the BdG spectrum rather than only the existence of a stationary solution.

13.6 Nonlinearity-Induced Topological Phase Transitions

13.6.1 Power-Driven Bandgap Closure and Chern Number Jump

A more subtle phenomenon is that nonlinear effects can **change the topology of a self-consistent branch** of the system. Consider a mean-field model that is topologically trivial in the linear regime, with effective Hamiltonian

$$\mathcal{H}(\mathbf{k}, P) = \mathcal{H}_0(\mathbf{k}) + P\mathcal{V}(\mathbf{k}), \quad (13.32)$$

when the power P increases to a certain critical value P_c , the self-consistent gap can close and reopen, and the branch Chern number can jump:

$$C_n(P < P_c) = 0 \quad \longrightarrow \quad C_n(P > P_c) = \pm 1. \quad (13.33)$$

This is often called a **nonlinearity-induced topological phase transition**. The statement is model-dependent: the transition requires a well-defined continuous nonlinear Bloch branch and a genuine gap closing in the corresponding linearized or self-consistent spectrum.

Estimate of the critical power P_c :

$$P_c \sim \frac{\Delta_0}{\gamma_{\text{eff}}} = \frac{\Delta_0 V_{\text{mode}}}{3\omega\chi^{(3)}}, \quad (13.34)$$

where Δ_0 is the linear bandgap and V_{mode} is the mode volume. The numerical value is highly platform-dependent because the relevant quantity is the intracavity intensity and field enhancement, not simply the launched optical power.

13.7 Extension Directions and Unified Picture

The Kerr model captures the central mechanism of nonlinear topological photonics, but several extensions are important. In $\chi^{(2)}$ media, the fundamental and second-harmonic fields can occupy different topological bands, so topological phase matching must track the compatibility of two frequency subspaces. In Floquet systems, nonlinearity can shift quasi-energy bands and may close either the 0-gap or the π -gap, producing intensity-dependent Floquet topology. In topological lasers, gain saturation is an effective nonlinearity, and edge-mode lasing depends on the competition between topological localization, gain overlap, radiative loss, and nonlinear saturation.

These directions are best viewed as applications of the same hierarchy:

1. Under weak nonlinearity, a smooth gapped branch can keep the invariant of its linear parent.
2. Nonlinearity changes dispersion and stability before it necessarily changes topology.
3. Strong nonlinearity can close the relevant gap or destabilize the branch, at which point a self-consistent and BdG-stability-aware analysis is required.

Thus nonlinear topological photonics should not be summarized by a single invariant. The physically meaningful question is always: which nonlinear branch is being followed, is the relevant gap open, and is that branch dynamically stable?

Chapter Summary

Starting from the nonlinear Maxwell equations, this chapter systematically constructed the theoretical framework of nonlinear topological photonics. In the weakly nonlinear limit, the topological generalization of the nonlinear Schrödinger equation was derived via the slowly varying envelope approximation, and the conditional stability of the Chern number under weak nonlinearity was established for smooth gapped nonlinear branches (Theorem 13.1). Going beyond perturbation theory, the self-consistent theory of nonlinear Bloch waves and the power-dependent band structure were introduced. The formation mechanism of topological edge solitons, the competition between topological protection and nonlinear instability, and possible nonlinearity-induced topological transitions were discussed. Finally, the general framework of nonlinear stability analysis was established through the BdG equation.

Core conclusion: Weak nonlinearity does not change the topological invariant of a well-defined gapped branch, but it can modify the dynamical properties of edge states (forming solitons, inducing instabilities). Strong nonlinearity can close the relevant gap and may induce

a self-consistent topological transition; whether this realizes robust switching must be checked together with branch existence and dynamical stability.

Chapter A

Differential Forms and Electromagnetism

Why Differential Forms Are Worth Using Here?

The main text (Chapters 2–4) uses vector analysis and Dirac notation as the primary language, which is sufficient for concrete calculations. However, at the following three levels, differential forms provide irreplaceable insight:

Level 1: Unification. The gradient, curl, and divergence—seemingly different operations in vector analysis—are unified in the language of differential forms into a single operator: the exterior derivative d . The concise identity $d^2 = 0$ unifies $\nabla \times (\nabla f) = 0$ and $\nabla \cdot (\nabla \times \mathbf{A}) = 0$. The homogeneous Maxwell equation $dF = 0$ expresses two vector equations ($\nabla \cdot \mathbf{B} = 0$ and Faraday’s law) in a single symbol, and locally implies the existence of electromagnetic potentials by the Poincaré lemma; global potentials require the relevant de Rham cohomology class to vanish.

Level 2: Geometric essence. In the vector-analysis framework, the Berry connection is merely a “vector field” $\mathcal{A}_\mu(k) = -i\langle u | \partial_{k^\mu} u \rangle$, whose gauge transformation rules and curvature computation easily lead to sign confusion in high-dimensional parameter spaces. In the fiber-bundle language, the Berry connection is a connection 1-form on a $U(1)$ principal bundle, the Berry curvature is a curvature 2-form, and the Chern number is a characteristic class—each step has a precise geometric meaning. This not only makes the notation more concise, but more importantly exposes the *origin* of topological quantization ($C \in \mathbb{Z}$): it comes from the compactness of the structure group $U(1)$ and the classification theory of bundles, not from any particular computational technique.

Level 3: Generalization power. When generalizing from the Abelian (single-band) to the non-Abelian (multi-band) case, the differential-forms framework only requires replacing $U(1)$ by $U(N)$ and supplementing the exterior derivative with a connection self-coupling term such as $i\mathbf{A} \wedge \mathbf{A}$ in the Hermitian Berry-connection convention. This generalization is almost impossible to express naturally in vector analysis. For higher-order topological invariants (such as the nested Wilson loop of Chapter 10), the fiber-bundle language provides the only rigorous mathematical foundation.

Remark (Reading suggestions). This appendix is organized in two progressive levels: §A.1–§A.2 are aimed at readers who need a “translation dictionary”—it suffices to compare the vector-analysis results of the main text with their differential-form counterparts. §A.3–§A.4 are aimed at readers seeking deeper understanding—some knowledge of abstract algebra and manifolds is needed, but we replace abstract arguments with concrete worked examples wherever possible.

This appendix provides a concise formulation of Maxwell’s equations in the language of differential forms. Although the main text uses vector analysis ($\nabla \times$, $\nabla \cdot$) as its primary tool, differential forms provide a more natural mathematical framework for discussing Berry geometry, Chern numbers, and gauge field theory.

A.1 Foundations of Exterior Algebra

A.1.1 Definition of Differential Forms

Definition A.1 (*p*-form). Let M be an n -dimensional smooth manifold. A p -form (differential p -form) on M is a smooth section of the exterior bundle $\bigwedge^p T^*M$. In local coordinates $(x^1, \dots, x^n)$, a p -form can be written as

$$\alpha = \frac{1}{p!} \alpha_{i_1 \dots i_p} dx^{i_1} \wedge \dots \wedge dx^{i_p}, \quad (\text{A.1})$$

where $\alpha_{i_1 \dots i_p}$ are the components of a totally antisymmetric tensor, and $\wedge$ is the wedge product.

Example A.2 (Low-order forms). • 0-form: a smooth function $f \in C^\infty(M)$.

- 1-form: $\alpha = \alpha_i dx^i$ (a cotangent vector field). In three-dimensional space, it corresponds to a vector field via the metric: $\alpha_i = g_{ij} A^j$.
- 2-form: $\beta = \frac{1}{2} \beta_{ij} dx^i \wedge dx^j$. In three-dimensional space, a 2-form corresponds through Hodge duality to a (pseudo-)vector field.
- 3-form: in three-dimensional space, it corresponds to a volume element $\rho dx^1 \wedge dx^2 \wedge dx^3$.

A.1.2 Wedge Product and Exterior Derivative

Definition A.3 (Wedge product). The wedge product of a p -form α and a q -form β is a $(p+q)$ -form $\alpha \wedge \beta$, satisfying:

$$\alpha \wedge \beta = (-1)^{pq} \beta \wedge \alpha. \quad (\text{A.2})$$

Definition A.4 (Exterior derivative). The exterior derivative $d : \Omega^p(M) \rightarrow \Omega^{p+1}(M)$ is defined in local coordinates as

$$d\alpha = \frac{1}{p!} \frac{\partial \alpha_{i_1 \dots i_p}}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p}. \quad (\text{A.3})$$

Theorem A.5 (Basic properties of the exterior derivative). 1. **Nilpotency:** $d^2 = 0$ (i.e., $d(d\alpha) = 0$ for any α).

2. **Leibniz rule:** $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^p \alpha \wedge d\beta$ (α being a p -form).

3. *Correspondence with vector analysis:*

- df corresponds to ∇f (gradient);
- $d\alpha$ (α a 1-form) corresponds to $\nabla \times \mathbf{A}$ (curl);

- $d\beta$ (β a 2-form) corresponds to $\nabla \cdot \mathbf{B}$ (divergence).

Proof. The nilpotency $d^2 = 0$ follows directly from the symmetry of partial derivatives, $\frac{\partial^2}{\partial x^i \partial x^j} = \frac{\partial^2}{\partial x^j \partial x^i}$, combined with the antisymmetry of the wedge product $dx^i \wedge dx^j = -dx^j \wedge dx^i$. In vector analysis, this corresponds to $\nabla \times (\nabla f) = 0$ and $\nabla \cdot (\nabla \times \mathbf{A}) = 0$. $\square$

A.1.3 Hodge Duality

Definition A.6 (Hodge star operator). On an n -dimensional oriented Riemannian manifold (M, g) , the Hodge star operator $\star : \Omega^p(M) \rightarrow \Omega^{n-p}(M)$ is defined by

$$\alpha \wedge \star \beta = \langle \alpha, \beta \rangle_g \text{vol}_g, \quad (\text{A.4})$$

where $\langle \cdot, \cdot \rangle_g$ is the inner product on p -forms induced by the metric g , and $\text{vol}_g = \sqrt{|g|} dx^1 \wedge \cdots \wedge dx^n$ is the volume form.

Example A.7 (Hodge duality in three-dimensional Euclidean space). Under the standard metric on $\mathbb{R}^3$: $\star dx = dy \wedge dz$, $\star dy = dz \wedge dx$, $\star dz = dx \wedge dy$; $\star(dy \wedge dz) = dx$, etc.; $\star 1 = dx \wedge dy \wedge dz$. Therefore the vector field $\mathbf{B} = B_x \hat{\mathbf{x}} + B_y \hat{\mathbf{y}} + B_z \hat{\mathbf{z}}$ corresponds to the 2-form $\mathcal{B} = B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy$, satisfying $\star \mathcal{B} = B_x dx + B_y dy + B_z dz$ (a 1-form).

Example A.8 (Explicit computation of the wedge product and comparison with the vector cross product). In $\mathbb{R}^3$, given two 1-forms

$$\alpha = 3 dx - 2 dy, \quad \beta = dx + 5 dz. \quad (\text{A.5})$$

Goal: compute the wedge product $\alpha \wedge \beta$ and compare with the vector cross product.

Step 1: Expand the wedge product. Using bilinearity and antisymmetry $dx^i \wedge dx^j = -dx^j \wedge dx^i$:

$$\begin{aligned} \alpha \wedge \beta &= (3 dx - 2 dy) \wedge (dx + 5 dz) \\ &= 3 dx \wedge dx + 15 dx \wedge dz - 2 dy \wedge dx - 10 dy \wedge dz. \end{aligned} \quad (\text{A.6})$$

Using $dx \wedge dx = 0$ and $dy \wedge dx = -dx \wedge dy$:

$$\boxed{\alpha \wedge \beta = -10 dy \wedge dz + 2 dx \wedge dy - 15 dx \wedge dz.} \quad (\text{A.7})$$

Rewriting in standard order ($dy \wedge dz, dz \wedge dx, dx \wedge dy$): $\alpha \wedge \beta = -10 dy \wedge dz + 15 dz \wedge dx + 2 dx \wedge dy$.

Step 2: Comparison with the vector cross product. Associating α and β with vectors $\mathbf{a} = (3, -2, 0)$ and $\mathbf{b} = (1, 0, 5)$,

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ 3 & -2 & 0 \\ 1 & 0 & 5 \end{vmatrix} = (-10, -15, 2). \quad (\text{A.8})$$

The three components of the wedge-product 2-form $(-10, 15, 2)$ (note the sign convention for the $dz \wedge dx$ term) precisely correspond to the components of the cross product.

Step 3: Antisymmetry verification. $\beta \wedge \alpha = (-1)^{1 \times 1} \alpha \wedge \beta = -\alpha \wedge \beta = 10 dy \wedge dz - 15 dz \wedge dx - 2 dx \wedge dy$, corresponding to $\mathbf{b} \times \mathbf{a} = -\mathbf{a} \times \mathbf{b}$. $\square$

Remark. The wedge product in $\mathbb{R}^3$ is *formally* identical to the cross product, but the wedge product is defined without reference to a metric or orientation—it is defined in any dimension and on any manifold, whereas the cross product is restricted to $\mathbb{R}^3$ (or $\mathbb{R}^7$). This is the core advantage of the differential-forms language: *coordinate-independent universality*.

Example A.9 (Hodge duality in four-dimensional Minkowski spacetime). In Minkowski spacetime (t, x, y, z) with metric signature $(-, +, +, +)$, the volume form is $\text{vol} = dt \wedge dx \wedge dy \wedge dz$. **Goal:** compute the Hodge dual $\star F$ of the electromagnetic field 2-form F (Eq. (A.16)).

Step 1: Hodge duals of basis 2-forms. For the signature $(-, +, +, +)$, the Hodge dual of a 2-form in 4 dimensions gives another 2-form. Using $\star(dx^\mu \wedge dx^\nu) = \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} g_{\rho\alpha} g_{\sigma\beta} dx^\alpha \wedge dx^\beta$, the key results are:

$$\star(dx \wedge dt) = -dy \wedge dz, \quad \star(dy \wedge dz) = -dx \wedge dt, \quad (\text{A.9})$$

$$\star(dy \wedge dt) = -dz \wedge dx, \quad \star(dz \wedge dx) = -dy \wedge dt, \quad (\text{A.10})$$

$$\star(dz \wedge dt) = -dx \wedge dy, \quad \star(dx \wedge dy) = -dz \wedge dt. \quad (\text{A.11})$$

Step 2: Computing $\star F$. Substituting the above results into F (Eq. (A.16)):

$$\begin{aligned} \star F &= E_x(-dy \wedge dz) + E_y(-dz \wedge dx) + E_z(-dx \wedge dy) \\ &\quad + B_x(-dx \wedge dt) + B_y(-dy \wedge dt) + B_z(-dz \wedge dt). \end{aligned} \quad (\text{A.12})$$

Rearranging:

$$\boxed{\star F = -B_x dx \wedge dt - B_y dy \wedge dt - B_z dz \wedge dt - E_x dy \wedge dz - E_y dz \wedge dx - E_z dx \wedge dy.} \quad (\text{A.13})$$

Comparing with F : $\star F$ maps $\mathbf{E} \rightarrow -\mathbf{B}$ and $\mathbf{B} \rightarrow -\mathbf{E}$. With the convention for F used here, this is the differential-forms expression of the electromagnetic duality map encoded by the Hodge star; changing the sign convention in the definition of F changes the apparent signs of the induced $(\mathbf{E}, \mathbf{B})$ map, but not the invariant statement $\star^2 = -1$ on two-forms in Lorentz signature.

Step 3: Verifying $\star \star F = -F$. Taking the Hodge dual of $\star F$ once more:

$$\begin{aligned} \star(\star F) &= \star[-\mathbf{B} \text{ (in the electric-field position)} - \mathbf{E} \text{ (in the magnetic-field position)}] \\ &= -F. \end{aligned} \quad (\text{A.14})$$

Thus,

$$\star \star F = -F. \quad (\text{A.15})$$

The minus sign comes from the Lorentz signature: $\star \star = (-1)^{p(n-p)+s}$, where $p = 2$, $n = 4$, $s = 1$ (one minus sign from the time dimension), giving $(-1)^{2 \cdot 2 + 1} = (-1)^5 = -1$. In Euclidean signature ($s = 0$), one would get $\star \star F = +F$ —this is a key physical effect of the Lorentz signature.

Remark. $\star \star = -\text{id}$ means that under the Lorentz signature, the eigenvalues of $\star$ are $\pm i$, and the corresponding 2-forms are called **self-dual** ($\star F = iF$) and **anti-self-dual** ($\star F = -iF$). Under Euclidean signature, the self-duality condition $\star F = F$ has real solutions—this is the basis of

Yang–Mills instanton theory, and also connects to certain self-dual structures in two-dimensional models in topological photonics.

A.2 Maxwell's Equations in Exterior Form

A.2.1 Electromagnetic Field as a Differential Form

In four-dimensional spacetime (t, x, y, z) , the electromagnetic field is naturally represented as a 2-form:

Definition A.10 (Electromagnetic field strength 2-form).

$$F = E_x dx \wedge dt + E_y dy \wedge dt + E_z dz \wedge dt + B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy. \quad (\text{A.16})$$

Equivalently, in the $3 + 1$ decomposition:

$$F = \mathcal{E} \wedge dt + \mathcal{B}, \quad (\text{A.17})$$

where $\mathcal{E} = E_i dx^i$ (1-form) and $\mathcal{B} = \frac{1}{2} B_{ij} dx^i \wedge dx^j$ (2-form).

A.2.2 Maxwell's Equations in Differential Forms

Theorem A.11 (Maxwell's equations (exterior-form version)). *In vacuum, or more generally after the constitutive relation has been encoded in a material Hodge operator, Maxwell's equations can be written as two groups:*

1. **Homogeneous equations** (source-free):

$$dF = 0. \quad (\text{A.18})$$

This single equation simultaneously contains $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0$.

2. **Inhomogeneous equations** (with sources):

$$d \star F = J, \quad (\text{A.19})$$

where J is the current 3-form. In vacuum $\star$ is the Hodge dual under the four-dimensional Minkowski metric; in media it is replaced by the appropriate material Hodge operator. This equation contains $\nabla \cdot \mathbf{E} = \rho$ and $\nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{J}$ in vacuum, or equivalently the usual $\nabla \cdot \mathbf{D} = \rho$ and $\nabla \times \mathbf{H} - \partial_t \mathbf{D} = \mathbf{J}$ after the constitutive map is applied.

Proof. Expanding dF according to Eq. (A.16):

$$dF = \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} + \frac{\partial B_z}{\partial t} \right) dx \wedge dy \wedge dt + \cdots + \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} \right) dx \wedge dy \wedge dz. \quad (\text{A.20})$$

$dF = 0$ requires each independent component to vanish: the $dx \wedge dy \wedge dz$ component gives $\nabla \cdot \mathbf{B} = 0$, and the $dx^i \wedge dx^j \wedge dt$ components give $(\nabla \times \mathbf{E})_k = -\partial_t B_k$. The proof of the inhomogeneous equation is similar, using $\star$ to map F to a dual 2-form before taking the exterior derivative. $\square$

Remark. Advantages of the differential-forms formulation:

1. $dF = 0$ automatically implies the existence of potentials—by the Poincaré lemma, $dF = 0 \Rightarrow F = dA$ (locally), where, with the sign convention of eq. (A.16), one may take $A = -\phi dt + A_i dx^i$ as the electromagnetic potential 1-form.
2. The gauge transformation $A \rightarrow A + d\chi$ (χ a 0-form) leaves $F = dA$ invariant, because $d(A + d\chi) = dA + d^2\chi = dA$ (using $d^2 = 0$).
3. The equation $dF = 0$ does not depend on the metric—it is *topological*. The equation $d \star F = J$ depends on the metric (through $\star$)—it contains the *dynamics* (constitutive relations).

A.2.3 Constitutive Relations

In linear media, the constitutive relations can be expressed as generalizations of the $\star$ operator:

Definition A.12 (Material Hodge operator).

$$\star_\varepsilon \mathcal{E} = \varepsilon \star \mathcal{E}, \quad \star_\mu^{-1} \mathcal{B} = \mu^{-1} \star \mathcal{B}. \quad (\text{A.21})$$

For non-isotropic materials, ε and μ are replaced by tensors, and the structure of the material Hodge operator is richer than that of the geometric Hodge operator.

A.3 Connections and Curvature in Fiber-Bundle Language

A.3.1 Principal Bundles and Connections

Definition A.13 ($U(1)$ principal bundle). Let M be the base space (e.g., the Brillouin zone T^2), and $P \rightarrow M$ a $U(1)$ principal bundle. Sections of P correspond to smooth extensions of Bloch wave functions over the entire M (if they exist).

Definition A.14 (Connection 1-form). A connection on P is a $\mathfrak{u}(1)$ -valued 1-form $\mathcal{A}$, which in a local section (gauge choice) is represented as

$$\mathcal{A}_n = i A_{n,\mu} dk^\mu, \quad A_{n,\mu} = -i \langle u_n | \partial_{k^\mu} u_n \rangle. \quad (\text{A.22})$$

Here $\mathcal{A}_n$ is the anti-Hermitian mathematical connection, while $A_n = A_{n,\mu} dk^\mu$ is the real Berry connection convention used in Chapter 4.

Definition A.15 (Curvature 2-form). The curvature 2-form of the connection is

$$\mathcal{F}_n = d\mathcal{A}_n = i F_n, \quad F_n = dA_n, \quad (\text{A.23})$$

where $F_n = \frac{1}{2} F_{n,\mu\nu} dk^\mu \wedge dk^\nu$ is the Berry curvature 2-form (in the Abelian case). $d\mathcal{F}_n = d^2 \mathcal{A}_n = 0$ is automatically satisfied (Bianchi identity).

A.3.2 The Chern Number as a Characteristic Class

Theorem A.16 (First Chern class and Chern number). *The first Chern class of a $U(1)$ bundle $P \rightarrow M$ (M a 2-dimensional compact manifold) is*

$$c_1(P) = \left[\frac{\mathcal{F}}{2\pi i} \right] = \left[\frac{F}{2\pi} \right] \in H^2(M, \mathbb{Z}), \quad (\text{A.24})$$

and its integral gives the first Chern number:

$$C_1 = \int_M c_1 = \frac{1}{2\pi} \int_M F. \quad (\text{A.25})$$

The Chern number $C_1 \in \mathbb{Z}$ is a topological invariant of the bundle, independent of the choice of connection (i.e., independent of gauge choice).

Proof sketch. The de Rham cohomology class of c_1 does not depend on the choice of connection: if $\mathcal{A}' = \mathcal{A} + i d\chi$, then $\mathcal{F}' = d\mathcal{A}' = d\mathcal{A} + i d^2\chi = \mathcal{F}$. The integrality comes from the compactness of the structure group $U(1)$ —the classification of bundles is given by $[M, BU(1)] \cong H^2(M, \mathbb{Z})$. For $M = T^2$, a constructive proof using a standard partition and Stokes' theorem can be given (see Chapter 4). $\square$

A.3.3 Non-Abelian Generalization

For the multi-band case (Chapter 4, §4.6), the structure group of the bundle is generalized from $U(1)$ to $U(N)$:

Definition A.17 (Non-Abelian connection and curvature).

$$\mathbf{A}_\mu = -i \langle \mathbf{u} | \partial_{k^\mu} | \mathbf{u} \rangle \quad (N \times N \text{ Hermitian matrix}), \quad (\text{A.26})$$

$$\mathbf{F} = d\mathbf{A} + i \mathbf{A} \wedge \mathbf{A}. \quad (\text{A.27})$$

Note that in the non-Abelian case $\mathbf{F} \neq d\mathbf{A}$ —the wedge-product term (corresponding to the self-interaction of the gauge field) is nonzero.

Definition A.18 (Non-Abelian Chern number).

$$C_1 = \frac{1}{2\pi} \int_M \text{tr } \mathbf{F}. \quad (\text{A.28})$$

A.4 Stokes' Theorem and Topological Quantization

Theorem A.19 (Generalized Stokes' theorem). *Let M be an n -dimensional oriented compact manifold with boundary, ∂M its boundary, and ω an $(n-1)$ -form; then*

$$\int_M d\omega = \oint_{\partial M} \omega. \quad (\text{A.29})$$

A.4.1 First Chern Number on a Magnetic Monopole Bundle: Step-by-Step Computation

The following example grounds the abstract definitions of connection, curvature, and Chern class on a concrete and physically important model—the Dirac magnetic monopole. This computation is the prototype for understanding the topological quantization of the Berry phase.

Example A.20 (First Chern number of a charge- n monopole bundle on S^2). Consider a magnetic monopole placed at the origin, whose magnetic field satisfies $\nabla \cdot \mathbf{B} = 4\pi n \delta^3(\mathbf{r})$. Define a $U(1)$ principal bundle $P_n \rightarrow S^2$, whose transition function on the equator $\theta = \pi/2$ is

$$g_{NS}(\phi) = e^{in\phi}, \quad n \in \mathbb{Z}. \quad (\text{A.30})$$

Goal: prove that the first Chern number is $C_1 = n$.

Step 1: Local connections on the northern and southern hemispheres.

On the northern patch $U_N = \{(\theta, \phi) \mid 0 \leq \theta < \pi\}$, choose a section that is non-singular at the north pole; the corresponding real $U(1)$ connection 1-form is

$$A_N = \frac{n}{2}(1 - \cos \theta) d\phi. \quad (\text{A.31})$$

On the southern patch $U_S = \{(\theta, \phi) \mid 0 < \theta \leq \pi\}$,

$$A_S = -\frac{n}{2}(1 + \cos \theta) d\phi. \quad (\text{A.32})$$

Note that A_N is non-singular at $\theta = 0$ but singular at the south pole, while A_S is non-singular at $\theta = \pi$ but singular at the north pole.

Step 2: Gauge transformation verification of the equatorial transition function.

On the equator $\theta = \pi/2$, the difference between the two connections should be given by the gauge transformation of the transition function:

$$A_N - A_S = \frac{n}{2}(1 - \cos \frac{\pi}{2}) d\phi + \frac{n}{2}(1 + \cos \frac{\pi}{2}) d\phi = n d\phi = -i g_{NS}^{-1} dg_{NS}. \quad (\text{A.33})$$

Here $g_{NS} = e^{in\phi}$ and we use the real physicist's connection convention. In the anti-Hermitian mathematical convention one instead writes $\mathcal{A} = iA$, so the transition law contains $g^{-1}dg$ without the prefactor $-i$.

Step 3: Curvature 2-form.

Taking the exterior derivative:

$$F = dA_N = \frac{n}{2} \sin \theta d\theta \wedge d\phi. \quad (\text{A.34})$$

This is exactly the area 2-form on the sphere multiplied by $n/2$. Note that F is globally defined on all of S^2 (even though A is not), because $dA_N = dA_S$ (the difference of the two connections $n d\phi$ is a closed form: $d(n d\phi) = 0$).

Step 4: Explicit integration for the Chern number.

$$C_1 = \frac{1}{2\pi} \int_{S^2} F = \frac{1}{2\pi} \cdot \frac{n}{2} \int_0^\pi \int_0^{2\pi} \sin \theta \, d\theta \, d\phi = \frac{1}{2\pi} \cdot \frac{n}{2} \cdot 4\pi = n. \quad (\text{A.35})$$

Step 5 (equivalent method): Stokes' theorem and equatorial integration.

Cut S^2 along the equator into upper and lower hemispheres D_N and D_S , with common boundary the equator S^1 :

$$\begin{aligned} \int_{S^2} F &= \int_{D_N} dA_N + \int_{D_S} dA_S \\ &= \oint_{S^1} A_N^{(\text{from } D_N)} + \oint_{S^1} A_S^{(\text{from } D_S)} \\ &= \oint_{S^1} (A_N - A_S) = \oint_0^{2\pi} n \, d\phi = 2\pi n, \end{aligned} \quad (\text{A.36})$$

where the orientation convention is that D_N and D_S take opposite orientations on the equator. Therefore $C_1 = 2\pi n / (2\pi) = n$. ✓

Remark (Dirac quantization condition). In physics, $C_1 = n \in \mathbb{Z}$ is equivalent to Dirac's magnetic charge quantization condition $eg = n\hbar c/2$ (Gaussian units). This integrality comes from the compactness of the structure group $U(1)$ —the equatorial transition function $g_{NS} : S^1 \rightarrow U(1)$ must have integer mapping degree, which is precisely the topological constraint $\pi_1(U(1)) = \mathbb{Z}$. In Berry geometry, replacing S^2 by the Brillouin zone T^2 and the connection by the Berry connection $\mathcal{A}_n(\mathbf{k})$, C_1 becomes the Chern number of the n -th band—the fundamental reason for its integrality is exactly the same as for the magnetic monopole.

This theorem unifies:

- The fundamental theorem of calculus ($n = 1$);
- Green's theorem, Stokes' curl theorem ($n = 2$);
- Gauss's divergence theorem ($n = 3$).

Application in Berry geometry: the proof of integrality of the Chern number (Chapter 4, Theorem 4.10) is essentially an application of Stokes' theorem on T^2 —cutting the Brillouin zone into a rectangle, the difference of Berry connection integrals on the boundary gives an integer multiple of 2π .

A.5 Further Applications of Differential Forms in Electromagnetism

This section demonstrates the powerful expressive capability of the differential-forms language through concrete electromagnetic problems, and establishes connections with Berry geometry.

A.5.1 Maxwell's Equations in Differential-Form Notation

Maxwell's equations have an extremely concise expression in the language of differential forms. Define the electromagnetic field 2-form F and its dual 2-form $\star F$:

$$F = E_x dx \wedge dt + E_y dy \wedge dt + E_z dz \wedge dt \\ + B_x dy \wedge dz + B_y dz \wedge dx + B_z dx \wedge dy, \quad (\text{A.37})$$

$$\star F = -B_x dx \wedge dt - B_y dy \wedge dt - B_z dz \wedge dt \\ - E_x dy \wedge dz - E_y dz \wedge dx - E_z dx \wedge dy. \quad (\text{A.38})$$

Maxwell's equations reduce to two extremely concise equations:

$$dF = 0, \quad d(\star F) = J, \quad (\text{A.39})$$

where J is the current 3-form. The first equation contains $\nabla \cdot \mathbf{B} = 0$ and Faraday's law $\nabla \times \mathbf{E} = -\partial_t \mathbf{B}$; the second equation contains Gauss's law $\nabla \cdot \mathbf{D} = \rho$ and Ampère's law $\nabla \times \mathbf{H} = \mathbf{J} + \partial_t \mathbf{D}$.

Topological significance: $dF = 0$ means that F is a closed form. By the Poincaré lemma, locally $F = dA$; globally this requires the cohomology class $[F] \in H^2(M)$ to vanish, where A is the electromagnetic potential 1-form $A = A_\mu dx^\mu$. The gauge transformation $A \rightarrow A + d\chi$ does not change F —this is precisely the geometric statement of $U(1)$ gauge symmetry.

A.5.2 Hodge Duality and Maxwell's Equations in Media

In anisotropic media, the $\star$ operator is modified by the constitutive relations of the medium. Define the generalized Hodge dual $\star_\varepsilon$:

$$(\star_\varepsilon F)_{ij} = \frac{1}{2} \varepsilon_{ijkl} \varepsilon^{km} F_{ml}, \quad (\text{A.40})$$

where ε^{km} are the components of the permittivity tensor.

For magneto-optical media ($\varepsilon_{xy} = -\varepsilon_{yx} = i\varepsilon_g$), the generalized Hodge dual introduces a nontrivial geometric structure—it can give photonic bands nonzero Berry curvature and, when the relevant bandgap and symmetry conditions are met, lead to topologically nontrivial photonic band structures.

This establishes a profound connection: **material response** $\leftrightarrow$ **modified Hodge structure** $\leftrightarrow$ **Berry geometry** $\leftrightarrow$ **possible topological invariants**.

A.5.3 Chern–Simons Forms and Topological Invariants

The Berry connection $A = -i\langle u_n | du_n \rangle$ is a 1-form on parameter space. The Berry curvature $F = dA$ is the corresponding 2-form. The first Chern number is

$$C_1 = \frac{1}{2\pi} \int_{\text{BZ}} F = \frac{1}{2\pi} \int_{\text{BZ}} dA. \quad (\text{A.41})$$

For higher-dimensional parameter spaces, one can define the Chern–Simons 3-form:

$$\omega_3 = \text{Tr} \left(\mathbf{A} \wedge d\mathbf{A} + \frac{2i}{3} \mathbf{A} \wedge \mathbf{A} \wedge \mathbf{A} \right), \quad (\text{A.42})$$

where $\mathbf{A}$ is the non-Abelian Hermitian Berry connection matrix. In the anti-Hermitian mathematical convention the same formula is written without the explicit factors of i .

The Chern–Simons form plays a key role in three-dimensional topological photonics:

- With a fixed normalization convention, the magnetoelectric polarizability θ (axion angle) can be written as a Chern–Simons invariant, schematically $\theta \propto \int_{\text{BZ}} \omega_3$ modulo 2π .
- When a gapped system has symmetry quantizing θ and $\theta = \pi \pmod{2\pi}$, it is in a three-dimensional topological-insulator-type phase.
- The quantization of θ is protected by time-reversal or spatial-inversion symmetry.

A.5.4 Fiber-Bundle Theory and Photonic Bands

The geometric structure of photonic bands can be precisely described by fiber bundles. The Brillouin zone BZ is the base space, and the Bloch states $|u_{n\mathbf{k}}\rangle$ at each $\mathbf{k}$ point constitute the fiber.

Trivial vs. nontrivial bundles: if the fiber bundle of an isolated band is trivial, then there exists a globally smooth gauge choice $|u_{n\mathbf{k}}\rangle$. If the fiber bundle is nontrivial ($C_n \neq 0$), then no globally smooth gauge exists—the gauge must have singularities at certain points in the Brillouin zone.

This mathematical fact has direct physical consequences: a Chern-nontrivial band cannot have Wannier functions that are simultaneously exponentially localized, periodic-gauge compatible, and form an orthonormal complete basis for that isolated band—this is the geometric origin of the Wannier obstruction for Chern bands. Other symmetry-protected obstructions, such as electronic $\mathbb{Z}_2$ obstructions, require their own symmetry constraints.

Chapter B

Fundamentals of Point Groups and Space Groups

Why Symmetry Must Be Made Explicit?

Throughout the main text, assertions such as “ C_{6v} symmetry requires a twofold degeneracy at the Γ point,” “the two branches of the E representation split into $A' + A''$ along the Γ – M line,” and “symmetry-protected BIC” appear repeatedly. The rigorous foundation of all these assertions comes from group representation theory—a purely algebraic framework that is independent of the specific physical system (photons, electrons, phonons). This appendix provides tools at three levels:

- **Origin of degeneracies:** the dimension of an irreducible representation directly determines the minimum degeneracy at a high-symmetry point. A two-dimensional E representation means *at least* twofold degeneracy—this is precisely the algebraic origin of the pseudo-spin in the Wu–Hu model (Chapter 6).
- **Rules of splitting:** compatibility relations reduce irreducible representations of a larger group to those of a smaller group, completely determining the splitting pattern of bands when moving away from a high-symmetry point—band topology “sticking” phenomena can be predicted without numerical computation.
- **Coupling selection rules:** the Wigner–Eckart theorem is the unified framework for selection rules; symmetry-protected BICs (Chapter 10) are essentially a direct consequence of the mismatch between the mode representation and the radiation representation.

Remark (Reading suggestions). If you only need to **use** character tables (look up and apply), you may skip the proof details of §B.2 and start directly from Example B.11 and Example B.12. If you need to **understand** the origin of character tables (e.g., to derive tables for new groups), you must master the orthogonality relations and Schur’s lemma. The newly added step-by-step C_{4v} derivation example bridges these two levels.

This appendix provides the minimal group-theory toolkit needed for symmetry analysis in photonic crystals, aimed at readers without a group-theory background. The focus is on irreducible representations of point groups, character tables, and their applications to band degeneracies and selection rules.

B.1 Basic Concepts of Groups

Definition B.1 (Group). A set G together with a binary operation $\cdot$ forms a **group** $(G, \cdot)$ if the following are satisfied:

1. **Closure**: $\forall g, h \in G, g \cdot h \in G$;
2. **Associativity**: $(g \cdot h) \cdot k = g \cdot (h \cdot k)$;
3. **Identity**: $\exists e \in G$ such that $e \cdot g = g \cdot e = g$;
4. **Inverse**: $\forall g \in G, \exists g^{-1}$ such that $g \cdot g^{-1} = e$.

The **order** $|G|$ of the group is the number of its elements.

Definition B.2 (Subgroup and coset). $H \subset G$ is a **subgroup** of G if H forms a group under the operation of G . The **left coset** of g with respect to H is $gH = \{gh : h \in H\}$. Lagrange's theorem: $|G| = |H| \cdot [G : H]$, where $[G : H]$ is the number of cosets (the index).

Definition B.3 (Conjugacy class). Elements $g, g' \in G$ are **conjugate** if $\exists h \in G$ such that $g' = hgh^{-1}$. Conjugacy is an equivalence relation, partitioning G into **conjugacy classes** $\{C_1, C_2, \dots, C_r\}$. The number of conjugacy classes r equals the number of irreducible representations.

B.2 Group Representation Theory

B.2.1 Basic Definitions

Definition B.4 (Group representation). A **representation** of a group G on a vector space V is a group homomorphism $D : G \rightarrow \text{GL}(V)$, i.e.,

$$D(g \cdot h) = D(g) \cdot D(h), \quad \forall g, h \in G. \quad (\text{B.1})$$

The **dimension** (degree) of the representation is $\dim V$.

Definition B.5 (Equivalent and irreducible representations). Two representations D and D' are **equivalent** if there exists an invertible matrix S such that $D'(g) = SD(g)S^{-1}$ for all g . A representation D is **irreducible** if V has no nontrivial G -invariant subspace.

Theorem B.6 (Maschke's theorem). *Any representation of a finite group G over the complex numbers can be completely reduced to a direct sum of irreducible representations.*

B.2.2 Characters

Definition B.7 (Character). The **character** of a representation D is defined as the trace of the matrix:

$$\chi(g) = \text{tr } D(g). \quad (\text{B.2})$$

Elements in the same conjugacy class have the same character (**class function**).

Theorem B.8 (Character orthogonality relations). *Let $\{\Gamma_\alpha\}_{\alpha=1}^r$ be all irreducible representations of G , and χ_α the corresponding characters; then*

$$\frac{1}{|G|} \sum_{g \in G} \chi_\alpha(g)^* \chi_\beta(g) = \delta_{\alpha\beta}. \quad (\text{B.3})$$

Theorem B.9 (Dimension relation). *The dimensions $d_\alpha = \chi_\alpha(e)$ of the irreducible representations satisfy*

$$\sum_{\alpha=1}^r d_\alpha^2 = |G|. \quad (\text{B.4})$$

B.3 Point Groups Commonly Used in Photonic Crystals

B.3.1 C_{nv} Groups

Definition B.10 (C_{nv} group). A point group containing an n -fold rotation axis C_n and n mirror planes σ_v containing that axis. $|C_{nv}| = 2n$.

Example B.11 (C_{4v} character table). The C_{4v} group (point group at the Γ point of a square lattice) has 8 elements and 5 conjugacy classes:

C_{4v}	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

Verification: $1^2 + 1^2 + 1^2 + 1^2 + 2^2 = 8 = |C_{4v}|$.

Example B.12 (C_{6v} character table). The C_{6v} group (point group at the Γ point of a triangular/honeycomb lattice) has 12 elements and 6 conjugacy classes:

C_{6v}	E	$2C_6$	$2C_3$	C_2	$3\sigma_v$	$3\sigma_d$
A_1	1	1	1	1	1	1
A_2	1	1	1	1	-1	-1
B_1	1	-1	1	-1	1	-1
B_2	1	-1	1	-1	-1	1
E_1	2	1	-1	-2	0	0
E_2	2	-1	-1	2	0	0

E_1 and E_2 are two-dimensional irreducible representations—corresponding to the twofold degeneracy at the Γ point of the Brillouin zone. The pseudo-spin in the Wu–Hu model (Chapter 6) is constructed from the two basis vectors of the E_1 or E_2 representation.

B.3.2 Time Reversal and Magnetic Point Groups

Definition B.13 (Magnetic point group). When anti-unitary time reversal is included, an ordinary point group can be extended to a **magnetic point group**:

$$G_M = H \cup \mathcal{T}(G - H), \quad (\text{B.5})$$

where H is the unitary subgroup. Ordinary nonmagnetic systems with full time-reversal symmetry are described by the gray group $G_M = G \times \{E, \mathcal{T}\}$, while magnetically biased systems generally belong to a smaller magnetic group that contains only the symmetry operations, unitary or anti-unitary, that leave the biased structure invariant.

Remark. In the photonic Chern insulator (Chapter 5), the gyromagnetic material breaks $\mathcal{T}$, and the magnetic point group does not contain $\mathcal{T}$, thereby allowing a nonzero Chern number. In the Wu–Hu model (Chapter 6), the system preserves $\mathcal{T}$ but constructs a pseudo- $\mathcal{T}$ —magnetic point group analysis helps rigorously determine which degeneracies are truly protected by $\mathcal{T}$.

B.4 Space Groups and the Symmetry of Bloch States

Example B.14 (Step-by-step derivation of the C_{4v} character table from scratch). This example shows how to construct the complete character table step by step, starting only from the group multiplication rules. This process requires no physical input—it is purely algebraic.

Step 1: List all symmetry operations. The C_{4v} group describes the symmetry of a square. The 8 operations are:

- The identity operation E ;
- Clockwise rotation by 90° about the z axis, C_4 , and its inverse C_4^3 (i.e., C_4^{-1} , counterclockwise 90°);
- Rotation by 180° , $C_2 = C_4^2$;
- Two mirror planes through the midpoints of opposite edges, σ_v and σ'_v (along the x and y axes);
- Two mirror planes through the diagonals, σ_d and σ'_d (along the $x = y$ and $x = -y$ directions).

Step 2: Determine the conjugacy classes. By computing hgh^{-1} (for all $h \in C_{4v}$), one obtains 5 conjugacy classes:

$$\{E\}, \quad \{C_2\}, \quad \{C_4, C_4^3\}, \quad \{\sigma_v, \sigma'_v\}, \quad \{\sigma_d, \sigma'_d\}. \quad (\text{B.6})$$

Key verification: C_4 and C_4^3 are conjugate (because $\sigma_v C_4 \sigma_v^{-1} = C_4^3$), but C_4 and C_2 are not conjugate (they have different orders). The number of conjugacy classes is $r = 5$, so there are exactly 5 irreducible representations.

Step 3: Use the dimension relation to determine the dimensions of the irreducible representations. The dimension formula:

$$\sum_{\alpha=1}^5 d_{\alpha}^2 = |C_{4v}| = 8. \quad (\text{B.7})$$

The unique non-negative integer solution is $d_{\alpha} = \{1, 1, 1, 1, 2\}$ —four one-dimensional representations plus one two-dimensional representation. (Verification: $1 + 1 + 1 + 1 + 4 = 8$.)

Step 4: Construct the four one-dimensional representations. One-dimensional representations are homomorphisms from the group to $\mathbb{C}^*$ ($D(g) \in \mathbb{C}$). Using $C_4^4 = E$, we get $D(C_4)^4 = 1$, so $D(C_4) \in \{1, -1, i, -i\}$. Simultaneously, $\sigma_v^2 = E$ requires $D(\sigma_v) = \pm 1$.

Using the group multiplication relation $\sigma_v C_4 = C_4^3 \sigma_v$ (mirror and rotation do not commute), taking absolute values in the one-dimensional representation: $D(\sigma_v)D(C_4) = D(C_4)^3 D(\sigma_v)$, i.e., $D(C_4) = D(C_4)^3$, so $D(C_4)^2 = 1$, ruling out $D(C_4) = \pm i$ and leaving only $D(C_4) = \pm 1$.

Similarly, $\sigma_d = C_4 \sigma_v$ gives $D(\sigma_d) = D(C_4)D(\sigma_v)$. Therefore the four one-dimensional representations are determined by the four combinations of $(D(C_4), D(\sigma_v))$:

	$D(C_4)$	$D(\sigma_v)$	$D(C_2)$	$D(\sigma_d)$	Name
$(+, +)$	1	1	1	1	A_1
$(+, -)$	1	-1	1	-1	A_2
$(-, +)$	-1	1	1	-1	B_1
$(-, -)$	-1	-1	1	1	B_2

Here $D(C_2) = D(C_4)^2 = 1$ holds for all one-dimensional representations, and $D(\sigma_d) = D(C_4)D(\sigma_v)$ is automatically determined by the group multiplication relation.

Step 5: Use orthogonality relations to determine the characters of the two-dimensional E representation. Let the characters of the E representation be $(\chi_E, \chi_{C_2}, \chi_{C_4}, \chi_{\sigma_v}, \chi_{\sigma_d})$. First, $\chi_E = d_E = 2$ (the character of the identity operation equals the dimension).

Row orthogonality (the E representation is orthogonal to A_1):

$$\frac{1}{8} [1 \cdot 2 + 1 \cdot \chi_{C_2} \cdot 1 + 2 \cdot 1 \cdot \chi_{C_4} + 2 \cdot 1 \cdot \chi_{\sigma_v} + 2 \cdot 1 \cdot \chi_{\sigma_d}] = 0. \quad (\text{B.8})$$

Similarly, orthogonality with A_2 , B_1 , B_2 gives three more equations. Combining the four orthogonality equations:

$$2 + \chi_{C_2} + 2\chi_{C_4} + 2\chi_{\sigma_v} + 2\chi_{\sigma_d} = 0, \quad (\text{B.9})$$

$$2 + \chi_{C_2} + 2\chi_{C_4} - 2\chi_{\sigma_v} - 2\chi_{\sigma_d} = 0, \quad (\text{B.10})$$

$$2 + \chi_{C_2} - 2\chi_{C_4} + 2\chi_{\sigma_v} - 2\chi_{\sigma_d} = 0, \quad (\text{B.11})$$

$$2 + \chi_{C_2} - 2\chi_{C_4} - 2\chi_{\sigma_v} + 2\chi_{\sigma_d} = 0. \quad (\text{B.12})$$

Eq. (B.9)+(B.10) gives $\chi_{C_4} = -(\chi_{C_2} + 2)/4$; Eq. (B.9)-(B.10) gives $\chi_{\sigma_v} = -\chi_{\sigma_d}$; Eq. (B.11)-(B.12) gives $\chi_{\sigma_v} = \chi_{\sigma_d}$. Therefore $\chi_{\sigma_v} = \chi_{\sigma_d} = 0$; substituting back gives $\chi_{C_2} = -2$, $\chi_{C_4} = 0$.

Complete C_{4v} character table						
C_{4v}	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$	Basis functions
A_1	1	1	1	1	1	$z; x^2 + y^2$
A_2	1	1	1	-1	-1	R_z
B_1	1	-1	1	1	-1	$x^2 - y^2$
B_2	1	-1	1	-1	1	xy
E	2	0	-2	0	0	$(x, y); (R_x, R_y)$

Remark. This derivation is entirely algebraic—the only place where physics enters is in determining which coordinate functions serve as basis functions for each irreducible representation (the last column). Physical meaning of the basis-function column: if the E_z component of a TM mode at the Γ point transforms as the E representation, then that mode is necessarily twofold degenerate.

B.4.1 Space Groups

Definition B.15 (Space group). A space group $\mathcal{G}$ is the group of all isometries of Euclidean space that leave a crystal invariant. Each element $\{R|\mathbf{t}\}$ consists of a point-group operation R and a translation $\mathbf{t}$:

$$\{R|\mathbf{t}\} : \mathbf{r} \rightarrow R\mathbf{r} + \mathbf{t}. \quad (\text{B.13})$$

Multiplication rule: $\{R_1|\mathbf{t}_1\}\{R_2|\mathbf{t}_2\} = \{R_1R_2|R_1\mathbf{t}_2 + \mathbf{t}_1\}$.

Definition B.16 (Little group). The **little group** $G_{\mathbf{k}}$ at a point $\mathbf{k}$ in the Brillouin zone is the set of point-group operations in the space group that leave $\mathbf{k}$ invariant (modulo a reciprocal lattice vector):

$$G_{\mathbf{k}} = \{R \in G : R\mathbf{k} = \mathbf{k} + \mathbf{G}\}. \quad (\text{B.14})$$

The degeneracy and symmetry of Bloch states at the $\mathbf{k}$ point are determined by the irreducible representations of $G_{\mathbf{k}}$.

B.4.2 Compatibility Relations

Definition B.17 (Compatibility relations). When moving from a high-symmetry point $\mathbf{k}_0$ (with a larger little group $G_{\mathbf{k}_0}$) along a high-symmetry line to $\mathbf{k}$ (with $G_{\mathbf{k}} \subset G_{\mathbf{k}_0}$), the irreducible representations of $G_{\mathbf{k}_0}$ may reduce upon restriction to the subgroup $G_{\mathbf{k}}$ —this determines whether and how bands split when departing from $\mathbf{k}_0$.

Example B.18 (Compatibility relations from Γ to M). In a square lattice, the point group at Γ is C_{4v} , and the little group along the Γ – M line is C_s (containing only one mirror plane). The two-dimensional E representation of C_{4v} , when restricted to C_s , reduces to $A' + A''$ —i.e., the twofold degeneracy at Γ generally splits upon departure from Γ . If it does not split (e.g., protected by additional symmetry), topological properties may be present.

B.5 Selection Rules and Radiative Coupling

Theorem B.19 (Wigner–Eckart theorem (simplified version)). *Let $|i\rangle$ belong to irreducible representation Γ_i , $|f\rangle$ belong to Γ_f , and operator $\hat{O}$ transform as Γ_O ; then the matrix element*

$$\langle f|\hat{O}|i\rangle \quad (\text{B.15})$$

can be nonzero only if Γ_f is contained in the direct-sum decomposition of $\Gamma_i \otimes \Gamma_O$.

Example B.20 (Symmetry selection rule for BICs). In Theorem 10.2 of Chapter 10, $|i\rangle$ is a mode of the photonic crystal slab (belonging to Γ_{mode}), $\hat{O}$ is the coupling operator (belonging to the identity representation A_1), and $|f\rangle$ is a radiating plane wave (belonging to Γ_{rad}). The matrix element is zero if and only if $\Gamma_{\text{mode}} \neq \Gamma_{\text{rad}}$ —in this case the mode is a symmetry-protected BIC.

Example B.21 (Transformation matrices for the E representation basis functions). Under the C_{4v} point group, the twofold-degenerate TM modes at the Γ point can use $\{p_x, p_y\}$ as the basis functions of the E representation (physically corresponding to the $p_{\pm}$ orbitals of the Wu–Hu model in Chapter 6). Below we explicitly construct the 2×2 representation matrices for all 8 operations.

Step 1: C_4 rotation matrix. C_4 maps $(x, y) \rightarrow (-y, x)$ (counterclockwise rotation by 90°), so $p_x \rightarrow p_y$, $p_y \rightarrow -p_x$ (the function evaluated at the inversely transformed point), giving

$$\boxed{D^{(E)}(C_4) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}, \quad \text{tr } D^{(E)}(C_4) = 0 = \chi_E(C_4) \quad \checkmark \quad (\text{B.16})$$

Step 2: Representation matrices for all 8 operations. Using $C_2 = C_4^2$, $C_4^3 = C_4^{-1}$, and the geometric definitions of the mirrors:

Operation	Geometric transform $(x, y) \rightarrow$	$D^{(E)}$	tr
E	(x, y)	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	2
C_4	$(-y, x)$	$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$	0
C_4^3	$(y, -x)$	$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$	0
C_2	$(-x, -y)$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$	-2
$\sigma_v(x)$	$(x, -y)$	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	0
$\sigma_v(y)$	$(-x, y)$	$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$	0
$\sigma_d(x=y)$	(y, x)	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	0
$\sigma_d(x=-y)$	$(-y, -x)$	$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$	0

The trace of every matrix is strictly consistent with the character table derived in Example B.14.

Step 3: Selection rule application—band splitting along $\Gamma \rightarrow X$. Along $\Gamma \rightarrow X$ (i.e., $k_y = 0$, k_x from 0 to π/a), the preserved symmetry operations are only $\{E, \sigma_v(x)\}$, and the little group reduces to C_s . The E representation restricted to C_s : $D^{(E)}(\sigma_v(x)) = \text{diag}(1, -1)$, whose eigenvalues are +1 and -1, corresponding to the two one-dimensional representations A' and A'' of C_s :

- p_x mode: symmetric about the x axis (σ_v eigenvalue +1), belongs to A' ;
- p_y mode: antisymmetric about the x axis (σ_v eigenvalue -1), belongs to A'' .

Therefore the twofold degeneracy at Γ *must split* into two branches upon departure from Γ (the dispersion slopes are determined by the effective velocities from $\mathbf{k} \cdot \mathbf{p}$ perturbation theory).

Step 4: Comparison along $\Gamma \rightarrow M$. Along $\Gamma \rightarrow M$ ($k_x = k_y$, diagonal direction), the preserved symmetry operations are $\{E, \sigma_d(x=y)\}$. $D^{(E)}(\sigma_d) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has eigenvalues ± 1 as well, but the basis functions are rotated: the eigenstates are $p_x \pm p_y$ (rather than p_x and p_y). The degeneracy likewise splits, but the *symmetry labels of the two branches are different*.

Remark. This example reveals a key fact: *using only the character table and compatibility relations*—entirely without numerical computation—one can predict the splitting pattern of the Γ -point twofold degeneracy. This is the core value of group theory in photonic band analysis. When the Wu–Hu model tunes the p - d band inversion through contraction/expansion, the two sets of basis functions of the E representation at the Γ point exchange positions, and the topological phase transition is realized precisely through this band inversion (Chapter 6).

B.6 Character Tables and Applications for Photonic Crystal Point Groups

This section gives the complete character tables of the point groups most commonly used in photonic crystals, and demonstrates through concrete examples how to use group theory to analyze degeneracies and coupling rules of photonic bands.

B.6.1 Character Table of the C_{6v} Point Group

The C_{6v} point group is the symmetry group at the Γ point of a honeycomb photonic crystal. It contains 12 symmetry operations, divided into 6 conjugacy classes: $\{E\}$, $\{C_6, C_6^{-1}\}$, $\{C_3, C_3^{-1}\}$, $\{C_2\}$, $\{3\sigma_v\}$, $\{3\sigma_d\}$.

Table B.1 Character table of the C_{6v} point group

	E	$2C_6$	$2C_3$	C_2	$3\sigma_v$	$3\sigma_d$
A_1	1	1	1	1	1	1
A_2	1	1	1	1	-1	-1
B_1	1	-1	1	-1	1	-1
B_2	1	-1	1	-1	-1	1
E_1	2	1	-1	-2	0	0
E_2	2	-1	-1	2	0	0

Application: At the Γ point of a honeycomb photonic crystal, p -orbital-type Bloch states belong to the E_1 representation (two-dimensional irreducible representation), so they form a twofold degeneracy. d -orbital-type states belong to the E_2 representation, also twofold degenerate. s -orbital-type states belong to the A_1 representation (one-dimensional), non-degenerate.

B.6.2 C_{3v} Point Group and K -Point Symmetry

At the K point of the honeycomb lattice, the symmetry group reduces to C_{3v} , containing 6 symmetry operations: $\{E\}$, $\{C_3, C_3^{-1}\}$, $\{3\sigma_v\}$.

Table B.2 Character table of the C_{3v} point group

	E	$2C_3$	$3\sigma_v$
A_1	1	1	1
A_2	1	1	-1
E	2	-1	0

The Dirac cone at the K point originates from the twofold degeneracy of the E representation. When spatial inversion symmetry is broken, the E representation splits into two one-dimensional representations, and the Dirac cone opens a bandgap—this is precisely the basic mechanism of valley photonic crystals.

B.6.3 C_{4v} Point Group and the Square Lattice

The symmetry group at the Γ point of a square-lattice photonic crystal is C_{4v} , containing 8 symmetry operations.

Table B.3 Character table of the C_{4v} point group

	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

B.6.4 Selection Rules and Coupling Matrix Elements

One of the core applications of group theory is determining which matrix elements vanish (selection rules). For two Bloch states $|n, \mathbf{k}\rangle$ and $|m, \mathbf{k}\rangle$, the necessary condition for the matrix element $\langle n, \mathbf{k} | \hat{V} | m, \mathbf{k} \rangle$ of a perturbation $\hat{V}$ to be nonzero is:

$$\Gamma_n^* \otimes \Gamma_V \otimes \Gamma_m \supset A_1, \quad (\text{B.17})$$

where $\Gamma_n, \Gamma_V, \Gamma_m$ are the irreducible representations to which $|n\rangle, \hat{V}, |m\rangle$ belong, respectively.

Example: In a C_{6v} -symmetric photonic crystal, consider the coupling between a p state in the E_1 representation and a d state in the E_2 representation. The perturbation $\hat{V}$ belongs to the E_1 representation (such as the linear term in $\mathbf{k} \cdot \mathbf{p}$ perturbation). The direct-product decomposition is

$$E_1^* \otimes E_1 \otimes E_2 = (A_1 \oplus A_2 \oplus E_2) \otimes E_2 = E_2 \oplus E_2 \oplus (A_1 \oplus A_2 \oplus E_2). \quad (\text{B.18})$$

Since the result contains A_1 , the matrix element is nonzero—the p and d states are coupled through the E_1 -type perturbation.

B.6.5 Time-Reversal Symmetry and Kramers' Theorem

In photonic systems, the time-reversal operator $\mathcal{T}$ acts as $\mathcal{T} : \mathbf{E} \rightarrow \mathbf{E}^*, \mathbf{H} \rightarrow -\mathbf{H}^*$. For systems with real dielectric constants, $\mathcal{T}^2 = +1$ (bosons).

The fundamental difference from electronic systems ($\mathcal{T}^2 = -1$, fermions) is:

- Electronic systems: Kramers' theorem guarantees that all energy levels are at least twofold degenerate.
- Photonic systems: no Kramers degeneracy; energy levels can be non-degenerate.

However, in photonic systems with specific crystalline symmetries and a suitable two-dimensional modal subspace (for example the pseudo-spin construction in the Wu–Hu model), a **pseudo-time-reversal symmetry** $\mathcal{T}_p$ can be constructed such that $\mathcal{T}_p^2 = -1$, thereby restoring Kramers-like degeneracy. This is not a universal consequence of C_6 symmetry alone; it also

requires negligible coupling between the chosen pseudo-spin sectors and the relevant crystalline representation structure.

Numerical Algorithm Implementation

Why Topological Photonics Needs Reproducible Algorithms?

The theoretical derivations in the main text establish a complete framework from Maxwell’s equations to topological invariants, but between *formulas* and *code* there are four typical pitfalls, each of which is sufficient to give a beginner completely wrong results on the first numerical attempt:

1. **Inverse-rule pitfall:** For TM polarization, one should expand $1/\varepsilon$ rather than ε ; for TE polarization, the opposite. But for gyromagnetic tensors, one must invert the Fourier matrix of the entire 2×2 tensor matrix as a whole (Chapter 5, Example 5.4), rather than processing each component separately—violating this rule can degrade convergence speed by two orders of magnitude.
2. **Gauge-choice pitfall:** The link variables of the FHS algorithm automatically achieve lattice gauge invariance through the principal-value selection of $\text{Im} \ln$, but if one instead uses finite-difference ∂_k , gauge smoothness of the eigenstates between adjacent k points must be ensured—which is nearly impossible to achieve manually near degeneracy points.
3. **BZ sampling pitfall:** An overly coarse grid not only affects accuracy but may also cause the true Berry flux through a plaquette to cross the principal-branch cut of $\text{Im} \ln$ (“plaquette overflow”), leading to incorrect Chern numbers. Rule of thumb: the computed flux of a single plaquette should stay well below π in magnitude; $\pi/2$ is a conservative warning threshold.
4. **Inner-product pitfall:** All calculations involving Bloch-state overlaps (FHS link variables, Wilson loop overlap matrices) must use the material-weighted inner product $\langle \cdot | \hat{M} | \cdot \rangle$, not the L^2 inner product—this is the core programming difference between photonic and electronic systems.

Remark (General debugging principle). Before obtaining any “meaningful” topological result, one must first validate the code on models with **known analytical solutions**: the one-dimensional bilayer medium (analytical dispersion relation), the Haldane model (Chern number $C = \pm 1$), and the Kane–Mele model ($\mathbb{Z}_2 = 1$). Only after the code is completely correct on these benchmarks may it be applied to photonic crystal systems.

This appendix provides pseudocode and implementation notes for the key numerical algorithms in the main text, enabling the reader to reproduce the main results on their own computational

platform. All algorithms are based on the material-weighted inner product framework established in the preceding chapters.

C.1 Plane-Wave Expansion Method (PWE)

C.1.1 Problem Setup

Goal: solve for the Bloch eigenfrequencies $\omega_n(\mathbf{k})$ and eigenmodes of a two-dimensional photonic crystal. Taking TM polarization ($\mathbf{E}$ along the z direction) as an example, the scalar eigenequation is (Chapter 3, Eq. (2.29)):

$$-(\nabla + i\mathbf{k}) \cdot \frac{1}{\varepsilon(\mathbf{r})} (\nabla + i\mathbf{k}) H_z(\mathbf{r}) = \frac{\omega^2}{c^2} H_z(\mathbf{r}). \quad (\text{C.1})$$

C.1.2 Fourier Expansion

Expand H_z and $1/\varepsilon$ in Fourier series over reciprocal lattice vectors $\mathbf{G}$:

$$H_z(\mathbf{r}) = \sum_{\mathbf{G}} h_{\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}}, \quad \frac{1}{\varepsilon(\mathbf{r})} = \sum_{\mathbf{G}} \eta_{\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}}. \quad (\text{C.2})$$

Substituting into Eq. (C.1) yields the matrix eigenvalue problem:

$$\sum_{\mathbf{G}'} |\mathbf{k} + \mathbf{G}| \eta_{\mathbf{G}-\mathbf{G}'} |\mathbf{k} + \mathbf{G}'| h_{\mathbf{G}'} = \frac{\omega^2}{c^2} h_{\mathbf{G}}. \quad (\text{C.3})$$

More precisely, the matrix elements are:

$$M_{\mathbf{G}\mathbf{G}'} = (\mathbf{k} + \mathbf{G}) \cdot \eta_{\mathbf{G}-\mathbf{G}'} (\mathbf{k} + \mathbf{G}'). \quad (\text{C.4})$$

C.1.3 PWE Algorithm Pseudocode

Algorithm 1 PWE_TM_Eigenvalue

1. **Input:** Lattice vectors $\mathbf{a}_1, \mathbf{a}_2$; the dielectric function $\varepsilon(\mathbf{r})$ sampled on the unit cell; truncation parameter $G_{\max}$ (determining the plane-wave basis size N_G); wave vector $\mathbf{k}$.
 2. Compute the reciprocal lattice vectors $\mathbf{b}_1, \mathbf{b}_2$, enumerate all reciprocal lattice vectors $\{\mathbf{G}_1, \dots, \mathbf{G}_{N_G}\}$ satisfying $|\mathbf{G}| < G_{\max}$.
 3. Compute the Fourier coefficients $\eta_{\mathbf{G}}$ of $1/\varepsilon(\mathbf{r})$ on the unit cell (via FFT: first sample $1/\varepsilon$ on a real-space grid, then perform a 2D FFT).
 4. Construct the $N_G \times N_G$ matrix M : $M_{ij} = (\mathbf{k} + \mathbf{G}_i) \cdot \eta_{\mathbf{G}_i - \mathbf{G}_j} (\mathbf{k} + \mathbf{G}_j)$.
 5. Solve the Hermitian matrix M for eigenvalues ω_n^2/c^2 and eigenvectors $h_{\mathbf{G}}^{(n)}$.
 6. **Output:** Eigenfrequencies $\omega_n = c\sqrt{\lambda_n}$, eigenmodes $H_z^{(n)}(\mathbf{r})$.
-

Remark (Implementation notes). • Choice of truncation N_G : convergence must be checked; typically $N_G \sim 100\text{--}1000$ is sufficient.

- For TE polarization, one must use the Fourier coefficients of $\varepsilon(\mathbf{r})$ rather than $1/\varepsilon(\mathbf{r})$, and the matrix form is different (see Chapter 3, Eq. (2.31)).
- Ho–Chan–Soukoulis truncation rule: for high dielectric contrast, directly truncating the Fourier series of $1/\varepsilon$ may cause Gibbs phenomena; a better approach is to compute the Fourier matrix of ε in real space and then invert it.

C.1.4 Complete Numerical Workflow for 1D Photonic Crystal PWE

Below we give the complete implementation of the TM-mode PWE for a one-dimensional layered photonic crystal. This “minimal verifiable model” is a *mandatory checkpoint* before debugging 2D/3D codes.

Example C.1 (End-to-end implementation of 1D PWE). **Model:** AB layered medium with period a , ε_A (thickness fa) and $\varepsilon_B = 1$ (thickness $(1-f)a$).

Step 1: Parameter definition

```

1 eps_A, eps_B = 12.0, 1.0    # Dielectric constants
2 f = 0.2                    # Fill fraction of layer A
3 a = 1.0                    # Normalized lattice constant
4 N_G = 31                   # Truncation order: G_n = 2*pi*n/a, n in [-N_G, N_G]
5 N = 2*N_G + 1              # Matrix dimension = 63
    
```

Step 2: Reciprocal lattice vector list

$$G_n = \frac{2\pi n}{a}, \quad n \in \{-N_G, -N_G + 1, \dots, N_G\}. \quad (\text{C.5})$$

Step 3: Fourier coefficients of the dielectric constant

The analytical expression for the Fourier coefficients of ε^{-1} (note: the inverse) of the 1D layered structure:

$$\eta_G \equiv (\widehat{\varepsilon^{-1}})_G = \begin{cases} \frac{f}{\varepsilon_A} + \frac{1-f}{\varepsilon_B}, & G = 0, \\ \left(\frac{1}{\varepsilon_A} - \frac{1}{\varepsilon_B} \right) \frac{\sin(\pi G f a / (2\pi))}{\pi G / (2\pi/a)}, & G \neq 0. \end{cases} \quad (\text{C.6})$$

```

1 import numpy as np
2 G_list = 2*np.pi*np.arange(-N_G, N_G+1)/a
3 eta = np.zeros(N, dtype=complex)
4 for i, Gi in enumerate(G_list):
5     if abs(Gi) < 1e-12:
6         eta[i] = f/eps_A + (1-f)/eps_B
7     else:
8         eta[i] = (1/eps_A - 1/eps_B)*np.sin(Gi*f*a/2)/(Gi*a/2)
    
```

Remark (The critical choice of inverse rule). The direct expansion of ε^{-1} above is *exactly correct* for the 1D TM mode, because the TM scalar equation (3.24) involves precisely ε^{-1} . But for TE modes or 2D problems, the **matrix inverse rule** must be used: first construct the Toeplitz matrix $[\varepsilon]_{mn} = \hat{\varepsilon}_{G_m - G_n}$ of ε , then perform `numpy.linalg.inv` to invert the entire matrix and obtain the $[\eta]$ matrix. *Directly expanding the Fourier series of ε^{-1} in the TE case leads to*

extremely slow convergence— $N_G = 200$ still has $> 5\%$ error, whereas the matrix inverse rule achieves $< 0.1\%$ at $N_G = 31$.

Step 4: Construction of the matrix $M(k)$

For the 1D TM mode, the PWE reduces to a standard eigenvalue problem

$$\boxed{M_{mn}(k) = (k + G_m) \eta_{m-n}(k + G_n)}, \quad \frac{\omega^2}{c^2} \mathbf{h} = M(k) \mathbf{h}. \quad (\text{C.7})$$

```

1 def build_M(k, G_list, eta_coeffs, N):
2     """Construct the N x N PWE matrix"""
3     M = np.zeros((N, N), dtype=complex)
4     for m in range(N):
5         for n in range(N):
6             idx = (m - n) + N_G    # Index offset for eta
7             M[m, n] = (k + G_list[m]) * eta_coeffs[idx] * (k + G_list[n])
8     return M

```

Step 5: Solve eigenvalues

```

1 from scipy.linalg import eigh
2 k_path = np.linspace(0, np.pi/a, 100) # Gamma -> X
3 bands = np.zeros((100, 5)) # First 5 bands
4 for i, k in enumerate(k_path):
5     M = build_M(k, G_list, eta_full, N)
6     eigvals = eigh(M, eigvals_only=True) # Real symmetric
7     omega = np.sqrt(np.abs(eigvals[:5])) * a / (2*np.pi) # Normalized frequency
8     bands[i] = omega

```

Step 6: Convergence test (mandatory!)

N_G	$\omega_1 a / 2\pi c$	$\omega_2 a / 2\pi c$	$\omega_3 a / 2\pi c$
5	0.2041	0.4368	0.6102
11	0.2033	0.4349	0.6078
21	0.2031	0.4345	0.6072
31	0.2031	0.4344	0.6071
41	0.2031	0.4344	0.6071

Criterion: frequency differences between adjacent rows $< 10^{-4}$ indicate convergence. $N_G = 31$ (63×63 matrix) is sufficient for these parameters; a 2D square lattice typically requires $N_G \sim 10\text{--}15$ (i.e., $(2N_G + 1)^2 \sim 500\text{--}900$ -dimensional matrices).

Step 7: Verification and debugging

- **Analytical check:** When $\varepsilon_A = \varepsilon_B = 1$ (homogeneous medium), the bands should be exact straight lines $\omega = c|k + G_n|$.
- **Bandgap check:** The bandgap width at $k = \pi/a$ (BZ boundary) should approximately be $\Delta\omega/\omega_{\text{mid}} \approx |\varepsilon_A - \varepsilon_B|/(\varepsilon_A + \varepsilon_B)$ (first-order perturbation theory).
- **Symmetry check:** $\omega(k) = \omega(-k)$ (time-reversal symmetry).

C.2 FHS Lattice Berry Curvature

Based on the Fukui–Hatsugai–Suzuki method (Chapter 4, FHS Theorem), the Chern number is computed on a discretized Brillouin zone.

C.2.1 FHS Algorithm Pseudocode

Algorithm 2 FHS_Chern_Number

1. **Input:** Brillouin zone grid $\{(k_{x,i}, k_{y,j})\}$ ($N_x \times N_y$), Bloch eigenstates $|u_n(\mathbf{k})\rangle$ at each grid point (already solved via PWE or FEM).
2. For each grid point (i, j) , compute four link variables:

$$U_x(i, j) = \frac{\langle u_n(\mathbf{k}_{i,j}) | \bar{M} | u_n(\mathbf{k}_{i+1,j}) \rangle}{|\langle u_n(\mathbf{k}_{i,j}) | \bar{M} | u_n(\mathbf{k}_{i+1,j}) \rangle|}, \quad (\text{C.8})$$

$$U_y(i, j) = \frac{\langle u_n(\mathbf{k}_{i,j}) | \bar{M} | u_n(\mathbf{k}_{i,j+1}) \rangle}{|\langle u_n(\mathbf{k}_{i,j}) | \bar{M} | u_n(\mathbf{k}_{i,j+1}) \rangle|}. \quad (\text{C.9})$$

Note: The inner product must use the material-weighted inner product $\langle \cdot | \bar{M} | \cdot \rangle$, not the L^2 inner product!

3. Compute the lattice Berry curvature for each plaquette:

$$\tilde{F}_{xy}(i, j) = \text{Im} \ln [U_x(i, j) U_y(i+1, j) U_x(i, j+1)^* U_y(i, j)^*]. \quad (\text{C.10})$$

Take the principal value of $\text{Im} \ln$ in $(-\pi, \pi]$.

4. Chern number:

$$C_n = \frac{1}{2\pi} \sum_{i,j} \tilde{F}_{xy}(i, j). \quad (\text{C.11})$$

Key: Due to the principal-value selection of $\tilde{F}_{xy}$, C_n is integer-valued for an admissible discretization with nonzero link overlaps and a well-isolated target band/subspace. A coarse grid can still return the wrong integer if it misses sharply concentrated Berry curvature or crosses a branch cut.

5. **Output:** Integer Chern number C_n .
-

Remark (Multi-band generalization). For the total Chern number of an isolated N -band subspace, replace the scalar overlap with the determinant of the $N \times N$ overlap matrix:

$$[S_\mu(i, j)]_{mn} = \langle u_m(\mathbf{k}_{i,j}) | \bar{M} | u_n(\mathbf{k}_{i,j} + \hat{\mu}) \rangle, \quad U_\mu(i, j) = \frac{\det S_\mu(i, j)}{|\det S_\mu(i, j)|}, \quad (\text{C.12})$$

and the lattice Berry curvature becomes

$$\tilde{F}_{xy}(i, j) = \text{Im} \ln [U_x(i, j) U_y(i+1, j) U_x(i, j+1)^* U_y(i, j)^*]. \quad (\text{C.13})$$

C.2.2 FHS Algorithm Debugging Checklist

In practice, although the FHS algorithm is conceptually simple, it has several subtle pitfalls. The following lists the eight most common types of errors and their fixes.

- Pitfall 1. Branch selection of the link variable.** $U_\mu(i, j) = \langle u(\mathbf{k}_i) | \hat{M} | u(\mathbf{k}_j) \rangle / |\langle u(\mathbf{k}_i) | \hat{M} | u(\mathbf{k}_j) \rangle|$ does not require manual argument branch selection—the $\text{Im} \ln \det$ in the FHS formula automatically restricts each plaquette’s curvature to the $(-\pi, \pi]$ interval. *Wrong approach:* computing $\arg$ of U_μ and then manually adding/subtracting 2π —this introduces uncontrollable errors.
- Pitfall 2. Insufficient Brillouin zone sampling density.** The FHS formula is gauge-invariant on the lattice, but a physically reliable result requires that the Berry flux on each plaquette not cross the principal-branch cut. Empirical criterion: $|\tilde{F}_{xy}(i, j)| < \pi/2$ for all plaquettes. If violated, the grid is too coarse and needs at least $2\times$ refinement. Typical safe values: Haldane model $N_k = 20 \times 20$, Wang model $N_k = 30 \times 30$.
- Pitfall 3. Handling degenerate or nearly degenerate bands.** When two bands are nearly degenerate at some $\mathbf{k}$ point ($|\omega_n - \omega_m|/\omega_n < 10^{-6}$), the single-band link variable U_μ exhibits random phase jumps. *Fix:* use multi-band FHS for the degenerate subspace—upgrade the link variable to an $N_{\text{occ}} \times N_{\text{occ}}$ matrix $[U_\mu]_{mn} = \langle u_m(\mathbf{k}_i) | \hat{M} | u_n(\mathbf{k}_j) \rangle$, and use $\text{Im} \ln \det$ to compute the argument of the determinant.
- Pitfall 4. Gauge smoothing—FHS does not need it!** The core advantage of the FHS method is *gauge invariance*: regardless of the random gauge of the eigenstates, the plaquette curvature from $\text{Im} \ln \det$ is invariant. However, intermediate steps of the Wilson loop method may require polar decomposition to ensure numerical stability—do not confuse the gauge requirements of the two methods.
- Pitfall 5. Chern number accuracy criterion.** The total Chern number

$$C = \frac{1}{2\pi} \sum_{i,j} \tilde{F}_{xy}(i, j) \quad (\text{C.14})$$

from the FHS method is integer-valued in exact arithmetic for an admissible lattice field, but floating-point arithmetic and near-zero overlaps can produce small deviations. Practical criterion:

$$|C - \text{round}(C)| < 0.01. \quad (\text{C.15})$$

If the deviation is between 0.01 and 0.1, the grid, overlap conditioning, or subspace tracking is usually insufficient; if the deviation is > 0.1 , there is very likely an eigenstate sorting error or an unexpected band crossing.

- Pitfall 6. Plaquette curvature overflow warning.** If a single plaquette’s $|\tilde{F}_{xy}|$ approaches π , $\text{Im} \ln$ cannot distinguish $+\pi$ from $-\pi$, leading to the loss of a 2π quantum. *Defensive programming:* add a warning when the plaquette curvature becomes too large:

```
if abs(F_plaq) > pi/2:
    warnings.warn("plaquette overflow")
```

- Pitfall 7. Continuity of eigenstate sorting.** Diagonalization routines (such as `eigh` in NumPy,) may return eigenstates with “label jumps” between different $\mathbf{k}$ points (sorting swaps when two bands cross). *Fix:* use the material-weighted overlap matrix $O_{mn} = |\langle u_m(\mathbf{k}) | \hat{M} | u_n(\mathbf{k} + \delta\mathbf{k}) \rangle|^2$ between adjacent $\mathbf{k}$ points to track band labels—if the maximum overlap is not on the diagonal, a swap has occurred. *Note:* The FHS method is much more robust

to sorting jumps than finite-difference methods (due to gauge invariance), but sorting errors will still cause the Chern number to be computed for the wrong band.

Pitfall 8. Periodicity of boundary conditions. The BZ boundaries $k_\mu = \pi/a$ and $k_\mu = -\pi/a$ represent the same physical state up to the chosen Bloch-periodic convention and a reciprocal-lattice translation. When constructing plaquette loops, the link variables of the last column/row must correctly “wrap around” to the first column/row: $U_x(N_x, j) = \langle u(k_{N_x}, k_{y,j}) | \hat{M} | u(k_1, k_{y,j}) \rangle / |\cdots|$. Omitting this step causes anomalous curvature in the last column of plaquettes.

Remark (Golden rule for debugging). For any newly implemented FHS code, first validate on *analytically known* models: (i) homogeneous medium (all bands $C = 0$); (ii) Haldane model $\phi = \pi/2$ ($C = \pm 1$); (iii) time-reversal invariant model (each band $C = 0$, but $\mathbb{Z}_2$ can be nontrivial). Only after all three tests pass should the code be applied to the target system.

C.3 Wilson Loop Method

The Wilson loop is used to compute Wannier center evolution and the $\mathbb{Z}_2$ invariant (Chapter 4, §4.6 and §4.8).

C.3.1 Wilson Loop Algorithm Pseudocode

Algorithm 3 Wilson_Loop

1. **Input:** N_x equally spaced k points along the k_x direction $\{k_{x,1}, \dots, k_{x,N_x}\}$ (with wrap-around $k_{x,N_x+1} = k_{x,1} + G_x$), at fixed k_y ; Bloch states $\{|u_n(\mathbf{k})\rangle\}_{n=1}^N$ of N bands at each k point.
2. Compute the overlap matrix between adjacent k points:

$$[F_i]_{mn} = \langle u_m(k_{x,i}, k_y) | \bar{M} | u_n(k_{x,i+1}, k_y) \rangle. \quad (\text{C.16})$$

3. (Optional but recommended) Perform polar decomposition $F_i = U_i P_i$ on each F_i , and take the unitary part U_i . This removes modulus information, retaining only phase/gauge information.
4. Compute the Wilson loop matrix:

$$\mathcal{W}(k_y) = U_{N_x} U_{N_x-1} \cdots U_2 U_1. \quad (\text{C.17})$$

$\mathcal{W}$ is an $N \times N$ unitary matrix.

5. Diagonalize $\mathcal{W}(k_y)$; the eigenvalues are $e^{i\theta_j(k_y)}$ ($j = 1, \dots, N$). $\theta_j/(2\pi)$ are the Wannier centers.
 6. Repeat steps 1–5 for each k_y value to obtain the Wannier bands $\{\theta_j(k_y)\}$.
 7. **Output:** Wannier bands.
 - If the winding number of the Wannier bands is nonzero $\rightarrow$ nonzero Chern number.
 - If the Wannier bands cross half the Brillouin zone an odd number of times $\rightarrow$ nontrivial $\mathbb{Z}_2$.
-

Remark (Nested Wilson loop). For higher-order topology calculations (see Chapter 10 for the discussion on BICs and other frontier directions), perform a second Wilson loop on the Wannier bands:

1. First-layer Wilson loop (along k_x) $\rightarrow$ Wannier bands $\theta_j(k_y)$;
2. Treat the Wannier bands as new “bands” and perform another Wilson loop in k_y space;
3. The winding number of the nested Wilson loop gives the quadrupole invariant q_{xy} .

C.4 Supercell Method and Edge States

Computing the dispersion relation of topological edge states.

C.4.1 Algorithm Concept

1. Construct a supercell that is finite along y and periodic along x (ribbon geometry).
2. The supercell consists of N_{cell} unit cells arranged along y , with the upper half in the topological phase ($C \neq 0$) and the lower half in the trivial phase ($C = 0$) or vacuum.
3. Apply Bloch periodicity along k_x ; retain finite degrees of freedom along y .
4. For each k_x , solve the supercell eigenvalue problem— bulk states form projected bands (many densely distributed eigenvalues), and edge states appear within the projected bandgap.

C.4.2 Supercell Edge State Algorithm Pseudocode

Algorithm 4 Supercell_Edge_States

1. **Input:** Dielectric function parameters for the topological and trivial phases; supercell size N_{cell} ; k_x grid.
 2. For each k_x :
 - (a) Construct the supercell dielectric function (upper half topological, lower half trivial);
 - (b) Solve for eigenfrequencies $\{\omega_n(k_x)\}$ ($n = 1, \dots, N_{\text{band}}$) using PWE or FEM;
 - (c) For each eigenstate, compute its center of mass along y : $\langle y \rangle_n$ — bulk states have their center in the middle, edge states near the interface.
 3. Plot the $\omega_n(k_x)$ dispersion relation, color-coded by $\langle y \rangle_n$ to distinguish bulk and edge states.
 4. **Output:** Projected band structure including edge state dispersion.
-

Remark (Finite-size effects). The supercell size N_{cell} must be large enough (typically > 20 unit cells) so that the edge states at the upper and lower interfaces do not couple via tunneling. Convergence can be checked by increasing N_{cell} .

C.5 Transfer Matrix Method (1D)

Used for analytical and numerical analysis of one-dimensional photonic crystals (Chapter 3, variational bandgap analysis).

C.5.1 Single-Layer Transfer Matrix

For a homogeneous dielectric layer (thickness d , refractive index n), the amplitudes (A, B) of the forward and backward propagating waves transform to (A', B') after passing through the layer:

$$\begin{pmatrix} A' \\ B' \end{pmatrix} = \underbrace{\begin{pmatrix} \cos(nk_0d) & -\frac{i}{n} \sin(nk_0d) \\ -in \sin(nk_0d) & \cos(nk_0d) \end{pmatrix}}_{T_{\text{layer}}} \begin{pmatrix} A \\ B \end{pmatrix}, \quad (\text{C.18})$$

where $k_0 = \omega/c$.

C.5.2 Transfer Matrix of One Period

For an AB -type bilayer periodic structure:

$$T_{\text{cell}} = T_B \cdot T_A. \quad (\text{C.19})$$

C.5.3 Bloch Condition and Dispersion Relation

Bloch's theorem requires the eigenvalues of T_{cell} to be $e^{\pm iKa}$ (K being the Bloch wave vector), i.e.,

$$\cos(Ka) = \frac{1}{2} \text{tr } T_{\text{cell}}. \quad (\text{C.20})$$

When $|\frac{1}{2} \text{tr } T_{\text{cell}}| > 1$, K is complex—a bandgap.

C.5.4 Zak Phase via the Transfer Matrix

The topological invariant of a one-dimensional system is the Zak phase $\gamma = \oint_{\text{BZ}} \mathcal{A}(k) dk$, with $\mathcal{A} = -i\langle u | \partial_k u \rangle$ in the convention of Chapter 4. It is quantized to 0 or π in the presence of spatial inversion symmetry.

C.6 Programming Recommendations

C.6.1 Recommended Tools

- **MIT Photonic Bands (MPB)**: The most mature frequency-domain photonic band solver, implementing the PWE method with support for 2D/3D and anisotropic materials. Open source, C language + Python/Scheme interface.
- **MEEP**: FDTD solver, in the same ecosystem as MPB. Suitable for computing transmission spectra, quality factors, and radiation modes.

Algorithm 5 Zak_Phase_Transfer_Matrix

1. Sample N_k uniformly spaced points on the Brillouin zone $k \in [-\pi/a, \pi/a]$.
2. For each k , solve for the Bloch state $|u(k)\rangle$ using the transfer matrix method.
3. Compute the overlap phase between adjacent k points (1D FHS method):

$$\gamma = -\text{Im} \sum_{j=1}^{N_k} \ln \langle u(k_j) | \hat{M} | u(k_{j+1}) \rangle. \quad (\text{C.21})$$

For scalar transfer-matrix models this reduces to the appropriate energy-normalized modal overlap.

4. Output $\gamma \bmod 2\pi$ (should be 0 or π).

- **COMSOL Multiphysics:** Commercial FEM software, suitable for complex geometries and anisotropic/dispersive materials.
- **Python ecosystem:**
 - **numpy/scipy:** matrix diagonalization, FFT, linear algebra;
 - **matplotlib:** band diagram and Berry curvature visualization;
 - **kwant:** topological calculations for tight-binding models (although designed for electronic systems, it can be directly used for photonic systems after tight-binding mapping).

C.6.2 Debugging and Verification Checklist

1. **Convergence test:** Increase N_G (PWE) or mesh density (FEM), check whether eigenfrequency changes are smaller than the target accuracy.
2. **Symmetry verification:** Do the computed bands respect the crystal symmetry (e.g., $\omega(k_x, k_y) = \omega(k_y, -k_x)$ under C_4 symmetry)?
3. **TRS verification:** Without magnetic field, $\omega(\mathbf{k}) = \omega(-\mathbf{k})$.
4. **Chern number integrality:** For an admissible grid the FHS method should give an integer up to roundoff; deviations beyond ± 0.01 indicate overlap zeros, too coarse a grid, or subspace-tracking errors.
5. **Bandgap closure check:** Near topological phase transition points, the k grid must be refined.
6. **Literature comparison:** For classic models (e.g., Wang model, Wu–Hu model), quantitatively compare with band diagrams in the literature.

C.7 Numerical Methods for Nonlinear Topological Photonic Systems

This section supplements the core numerical algorithms involved in nonlinear topological photonics (Chapter 12), providing a directly implementable computational framework for the reader.

C.7.1 Self-Consistent Iteration Algorithm for Nonlinear Bloch Waves

The standard method for solving the nonlinear eigenequation

$$\nabla \times (\nabla \times \mathbf{E}_{\mathbf{k}}) = \frac{\omega^2}{c^2} [\varepsilon(\mathbf{r}) + \chi^{(3)}(\mathbf{r}) |\mathbf{E}_{\mathbf{k}}|^2] \mathbf{E}_{\mathbf{k}} \quad (\text{C.22})$$

is self-consistent iteration. The complete steps of the algorithm are as follows:

Algorithm C.1: Nonlinear Bloch wave self-consistent iteration

1. **Initialization:** Solve the linear problem ($\chi^{(3)} = 0$), obtaining the Bloch mode $\mathbf{E}_{\mathbf{k}}^{(0)}$ and frequency $\omega^{(0)}$ of the n -th band. Set the target power P_{target} and convergence threshold δ_{tol} .
2. **Power normalization:** Normalize $\mathbf{E}_{\mathbf{k}}^{(j)}$ to the target power:

$$\mathbf{E}_{\mathbf{k}}^{(j)} \leftarrow \sqrt{\frac{P_{\text{target}}}{\int_{\text{cell}} |\mathbf{E}_{\mathbf{k}}^{(j)}|^2 d^3r}} \mathbf{E}_{\mathbf{k}}^{(j)}. \quad (\text{C.23})$$

3. **Effective dielectric constant update:**

$$\varepsilon_{\text{eff}}^{(j)}(\mathbf{r}) = \varepsilon(\mathbf{r}) + \chi^{(3)}(\mathbf{r}) |\mathbf{E}_{\mathbf{k}}^{(j)}(\mathbf{r})|^2. \quad (\text{C.24})$$

4. **Linear solve:** Using $\varepsilon_{\text{eff}}^{(j)}$ as the dielectric constant, solve the linear Maxwell eigenequation to obtain $\mathbf{E}_{\mathbf{k}}^{(j+1)}$ and $\omega^{(j+1)}$.
5. **Mixing** (to accelerate convergence):

$$\tilde{\mathbf{E}}_{\mathbf{k}}^{(j+1)} = \alpha \mathbf{E}_{\mathbf{k}}^{(j+1)} + (1 - \alpha) \mathbf{E}_{\mathbf{k}}^{(j)}. \quad (\text{C.25})$$

Compute the relative error

$$\delta = \frac{\|\tilde{\mathbf{E}}_{\mathbf{k}}^{(j+1)} - \mathbf{E}_{\mathbf{k}}^{(j)}\|}{\|\mathbf{E}_{\mathbf{k}}^{(j)}\|}. \quad (\text{C.26})$$

If $\delta < \delta_{\text{tol}}$ (typically 10^{-8}), then converged; otherwise return to step 2.

Remark (Convergence and power continuation). For weak nonlinearity ($\chi^{(3)} |\mathbf{E}|^2 \ll \varepsilon$), the above algorithm typically converges within 5–20 iterations. For strong nonlinearity, starting directly from high power may not converge. In this case, a **power continuation** strategy should be used: start from low power P_0 and gradually increase P , using the converged solution from the previous step as the initial guess for the next. The step size ΔP should be small enough to ensure continuity.

C.7.2 Numerical Solution of the Bogoliubov–de Gennes Equation

Numerical solution of the BdG equation requires constructing a $2N \times 2N$ matrix (N being the number of spatial discretization points), then diagonalizing it.

Algorithm C.2: BdG stability analysis

1. After self-consistent iteration converges, obtain the nonlinear Bloch wave $\mathbf{E}_0(\mathbf{r})$.
2. Construct the BdG matrix:

$$\mathcal{M}_{\text{BdG}} = \begin{pmatrix} L & M \\ -M^* & -L^* \end{pmatrix}, \quad (\text{C.27})$$

where $L_{ij} = \Theta_{ij} - (\omega^2/c^2)[\varepsilon_i + 2\chi_i^{(3)}|E_{0,i}|^2]\delta_{ij}$, $M_{ij} = -(\omega^2/c^2)\chi_i^{(3)}E_{0,i}^2\delta_{ij}$.

3. Numerically diagonalize $\mathcal{M}_{\text{BdG}}$, obtaining eigenvalues $\{\lambda_j\}$ and eigenvectors $\{(u_j, v_j)^T\}$.
4. Check the real parts of all eigenvalues: if $\max_j \text{Re}(\lambda_j) > 0$, then the nonlinear Bloch wave is unstable.
5. Verify the result using particle-hole symmetry: eigenvalues should appear in pairs $(\lambda, -\lambda^*)$. If this is not satisfied (beyond numerical error), the discretization accuracy is insufficient.

C.7.3 Implementation Details of the Split-Step Fourier Method

The split-step Fourier method is the standard tool for simulating nonlinear optical propagation. For the nonlinear Schrödinger equation with a topological potential

$$i\frac{\partial\psi}{\partial z} = -\frac{1}{2k_0}\nabla_{\perp}^2\psi + V(\mathbf{r}_{\perp})\psi + g|\psi|^2\psi, \quad (\text{C.28})$$

the split-step algorithm is:

Algorithm C.3: Symmetric split-step Fourier

1. **Nonlinear half-step:** $\psi^{(1)} = \psi^{(n)} \exp[-i\frac{\Delta z}{2}(V + g|\psi^{(n)}|^2)]$.
2. **Linear full step** (in Fourier space): $\hat{\psi}^{(2)} = \hat{\psi}^{(1)} \exp[-i\Delta z\frac{|\mathbf{k}_{\perp}|^2}{2k_0}]$.
3. **Nonlinear half-step:** $\psi^{(n+1)} = \psi^{(2)} \exp[-i\frac{\Delta z}{2}(V + g|\psi^{(2)}|^2)]$.

Remark (Step size selection). The local truncation error of the symmetric split-step method is $O(\Delta z^3)$. The step size Δz must satisfy two conditions: (1) linear condition $\Delta z < k_0\Delta x^2/\pi$ (Δx being the spatial grid spacing); (2) nonlinear condition $\Delta z < 1/(g|A_{\text{max}}|^2)$. In practice, take the smaller of the two.

C.7.4 Imaginary-Time Evolution for Topological Edge Solitons

Algorithm C.4: Imaginary-time evolution

1. Set the topological potential $V(\mathbf{r})$ (e.g., SSH potential or Haldane potential).
2. Choose an initial guess $\psi_0(\mathbf{r})$, localized near the topological edge. Typically take the linear topological edge state multiplied by an appropriate amplitude.
3. Replace $z \rightarrow -i\tau$ (imaginary time); the evolution equation becomes diffusion-type:

$$\frac{\partial\psi}{\partial\tau} = \frac{1}{2k_0}\nabla_{\perp}^2\psi - V\psi - g|\psi|^2\psi. \quad (\text{C.29})$$

4. After each time step $\Delta\tau$, normalize to the target power: $\psi \leftarrow \psi\sqrt{P_{\text{target}}/\int|\psi|^2d\mathbf{r}}$.

5. Monitor the convergence of the chemical potential $\mu = -\langle \psi | \hat{H}_{\text{NL}} | \psi \rangle / P$.
6. The converged $\psi(\mathbf{r})$ is the nonlinear topological edge soliton.

C.7.5 Supercell Method for 3D Topological Photonic Crystals

For three-dimensional topological photonic crystals containing surfaces or interfaces, the supercell method must be used.

Algorithm C.5: 3D supercell computation

1. Construct a supercell along the z direction: N_z layers of photonic crystal + N_{vac} layers of vacuum. Typically $N_z = 10\text{--}20$, $N_{\text{vac}} = 5\text{--}10$.
2. Select sampling points in the (k_x, k_y) plane.
3. For each (k_x, k_y) , solve the supercell eigenfrequencies using the plane-wave expansion method.
4. Distinguish bulk and surface states by analyzing the spatial distribution of eigenmodes:

$$\eta_{\text{surf}} = \frac{\int_{|z| < d} |\mathbf{E}|^2 d^3r}{\int_{\text{cell}} |\mathbf{E}|^2 d^3r}, \quad (\text{C.30})$$

where d is the surface layer thickness. States with $\eta_{\text{surf}} > 0.5$ are classified as surface states.

5. Plot the dispersion relation $\omega(k_x, k_y)$ of the surface states, which constitutes the Fermi arc.

Remark (Convergence of supercell size). The key parameters of the supercell method are N_z and N_{vac} . N_z determines the resolution of the bulk dispersion (larger N_z gives denser bulk sub-bands). N_{vac} determines the coupling strength between the upper and lower surface states—when N_{vac} is insufficient, the states on the two surfaces hybridize, causing spurious bandgaps in the Fermi arcs. In practice, convergence with respect to N_z and N_{vac} should be checked.

C.7.6 Efficient Numerical Computation of Berry Curvature

The standard method for computing Berry curvature in the 3D Brillouin zone is the Fukui–Hatsugai–Suzuki discretization scheme. For the Chern number on a 2D slice BZ_{k_z} :

Algorithm C.6: Discrete Berry curvature and Chern number

1. Discretize the (k_x, k_y) plane into an $N_x \times N_y$ grid.
2. At each grid point $\mathbf{k}_{ij}$, solve the Bloch equation to obtain eigenstates $|u_{n,ij}\rangle$.
3. Compute the link variables between grid points:

$$U_x(i, j) = \frac{\langle u_{n,ij} | u_{n,i+1,j} \rangle}{|\langle u_{n,ij} | u_{n,i+1,j} \rangle|}, \quad U_y(i, j) = \frac{\langle u_{n,ij} | u_{n,i,j+1} \rangle}{|\langle u_{n,ij} | u_{n,i,j+1} \rangle|}. \quad (\text{C.31})$$

4. Compute the Berry flux for each plaquette:

$$F_{ij} = \text{Im} \ln [U_x(i, j) U_y(i+1, j) U_x^*(i, j+1) U_y^*(i, j)]. \quad (\text{C.32})$$

5. The Chern number is the sum of Berry fluxes over all plaquettes:

$$C = \frac{1}{2\pi} \sum_{i,j} F_{ij}. \quad (\text{C.33})$$

Since F_{ij} has range $(-\pi, \pi]$, C is automatically an integer (on a sufficiently fine grid).

C.8 Implementation Details for Wilson Loop and Chern Number Computation

To translate the topological invariant formulas in the main text into reproducible experimental and numerical results, this section supplements the implementation details and error control principles for Wilson loop and discrete Berry curvature algorithms in actual programming.

C.8.1 Stabilization of the Discrete Berry Curvature Algorithm

On the 2D Brillouin zone grid $\mathbf{k}_{ij}$, the Fukui–Hatsugai–Suzuki algorithm defines link variables

$$U_\mu(\mathbf{k}_{ij}) = \frac{\langle u_n(\mathbf{k}_{ij}) | u_n(\mathbf{k}_{ij} + \hat{\mu}) \rangle}{|\langle u_n(\mathbf{k}_{ij}) | u_n(\mathbf{k}_{ij} + \hat{\mu}) \rangle|}. \quad (\text{C.34})$$

The plaquette Berry curvature is

$$F_{12}(\mathbf{k}_{ij}) = \text{Im} \ln [U_1(\mathbf{k}_{ij}) U_2(\mathbf{k}_{ij} + \hat{1}) U_1^{-1}(\mathbf{k}_{ij} + \hat{2}) U_2^{-1}(\mathbf{k}_{ij})]. \quad (\text{C.35})$$

The most critical issue in implementation is not the formula itself, but the gauge jumps of adjacent eigenvectors. If a numerical diagonalization routine is called directly, the phases of eigenstates returned at different $\mathbf{k}$ points are typically random. The link variable method is stable precisely because it automatically eliminates local gauge phases, but the prerequisite is that the mode correspondence between adjacent grid points has not been disrupted by band crossings.

C.8.2 Subspace Projection for Nearly Degenerate Bands

When two or more bands are numerically nearly degenerate, computing Berry curvature band by band is often unstable. In this case, one should switch to the subspace projection matrix

$$P(\mathbf{k}) = \sum_{n \in \text{occ}} |u_n(\mathbf{k})\rangle \langle u_n(\mathbf{k})|, \quad (\text{C.36})$$

and construct the non-Abelian Wilson loop using overlap matrices. For a closed path $\mathcal{C}$, define

$$\mathcal{W}[\mathcal{C}] = \prod_{\mathbf{k} \in \mathcal{C}} M(\mathbf{k}, \mathbf{k} + \Delta\mathbf{k}), \quad M_{mn}(\mathbf{k}, \mathbf{k} + \Delta\mathbf{k}) = \langle u_m(\mathbf{k}) | u_n(\mathbf{k} + \Delta\mathbf{k}) \rangle. \quad (\text{C.37})$$

Diagonalizing $\mathcal{W}$ to obtain eigenvalues $e^{i\theta_m}$ gives the Wannier center flow. Its winding number directly corresponds to the Chern number.

C.8.3 Numerical Errors and Grid Convergence

In high-precision calculations, the following three classes of error sources should be checked simultaneously:

1. **Basis truncation error:** Finite basis truncation alters eigenfrequencies and eigenvectors; frequency convergence tests should be completed at a fixed $\mathbf{k}$ point first.
2. **Brillouin zone grid error:** If Berry curvature is highly localized (e.g., near a Dirac point), local grid refinement is needed; otherwise, although the integer Chern number may remain unchanged, the local distribution will be distorted.
3. **Band sorting error:** Sorting by frequency in nearly degenerate intervals causes mode label swaps; the projection-overlap maximum principle should be used for continuous band tracking.

In practice, one typically first uses a coarse grid to locate Berry curvature hotspots, then performs adaptive refinement to balance efficiency and accuracy.

C.8.4 Pseudocode Framework

Below is the recommended implementation framework for the Wilson loop algorithm:

1. For each k_y , sample a closed path along the k_x direction;
2. At each sampling point, obtain the eigenstate matrix $U(\mathbf{k})$ of the target subspace;
3. Compute the overlap matrix $M(\mathbf{k}_j, \mathbf{k}_{j+1}) = U^\dagger(\mathbf{k}_j)U(\mathbf{k}_{j+1})$ between adjacent points;
4. Perform polar decomposition or SVD orthogonalization on each M to remove numerical noise;
5. Multiply sequentially to obtain the Wilson matrix $\mathcal{W}(k_y)$;
6. Extract the phase spectrum $\theta_m(k_y)$ and plot the Wannier flow;
7. Count the net winding number across the reference line to obtain the Chern number or $\mathbb{Z}_2$ invariant.

C.9 Unified Implementation Framework for Wilson Loop, Polar Decomposition, and Numerical Stability

This section supplements the link between the main text and the preceding algorithm entries that has not been fully developed: *why* the Wilson loop algorithm mathematically requires polar decomposition of the overlap matrices, and how in actual code to unify gauge smoothing, subspace tracking, singular value truncation, and error estimation into a single reproducible numerical workflow.

Let a set of occupied bands (or the band group to be analyzed) provide Bloch eigenstates on a discrete wave-vector chain

$$\mathbf{k}_0 \rightarrow \mathbf{k}_1 \rightarrow \cdots \rightarrow \mathbf{k}_{N-1} \rightarrow \mathbf{k}_N = \mathbf{k}_0 + \mathbf{G} \quad (\text{C.38})$$

as $\{|u_n(\mathbf{k}_j)\rangle\}_{n=1}^{N_{\text{occ}}}$. Under the material-weighted inner product, define the overlap matrix between adjacent points

$$[\mathcal{M}_j]_{mn} = \langle u_m(\mathbf{k}_j), u_n(\mathbf{k}_{j+1}) \rangle_M. \quad (\text{C.39})$$

If the eigenstate gauge is continuous and the discrete step size is sufficiently small, then $\mathcal{M}_j$ should be close to unitary; however, finite-precision diagonalization, nearly degenerate band mixing, and arbitrary phase/subspace rotations of eigenvectors at different $\mathbf{k}$ points all cause $\mathcal{M}_j$ to deviate from unitarity. If all $\mathcal{M}_j$ are directly multiplied, numerical errors accumulate exponentially along the chain, eventually causing spurious drift in the Wannier spectrum. The standard fix is to perform polar decomposition on each $\mathcal{M}_j$

$$\mathcal{M}_j = \mathcal{Q}_j \mathcal{P}_j, \quad (\text{C.40})$$

where $\mathcal{Q}_j$ is unitary and $\mathcal{P}_j = (\mathcal{M}_j^\dagger \mathcal{M}_j)^{1/2}$ is positive semi-definite. The discrete approximation of the Wilson loop takes not $\mathcal{M}_j$ itself but its closest unitary factor $\mathcal{Q}_j$:

$$\mathcal{W} = \mathcal{Q}_{N-1} \mathcal{Q}_{N-2} \cdots \mathcal{Q}_0. \quad (\text{C.41})$$

The mathematical significance of this approach is explained by the following proposition.

Proposition C.2 (Polar decomposition gives the closest unitary approximation). *Let $\mathcal{M}$ be a non-singular complex matrix with polar decomposition $\mathcal{M} = \mathcal{Q}\mathcal{P}$. Then for any unitary matrix U , under the Frobenius norm,*

$$\|\mathcal{M} - \mathcal{Q}\|_F \leq \|\mathcal{M} - U\|_F. \quad (\text{C.42})$$

Therefore replacing $\mathcal{M}$ by $\mathcal{Q}$ is the optimal choice that minimizes the local non-unitary error while preserving the phase transport information.

Proof. By the singular value decomposition $\mathcal{M} = V\Sigma W^\dagger$, we have $\mathcal{Q} = VW^\dagger$. Any unitary matrix can be written as $U = V\tilde{U}W^\dagger$, where $\tilde{U}$ is still unitary. Then

$$\|\mathcal{M} - U\|_F = \|\Sigma - \tilde{U}\|_F. \quad (\text{C.43})$$

Since Σ is a non-negative diagonal matrix and all singular values of a unitary matrix are 1, the element-wise optimum is clearly achieved at $\tilde{U} = I$, i.e., $U = VW^\dagger = \mathcal{Q}$. Hence the proposition holds. $\square$

Algorithmically, the most robust implementation does not explicitly compute the matrix square root, but uses the singular value decomposition

$$\mathcal{M}_j = V_j \Sigma_j W_j^\dagger, \quad \mathcal{Q}_j = V_j W_j^\dagger. \quad (\text{C.44})$$

If the minimum singular value $\sigma_{\min}(\mathcal{M}_j)$ is too small, it indicates that subspace tracking has failed at that step, a band crossing has not been correctly handled, or the discrete step size is too large. In this case, one should not blindly continue computing the Wilson loop, but should trigger **adaptive repair**:

1. If $\sigma_{\min} < 10^{-10}$, first check whether there is near-degeneracy in the vicinity;

2. If so, expand the band subspace dimension N_{occ} and switch to multi-band Wilson loop;
3. If no obvious degeneracy exists, double the k grid density in that direction;
4. If the singular value remains anomalous after refinement, re-check whether the material-weighted inner product implementation is correct, especially whether $\bar{M}$ has been omitted or the standard L^2 inner product has been mistakenly used.

On this basis, a practically executable robust Wilson loop workflow can be given.

Algorithm 6 Stable_Wilson_Loop_via_Polar_Decomposition

1. Input: closed discrete path $\{\mathbf{k}_j\}_{j=0}^N$, target band group eigenstates $U(\mathbf{k}_j)$ at each point, material matrix $\bar{M}$, singular value threshold σ_{tol} .
 2. For each adjacent pair $(\mathbf{k}_j, \mathbf{k}_{j+1})$, construct the overlap matrix $\mathcal{M}_j = U(\mathbf{k}_j)^\dagger \bar{M} U(\mathbf{k}_{j+1})$.
 3. Compute SVD: $\mathcal{M}_j = V_j \Sigma_j W_j^\dagger$.
 4. If $\min \text{diag}(\Sigma_j) < \sigma_{\text{tol}}$, halt and return the diagnostic message “grid needs refinement / band group needs expansion.”
 5. Set $\mathcal{Q}_j = V_j W_j^\dagger$, and accumulate the product $\mathcal{W} = \prod_j \mathcal{Q}_j$.
 6. Diagonalize $\mathcal{W}$: $\mathcal{W}|\nu_m\rangle = e^{i2\pi\bar{x}_m}|\nu_m\rangle$.
 7. Output the Wannier centers $\bar{x}_m \in [0, 1)$ and their evolution curves as a function of the transverse parameter.
-

It is worth emphasizing that *the topological information of the Wannier spectrum is not in individual point values, but in its overall manifold structure*. For example, in a 2D TRS system, if the Wilson loop eigenphase curves over half the Brillouin zone exhibit an odd number of “partner switchings,” this corresponds to a nontrivial $\mathbb{Z}_2$ phase; in a Chern system, the net winding number of the Wannier centers with respect to the transverse momentum equals the total Chern number of the band group. Therefore, in numerical post-processing, one should not only look at the eigenphases at a single k_y cross-section, but should check whether the entire spectral flow is continuous, whether it crosses the reference branch, and whether its integer winding number is stable under grid refinement.

Finally, two practical criteria are given:

1. **Local unitarity criterion:** For all j , $\|\mathcal{Q}_j^\dagger \mathcal{Q}_j - I\|_F < 10^{-12}$;
2. **Global convergence criterion:** After doubling the number of discrete path points N , the maximum displacement of all Wannier center curves should satisfy

$$\max_m \max_{k_\perp} |\bar{x}_m^{(2N)} - \bar{x}_m^{(N)}| < 10^{-3}. \quad (\text{C.45})$$

If both are simultaneously satisfied, then the Wilson loop results are usually sufficient to support the topological determinations in the main text, without needing to resort to the more expensive full-Brillouin-zone direct Berry curvature integration.

C.10 Error Estimation and Convergence Diagnostics

The preceding sections have given the basic workflows for PWE, FHS, Wilson loop, and supercell calculations, but in actual programming, what truly determines whether results are “trustworthy” is often not the algorithm name itself, but whether error control, preconditioning, and convergence diagnostics form a coherent system. Below we organize these seemingly scattered rules of thumb into a unified framework.

C.10.1 Triple Decomposition of Discretization Error

For nearly all numerical tasks in topological photonics, the total error can be written as

$$\mathcal{E}_{\text{tot}} \leq \mathcal{E}_{\text{basis}} + \mathcal{E}_{\text{sampling}} + \mathcal{E}_{\text{solver}}, \quad (\text{C.46})$$

where:

1. $\mathcal{E}_{\text{basis}}$: basis expansion truncation error, e.g., G_{max} truncation in PWE, mesh size h in FEM, multipole order truncation in BEM;
2. $\mathcal{E}_{\text{sampling}}$: parameter-space sampling error, e.g., insufficiently fine BZ k grid, frequency sweep step too coarse, supercell thickness insufficient leading to edge state aliasing;
3. $\mathcal{E}_{\text{solver}}$: linear algebra solver error, e.g., iterative eigensolver residual not converged, polar decomposition/QR/SVD threshold set improperly, double-precision rounding error accumulation.

The significance of this decomposition is: if one does not first identify the error source, it is very easy to “refine the wrong thing”—for example, the real problem comes from eigenstate sorting confusion, yet one blindly increases the k grid from 40×40 to 200×200 , which not only fails to solve the problem but causes computational cost to explode.

Proposition C.3 (Minimal diagnostic loop for trustworthy computation). *For any target quantity Q (frequency, bandgap, Chern number, Wilson spectrum, edge state group velocity), at least the following three checks should be completed simultaneously:*

$$\begin{aligned} \text{diagnostic loop} = & \text{residual convergence} \\ & + \text{parameter-refinement convergence} \\ & + \text{symmetry/conservation-law verification.} \end{aligned} \quad (\text{C.47})$$

If any one of these is missing, the result can only be regarded as a “candidate result,” not a “verified result.”

Proof. Residual convergence only shows that the discrete linear algebra problem has been solved accurately, but does not show that the discretization itself is correct; parameter refinement convergence shows that the discretization is stable, but does not guarantee the absence of systematic sign errors in the code; symmetry/conservation law verification can rule out a large class of “seemingly converged but actually wrong” situations, such as obtaining $\omega(\mathbf{k}) \neq \omega(-\mathbf{k})$ without a magnetic field, or a non-integer Chern number. All three are indispensable. $\square$

C.10.2 Residuals and A Posteriori Error Estimates for Eigenvalue Problems

Let the discrete eigenpair satisfy

$$A_h \mathbf{x}_h = \lambda_h B_h \mathbf{x}_h, \quad (\text{C.48})$$

where $A_h = A_h^\dagger$, $B_h = B_h^\dagger > 0$. Define normalization $\mathbf{x}_h^\dagger B_h \mathbf{x}_h = 1$ and residual

$$\mathbf{r}_h = A_h \mathbf{x}_h - \lambda_h B_h \mathbf{x}_h. \quad (\text{C.49})$$

Then, when the spectral gap is not too small, there exists a constant C such that

$$|\lambda_h - \lambda| \leq C \|\mathbf{r}_h\|_{B_h^{-1}}. \quad (\text{C.50})$$

Here $\|\mathbf{r}\|_{B_h^{-1}}^2 = \mathbf{r}^\dagger B_h^{-1} \mathbf{r}$. This theorem tells us: printing “solver converged” is far from sufficient; the residual norm must be explicitly output.

In PWE, if a dense Hermitian diagonalizer is used directly, the residual is typically near machine precision; in large-scale 3D problems, iterative solvers such as LOBPCG, ARPACK, or Jacobi–Davidson are commonly used, and for each target frequency band the following diagnostic table should be saved:

Mode n	$\omega_n a / 2\pi c$	Relative residual $\ \mathbf{r}_n\ / \ A_h \mathbf{x}_n\ $	Remark
1	0.28413	1.1×10^{-11}	Pass
2	0.31742	2.3×10^{-10}	Pass
3	0.48107	4.8×10^{-7}	Needs review

Empirically, the occupied-band subspace involved in topological invariant calculations should have residuals better than 10^{-8} ; if the band gap is extremely small, even stricter standards are needed.

C.10.3 Why Preconditioners Are Especially Important in Maxwell Problems

The ill-conditioning of discrete Maxwell matrices typically comes from two sources:

1. High refractive-index contrast makes the spectral span of the material matrix very large;
2. The curl operator has a near-null space of long-wavelength components, causing iterative solvers to stagnate in “near-gradient directions.”

Therefore a preconditioner is not an optional “acceleration plugin” but a core component that determines whether the algorithm is scalable.

Diagonal preconditioning in PWE. For the matrix elements

$$M_{\mathbf{G}\mathbf{G}'} = (\mathbf{k} + \mathbf{G}) \cdot \eta_{\mathbf{G}-\mathbf{G}'} (\mathbf{k} + \mathbf{G}') \quad (\text{C.51})$$

the simplest yet effective preconditioner retains only the diagonal part:

$$P_{\mathbf{G}\mathbf{G}'} = \delta_{\mathbf{G}\mathbf{G}'} (|\mathbf{k} + \mathbf{G}|^2 + \sigma), \quad (\text{C.52})$$

where $\sigma > 0$ prevents small denominators. The idea is simple: in the high- $|\mathbf{G}|$ region, the dominant stiffness is already of Laplacian-type growth, so the diagonal terms capture the hardest spectral directions.

H(curl) preconditioning in FEM. For edge-element discretization of the Maxwell eigenvalue problem, a more natural approach is the Hiptmair–Xu-type auxiliary space preconditioner, which decomposes the $H(\text{curl})$ problem into a vector Laplacian problem plus a gradient correction. The deep idea behind this preconditioner is completely consistent with the operator perspective of Chapter 1 of this book: first separate the “truly propagating curl degrees of freedom” from the “gauge-redundant/near-gradient degrees of freedom,” then handle them separately.

Remark (Why topological calculations are more sensitive to ill-conditioning than ordinary band structure). Ordinary band structure plots only require frequency values to look “roughly right”; but Chern numbers, Wilson spectrum winding, and edge state group velocities are all extremely sensitive to eigenstate phases and overlaps between adjacent k points. A 10^{-5} -level eigenstate contamination invisible to the eye can be amplified near degeneracies into an incorrect topological classification. Therefore topological calculations typically require higher standards for preconditioning and residual control than “just plotting bands.”

C.10.4 Polar Decomposition Error Control in Wilson Loops

Let the overlap matrix between adjacent k points be F_i . Under an ideal continuous gauge, F_i should be close to unitary, but discretization, numerical errors, and near-degeneracy all cause deviations. Let its SVD be

$$F_i = W_i \Sigma_i V_i^\dagger, \quad \Sigma_i = \text{diag}(\sigma_1, \dots, \sigma_N), \quad (\text{C.53})$$

then the unitary part from polar decomposition is

$$U_i = W_i V_i^\dagger. \quad (\text{C.54})$$

Define the non-unitarity deviation

$$\epsilon_i^{\text{uni}} = \|F_i^\dagger F_i - I\|_F. \quad (\text{C.55})$$

If ϵ_i^{uni} is too large, the k step size is too coarse or subspace tracking has failed. A practical rule of thumb:

$$\max_i \epsilon_i^{\text{uni}} < 10^{-6} \implies \text{Wilson spectrum is usually trustworthy.} \quad (\text{C.56})$$

If above 10^{-4} , one should first refine the k grid rather than “pushing through” the Wilson matrix multiplication.

Furthermore, the Wilson matrix product

$$\mathcal{W}_N = U_N U_{N-1} \cdots U_1 \quad (\text{C.57})$$

itself may accumulate rounding errors from the long chain multiplication. In practice, periodic re-unitarization should be applied, e.g., performing a QR decomposition every 10 matrix mul-

tifications and using Q as the new accumulated matrix. This is the same numerical idea as re-orthogonalization in time-stepping algorithms.

C.10.5 Spurious Edge State Criteria in Supercell Calculations

In topological edge state calculations, the most dangerous misjudgment comes from “finite-size-induced spurious edge states.” A genuine topological edge state should satisfy at least the following three criteria:

1. **Spatial localization:** The mode intensity decays exponentially in the interface normal direction, rather than being uniformly spread across the entire supercell;
2. **Size robustness:** After increasing the supercell thickness, its frequency undergoes only exponentially small corrections;
3. **Phase connectivity:** It evolves continuously with the parallel momentum $k_{\parallel}$ and connects to adjacent projected bulk bands, rather than floating in isolation within the bandgap.

To this end, for each candidate edge mode it is recommended to record

$$y_c = \frac{\int y u(\mathbf{r}) d\mathbf{r}}{\int u(\mathbf{r}) d\mathbf{r}}, \quad \xi^{-2} = \frac{\int (y - y_c)^2 u(\mathbf{r}) d\mathbf{r}}{\int u(\mathbf{r}) d\mathbf{r}}, \quad (\text{C.58})$$

where $u(\mathbf{r})$ can be taken as the energy density. y_c identifies which interface it is located at, and ξ characterizes the localization length. If increasing the supercell thickness leaves ξ and the frequency nearly unchanged while the splitting between modes at the upper and lower interfaces rapidly decreases, then the edge state identification can be considered reliable.

C.10.6 Comparison of PWE, FEM, and FDTD Error Sources in Topological Calculations

Different numerical methods have different error structures. Without understanding the error types of the method itself, it is easy to misidentify “method differences” as “novel physical phenomena.”

Method	Dominant error source	Advantage for topological calculations	Main risk
PWE	Basis truncation, interface Gibbs oscillation, inverse rule misuse	Frequency-domain eigenvalue problem directly suited for bands and Berry curvature	Slow convergence for high dielectric contrast and complex boundaries
FEM	Mesh discretization error, numerical dispersion, spurious modes	Flexible for complex geometries and anisotropic materials	Careful treatment of divergence constraints and boundary conditions needed
FDTD	Time discretization error, absorbing boundary reflections, long-time accumulation error	Suitable for pulse propagation, edge state transport, and open systems	Difficult to directly extract high-precision topological invariants

For most topological invariant calculations in this textbook, the recommended strategy is:

$$\begin{aligned}
& \text{PWE/FEM for eigenstates} \\
& \implies \text{FHS or Wilson loop for topological quantities} \\
& \implies \text{FDTD for transport verification.}
\end{aligned} \tag{C.59}$$

This division of “frequency domain for topology, time domain for transport” minimizes the mixing of method errors.

C.10.7 Executable Unified Diagnostic Workflow

Combining the above discussions, the numerical diagnostics for Wilson loop and surface state calculations can be written as the following workflow:

1. Fix the target frequency range and choose the occupied subspace dimension M ;
2. Compute the overlap matrices F_j between adjacent k points and record Q_C ;
3. If $Q_C < Q_{\min}$, refine the grid or expand the subspace;
4. Construct the unitary transport matrices $\tilde{U}_j$ via SVD or polar decomposition;
5. Compute Wilson eigenphases on both N_k and $2N_k$ grids, and check (C.83);
6. If studying surface states, solve for both supercell thicknesses L and $2L$, and check (C.94);
7. Only when subspace quality, discretization convergence, and edge state credibility all pass simultaneously may the final topological conclusion be reported.

The philosophy of this workflow is very simple: **The integrality of topological invariants does not automatically guarantee computational trustworthiness**; an integer result can also come from the wrong subspace, wrong boundary conditions, or wrong convergence regime.

Truly reliable numerical topology must elevate “computing an integer” to “proving that this integer is stable with respect to grid, basis, and boundary conditions.”

C.11 Numerical Tools for 3D Topological Photonics

This section supplements two classes of numerical techniques most commonly used in 3D topological photonics but often abbreviated to “black-box steps” in textbooks: first, the surface Green’s function iteration for semi-infinite geometries, used to directly extract surface state dispersion, local density of states, and Fermi arcs; second, the discrete flux integration of Weyl point topological charges, used for rigorous determination of candidate band crossings in the 3D Brillouin zone. We adhere to the Maxwell-first principle of this book: all algorithms should trace back to the spectral properties of the frequency-domain Maxwell operator, rather than merely copying notation from tight-binding electronic systems.

C.11.1 From Block-Tridiagonal Discrete Maxwell Operator to Surface Green’s Function

Consider a structure layered along z , with Bloch periodicity maintained in the directions (x, y) parallel to the surface. For fixed surface momentum $\mathbf{k}_{\parallel} = (k_x, k_y)$ and frequency ω , after PWE, FEM, or finite-difference discretization, the Maxwell equation can be written in block-tridiagonal form

$$\mathcal{L}(\mathbf{k}_{\parallel}, \omega) = \begin{pmatrix} A_0 & B_0 & 0 & 0 & \cdots \\ C_0 & A_1 & B_1 & 0 & \cdots \\ 0 & C_1 & A_2 & B_2 & \cdots \\ 0 & 0 & C_2 & A_3 & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \quad (\text{C.60})$$

where A_j describes the internal degrees of freedom of the j -th layer, and B_j, C_j describe inter-layer coupling. If the bulk region becomes periodic after a certain number of layers, there exists an integer j_* such that

$$A_j = A, \quad B_j = B, \quad C_j = C, \quad j \geq j_*. \quad (\text{C.61})$$

This is precisely the numerical entry point for a semi-infinite crystal: the surface layers can be arbitrarily complex, but the deep interior must return to the periodic bulk material.

For a given ω , we define the analytically continued resolvent

$$G(\mathbf{k}_{\parallel}, \omega) = [\mathcal{L}(\mathbf{k}_{\parallel}, \omega) + i\eta \mathbb{I}]^{-1}, \quad \eta > 0, \quad (\text{C.62})$$

where η is a small regularization parameter. The surface Green’s function is its upper-left block

$$g_s(\mathbf{k}_{\parallel}, \omega) = (G)_{00}. \quad (\text{C.63})$$

Once g_s is obtained, the surface local density of states can be written as

$$\rho_s(\mathbf{k}_{\parallel}, \omega) = -\frac{1}{\pi} \text{Im Tr} g_s(\mathbf{k}_{\parallel}, \omega). \quad (\text{C.64})$$

The Fermi arc or iso-frequency arc corresponds precisely to the ridge of peaks of ρ_s within the projected bandgap.

C.11.2 Schur Complement and Self-Energy Fixed-Point Equation

Due to the block-tridiagonal structure, g_s can be obtained by recursive Schur complement. If we temporarily separate layer 0 from all subsequent layers, then

$$g_s = \left[A_0 + i\eta \mathbb{I} - B_0 \Sigma_1 C_0 \right]^{-1}, \quad (\text{C.65})$$

where Σ_1 is the effective propagator of the “remaining semi-infinite body” projected onto layer 1. When the deep bulk region is strictly periodic, Σ_1 converges to the translationally invariant fixed point g_b , satisfying

$$g_b = \left[A + i\eta \mathbb{I} - B g_b C \right]^{-1}. \quad (\text{C.66})$$

Equivalently, this can be written as the nonlinear matrix equation

$$(A + i\eta \mathbb{I})g_b - B g_b C g_b = \mathbb{I}. \quad (\text{C.67})$$

This shows that the surface Green’s function problem is essentially a matrix Riccati-type fixed-point problem. For frequencies falling within the bulk bandgap, the physical solution corresponds to the branch decaying along $+z$; the task of the numerical algorithm is to stably select this branch, rather than mixing the growing solution into the surface propagator.

Remark (Why one cannot simply truncate a finite-thickness supercell to replace the semi-infinite body). For Weyl photonic crystals, Fermi arcs are typically localized on one surface, with decay lengths that vary with surface momentum. If a finite-thickness supercell is directly used to replace the semi-infinite body, the two surfaces will have exponentially small but non-negligible coupling, distorting the originally open iso-frequency arcs into weakly hybridized closed loops. The advantage of the semi-infinite surface Green’s function is precisely that it eliminates the existence of the other surface by construction, and therefore provides a cleaner single-surface spectrum.

C.11.3 Layer Decimation and Sancho–Rubio-Type Doubling Iteration

If the bulk region is completely periodic from layer j_* onward, iterative decimation can be performed on the repeating blocks. Define the effective blocks $(\alpha_n, \beta_n, \gamma_n)$ at the n -th iteration round, initialized as

$$\alpha_0 = A, \quad \beta_0 = B, \quad \gamma_0 = C. \quad (\text{C.68})$$

By eliminating even-numbered layers, the doubling recursion is obtained:

$$\alpha_{n+1} = \alpha_n - \beta_n \alpha_n^{-1} \gamma_n - \gamma_n \alpha_n^{-1} \beta_n, \quad (\text{C.69})$$

$$\beta_{n+1} = -\beta_n \alpha_n^{-1} \beta_n, \quad (\text{C.70})$$

$$\gamma_{n+1} = -\gamma_n \alpha_n^{-1} \gamma_n. \quad (\text{C.71})$$

When $\|\beta_n\|$ and $\|\gamma_n\|$ decay below a threshold, the remaining propagation has been effectively localized, and one can take

$$g_b \approx \alpha_n^{-1}, \quad \text{and back-substitute to obtain } g_s. \quad (\text{C.72})$$

Strictly speaking, Eqs. (C.69)–(C.71) are not the only possible form; different discretization bases and gauges yield equivalent variants; but their common core is: each iteration round doubles the propagation distance, so for in-gap decaying modes the convergence is exponentially fast.

Proposition C.4 (Credible convergence criterion for the doubling iteration). *Suppose α_n remains invertible throughout, and there exists some n_0 such that for all $n \geq n_0$,*

$$\max\{\|\beta_n\|, \|\gamma_n\|\} < \varepsilon_{\text{hop}}, \quad (\text{C.73})$$

where $\varepsilon_{\text{hop}} \ll 1$ is a user-set threshold. Then the contribution of the omitted far layers to g_b is at most $O(\varepsilon_{\text{hop}})$, and the surface spectral peak positions constructed from α_n^{-1} have first-order credibility.

Proof. After each decimation round, β_n and γ_n represent the retained effective forward/backward inter-layer couplings. When their operator norms are already much smaller than the spectral scale of α_n , any further propagation into deeper layers can only enter the Schur complement through at least one insertion of β_n or γ_n , and therefore the correction to g_b is a higher-order small quantity. If $\|\alpha_n^{-1}\|$ has a uniform upper bound, the Neumann series gives the error estimate $\|\delta g_b\| \lesssim C\varepsilon_{\text{hop}}$. Hence Eq. (C.73) serves as a practical a posteriori convergence diagnostic. $\square$

In implementation, one should also record

$$r_n = \|\mathbb{I} - \alpha_n g_b^{(n)}\|, \quad (\text{C.74})$$

as well as the fixed-point residual

$$R_n = \|g_b^{(n)} - [A + i\eta\mathbb{I} - Bg_b^{(n)}C]^{-1}\|. \quad (\text{C.75})$$

Only when ε_{hop} , r_n , and R_n are all simultaneously small enough should we interpret the obtained surface peaks as credible topological surface states, rather than numerical iteration residuals or pseudo-spectral effects.

C.11.4 Discrete Flux Integration of Weyl Point Topological Charge

Suppose a candidate point $\mathbf{k}_c$ has been found in the 3D Brillouin zone via a minimum-gap search. To determine whether it is truly a Weyl point, rather than an accidental near-degeneracy or numerical noise, one can take a small closed surface S around $\mathbf{k}_c$ and perform a discrete flux integration of the lower-band Berry curvature:

$$\chi = \frac{1}{2\pi} \oint_S \boldsymbol{\Omega}(\mathbf{k}) \cdot d\mathbf{S}. \quad (\text{C.76})$$

If $\chi = \pm 1$, the point is a simple Weyl point; if $\chi = 0$, it is not an isolated monopole source.

Numerically, the most robust method is not to directly finite-difference the continuous Berry curvature, but to discretize the closed surface into six faces (cube) or a latitude-longitude grid (sphere), and use FHS-type discrete flux on each small face. Taking the six-face cube partition as an example, the discrete Berry phase on each face is

$$\Phi_{\square} = \text{Im} \ln [U_1 U_2 U_3^{-1} U_4^{-1}], \quad (\text{C.77})$$

and the total charge is

$$\chi_{\text{disc}} = \frac{1}{2\pi} \sum_{\square \in S} \Phi_{\square}. \quad (\text{C.78})$$

Here the U_j must be constructed using the material-weighted inner product; otherwise the gauge geometry of the photonic Bloch states will be incorrectly projected under the Euclidean inner product.

In practice, one should not report just a single numerical value, but should output a stability plateau of the charge versus the enclosing size ℓ . If 0.97 or 1.04 is obtained only at one isolated size, and immediately drifts to 0.4 or 1.8 upon slight adjustment of ℓ , then the candidate point cannot yet be confirmed as a genuine Weyl point.

C.12 Error Control in Wilson-Loop Calculations

The Wilson loop algorithm given earlier explained how to extract Wannier centers from discrete overlap matrices, but in actual 3D topological photonic crystal calculations, *obtaining a curve* does not mean *obtaining a trustworthy topological conclusion*. The real difficulty of numerical implementation lies not in the matrix chain multiplication itself, but in how the following three classes of errors are rigorously distinguished:

1. **Discretization error:** the path-ordered exponential approximation error due to finite k -space step size;
2. **Subspace error:** truncation of nearly degenerate bands causing the occupied subspace to rotate or leak;
3. **Spectral pollution:** spurious surface states arising from finite supercells, finite basis sets, or finite-thickness boundary conditions.

If these three types of errors are not discussed separately, the so-called “Wilson band winding” may be nothing more than a numerical artifact. Therefore, this section establishes a framework of a **posteriori error estimation** and **spectral pollution diagnosis**, enabling the reader to assign a credibility level to each numerical topological conclusion.

C.12.1 Discrete Wilson Loop and the Continuous Limit

Let the continuous Wilson loop be defined as

$$\mathcal{W}[\mathcal{C}] = \mathcal{P} \exp \left[i \int_{\mathcal{C}} A_{\mu}(\mathbf{k}) dk_{\mu} \right], \quad (\text{C.79})$$

where $A_{\mu} = -i \langle u | \partial_{k_{\mu}} u \rangle$ is the non-Abelian Berry connection on the occupied subspace in the convention of Chapter 4. Discretizing the closed path $\mathcal{C}$ into the point sequence

$\{\mathbf{k}_0, \mathbf{k}_1, \dots, \mathbf{k}_{N-1}, \mathbf{k}_N = \mathbf{k}_0 + \mathbf{G}\}$, the discrete Wilson loop is written as

$$\mathcal{W}_N = \tilde{U}_{N-1} \tilde{U}_{N-2} \cdots \tilde{U}_1 \tilde{U}_0, \quad (\text{C.80})$$

where $\tilde{U}_j$ is the unitary part of the overlap matrix F_j after polar decomposition or SVD correction. As the step size $\Delta k = \max_j |\mathbf{k}_{j+1} - \mathbf{k}_j|$ tends to zero, $\mathcal{W}_N$ should converge to the continuous Wilson loop.

Proposition C.5 (First-order estimate of Wilson discretization error). *Let the occupied subspace projection operator $P(\mathbf{k})$ be twice continuously differentiable on the path $\mathcal{C}$, and let the minimum direct bandgap satisfy $g_{\min} > 0$. Then there exists a constant $C_{\mathcal{C}}$ related to the path length and $\|\partial_k P\|$, $\|\partial_k^2 P\|$, such that*

$$\|\mathcal{W}_N - \mathcal{W}[\mathcal{C}]\| \leq C_{\mathcal{C}} \Delta k + O(\Delta k^2). \quad (\text{C.81})$$

In particular, if the grid is refined by a factor of two, the Wilson eigenphase error should approximately halve.

Proof. Writing the overlap between adjacent k points of the occupied subspace as

$$F_j = \mathbf{1} + iA_{\mu}(\mathbf{k}_j) \Delta k_j + O(\Delta k_j^2), \quad (\text{C.82})$$

the unitary part $\tilde{U}_j$ from polar decomposition differs from F_j only by second-order terms. Therefore the discrete product (C.80) is a first-order Lie–Trotter approximation of the continuous path-ordered exponential. Since the projection operator is smooth along the entire path and the bandgap does not close, the local errors accumulate linearly in the product, yielding the estimate (C.81). A more rigorous proof can be given via Dyson expansion and discrete product integral theory. $\square$

The true value of Proposition C.5 lies not in the formal error order, but in providing a simple and actionable **a posteriori criterion**: compute the Wilson spectrum on the same path with N_k and $2N_k$ grids; if all eigenphases satisfy

$$\max_j \min_{m \in \mathbb{Z}} |\theta_j^{(2N_k)} - \theta_j^{(N_k)} + 2\pi m| < \delta_{\theta}, \quad (\text{C.83})$$

then the discretization error is controlled within δ_{θ} . In textbook practice, $\delta_{\theta} = 10^{-3} \sim 10^{-2}$ is commonly used.

C.12.2 Nearly Degenerate Subspaces and Parallel Transport Gauge

In 3D Weyl systems, Dirac systems, and linearized nonlinear BdG problems, a common difficulty is that the target subspace is not composed of a single isolated band, but of several nearly degenerate bands. If one still computes phases “band by band,” violent gauge jumps occur near band crossings. Therefore the computation object must be elevated from individual eigenstates to the **occupied subspace projection**.

Let $\mathcal{S}(\mathbf{k}) = \text{span}\{|u_1(\mathbf{k})\rangle, \dots, |u_M(\mathbf{k})\rangle\}$ be the M -dimensional target subspace. Define the overlap matrix between adjacent points

$$[F_j]_{mn} = \langle u_m(\mathbf{k}_j) | \hat{M} | u_n(\mathbf{k}_{j+1}) \rangle, \quad m, n = 1, \dots, M. \quad (\text{C.84})$$

If $\sigma_{\min}(F_j)$ is very small, it indicates that the target subspaces at adjacent grid points have nearly lost correspondence; this typically means one of three possibilities:

1. A genuine bandgap closure near the path, with the topological phase undergoing a change;
2. The grid is too coarse to resolve the rapidly varying Berry connection;
3. Eigenstate sorting errors have caused inconsistent subspace selection at adjacent points.

Therefore, the minimum singular value $\sigma_{\min}(F_j)$ is itself an important diagnostic quantity.

Definition C.6 (Wilson subspace quality factor). Define the subspace quality factor on the path $\mathcal{C}$:

$$Q_{\mathcal{C}} = \min_j \sigma_{\min}(F_j). \quad (\text{C.85})$$

If $Q_{\mathcal{C}} \ll 1$, the discrete Wilson spectrum is no longer reliable, and the grid must be refined or the target subspace expanded.

Empirically,

$$Q_{\mathcal{C}} > 10^{-2} \quad (\text{C.86})$$

is usually sufficient for stable Wilson phase ordering; when $Q_{\mathcal{C}} < 10^{-6}$, one has typically entered the numerically ill-conditioned regime.

To further reduce gauge fluctuations, the **parallel transport gauge** can be adopted. Its idea is: at each step, use polar decomposition

$$F_j = U_j H_j, \quad (\text{C.87})$$

take the unitary part U_j as the optimal transport map from $\mathbf{k}_j$ to $\mathbf{k}_{j+1}$, and update the basis at the next point accordingly, making the overlap matrix as close to the identity as possible. Under this gauge, numerical phase jumps are minimized, and continuous tracking of Wannier centers is significantly stabilized.

Remark (Parallel transport is not “beautifying the image”). The goal of the parallel transport gauge is not to make plots smoother, but to reduce the gauge ill-conditioning number on the discrete path. As long as the final Wilson loop is constructed from subspace projections, its topological information is gauge-independent; but different gauges dramatically affect the *numerical stability* at finite precision. Therefore, parallel transport is a numerical analysis issue, not a physical definition issue.

C.12.3 Comparison of Polar Decomposition, QR, and SVD Correction

When unitarizing the overlap matrix F_j , the three most common strategies are:

1. **Polar decomposition:** $F_j = U_j H_j$, take U_j ;
2. **QR decomposition:** $F_j = Q_j R_j$, take Q_j ;
3. **SVD correction:** $F_j = W_j \Sigma_j V_j^\dagger$, take $W_j V_j^\dagger$.

From the topological perspective, all three attempt to extract the “closest unitary matrix to F_j ”; from the numerical analysis perspective, their behavior in ill-conditioned cases is not the same.

Method	Mathematical definition	Advantages	Risks
Polar decomposition	$U = F(F^\dagger F)^{-1/2}$	Directly corresponds to the closest unitary matrix problem; clear gauge meaning	Requires stable matrix square root computation
QR decomposition	$F = QR$, take Q	Simple implementation, fast	Sensitive to column ordering and pivoting; not robust enough near degeneracies
SVD correction	$F = W\Sigma V^\dagger$, take $WV^\dagger$	Most robust when singular values approach zero; can directly monitor ill-conditioning	Relatively higher computational cost

Proposition C.7 (Optimality of SVD unitarization). *Under the Frobenius norm, $W_j V_j^\dagger$ is the unitary matrix closest to F_j , i.e.,*

$$\|F_j - W_j V_j^\dagger\|_F = \min_{U^\dagger U = \mathbf{1}} \|F_j - U\|_F. \quad (\text{C.88})$$

Therefore, when the overlap matrix is ill-conditioned, SVD correction is the most robust default choice.

Proof. Let $F_j = W_j \Sigma_j V_j^\dagger$ be the SVD, and let $\tilde{U}$ be any unitary matrix. By unitary invariance,

$$\|F_j - \tilde{U}\|_F = \|\Sigma_j - W_j^\dagger \tilde{U} V_j\|_F. \quad (\text{C.89})$$

Since $W_j^\dagger \tilde{U} V_j$ is still unitary, the problem reduces to finding the closest unitary matrix to the non-negative diagonal matrix Σ_j . Element-wise minimization shows the optimal choice is the identity matrix, so the optimal $\tilde{U} = W_j V_j^\dagger$. $\square$

In practice, the recommended priority order is:

$$\text{SVD correction} \succ \text{Polar decomposition} \succ \text{QR correction}. \quad (\text{C.90})$$

Here “ $\succ$ ” does not mean absolute mathematical superiority, but rather the engineering priority in photonic topological calculations considering stability, diagnosability, and implementation complexity.

C.12.4 Credible Surface State Criteria in Supercell Calculations

For numerical computations of Fermi arcs, topological surface Dirac cones, and interface states, a common approach is to truncate along one direction using a supercell, approximating the semi-infinite boundary problem as a large but finite matrix eigenvalue problem. However, finite supercells have a core risk: **spectral pollution**. Spectral pollution means that an eigenvalue

of the finite matrix falls within the bulk bandgap, but its corresponding eigenstate is not a genuine localized surface state, but rather a spurious state created by finite thickness, boundary conditions, or basis truncation.

To distinguish genuine surface states from spurious ones, the following three a posteriori indicators are introduced:

1. **Edge localization**

$$\Lambda_{\text{edge}} = \frac{\int_0^{y_c} w(y) dy + \int_{L-y_c}^L w(y) dy}{\int_0^L w(y) dy}, \quad (\text{C.91})$$

where $w(y)$ is the projection of the mode energy density along the normal direction, and y_c is the boundary layer thickness;

2. **Thickness stability**

$$\delta\omega_{L \rightarrow 2L} = |\omega^{(2L)} - \omega^{(L)}|; \quad (\text{C.92})$$

3. **Projection residual**

$$r = \frac{\|\hat{\Theta}_L u_L - \lambda_L u_L\|}{\|u_L\|}, \quad (\text{C.93})$$

where $\hat{\Theta}_L$ is the finite supercell discrete operator.

Definition C.8 (Empirical criterion for credible surface states). If an in-gap state simultaneously satisfies

$$\Lambda_{\text{edge}} > 0.8, \quad \delta\omega_{L \rightarrow 2L} < 10^{-3}, \quad r < 10^{-8}, \quad (\text{C.94})$$

then it can be classified as a “high-credibility” surface state; if only one or two conditions are met, boundary-condition dependence must be further investigated.

Here Λ_{edge} rules out bulk state contamination, $\delta\omega_{L \rightarrow 2L}$ rules out finite-size coupling spurious states, and r checks whether the discrete eigenvalue problem has been truly solved. All three are indispensable.

C.12.5 From “Program Output” to “Textbook Figures”: Selection Criteria

Not every computed dataset deserves to be placed in a textbook. To ensure that figures have both pedagogical value and credibility, the following four-step selection is recommended:

1. First pass residual and refinement tests to obtain “numerically credible” candidate data;
2. Then check whether it manifests a clear physical law (e.g., bandgap opening, winding, unidirectional edge states);
3. Label every curve or color map in the figure with parameters, normalization units, and sampling method;
4. If a result holds only in an extremely narrow parameter window, this fragility must be stated in the text, not packaged as a “universal law.”

This step is particularly applicable to topological photonics, because the field easily produces “images that look like topological phenomena but are actually just ordinary cavity-mode interference” misreadings. What textbook writing needs is not pretty pictures, but *pictures that withstand reproducibility and diagnostic chain scrutiny*.

Chapter D

Supplement on Maxwell Spectral Theory

This appendix collects advanced spectral-theoretic tools that were moved out of Chapter 2, including the Green’s function of the Maxwell operator, the Purcell effect, the variational principle, and the Courant–Fischer principle. These topics are not essential for the main line of topological photonics derivations, but have independent value in applications such as inverse design, spontaneous emission enhancement, and bandgap optimization.

D.1 Green’s Function of the Maxwell Operator

Let the eigenmodes of $\hat{N}\mathbf{f}_n = \omega_n \bar{M}\mathbf{f}_n$ form a complete orthogonal set $\{\mathbf{f}_n\}$ (Chapter 2, theorem 2.11); then the Green’s function of the Maxwell operator has the spectral representation:

$$\mathbb{G}(\mathbf{r}, \mathbf{r}'; \omega) = \sum_n \frac{\mathbf{f}_n(\mathbf{r}) \mathbf{f}_n^\dagger(\mathbf{r}') \bar{M}(\mathbf{r}')}{(\omega - \omega_n) \langle \mathbf{f}_n, \mathbf{f}_n \rangle_M}, \quad (\text{D.1})$$

where the summation runs over all eigenmodes (including both positive and negative frequency branches). If the modes are normalized by $\langle \mathbf{f}_n, \mathbf{f}_n \rangle_M = 1$, the denominator factor is of course absent. The poles of the Green’s function are the eigenfrequencies ω_n , and the residues give the projection operators onto the corresponding modes.

D.2 Local Density of States and the Purcell Effect

The **local density of states** (LDOS) is defined as

$$\rho_E(\mathbf{r}, \omega) = \frac{2\omega}{\pi c^2} \text{Im}[\text{Tr } \mathbb{G}_{EE}(\mathbf{r}, \mathbf{r}; \omega + i0^+)], \quad (\text{D.2})$$

for the electric LDOS, where $\mathbb{G}_{EE}$ is the electric–electric block of the retarded Green tensor. Overall signs and prefactors depend on the Green-function convention; the invariant physical statement is that the emission rate is proportional to the imaginary part of the retarded Green tensor projected along the dipole orientation. Near a resonant frequency ω_n , the LDOS is

dominated by that mode:

$$\rho_E(\mathbf{r}, \omega) \approx \frac{|\mathbf{E}_n(\mathbf{r})|^2}{\langle \mathbf{f}_n, \mathbf{f}_n \rangle_M} \frac{\gamma_n/\pi}{(\omega - \omega_n)^2 + \gamma_n^2}, \quad (\text{D.3})$$

where γ_n is the mode linewidth (in a closed system, $\gamma_n \rightarrow 0^+$).

The **Purcell factor** quantifies the enhancement of the spontaneous emission rate by the photonic density of states:

$$F_P = \frac{\rho_E(\mathbf{r}_0, \omega_0)}{\rho_{E,0}(\omega_0)} = \frac{3}{4\pi^2} \left(\frac{\lambda}{n} \right)^3 \frac{Q}{V_{\text{eff}}}, \quad (\text{D.4})$$

where Q is the quality factor, V_{eff} is the effective mode volume, and $\rho_{E,0}$ is the free-space electric LDOS.

D.3 Variational Principle and the Rayleigh Quotient

For the self-adjoint generalized eigenvalue problem $\hat{N}\mathbf{f} = \omega \bar{M}\mathbf{f}$, the **Rayleigh quotient** is defined as

$$R[\mathbf{f}] = \frac{\langle \mathbf{f}, \hat{N}\mathbf{f} \rangle_{L^2}}{\langle \mathbf{f}, \mathbf{f} \rangle_M}. \quad (\text{D.5})$$

By self-adjointness, $R[\mathbf{f}]$ is stationary at the eigenmodes $\mathbf{f}_n$ with stationary value ω_n . However, because the first-order Maxwell spectrum contains positive, negative, and possible zero-frequency branches, the lowest positive-frequency eigenvalue is not obtained by an unconstrained global minimum of $R[\mathbf{f}]$. A minimax statement is valid only after restricting to the positive-frequency spectral subspace, or after passing to a positive second-order formulation such as the nonmagnetic **H** equation. In that restricted setting one may write schematically

$$\omega_1^+ = \min_{\mathbf{f} \in \mathcal{H}_+, \mathbf{f} \neq 0} R[\mathbf{f}], \quad (\text{D.6})$$

where $\mathcal{H}_+$ denotes the positive-frequency subspace. This qualification is essential for variational estimates of photonic bandgaps.

D.4 The Courant–Fischer Principle and Bandgap Sensitivity

Theorem D.1 (Courant–Fischer minimax principle). *The n -th eigenfrequency satisfies*

$$\omega_n^+ = \min_{\substack{V \subset \mathcal{H}_+ \\ \dim V = n}} \max_{\mathbf{f} \in V, \mathbf{f} \neq 0} R[\mathbf{f}], \quad (\text{D.7})$$

where V ranges over all n -dimensional subspaces of the positive-frequency spectral subspace. Equivalently, one can formulate the usual Courant–Fischer theorem for a positive second-order Maxwell operator.

A direct corollary in nondispersive, nonmagnetic scalar media is: increasing the local dielectric constant weakly lowers the eigenfrequencies of the positive second-order Maxwell problem, while decreasing it weakly raises them, assuming the ordering of the relevant isolated bands is

maintained. The first-order frequency sensitivity is

$$\frac{\delta\omega_n}{\omega_n} = -\frac{1}{2} \frac{\int \delta\varepsilon(\mathbf{r}) |\mathbf{E}_n(\mathbf{r})|^2 d^3r}{\int \varepsilon(\mathbf{r}) |\mathbf{E}_n(\mathbf{r})|^2 d^3r}. \quad (\text{D.8})$$

This formula is the source of gradient information in inverse design of photonic crystals. For anisotropic, magnetic, bianisotropic, or dispersive media the perturbation formula must be replaced by the corresponding energy-weighted expression involving the full material matrix (or the Brillouin energy density in dispersive systems).

Bibliography

- [1] S. Raghu and F. D. M. Haldane. Analogs of quantum-Hall-effect edge states in photonic crystals. *Physical Review A*, 78(3):033834, 2008.
- [2] 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):013904, 2008.
- [3] Ling Lu, John D. Joannopoulos, and Marin Soljačić. Topological photonics. *Nature Photonics*, 8(11):821–829, 2014.
- [4] 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):015006, 2019.
- [5] Hannah Price, Y. D. Chong, Alexander B. Khanikaev, Henning Schomerus, Lukas J. Maczewsky, Mark Kremer, Matthias Heinrich, Alexander Szameit, Oded Zilberberg, Yihao Yang, Baile Zhang, Alexander Cerjan, et al. Roadmap on topological photonics. *Journal of Physics: Photonics*, 4(3):032501, 2022.
- [6] Zheng Wang, Y. Chong, John D. Joannopoulos, and Marin Soljačić. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature*, 461(7265):772–775, 2009.
- [7] 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, 2013.
- [8] Li-Heng Wu and Xiao Hu. Scheme for achieving a topological photonic crystal by using dielectric material. *Physical Review Letters*, 114(22):223901, 2015.
- [9] Ling Lu, Liang Fu, John D. Joannopoulos, and Marin Soljačić. Weyl points and line nodes in gyroid photonic crystals. *Nature Photonics*, 7(4):294–299, 2013.
- [10] Ling Lu, Zhiyu Wang, Dexin Ye, Lixin Ran, Liang Fu, John D. Joannopoulos, and Marin Soljačić. Experimental observation of Weyl points. *Science*, 349(6248):622–624, 2015.
- [11] 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):eaar4005, 2018.

- [12] 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):eaar4003, 2018.
- [13] Nobuyuki Okuma and Masatoshi Sato. Non-Hermitian topological phenomena: A review. *Annual Review of Condensed Matter Physics*, 14:83–107, 2023.
- [14] 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, 2013.
- [15] Luqi Yuan, Qian Lin, Meng Xiao, and Shanhui Fan. Synthetic dimension in photonics. *Optica*, 5(11):1396–1405, 2018.
- [16] John D. Joannopoulos, Steven G. Johnson, Joshua N. Winn, and Robert D. Meade. *Photonic Crystals: Molding the Flow of Light*. Princeton University Press, Princeton, 2nd edition, 2008.
- [17] Ching-Kai Chiu, Jeffrey C. Y. Teo, Andreas P. Schnyder, and Shinsei Ryu. Classification of topological quantum matter with symmetries. *Rev. Mod. Phys.*, 88:035005, 2016.
- [18] Mário G. Silveirinha. Topological classification of Chern-type insulators by means of the photonic Green function. *Physical Review B*, 97(11):115146, 2018. See also: Silveirinha, Phys. Rev. B 92, 125153 (2015) for continuous media Chern invariants.
- [19] Mikio Nakahara. *Geometry, Topology and Physics*. CRC Press, 2nd edition, 2003.
- [20] K. M. Ho, C. T. Chan, and C. M. Soukoulis. Existence of a photonic gap in periodic dielectric structures. *Phys. Rev. Lett.*, 65:3152–3155, 1990.
- [21] 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):013905, 2008.
- [22] B. Andrei Bernevig, Taylor L. Hughes, and Shou-Cheng Zhang. Quantum spin Hall effect and topological phase transition in HgTe quantum wells. *Science*, 314(5806):1757–1761, 2006.
- [23] Mohammad Hafezi, Eugene A. Demler, Mikhail D. Lukin, and Jacob M. Taylor. Robust optical delay lines with topological protection. *Nature Physics*, 7(11):907–912, 2011.
- [24] Mohammad Hafezi, Sunil Mittal, Jin Fan, Alan Migdall, and Jacob M. Taylor. Imaging topological edge states in silicon photonics. *Nature Photonics*, 7(12):1001–1005, 2013.
- [25] Kai Zhang, Zhesen Yang, and Chen Fang. Correspondence between winding numbers and skin modes in non-Hermitian systems. *Physical Review Letters*, 125(12):126402, 2020.
- [26] 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):257401, 2014.
- [27] Chia Wei Hsu, Bo Zhen, A. Douglas Stone, John D. Joannopoulos, and Marin Soljačić. Bound states in the continuum. *Nature Reviews Materials*, 1(9):16048, 2016.

-
- [28] Jian-Wen Dong, Xiao-Dong Chen, Hanyu Zhu, Yuan Wang, and Xiang Zhang. Valley photonic crystals for control of spin and topology. *Nature Materials*, 16:298–302, 2017.
 - [29] Xiao-Dong Chen, Fuli Zhao, Min Chen, and Jian-Wen Dong. Valley-contrasting physics in all-dielectric photonic crystals: Orbital angular momentum and topological propagation. *Physical Review B*, 96:020202, 2017.
 - [30] 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*, 14:140–144, 2018.
 - [31] Jiuyang Lu, Chunyin Qiu, Liping Ye, Xianyu Fan, Manzhu Ke, Fan Zhang, and Zhengyou Liu. Observation of topological valley transport of sound in sonic crystals. *Nature Physics*, 13:369–374, 2017.
 - [32] Guo-Jing Tang, Xin-Tao He, Fu-Long Shi, Jian-Wei Liu, Xiao-Dong Chen, and Jian-Wen Dong. Topological photonic crystals: Physics, designs, and applications. *Laser and Photonics Reviews*, 16:2100300, 2022.
 - [33] Daria Smirnova, Daniel Leykam, Yidong Chong, and Yuri Kivshar. Nonlinear topological photonics. *Applied Physics Reviews*, 7:021306, 2020.