




11-1-2019

Physics 516: Electromagnetic Phenomena (Fall 2019)

Philip C. Nelson

University of Pennsylvania, nelson@physics.upenn.edu

Follow this and additional works at: https://repository.upenn.edu/physics_papers

 Part of the [Astrophysics and Astronomy Commons](#), [Atomic, Molecular and Optical Physics Commons](#), [Biological and Chemical Physics Commons](#), [Condensed Matter Physics Commons](#), [Elementary Particles and Fields and String Theory Commons](#), [Optics Commons](#), and the [Statistical, Nonlinear, and Soft Matter Physics Commons](#)

Recommended Citation

Nelson, P. C. (2019). Physics 516: Electromagnetic Phenomena (Fall 2019). *n/a*, Retrieved from https://repository.upenn.edu/physics_papers/644

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/physics_papers/644
For more information, please contact repository@pobox.upenn.edu.

Physics 516: Electromagnetic Phenomena (Fall 2019)

Abstract

These course notes are made publicly available in the hope that they will be useful. All reports of errata will be gratefully received. I will also be glad to hear from anyone who reads them, whether or not you find errors: pcn@upenn.edu.

Keywords

electrodynamics

Disciplines

Astrophysics and Astronomy | Atomic, Molecular and Optical Physics | Biological and Chemical Physics | Condensed Matter Physics | Elementary Particles and Fields and String Theory | Optics | Physical Sciences and Mathematics | Physics | Statistical, Nonlinear, and Soft Matter Physics

PHYS516 notes
University of Pennsylvania P. Nelson
1 January 2020

These notes are made publicly available in the hope that they will be useful. All reports of errata will be gratefully received. I will also be glad to hear from anyone who reads them, whether or not you find errors: pcn@upenn.edu.



James Clerk Maxwell.

Brief contents

PART I Preliminaries

Prologue 2

Chapter 1 | Warmup: Newtonian Gravitation 14

PART II Static and Almost-static

Chapter 2 | Electrostatics Introduced 21

Chapter 3 | Electrostatic Multipole Expansion 31

[FRET] 46

Chapter 4 | Curvilinear Coordinates and Separation of Variables 47

Chapter 5 | Capacitors 55

Chapter 6 | Electrohydrostatics 67

Chapter 7 | Charge Flux, Continuity Equation, Ohmic Conductors 81

Chapter 8 | Cell Membrane Capacitance 91

Chapter 9 | Electrostatics in Solution 97

Chapter 10 | Cable Equation 119

Chapter 11 | Nerve Impulses 129

Chapter 12 | Examples of 3-tensors in Physics 140

Chapter 13 | Tensors from Heaven 149

- Chapter 14 | Magnetostatics 160
- Chapter 15 | Units in Electrodynamics 175
- Chapter 16 | Magnetostatic Multipole Expansion 187

PART III Dynamic

- Chapter 17 | Beyond Statics 196
- [Johnson noise] 217
- Chapter 18 | First Look at Energy and Momentum Transport by Waves 218
- Chapter 19 | Ray Optics and the Eikonal 226
- [Waves in plasmas] 243
- Chapter 20 | Diffraction 244
- Chapter 21 | Rainbows and Other Caustics 253
- Chapter 22 | Partial Polarization 256
- Chapter 23 | First Look at Radiation 260
- Chapter 24 | [Waveguides] 270

PART IV Relativity: Low Tech

- Chapter 25 | Galilean Relativity 272
- Chapter 26 | Springs and Strings 282
- Chapter 27 | Einstein's Version of Relativity: Overview 288
- Chapter 28 | Provisional Lorentz Transformations and the Fizeau Experiment 294
- Chapter 29 | Aberration of Starlight and Doppler Effect 305
- Chapter 30 | Relativistic Momentum and Energy of Particles 324

PART V Relativity: High Tech

- Chapter 31 | Four-vectors and -tensors 334
- Chapter 32 | The Faraday Tensor 351
- Chapter 33 | Manifestly Invariant Form of Maxwell 361
- Chapter 34 | Energy and Momentum of Fields 381
- Chapter 35 | Faraday's Field Lines 392
- Chapter 36 | Plane Waves in 4D Language 397
- Chapter 37 | A Simple Spherical Wave 404
- Chapter 38 | Beams 409
- Chapter 39 | Variational Formulation 414

PART VI Radiation

- Chapter 40 | Radiation Green Function 421
- Chapter 41 | J. J. Thomson's Pictorial Explanation of Radiation 430
 - [Radiation Reaction] 436
- Chapter 42 | Electric Dipole Radiation 437
- Chapter 43 | Higher-Multipole Radiation 449
 - [Transition radiation] 456
- Chapter 44 | Liénard–Weichert Formula and Synchrotron Radiation 457
- Chapter 45 | The Microwave Polarizer 466
- Chapter 46 | Scattering by Free and Bound Charges 470
 - [Thomson Scattering of Unpolarized Light] 477

Chapter 47 | [Scattering by Many Objects] 478

[Scattering by a Dielectric Sphere] 479

PART VII Light in Materials

Chapter 48 | Index of Refraction and Optical Rotatory Power 481

Chapter 49 | Isotropic, Linear Media 496

Chapter 50 | Anisotropic Media 514

Chapter 51 | Čerenkov radiation 519

Chapter 52 | Energy in Media 526

Chapter 53 | Photonic Bandgap Materials 527

Chapter 54 | Metamaterials 531

[Nonlinear Media and Optical Solitons] 534

Chapter 55 | Field Quantization, Polarization, and the Orientation of a Single Molecule 535

Chapter 56 | Einstein's Gravitation 547

[Yang–Mills] 548

Epilogue 549

Appendix A | Units and Dimensional Analysis 554

Appendix B | Global List of Symbols 558

Appendix C | Numerical Values 567

Appendix D | Formulas 569

Detailed contents

Web resources xxix
To the student xxx
To the instructor xxxv

PART I Preliminaries

Prologue 2

- 0.1 In Their Glory 2
 - 0.1.1 The Maxwell equations 2
 - 0.1.2 Lorentz force law 3
 - 0.1.3 In words 3
- 0.2 Explanation of Symbols 4
 - 0.2.1 3-vectors 4
 - 0.2.2 Concerning right-hand rules 5
 - 0.2.3 Kronecker symbol 7
- 0.3 Mathematical Miscellany 7
 - 0.3.1 Divergence theorem 7
 - 0.3.2 Stokes theorem 7
 - 0.3.3 Two useful lemmas 8
 - 0.3.4 Euler theorem 8
 - 0.3.5 Angle and solid angle 8
 - 0.3.6 Delta function 9
- 0.4 What Lies Ahead 9
 - 0.4.1 Einstein's critique 9
 - 0.4.2 Some more hanging questions 11
- Track 2 12
 - 0.1.2' Point charge 12
- Problems 13

Chapter 1 | Warmup: Newtonian Gravitation 14

- 1.1 Equations 14
- 1.2 A Tricky Point 15
- 1.3 Earth/Moon 16
- 1.4 *Plus Ultra* 17
- 1.5 Meaning 17
- 1.6 More Hanging Questions 18
- Problems 18

PART II Static and Almost-static

Chapter 2 | Electrostatics Introduced 21

- 2.1 Framing 21
- 2.2 Rephrase in Terms of a Potential 22
 - 2.2.1 Field equations 22
 - 2.2.2 Force law 22
 - 2.2.3 Step back 23
- 2.3 Differences from Gravitation 23
- 2.4 Another Look Ahead 23
 - 2.4.1 Reality 23
- 2.5 Basic Solutions 24
 - 2.5.1 Point charge 24
 - 2.5.2 Continuous charge distribution 24
- 2.6 Conductors 25
- 2.7 Upcoming 26
 - 2.7.1 Quasi-static 26
 - 2.7.2 Beyond static 26
- Track 2 27
 - 2.2' 27
- Problems 28

Chapter 3 | Electrostatic Multipole Expansion 31

- 3.1 What is to Be Shown 31
- 3.2 Some Taylor expansions 32
- 3.3 Prove the Electrostatic Multipole Formula 33
- 3.4 Why Show This 33
- 3.5 More Remarks 33
 - 3.5.1 33
 - 3.5.2 From potentials to fields 34
 - 3.5.3 Apparent singularity 34
 - 3.5.4 Base point dependence 34
 - 3.5.5 Spherical distributions 34
 - 3.5.6 Moments that must equal zero by symmetry 34
 - 3.5.7 Pure dipole 35
- 3.6 Force and Torque on a Charge Distribution 36
 - 3.6.1 Force 36
 - 3.6.2 Torque 36
 - 3.6.3 Some famous molecules 37
 - 3.6.4 Induced dipole moment 38
- Track 2 39
 - 3.1'a Counting 39
 - 3.1'b Connection to spherical harmonics 39
 - 3.6.3'a 39
 - 3.6.3'b 40
- Problems 41

[FRET] 46**Chapter 4** | Curvilinear Coordinates and Separation of Variables 47

4.1	Separation of Variables in the Laplace Equation	47
4.2	Familiar Stuff	47
4.2.1	Cartesian coordinates	47
4.2.2	Plane polar coordinates	47
4.2.3	Plane polar payoff	49
4.2.4	Some food for thought	49
4.2.5	Three dimensions	49
4.3	A Spherical Conductor in a Uniform Field	50
4.4	Lightning Rod	50
4.5	Divergence Operator	51
4.5.1	Hard way	51
4.5.2	Easy way	52
4.6	<i>Plus Ultra</i>	52
	Problems	52
Chapter 5	 Capacitors	55
5.1	Framing	55
5.2	Parallel Plates in Vacuum	55
5.3	Energy Stored	55
5.4	Cylindrical Conductors in Vacuum	56
5.5	Parallel Plates With Medium	56
5.6	Energy Puzzle	58
5.7	Dense Medium	59
5.8	Polar Fluid Medium	61
5.9	Partitioning of Ions at a Fluid Interface or Cell Membrane	62
5.10	Boundary Conditions	63
	Problems	64
Chapter 6	 Electrohydrostatics	67
6.1	Framing	67
6.2	Some geometry	68
6.2.1	Curves in a plane	68
6.2.2	Mechanical equilibrium of an interface in a plane	70
6.2.3	Surfaces in space	71
6.2.4	The Young–Laplace formula	73
6.3	Effect of Electric Field	74
6.3.1	New contribution to energy balance	74
6.3.2	Electrostatics near a conical point	74
6.4	Technological applications	76
6.5	<i>Plus Ultra</i>	76
6.5.1	A look ahead	76
6.5.2	Other physical surfaces	76
6.5.3	A glimpse of general relativity	76
Track 2		78
6.2.3'		78
6.2.4'		78
	Problems	80

Chapter 7 | Charge Flux, Continuity Equation, Ohmic Conductors 81

- 7.1 One Dimension 81
- 7.2 Two or More Dimensions 81
- 7.3 Remarks 83
- 7.4 Nonstatic Situations 83
- 7.5 Quasistatic Situations 84
- 7.6 Electroencephalogram/Electrocardiogram 85
 - 7.6.1 EEG 85
 - 7.6.2 EKG 87
- Problems 88

Chapter 8 | Cell Membrane Capacitance 91

- 8.1 Framing 91
- 8.2 Fricke Experiment 91
 - 8.2.1 93
 - 8.2.2 93
- Problems 94

Chapter 9 | Electrostatics in Solution 97

- 9.1 97
 - 9.1.1 The Nernst relation sets the scale of membrane potentials 97
 - 9.1.2 The electrical resistance of a solution reflects frictional dissipation 99
- 9.2 A Repulsive Interlude 100
 - 9.2.1 Electrostatic interactions are crucial for proper cell functioning 100
 - 9.2.2 The Gauss law 103
 - 9.2.3 A charged surface in pure water is surrounded by a neutralizing ion cloud 103
 - 9.2.4 Excess salt shrinks the electric double layer 108
 - 9.2.5 The repulsion of like-charged surfaces arises from compression of their ion clouds 108
 - 9.2.6 Oppositely charged surfaces attract by counterion release 111
- Track 2 112
 - 9.1.2' 112
 - 9.2.3' 112
 - 9.2.5' 114
- Problems 116

Chapter 10 | Cable Equation 119

- 10.1 Cables and Their Equations 119
 - 10.1.1 The ill-fated transatlantic cable 119
 - 10.1.2 Setup 121
 - 10.1.3 Discretize 122
 - 10.1.4 Cable equation 122
- 10.2 Neurons 123
 - 10.2.1 A new phenomenon 123
 - 10.2.2 Ions in and out of equilibrium 124
 - 10.2.3 Linear cable equation for an axon 126
 - 10.2.4 Day breaks 127

Chapter 11 | Nerve Impulses 129

11.1	129
11.1.1	The time course of an action potential suggests the hypothesis of voltage gating 129
11.1.2	Voltage gating leads to a nonlinear cable equation with traveling wave solutions 132
Track 2	137
11.1.2'	137
Problems	139
Chapter 12	Examples of 3-tensors in Physics 140
12.1	Framing 140
12.2	Rank Zero; Rank One 140
12.3	Rank Two 141
12.3.1	Tensors as linear vector-valued functions of a vector 141
12.3.2	Symmetric tensor as a scalar-valued, quadratic function of a vector 143
12.3.3	Some linear vector functions, but not all, arise as the derivative of a quadratic scalar function 144
12.4	Rank Three 144
12.4.1	A vector-valued bilinear function of vectors 145
12.4.2	A scalar-valued trilinear function of vectors 145
Track 2	147
12.2'	Tensor properties of probability density functions 147
Problems	148
Chapter 13	Tensors from Heaven 149
13.1	Framing 149
13.2	The Components of a Tensor Transform Upon Change of Coordinates 149
13.2.1	Linear coordinate changes 149
13.2.2	Cartesian coordinates and their transformations 150
13.2.3	A reformulation 150
13.3	3D Levi-Civita Tensor 151
13.3.1	Coordinate version 151
13.3.2	Caveat 152
13.4	Connect to Elementary Things 153
13.4.1	Dot product 153
13.4.2	Cross product 153
13.5	Useful Identities 154
13.5.1	Swap dot and cross 154
13.5.2	Triple cross product 154
13.6	<i>Plus Ultra</i> 154
Track 2	156
13.3'	156
13.4'	156
Problems	159
Chapter 14	Magnetostatics 160
14.1	A New Force Awakens 160
14.2	Vector Potential 162
14.2.1	Preliminary 162
14.2.2	No scalar potential this time 162
14.2.3	Revisit electrostatics 162

14.2.4	Poincaré lemma	163
14.3	Gauge Invariance	164
14.4	Back to Physics	165
14.4.1	Steady currents	165
14.4.2	General solution	166
14.4.3	Self-consistency	166
14.4.4	Counting equations	167
14.5	Down From the Mountain: the Oersted Problem	167
14.6	Biot-Savart Law	167
14.6.1	General form	168
14.6.2	More general wires	168
14.7	Boundary Conditions	169
14.8	Magnetoencephalography	170
14.9	<i>Plus Ultra</i>	170
Track 2		171
14.9'a	What about magnetic monopoles?	171
14.9'b	Against pseudovectors	171
14.9'c	Totally antisymmetric tensors are so useful...	171
14.9'd	What about angular momentum conservation?	172
Problems		173

Chapter 15 | Units in Electrodynamics 175

15.1		175
15.2	Units in Mechanics	175
15.3	Units in Electrodynamics	177
15.3.1	SI units	178
15.3.2	Derived units	179
15.3.3	Gaussian system	180
15.3.4	One final confusion	182
15.4	Remarks	183
Track 2		184
15.2'	Elimination of more units	184
Problems		185

Chapter 16 | Magnetostatic Multipole Expansion 187

16.1	Tensor Preliminaries	187
16.2	Far Fields of a Steady, Localized Current Distribution	188
16.2.1	Magnetic dipole potential	188
16.2.2	A Familiar Example	188
16.3	Remarks	189
16.3.1	Higher moments	189
16.3.2	No base point ambiguity	189
16.4	Force and Torque on a Magnetic Dipole	189
16.4.1	Fixed dipole strength	189
16.4.2	Diamagnetism, paramagnetism, ferromagnetism	191
16.4.3	Magnetic levitation of objects at room temperature	191
Problems		191

PART III Dynamic

Chapter 17 | Beyond Statics 196

- 17.1 Statics Review 196
 - 17.1.1 Field equations 196
 - 17.1.2 A worked example 196
- 17.2 Faraday Law 197
- 17.3 198
 - 17.3.1 Back-EMF 198
 - 17.3.2 Cables, again 199
 - 17.3.3 Energy of magnetic fields 200
- 17.4 Maxwell's Modification to Ampère's Law 201
 - 17.4.1 A bold prediction 201
 - 17.4.2 Boundary conditions 202
- 17.5 Wave Solutions 203
- 17.6 204
- 17.7 Complex Notation 205
 - 17.7.1 Electric Gauss law 206
 - 17.7.2 Faraday law 206
 - 17.7.3 Magnetic Gauss law 207
 - 17.7.4 Ampère law 207
 - 17.7.5 Summary 207
- 17.8 Potentials 207
 - 17.8.1 Representation of \vec{E} and \vec{B} 207
 - 17.8.2 Gauge invariance 208
 - 17.8.3 Coulomb gauge 208
 - 17.8.4 The case of zero charge density 209
- 17.9 Waves, Again 210
- 17.10 Complex Polarizations 210
 - 17.10.1 Linear, circular, elliptical 210
 - 17.10.2 Circular polarization basis 211
- 17.11 Spherical Waves 211
- 17.12 *Plus Ultra* 211
- Problems 212

[Johnson noise] 217

Chapter 18 | First Look at Energy and Momentum Transport by Waves 218

- 18.1 Linear Polarization 218
 - 18.1.1 Energy transport 218
 - 18.1.2 Momentum 219
 - 18.1.3 Some electromagnetic phenomena 220
- 18.2 Light Cannot Be Interpreted As a Stream of Newtonian Particles 220
- 18.3 Other polarizations 221
- 18.4 Angular Momentum 221
- Track 2 223
 - 18.1.2' Pondermotive force 223
- Problems 224

Chapter 19	Ray Optics and the Eikonal	226
19.1	Framing	226
19.2	Uniform Medium	227
19.3	Piecewise-uniform Medium	227
19.3.1	Refraction law	227
19.3.2	Optical tweezers	228
19.3.3	Spherical aberration	228
19.3.4	Total internal reflection	229
19.4	Gradient-index Medium	230
19.4.1	Streamlines	230
19.4.2	Almost-plane waves	230
19.4.3	Eikonal equation	231
19.4.4	Rays in vacuum	232
19.4.5	Rays in an inhomogeneous medium	233
19.4.6	An example	234
19.5	More Applications	235
19.5.1	Mirage	235
19.5.2	A spherical lens with minimal spherical aberration	236
19.5.3	Gravitational lensing	236
19.6	<i>Plus Ultra</i>	237
	Problems	238

[Waves in plasmas] 243

Chapter 20	Diffraction	244
20.1	Waves	244
20.2	One Thin Slit	247
20.3	Two or More Thin Slits	247
20.4	One Fat Slit	249
20.5	Ray Optics	250
20.6	Vista	252

Chapter 21	Rainbows and Other Caustics	253
	Problems	254

Chapter 22	Partial Polarization	256
22.1	Light is Usually Chaotic	256
22.2	Fully Polarized Case	257
22.3	Unpolarized Case	258
22.4	Partial Polarization	258
22.5	How to Measure the Stokes Parameters	258
	Problems	259

Chapter 23	First Look at Radiation	260
23.1	The Wave Equation	260
23.2	Recall Magnetostatics	260

23.3	A Physically Motivated Guess	261
23.4	Check the Guess	261
23.4.1	Preliminary results	262
23.5	Our First Antenna	262
23.5.1	Far fields	262
23.5.2	Energy loss	265
23.5.3	Directionality	265
23.6	<i>Plus Ultra</i>	265
	Problems	266

Chapter 24 | [Waveguides] 270

PART IV Relativity: Low Tech

Chapter 25 | Galilean Relativity 272

25.1	Principle of Relativity	272
25.2	A Simple System	272
25.3	Active Viewpoint: Symmetry	273
25.4	Passive Viewpoint: Invariance	273
25.5	Rotations	274
25.6	Galilean Group	275
25.6.1	Some coordinate systems on spacetime are preferred	275
25.6.2	Boosts	276
25.6.3	Matrix notation	276
25.6.4	Group structure	276
25.6.5	Physical significance of invariance	277
25.6.6	Light cannot be interpreted as a stream of newtonian particles, part 2	278
25.7	1905 and All That	279
	Track 2	280
	25.1'	280
	Problems	281

Chapter 26 | Springs and Strings 282

26.1	Equation of Motion	282
26.2	Transverse Waves	283
26.3	Invariance Lost	283
26.3.1	Active viewpoint	283
26.3.2	Passive viewpoint	284
26.4	Invariance Regained	284
26.5	Connection to Electromagnetism	285
26.6	Energy and Momentum	285
26.6.1	Continuity equations	286
26.6.2	The case of harmonic waves	287
26.7	<i>Plus Ultra</i>	287
	Problems	287

Chapter 27 | Einstein's Version of Relativity: Overview 288

27.1	The Æther Hypothesis	288
27.2	The No-æther Hypothesis	289
27.3	Where We are Heading	291
Track 2		292
27.1'		292
27.2'	Poincaré's work	292

Chapter 28 | Provisional Lorentz Transformations and the Fizeau Experiment 294

28.1	Review	294
28.1.1	Galilean	294
28.1.2	The problem in a nutshell	295
28.2	Graphical Explorations	296
28.3	The Wave Equation is Invariant Under Provisional Lorentz Transformations	297
28.3.1	Active viewpoint	298
28.3.2	Passive viewpoint	298
28.4	Velocity Addition	298
28.5	A Nonnull, Falsifiable Prediction	300
28.6	Postscript	302
Problems		304

Chapter 29 | Aberration of Starlight and Doppler Effect 305

29.1	Framing	305
29.2	No Dilatation Invariance	305
29.3	Lorentz Transformations in One Space Dimension	307
29.3.1	A subgroup that excludes dilatations	307
29.3.2	Rapidity parameter	308
29.4	A Typical Paradox and its Resolution	309
29.5	Lorentz Transformations in Three Space Dimensions	310
29.6	More Key Experiments: Aberration of Starlight and Doppler Shift	311
29.6.1	Bending of light-speed trajectories	311
29.6.2	Effect on frequency	312
29.7	An Enormous Generalization	313
29.7.1	Lorentz invariance must apply to all of physics	313
29.7.2	Muon lifetime, CMBR dipole, and more	314
29.8	What's Next	316
Track 2		317
29.3'a	Light-cone coordinates	317
29.3'b	Invariant interval	317
29.3'c	Velocity addition	317
29.3'd	Relation to rapidity	317
29.6.2'		318
29.7.2'		318
Problems		320

Chapter 30 | Relativistic Momentum and Energy of Particles 324

30.1	Conservation Laws Lost	324
30.2	Conservation Laws Recovered	325

30.2.1	An unexpected analogy	325
30.2.2	What has/has not been shown	328
30.2.3	A geopolitical consequence	328
30.3	Particles with Speed at or Near c	329
30.4	<i>Plus Ultra</i>	330
	Problems	330

PART V Relativity: High Tech

Chapter 31 | Four-vectors and -tensors 334

31.1	How to Avoid Reading This Chapter and the Next One	334
31.2	3D Prelude	336
31.2.1	Rotations	336
31.2.2	3-vectors	337
31.2.3	3-tensor transformation rule	337
31.2.4	Symmetric and antisymmetric 3-tensors	338
31.3	Other Rotationally Invariant Systems in Mechanics	339
31.3.1	Gravitation	339
31.3.2	Angular momentum	339
31.3.3	Field equations in 3D	339
31.4	Summary: The Rules in 3D	340
31.5	Four Dimensions	341
31.5.1	Lorentz transformations and the invariant interval	342
31.5.2	Other invariant quantities	343
31.5.3	Four-velocity	344
31.5.4	Summary and first payoff	345
31.6	Momentum and Energy Revisited	345
31.6.1	Aside on quantum mechanics	346
31.6.2	$\mathcal{E} = mc^2$	346
31.6.3	Massless particles	347
31.6.4	Particle creation and destruction	347
	Problems	348

Chapter 32 | The Faraday Tensor 351

32.1	A New Way of Thinking	351
32.2	Lorentz Force Law	352
32.2.1	The Faraday tensor	352
32.2.2	Relate to traditional form	354
32.2.3	Summary	354
32.2.4	On beauty	355
32.2.5	Better than beauty: an experimental consequence	356
32.3	Transformation of the Faraday Tensor	356
32.3.1	Electric and magnetic fields mix under Lorentz boosts	356
32.3.2	A charge in uniform, straight-line motion	357
32.4	<i>Plus Ultra</i>	358
	Track 2	359
32.2.4'	More on beauty	359
32.4'a	Bigger symmetry groups	359

xx

32.4'b Supersymmetry 359

Problems 360

Chapter 33 | Manifestly Invariant Form of Maxwell 361

33.1 Framing 361

33.2 Field Equations in 4D 361

33.2.1 Transformation of spacetime derivatives 361

33.2.2 The wave operator 362

33.3 General 4-tensors 363

33.3.1 Rank 363

33.3.2 Symmetry 363

33.3.3 The metric is itself a tensor 364

33.4 Summary: The Rules in 4D 364

33.5 Vacuum Maxwell Equations 365

33.6 The Charge Flux 4-Vector 367

33.6.1 A geometrical formulation 367

33.6.2 \underline{J} is a 4-vector 368

33.7 Complete, Invariant Maxwell Equations 369

33.8 Four-vector Potential 370

33.8.1 370

33.8.2 Particle in uniform motion revisited 370

33.9 More About \underline{J} 371

33.9.1 A property of the delta function 371

33.9.2 A useful alternative formulation of \underline{J} 372

33.9.3 Another proof that \underline{J} is a 4-vector 373

33.10 A Dizzying Vista 373

33.11 *Plus Ultra* 374

Track 2 376

33.4' Inversions 376

33.7' Degeneracy of Maxwell equations 376

33.8.1' Counting equations, again 376

33.9' Geometric status of the charge flux 377

33.11' Spinors 377

Problems 380

Chapter 34 | Energy and Momentum of Fields 381

34.1 Framing 381

34.2 What Needs to Be Shown and Why 381

34.3 Continuity Equation for Energy and Momentum in the Absence of Long-Range Forces 382

34.4 Interactions Seem to Spoil Local Conservation 383

34.4.1 Long-range forces 383

34.4.2 Nonconservation of particle energy and momentum 384

34.5 Accounting for Field Contributions Restores Local Conservation of Energy and Momentum 385

34.6 What Has Been Accomplished 387

Track 2 388

34.5' 388

Problems 389

Chapter 35 | Faraday's Field Lines 392

- 35.1 Framing 392
- 35.2 Field Lines 393
- 35.3 Electric and Magnetic Forces 394
- 35.4 Forces Via the Stress 3-Tensor 394
 - 35.4.1 Magnetic forces 395
- 35.5 Magnetic Induction 395
- Problems 396

Chapter 36 | Plane Waves in 4D Language 397

- 36.1 Lorenz Gauge 397
 - 36.1.1 It's useful 397
 - 36.1.2 It's permitted 398
- 36.2 Plane Waves 398
- 36.3 Energy and Momentum 399
- Track 2 400
 - 36.1'a Gravity waves 400
 - 36.1'b Spin versus polarization 400
- Problems 401

Chapter 37 | A Simple Spherical Wave 404

- 37.1 Framing 404
- 37.2 Spherical Wave 404
 - 37.2.1 Exact solution 404
 - 37.2.2 Far fields 405
 - 37.2.3 Near field 405
- 37.3 A Circularly Polarized Spherical Wave? 406
- 37.4 Other Kinds of Spherical Waves 406
- 37.5 Interference 406
- 37.6 Summary 406
- Problems 407

Chapter 38 | Beams 409

- 38.1 Gaussian Beams 409
- 38.2 Bessel Beams 409
- 38.3 Optical-Vortex Beams 409
- 38.4 Transfer of spin angular momentum to a sphere 409
- Problems 410

Chapter 39 | Variational Formulation 414

- 39.1 Framing 414
- 39.2 Newtonian mechanics 414
- 39.3 Field Equations 415
 - 39.3.1 Scalar field 415
 - 39.3.2 Maxwell/Lorentz system 416
 - 39.3.3 Fields plus particles 416

- 39.4 Noether theorem 417
 - 39.4.1 Scalar field example 417
 - 39.4.2 Consequences of invariance 418
 - 39.4.3 Electrodynamics 419
- 39.5 *Plus ultra* 419

PART VI Radiation

Chapter 40 | Radiation Green Function 421

- 40.1 Framing 421
- 40.2 Time Ordering and Causality 421
- 40.3 Retarded Green Function 422
 - 40.3.1 Scalar wave equation 423
 - 40.3.2 Reformulate and confirm 424
- 40.4 Remarks 425
 - 40.4.1 Upgrade to 4-vector fields 425
 - 40.4.2 Check self-consistency 425
- Track 2 427
 - 40.3'a Alternative derivation in Fourier space 427
 - 40.4' 428

Chapter 41 | J. J. Thomson's Pictorial Explanation of Radiation 430

- 41.1 Framing 430
- 41.2 Electric Fields From an Accelerating Charge 430
- 41.3 Magnetic Fields 432
- Problems 434

[Radiation Reaction] 436

Chapter 42 | Electric Dipole Radiation 437

- 42.1 Framing 437
- 42.2 Far Field Approximation 437
- 42.3 Harmonic Time Variation 438
- 42.4 Multipole Approximation 438
- 42.5 Leading Order: Electric Dipole Radiation 439
 - 42.5.1 A time-varying ED moment leads to $1/r$ potentials 439
 - 42.5.2 Pure dipole limit 439
- 42.6 Radiation 440
 - 42.6.1 Electric and magnetic fields 440
- 42.7 Concrete Examples 441
 - 42.7.1 Electric dipole antenna 441
 - 42.7.2 Greenhouse gases 442
- 42.8 Energy Flux 442
- 42.9 Linear Polarization 444
- Track 2 446
 - 42.8' Only one rank-2 tensor from Heaven 446

Problems 447

Chapter 43 | Higher-Multipole Radiation 449

43.1 Framing 449

43.2 Next-order terms 449

43.2.1 Order-one terms in ϵ_{multi} can be divided into two tensor structures 449

43.2.2 Antisymmetric part of the moment 450

43.2.3 Symmetric part of the moment 451

43.3 Higher orders 452

43.4 *Plus Ultra* 452

Problems 453

[Transition radiation] 456

Chapter 44 | Liénard–Weichert Formula and Synchrotron Radiation 457

44.1 Solving the Inhomogeneous Maxwell Equations II: Finding the Fields of an Accelerating Charge 457

44.2 The Intensity Distribution 459

44.3 Analysis of the Distribution 462

44.4 Other Matters 463

Problems 465

Chapter 45 | The Microwave Polarizer 466

45.1 Framing 466

45.2 A Simplified Case 466

45.3 Effect on Arbitrarily Polarized Wave 468

45.4 Regeneration 468

Problems 468

Chapter 46 | Scattering by Free and Bound Charges 470

46.1 Shake It 470

46.2 Thomson Cross Section 470

46.3 An Astrophysical Application 472

46.4 Polarized Incoming Light 472

46.5 Unpolarized Incoming Light 473

46.6 Bound Charges 473

46.6.1 Rayleigh scattering cross section 473

46.6.2 Blue sky 473

46.6.3 A demo 474

Track 2 475

46.2' 475

Problems 476

[Thomson Scattering of Unpolarized Light] 477

Chapter 47 | [Scattering by Many Objects] 478

[Scattering by a Dielectric Sphere] 479

PART VII Light in Materials

Chapter 48 | Index of Refraction and Optical Rotatory Power 481

- 48.1 Introduction 481
- 48.2 Forward Scattering by Many Dipoles 481
- 48.3 Remarks on Index of Refraction 484
 - 48.3.1 About units 484
 - 48.3.2 About solutions 485
 - 48.3.3 About refraction 485
 - 48.3.4 About absorption 486
- 48.4 Scattering in an Anisotropic, Crystalline Medium 486
 - 48.4.1 Birefringence 487
 - 48.4.2 Linear dichroism 487
- 48.5 Scattering in a Fluid 488
- 48.6 Optical Rotatory Power 489
- 48.7 Coda 494

Chapter 49 | Isotropic, Linear Media 496

- 49.1 Framing 496
 - 49.2 Polarizable Media 496
 - 49.2.1 Electric 496
 - 49.2.2 Magnetic 497
 - 49.2.3 Maxwell 498
 - 49.2.4 Boundary conditions 499
 - 49.3 Linear Regime 499
 - 49.3.1 Electric 500
 - 49.3.2 Magnetic 500
 - 49.3.3 Maxwell 501
 - 49.3.4 Macroscopic physical realizations 501
 - 49.3.5 Remarks and further examples 501
 - 49.4 Fresnel Equations, “Total” Internal Reflection, and the Evanescent Wave 502
 - 49.5 Circular Birefringence 502
 - 49.6 Cross-susceptibility 503
 - 49.6.1 Macroscopic physical realization 503
 - 49.6.2 General form 504
 - 49.7 The Origin of Circular Birefringence 505
 - 49.8 Demo 506
 - 49.9 Coda 506
 - 49.10 *Plus Ultra* 507
- Track 2 509
- 49.2' 509
 - 49.2.1' Dissipation and frequency dependence 509
 - 49.6'a Just two enantiomers 510
 - 49.6'b Relativistic formulation 510

Problems 512

Chapter 50 | Anisotropic Media 514

- 50.1 Ordinary Birefringence 514
 - 50.1.1 Half-wave plate 515
 - 50.1.2 Linear dichroism 516
- 50.2 Optical Torque Wrench 516
- 50.3 *Plus Ultra* 516
- Track 2 517
 - 50.3' Magnetic anisotropy 517
- Problems 518

Chapter 51 | Čerenkov radiation 519

- 51.1 Framing 519
- 51.2 Charged Particle in Vacuum 519
- 51.3 Charged Particle in a Dielectric Medium 523
- 51.4 Interpretation 523
- Problems 524

Chapter 52 | Energy in Media 526**Chapter 53** | Photonic Bandgap Materials 527

- 53.1 Framing 527
- 53.2 1D Stack 528
- Problems 528

Chapter 54 | Metamaterials 531

- Problems 531

[Nonlinear Media and Optical Solitons] 534**Chapter 55** | Field Quantization, Polarization, and the Orientation of a Single Molecule 535

- 55.1 A Single Molecule Emits Photons in a Dipole Distribution 535
- 55.2 Maxwell Equations as Decoupled Harmonic Oscillators 535
- 55.3 Quantization Replaces Field Variables by Operators 538
- 55.4 Photon States 540
 - 55.4.1 Basis states can be formed by applying creation operators to the vacuum state 540
 - 55.4.2 Coherent states mimic classical states in the limit of large occupation numbers 541
- 55.5 Interaction with Electrons 542
 - 55.5.1 Classical interactions involve adding source terms to the field equations 542
 - 55.5.2 Electromagnetic interactions can be treated perturbatively 543
 - 55.5.3 The dipole emission pattern 544
- 55.6 Vistas 545
 - 55.6.1 Some transitions are far more probable than others 545
 - 55.6.2 Lasers exploit a preference for emission into an already occupied state 546

Chapter 56 | Einstein's Gravitation 547

[Yang–Mills] 548

Epilogue 549

Acknowledgments 553

Appendix A | Units and Dimensional Analysis 554

- A.1 Base Units 554
- A.2 Dimensions Versus Units 555
- A.3 About Graphs 556
 - A.3.1 Arbitrary units 557
 - A.3.2 Angles 557
- A.4 Payoff 557

Appendix B | Global List of Symbols 558

- B.1 Mathematical Notation 558
- B.2 Named Quantities 561

Appendix C | Numerical Values 567

- C.1 Fundamental Constants 567
- C.2 Optics 567
 - C.2.1 Index of refraction for visible light 567
 - C.2.2 Miscellaneous 567

Appendix D | Formulas 569

- D.0 Prologue 569
 - D.0.1 Vector calculus 569
 - D.0.2 The Maxwell equations 569
 - D.0.3 Lorentz force law 569
- D.1 Newtonian Gravity 570
- D.2 Electrostatics 570
- D.3 Electrostatic Multipole Expansion 570
 - D.3.1 Electrostatic force and torque on a charge distribution 572
- D.4 Curvilinear Coordinates 572
 - D.4.1 Laplace operator 572
- D.5 Capacitor 573
 - D.5.1 573
- D.6 Electrohydrostatics 574
- D.7 574
 - D.7.1 Continuity equation 574
 - D.7.2 Quasi-static 574
- D.8 Cell Membrane Capacitance 574
- D.9 Aqueous Solutions 575
- D.10 Cable Equation 575
- D.11 Nerve Impulses 575
- D.12 3-tensors 575
 - D.12.1 Tensors and spherical symmetry 575

D.13	3-tensors From Heaven	576
D.14	Magnetostatics	576
	D.14.1 Poincaré lemma	576
	D.14.1.1 For theory enthusiasts only	577
	D.14.2 Charge and flux	578
	D.14.3 Gauge freedom; fundamental solutions	578
D.15	Units	579
D.16	Magnetic Multipole Expansion	579
	D.16.1 More about magnetic dipoles	580
	D.16.2 Magnetic polarizability	580
D.17	Nonstatic Fields	580
	D.17.1 Potentials, nonstatic case	581
	D.17.2 Plane wave solutions	581
	D.17.3 Polarization of waves	581
D.18	Energy and Momentum Transport by Waves	582
D.19	Ray Optics	582
D.20	Diffraction	582
D.21	Rainbows	583
D.22	Partial Polarization	583
D.23	Generation of Radiation	583
D.24	Galilean Relativity	583
	D.24.1 Rotations	583
	D.24.2 Active viewpoint	583
	D.24.3 Passive viewpoint	583
	D.24.4 Boosts	584
D.25	Strings/Springs	584
D.26		585
D.27	Provisional Lorentz Transformations	585
	D.27.1 Addition of velocities	585
D.28	Full Lorentz Transformations	585
D.29	Relativistic Particle Energy and Momentum	586
	D.29.1 Some experimentally testable consequences	586
D.30	Four-vectors and -tensors	587
	D.30.1 More about 3-vectors and 3-tensors	587
	D.30.1.1 For theory enthusiasts only	588
	D.30.2 3-scalar, 3-vector, and 3-tensor fields	589
	D.30.3 4-vector notation	589
	D.30.4 Relativistic particles	591
	D.30.4.1 Massive	591
	D.30.4.2 Massless	592
D.31	The Faraday Tensor	592
	D.31.1 Manifestly invariant form of Lorentz force law	592
D.32	Manifestly invariant form of Maxwell	593
	D.32.1 4-covectors et al.	593
	D.32.2 4-tensor fields	594
	D.32.3 Invariants and invariant equations	595
	D.32.3.1 For theory enthusiasts only	596
	D.32.4	596
	D.32.4.1 For theory enthusiasts only	596

D.32.5	Four-vector potential	597
D.32.6	More about the charge–flux 4-vector field	597
D.32.6.1	For theory enthusiasts only	597
D.33	Energy and Momentum of Fields, General	598
D.33.1	598	
D.33.1.1	For theory enthusiasts only: Angular momentum of fields	598
D.34	598	
D.35	Plane Waves in 4D Language	598
D.35.1	Scalar waves	599
D.35.1.1	For theory enthusiasts only	599
D.35.2	599	
D.35.2.1	For theory enthusiasts only	600
D.36	Spherical Waves	600
D.37	Beams	600
D.38	Variational Formulation	600
D.39	Radiation Green Function	600
D.39.1	600	
D.40	601	
D.41	Electric Dipole Radiation	601
D.41.1	Far fields of an oscillating electric dipole	601
D.42	Higher Multipole Radiation	602
D.43	602	
D.44	603	
D.45	603	
D.46	603	
D.47	603	
D.47.1	Isotropic, Linear Media	603
D.47.2	Magnetic materials	604
D.47.3	Maxwell equations in media	604
D.47.4	Cross-susceptibility	605
D.47.4.1	For theory enthusiasts only	605
D.47.5	Waves in media	606
D.47.6	Optical activity	607
D.48	Anisotropic Media	607
D.48.1	Ordinary birefringence	607
D.49	Čerenkov	608
D.50	Poynting Theorem in Medium	608
D.51	Photonic Bandgap Materials	608
D.52	Metamaterials	608
D.53	Field Quantization	608
	<i>Bibliography</i>	610
	<i>Credits</i>	617
	<i>Index</i>	623

To the Student

27 January 1884. Thought about electromagnetic rays.
11 May. Hard at work on Maxwellian electromagnetics.
13 May. Nothing but electromagnetics.
16 May. Worked on electromagnetics all day.
8 July. Electromagnetics, still without success.
17 July. Depressed; could not get on with anything.
24 July. Did not feel like working.
7 August. Saw from Ries's book that most of what I have found so far is already known.

— *From the Diary of Heinrich Hertz*

The first goal of these notes is to give you the foundations, working knowledge, and fluency in some core theory ideas that even the most hard-nosed experimentalist must know. The second goal is to give you the foundation, working knowledge, and fluency in some key real-world phenomena that even the most abstruse theorist must know. My choices of what, precisely, constitute that dual core are what distinguish this treatment from the dozens of others available.

For example, I have a nice book that belonged to my mother's grandmother, titled *Fourteen Weeks in Physics* by J. D. Steele (1878). Now, first it is interesting and pleasing to me that a young woman in Camanche, Iowa in the 1890s was encouraged, or even permitted, to study Physics. Second, notice how little has changed—a semester is still 14 weeks in physics.

And yet, *everything* has changed. The book has chapters on electricity, and magnetism, and optics—but no inkling that these are the *same thing*. For this revolution we must thank not only Maxwell, but also the almost-forgotten Heinrich Hertz (now he's a unit!). Partly he lost out on icon status because he was dead at 36, couldn't sell himself.

Maxwell's work came out 1864, well before that textbook was written. Why wasn't it mentioned? For one thing, Hertz's *systematic experimental validation* took some time.¹ But perhaps also the author said, "This fancy stuff will blow over; students don't need to know about it." Einstein described his frustration as a student, even much later: "We were desperate to learn Maxwell's theory, but the old farts didn't think it was important." There's a parallel in my own life: After my first year in grad school, I went to a summer school where people who would later become household

¹Less systematic observations had been available for some time. For example, long before Hertz or Maxwell, Joseph Henry wrote, "A single spark from the prime conductor of a machine, of about an inch long, thrown on to the end of a circuit of wire in an upper room, produced an induction sufficiently powerful to magnetize needles in a parallel circuit of iron placed in the cellar beneath, at a perpendicular distance of 30 feet, with two floors and ceilings, each 14 inches thick, intervening" in 1842. [Maxwell was eleven years old!]

names told me about “string theory.” When I returned, the old farts (some of whom were under 30) mostly said, “This fancy stuff will blow over; it’s not important.”

I feel a point coming on. When your elders say, “Here are some tools and frameworks to solve difficult important problems,” you should probably pay attention to them. But when they say, “That new stuff isn’t important... Somebody already tried that and it doesn’t work...” listen politely, then make your own decision. Even if such naysaying is *usually* right, you don’t want to miss the exceptional opportunities.

Moreover, Maxwell originally wrote *twenty* equations! Later, when writing his magnum opus, he fell under the sway of an evil mathematician (Tait) who convinced him to rephrase everything in *quaternions*! That guaranteed that no ordinary scientist would bother to understand them. Only later did some practical American (Gibbs) write them in a vector form at all recognizable today. No wonder reasonable people initially saw it as inscrutable academic fluff. Also, Gibbs didn’t *publish promptly*. Priority therefore falls to Heaviside,² years after Steele’s book was published. Maxwell might not have even gotten credit had he not been such a towering figure for other work.

I feel a second point coming on: Someday if *you* revolutionize physics, make sure you say it in a way that others can grasp.

Last point: If it becomes hard, take heart from Hertz’s struggles (above). Everything worth doing is hard at first. Every physicist has a story of bottoming-out at some point.³ It never gets easier, but if you keep the fire on, eventually the kettle will boil, even if nothing seems to be happening for a long time. Later, you get to remember the previous occasions, and how you overcame them. Ask for help, and don’t wait till just before an exam or due date.

Some goals of this course

1. Wild intellectual romp; survey of remarkable phenomena (recall our course’s title). Finally do relativity “right,” i.e. the way physicists actually think about it.
2. Organize, systematize, integrate, consolidate. In particular, systematize the notion of symmetry, and its connection to tensors and tensor calculus (what is the cross product *really*?). We’ll start in 3D because most of us grew up in a (seemingly) 3D world. But then we’ll see the vast advantages when we bump vectors and tensors up to 4D (and beyond). Also you know there’s a relation between symmetry and conservation laws—we’ll make it precise.
3. Forge links to other kinds of physics, do problems that borrow from those fields instead of working in a hermetically sealed silo.
4. Meditation on “Where do good theories come from?” with glimpses of QED, Yang-Mills, GR. Electrodynamics is the gateway to all of current physics.
5. Specifically, systematically explore the physical hypothesis that Nature has certain symmetries. (Why is that cross product there? Must it be there?)
6. Applications (hence our course’s title). If you’re in the PhD program, your #1

²In 1882. Heaviside also introduced many familiar terms, including “impedance,” “inductance,” and “attenuation.”

³Mine involved spinor algebra.

question must be, not truth/beauty, but rather “what will I do my PhD on?” Vistas to all the research in this department.

7. Mathematical modeling. Of course that’s just what physicists do, but it’s worthwhile from time to time to examine our deep habits and then renew our vows (or update them). (They value this skill on Wall St too.)
8. A PhD is about research, and in research you keep getting stuck. You need problem-solving skills. This class is an opportunity to strengthen this generically useful faculty, but with more real-world problems than you may be accustomed to. You may have done a lot of cookbook problems, but this approach requires additional skills. Which are learnable. If you make an effort.
9. Among those skills are computation/data visualization.

Some uncomfortable questions

I might as well mention some *unmentionable topics*, since some of you are surely thinking them.

- “OMG, why must we take this course a *fourth* time?” (HS, Freshman, Junior, and again now.) “It’s not even quantum—it’s pre-internet! pre-electroweak! pre-TV! So old!” Well, but I looked around this department and found that quite a lot of research rested on understanding electromagnetic phenomena. Microwaves—there’s the CMBR. There’s the RF sent through your body in a MRI scan. Etc. Look at the Contents of these notes—the applications we’ll do are not old-fashioned stuff.
- And anyway, your understanding is still incomplete: I’m sorry if I’m the first one to tell you this, but *there’s no such thing as the electric field*. Nor the magnetic field. Neither has an independent meaning. But (in the classical approximation) there is an *electromagnetic* field, a union. We need to understand that. Seems absurd—they seem to play such different roles. Hmm, space and time also seem to play different roles, yet they, too, will get unified. Stay tuned.
- Some of you may be thinking, “OMG, no physicist believes that classical physics is true. Why discuss lies?” Similarly, alumni of 280 may say, “OMG, what’s this approach got to do with that approach to the same phenomena?”

In a nutshell, PHYS280 studied a quantum theory of light, but at the single-photon level. That picture was valuable for many phenomena relevant for biophysics, such as how single molecules absorb and reemit light. But some other simple phenomena, like what is in the space outside a permanent magnet, are *not* easy to describe in this way. To go further, we must invent a more detailed version of electrodynamics. This course will introduce its classical limit. Ultimately, the complete picture does require that we quantize that theory. Then the state created outside a permanent magnet can be regarded as a “coherent state” of the quantum field. I will only have time to hint at that complete theory at the end of the semester, because until then we’ll be busy with the many important electromagnetic phenomena that are adequately described by the classical theory.

The virtue of classical electromagnetism is that for many advanced applications it is a *fantastically accurate approximation* to the full quantum world

and *much simpler to handle*. Specifically, *polarization* effects are awkward in the approach of 280, yet important for many applications.⁴ The coherent response of many electrons in an antenna to a coherent state of EM radiation is another example. We like simple theories not (just) because we're lazy, but because with them we can see further without getting lost in formalism.

Finally, we'll see in Chapter 55 that the full structure of the classical theory is needed as the first step to quantization.

- “OMG—How can we cram two semesters into one?” Well, it's an ambitious course. If you haven't taken an intermediate-level E&M class, consider PHYS561–2 instead. If you're ready for 516, I'm committed to helping you teach yourself these difficult things. And I'll expect similar commitment from you.

Features of these notes

- Many chapters end with an appendix labeled “Track 2.” These sections are For Theory Enthusiasts Only. There are also Track 2 footnotes and problems, marked with the symbol $\boxed{2}$.
- Appendix B summarizes mathematical notation, then lists key symbols that are used consistently throughout the book.
- The notations “Equation x.y” and “Idea x.y” both refer to the same numbered series.
- Units appear in sans-serif font, dimensions in blackboard-bold. This way, you can visually distinguish between m (meters), \mathbb{M} (dimension of mass), and m (a variable that could denote a particular object's mass, or an integer index, etc.). In handwriting, I personally can't do a distinct sans font, so I sometimes find it helpful not to use standard abbreviations for units (that is, say “meters,” “sec,” and “coul” instead of m , s and C) to avoid confusion. In fact, even in these notes I use coul for coulombs and volt for volts.
- The square root of minus one is indicated in roman type (i) to distinguish it from say, an index. (Some software packages instead refer to this quantity as I or as j .) The base of natural logarithms is indicated in roman type (e) to distinguish it from the charge on a proton (e), a constant of Nature. The differential operation is indicated in roman type (d) to distinguish it from any variable called d , which could denote a distance.

On accuracy

Everybody makes errors. Some people *seem* to make fewer errors because they catch them.

- Step 0 is of course to carry units everywhere (see Chapter 15). That's really important, but just the start. What if your units are correct but you dropped a term?
- Step 1: You can lock your work in a drawer and do it over from scratch, then reconcile. That's a really good approach too, but it won't help if you've got a conceptual problem. And/or you can get symbolic software to carry out steps for

⁴However, we will see how they fit into the quantum theory in Chapter 55.

you (same problem). And/or you can collaborate, hoping that your collaborator will make a disjoint set of errors, then reconcile. And/or you can come to office hours and ask your instructor or teaching assistant. But that stops working when the course is over.

- Step 2: You should impose general reasonableness tests—features the correct solution must have.

These steps will take you a long way, but... You need the secret weapon, the most powerful of the Rings of Power. You need *Step three*.

- Step 3 is, you identify limiting cases in which the answer is obvious, or at least known, or at least doesn't require computer math. Specialize your answer to such a case and reconcile if necessary.

For example, suppose that you are asked to compute the near- and far-fields of an oscillating dipole. Work hard, but then specialize your answer to the limiting case of the *static* dipole *far* fields and compare to the answer you know. Next, work up to considering the far fields of an *oscillating* dipole and compare to your physical expectations, and so on.

Books

Not all of these sources use the same units and notation as the present notes, so beware.

Indispensable: Pollack & Stump, 2002 (referred to as “P+S” in these notes); Landau & Lifshitz, 1979, Purcell & Morin, 2012; Feynman et al., 2010a; Feynman et al., 2010b.

Tensor methods: Fleisch, 2012; Foster & Nightingale, 2006.

Other math: Garrity, 2002; Shankar, 1995.

Other personal favorites: Vanderlinde, 2004; Lorrain et al., 1988; Fleisch, 2008.

Historical: Mahon, 2003; Hirshfeld, 2006; Nahin, 1988; Mahon, 2017

Technical: Zangwill, 2013; Smith, 1997; ?; Landau et al., 1984.

Computers:

MATLAB: Nelson & Dodson, 2015 = <https://github.com/NelsonUpenn/PMLS-MATLAB-Guide>

Python: Kinder & Nelson, 2018.

Möller, 2007.

Let's get started.

PART I

Preliminaries



It is impossible to study this wonderful theory without feeling as if the mathematical equations had an independent life and an intelligence of their own, as if they were wiser than ourselves, indeed wiser than their discoverer, as if they gave forth more than he had put into them. And this is not altogether impossible; it may happen when the equations prove to be more correct than their discoverer could have known with certainty. It is true that such comprehensive and accurate equations only reveal themselves to those who with keen insight pick out every indication of the truth that is only faintly visible in nature. – Heinrich Hertz, 1896

Prologue

0.1 IN THEIR GLORY

I assume that you have already encountered the basic equations of electrodynamics, and the symbols in which they are formulated, in previous classes. This short chapter will just establish some notation. Later chapters will:

- Motivate the form of each equation based on simple electromagnetic phenomena;
- Explore less simple phenomena that can be understood on the basis of these equations;
- Reformulate them in ways that for some purposes are more powerful; and
- Extend their reach by incorporating some idealized forms of macroscopic media.

0.1.1 The Maxwell equations

Maxwell did not write them in this form.¹ Each equation is named for somebody prior to Maxwell; besides systematizing everything, we will see that Maxwell also made a crucial modification to “Ampère’s” law.

$$\vec{\nabla} \cdot \vec{E} = \rho_q / \epsilon_0 \quad \text{electric Gauss} \quad (0.1)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{magnetic Gauss} \quad (0.2)$$

$$\vec{\nabla} \times \vec{E} + \dot{\vec{B}} = 0 \quad \text{Faraday} \quad (0.3)$$

$$\vec{\nabla} \times \vec{B} - \mu_0 \epsilon_0 \dot{\vec{E}} = \mu_0 \vec{j}. \quad \text{Ampère} \quad (0.4)$$

These equations can be solved for the electric and magnetic fields if we know the motions of charged particles.

The dots represent $\partial/\partial t$. The constants have numerical values $\mu_0 \approx 4\pi \cdot 10^{-7} \text{ m kg coul}^{-2}$ (the **magnetic permeability of vacuum**), and $\epsilon_0 \approx 8.85 \cdot 10^{-12} \text{ coul}^2 \text{ N}^{-1} \text{ m}^{-2}$ (the **electric permittivity of vacuum**).²

Later chapters will define the **charge density** ρ_q and **charge flux** \vec{j} in terms of the positions and motions of charged particles.³

The official name for \vec{E} is “electric field intensity”; \vec{B} is the “magnetic induction.” We’ll just call them the **electric and magnetic fields**. Some formulas are neater when

¹Even Einstein’s original relativity paper used different names for each cartesian component, today considered horrible.

²Chapter 15 will discuss units in greater detail, and explain why the value of μ_0 stopped being exact, and became only approximate, in 2019.

³Some authors call \vec{j} the “current density.”

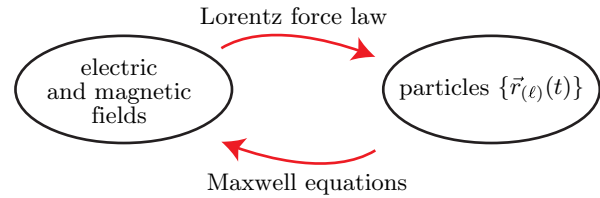


Figure 0.1:

expressed in terms of a quantity I'll call $\check{B} \equiv c\vec{B}$,⁴ because this quantity has the same dimensions as \vec{E} .

0.1.2 Lorentz force law

Reciprocally, the **Lorentz force law** describes the motions of a charged point particle if the fields are known:

$$\frac{d}{dt}\vec{p} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) + \vec{f}_{\text{other}}. \quad (0.5)$$

This time, d/dt represents the ordinary time derivative along a particle's trajectory. The fields \vec{E} , \vec{B} are to be evaluated at some time t and at the position $\vec{r}(t)$ of the particle at that time; $\vec{v} = d\vec{r}/dt$ at that time. q and m are constants completely characterizing the point charge. \vec{f}_{other} represents any non-electromagnetic force acting on the charged bodies in the system.⁵ And the momentum $\vec{p}(t) = m\vec{v}(t)$, at least for velocities much smaller than 10^8 m/s. (See Chapter 30 later.)

A **test body** refers to a limiting case of a point object with charge and mass infinitesimally small, but q/m a finite constant. In practice, a test body is a point charge so small that does not significantly perturb surrounding fields set up by other charges.

[\[D\]](#) Section 0.1.2' (page 12) discusses the notion of "charged point particle."

0.1.3 In words

- The electric Gauss law says, "Charges give rise to electric fields with some constant of proportionality $1/\epsilon_0$."
- The magnetic Gauss law says, "No point sources of magnetic fields."
- The Faraday law says, "Time-dependent magnetic fields are another source giving rise to electric fields."
- The Ampère law says, "Currents give rise to magnetic fields with some constant of proportionality μ_0 . Time-dependent electric fields are another source giving rise to magnetic fields."
- The Lorentz force law says, "A charged particle experiences a position-dependent electric force per charge, as well as a position and velocity dependent magnetic force per charge. The latter force is always directed perpendicular to the velocity."

⁴We won't give this quantity any particular identifying phrase. (Confusingly, gaussian people call it "the magnetic induction.")

⁵Sometimes it's appropriate to instead introduce a constraint. For example, we can imagine a situation in which a static charge is fixed onto on a spinning disk.

0.2 EXPLANATION OF SYMBOLS

0.2.1 3-vectors

A point in 3-space can be specified by choosing a “good” coordinate system (in particular, a cartesian system⁶) and quoting its components:

$$\vec{r}_i = \begin{bmatrix} x \\ y \\ z \end{bmatrix}_i, \quad i = 1, 2, 3. \quad (0.6)$$

That is, the symbol \vec{r} can either represent an abstract geometric object (an arrow), or it can represent a set of three numbers, called $\vec{r}_1 = x$, $\vec{r}_2 = y$, and $\vec{r}_3 = z$, regarded as a column (3×1 matrix). Note that a Latin subscript on a 3-vector indicates that only one of its components (an ordinary number) is meant. Again: The overarrow notation implies that we mean specifically cartesian coordinates.⁷ We won’t ever use the 3-vector notation \vec{r}^i (upper index) in these notes.⁸

Other quantities with an overarrow are understood to be triples of numbers with the same transformation under rotation of the spatial axes as \vec{r} , that is, **3-vectors**. The **3-scalar product** (also called **dot product**) is $\vec{a} \cdot \vec{b} = \sum_{i=1}^3 \vec{a}_i \vec{b}_i = \vec{a}^t \vec{b}$. We denote $\vec{r} \cdot \vec{r}$ by $\|\vec{r}\|^2$, \vec{r}^2 , or simply r^2 ; so $r \equiv \sqrt{\vec{r}^2}$. Thus, $\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos \theta$, where θ is the angle between the vectors.

The vector $\hat{r} = \vec{r}/r$ has length equal to one. More generally, a circumflex instead of an overarrow implies that a vector has been **normalized**, that is, divided by its length to convert it to a unit vector. Some standard unit vectors include the coordinate-axis directions \hat{x} , \hat{y} , \hat{z} (P+S call them **i**, **j**, and **k**).

The components of a vector *field*, such as $\{\vec{E}_i\}$, are themselves functions on spacetime, i.e. $\vec{E}_i(t, \vec{r})$ etc. We differentiate them with the vector of operators⁹

$$\vec{\nabla}_i = \begin{bmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{bmatrix}_i.$$

The dot product of $\vec{\nabla}$ with itself is the **Laplace operator** (or **laplacian**), written as¹⁰ ∇^2 .

The dot product of $\vec{\nabla}$ acting on a vector field is called the **divergence operator** and denoted $\vec{\nabla} \cdot \vec{V}$. Note that $\vec{\nabla} \cdot \vec{V}$ is an ordinary function, whereas $\vec{V} \cdot \vec{\nabla}$ is an

⁶Thus, curvilinear coordinates such as spherical polar are not “good” in this sense. Why make this restriction? For now, our answer is, “Because these are the coordinate systems in which Maxwell’s equations look nice, and we’re studying Maxwell’s equations.” Later we’ll consider how the representation of a vector changes when we switch from one “good” system to another or to a less “good” system.

⁷Most authors drop the overarrow when explicitly writing the index on a vector, but in these notes we retain it for clarity. Later we will sometimes append a sub- or superscript in parenthesis to the name of a vector. In this case we don’t mean to refer to a component; the label in parenthesis indicates which one of a *set of* related 3-vectors is meant (see for example Section 1.1 later).

⁸Such notation may, however, be useful when dealing with curvilinear coordinates. Later, when we define 4-vectors, we *will* introduce an upper-index notation, distinct from lower indices.

⁹Again, one can also set up a curvilinear coordinate system for expanding vectors, and find corresponding vector differential operators, but we’ll rarely use such systems: We are constructing tensor analysis on flat spaces, usually in the restricted class of cartesian coordinate systems.

¹⁰Mathematicians use the symbol Δ for the laplacian, but physicists don’t. It’s too easy to confuse that with Δ , the physicists’ symbol for a change in some quantity.

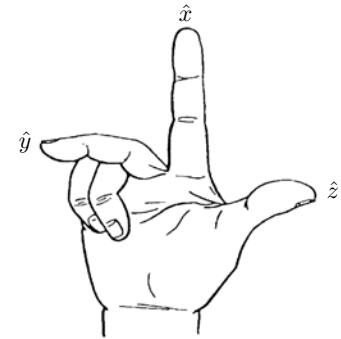


Figure 0.2: If we agree to call this a “right hand,” then the triad of unit vectors indicated is “right-handed.”

operator that acts on whatever sits to its right and does not involve any derivatives of \vec{V} . In fact, $(\vec{V} \cdot \vec{\nabla})f$ is the directional derivative of f along \vec{V} .

Consider a trajectory parameterized as a function of time, i.e. $\vec{r}(t)$. The 3-velocity is then $\vec{v} = d\vec{r}/dt$.

0.2.2 Concerning right-hand rules

The two best things in Italy are spaghetti and [Tullio] Levi-Civita.

— *Einstein*

Equations 0.3–0.5 given above assume that we have chosen a convention for “right hand.” This is the same thing as selecting a reference coordinate system on space whose unit vectors \hat{x} , \hat{y} , and \hat{z} are mutually perpendicular. To see the equivalence, note that with such a choice made, we can say which of your hands should be called “right” by the following procedure (Figure 0.2):

- Hold one hand flat with the fingers initially pointing along \hat{x} .
- Orient the hand so that when you bend your fingers by 90 degrees they now point along \hat{y} .
- If with that orientation, your thumb is pointing along \hat{z} , then that hand will be called “right” according to that coordinate system. If your thumb is pointing along $-\hat{z}$, then that hand will be called “left.”

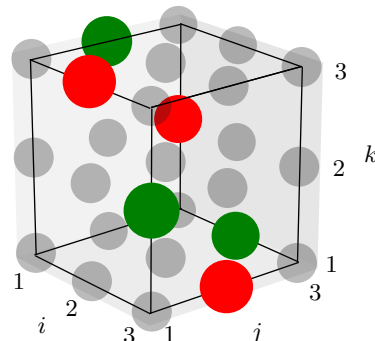
Alternatively, we could start by choosing one particular hand (for example, the one farthest from the heart of a normal human¹¹), and use it to classify coordinate systems as “right handed” or not.

The vector operators above are then defined by their usual (cartesian) formulas in a right-handed coordinate system. For example, the cross product can be expressed by saying that $\vec{a} \times \vec{b} = \hat{c} \|\vec{a}\| \|\vec{b}\| \sin|\theta|$, where θ is the angle from \vec{a} to \vec{b} and \hat{c} is a unit vector perpendicular to each of them. There are two such unit vectors; we choose the one specified by the right-hand rule.¹²

¹¹Less anthropocentrically, we could use the helical structure of the DNA of *any* (terrestrial) organism.

¹²That is, we choose \vec{c} that completes $\{\vec{a}, \vec{b}\}$ to a right-handed basis in the sense just given. If the

Figure 0.3: The 27 numbers that make up the Levi-Civita symbol ε_{ijk} , represented as a stack of balls. Three entries are +1 (green balls), three are -1 (red balls), and 21 are zero (transparent gray balls).



There is an equivalent formulation of the cross product that will be helpful throughout these notes. Although we are not ready to prove the equivalence (see Chapter 13), we will at least state the definition here, via the formula

$$(\vec{a} \times \vec{b})_i = \sum_{j,k=1}^3 \varepsilon_{ijk} \vec{a}_j \vec{b}_k. \quad (0.7)$$

The formula involves the **3D Levi-Civita symbol** ε_{ijk} , which is shorthand for $3^3 = 27$ numbers. Most of those entries equal zero: $\varepsilon_{ijk} = 0$ if any two of the indices match, for example, ε_{112} . If all three are different, then i, j, k must be a permutation of 1, 2, 3; ε_{ijk} is then defined using the parity of that permutation. Thus $\varepsilon_{123} = +1$, $\varepsilon_{231} = +1$, $\varepsilon_{132} = -1$ etc. (Figure 0.3).

The entries \vec{a}_j and \vec{b}_k in Equation 0.7 refer to the components of the vectors in any right-handed coordinate system, and the formula yields the components of the resulting vector in that same system.

We are not ready yet to prove that Equation 0.7 is independent of *which* right-handed coordinate system we chose, and indeed equivalent to the geometric definition. But you can readily generate some evidence:

Your Turn 0A

- Use Equation 0.7 to show that $\vec{a} \times \vec{a} = 0$ for any vector, in agreement with the geometric definition.
- The geometric definition clearly depends on which hand we declare to be “right.” Show that Equation 0.7 also has this (undesirable) feature. [Hint: Let $u = x$, $v = y$, and $w = -z$, and construct the corresponding unit vectors. Then a vector \vec{a} will have components with $\vec{a}'_1 = \vec{a}_1$, $\vec{a}'_2 = \vec{a}_2$, and $\vec{a}'_3 = -\vec{a}_3$. Writing \times' for the alternate version, we find $(\vec{a} \times' \vec{b})'_3 = \vec{a}'_1 \vec{b}'_2 - \vec{a}'_2 \vec{b}'_1$ and so on. Are these the primed components of the vector $\vec{a} \times \vec{b}$ defined in the usual way?]

One advantage of the algebraic formulation, Equation 0.7, is that it will show

angle between \vec{a} and \vec{b} equals zero or 180° then the choice is ambiguous—but in that case $\sin \theta = 0$, so the ambiguity doesn't matter.

us how, and in what sense, we may generalize the cross product to more than three-dimensional spaces.¹³

The cross product of $\vec{\nabla}$ acting on a vector field \vec{V} is a new vector field called the **curl** of \vec{V} , denoted $\vec{\nabla} \times \vec{V}$.

0.2.3 Kronecker symbol

There's also the more familiar **Kronecker symbol** δ_{ij} , which is defined to be +1 if $i = j$ and 0 otherwise.

0.3 MATHEMATICAL MISCELLANY

From now on we will employ the **summation convention**: When a vector index appears exactly twice in a formula, we mean for it to be summed over all its values, even if we don't explicitly write the summation symbol. Thus we abbreviate Equation 0.7 as $(\vec{a} \times \vec{b})_i = \varepsilon_{ijk} \vec{a}_j \vec{b}_k$. Such a summed index is also called a **dummy index**. If an index appears just once in an expression, it's called a **loose index** and is not summed. An expression with one or more loose indices really means several expressions, one for each set of index values. A loose index on one side of an equation must match a loose index on the other side (unless the other side is zero, in which case we mean that it's zero for *all* values of the index).

A summed pair of indices must each be named with the same letter of the alphabet. We can rename them both if we like, as long as they still agree with each other. When we combine formulas, we sometimes need to rename some index pairs in this way, to avoid ambiguity. Thus the product of $\vec{a}_i \vec{b}_i$ times $\vec{c}_i \vec{d}_i$ should be rewritten $\vec{a}_i \vec{b}_i \vec{c}_j \vec{d}_j$ (or $(\vec{a} \cdot \vec{b})(\vec{c} \cdot \vec{d})$).

Two crucial theorems from vector calculus are both beefed-up versions of the Fundamental Theorem of Calculus:

0.3.1 Divergence theorem

$$\int_V d^3r \vec{\nabla} \cdot \vec{E} = \int_{\partial V} d^2\vec{\Sigma} \cdot \vec{E}. \quad (0.8)$$

Here d^3r is a volume element.

V is a finite volume and ∂V is the closed surface bounding it. Any small element of that surface, $d^2\Sigma$, has two perpendicular directions called **normals** or (**normal vectors**). The surface separates space into "inside" and "outside," so one of the normals is the "outward-pointing normal." We convert an area element $d^2\Sigma$ into a vector, $d^2\vec{\Sigma}$, by multiplying it by the unit outward-pointing normal vector.

0.3.2 Stokes theorem

$$\int_{\Sigma} d^2\vec{\Sigma} \cdot (\vec{\nabla} \times \vec{E}) = \oint_{\partial\Sigma} d\vec{\ell} \cdot \vec{E}. \quad (0.9)$$

¹³And even to curved spaces.

Here $d\vec{\ell}$ is a vector line element. Σ is a surface, and $\partial\Sigma$ is its boundary (a curve in space). An open patch of surface has no “inside/outside” distinction, so we may choose either face as “outward” when defining the sign of $d^2\vec{\Sigma}$. Then the line integral along $\partial\Sigma$ must be traversed in the direction selected by applying a right-hand rule to the chosen outward-pointing vector.¹⁴

Please get (re)acquainted with these formulas, and with the specific conventions they contain concerning choice of handedness.

Your Turn 0B

- Show that, if you instead make the opposite choice of outward direction, then each side of Equation 0.9 changes sign, and the formula is still valid.
- Similarly, show that if we change our convention for which hand is “right,” then again we get canceling minus signs on each side.

If $\vec{\nabla} \times \vec{E} = 0$, we call \vec{E} a **curl-free vector field**. Then its contour integral depends only on the two endpoints. Another generalization of the Fundamental Theorem of Calculus says that in this situation, $\oint_{\vec{r}_0}^{\vec{r}} d\vec{\ell} \cdot \vec{E}(\vec{\ell})$, regarded as a function of the final endpoint \vec{r} , has gradient equal to $\vec{E}(\vec{r})$.

0.3.3 Two useful lemmas

Your Turn 0C

If the cartesian components \vec{V}_i of a vector field depend on position \vec{r} only via its distance r to the origin of coordinates, then show that

- $\vec{\nabla} \times \vec{V} = \hat{r} \times d\vec{V}_i/dr$, and
- $\hat{r} \cdot (\vec{\nabla} \times (\vec{r} \times \vec{V})) = -2\hat{r} \cdot \vec{V}$.

0.3.4 Euler theorem

When studying time-varying quantities, it’s useful to know that $e^{-i\omega t} = \cos(\omega t) - i \sin(\omega t)$. Thus we can represent both sines and cosines in a unified way: Either one can be written as $\frac{1}{2}[\bar{b}e^{-i\omega t} + \text{c.c.}]$ where “c.c.” stands for “complex conjugate.” If we choose $\bar{b} = 1$, then this expression equals $\cos(\omega t)$; if we choose $\bar{b} = i$, then it equals $\sin(\omega t)$; if \bar{b} is complex, then we get a sinusoid with frequency ω but with some phase shift relative to sine or cosine.

0.3.5 Angle and solid angle

A short line element $d\vec{\ell}$, seen from a distance, subtends an angle $d\theta = \|d\vec{\ell} \times \hat{r}\|/r$, where \hat{r} is the vector from the observer to the line element. This expression is dimensionless, but sometimes we add “radians” to an angle to emphasize that we are not using some weird units (like milliradians or degrees). Similarly, a small surface element $d^2\vec{\Sigma}$, seen from a distance, subtends a **solid angle**¹⁵ $d\Omega = d^2\vec{\Sigma} \cdot \hat{r}/r^2$. This expression is

¹⁴Point the thumb of your right hand along the chosen normal, then traverse the boundary in the sense that follows the curve of your fingers.

¹⁵A better name for this quantity might be **angular area**.

dimensionless, but sometimes we add “steradians” to a solid angle to emphasize that we are not using some weird units (like milliradians or square degrees).

0.3.6 Delta function

See Pollack & Stump, 2002, §3.5.1 for the definition of the **delta function**.¹⁶ Technically, it’s not really a function at all: When $\delta(x)$ is integrated over x , it’s a linear machine that eats an ordinary function and returns its value at zero:

$$\int dx \delta(x) f(x) = f(0).$$

Thus, the dimensions of $\delta(x)$ are always inverse to those of its argument x .

For our purposes it will usually suffice to regard $\delta(x)$ in a sloppy way as limit of a bump function, for example $e^{-x^2/(2\sigma^2)}/(2\pi\sigma)$, as it becomes sharply peaked holding the area under the curve fixed to 1. That viewpoint also makes it clear that the dimensions are inverse to those of x .

Section 33.9.1 will show that

$$\delta(f(x)) = \frac{1}{|f'(x_*)|} \delta(x - x_*).$$

Here we suppose that the function f has one zero at x_* ; if there’s more than one, the right hand side becomes a sum of terms for each zero. In multiple variables, the denominator of the prefactor gets replaced by the absolute value of the determinant of the Jacobian matrix.

0.4 WHAT LIES AHEAD

[Einstein’s first relativity paper] says that the usual formulation of the law of induction contains an asymmetry which is artificial, and does not correspond to facts. According to observation, the current induced depends only on the relative motion of the conducting wire and the magnet, while the usual theory explains the effect in quite different terms according to whether the wire is at rest and the magnet moving or vice versa.

— *Max Born*

The Maxwell equations are two vector PDEs, plus two more scalar PDEs. That’s a lot of complexity, even though the equations are linear. We will consider various reduced special cases before we start analyzing them in earnest, and some practical applications that can be understood using those simplified versions.

0.4.1 Einstein’s critique

If we know the equations, and accept that they are “true,” aren’t we done? Can’t we in principle just slap them on some big computer and find what they predict? In fact, it’s

¹⁶Sometimes called “Dirac delta function.”

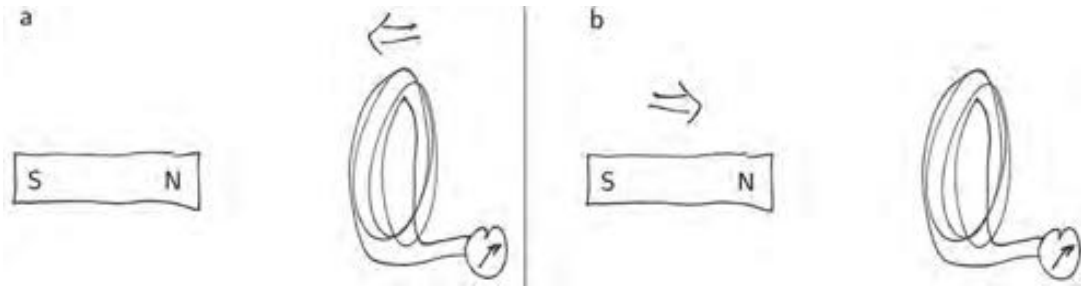


Figure 0.4: (a) A coil of wire is pushed over a stationary permanent magnet. (b) A permanent magnet is plunged into a stationary coil of wire.

fair to say that nobody understood the real content of Maxwell (certainly not Maxwell himself), until Einstein demonstrated a key hidden feature, an invariance property (or “symmetry”) that was there all along, buried in poor notation. Unfortunately, nobody understood *Einstein*, till Minkowski and successors found the appropriate generalization of vector notation to make this invariance manifest.¹⁷

One point that everybody *could* understand, mentioned right at the start of Einstein’s first paper on relativity, concerned what happens when a bar magnet enters a coil of wire (Figure 0.4).

- Every freshman knows how to explain the first of the two setups shown: Charges in the wire are free to move within it, but they are constrained not to leave it. When the wire is pushed sideways, as in Figure 0.4a, these charges must also move sideways. The Lorentz force law (Equation 0.5) then predicts a force perpendicular to that motion and to \vec{B} , so a charge initially in the plane of the page gets pushed out of the page, ultimately creating a current measured on the meter.
- When the coil is stationary (Figure 0.4b), then its charge carriers are not required to move, so $\vec{v} \times \vec{B}$ has no reason to be nonzero. In this case, however, the \vec{B} field is time-dependent. Faraday’s law (Equation 0.3) then implies an \vec{E} field, which *can* push charges that were initially at rest, again in the direction running along the wire. Again the meter responds.

Einstein said (paraphrasing), it’s *crazy* to offer two such totally different explanations of what is *obviously just one phenomenon*. After all, if you walk alongside the moving magnet, it appears stationary to you and the coil appears to move, and vice versa.¹⁸

In fact, why should we even invoke a *dynamical* explanation (rooted in equations of motion) for this equivalence, which ought rather to be *kinematic* in character? It will take us a while to arrive at Einstein’s answer to this question, but for now, suffice

¹⁷A good lesson: We old teachers should, like Minkowski, stay interested in our former students’ work. By the way, how did Einstein get through peer review, if nobody understood him at first? It’s simple. At that time, peer review was: Planck was the journal editor. He read the manuscript, said “I don’t understand it, but it looks good,” and that was that.

¹⁸Actually, at the time everybody other than Einstein would have agreed that *he* was crazy: “Obviously” the two situations were *not* equivalent, because at most one of them could be at rest with respect to the “luminiferous æther.” We’ll see later what Einstein said about that argument.

to note that *relativity was born out of frustration with electrodynamics*. We will leave it as a Hanging Question:

Hanging #A: Can we eliminate the asymmetry between our explanations of the coil/magnet phenomena? (Eventually we'll rephrase that as: Can we make full Lorentz symmetry manifest in the equations?)

0.4.2 Some more hanging questions

Section 0.4.1 raised a question that we won't answer for some time. Here are several more. Keep them in mind as we work through to their resolutions.

Hanging #B: Why must the Maxwell equations have exactly that (arbitrary-looking) form?

Hanging #C: How can \vec{E} and \vec{B} be “two parts of a single object” when they appear in such non-parallel ways?

Hanging #D: How can we solve the eight Maxwell equations with only *six* unknown functions \vec{E} , \vec{B} ?

Hanging #E: Our equations are full of cross products, which depend on an arbitrary choice of which is our “right” hand.¹⁹ Can we formulate electrodynamics in a way that doesn't conceal its invariance under spatial inversions?

¹⁹“The one farthest from the heart of a normal human” isn't very universal! Even “the one that describes DNA in all living organisms on Earth” is too Earth-centric to have fundamental significance.

 T_2 Track 2**0.1.2' Point charge**

A “point charge” is an idealization, having no multipole moments other than its total charge. Alternatively, if higher moments are present we assume their effects are negligible because surrounding fields are slowly varying. In classical electrodynamics, we assume that any charged macroscopic body can be regarded as a collection of point charges. But in a strong enough field gradient, even an electron cannot be regarded as a point charge, because it has a magnetic dipole moment! Similarly a neutron, although electrically neutral, can be pushed by a magnetic field gradient, etc.

PROBLEMS

0.1 *All Greek to me*

Now's a good time to learn the Greek alphabet. Here are the letters most often used by scientists. The following list gives both lowercase and uppercase (but omits the uppercase when it looks just like a Roman letter):

$\alpha, \beta, \gamma/\Gamma, \delta/\Delta, \epsilon, \zeta, \eta, \theta/\Theta, \kappa, \lambda/\Lambda, \mu, \nu, \xi/\Xi, \pi/\Pi, \rho, \sigma/\Sigma, \tau, \upsilon/\Upsilon, \phi$ (sometimes written φ)/ $\Phi, \chi, \psi/\Psi, \omega/\Omega$

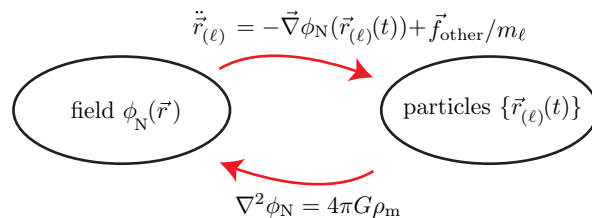
When reading aloud we call them alpha, beta, gamma, delta, epsilon, zeta, eta, theta, kappa, lambda, mu, nu, xi (English speakers pronounce it “k’see”), pi, rho, sigma, tau, upsilon, phi, chi (pronounced “ky”), psi, omega. Don’t call them all “squiggle.” Sometimes we will use the variant form φ for phi and ϑ for theta.

Practice by examining a quotation by D’Arcy Thompson: “Cell and tissue, shell and bone, leaf and flower, are so many portions of matter, and it is in obedience to the laws of physics that their particles have been moved, moulded, and conformed. They are no exception to the rule that $\Theta\epsilon\delta\varsigma\ \alpha\epsilon\iota\ \gamma\epsilon\omega\mu\epsilon\tau\rho\epsilon\iota$.” From the sounds made by each letter, can you guess what Thompson was trying to say? [*Hint*: ς is an alternate form of σ .]

CHAPTER 1

Warmup: Newtonian Gravitation

Newtonian gravitation isn't the subject of this course, but it's useful to introduce some themes with a scalar field theory before we move up to a vector field theory.



The cartoon above can be summarized as “fields tell particles how to move; particles tell fields what to be.” Let’s unpack that slogan.

1.1 EQUATIONS

The **newtonian potential** ϕ_N is a function that obeys

$$\nabla^2\phi_N = 4\pi G_N\rho_m. \quad (1.1)$$

Here G_N is a universal constant of Nature and ρ_m is the mass density of matter.

We can think of matter as a collection of N point masses m_ℓ following trajectories $\vec{r}_{(\ell)}(t)$. Here m_ℓ is a constant characterizing particle number ℓ . In the notation $\vec{r}_{(\ell)}$, the particle number ℓ appears in parentheses to avoid confusing it with a vector index labeling which component we’re discussing; the vector index has been suppressed to make the formula more compact.

With that notation understood, then we can finish specifying Equation 1.1 by constructing the mass density distribution as

$$\rho_m(t, \vec{r}) = \sum_{\ell} m_\ell \delta^{(3)}(\vec{r} - \vec{r}_{(\ell)}(t)). \quad (1.2)$$

In this formula, $\delta^{(3)}$ denotes the product of three delta functions. Notice the big distinction between \vec{r} and $\vec{r}_{(\ell)}$:

- \vec{r} labels the “field point,” that is, the point where we wish to evaluate ρ_m .
- The $3N$ functions of time, $\vec{r}_{(\ell)}(t)$, specify the N particle trajectories.

Often it’s a good approximation to think of ρ_m as a continuous function of position, effectively smearing the many delta functions together. Then sums over ℓ turn into integrals over d^3r .

Solving Equation 1.1 gives us the newtonian potential function if we know what all the masses are doing. Conversely, Newton’s second law amounts to $3N$ equations of motion that tell what the masses will do, given the potential:

$$\ddot{\vec{r}}_{(\ell)} = -\vec{\nabla}\phi_{\text{N}}(\vec{r}_{(\ell)}(t), t) + \vec{f}_{\text{other}}/m_{\ell}. \quad (1.3)$$

Thus we get a closed system of equations that, when solved together, tells us the future evolution of the system from initial conditions—the goal of classical physics.

The standard terminology is confusing: The “newtonian potential” is *not* the potential energy of particle ℓ . Instead, ϕ_{N} determines each particle’s potential energy by $U_{\ell} = m_{\ell}\phi_{\text{N}}$; it is potential energy *per unit mass* of a test body.

The term \vec{f}_{other} allows us to incorporate non-gravitational forces. Sometimes it’s an adequate approximation to instead introduce a constraint. Here the idea is that internal stresses supply whatever force is needed to maintain that constraint. For example, such stresses prevent the Earth from collapsing to a point, so that we may treat it as a fixed mass distribution. Other constraints ensure that the length of a pendulum remains constant, and so on.

But what is the field “really?” Newton’s successors eventually gave up fiddling with vortices in the æther and other mechanistic explanations, and just said, “it’s *really* a function on space and time, *period*. We don’t need a more explicit mechanical explanation to get on with making testable predictions. We don’t need to know if it’s really about vortices, or quantum coherent states of gravitons, or condensates of superstrings. . . . All we need to do is tell how to measure it operationally. If every time anybody measures it they find that it obeys the equations, then they are good and useful equations.”

1.2 A TRICKY POINT

To connect these formulas to first-year physics, you can find the solution to Equation 1.1 for the case of a point mass M (“the Earth”) at rest at the origin of coordinates. You already know the answer is $\phi_{\text{N}}(t, \vec{r}) = -MG_{\text{N}}/r$, but let’s just confirm that formula, using steps that we’ll need *again and again* in this course.

Equation 1.1 tells us to compute the laplacian of ϕ_{N} , that is, the divergence of the gradient. Let’s start with the gradient, and drop the prefactor $-MG_{\text{N}}$. So we want to find $\vec{\nabla}\left(\frac{1}{r}\right)$, where $r = \|\vec{r}\|$ is the length of the vector \vec{r} from the point mass to the observer. The first component of the gradient is

$$\frac{\partial}{\partial x}(x^2 + y^2 + z^2)^{-1/2} = -\frac{1}{2}(x^2 + y^2 + z^2)^{-3/2}2x = -x/r^3.$$

Notice that $1/r$ has units of inverse meters, as does $\vec{\nabla}$, so it’s right and proper that our answer has units of m^{-2} . Proceeding similarly with the other two components, and reinstating the constants, gives

$$\vec{\nabla}\phi_{\text{N}} = (-MG_{\text{N}})(-\vec{r}/r^3) = MG_{\text{N}}\hat{r}/r^2,$$

a familiar result. Here $\hat{r} = \vec{r}/r$ is the unit vector pointing to \vec{r} .

Now we want to compute the divergence: $\vec{\nabla} \cdot (\vec{\nabla} r^{-1}) = -\vec{\nabla} \cdot (\vec{r}/r^3)$. We use the Leibnitz property of derivatives (“product rule”) to write this as

$$-r^{-3}\vec{\nabla} \cdot \vec{r} - \vec{r} \cdot \vec{\nabla}(r^{-3}). \quad (1.4)$$

The first term is easy because $\vec{\nabla} \cdot \vec{r} = \frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3$. For the second term, adapt the previous result:

$$\vec{\nabla}(x^2 + y^2 + z^2)^{-3/2} = -\frac{3}{2}(x^2 + y^2 + z^2)^{-5/2} \begin{bmatrix} 2x \\ 2y \\ 2z \end{bmatrix} = -3\vec{r}/r^5.$$

So Equation 1.4 becomes $\nabla^2(r^{-1}) = -3r^{-3} - \vec{r} \cdot (-3\vec{r}/r^5) = 0$.

Oops. We succeeded too well. We wanted the laplacian to vanish away from the point mass at the origin, but we seem to have proved instead that it vanishes *everywhere*. The problem is that everything we’ve done is invalid right at $r = 0$, where the potential function is singular. To handle that point, consider a spherical surface surrounding it and use the divergence theorem:¹

$$\int_{\text{surf}} d^2\vec{\Sigma} \cdot \vec{\nabla}(r^{-1}) = (4\pi r^2 \hat{r}) \cdot (-\hat{r}/r^2) = -4\pi.$$

So the integral of $\nabla^2(-G_N M/r)$ over any spherical volume containing the origin is always $4\pi G_N M$, even though $\nabla^2(-G_N M/r) = 0$ everywhere other than the origin. The same things can be said of $4\pi G_N \rho_m$ for a point mass (that is, $\rho_m(\vec{r}) = M\delta^{(3)}(\vec{r})$), so we see that the familiar newtonian potential (which gives rise to the familiar newtonian force) really does solve Equation 1.1 for a point mass.

1.3 EARTH/MOON

The $1/r$ potential gives the equation of motion for a test particle (that is, a mass too small to itself pull appreciably on M):

$$\ddot{\vec{r}} = -\vec{\nabla}\phi_N(t, \vec{r}) = -MG_N\hat{r}/r^2. \quad (1.5)$$

That’s the familiar formula that gives rise to Kepler’s laws.

Just to find the simplest solution, recall that uniform circular motion has $\ddot{\vec{r}} = -\omega^2\vec{r}$ where ω is the angular frequency. Taking the value of ω that corresponds to a sidereal month, and r to be the Earth–Moon distance, and substituting into Equation 1.3 gave Newton a rough² numerical value for the quantity $G_N M$.

Newton also knew the acceleration of gravity for an object dropped near Earth’s surface. Knowing the radius of the Earth (a bit inaccurately, at the time) gave him another, independent estimate of $G_N M$. With historic understatement, Newton wrote

¹See Equation 0.8. To be a bit more precise, imagine the mass distribution not as a singular point, but spread over a very small volume. Take the spherical surface to lie outside this occupied region. Then the radius of that region drops out of the formulas, so we can take the limit where it, and the surface, shrink to zero size.

²It’s not really a good approximation to suppose the Earth stationary, because the Moon’s mass is not much less than Earth’s. You can do better by using the “reduced mass.”

that these two estimates “answered pretty nearly.” That was the first grand unified theory—of celestial and terrestrial motions.

By the way, the fact that Earth is not quite spherical is easy to incorporate into our assumed mass density function. Then we can solve the field equation and find the not-quite-spherical potential surrounding Earth, and from there the not-quite-Keplerian orbits of, say, spy satellites. Stay tuned for this idea in the context of electrostatics.

1.4 PLUS ULTRA

1. It’s true that we only found the solution to the field equation for a point mass, but perhaps surprisingly that’s all we need. Because the field equation is a *linear* PDE, and also invariant under spatial translations, we can subdivide any complicated distribution of mass into small chunks, apply the fundamental solution to each chunk, then use superposition to assemble all the sub-solutions into the full solution for ϕ_N .

We’ll similarly exploit the linearity of Maxwell’s equations for a similar win. The fundamental solution that must be integrated is called the **Green function** for whatever field equation we seek. We’ll find simplified Green functions for electro- and magnetostatics, then a more elaborate one for the full Maxwell equations.

2. Why introduce the potential function? Why not just work directly with the forces? One huge practical advantage is that the potential is a *scalar*. Combined with the preceding point, this means that we can conveniently integrate contributions from a complicated source (the ellipsoidal Earth, etc.), then at the very end compute the gradient, instead of having to carry around vector quantities throughout the calculation.

3. From this promising start, Newton and his successors³ proceeded to explain planetary motion, motions of moons around other planets, comet orbits, tides, the nonspherical shape of the Earth, phase-locking of Mercury and of Earth’s Moon, precession of Earth’s axis, effects of Jupiter on other planets—a fantastic wealth of testable predictions from very few assumptions.⁴

Once the idea sank in that Nature was governed by laws, on Earth as it is in Heaven, the seeds were sown for the Enlightenment and all that entailed. Newton’s biggest fan in France was Voltaire, who thought that if Nature itself is subject to natural laws, not the whims of a supernatural being, then the divine rights of capricious kings looked a bit silly. But that is another story.

1.5 MEANING

Before we can claim that Equations 1.1–1.3 make testable predictions, we need to give meaning to all the quantities that they relate. Later developments showed that even

³Notably Euler and Laplace.

⁴Later in the course, we will follow Einstein’s steps as he realized that the equations in this chapter, and in particular their invariances, are not quite correct. We are reviewing them because the cartoon at the start of this handout is still a good way to think about more advanced theories.

the very coordinates $\vec{r} = (x, y, z)$ and t require careful interpretation.⁵

Newton wrote some mumbo-jumbo about absolute space and time, but a more fruitful attitude emerged slowly. Today we say that what the equations are claiming is merely that *there exists a way* of labeling events by sets of four numbers, such that that any motion of any set of masses, with any initial conditions, corresponds to a solution of the equations.

This may sound like a big loss of predictive power—maybe there’s a physical motion that fails to satisfy the equations, but we could rescue them by merely relabeling the points! But even in this weakened form, the equations have the character of an interlocking web of many predictions: One *single* coordinate choice is supposed to handle any conceivable apparatus that we might wheel into the lab,⁶ any initial conditions we may set on that apparatus, etc.

Interestingly, and important for our later discussion, once we find one set of “good” coordinates on spacetime (that is, coordinates for which all phenomena obey the equations in their usual form), then there will also be *other* such “good” coordinate systems with the same property. You won’t be surprised to know that rigidly shifting or rotating x, y, z (leaving t unchanged) gives a new “good system.” Also, shifting $t' = t + t_0$ works, and so does negating any or all of x, y, z , or t . Later we’ll investigate just how big the set of “good” systems is. For now, we content ourselves with the statement that *the content of newtonian physics includes the claim that at least one “good” coordinate system exists.*

1.6 MORE HANGING QUESTIONS

Hanging #F: Can we introduce a potential function for electromagnetism analogous to the gravitational potential, and reap benefits analogous to the ones we got in that situation?

Hanging #G: What physically makes some coordinate systems “good” and others not?

PROBLEMS

1.1 *The first grand unification*

Repeat Newton’s early unification: Look up the radius and period of the Moon’s orbit, calculate its acceleration, and estimate the quantity $G_N M_{\text{earth}}$. (Make the approximation that the orbit is circular. You can also ignore the reduced-mass effect, that is, make the approximation that the Moon is much less massive than Earth.) Next look up the Earth’s radius and again estimate $G_N M_{\text{earth}}$, this time based on the terrestrial acceleration of gravity. Compare the two values you found for $G_N M_{\text{earth}}$.

⁵A system of coordinates on spacetime is sometimes called a **frame of reference** or **coordinate frame**, but those terms can get confusing: Some authors restrict them to refer only to *inertial* (“good”) coordinate systems, but others don’t. Still other authors use the shorter term **observer**, but this, too, can get confusing: An observer is a *person* (or apparatus).

⁶Henry Cavendish designed a gravitational experiment that fits in a room.

1.2 *Flyby*

The text claimed that the birth of Western science was when Newton solved the planetary orbit problem, deriving Kepler's empirical observations as predictions and predicting the return of Halley's Comet (among many other things). Because of the similarity between electrostatics and gravitation, we get to revisit this highlight in this course, as it's mathematically the same problem as one needed to understand proton therapy.⁷ In this problem, assume that everything is moving much more slowly than the speed of light; thus you may use familiar nonrelativistic mechanics.

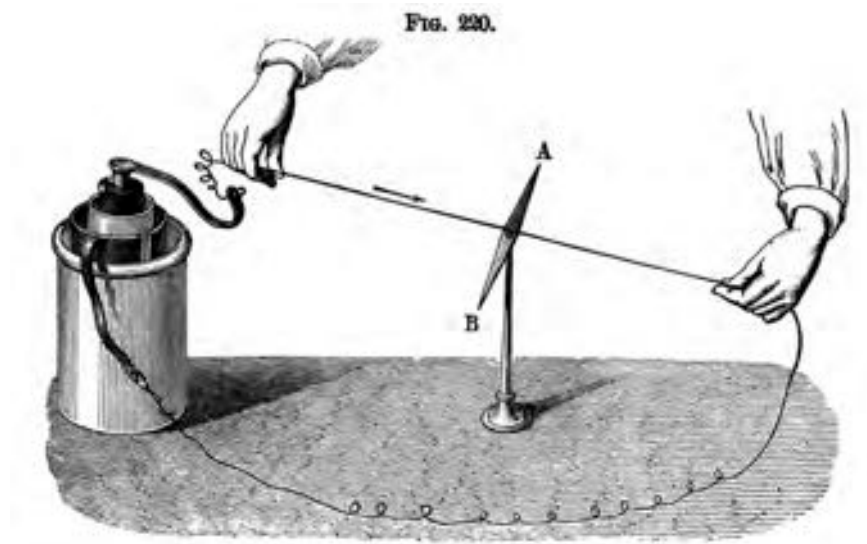
A heavy object M sits at the origin of coordinates. We will neglect any perturbation to its position during this problem, because the other object in the collision, m , is much lighter. The lighter object comes initially along a straight line parallel to the \hat{x} axis, moving from negative to positive x . If it were not deflected by M , the trajectory would pass within distance A of M ; that is, its initial trajectory is $x(t) = v_0 t$, $y(t) = A$ when $t \rightarrow -\infty$. Set up polar coordinates centered on M , in which φ is measured clockwise from the $-\hat{x}$ axis. Thus the incoming body starts with $\varphi = 0$, and φ increases with time. If M were not present, then the trajectory would have $\varphi \rightarrow \pi$ at $t \rightarrow +\infty$.

- Express the angular momentum of m about the origin, and the kinetic energy, both in terms of $r(t)$ and $\varphi(t)$. Use the constancy of the angular momentum to eliminate $\dot{\varphi}$ from the KE.
- Write the potential energy as $-K/r$. Thus $K = G_N M m$ for celestial mechanics, and $K = -q_1 q_2 / (4\pi\epsilon_0)$ for the nuclear case. Find the equation for the shape of the trajectory, that is, for $dr/d\varphi$.
- Substitute $u(t) = r(t)^{-1}$; that is, get an equation for $du/d\varphi$. We will be solving this equation for $u(\varphi)$.
- Initially $u \rightarrow 0$. Work out the initial value for $du/d\varphi$ from the fact that initially m is moving in uniform straight-line motion.
- Solve the equation given the initial conditions. Determine the value of φ at which u stops increasing and turns around. Double this angle to find the total angular deflection during the encounter. Be sure that your answer covers both the attractive and repulsive cases.
- An electron flies past a stationary proton with $A = 100 \text{ pm}$ and $v_0 = 0.01c$. (A picometer is 10^{-12} m .) What is the total deflection?
- A proton flies past an initially stationary electron. The proton's path is approximately unaffected by the electron, but the electron gains some kinetic energy. Find how much. Evaluate your answer for the illustrative case $A = 100 \text{ pm}$ and $v_0 = 0.01c$.

⁷One can also argue that the birth of modern Physics was Rutherford's discovery of the atomic nucleus; this problem is relevant for that discovery as well.

PART II

Static and Almost-static



“In the Middle Ages it had been the custom to ring peals of bells in the church steeples to disperse the thunder, as a result of which a high number of bell-ringers had been electrocuted. Indeed as late as 1786 the Parlement of Paris enforced an edict forbidding the practice, because over the previous thirty-three years no fewer than 103 unfortunates had been killed on the ends of their wet bell-ropes.... In 1708 Dr. Wall, in England, wrote that electricity “seems in some degree to represent thunder and lightning...” Concern soon focused on lightning strike and the danger it presented to gunpowder arsenals all over Europe. The row over exactly how to protect them began with the work of a hitherto obscure fifteenth child of a Bostonian soap-boiler, Benjamin Franklin.... He suggested that a church steeple be used to prove his theory. The Royal Society was not interested....

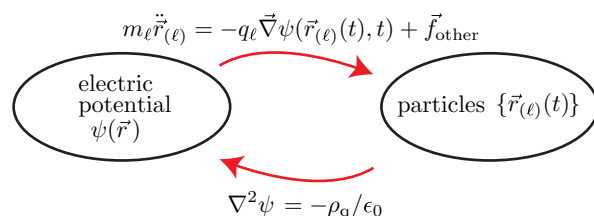
The explosion of an arsenal in Brescia, northern Italy, in 1769 made the rod a political issue. An estimated 175 000 pounds of powder exploded, destroying 190 houses within a radius of 639 feet from the explosion. The Brescia authorities asked the Royal Society for help in preventing a further disaster, and a committee was set up, of which Franklin was a member. An issue developed over whether the rods should be pointed at the top, as Franklin said, or round. The British settled for the round variety, on the grounds that Franklin was a revolutionary. Conductors sprang up all over Europe. There was even a chapeau paratonnerre - an anti-lightning hat—for the ladies of Paris in 1778.” – James Burke

CHAPTER 2

Electrostatics Introduced

2.1 FRAMING

Maxwell’s equations simplify a lot if we consider a static, or nearly-static, situation. That is, all charges are either motionless or slowly moving.¹ We will arrive at a system of equations of the form:



The cartoon above looks a lot like the one at the start of Chapter 1, but now each particle is characterized by *two* items of intrinsic information, called “mass” m (as before) and “charge” q . Each has a density: ρ_m (as before) and ρ_q respectively. Again, “fields tell particles how to move; particles tell fields what to be.” Let’s work through the second part of that slogan.

Although the equations in the figure are in principle complete, later we will find it useful to modify them in ways that approximately treat complicated systems in simpler, tractable ways:

- In this chapter and the next, we imagine the “other” forces to be *constraints*, that is, whatever is required to keep the charges at rest. In that case, the distribution of charge is invariant under both time shift and time reversal.²
- Later Chapter 5 will introduce dielectric media, containing molecules (deformable distributions of charge). Instead of treating these “bound” charges explicitly, we will summarize them with a modified value of the permittivity.
- Next Chapter 7 will go beyond statics, introducing conductors, in which mobile charges are impeded by their surroundings, and effectively obey a dissipative law of motion (Ohm’s “law”) instead of the ballistic one in the figure.
- Then Chapter 9 will introduce thermal agitation, which changes the equation of motion for the charges by adding a statistical-physics aspect. Charges will still be assumed to be slowly moving

¹Eventually we’ll say more precisely “slowly enough that we may neglect magnetic field effects.”

²In contrast, current flowing steadily through a wire is invariant under time shifts but not under time reversal—that’s called “stationary,” not static.

- Later, we'll consider situations in which, although charges move slowly, nevertheless they are so numerous that magnetic effects may not be ignored (Chapter 14).
- Chapter 17 will begin our study of charges with general motions.

For now, however, we will stick to the most basic situation: electrostatics. Thus we focus on the lower arrow on the figure.

2.2 REPHRASE IN TERMS OF A POTENTIAL

2.2.1 Field equations

Because $\vec{j} = 0$ and $\dot{\vec{E}} = 0$, there are no magnetic fields ($\vec{B} = 0$), and all we have left of Maxwell are

$$\vec{\nabla} \cdot \vec{E} = \rho_q / \epsilon_0, \quad \vec{\nabla} \times \vec{E} = 0. \quad (2.1)$$

Here ρ_q is electric charge density and ϵ_0 is a proportionality constant. Some such constant is needed for dimensional reasons: Because charge carries a new kind of dimension that cannot be converted to length, time, or mass, and \vec{E} is force *per charge*, ϵ_0 must among other things cancel two powers of charge units.

Our equations look much more complicated than those of newtonian gravity! Let's first address that defect.

Choose any fixed "reference point" \vec{r}_0 in space and define the **electrostatic potential** as the scalar function

$$\psi(\vec{r}) = - \int_{\vec{r}_0}^{\vec{r}} d\vec{r} \cdot \vec{E}. \quad (2.2)$$

Here the notation denotes the line integral along any path that starts at the reference point \vec{r}_0 and ends at the "field point" \vec{r} . It doesn't matter which such path we choose. Any two such paths differ by a closed loop, so switching to a different path changes ψ by the integral $-\oint d\vec{r} \cdot \vec{E}$ around that closed loop. By Stokes's theorem, this can be written as a surface integral of $\vec{\nabla} \times \vec{E}$, which is always zero by Equation 2.1.

It is true that the potential depends on the arbitrary choice of reference point, but in a trivial way: Changing \vec{r}_0 just adds a *constant* to ψ . We won't explicitly indicate the dependence on \vec{r}_0 , because we are already accustomed to the fact that potential energy is only well defined up to an additive constant.

By the Fundamental Theorem of Calculus, we then have that $\vec{E} = -\vec{\nabla}\psi$. And the other Maxwell equation above becomes the **Poisson equation**

$$\nabla^2 \psi = -\rho_q / \epsilon_0. \quad (2.3)$$

In a region with no net charges, the right hand side is zero and the equation is often rechristened the **Laplace equation**.

2.2.2 Force law

The Lorentz force law with no magnetic field becomes:

$$\dot{\vec{p}}_{(\ell)} = -q_\ell \vec{\nabla} \psi(\vec{r}_{(\ell)}(t)). \quad (2.4)$$

Here q_ℓ is the electric charge, a fixed quantity that is attached to particle ℓ .

The electrostatic potential $\psi(\vec{r})$ is the potential energy *per unit charge* of a test body located at \vec{r} . Its units are therefore joules per coulomb, which is the definition of “volt.” Most authors abbreviate this unit “V,” but that could lead to confusion with volume or something, so I’ll write it **volt**.

The electric field $-\vec{\nabla}\psi$ therefore has units of newtons per coulomb, or equivalently volts per meter.

2.2.3 Step back

We have transformed electrostatics from a set of four linear PDEs in the three unknown functions \vec{E} (Equation 2.1) to *one* linear PDE in *one* unknown function ψ , a considerable simplification. Indeed, it’s the same equation as in newtonian gravitation.

Our success relied on establishing an **integrability lemma**: While clearly any gradient has zero curl, we found that conversely any curl-free vector field can be written as a gradient via Equation 2.2. We will upgrade this argument when it’s time to find a potential for magnetostatics (Chapter 14), and then again when it’s time to find a 4-vector potential for electrodynamics (Chapter 36).

2.3 DIFFERENCES FROM GRAVITATION

There is an obvious big difference between newtonian gravity and electrostatics: The mass density ρ_m must always be nonnegative (everything attracts everything), but charge density ρ_q need not be nonnegative (some pairs of things attract but others repel). (Placement of the 4π factor is just a convention. In gravity we put it into the Poisson equation; in electrostatics, it’s conventional to bury it in the definition of the constant ϵ_0 .)

2.4 ANOTHER LOOK AHEAD

2.4.1 Reality

“But what is the electric field really?” This question turned out for many practical purposes to be as unnecessary as the similar one about the newtonian gravitational field. In this class, $\vec{E}(t, \vec{r})$ is a set of three *functions on spacetime*, period.

But another kind of “reality” question deserves comment. We could imagine saying, “there’s no such thing as the electric field, just action at a distance between charges via Coulomb’s law.” Today physicists find such nonlocal hypotheses to be repugnant, but that could be prejudice. Must we attribute independent reality to \vec{E} ? Occam’s Razor would say, “not if you can avoid doing so.” (Especially we should avoid introducing entities that you cannot see, hear, feel, smell, or taste.)

Let’s look ahead a bit. When we graduate to full electrodynamics, we’ll find wave solutions that are “real” (e.g. transport real energy) even after the charges that generated them have stopped moving or even *ceased to exist*. For example, dipping into quantum phenomena for a moment, consider the atom-like bound state of an

electron and a positron. At some moment the electron and positron annihilate each other, as for example in positron emission tomography (PET) imaging. Now nothing remains of them, nothing that could be exerting forces on distant charges—and yet, distant detectors eventually receive any radiation that the electron and positron gave off when they formed that bound state. It would be contrived at best to attempt to represent this situation as action at a distance from charges that no longer exist at the time of detection!

Occam says don't add new entities *unnecessarily*. But this example shows that the field concept is unavoidable, if we want to live in a world in which energy is locally conserved.³ Of course, “wanting” isn't enough. Eventually we'll need to *prove* some mathematical result about local conservation.

Hanging #H: Where is the energy in between emission and absorption of radiation? What continues to carry that energy even after the source no longer exists? Is there even a useful concept of “electromagnetic energy,” and for that matter, what does “useful” mean?

Chapter 34 will show that there is a unique way to attribute energy to fields in such a way that the total energy (particles plus fields) is locally conserved. As a bonus, we'll also get similar results for momentum and angular momentum.

2.5 BASIC SOLUTIONS

2.5.1 Point charge

One solution of the Poisson equation is the one we found in gravitation: A point charge of strength q located at the origin gives $\psi(\vec{r}) = q/(4\pi\epsilon_0 r)$, or more generally

$$\psi(\vec{r}) = \frac{q}{4\pi\epsilon_0 \|\vec{r} - \vec{r}_*\|}$$

if the charge is located at \vec{r}_* .⁴ You should find the negative gradient of this function, then go back via Equation 2.2 to see how it all fits together.

The minus sign in the Poisson equation tells that a + charge creates a $+1/r$ potential, that is, a potential energy hill for another + charge. Hence similar charges repel, unlike in gravity.

2.5.2 Continuous charge distribution

The Poisson equation is linear in ψ , so we can quickly generalize our point-charge solution to the case of a continuous distribution with charge density $\rho_q(\vec{r}_*)$. Simply subdivide charge into small elements $dq = \rho_q(\vec{r}_*)d^3r_*$ and add up their contributions. We'll call \vec{r}_* the **source point**, to distinguish it from the point \vec{r} where we wish to

³[\[T2\]](#) This example is not quite satisfying—if we're going to allow quantum ideas into the world, then we can say that light is a stream of little particles that *inhabit* space and that carry energy. But those particles do a lot of weird things not very much like, say, baseballs. The only way so far found to understand all light phenomena is to introduce fields, then quantize them (Chapter 55).

⁴The 4π had to pop up somewhere! We banished it from the Poisson equation, so it appears here.

know the potential (the **field point**). Thus the potential at the field point becomes an integral over source points:

$$\psi(\vec{r}) = \int d^3r_* \frac{\rho_q(\vec{r}_*)}{4\pi\epsilon_0\|\vec{r} - \vec{r}_*\|}. \quad (2.5)$$

This expression gives the general solution to the Poisson equation. It is called a **Green function solution**, and $1/(4\pi\|\vec{r} - \vec{r}_*\|)$ is called the **Green function** of the Laplace operator.

We should think carefully before applying a formula of this type. Suppose that we wish to know the potential at a field point somewhere inside the distribution, that is, a point where $\rho_q(\vec{r}) \neq 0$. The expression in Equation 2.5 seems to involve $1/0$ when $\vec{r}_* = \vec{r}$! But consider the integrand close to that point. Let $\vec{R} = \vec{r} - \vec{r}_*$. Then the suspicious part of the integral is d^3r/R , times the smooth function $\rho_q(\vec{r} - \vec{R})$. And $d^3r/R = R d\varphi d(\cos\theta)$ presents no problems near $R \rightarrow 0$.

2.6 CONDUCTORS

Another difference from gravity concerns “conductors.” These are a class of macroscopic bodies for which it’s a good approximation to say that charges (eventually) arrange themselves freely inside the body, without leaving it.⁵

It may seem a nightmare to handle problems of this sort—we can’t find the fields until we know where the charges go, and vice versa. In practice, however, the method of potentials gives an elegant approach: The free charges in a conductor just scoot around till they no longer feel any net force, that is, until $\vec{E} = 0$ everywhere inside the conducting body (and hence $\psi = \text{constant}$). Because ψ is a potential energy per test charge, it cannot change discontinuously across the conductor’s surface. Thus we get a boundary condition on the potential’s gradient (the derivatives of the potential parallel to the surface equal zero):

$$\vec{E}_{\parallel} = 0. \quad \text{just outside a good conductor, static} \quad (2.6)$$

The normal component E_{\perp} need not be zero at the surface; by the Gauss law, E_{\perp} tells us about the surface charge density.

Often we don’t even need to know the surface charge distribution. But if we do, we can find it by computing $\epsilon_0 \vec{\nabla}_{\perp} \psi$ once we have solved the boundary-value problem for the potential.

Chapter 9 will modify the preceding comments, acknowledging that they are true only at zero temperature. At nonzero temperature, thermal fluctuations are constantly knocking surface charges away from the surface, so there will be a thin layer with nonzero interior electric field even in equilibrium. That’s called a **depletion layer** in semiconductors, or **electric double layer** in soft matter.

⁵“Eventually” because charges may rearrange slowly, due to friction.

2.7 UPCOMING

2.7.1 Quasi-static

We'll see in Section 7.5 that many situations of interest are not precisely static, but may nevertheless be regarded as such because charges are moving slowly.

2.7.2 Beyond static

When things are moving fast, so that we're not even approximately static, it may seem that we can't get to first base: The electric field won't be curl-free, which seems to preclude introducing a potential. Luckily that's not true—later we'll construct a version of the potential that applies in this case as well. It won't have any interpretation as potential energy per unit charge, but nevertheless it will still be called a “potential.” Sorry for that misleading, but standard, terminology.

T₂ Track 2

2.2'

Equations 0.1–0.5 simultaneously give operational meaning to the electric and magnetic fields, and to the charge/mass ratios of the charged bodies, *and* to the choice of good coordinates on spacetime. In addition to defining the quantities they contain, they also make falsifiable predictions about relations between those quantities! The way this works is that the formulas have the character of an interlocking web of many predictions:

α . Suppose that we have reproducible classes of test bodies (e.g. protons, muons...), and an apparatus that creates repeatable situations. Then there exists at least one coordinate system on spacetime, and a number q_ℓ/m_ℓ characterizing each test body ℓ (but independent of the apparatus and the test body's motion), and a set of six functions $\vec{E}(t, \vec{r})$, $\vec{B}(t, \vec{r})$ characterizing the apparatus but independent of the test body and its initial conditions, such that *any physically realizable trajectory of any test body is a solution to Equation 0.5*.

β . If the apparatus consists of charges executing specified motions, then the functions \vec{E} and \vec{B} , measured as described in (α) above, *are not arbitrary, but are solutions to the partial differential Equations 0.1–0.4* with sources determined by the charges.

γ . If the apparatus consists of point charges which are themselves free (other than being influenced by EM fields and known forces \vec{f}_{other}), then the combined history of the fields and charges is a *self-consistent solution of Equations 0.1–0.5*, with sources given by formulas in Section 7.2 and Section 33.6.1 (specifically Equation 33.8 (page 368)).

Once we find one set of “good” coordinates on spacetime (that is, coordinates for which all phenomena obey the equations in their usual form), then there will also be *other* such “good” coordinate systems with the same property. You won't be surprised to know that rigidly shifting or rotating x, y, z (leaving t unchanged) gives a new “good system.” Also, shifting $t' = t + t_0$ works, and so does negating any or all of x, y, z , or t . Later we'll investigate just how big the set of “good” systems is. For now, we content ourselves with the statement that *the content of the Maxwell/Lorentz equations includes the assertion that at least one “good” coordinate system exists*.

Einstein called a “good” coordinate system on spacetime an **inertial frame of reference**. Note that all “good” coordinate systems in the above sense are, in particular, cartesian in x, y , and z and non-accelerating. One can extend the definitions of the vector operators, dot product, etc. to accommodate curvilinear or accelerated coordinates, but the very fact that those formulas look different from the usual cartesian form means that the Maxwell and Lorentz equations are not form-invariant under arbitrary change of coordinate systems. There is something special about inertial coordinate systems.

For sure, there will also be *bad* coordinate systems, in which the equations as written are *not* valid (just as there are accelerating systems in newtonian physics). What Einstein found illuminating, however, was the transformations *between* the presumed good systems, which were not what everybody had expected.

PROBLEMS

2.1 *Statics basics*

A static charge distribution produces a radial electric field $\vec{E} = Ar^{-2}e^{-br} \hat{r}$, where A, b are constants. \hat{r} is the unit vector in the radial direction.

- a. What is the total charge Q ?
- b. What is the charge density? Let $g(r)dr$ denote the charge located in a shell between radius r and $r + dr$, and sketch a graph of $g(r)$.

2.2 *Scalar potential*

- a. Suppose that far from a source we measure the static electrostatic potential

$$\psi(\vec{r}) = \frac{K}{r^5}(2x^2 - y^2 - z^2),$$

where $\vec{r} = (x, y, z)$, $r = \sqrt{\vec{r}^2}$, and K is a constant. Working in cartesian coordinates, derive a formula for the electric field $\vec{E}(\vec{r})$.

- b. Compute $\vec{\nabla} \cdot \vec{E}$ for the field you found in (a) and comment. To what class of potential functions does this one belong?
- c. Could this function describe the newtonian gravitational potential far from a localized distribution of mass?

2.3

Figure 2.1 represents the electric field lines outside a static charge distribution that is overall neutral. (The gray disks are singular regions, where I have not drawn the field lines.)

- a. Sketch a charge distribution that could result in such a field.
- b. If the electric field's magnitude falls with distance as $\vec{E} \sim r^\nu$, what is ν ?

2.4 *Proton therapy*

This problem continues Problem 1.2. In that problem, you found a formula for the deflection angle when an electron flies by a stationary proton. Your formula involved the quantity

$$Y \equiv K/(mAv_0^2)$$

where $K = e^2/(4\pi\epsilon_0)$, m =electron mass, and v_0 =magnitude of initial velocity. A is the perpendicular distance from the proton to the electron's initial trajectory.

- a. You also considered the related problem in which a proton flies by an initially stationary electron; you got a formula for the electron's final kinetic energy W . Do a little trigonometry to express the answer in the form

$$W = (\text{stuff})/(A^2 + (\text{more stuff})),$$

where the factors in parentheses don't depend on A ; you are to find them.

- b. When a proton flies through a gas of many initially stationary electrons⁶, it occasionally encounters one with a small value of A and gives it a significant kick.

⁶We will neglect screening in this problem.

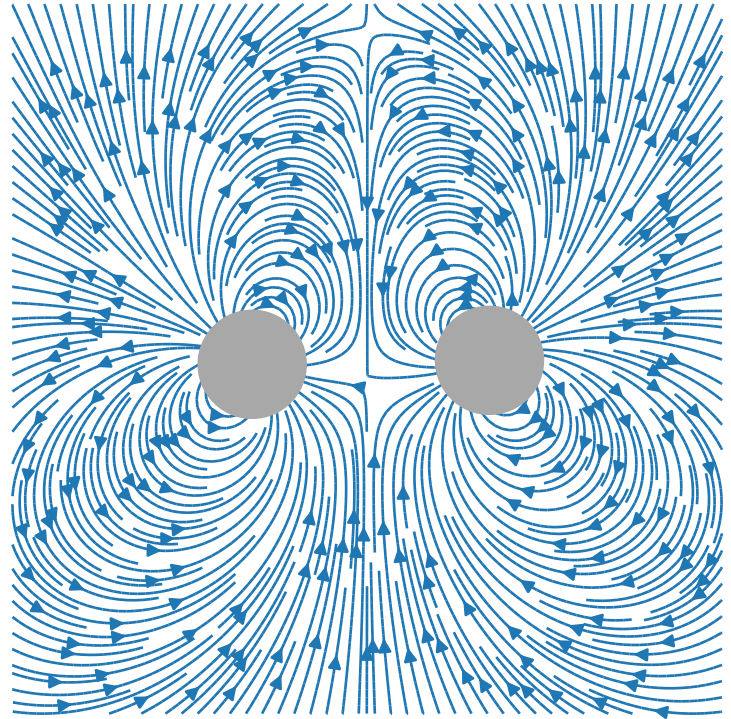


Figure 2.1:

During that encounter we'll pretend that the proton is unaffected, but over many collisions it will lose energy, about equal to the sum of all the W values for each encounter.

Suppose that the medium has a uniform number density of electrons, ρ_e . Initially the incoming proton is at depth $x = 0$ within the tissue, and has kinetic energy T_0 . After passing through to depth x , its velocity has fallen to some value $v(x) < v_0$ due to many encounters, and so its kinetic energy has also fallen to $T(x)$. Neglect the fact that the proton's direction will also change; suppose it is always moving in the same direction.

In the next dx , there are electrons at various values of A . Of these,

$$(2\pi A dA dx) \rho_e$$

have A values lying between A and $A + dA$. Write a formula for the total energy loss to the proton due to these electrons, and integrate it over A to get the energy loss per depth, dT/dx .

- c. Uh-oh. You found an infinite result; the integral is divergent. But wait. The electrons in human tissue aren't free; they are bound into molecules. If the energy transfer exceeds the binding energy, then maybe it's OK to neglect that fact, as we have done. But otherwise, the passage of the proton just deforms the molecule temporarily without necessarily any net loss of energy; your formula from (a) is not applicable in this case.

We'll take this complication into account crudely by just cutting off the integral

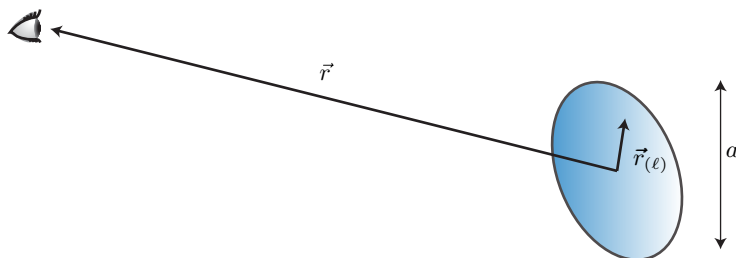
- in (b) at the value A_{\max} at which W equals the ionization energy I of a molecule. Find a formula for A_{\max} .
- d. Now do the integral over A , that is, find dT/dx in terms of T , I , ρ_e , and constants. Note that $T = \frac{1}{2}M_p v^2$, where M_p is the proton mass (i.e., it's not $\frac{1}{2}mv^2$).
- e. Simplify your expression by defining a suitable length scale and expressing x in terms of it. Also substitute some numbers:
 You know $K = e^2/(4\pi\epsilon_0) = 1.4 \text{ eV nm}$.
 You know the electron and proton masses.
 Suppose $I \approx 10 \text{ eV}$.
 Tissue is mostly water. You know how to compute the electron density ρ_e of water. (Assume that all the electrons have the same ionization energy.)
 Suppose that the proton initially has $T_0 = 100 \text{ MeV}$.
- f. Now you can find the relation between x and $T(x)$. This will involve solving the differential equation $dT/dx = (\text{expression you found})$. Luckily that equation can be solved just by doing an integral. Unluckily I don't know how to do that integral. Ask some computer to evaluate it numerically, and hence find the x values corresponding to a set of T 's starting at T_0 and decreasing to, say, $T_0/50$. Plot your answer as a graph of remaining kinetic energy T versus x .
- g. Actually, we are more interested in the deposition of energy as a function of depth. Make a second plot showing dT/dx as a function of x , and comment on its general form.

CHAPTER 3

Electrostatic Multipole Expansion

Electrostatics is easy if you are told where the charges are (fixed charge distribution). And that's often a reasonable approximation when we study *molecules*, for example H₂O or CO₂. The charge distributions on these molecules come from quantum mechanics, but given that, we can ask what electrostatic fields they create, and what qualitative conclusions we can draw.¹ Moreover, often we are only interested in the fields *far* from a molecule. It's convenient to be able to summarize the distribution for such purposes with just a few numbers. Here we systematize that procedure.

Besides bringing technical and conceptual benefits, Chapter 14 will extend these ideas to get a similarly useful magnetic multipole expansion. Then it will come around a third time, when we study radiation in Chapter 42. It's a powerful method.



3.1 WHAT IS TO BE SHOWN

Consider an isolated, static charge distribution confined to a region of size $\approx a$, viewed from far away; that is, $r \gg a$. We'll choose an origin of coordinates somewhere inside that region, so that charge $\# \ell$ sits at a position $\vec{r}_{(\ell)}$ with $r_{(\ell)} \ll r$. The goal is to show that the electrostatic potential at \vec{r} can be expanded in powers of a/r as

$$\psi(\vec{r}) = q_{\text{tot}}\psi^{[0]}(\vec{r}) + \vec{\mathcal{D}}_{\text{E}} \cdot \vec{\psi}^{[1]}(\vec{r}) + \sum_{ij} \left[\vec{\mathcal{Q}}_{\text{E},ij} \vec{\psi}_{ij}^{[2]}(\vec{r}) \right] + \mathcal{O}(a^3/r^4). \quad (3.1)$$

In this formula, q_{tot} is a scalar constant called **electric monopole moment** or “zeroth moment of charge.” The three constants $\vec{\mathcal{D}}_{\text{E}}$ form a vector called **electric dipole moment** or “first moment of charge.” The constants $\vec{\mathcal{Q}}_{\text{E}}$ are called the **electric quadrupole tensor** or “traceless part of the second moment of charge.” These quantities are defined by²

$$q_{\text{tot}} = \sum_{\ell} q_{\ell}, \quad \vec{\mathcal{D}}_{\text{E},i} = \sum_{\ell} q_{\ell} \vec{r}_{(\ell)i}, \quad \vec{\mathcal{Q}}_{\text{E},ij} = \sum_{\ell} q_{\ell} (3\vec{r}_{(\ell)i} \vec{r}_{(\ell)j} - r_{(\ell)}^2 \delta_{ij}). \quad (3.2)$$

¹It's true that a molecule is not quite fixed—it can deform, e.g. polarize, but for many purposes we don't need that level of detail.

²Some authors move a factor 1/2 from the quadrupole field into the definition of $\vec{\mathcal{Q}}_{\text{E}}$; others instead use the convention given here.

Although the indices on the quadrupole tensor each run from 1 to 3, so that it has nine entries, only five of these have independent values. That's because $\vec{\vec{Q}}_{E,ij}$, regarded as a matrix, is always symmetric and traceless.

For a continuous charge distribution, we have analogously

$$q_{\text{tot}} = \int d^3r_* \rho_q(\vec{r}_*), \quad \vec{D}_E = \int d^3r_* \rho_q(\vec{r}_*) \vec{r}_*,$$

the **zeroth** and **first moments** of the charge distribution with respect to the chosen reference point, and similarly for $\vec{\vec{Q}}_E$.

Continuing to unpack Equation 3.1, the multipole potentials are universal functions of observer position (independent of the nature of the charge distribution):³

$$\psi^{[0]}(\vec{r}) = \frac{1}{4\pi\epsilon_0 r}; \quad \vec{\psi}_i^{[1]}(\vec{r}) = \frac{1}{4\pi\epsilon_0 r^2} \hat{r}_i; \quad \vec{\psi}_{ij}^{[2]}(\vec{r}) = \frac{1}{8\pi\epsilon_0 r^3} (\hat{r}_i \hat{r}_j - \frac{1}{3} \delta_{ij}). \quad (3.3)$$

These formulas define a single monopole field, a set of three dipole fields, and a set of five independent quadrupole fields.

3.2 SOME TAYLOR EXPANSIONS

We need to prove Equation 3.1. First recall some useful facts.

We will often use the series expansions for the functions $(1 + \epsilon)^{\pm 1/2}$ near $\epsilon = 0$:

$$\begin{aligned} \sqrt{1 + \epsilon} &= 1 + \frac{1}{2}\epsilon - \frac{1}{8}\epsilon^2 + \dots \\ 1/\sqrt{1 + \epsilon} &= 1 - \frac{1}{2}\epsilon + \frac{3}{8}\epsilon^2 + \dots \end{aligned}$$

It is good to know how to get these from Taylor's theorem.

Your Turn 3A

You may wonder *how good* these approximations are, how small ϵ must be, and so on. (a) Get a computer to make a graph of the residuals: $f_0(\epsilon) = \sqrt{1 + \epsilon} - 1$, $f_1(\epsilon) = \sqrt{1 + \epsilon} - (1 + \epsilon/2)$, $f_2(\epsilon) = \sqrt{1 + \epsilon} - (1 + \epsilon/2 - \epsilon^2/8)$ and comment. (b) Repeat for $(1 + \epsilon)^{-1/2}$.

Now suppose that the small quantity ϵ is itself given in terms of another small quantity: $\epsilon = \delta + A\delta^2$, and we wish to organize our result as a series in δ . Substituting gives

$$1/\sqrt{1 + \delta + A\delta^2} = 1 - \frac{1}{2}\delta + \delta^2(-\frac{1}{2}A + \frac{3}{8}) + \dots$$

Note how part of the term that was first order in ϵ has entered into the term that is second order in δ .

³The δ_{ij} terms in Equations 3.2 and 3.3 are redundant: You can omit either (but not both) without changing ψ . I included both to emphasize that: (a) The potential at order r^{-3} has a traceless character, no matter what the charge distribution; and (b) the trace of the second moment of charge cannot contribute at all to the parts of the field that are of order r^{-3} or lower.

3.3 PROVE THE ELECTROSTATIC MULTIPOLE FORMULA

Now that we have a precise claim (Equations 3.1–3.3), it’s time to prove it starting from the basic solution for the potential around a point charge. Use Taylor’s theorem to expand the $1/r$ factor:

$$\begin{aligned}\psi(\vec{r}) &= \sum_{\ell} \frac{q_{\ell}}{4\pi\epsilon_0} (\|\vec{r} - \vec{r}_{(\ell)}\|)^{-1/2} = \sum_{\ell} \frac{q_{\ell}}{4\pi\epsilon_0} (r^2)^{-1/2} \left(\left(\frac{\vec{r}}{r} - \frac{\vec{r}_{(\ell)}}{r} \right)^2 \right)^{-1/2} \\ &= \frac{1}{4\pi\epsilon_0 r} \sum_{\ell} q_{\ell} \left(\hat{r}^2 - 2 \frac{\vec{r} \cdot \vec{r}_{(\ell)}}{r^2} + \frac{(\vec{r}_{(\ell)})^2}{r^2} \right)^{-1/2} \\ &= \frac{1}{4\pi\epsilon_0 r} \sum_{\ell} q_{\ell} \left(1 - \frac{1}{2} \left(-2 \frac{\vec{r} \cdot \vec{r}_{(\ell)}}{r^2} + \frac{r_{(\ell)}^2}{r^2} \right) + \left(\frac{3}{8} \right) \left(-2 \frac{\vec{r} \cdot \vec{r}_{(\ell)}}{r^2} + \dots \right)^2 + \dots \right).\end{aligned}$$

The ellipses indicate terms that fall off at long distance faster than r^{-2} .

$$\begin{aligned}&= \frac{1}{4\pi\epsilon_0 r} \sum_{\ell} q_{\ell} \left(1 + \frac{\vec{r} \cdot \vec{r}_{(\ell)}}{r^2} + \frac{1}{r^2} \left(-\frac{1}{2} r_{(\ell)}^2 + \frac{3}{2r^2} (\vec{r} \cdot \vec{r}_{(\ell)})^2 \right) + \dots \right) \\ &= \frac{1}{4\pi\epsilon_0} \sum_{\ell} \left(\frac{q_{\ell}}{r} + \frac{q_{\ell} \hat{r} \cdot \vec{r}_{(\ell)}}{r^2} + \frac{q_{\ell}}{2r^3} \hat{r}_i \hat{r}_j (3\vec{r}_{(\ell)i} \vec{r}_{(\ell)j} - r_{(\ell)}^2 \delta_{ij}) \right).\end{aligned}$$

This result is nearly the one announced earlier. We only need to note that the difference between the last formula and Equation 3.1 is $1/(8\pi\epsilon_0 r^3)$ times

$$-\frac{1}{3} \vec{\mathcal{Q}}_{\text{E},ij} \delta_{ij} = -\frac{1}{3} \sum_{\ell} q_{\ell} (3\vec{r}_{(\ell)i} \vec{r}_{(\ell)j} - r_{(\ell)}^2 \delta_{ij}) \delta_{ij} = 0.$$

3.4 WHY SHOW THIS

Now that we’ve proved the result, it’s worthwhile to ask if it was worthwhile.

The virtue of Equation 3.1 is that each term has been written as a product of a universal, archetypal field (one of the $\psi^{[p]}$ ’s) times a number (one of the moments). The fields are completely independent of details of the source—they are just the possible solutions of the Laplace equation. The moments are independent of position \vec{r} —they are saying how much of each field type is present. The first few moments are a convenient *summary* of the *relevant aspects* of the source for purposes of finding its far fields. Specifically, keeping up to order p (the “ 2^p -pole approximation”) tells us the distant potential up to order $(a/r)^{p+1}$, or equivalently the electrostatic field up to order $(a/r)^{p+2}$. It can be more convenient to work with just these few moments than to include all the irrelevant other details of the full charge distribution.

It also gives us insight to connect *symmetry* of, say, a molecule to the character of its long-range forces (see Section 3.6.3).

3.5 MORE REMARKS

3.5.1

Without a lot of fancy math, we have found that

- If a static, localized charge distribution has any part of its potential that falls as $1/r$, that part of the field must be spherically symmetric.
- If it has any $1/r^2$ term, that part of the field must have a specific angular dependence (it must be dipolar), and so on.

3.5.2 From potentials to fields

This derivation would have been a nightmare had we worked directly with the electric field. So the potential method has practical advantages. After finding the quadrupole fields from the moments, *then* we can take a negative gradient and find the field, should we wish that.

Your Turn 3B

Find the contributions to the electric field coming from the dipole and quadrupole potentials $\vec{\psi}^{[1]}$ and $\vec{\psi}^{[2]}$ appearing in Equation 3.3.

3.5.3 Apparent singularity

Every term in the multipole expansion of ψ is singular at $r = 0$. The corresponding singularities in the electric field are worse still. Is that a problem? No: The expansion is a power series in a/r , so it breaks down (becomes inaccurate) at $r \rightarrow 0$. A smooth distribution of charge will have smooth potential and field. [Similarly, the Earth's gravitational potential looks like $1/r$ outside the Earth, but that doesn't imply there's a black hole at the center.]

3.5.4 Base point dependence

Our expansion of ψ depends implicitly on our choice of the origin of coordinates. If we choose a different origin, then q_{tot} won't change, but in general $\vec{\mathcal{D}}_E$ will (unless $q_{\text{tot}} = 0$), and so will $\vec{\mathcal{Q}}_E$ (unless $q_{\text{tot}} = 0$ and $\vec{\mathcal{D}}_E = 0$), etc. It's not really about coordinate choice: we could have defined moments relative to any reference point \vec{h} via

$$\vec{\mathcal{D}}_E = \sum_{\ell} q_{\ell}(\vec{r}_{(\ell)} - \vec{h}), \text{ etc.}$$

3.5.5 Spherical distributions

Any spherically-symmetric distribution of charge trivially has $\vec{\mathcal{D}}_E = 0$, and not so trivially $\vec{\mathcal{Q}}_E = 0$ also.⁴ In fact *all* moments beyond the 0th are zero: $\psi = q_{\text{tot}}/(4\pi\epsilon_0 r)$ outside any such distribution (Birkhoff's theorem).

3.5.6 Moments that must equal zero by symmetry

Even without spherical symmetry, we sometimes have a shortcut to seeing that some moments must equal zero.

⁴See Problem 3.2.

Any static charge distribution with an inversion symmetry through a point will have $\vec{\mathcal{D}}_E = 0$ when evaluated with respect to that point. Any distribution with a plane of reflection symmetry will have $\vec{\mathcal{D}}_E \cdot \hat{n} = 0$ where \hat{n} is the normal to that plane.

Next, suppose that $+q$ is located at $(0, 0, a)$ and $-q$ is at $(0, 0, -a)$. Then $\vec{\mathcal{D}}_E = (2qa)\hat{z}$. You should compute that $\vec{\mathcal{Q}}_E = 0$ directly from the definition, but here is a more insightful, and generalizable, argument.

Consider any arbitrary static charge distribution. Create a new charge distribution obtained from the given one by inverting all positions, $\vec{r}'_{(\ell)} = -\vec{r}_{(\ell)}$ and also reversing the signs of each charge, $q'_\ell = -q_\ell$.

- The new distribution has $q'_{\text{tot}} = -q_{\text{tot}}$.
- The new distribution has two minus signs in the dipole moment, so $\vec{\mathcal{D}}'_E = \vec{\mathcal{D}}_E$.
- The new distribution has three minus signs in the quadrupole moment, so $\vec{\mathcal{Q}}'_E = -\vec{\mathcal{Q}}_E$.

Back to the specific case with $+q$ located at $(0, 0, a)$ and $-q$ at $(0, 0, -a)$: The transformation described above leaves this distribution unchanged, so every multipole moment is also unchanged.

- We conclude that $q_{\text{tot}} = q'_{\text{tot}} = -q_{\text{tot}}$, so q_{tot} must equal zero (as it does).
- We also find that $\vec{\mathcal{D}}_E = \vec{\mathcal{D}}'_E = \vec{\mathcal{D}}_E$, which is a tautology, so there is no restriction on the dipole moment.
- And we get that $\vec{\mathcal{Q}}_E = \vec{\mathcal{Q}}'_E = -\vec{\mathcal{Q}}_E$, so the quadrupole moment equals zero. In fact, every 2^p -pole moment with p an even integer must be zero.

Your Turn 3C

But octupole, for example, is up for grabs; we cannot conclude it's zero. In fact, try carrying the Taylor expansion out to next order for *this special case* (don't attempt the general case), and find a formula for the part of the potential that falls as r^{-4} .

Your Turn 3D

Think up a charge distribution that, under the above transformation, becomes *minus* itself. Explain why, for any such distribution, every 2^p -pole moment with p an *odd* integer must equal zero. [Hint: Try four point charges all in the xy plane.]

3.5.7 Pure dipole

The two-charge distribution discussed in the previous subsection must have vanishing quadrupole moment, but as mentioned, nothing prevents it from having *octupole* and higher moments.

If we want to get a *purely* dipole field, then we must consider a limiting case, in which the separation $2a$ between the two point charges is sent to zero while increasing the charge so as to hold the dipole moment fixed. Thus in this limit the charge $q = \mathcal{D}_E/(2a) \rightarrow \infty$. That singular limit is called the **pure dipole** or **point dipole** distribution.

3.6 FORCE AND TORQUE ON A CHARGE DISTRIBUTION

3.6.1 Force

Suppose that a localized charge distribution (subsystem 1) sits in an externally created electric potential ψ^{ext} (from system #2). The distribution has a reference point which we take as the origin of coordinates, and consists of charges q_ℓ located at offsets $\vec{r}_{(\ell)}$ from that point.

Suppose that the potential is slowly varying over the size of subsystem 1, and also that subsystem 2 is not significantly distorted by the presence of #1. Consider rigidly displacing each element of #1 by the same vector $\Delta\vec{r}$. Then we can write the potential energy as

$$U(\vec{r}) = \sum_{\ell} q_{\ell} \psi^{\text{ext}}(\vec{r}_{(\ell)} + \Delta\vec{r}).$$

Use a Taylor expansion to show that

$$U = \sum_{\ell} q_{\ell} \psi^{\text{ext}}(0) + \sum_{\ell} q_{\ell} \left. \frac{d\psi^{\text{ext}}}{d\vec{r}} \right|_0 \cdot (\vec{r}_{\ell} + \Delta\vec{r}) + \frac{1}{2} \sum_{\ell} q_{\ell} \left. \frac{d^2\psi^{\text{ext}}}{d\vec{r}_i d\vec{r}_j} \right|_0 (\vec{r}_{\ell} + \Delta\vec{r})_i (\vec{r}_{\ell} + \Delta\vec{r})_j + \dots \quad (3.4)$$

$$= \text{const} + \Delta\vec{r} \cdot (-\vec{E}^{\text{ext}}|_0) \sum_{\ell} q_{\ell} + \left(-\frac{\partial \vec{E}_i}{\partial r_j}\right) \Delta r_i \sum_{\ell} q_{\ell} r_{(\ell)j} + \dots, \quad (3.5)$$

where the dots are subleading terms.

We can now compute the negative gradient to find the force on the charge distribution. In addition to the expected $q_{\text{tot}} \vec{E}^{\text{ext}}$, there is now a new term, which we can write either as $\vec{D}_{E,j} \vec{\nabla} \vec{E}_j^{\text{ext}}$ or as $\vec{D}_{E,j} \vec{\nabla}_j \vec{E}^{\text{ext}}$ (plus terms with higher derivatives). If the distribution is rigid, these formulas can also be written $\vec{\nabla}(\vec{D}_E \cdot \vec{E})$.

Your Turn 3E

Suppose that subsystem 2 is itself a distant, static electric dipole, with net charge zero and dipole moment $\vec{D}_{E(2)}$. Suppose that it is far from dipole 1.

- Find the leading-order contribution to the interaction potential energy, $U(r)$. How does it depend on the separation r ?
- Holding r fixed, consider four possible orientations of the two dipoles: $(\uparrow \cdots \uparrow)$; $(\rightarrow \cdots \rightarrow)$; $(\uparrow \cdots \downarrow)$; $(\rightarrow \cdots \leftarrow)$. Rank-order these according to their interaction potential energy and say which feel attractive and which feel repulsive forces.

Your Turn 3F

Even if the net charge and dipole moment of dipole 1 are both zero, in general there will nevertheless be some interaction: Continue the Taylor expansion, Equation 3.4, to the next order and describe what you get.

3.6.2 Torque

Until now we have allowed the charge distribution to translate (that is, to change its position \vec{r}) but not rotate. If its potential energy changes upon rotation about some

point, then it will experience a net *torque* about that point. To be concrete, consider rotation by $d\theta$ about an axis parallel to \hat{z} and passing through the reference point we used to define the multipole expansion. Then $-dU/d\theta$ is the z component of torque, $\vec{\tau}_3$, computed about the reference point. To find it, we displace each constituent charge from $\vec{r}_{(\ell)}$ to $S\vec{r}_{(\ell)}$, where the infinitesimal **rotation matrix** S is defined by

$$S = \begin{bmatrix} 1 & -d\theta & 0 \\ d\theta & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \cdots = \mathbf{1} + d\mathbf{T} + \cdots .$$

The ellipses denote terms of second and higher order in small quantities. The infinitesimal matrix $d\mathbf{T}$ is $d\theta$ times a matrix called the **generator** of the rotation S . To first order, the potential energy is then

$$\begin{aligned} U &= \sum_{\ell} q_{\ell} \psi^{\text{ext}}(\vec{r}_{\ell} + d\mathbf{T} \cdot \vec{r}_{\ell}) \\ &= \text{const} + \sum_{\ell} q_{\ell} \left. \frac{\partial \psi^{\text{ext}}}{\partial \vec{r}_i} \right|_0 (\vec{r}_{\ell} + d\mathbf{T} \cdot \vec{r}_{\ell}) + \cdots \end{aligned}$$

The change as we rotate is

$$dU = -\vec{E} \cdot d\mathbf{T} \cdot \vec{\mathcal{D}}_E.$$

Notice that the antisymmetric matrix can be written $d\mathbf{T}_{ij} = -\varepsilon_{ij3}d\theta$. So

$$\vec{\tau}_3 = -dU/d\theta = -\varepsilon_{ij3} \vec{E}_i \vec{\mathcal{D}}_{E,j} = (\vec{\mathcal{D}}_E \times \vec{E})_3.$$

More generally, $\vec{\tau} = \vec{\mathcal{D}}_E \times \vec{E}$.

In short, a neutral dipole free to rotate in an external field tends to align with that field: It feels a torque that vanishes when $\vec{\mathcal{D}}_E \parallel \vec{E}$. When aligned, we already found in Equation 3.5 that it further tends to migrate toward stronger $\|\vec{E}\|$ (it feels a force).

3.6.3 Some famous molecules

- Sodium chloride is just a lot of ions (electric monopoles).
- Water and HCl consist of molecules that are neutral but that have big net dipole moments.
- CO₂ has no dipole moment but nonzero quadrupole moment.
- Methane is a tetrahedron.
- Neon is a perfectly spherical charge distribution; all of its multipole moments vanish.

Hmm, that list is also ordered in terms of boiling points! For example, the dipole-dipole attraction of water molecules for each other gives them a strong cohesive force that discourages them from separating (vaporizing). As we go down the list, the intermolecular forces fall faster with distance and the boiling point goes down.

The preceding paragraph is a bit glib. The reasoning may not seem applicable to molecules with permanent dipole moment in liquid state, e.g. water at room temperature, because they are thermally randomized; the average $\langle \vec{\mathcal{D}}_E \rangle = 0$. However,

the random thermal fluctuations of neighboring molecules will be partially *correlated*, leading to nonzero $\langle \vec{\mathcal{D}}_{E(1)} \cdot \vec{\mathcal{D}}_{E(2)} \rangle \neq 0$, and hence decreased energy via Equation 3.5. So there will be a net attraction, in this context sometimes called **Keesom interaction**.

Even neon does liquefy, albeit at a low temperature. So its atoms do have *some* attraction, despite being perfectly spherical in the ground state! To understand this qualitatively, remember that even though the dipole moment's *expectation* is zero, still its instantaneous value will have quantum fluctuations. And these quantum fluctuations again have an energetic tendency to correlate with those of a neighboring atom. This source of electrostatic attraction is sometimes called **London force** or **dispersion interaction**.

Together, the quantum and the statistical correlation attraction effects are sometimes called the **van der Waals** interaction. Van der Waals interactions play a dominant role in some soft matter systems.

3.6.4 Induced dipole moment

We can summarize the preceding remarks by saying that an atom or molecule may have no average dipole moment in isolation, yet may nevertheless develop one in the presence of an external field.

Moreover, real atoms and molecules are not perfectly rigid; they may deform in the presence of an external field, *developing* a nonzero average dipole moment. Although our preceding energy arguments assumed a rigid distribution of charge, they can be equivalently stated by computing forces on each element of the distribution; then we see that the force and torque we computed remain valid even in this situation.

In the simplest situation, the resulting **induced dipole** moment will align with the external field, and then experience a force pushing it toward regions of higher field strength, even if the atom or molecule is neutral and had no dipole moment to begin with.

To get intuition, imagine the molecule as two charges on a Hooke-law spring. Then the induced dipole moment is linearly proportional to the imposed electric field: $\vec{\mathcal{D}}_E = \alpha \vec{E}$, where α is a constant called the molecular **polarizability**. That induced moment in turn feels a force $\alpha \vec{E}_i \vec{\nabla}(\vec{E}_i) = \frac{1}{2} \alpha \vec{\nabla}(E^2)$ directed toward the region of higher field strength.

Note that the electric field appears squared. If it changes sign, that doesn't affect the force. So even the rapidly-varying electric field of a laser beam will create a net force pulling a polarizable object into the beam. This observation is one way to think about **optical tweezers**, which can pull a micrometer-scale object with precisely controlled, piconewton-scale forces. Typically the object is not in vacuum, but what matters is the *difference* between its polarizability and that of the surrounding water (at optical frequency).

FURTHER READING

General: Pollack & Stump, 2002, §3.8; Zangwill, 2013, ch. 4.

Optical tweezers: Perkins, 2014.

Van der Waals interactions: Butt & Kappl, 2018; Israelachvili, 1992.

T₂ Track 2

3.1'a Counting

There's only one kind of monopole field, characterized by only one overall constant of proportionality, q_{tot} . There's *essentially* only one kind of dipole field: You can convert any of the $\vec{\psi}_i^{[1]}$ into any other just by rotating and rescaling, or in other words you can place any dipole in a standard orientation, normalize its overall strength; it then resembles any other.

Quadrupole fields are more interesting. Even if we choose a standardized normalization, the quadrupole tensor $\vec{\mathcal{Q}}_E$ has $5 - 1 = 4$ independent degrees of freedom, too many to be reduced to a standard form by the action of just three rotations.

Indeed, a symmetric matrix like $\vec{\mathcal{Q}}_{E,ij}$ has three real eigenvalues, each of which is rotation-invariant. One of these is redundant because $\vec{\mathcal{Q}}_E$ is traceless, but the ratio of the other two invariantly characterizes the quadrupole. Qualitatively, we may say that some quadrupoles have more symmetry than others, because there is an invariant distinction between those for which two eigenvalues match (**uniaxial symmetry**) and those for which no two match (**biaxial symmetry**).⁵ Try to find concrete examples of each case.

It's an example of the unity of physics that these same concepts arise in liquid crystals.

3.1'b Connection to spherical harmonics

We won't say much about the spherical harmonic functions $Y^{\ell m}$ in this course, but take a moment to examine the quadrupole fields (Equation 3.3), and show that:

- The angular dependence of $\vec{\psi}_{zz}^{[2]}$ is the same as that of Y^{20} .
 - The angular dependence of $\vec{\psi}_{xx}^{[2]}$ is the same as a linear combination of $Y^{2,\pm 2}$ and Y^{20} .
 - The angular dependence of $\vec{\psi}_{xy}^{[2]}$ is the same as a linear combination of $Y^{2,\pm 2}$.
 - The angular dependence of $\vec{\psi}_{zx}^{[2]}$ and $\vec{\psi}_{zy}^{[2]}$ are the same as linear combinations of $Y^{2,\pm 1}$.
- (The dipole fields $\vec{\psi}_i^{[1]}$ are easier to identify with the Y^{1m} .)

If you've studied spherical harmonics, you probably found them at the end of a tortuous derivation in spherical polar coordinates, involving Legendre polynomials and so on. So it's remarkable to see them just pop out effortlessly when we apply Taylor's theorem to a superposition of $1/r$ potentials in cartesian coordinates.

T₂ Track 2

3.6.3'a

Pursuing the quadrupole term may seem like hairsplitting—it's subleading in powers of the small quantity. But:

- Sometimes the dipole moment of a neutral atom or molecule is zero for symmetry reasons, e.g. in CO_2 . In that case, the quadrupole term is the dominant one.
- There is also a multipole expansion for electromagnetic *radiation*, as we'll see. Here, too, if the transition dipole moment is zero, still the atom or molecule can radiate via its quadrupole moment. But that radiation is weaker in classical electrodynamics (the

⁵Why can't all three match?

rate is smaller in the quantum version), a reflection of its higher-multipole character, just as we found that the static quadrupole field falls faster than a dipole field.

- In *gravitational* radiation, there's *never* a dipole component; the leading order behavior involves the time-dependence of the quadrupole moment of mass (unless that's zero).

3.6.3'b

Interestingly, no fundamental particle is known to have a permanent electric dipole moment. A nonzero moment would break “CP” symmetry, and although the Standard Model predicts such breaking, it does so very weakly. For example, the predicted moment for the electron is $\approx (10^{-38} e) \text{ cm}$, whereas in 2018 the experimental bound was $\mathcal{D}_E \lesssim (10^{-29} e) \text{ cm}$. (In contrast, many fundamental particles, such as electrons and neutrons, have readily measurable *magnetic* dipole moments, which do not violate CP symmetry.)

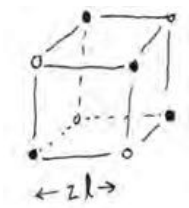


Figure 3.1:

PROBLEMS

3.1 *Electrostatic multipole*

- Find the electrostatic potential far away from two point charges, q and $-q$ fixed on the z -axis at $z = a$, $-a$ respectively. Give only the first two nonzero terms in the expansion of the potential in powers of r/a . Comment on why your answer “had to” behave this way.
- Consider point charges q , $-2q$, q located on the z -axis at $z = a$, 0 , $-a$ respectively. Discuss the part of the electrostatic potential at $r \gg a$ that falls off as r^{-4} .

3.2 *Just a moment*

Consider a spherically symmetric charge distribution: $\rho_q(\vec{r}) = f(r)$ is independent of the polar and azimuthal angles.

- Clearly such a distribution must have vanishing dipole moment, because there’s no spherically-symmetric vector (other than zero). But work this out directly from the definition of dipole moment as an integral over the distribution.
- More precisely, the dipole moment *computed about the point of symmetry* must be zero. Repeat your calculation but this time suppose that the distribution, while spherically symmetric, is centered about some point \vec{h} other than the origin.
- Repeat (a) but for the distribution’s quadrupole moment. This time we can’t just say, “It must equal zero because there’s no such thing as a rotationally-invariant rank-2 tensor,” because that’s *not a true statement*. So work it out and then discuss.
- Repeat (b) for the quadrupole moment.

3.3 *Tetrahedron*

- Consider four identical point charges q rigidly fixed at the vertices of a tetrahedron (solid dots shown in Figure 3.1), and $-4q$ fixed at its center. The distance from the center to any vertex is a . Find the dipole moment and quadrupole tensor for this distribution. What do these results imply about the behavior of the electric field to leading order in a/r ? [Remark: A convenient construction of a tetrahedron begins with a cube centered on the origin, that is, with vertices $(\pm\ell, \pm\ell, \pm\ell)$ for some length ℓ related to a . You can select four of the cube’s eight vertices and use them as the vertices of the desired tetrahedron.]
- Does your result appear to be relevant to the behavior of some well known small molecule? Does it explain a big qualitative difference between that molecule’s properties and those of, say, water?

3.4 Benzene I

We can idealize an isolated aromatic molecule, such as benzene, as follows. Charge $-q$ is spread uniformly throughout a thin ring (annulus) in the xy plane, that is, the region $w < \sqrt{x^2 + y^2} < 2w$. A point charge $+q$ is all concentrated at the center of the ring. Find the static electric potential far from this charge distribution to leading nontrivial order in powers of r^{-1} for $r \gg w$. Also find the static electric field \vec{E} in the same approximation.

3.5 Benzene II

Let us idealize the benzene molecule as six positive point charges q in the xy plane at the vertices of a regular hexagon, each a distance a from the origin. There is also a washer-shaped ring (an annulus) with uniform charge density and total charge $-6q$, also in the xy plane, extending from radius b out to c .

Find the dipole and quadrupole moments of this charge distribution in terms of q , a , b , and c and comment. For example, maybe your result has something to do with the fact that benzene is more volatile than water, despite being a more massive molecule.

3.6 Ab Ovo

a. A rigid ellipsoid is defined by the equation $(x/a)^2 + (y/b)^2 + (z/c)^2 \leq 1$. It has net charge q uniformly distributed throughout its volume, balanced by a point charge $-q$ at the center. Find the quadrupole tensor of this charge distribution. [Hint: Take the reference point to be its center.]

Suppose that the ellipsoid in (a) has $a = c = 1$ m and $b = 0.5$ m (so it's "oblate"). The center of the ellipsoid is placed at the origin of coordinates, in an external electrostatic potential $\psi(\vec{r}) = \vec{\alpha} \cdot \vec{r} + \beta yz + \gamma(x^2 - y^2)z$. Here $\vec{r} = (x, y, z)$ and $\vec{\alpha}$, β , and γ are constants with appropriate dimensions.

- b. Under what conditions may we use the multipole expansion to calculate the force on this charge distribution?
- c. Assuming the condition in (b) is met, find the force on the ellipsoid exerted by this field to leading order in the multipole approximation.

3.7 Multipole math

Get a formula for $\nabla^2(r^{-5}\vec{r}_i\vec{r}_j)$ where ∇^2 is the Laplace operator and the indices $i, j = 1, 2, \text{ or } 3$. If your answer is nonzero, explain how the expression $r^{-5}\vec{r}_i\vec{r}_j$ is admissible as a term in the multipole expansion of the electrostatic field.

3.8 Quadrupole

Four point charges are placed in the xy plane as follows:

- 1,2: Charges $+q$ are placed at points $(0, \pm a, 0)$.
- 3,4: Charges $-q$ are placed at points $(\pm a, 0, 0)$.

An observer sits at a position \vec{r} , with $r \gg a$.

- a. Work out the monopole, dipole, and quadrupole moments of this distribution. Is it uniaxial or biaxial?
- b. Substitute the nonzero moment(s) into the general formula to find the far-potential of this static distribution to leading order in $1/r$. *Hint:* If you don't remember the formula, you can instead write out the four contributions to the point-charge

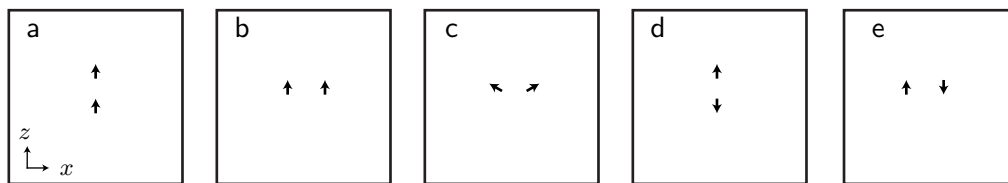


Figure 3.2:

potential felt at \vec{r} , approximate each one by using Taylor's theorem, and add them, keeping only the leading term in the expansion in powers of a/r .

- Differentiate your answer to (b) to get an analytic formula for the electric field. Simplify by evaluating only on the plane $z = 0$.
- Use a computer to display this vector field, after first normalizing it to unit length. On the same axes, but in a different color, display the exact answer for the electric field of the charge distribution (1-4) and comment.

3.9 Pictures at an exhibition

In this problem, you are to make graphical representations of electrostatic fields corresponding roughly to charge distributions encountered in simple molecules. Section 3.5.7 described a limiting charge distribution whose potential consists of only the dipole term of Equation 3.1. By computing minus the gradient of such a function you can find the corresponding electric field. In this problem, you are to find and display exact expressions for the fields outside *pairs* of pure dipoles that are *not* located at the origin.

- Figure 3.2a represents two dipoles of equal strength, both directed along $+\hat{z}$, but located at $(0, 0, \pm a)$. Write an expression for the electric field. Use a computer to evaluate this vector field on a suitable grid of points in the xz plane covering the region $-3a < x < +3a$ and $-3a < z < +3a$.⁶ (Arrange your grid so that the two singular points $(0, 0, \pm a)$ are not themselves grid points.) Normalize the vector field to a constant length, to make it easier to see each arrow,⁷ and display it. Then get your computer to find and show some representative streamlines of the vector field in a separate plot. You don't need a specific value for the length scale a (why not?), but for a molecule it could be, say, 0.2 nm. You also don't need a specific value for the strength of the dipoles (why not?).
- Repeat for the situation in (b): two dipoles directed along $+\hat{z}$ located at $(\pm a, 0, 0)$.
- Repeat for (c): two dipoles tilted $\pm 60^\circ$ away from \hat{z} towards the $\pm x$ -axis and located at $(\pm a, 0, 0)$. What familiar molecule might this model?
- Repeat for (d): similar to (a), but the dipoles oppose each other. What familiar molecule might this model?
- Repeat for (e): similar to (b), but the dipoles oppose each other. This might model two familiar molecules electrostatically sticking to each other (like what?)

⁶One way to approach the problem is to evaluate the potential first, then compute its gradient numerically. You'll get numerically better results, however, if you instead evaluate the electric *field* directly.

⁷Changing the normalization does not change the streamlines.

- f. The fields in examples (a–c) all fall into one group, and examples (d–e) into a different group, based on some common characteristic. What is it and what does it mean physically?

3.10 *Electrostatic multipoles II*

a. Consider equal but opposite point charges q and $-q$ held at a fixed positions $(0, 0, \pm d/2)$. Expand the electrostatic potential ψ at a distant point \vec{r} in power series in d/r , and find the *first two* nonvanishing terms in the series. Then comment.

b. Now flesh out the general argument I made to get the same conclusion: Consider the combined operation that reverses the sign of all charges, and also inverts all positions through the origin. The charge distribution in (a) is unchanged by this operation. Examine the behavior of the generic expressions for the monopole, dipole, and quadrupole moments of a charge distribution (formulas in Brau), and find their behavior under this operation. Thus argue that some of them must equal zero for the particular distribution in (a).

3.11 *3D field line plot*

Learn how to get a computer to create 3d streamplots, and show them for an electric dipole field. Look at various viewing angles till you find one that is most informative.

3.12

[Not ready yet.]

3.13 *Classical model of FRET*

We can get some insight into fluorescence resonance energy transfer by using ideas from newtonian physics. Imagine an oscillator representing the charge cloud (electric dipole moment) of a donor fluorophore. The donor gives rise to an electrostatic force on a second oscillator, which represents the acceptor fluorophore. Suppose that this force $f(t)$ has fixed angular frequency ω_D (determined by the donor's excited state), and amplitude J (determined by the donor's state and the distance to the acceptor):

$$f(t) = J \cos(\omega_D t). \quad (3.6)$$

We model the acceptor's electron cloud as a point object with mass m . It's attached to a fixed object (representing the molecule's heavy nuclei) by a spring, with spring constant k . Moreover, the acceptor slowly dissipates energy to "friction," which represents energy loss from the acceptor, for example by fluorescence. Calling the friction constant η , Newton's law $f = ma$ states that the donor's position $x(t)$ obeys

$$m \frac{d^2}{dt^2} x = -kx - \eta \frac{d}{dt} x + f. \quad (3.7)$$

To simplify this equation, define new symbols $\omega_A = \sqrt{k/m}$, $\eta = \eta/m$, and $K = J/m$, and eliminate k , η , and J by writing them in terms of the new symbols.

- a.** After a short transient, the solution $x(t)$ will oscillate at frequency ω_D . So consider the trial solution $x(t) = A \cos(\omega_D t) + B \sin(\omega_D t)$. Find the constants A and B in terms of K , η , ω_D , and ω_A .
- b.** In the steady state that we are studying, the rate at which the acceptor gets energy from the donor must equal the rate at which it loses energy to dissipation, which is $\mathcal{P} = \eta(dx/dt)^2$. Evaluate this for your solution.

- c. The quantity you found in (b) is always positive, but it oscillates. Actually, we only need its time-average $\langle \mathcal{P}(\omega_D, \omega_A) \rangle$, which is given by a simpler expression than the answer to (b). Find that.
- d. Actually, the donor and acceptor are not in precisely known states: Rather, each moves within a *distribution* of possible states, with varying values of ω_D, ω_A . The average rate of energy transfer is then the average of the quantity you found in (c), weighted by the corresponding distributions $\wp_D(\omega_D)$ and $\wp_A(\omega_A)$:

$$\langle\langle \mathcal{P} \rangle\rangle = \int d\omega_D \wp_D(\omega_D) \int d\omega_A \wp_A(\omega_A) \langle \mathcal{P}(\omega_D, \omega_A) \rangle.$$

To simplify this expression, suppose that the damping η is very small. Then your expression from (c) is very sharply peaked near $\omega_D = \omega_A$. Exploit this fact by letting

$$\omega_D = \bar{\omega} - \frac{1}{2}\Delta\omega; \quad \omega_A = \bar{\omega} + \frac{1}{2}\Delta\omega,$$

and changing integration variables from ω_D, ω_A to $\bar{\omega}, \Delta\omega$. Then approximate your answer to (c) by replacing $\Delta\omega$ by 0 everywhere, except for the one term in the denominator responsible for making the sharp peak. With this approximation, you can readily do the integral over $\Delta\omega$.

- e. The donor creates a dipole field, which shakes charges on the acceptor. Imagine the acceptor as having a fixed axis \hat{p}_A and a charge q that is only able to move along that axis. Then the force driving that charge's motion is the product of the charge times the component of the donor's electric field along \hat{p}_A . From this information, the behavior of dipole fields, and your calculations, comment on how the energy transfer rate depends on the separation and relative orientation of donor and acceptor.

CHAPTER 4

Curvilinear Coordinates and Separation of Variables

4.1 SEPARATION OF VARIABLES IN THE LAPLACE EQUATION

Earlier sections stressed that cartesian coordinates are “good” because Maxwell’s equations look exactly the same in any cartesian coordinate system. We will say much more on that subject later. But you already know that some non-cartesian coordinate systems are “pretty good” because Maxwell’s equations look *almost* the same in them, and that such a system can be extremely convenient for certain kinds of problems, for example, those with certain symmetries. Here we will sharpen the notion of “pretty good” to introduce systems for which the Laplace operator is **separable**.

The potential method is especially useful for electrostatic problems involving conductors because $\psi = \text{constant}$ is a nice boundary condition for the Laplace equation. However, the Laplace equation is still a 3D partial differential equation. It’s nicer still when we can reduce it to a few ordinary differential equations. We’ll see that this works if we choose coordinates with the separability property, and if the boundary of our conductor looks simple in those coordinates.

In fact, the Laplace operator is separable in ordinary cartesian coordinates x, y, z , but many problems have boundaries that don’t look simple in those coordinates. So we will sometimes use other separable coordinate systems, collectively called **curvilinear**, in which the Laplace operator is again separable but certain frequently-encountered shapes look nice.

4.2 FAMILIAR STUFF

4.2.1 Cartesian coordinates

Because the Laplace operator is the sum of a term not involving y, z , plus a term not involving x, z , plus a term not involving x, y , we can look for solutions of the form $A(x)B(y)C(z)$, where $A'' = \kappa A$, $B'' = \lambda B$, $C'' = \nu C$, and $\kappa + \lambda + \nu = 0$. If our boundaries also look nice in cartesian coordinates, then this coordinate choice can be useful.

4.2.2 Plane polar coordinates

For simplicity, let’s warm up with just two dimensions. Let $x = r \cos \varphi$ and $y = r \sin \varphi$ as usual. You already know what the Laplace operator looks like in these coordinates, but let’s redo that derivation in a way that will generalize easily.

Define two vector fields $\vec{e}_{(r)}(r, \varphi)$ and $\vec{e}_{(\varphi)}(r, \varphi)$ as the motions we make as we

vary one or the other of the new coordinates:

$$\vec{e}_{(r)} = \frac{\partial \vec{r}}{\partial r}, \quad \vec{e}_{(\varphi)} = \frac{\partial \vec{r}}{\partial \varphi}. \quad (4.1)$$

Note that the first of these is the same as the unit vector \hat{r} , but the second is not the same as $\hat{\varphi}$. Instead, $\vec{e}_{(\varphi)} = r\hat{\varphi}$. It will soon be convenient that these two vector fields are everywhere perpendicular to each other.

We want to formulate the Laplace operator in terms of the new variables. Let f be a function on the plane, and abbreviate $f_r = \partial f / \partial r$, $f_\varphi = \partial f / \partial \varphi$, etc. The cartesian components of the gradient can be written

$$\vec{\nabla} f = \mathbf{J} \begin{bmatrix} f_r \\ f_\varphi \end{bmatrix}, \quad \text{where } \mathbf{J} = \begin{bmatrix} \frac{\partial r}{\partial x} & \frac{\partial \varphi}{\partial x} \\ \frac{\partial r}{\partial y} & \frac{\partial \varphi}{\partial y} \end{bmatrix}. \quad (4.2)$$

Here's a useful trick to get an expression for the Laplace operator reexpressed in terms of our new coordinates. Let g be any function and let f be a function that is zero everywhere except in some small region. Then Equation 4.2 gives that

$$\int d^2 \vec{r} \vec{\nabla} f \cdot \vec{\nabla} g = \int d^2 \vec{r} [f_r, f_\varphi] \mathbf{J}^t \mathbf{J} \begin{bmatrix} g_r \\ g_\varphi \end{bmatrix}. \quad (4.3)$$

However, we also have

$$\int d^2 \vec{r} \vec{\nabla} f \cdot \vec{\nabla} g = \int d^2 \vec{r} [\vec{\nabla} \cdot (f \vec{\nabla} g) - f \nabla^2 g] = - \int d^2 \vec{r} f \nabla^2 g. \quad (4.4)$$

In the last step, we used the divergence theorem to express the first term as an integral over the boundary. That term is zero because of our assumption about f .

We have found two expressions that must agree for any choice of f . To get a formula for $\nabla^2 g$, then, we will just manipulate the right-hand side of Equation 4.3 until there are no more derivatives on f , then compare to the right-hand side of Equation 4.4.

First we need an explicit formula for the 2×2 matrix $\mathbf{J}^t \mathbf{J}$. It's messy to compute \mathbf{J} directly, because once we compute $\partial r / \partial x$ etc. we must then reexpress everything as functions of r and φ . Luckily there's a shortcut to make that step unnecessary. Note that \mathbf{J}^{-1} is the matrix

$$\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi} \end{bmatrix}.$$

(Proof: The stated matrix transforms cartesian derivatives to polar, the opposite of what \mathbf{J} does.) The nice property about \mathbf{J}^{-1} is that its rows are the components of $\vec{e}_{(r)}$ and $\vec{e}_{(\varphi)}$ defined by Equation 4.1. Thus, we may write

$$\mathbf{J}^{-1} (\mathbf{J}^{-1})^t = \begin{bmatrix} \|\vec{e}_{(r)}\|^2 & \vec{e}_{(r)} \cdot \vec{e}_{(\varphi)} \\ \vec{e}_{(r)} \cdot \vec{e}_{(\varphi)} & \|\vec{e}_{(\varphi)}\|^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.$$

We want $\mathbf{J}^t \mathbf{J}$, which is the inverse of the preceding result. But the inverse of a diagonal matrix is easy. That's the benefit we get from the fact that the $\vec{e}_{(i)}$'s are everywhere perpendicular to each other.

Now we can return to Equations 4.3–4.4. For the first of these, we use the result just found for J^tJ then integrate by parts:

$$\int (r \, dr \, d\varphi) (f_r g_r + f_\varphi r^{-2} g_\varphi) = - \int (dr \, d\varphi) \left(f \frac{\partial}{\partial r} (r g_r) + f r^{-1} \frac{\partial}{\partial \varphi} g_\varphi \right).$$

We want to rephrase this expression into a form resembling the right side of Equation 4.4, so multiply and divide by r :

$$= - \int (r \, dr \, d\varphi) f \left[r^{-1} \frac{\partial}{\partial r} (r g_r) + r^{-2} g_{\varphi\varphi} \right]. \quad (4.5)$$

Equation 4.4 says that the last expression equals $-\int d^2\vec{r} \, f \nabla^2 g$ for any function f that vanishes outside a small region. For example, f could be a bump function localized anywhere. The only way that these expressions could be equal for arbitrary f if the terms in square brackets of Equation 4.5 are equal to $\nabla^2 g$, and of course this is a familiar formula:

$$\nabla^2 g = r^{-1} \frac{\partial}{\partial r} \left(r \frac{\partial g}{\partial r} \right) + r^{-2} \frac{\partial^2 g}{\partial \varphi^2}.$$

4.2.3 Plane polar payoff

If we have a circularly-symmetric, 2D problem, we can entertain trial solutions of the form $\psi(\vec{r}) = A(r)B(\varphi)$. Then $\nabla^2 \psi = 0$ becomes

$$0 = \frac{B_{\varphi\varphi}}{B} + \frac{r}{A} \frac{\partial}{\partial r} (r A_r).$$

The first term is completely independent of r . The second term is completely independent of φ . Their sum is the constant 0, so each term must *separately* be a constant. That reduces our problem to two decoupled ordinary differential equations.

If moreover our boundary conditions can be stated simply in these coordinates, for example as $A(R) = 1$, then we win.

4.2.4 Some food for thought

It is definitely *not* the case that the Laplace operator can be written as $\partial^2/\partial r^2 + \partial^2/\partial \varphi^2$! Einstein asked himself, “What’s special about some coordinate systems (such as cartesian) that makes the Laplace operator look simpler in them than in others (such as polar)?” Following that road led him into general relativity.

For now, we just notice that in polar coordinates the Laplace operator still looks *fairly* simple, whereas in general coordinates it does not.

4.2.5 Three dimensions

Your Turn 4A

Run through all these steps for cylindrical and spherical polar coordinates, to see how they yield the rather mysterious formulas found on the inside cover of any E&M textbook.

4.3 A SPHERICAL CONDUCTOR IN A UNIFORM FIELD

Consider a spherical conductor of radius R between two distant, infinite, flat, parallel, charged plates. We choose an origin of coordinates centered on the center of the sphere and set up spherical polar coordinates with polar axis along \hat{z} , which is normal to the planes.

At the sphere, $\psi(r = R)$ must be independent of θ and φ ; we may take its value to be zero. Far from the sphere, we get the same uniform electric field we'd have had without the sphere, so $\psi \rightarrow Cz$ where C is a constant related to how much charge is on the plates. Now we want ψ everywhere.

Our problem isn't spherically symmetric, but at least it's axially symmetric, so we get a shortcut: ψ will be independent of φ . The boundary condition is simple in polar coordinates (it involves only r), so let's seek a solution of the form $A(r)B(\theta)$. In order to solve the Laplace equation in the space between sphere and plates, we need functions A, B satisfying

$$A^{-1}(r^2 A')' = \lambda \quad \text{for } r \geq R \quad \text{and} \quad B^{-1} \frac{1}{\sin \theta} (\sin \theta B')' = -\lambda \quad \text{for } 0 \leq \theta \leq \pi.$$

In the first equation, prime means d/dr ; in the second one, prime means $d/d\theta$. Let $\mu = \cos \theta$, so $d\mu = -\sin \theta d\theta$. Thus, the second equation becomes the **Legendre equation**:

$$B^{-1} \frac{d}{d\mu} \left((1 - \mu^2) \frac{dB}{d\mu} \right) = -\lambda. \quad (4.6)$$

One solution is $B = \text{const}$, which gives $\lambda = 0$. But that's a spherically symmetric solution, and our boundary condition is not spherically symmetric. The next most complicated solution to Equation 4.6 is $B = \mu$, which gives $\lambda = 2$. Put that back into the equation for A :

$$(r^2 A')' = 2A.$$

This equation is homogeneous, so maybe a power-law solution will work: $A(r) = r^p$. Substituting shows that $p = 1$ or -2 both work:

$$\psi(r, \theta) = (Cr + Dr^{-2}) \cos \theta.$$

Indeed, $\psi \rightarrow Cz$ at $r \rightarrow \infty$, as desired.¹ And we can satisfy the boundary condition by choosing $D = -RC$.

We're done. The second term is familiar from the multipole expansion, but the first term is new: Multipole missed it because it does not fall off with distance.

4.4 LIGHTNING ROD

Our Founder, Ben Franklin, was not the first to discover that electric discharges tend to occur at sharp points. It's not at all clear that he even did the dangerous and stupid experiment that he almost, but not quite, claimed to have done with the kite.

¹The apparent singularity at $r \rightarrow 0$ is not a problem because this solution is only to be used outside the sphere.

(Others did it, and not all survived.) Ben's breakthrough was to point out that putting a grounded, pointy conductor on top of a house could save it from burning down.²

How shall we find the electric field just outside a charged ellipsoid? Conformal transformation only works for 2D problems. The multipole expansion only gives the potential far away from an object, and even then requires that we know the charge distribution in advance. Spherical harmonic expansion goes bad in the limit of interest, where the ellipsoid is very pointy.³ Finite-grid numerical solution also loses accuracy in that limit.

Really, we'd like an exact solution. Spherical polar coordinates were good for spherical conductors. Remarkably, a different curvilinear system exists that can handle an ellipsoid. Problem 4.1 introduces a coordinate system in which an ellipsoid is the level set of one of the coordinates. By following the steps in this chapter, you can also discover that the Laplace operator is separable in those coordinates as well, and hence get an exact solution for the lightning-rod problem almost as readily as in Section 4.3.

4.5 DIVERGENCE OPERATOR

So far we have restricted attention to the Laplace operator, but the rest of vector calculus can be cast into curvilinear coordinates when that's helpful. Just remember that as always, it's better to know where the formulas came from, so that you can use them correctly.

We will eventually want to know the divergences of special vector fields such as

$$\vec{V} = \frac{1}{r} \hat{x} e^{ikr}. \quad (4.7)$$

Here r is distance from the origin, and k is a scalar constant. We will first do this the hard way, just to highlight how much easier our second approach is.

4.5.1 Hard way

The hard way at first seems easier: Just look in any book and find the formula

$$\vec{\nabla} \cdot \vec{V} = r^{-2} \frac{\partial}{\partial r} (r^2 \vec{V}_r) + \frac{1}{r \sin \theta} \left[\frac{\partial}{\partial \theta} (\sin \theta \vec{V}_\theta) + \frac{\partial \vec{V}_\varphi}{\partial \varphi} \right].$$

But it's tricky to apply this formula to Equation 4.7 correctly. Note that $\hat{r} = [\sin \theta \cos \varphi, \sin \theta \sin \varphi, \cos \theta]^t$, so

$$\hat{x}_r = \hat{r} \cdot \hat{x} = \sin \theta \cos \varphi, \quad \hat{x}_\theta = \hat{\theta} \cdot \hat{x} = \cos \theta \cos \varphi, \quad \hat{x}_\varphi = \hat{\varphi} \cdot \hat{x} = -\sin \varphi.$$

After you substitute these into the famous formula, you still must do a lot of algebra to find

$$\vec{\nabla} \cdot \vec{V} = \sin \theta \cos \varphi (-r^{-2} + ik/r) e^{ikr}.$$

²Some people objected—lightning strikes were manifestations of divine will, which humans would defy at their peril. Ben was persistent.

³It's also bad in another limit of interest, where the ellipsoid is squashed very flat to a thin disk.

4.5.2 Easy way

Instead of a black-box formula, let's do it from scratch. Use the product rule, $\vec{\nabla} \cdot (f\hat{x}) = \hat{x} \cdot \vec{\nabla} f + f\vec{\nabla} \cdot \hat{x}$ and take $f = r^{-1}e^{ikr}$. The second term is zero because the cartesian components of \hat{x} are all constants. Thus

$$\vec{\nabla} \cdot \vec{V} = \hat{x} \cdot \hat{r} \frac{\partial}{\partial r} (r^{-1}e^{ikr}) = \sin\theta \cos\varphi (-r^{-2} + ik/r)e^{ikr}.$$

4.6 PLUS ULTRA

There are a total of 11 coordinate systems in which the 3D Laplace operator is separable, plus two more that are almost as good. See the references.

FURTHER READING

Pollack & Stump, 2002, §5.2.1.

About the famous list of 13 separable coordinate systems: See Weisstein, Eric W. 'Laplace's Equation': <http://mathworld.wolfram.com/LaplacesEquation.html>.

Also see books: Landau and Lifshitz *Mechanics* ch. VII; Arfken; and Morse and Feshback.

PROBLEMS

4.1 NSOM probe

In class I motivated the study of a long, thin metal probe in a uniform background electric field, which is relevant to apertureless nearfield scanning optical microscopy.

We can define an ellipse as the locus of points in the xz plane that solve

$$(x/\alpha)^2 + (z/\beta)^2 = 1,$$

where the constants α and $\beta > \alpha$ are called the "semimajor" and "semiminor" axes, respectively. Thus 2β , the "major axis," is the distance between the two most distant antipodal points (the "poles"), and 2α , the "minor axis," is the distance between the two least distant antipodal points.

Define two points P_{\pm} on the \hat{z} axis, located at $z = \pm\sigma$. For any other point, let r_{\pm} be the distances from that point to P_{\pm} . We can specify this point either by its x, y, z values, or its cylindrical polar coordinates ρ, φ, z , or by new coordinates ξ, η , and φ . Here φ is the same as in cylindrical coordinates and

$$\xi = (r_+ + r_-)/(2\sigma) \quad \eta = (r_+ - r_-)/(2\sigma).$$

- Show that the surface $\{\xi = \xi_0\}$ is what you get by rotating an ellipse about its axis. Find its major and minor axes in terms of σ and ξ_0 .

We wish to find the field outside a conductor whose surface is the one in (a), in the presence of a background electrostatic field that's uniform at infinity. But first some math.

- b. Express ρ and z in terms of ξ and η . [*Hint:* Express $\xi\eta$ and $(\xi^2 - 1)(1 - \eta^2)$ in terms of ρ and z , then think.]
- c. Thus express x, y, z in terms of ξ, η , and φ . Differentiate to find the vector $\vec{e}_{(\xi)} \equiv \partial\vec{r}/\partial\xi$, and similarly $\vec{e}_{(\eta)}$ and $\vec{e}_{(\varphi)}$. These three vectors have a very nice property similar to the one we found in class for plane polar coordinates — what is it?
- d. Use (c) to express the volume element d^3r in terms of $d\xi d\eta d\varphi$. Find the region in ξ - η plane corresponding to the region outside the surface in (a).
- e. Use (c,d) to express the integral $\int d^3r \vec{\nabla}\Upsilon \cdot \vec{\nabla}\psi$ in the coordinates ξ, η , and φ . Here ψ is any function independent of φ , while Υ , also independent of φ , is nonzero only in some small region of ξ and η .
- f. Use integration by parts to work out the Laplace operator $\nabla^2\psi$ in these coordinates, for the case where ψ is independent of φ .
You're ready to begin the problem, which is to find the electrostatic potential in the region outside the conductor, subject to the boundary conditions:

$$\psi = 0 \text{ on the surface, } \quad \psi \rightarrow -E_\infty z \text{ far away.}$$

We seek an exact solution $\psi = A(\xi)B(\eta)$ by separation of variables.

- g. Translate the boundary conditions above into conditions on A and B . Find a solution to the equation for B meeting those conditions.
- h. Now that you know the dependence on η , write the required ordinary differential equation and boundary conditions on the function A .
- i. The equation is second order, so it has two independent solutions. You can readily guess one of them from the boundary condition at infinity, and substitute to confirm that it works.
- j. But we need the other solution too, in order to enforce the surface boundary condition. I don't remember how to find the other solution, but *Mathematica*, Maple, and Wolfram Alpha do. So ask one of them (unless you know all about obscure special functions).
- k. Finish the problem: Work out the magnitude of the electric field just outside the conductor at its two poles, and compare this value to the applied E_∞ .
- l. Consider a conductor with major axis $100 \mu\text{m}$ and minor axis $0.5 \mu\text{m}$ and evaluate your expression in (k) for the field ratio numerically. Then make a contour plot of the normalized electrostatic potential ψ/E_∞ in the xz -plane.
- m. Here is a related problem that's easy after you invent the above formalism: Now consider a metal ellipsoid carrying *nonzero* net charge Q but totally isolated, that is, the electric field approaches zero at infinity. Adapt the procedure of parts (a–j) to find the exact solution for the potential. Then make a contour plot of the electrostatic potential ψ/Q in the xz -plane for the same geometry as in (l).

[*Remark:* At optical frequencies, most metals are not really well described by our assumption of perfect conductors. Moreover, the geometry of a probe approaching a surface is probably closer to a hyperboloid near a plane than to the geometry assumed in this problem. Nevertheless, ξ - η coordinates are still useful in realistic treatments of NSOM probes and their field-focusing properties.]

4.2 *Razor's edge*

A thin metal plate in vacuum is placed in the half-plane $y \approx 0$, $x < 0$ for all z . Thus the edge of the plate is the z axis. The electrostatic potential ψ is constant everywhere on the plate, but the plate may be charged.

We can seek a solution by using separation of variables in cylindrical coordinates, for which the plate occupies the half-plane with $\varphi = \pm\pi$:

$$\psi(\rho, \varphi, z) = f(\rho) \cos(\varphi/2).$$

This guess for the angular dependence satisfies the boundary condition $\psi(\rho, \pm\pi, z) = 0$. Write and solve the equation satisfied by the radial function f . Comment on how your solution behaves near the edge.

CHAPTER 5

Capacitors

5.1 FRAMING

Section 2.1 pointed out that what makes electrodynamics physics, not math, is that we must constantly seek idealizations of systems too complex to handle explicitly. Thus, in an electron beam we may be able to use Newton's laws of motion with electrostatic forces on each electron, but many other situations involve condensed (solid or liquid) matter, which is packed with too many charges to handle explicitly. This chapter will introduce one sort of idealization that is useful in many real situations: A dielectric material.

5.2 PARALLEL PLATES IN VACUUM

Charge q is placed on a flat planar conductor with area Σ . Charge $-q$ is placed on another such conductor, parallel and a distance a away from the first in the $+x$ direction. Both conductors are much bigger in y and z than a . By symmetry, the electric field must point along \hat{x} . Let $\sigma_q = q/\Sigma$ be the surface charge density on the left plate.

Use the electric Gauss law to find that between the planes, $\vec{E}_x = \sigma_q/\epsilon_0$. Integrate $-\vec{E}$ along x to find the potential throughout the gap, and its total change $\Delta\psi = \psi(0) - \psi(a) = \sigma_q a/\epsilon_0$. We define the **capacitance** as the constant of proportionality relating charge and potential:

$$C = q/\Delta\psi. \quad (5.1)$$

Mnemonic: If you have large "capacity," you can store lots of charge without developing a big potential. That's why q is in the numerator and $\Delta\psi$ is in the denominator.

For this system, neglecting edge effects, $C = \epsilon_0 \Sigma/a$ per unit area. The natural SI unit for capacitance is coulombs per volt, which is called the farad: $1 \text{ F} = 1 \text{ coul/volt}$.

5.3 ENERGY STORED

We can now imagine pulling a charge dq away from the negative plate and depositing it on the positive plate. If dq is positive, then we must do work against the electric field to accomplish this: $(dq)\Delta\psi = dq(q/C) = d(\frac{1}{2}q^2/C)$. If we wish to build up charge starting from zero, then we must do a total amount of work

$$\mathcal{E} = \frac{1}{2}q^2/C.$$

Rephrasing using Equation 5.1 gives the stored electrostatic potential energy as

$$\mathcal{E}/(\text{volume}) = \frac{\epsilon_0}{2} \vec{E}^2. \quad (5.2)$$

That’s interesting: The total energy is proportional to the volume, as though it were stored in *empty space* with a density depending quadratically on the field. We’ll need to do a lot more work before we can be confident about this suspicion, however.

5.4 CYLINDRICAL CONDUCTORS IN VACUUM

Consider a long, straight metal cylinder (“wire”) carrying linear charge density $\rho_q^{(1D)}$ (coulombs per meter). Inside any good conductor the electric field must equal zero, so the potential must be a constant. Outside, the potential must obey the Laplace equation: $\nabla^2\psi = 0$. Cylindrical coordinates make this problem straightforward: $\psi(r, \varphi, z) = B \ln(r/r_0)$ outside the cylinder (and uniform inside). Here B is a constant related to $\rho_q^{(1D)}$ and to the radius (thickness) of the wire.

Your Turn 5A

Find that relation.

r_0 is an uninteresting constant; changing it just adds a constant to the potential.

Next consider *two* long, parallel cylinders with charge densities $\pm\rho_q^{(1D)}$. We can superpose two solutions of the above form. The result will again solve the Poisson equation with charge 0 outside each cylinder. It won’t be exactly constant on the two cylinders’ surfaces, but it will be approximately so if their radii are much smaller than their separation.

Your Turn 5B

Work out the potential difference in this approximation between the two wires as a function of $\rho_q^{(1D)}$, the radii, and the separation. From this, work out an approximate formula for the capacitance per unit length of this “twinlead” cable.

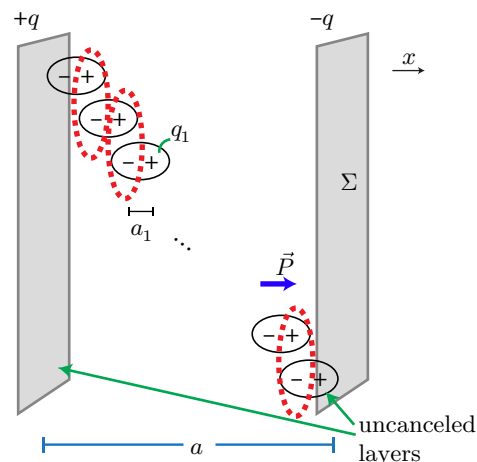
5.5 PARALLEL PLATES WITH MEDIUM

Now fill the gap between conductors with some nonpolar atoms or molecules, maybe liquid neon, or more prosaically some kind of oil. What matters is that there be no free charges, so that the material is an insulator. In this context such a material is generically called a **dielectric**.

Each atom/molecule has no dipole moment in isolation, but nevertheless it can deform under the influence of an external field, and so develop an *induced* dipole moment.¹ Figure 5.1 suggests that the resulting uniform polarization density will lead to canceling net charge density in the interior (see the dashed red lines in the figure), but not on the two boundaries of the medium. Suppose that each molecule separates

¹Recall Section 3.6.4.

Figure 5.1: Polarizable “molecules” with density ρ_{stuff} fill the gap between parallel conducting plates, creating a density of dipole moment $\vec{P} = q_1 a_1 \rho_{\text{stuff}} \hat{x}$. On the left, a layer of thickness a_1 contains uncanceled $-q_1$ per molecule, so the total bound charge near that plate is $(a_1 \rho_{\text{stuff}} \Sigma)(-q_1)$. The bound surface charge density on the left plate is therefore $\sigma_b = \hat{n} \cdot \vec{P}$, because the outward-pointing unit vector is $\hat{n} = -\hat{x}$. Similarly, on the right side there is again a partial cancellation of free and bound charges.



charge q_1 by distance a_1 , and that they are packed with volume density ρ_{stuff} . Then the uncanceled net charge forms a thin “bound” layer at the interface, with areal density

$$\sigma_b = \hat{n} \cdot \vec{P}, \quad (5.3)$$

where \vec{P} is the volume density of induced dipole moment (polarization) and \hat{n} is the unit vector perpendicular to the surface and directed away from the medium.² We will refer to σ_b as the **bound surface charge density**, because it can’t escape from the medium, nor even move freely within it; in contrast, the **free charge** on either plate could be moved elsewhere by connecting a wire to the plate. We’ll call the areal density of free surface charge σ_f .

On the left side of the figure above, \vec{P} and $-\hat{n}$ point rightward, so the bound charge is negative and indeed partially cancels the charge we put on the left plate.

Most dielectric materials have zero polarization in the absence of an externally applied field.³ So it’s natural to suppose that it will have a Taylor expansion, whose leading term is $\vec{P} \propto \vec{E}$. The constant of proportionality is called the **bulk polarizability** of the medium. It is traditionally expressed as $\epsilon_0 \chi_e$, where χ_e is called the **dielectric susceptibility**. The relation $\vec{P} = \epsilon_0 \chi_e \vec{E}$ is our first example of a **response function**. Unlike laws of Nature, it is approximate (for example, we assumed the response was linear in the field strength) and nonuniversal (different materials will have different values of χ_e).⁵

Applying the electric Gauss law at the left-hand plate gives

$$\vec{E}_x = (\sigma_f + \hat{n} \cdot (\epsilon_0 \chi_e \vec{E})) / \epsilon_0 \quad \text{where} \quad \hat{n} = -\hat{x}.$$

²Chapter 49 will look at the general situation, where the polarization density may be nonuniform.

³Nothing forbids a permanent electric polarization, analogous to the phenomenon of permanent magnetism, and indeed materials with this property, called **ferroelectrics** or **ferroelectric**, are known.⁴ Other materials develop permanent e.d.m. when mechanically strained; they are called **piezoelectric**.

⁵We also assumed that the induced polarization points parallel to the applied field. Chapter 50 will explore interesting phenomena that arise in materials that don’t obey that assumption.

Solving gives

$$\vec{D}_x = \sigma_f, \quad (5.4)$$

where the **electric displacement** is defined as⁶

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P} \quad (\text{generally}). \quad (5.5)$$

In the context of our specific model we have

$$\vec{D} = \epsilon \vec{E}, \quad (\text{linear isotropic medium}) \quad (5.6)$$

where the **permittivity** ϵ of the medium⁷ is

$$\epsilon = (1 + \chi_e)\epsilon_0. \quad (5.7)$$

Thus, the effect of the medium is simply to replace the vacuum permittivity ϵ_0 by a *larger* effective value. Instead of accounting explicitly for every charge in the medium, we can simplify by *forgetting* it and making this one substitution. Equation 5.6 is called a **constitutive relation** for the material in the capacitor.

The same argument as earlier now gives capacitance as

$$C = \epsilon \Sigma / a, \quad (5.8)$$

which is greater than the vacuum value.

5.6 ENERGY PUZZLE

Can we still maintain our idea of energy as stored in the space between the plates? At first it looks bad: Our previous formula gave $\frac{1}{2}\epsilon_0 \vec{E}^2$. We could minimize this expression by assuming enough polarization to completely neutralize the applied charge, and hence get *zero* energy storage! That doesn't seem right.

To see what went wrong, remember that the polarization surface charge arose from *deformation* of molecules (or atoms) throughout the gap. The molecules will resist this deformation. They therefore store energy; the final polarization must involve optimizing the *total* energy (field plus deformation). For weak deformation we may assume a Hooke-law (linear) force law.

To keep things simple, this section will make the unjustified assumption that each dipole responds to the spatially-averaged electric field \vec{E} .⁸ Again imagine an individual molecule as a pair of charges $\pm q_1$, with a Hooke-law spring constant k_1 controlling their separation. Thus $a_1 = q_1 \vec{E}_x / k_1$. Suppose that the polarizable objects are distributed with density ρ_{stuff} .

⁶To understand this name, notice that the second term of this expression really involves the movement of charges in the dielectric. Maxwell initially imagined the first term as having a similar origin, a “displacement” of charge in the æther.

⁷Many authors use the notation advocated here. Beware, however, that some older works write the permittivity as $\epsilon\epsilon_0$, so for them the symbol ϵ is what we would call ϵ/ϵ_0 , a *dimensionless* quantity often called the **dielectric constant**. To avoid confusion, we will not introduce any symbol for dielectric constant.

⁸Section 5.7 will show that this is justified when the dielectric is of low density, and will give an improved derivation for dense matter.

Your Turn 5C

Show that in this model, $\vec{P}_x = q_1^2 \vec{E}_x \rho_{\text{stuff}} / k_1$, and so

$$\epsilon_0 \chi_e = q_1^2 \rho_{\text{stuff}} / k_1$$

for the low-density medium we are studying.

To understand this result from an energy viewpoint, let's write down the total stored energy (electric field plus elastic deformation energy):

$$\mathcal{E}/(\text{volume}) = \frac{1}{2} \epsilon_0 \vec{E}_x^2 + \frac{1}{2} k_1 a_1^2 \rho_{\text{stuff}} = \frac{1}{2} \left(\epsilon_0 + \frac{q_1^2}{k_1} \rho_{\text{stuff}} \right) \vec{E}_x^2 \quad (5.9)$$

$$= \frac{1}{2} (\epsilon_0 + \epsilon_0 \chi_e) \vec{E}_x^2 = \frac{1}{2} \epsilon \vec{E}_x^2 = \frac{1}{2} \frac{\sigma_f^2}{\epsilon}. \quad (5.10)$$

We see that, for fixed free charge introduced on the plates, the system finds an equilibrium: a compromise between minimizing the two kinds of energy. The energy is smaller than it would have been with no polarization at all (because $\epsilon > \epsilon_0$). The energy is also lower than it would have been if the material had polarized enough to eliminate the electric field altogether. But it's not *zero*, as suggested at the start of this section!

Instead of using the Hooke law, we could have left a_1 arbitrary. Then Equation 5.9 has two terms that are analogous to a mechanical system: two springs in series. We know that that system minimizes its total energy by distributing overall deformation between the springs, rather than assigning all of it to just one of them. Similarly, our capacitor will minimize total energy by canceling some, but not all, of its imposed free charge with bound charge, resolving the paradox at the start of this section.

The reduction of total energy when we introduce a dielectric material at fixed free charge implies a *force* that pulls that material into the gap. For example, a fluid dielectric will be pulled into the space between charged plates, even if it must overcome gravity to do so.⁹

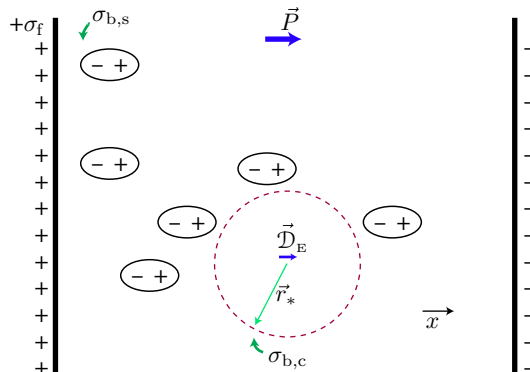
5.7 DENSE MEDIUM

The preceding section warned that it is not really justified to assume that each polarizable molecule responds to the spatially-averaged field. This may be surprising: Often, when a medium is uniform on macroscopic length scales, we may work with spatially averaged quantities, such as the local velocity in fluid mechanics. This section will make some more ad hoc assumptions, but we will at least see why this reasoning breaks down in the presence of long-range forces such as electrostatics.

We again imagine a parallel-plate capacitor with a uniform, polarizable medium between the plates. This time, however, we will single out one particular molecule for study, and set up polar coordinates centered on it. This dipole of interest responds to the net electric field created by all charges *except itself*. Those charges include the free charge on the distant plates, as well as bound charges in the medium. To

⁹See Problem 5.4.

Figure 5.2: A spherical surface has been drawn surrounding one polarizable molecule in a medium. We regard the interior of this surface as a “cavity” containing only a point dipole representing the molecule.



improve, if only slightly, on our previous derivation, we now suppose that we may treat the medium as continuously and uniformly polarized, *except* in a spherical cavity surrounding the dipole of interest, with volume equal to $1/\rho_{\text{stuff}}$. After all, surely it is foolish to insist on a continuum distribution below the molecular scale.¹⁰

Figure 5.2 illustrates our idealization. An induced dipole of unknown moment \vec{D}_E sits at the center of a spherical cavity. It feels a local field \vec{E}_{loc} with three contributions: from free charge with surface density $\pm\sigma_f$ at the plates, from bound charges $\pm\sigma_{b,p}$ at the plates, and from bound charges $\sigma_{b,c}$ on the surface of the cavity. The free charge density is given, but we must find all of the bound charge densities and the average dipole moment density \vec{P} .

Because \vec{P} points to the right in the figure, we define b as its magnitude via $\vec{P} = b\hat{x}$. The same reasoning as in Section 5.5 gives the bound charge density at the left plate as $\sigma_{b,p} = (-\hat{x}) \cdot \vec{P} = -b$.

Let \hat{r}_* be the unit vector from the dipole of interest to a point on the surface of the cavity and let a be its radius. Then the unit vector perpendicular to the surface and “outward” (away from the bulk material) is $-\hat{r}_*$, and the bound surface charge at the cavity is $\sigma_{b,c} = (-\hat{r}_*) \cdot \vec{P} = -b \cos \theta_*$, where θ_* is polar angle measured from \hat{x} . The figure illustrates why the cosine factor is needed: For example, at $\theta_* = \pi/2$ the molecular distortion is parallel to the surface and no net bound surface charge arises.

We wish to find the electric field at the center of the cavity, $\vec{E}_{\text{tot}}(0)$, because that is what acts on the molecule we are studying. It receives a contribution from the charges on the plates:

$$\vec{E}_{\text{plate}} = \frac{\sigma_f + \sigma_{b,p}}{\epsilon_0} \hat{x} = \frac{\sigma_f - b}{\epsilon_0} \hat{x}. \quad (5.11)$$

The other contribution, \vec{E}_{cav} , comes from $\sigma_{b,c}$. To find it, first use the potential formula Equation 2.5

$$\psi_{\text{cav}}(\vec{r}) = \frac{1}{4\pi\epsilon_0} \int_{\text{sphere}} d^2\Sigma \frac{\sigma_{b,c}(\vec{r}_*)}{\|\vec{r} - \vec{r}_*\|}.$$

¹⁰More sophisticated treatments consider a spherical hole that is much larger than the molecular scale, but in the end they still make assumptions, and still give only rough answers except for extremely special media such as liquid helium. For a *much* more sophisticated treatment see Zangwill, 2013, chapt. 6.

$$\begin{aligned}\vec{E}_{\text{cav}}(\vec{r}) &= -\vec{\nabla}\psi_{\text{cav}} = \frac{-1}{4\pi\epsilon_0} \int_{\text{sphere}} a^2 d(\cos\theta_*) d\varphi_* \frac{-b \cos\theta_*}{(-2)\|\vec{r} - \vec{r}_*\|^3} 2(\vec{r} - \vec{r}_*) \\ \vec{E}_{\text{cav}}(\vec{0}) &= \frac{-b}{4\pi\epsilon_0} \frac{a^2}{a^3} \int_{\text{sphere}} \cos\theta_* d(\cos\theta_*) d\varphi_* (-\vec{r}_*) \\ &= \hat{x} \frac{b2\pi}{4\pi\epsilon_0} \int_{-1}^1 \cos^2\theta_* d(\cos\theta_*) = \hat{x}b/(3\epsilon_0).\end{aligned}$$

The induced dipole moment equals the total field times the molecular polarizability:

$$\vec{\mathcal{D}}_{\text{E}} = \alpha(\vec{E}_{\text{plate}} + \vec{E}_{\text{cav}}(\vec{0})) = \hat{x}\alpha\left(\frac{\sigma_{\text{f}} - b}{\epsilon_0} + \frac{b}{3\epsilon_0}\right).$$

We have now established a connection between the induced moment $\vec{\mathcal{D}}_{\text{E}}$ and the strength b of the average polarization \vec{P} . But the same connection applies to every molecule, so we also have

$$\vec{P} = \rho_{\text{stuff}}\vec{\mathcal{D}}_{\text{E}}.$$

Combining the last two displayed equations and recalling that $\vec{P} = b\hat{x}$ gives

$$\frac{b}{\rho_{\text{stuff}}} = \frac{\alpha}{\epsilon_0}\left(\sigma_{\text{f}} - \frac{2b}{3}\right).$$

Solving for b gives

$$\vec{P} = \hat{x}\sigma_{\text{f}}\left(\frac{2}{3} + \frac{\epsilon_0}{\alpha\rho_{\text{stuff}}}\right)^{-1}.$$

Now compare the last formula for \vec{P} to Equation 5.11 to find

$$\vec{P} = \vec{E}_{\text{plate}} \frac{\epsilon_0}{\alpha\rho_{\text{stuff}} - \frac{1}{3}}.$$

Writing this as $\epsilon_0\chi_{\text{e}}\vec{E}_{\text{plate}}$ at last gives the dielectric susceptibility in terms of molecular polarizability:

$$\chi_{\text{e}} = \frac{\alpha\rho_{\text{stuff}}}{\epsilon_0 - \alpha\rho_{\text{stuff}}/3}. \quad \text{Clausius-Mossotti formula} \quad (5.12)$$

Many materials conform to approximate versions of this formula, although with other factors of order unity in place of the factor of 1/3 that came from our simplified approach.

Returning to the start of this section, consider subdividing a substance more and more finely, $\rho_{\text{stuff}} \rightarrow \infty$ while holding $\alpha\rho_{\text{stuff}}$ fixed. Equation 5.12 shows that even in this limit, the susceptibility disagrees with the naive continuum version in Your Turn 5C.

5.8 POLAR FLUID MEDIUM

So far we have considered molecules that have no intrinsic dipole moment, but that can polarize by deforming slightly. We can also consider a medium consisting of polar

molecules that initially are randomly oriented; an important example is liquid water. No energetic penalty must be paid to polarize such a medium if it is a fluid; the molecules can simply align to create net polarization.

So once again we may at first worry that this system would always cancel an applied \vec{E} , at least up until the molecules had reached perfect alignment. But there is a price to be paid, even if it is not energy: Aligning the molecules costs *entropy*, or equivalently raises the *free* energy of the system. In a weak field, the compromise between free energy cost and electrostatic energy reduction will be mathematically similar to what we previously worked out, again leading to an incomplete cancellation of the electric field. Although the polarizability is therefore not infinite, for water at room temperature it is quite high: $\epsilon \approx 80\epsilon_0$. Interestingly, solid water (ice) has a much smaller permittivity, because its molecules are not free to reorient. Like any other molecules, they may also deform, but the effective spring constant for deformation is much stiffer than the one for alignment.

The reorientation of water molecules is accompanied by frictional loss as they rub against their neighbors. When \vec{E} is shaken at microwave frequency, the associated heating can be considerable, and indeed you know that a microwave oven heats pure liquid water, with its strong and mobile dipoles, much faster than it does glass, plastic, or even ice.¹¹

At higher frequencies, however, the reorientation response is too slow to follow the field fluctuations. That is, the permittivity of water is strongly frequency-dependent and much closer to that of ice in the optical range than at lower frequency. Later we will find that the polarizability of a medium slows the transmission of light, and indeed, the velocity of light in liquid water is only a little slower than in vacuum (3/4 as fast).

5.9 PARTITIONING OF IONS AT A FLUID INTERFACE OR CELL MEMBRANE

Imagine an oil-water interface. An ion, for example Na^+ , is dissolved in the water. Although Section 5.5 only considered uniform fields, a similar argument suggests that the ion's surrounding electric field will be reduced relative to vacuum by the polarizability of water. Hence its net energy, in this situation called the ion's **Born self-energy**, is also reduced. For water, that reduction will be substantial.

Now suppose that the ion crosses the interface to the oil side. The low polarizability of oil means that the self-energy *increases*. Thus, even though there is no material barrier at the interface, *ions will segregate to the water side*, following the Boltzmann probability rule.

Living cells are surrounded by a few-nanometer thick membrane that is essentially oil. The water either side of this **bilayer membrane** contains lots of ions, but they will not cross the membrane because of the high Born self-energy they would incur in the intermediate states while crossing. That is, *cell membranes are electrically insulating*, despite being so thin. Because of that thinness, such membranes have extremely high

¹¹Food contains salty water, which is a conductor. The electric fields in the applied microwaves therefore induce currents, which give rise to additional heating by the usual resistive mechanism.

capacitance per unit area (Equation 5.8).¹² The passage of ionic current into or out of a cell is strictly controlled by **ion channels** embedded in the membrane. Chapters 10–11 will show how the interplay of high capacitance and controlled passage leads to the phenomenon of nerve impulses.

5.10 BOUNDARY CONDITIONS

The same discussion we gave at an interface between a conductor and vacuum continues to hold at interfaces between a conductor and a dielectric, a dielectric and vacuum, or between two different dielectrics (Section 2.6): E_{\perp} can jump at such an interface, because free charges (in a conductor) or bound charges (in one or both dielectrics) can be localized at the surface.

For example, suppose that a dielectric material 1 faces vacuum and let \hat{n} be the perpendicular to a point on the surface that points away from the material. Suppose there is no free surface charge; for example, the dielectric could have been neutral before an external field was applied. Then Equation 5.3 and the electric Gauss law give that (Figure 5.3a)

$$\hat{n} \cdot (\vec{E}^{[\text{vac}]} - \vec{E}^{[1]}) = \hat{n} \cdot \vec{P}^{[1]}/\epsilon_0. \quad (5.13)$$

If we know the polarization in terms of the electric field, for example via $\vec{P}^{[1]} = \epsilon_0 \chi_e \vec{E}^{[1]}$, then we get a condition for how E_{\perp} jumps. Rephrasing in terms of the displacement gives the simple form

$$\Delta D_{\perp} = 0. \quad \text{dielectric boundary} \quad (5.14)$$

Turning now to the components of \vec{E} that are tangential to the surface of a conductor, integrating both sides of $\vec{\nabla} \times \vec{E} = 0$ over a small area that passes through the interface shows that \vec{E}_{\parallel} may *not* jump as we cross the boundary (Figure 5.3b):

$$\Delta \vec{E}_{\parallel} = 0. \quad \text{dielectric boundary} \quad (5.15)$$

For example, these two components must equal zero just outside a conductor.

FURTHER READING

See also Pollack & Stump, 2002, ch. 4 and 6.

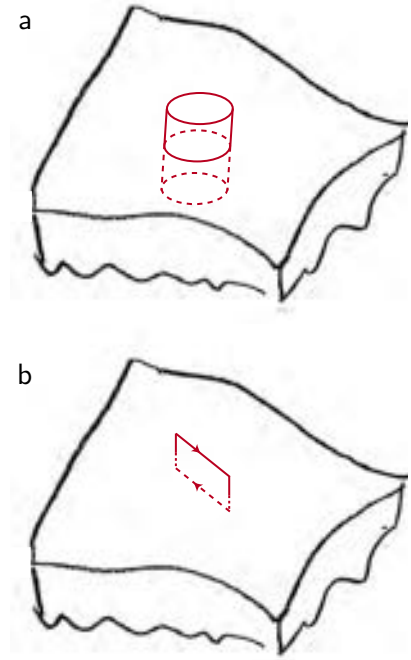
Electrets: <https://en.wikipedia.org/wiki/Electret>

Piezoelectricity: <https://en.wikipedia.org/wiki/Piezoelectricity>

Bioelectricity, Coulter counter: Grodzinsky, 2011.

¹²Chapter 8 will discuss how this prediction was confirmed experimentally.

Figure 5.3: [Sketches.] **Boundary conditions near a conductor.** (a) The short *red cylinder* has one end cap just outside a conductor and the other just inside. Integrating the electric Gauss law over it, and using the divergence theorem, shows that the component of \vec{E} perpendicular to the surface can have different values just inside and outside the conductor, at least over length scales larger than the thickness of the surface charge layer. (b) The *red rectangle* has one of its longer edges just outside a conductor and the other just inside. Integrating the curl-free condition, and using Stokes's theorem, shows that any component of \vec{E} parallel to the surface must have the same values just inside and just outside the conductor.



PROBLEMS

5.1 Capacitor fun

A simple capacitor is a device formed by two insulated conductors adjacent to each other in vacuum. If equal and opposite charges are placed on the conductors, there will be an electrostatic potential difference between them. The ratio of the magnitude of charge on one of them to the magnitude of $\Delta\psi$ is called capacitance, measured in farads (or F). Using the electric Gauss law, calculate the capacitance of:

- a. Two large flat conducting sheets of area A separated by small distance d .
- b. Two concentric conducting spheres with radii a and b .
- c. Two concentric conducting cylinders of length L , large compared to their radii a, b .
- d. What is the diameter of the outer conductor in a vacuum-filled coaxial cable whose central conductor is a cylindrical wire of diameter 1 mm and whose capacitance per unit length is $0.5 \mu\text{F}/\text{cm}$?

5.2 Twinlead cable

Two long, cylindrical conductors of radii a, b are parallel and separated by distance d , which is much bigger than either a, b .

- a. Let $c = \sqrt{ab}$ and show that the capacitance per length is approximately proportional to $(\ln(d/c))^{-1}$. Find the constant of proportionality.
- b. What diameter wire (in mm) would be necessary to obtain $0.1 \text{pF}/\text{cm}$ if the separation is $d = 5 \text{mm}$? (That symbol means 10^{-12}F .)

5.3 Can you take the pressure?

Let us understand why there is such a strong tendency for matter to be electrically neutral. Consider a spherical balloon filled with gas. At atmospheric pressure and room temperature, a balloon of radius $R = 17\text{ cm}$ contains about one mole of gas.

- Now assume that we remove *one* electron from *one out of every million* gas atoms, while holding R fixed. The remaining uncompensated charges will repel each other, so they will distribute themselves on the surface of the sphere. Find the electrostatic self-energy of this assembly of charge, in joules.
- Differentiate your result in (a) to find the pressure (change of energy per change of volume) exerted on the balloon. Express your answer as a multiple of atmospheric pressure, which is about 10^5 N/m^2 .

5.4 Fluid-filled capacitor

[Not ready yet.]

5.5 Biocapacitor

- Show that the electric field outside a line of charge in vacuum is $\vec{E} = \hat{r}\rho_q^{(1D)}/(2\pi\epsilon_0 r)$. Here r is the distance from the observation point to the line and \hat{r} is the unit vector pointing from the line, perpendicular to it, and passing through the observation point. $\rho_q^{(1D)}$ is the charge per unit length on the line, which we assume to be uniform.
- Suppose that instead the charge is distributed on a cylinder of radius R_1 , and that an equal and opposite charge is distributed on a larger cylinder, with radius R_2 . The two cylinders are concentric (they have the same centerline). Use (a) to state the capacitance per unit length of this coaxial “cable.”
- The neurons in your body each have a long cylindrical “output line” called the axon. It’s got a good conductor inside (**axoplasm**) and outside (salt water), separated by a thin insulating layer (cell membrane). The insulating layer has permittivity ϵ , which may be different from ϵ_0 , but with this modification we ought to be able to apply your result in (b) to find the capacitance of the membrane. And yet, people always use a formula that looks quite different from yours, namely, the parallel-plate capacitor formula $C = \Sigma\epsilon/\delta$. Here Σ is the total area of the membrane and δ is its thickness. Resolve this apparent discrepancy. *Hint:* An axon may typically have diameter $1\ \mu\text{m}$. Its membrane may typically have thickness $2\ \text{nm}$.

5.6 Microwave heating

[Not ready yet.]

5.7 Measure ϵ_0

If you know about fringe fields, neglect them in this problem.

- A flat, circular plate, of radius $r = 14\text{ cm}$, in vacuum, carries total charge q . Write an approximate expression for the electric field strength \vec{E} very near the plate as a function of distance d to the plate (so $d \ll r$).
- A second such plate is held close to, and parallel to, the first one, and carries total charge $-q$. Find the force $d\vec{f}$ on each surface area element dA of the second plate due to the charge on the first plate.
- Find the electric field strength and the electrostatic potential drop $\Delta\psi$ between the plates as a function of their separation d .

- d. A mechanical force f is required to maintain the second plate at a fixed distance d . Find this force as a function of r , d , $\Delta\psi$ and physical constants.
- item I did this experiment and measured the $\Delta\psi$ needed to balance a force of 10^{-2}N at separation $d = 0.5\text{ cm}$. I got $\Delta\psi \approx 1055\text{ volt}$. Use this information to determine the approximate numerical value of ϵ_0 (that is, *don't* use the standard value listed in books). [Note: The plates are were air, so really you are finding the dielectric susceptibility ϵ_{air} . But air is similar to vacuum.]
- e. Make a prediction for the speed of light based on your answer to (e).

CHAPTER 6

Electrohydrostatics

It is a pleasure to find out how different observable phenomena of the physical world fit together. . . . It is the discovering of the connection between physical phenomena and describing them by mathematical analysis, rather than the analysis itself, which is interesting.

— *G. I. Taylor*

6.1 FRAMING

Think about soap bubbles you have observed. The usual closed ones come to a hydrostatic equilibrium where they stop wobbling and assume a spherical shape (Figure 6.1a). Since childhood people have told you, “A sphere has the smallest surface area for given volume, so surface tension dictates that shape.” Indeed, when we see videos of astronauts on the Space Station creating zero-gravity blobs of soup and then slurping them up, the equilibrium shapes are spherical, again due to the air–liquid interfacial tension. Even with gravity and wind resistance, raindrops are also roughly spherical.

Think some more. A wire frame dipped in soap solution can lead to other kinds of equilibrium surface shapes. Dip a frame shaped like a Pringle potato chip and you get a saddle-shaped film (figure panel (b)). Dip two circular rings and if you’re careful you can get a catenary-type surface spanning them (figure panel (c)). (With even greater care, you could in principle get a cylinder with closed caps.)

But many other shapes *never* arise: You never get a free, sharp point, nor indeed any sort of cone (Figure 6.2e)—not for open nor closed soap films, nor for water droplets.¹ And yet, Figure 6.3 shows a conical surface of a fluid–fluid interface, in equilibrium, displaying a sharp point. We’d like to answer questions like:

- How is this possible at all?
- Are there restrictions on the sort of conical shapes we can realize?
- Is there technological relevance? (Ans: Yes, lots.)

Another goal of this chapter is to foreshadow some ideas about tensors for future elaboration.

The next sections introduce many symbols, which we summarize here for reference:

¹It is possible to obtain equilibrium soap films with *line* singularities that terminate on corners. We will tacitly exclude these from consideration in this chapter. Also, crystals of *frozen* water are a completely different matter.

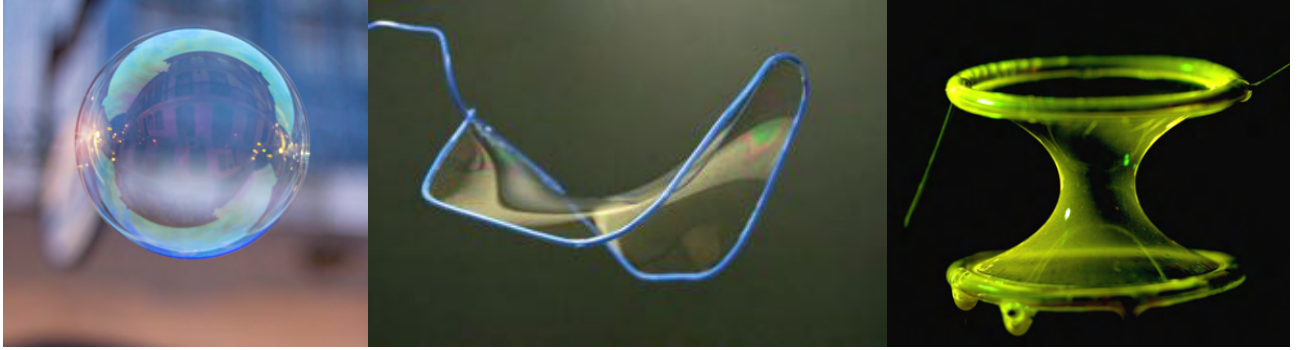


Figure 6.1: Some fluid-fluid-fluid interfaces in mechanical equilibrium. (a) Closed (distinct inside and outside regions). (b) Open. (c) Open. [(c): Photo by RE Goldstein, A Pesci, and K Moffatt.]

s	arclength coordinate along a curve
$\Delta(s)$	deviation from tangent line
κ	curvature of a curve in a plane
ξ	small perpendicular displacement of a curve or surface
ΔT	jump in surface tension upon crossing a 1D barrier
F	line tension
L	total arclength of a curve in a plane
u, v	local coordinates centered on a point \mathbf{P}
B_{ij}	description of surface shape near a point; k_i , its eigenvalues when expressed in normal coordinates
$H = (k_1 + k_2)/2$	mean curvature of a surface in 3-space
$G = k_1 k_2$	Gauss curvature of a surface in 3-space
Δp	jump of fluid pressure upon crossing a 2D interface
T	interfacial tension of a fluid-fluid interface or free film
$N(r), M(\theta)$	functions used in separation of variables
θ_0	polar angle for a cone with opening angle $2(\pi - \theta_0)$

6.2 SOME GEOMETRY

6.2.1 Curves in a plane

Before we discuss surfaces with interfacial tension, let's warm up by studying *curves* in a *plane*, possibly with *line* tension, for example, a stretched rubber band. Consider the curve shown in Figure 6.4a. At the point \mathbf{P} , construct the tangent line as shown. As we walk away from that point, the perpendicular distance $\Delta(s)$ from the tangent to the curve begins to grow as a function of arclength s (unless the line is straight at \mathbf{P}). The Taylor series of $\Delta(s)$ has no linear term (that's what it means to be tangent). The quadratic term describes whether the curve is straight or not at \mathbf{P} . Writing that term as $\frac{1}{2}\kappa(\mathbf{P})s^2$, the coefficient $\kappa(\mathbf{P})$ has dimensions \mathbb{L}^{-1} and is called the **curvature** at \mathbf{P} .

The curvature as just defined also controls how a new curve, obtained by displacing

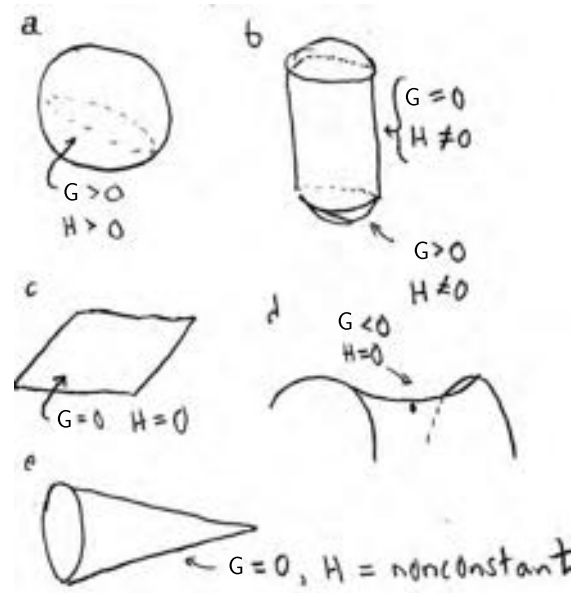


Figure 6.2: Some illustrative 2-surfaces. (a–b) A closed soap bubble can reach hydrostatic equilibrium as a surface of constant curvature, possibly confined on a wire frame: (a), a free-standing sphere; (b), a cylinder with bulging caps. (c–d) An open soap film can reach equilibrium as a surface of constant *zero* curvature: (c), flat plane; (d), saddle. (e) A sharp conical point never arises as an equilibrium shape—except in the situation shown in Figure 6.3.

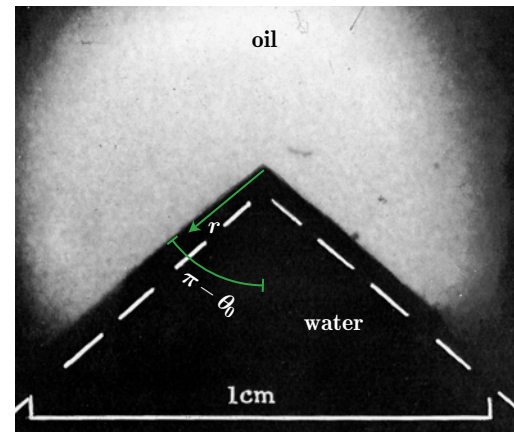


Figure 6.3: Conical point of an oil–water interface (side view). The surface with polar angle θ_0 is a cone with half-angle $\pi - \theta_0$. [From Taylor, 1964, Fig. 9.]

the original by an amount $\xi(s)$ along the perpendicular, will have slightly different arclength from the original; see Figure 6.4b. Intuitively, a straight line is the shortest curve joining two given points,² because if there’s a bend, “you could instead take a

²In flat euclidean space.

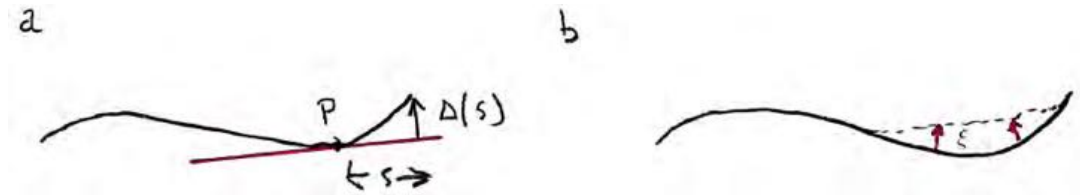


Figure 6.4: (a) $\Delta(s)$ is perpendicular distance from a curve to its tangent line at **P**, after we travel arclength s away from **P** within the curve. The quadratic part of $\Delta(s)$ is a measure of curvature at **P**. (b) The curve has been shortened (*dashed*) by displacing it a perpendicular distance $\xi(s)$.

shortcut.” To make that more precise, in Problem 6.1 you’ll show that:

- a. To first order in ξ , the total length change is the integral along the curve of arclength times $-\kappa\xi$ (a local formula).
 - b. In contrast, the **area** in the plane occupied by one side of the curve grows, and the other side shrinks, by an amount proportional to the line integral of arclength times ξ without any factor of curvature (another local formula).
- (6.1)

6.2.2 Mechanical equilibrium of an interface in a plane

Now imagine a floating skimmer designed to contain an oil slick. If you pin it between two fixed points and put it under **line tension** F , and there’s no oil slick, then it will minimize length by assuming a curve of constant, zero curvature (a straight line). If one side confines an oil slick, however, then the skimmer will bulge out: It is pulled sideways by the higher air–water **interfacial tension** on the oil-free side.³ It now assumes a shape that is a circular arc: Constant, but not zero, curvature. Let’s understand why.

To understand the situation, think in terms of energy. In mechanical equilibrium, the line tension F that we apply to the skimmer is constant along its length. The interfacial tension difference ΔT is also constant, set by properties of water and oil. Mechanical equilibrium also requires that the curve’s shape minimize total energy. We described a small shape disturbance by a function $\xi(s)$. Idea 6.1 says that the corresponding first-order change in energy has two parts: The interfacial tension difference ΔT multiplies $\int ds \xi$, whereas the line tension F multiplies⁴ $-\int ds \xi \kappa$. In mechanical equilibrium, the net first-order variation of free energy must be zero:

$$0 = \int ds (\Delta T - F\kappa)\xi.$$

This relation must hold for any displacement $\xi(s)$, so:

$$\text{Mechanical equilibrium selects a shape that has constant curvature } \kappa = (\Delta T)/F. \quad (6.2)$$

Idea 6.2 explains two familiar situations:

³Try floating a thin loop of string on water and add a drop of detergent to the inner surface.

⁴The minus sign reflects a particular choice of which direction of deviation from the tangent will be called positive (Figure 6.4a,b). Strictly speaking, interfacial tension involves the *free* energy cost.

- When we float an open string on a surface, $\Delta T = 0$. If we pull the ends, then $F > 0$ and the string stretches out straight ($\kappa = 0$).
- When we float a closed loop of string on a surface, then add some oil or detergent to the water it encloses, then $\Delta T > 0$. The string jumps outward to form a circle (κ constant).

6.2.3 Surfaces in space

Our real goal is to understand mechanical equilibrium of a 2D *surface* in 3D space. So we must make some substitutions in the preceding discussion:

- Line tension F along a skimmer \rightsquigarrow surface tension T of a soap film, or the interfacial tension of a fluid–fluid interface. Again, in mechanical equilibrium it will be constant throughout the surface.
- Interfacial tension difference ΔT between two sides of skimmer \rightsquigarrow pressure difference Δp between sides of our surface. Again, in mechanical equilibrium it will be constant throughout the volume on each side of the surface.
- Curvature of a curve in a plane $\rightsquigarrow \dots$ what?

To make progress, we must generalize Idea 6.1a to get a simple local formula for the change in *area* of a curved *surface* to first order in a small perpendicular displacement ξ . We can proceed as before: At any chosen point \mathbf{P} , set up a tangent *plane*. Then measure the perpendicular displacement $\Delta(u, v)$ between that tangent and the actual surface, where u, v are two surface coordinates (e.g. latitude and longitude on a sphere). We'll require that u and v be centered on \mathbf{P} .

- As before, the Taylor series expansion of Δ will again have no linear terms: That's what tangency means.
- To leading order, then, Δ is a quadratic function of the two small excursions u, v . That function is zero if the surface is flat, so it describes curvature.
- More generally, the quadratic part may be expressed as

$$\Delta^{[2]}(u, v) = \frac{1}{2}(B_{11}u^2 + 2B_{12}uv + B_{22}v^2). \quad (6.3)$$

Unfortunately, the coefficients B_{11} , B_{12} , and B_{22} depend on our choice of coordinate system u, v for the surface. In one dimension, we removed this ambiguity by specifying arclength as the coordinate s . But what's the analog of that choice on a 2D surface?

Although there is no unique, standard coordinate system, we may at least restrict the choice by requiring that if we start at \mathbf{P} and move to a nearby point, then the arclength squared of the resulting curve within the surface must take the form

$$ds^2 = du^2 + dv^2 + \dots, \quad (6.4)$$

where the ellipsis is terms of higher than quadratic order.⁵ If our coordinates don't

⁵The presence of the higher-order terms may surprise you—isn't Equation 6.4, without any higher terms, just the pythagorean theorem? Indeed, on a flat plane we may choose cartesian coordinates, in which the usual formula is exactly true. Certain curved surfaces may also admit such special coordinates; however, in general they don't exist and Equation 6.4 is the best we can do.

have that property, we can always find new coordinates that do have it just by applying a linear transformation to u, v . We'll call such a choice **normal coordinates** in the surface at \mathbf{P} .

Your Turn 6A

- Look up the latitude $(\pi/2) - \theta_0$ and longitude $-\varphi_0$ of your hometown. You could choose $u = \theta - \theta_0$ and $v = \varphi - \varphi_0$, but do they satisfy Equation 6.4? If not, find a linear transformation that turns them into good coordinates.
- Even with the choice you made in (a), does Equation 6.4 hold exactly, that is, without higher-order terms?

Once we find local coordinates that meet our criterion, however, they will still not be unique: Other choices will *also* obey Equation 6.4. However, all such choices are of the form

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \mathbf{S} \begin{bmatrix} u \\ v \end{bmatrix} + \dots,$$

where \mathbf{S} is a 2D rotation matrix and the ellipsis again denotes possible higher-order terms. If we reexpress Equation 6.3 in terms of u', v' , then it will involve three new coefficients $B'_{11}, B'_{12}, B'_{22}$. That is, none of these quantities *invariantly* characterizes the surface near \mathbf{P} , due to the residual coordinate freedom.

Luckily, there is a way out. The quadratic function $\Delta^{[2]}(u, v)$ can be expressed in terms of a matrix $\mathbf{B} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$. Its new form involves a new matrix $\mathbf{B}' = (\mathbf{S}^{-1})^t \mathbf{B} \mathbf{S}^{-1}$. But 2D rotations have the special property that $\mathbf{S}^t = \mathbf{S}^{-1}$, so \mathbf{B}' is related to \mathbf{B} via a *similarity transformation*. And any matrix has two famous properties that are invariant under similarity transformation, and hence *don't care* which local coordinates we chose (as long as they obeys Equation 6.4).

In the present context, those invariants are called the **Gauss curvature**, $G = \det \mathbf{B}$ and the **mean curvature**, $H = \frac{1}{2} \text{Tr} \mathbf{B}$. Put differently, the two eigenvalues of \mathbf{B} are called **principal curvatures**. Both are invariant; we just repackage them into $G = k_1 k_2$ and $H = (k_1 + k_2)/2$. Now examine Figure 6.2. Panel (c) shows a case where both principal curvatures are zero. Panels (b,e) show cases where $k_1 = 0$ while k_2 is not zero but constant (lateral surfaces in panel (b)) or nonconstant (panel (e)). Panel (a) shows both k_1 and k_2 nonzero with the same sign; (d) shows opposite signs. Thus, the mean curvature is zero in (c), and potentially also (d) (if $k_1 = -k_2$ exactly). The Gauss curvature is zero in panels (b,c,e).

Your Turn 6B

Assume a spherical Earth. Continuing Your Turn 6A, find Earth's two principal curvatures at your hometown. (What about *my* hometown?)

A cone has a sharp apex, so it shouldn't surprise you that its mean curvature is infinite there, and hence nonconstant elsewhere. In fact, if we let r denote distance from the apex to \mathbf{P} , then axial symmetry implies that $H = H(r)$, and you'll show in Problem 6.2 that $H \propto r^{-1}$.

6.2.4 The Young–Laplace formula

As in Section 6.2.1, we now imagine distorting a surface to a nearby one by moving each point \mathbf{P} a distance ξ perpendicular to the surface at \mathbf{P} .⁶ We can now state the result we need, analogous to Idea 6.1:

- a. To first order in perpendicular displacement ξ , the total area change is the integral over the surface of its area element times $-2H\xi$ (a local formula).
 b. In contrast, the **volume** occupied by one side grows, and the other side shrinks, by the integral over the surface of its area element times ξ without any factor of curvature (another local formula).

(6.5)

We won't prove Idea 6.5,⁷ but look at the example surfaces in Figure 6.2a,b,e: Flatten- ing a patch of any of these surfaces will reduce the surface area. So Gauss curvature cannot be what controls this loss, because it's zero for panels (b) and (e). Instead, all three of these surfaces have nonzero *mean* curvature. In contrast, panel (c) has extremal area and also zero mean curvature. So it's reasonable to suppose that mean curvature controls the first-order change in area.

Now imitate the argument in Section 6.2.2, modified as at the start of Section 6.2.3. A soap bubble, or a fluid–fluid interface, costs some energy proportional to its surface area; the constant T is called surface or interfacial tension.⁸ In mechanical equilibrium it's constant, because molecules can rearrange parallel to the surface. A closed surface (closed soap bubble or liquid drop boundary) separates two sides that can have different hydrostatic pressures; this pressure difference Δp is also constant in equilibrium.⁹ The equilibrium surface shape must minimize total free energy. Arguing as before (Idea 6.2 but with Idea 6.5) now gives

Mechanical equilibrium selects a shape that has constant mean curvature. The value of mean curvature will be zero for an open soap film, or more generally $2H = \Delta p/T$ for a closed bubble or fluid–fluid interface.

Young–Laplace
formula

(6.6)

Pressure is measured in newtons per meter squared, whereas interfacial tension is in newtons per meter, so Equation 6.6 is at least dimensionally correct.

Although we didn't prove the mathematical result Idea 6.5, it has led to Idea 6.6, which does accord with experience. Look at the examples in Figure 6.2 and note how the Young–Laplace formula applies to each one: Each is a possible equilibrium surface, *except not (e)*.

⁶Also as in 2D, our convention is that the deviation is positive if the perpendicular from the tangent plane to the surface points parallel to the perpendicular chosen when defining ξ .

⁷[\[T2\]](#) See Section 6.2.4' (page 78).

⁸More precisely, the tension is the *free* energy cost per area.

⁹Pressure can be nonconstant only if a “body force” like gravity acts on the bulk of the fluid. In the experiments we are studying, the net effect of gravity involves the density difference of the two fluids and is negligibly small. Surface tension can also be nonconstant, for example in the presence of temperature or chemical gradients (Marangoni effect), but those are nonequilibrium situations.

6.3 EFFECT OF ELECTRIC FIELD

6.3.1 New contribution to energy balance

So, great: Now you know why you have never seen a conical soap bubble or fluid–fluid interface. The only problem is . . . you *have* seen one in Figure 6.3. Contrary to Idea 6.6, this shape has mean curvature that is nonconstant and indeed diverges at the cone’s apex. What physics have we forgotten to include?

The new physics is that the lower fluid in the photo was electrically conducting, and the system was subjected to a strong electrostatic field. To see how that matters, recall that there is no electric field inside a conducting body, and hence no electric field energy there; any dielectric properties of the fluid are immaterial. But there *is* field energy in empty space or an insulator, and unlike hydrostatic pressure, its density *need not be uniform*. Indeed, Chapter 5 argued provisionally that that density equals¹⁰ $\epsilon \|\vec{E}\|^2/2$. If we deform the interface, then this energy cost grows or shrinks proportional to the change of volume on the side with nonzero field.

In the region near the point, Our Founder Ben Franklin told us to expect a nonconstant electric field. Moreover, the field becomes huge near the point, so we can neglect any hydrostatic pressure difference (set $\Delta p = 0$) and attempt to balance electric field energy against interfacial tension.

6.3.2 Electrostatics near a conical point

Before we ask about mechanical equilibrium, let’s first ask what sort of static electric field could exist outside a cone-shaped conductor. It will be convenient to use spherical polar coordinates, because the Laplace equation is separable in such coordinates, axial symmetry is easy to implement, and our boundary condition is simply that the cone with one particular value θ_0 of polar angle must be an equipotential:

$$\psi(r, \theta_0, \varphi) = 0 \quad \text{for all } r \text{ and } \varphi. \quad (6.7)$$

Following Chapter 4, let us therefore look for potentials of the form

$$\psi(r, \theta, \varphi) = CN(r)M(\cos \theta) \quad \text{where} \quad M(\cos \theta_0) = 0. \quad (6.8)$$

Here C is an unknown overall constant. Our conducting fluid cone is the region $\theta \geq \theta_0$, so its half opening angle is $\pi - \theta_0$.

If a solution of the form Equation 6.8 exists, the function N must obey¹¹ $2rN' + r^2N'' = \lambda N$ for some constant λ . Moreover, we know how N must diverge at $r \rightarrow 0$. The electric field energy density involves $\|\vec{\nabla}\psi\|^2$, and our generalized form of the Young–Laplace formula says it must balance the mean curvature, which diverges as r^{-1} . So the potential must itself diverge as $r^{1/2}$. Substituting that trial solution into the radial equation yields that the behavior is in fact exactly $r^{1/2}$ and the eigenvalue λ equals $3/4$.

Meanwhile, the angular function obeys Equation 4.6 (page 50):

$$((1 - \mu^2)M')' = -\lambda M,$$

¹⁰See Equation 5.10 (page 59). Detailed derivations must wait for Chapters 34 and 52.

¹¹See Section 4.3 (page 50).

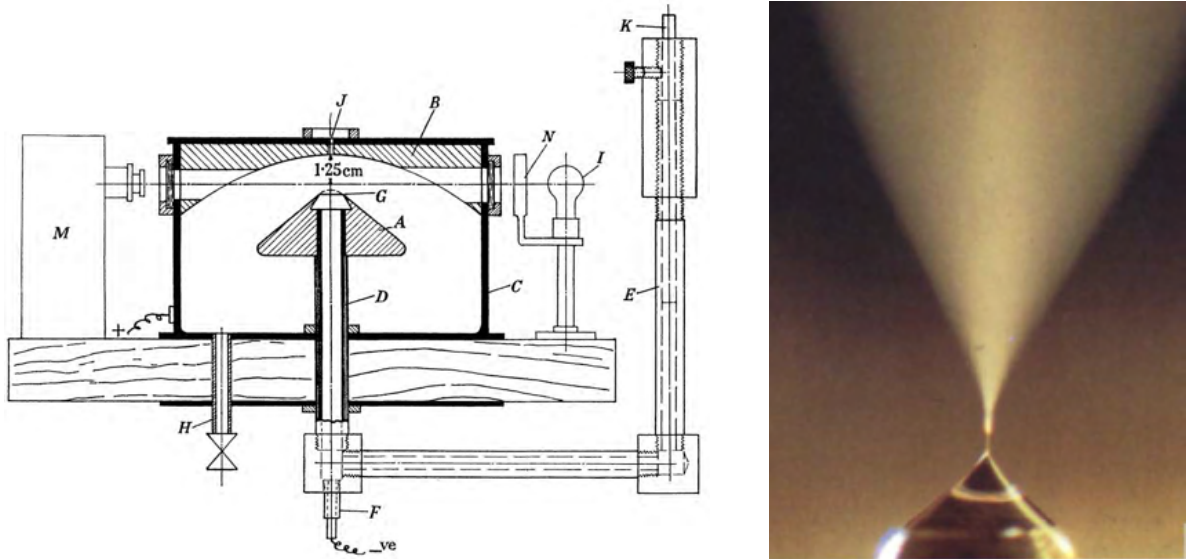


Figure 6.5: *Left:* Cross-section of Taylor's apparatus, showing the curved electrodes (*hatched*). "The cone **A** was truncated so that its upper edge was a horizontal circle 1 cm diameter which could form the lower edge of a conical liquid surface if such a surface could in fact be formed." [From Taylor, 1964, Fig. 5.] *Right:* Taylor cone (*bottom*) giving rise to a jet of fluid (methanol with a small amount of hydrochloric acid). [From Pantano et al., 1994.]

where now prime indicates $d/d\mu$ and $\mu = \cos\theta$. We know the solutions to this equation are Legendre polynomials, at least for integer λ . Indeed, the standard form of the Legendre equation is

$$(1 - \mu^2) M'' - 2\mu M' + [\ell(\ell + 1)] M = 0.$$

Comparing shows that we need the case $\ell = 1/2$. Thus, the function M that we need is a Legendre function of order $1/2$. It's not a finite polynomial like the ones we're used to, but it's a perfectly well defined function. You'll evaluate it in Problem 6.3, but first we can say some simpler things.

We have found a function, Equation 6.8, that satisfies the Laplace equation, is axisymmetric, and has the right kind of singularity at $r \rightarrow 0$. But we haven't yet enforced the boundary condition Equation 6.7, which also requires $M(\theta_0) = 0$. So remarkably, *there is only one possible angle* for an equilibrium cone singularity, regardless of the value of the interfacial tension. When you evaluate it in Problem 6.3, you'll see that experimentally, the angle in Figure 6.3 really is as predicted.

Figure 6.5a shows two electrodes shaped approximately as equipotentials of the solution to our equation, apart from a missing conical bit at the point labeled **G**. At the appropriate value of potential difference, a puddle of conducting fluid at **G** was observed to rise up and form the sharp point shown in Figure 6.3.

6.4 TECHNOLOGICAL APPLICATIONS

In 2002, John B. Fenn shared a Nobel Prize, not for discovering the cone state, but for applying it. Fenn knew that at high applied potential, a molecular-scale jet of fluid can emerge from the apex of the cone (Figure 6.5b). This proved to be a convenient way to gently isolate and ionize dissolved macromolecules without breaking them; it led to a big advance in mass spectrometry.

More broadly, Derek Stein wrote me, “There’s a lot of beautiful physics involved in electrospray. For instance, how it provokes a cascade of ‘Coulomb explosions,’ with charged droplets shrinking due to evaporation and then spontaneously developing one or more Taylor cones from which even smaller charged droplets emerge. Another example is the formation of a fluid jet at the end of the Taylor cone from which those charged droplets detach; the detachment process very naturally creates singularly sharp points. We’re using Taylor cones to deliver ionized biomolecules from liquid into the vacuum environment of a mass spectrometer. We’re looking into an interesting regime of Taylor cones developing at the end of a nanoscale capillary: we find that single ions jump directly from the liquid into the vacuum. This is quite different from a conventional electrospray, where a cone-jet mode delivers charged droplets into a mass spectrometer.”

6.5 PLUS ULTRA

6.5.1 A look ahead

Once again, a tensor quantity has popped out in the course of other business. Previously this happened when we invented the quadrupole tensor; this time, the quadratic function $\Delta^{[2]}(u, v)$ involved the coordinate-dependent, symmetric matrix B . Later chapters will extend and systematize the notions introduced informally in this chapter.

6.5.2 Other physical surfaces

We have barely scratched the surface of surfaces. Soap films and simple interfaces are characterized by a single parameter, the interfacial tension T . A cross-linked surface, such as the bacterial cell wall, will also resist shear deformation as well as local changes in area. Other membranes, such as lipid bilayers, have no such shear moduli, but may nevertheless have a preferred value of mean curvature, for example zero. They resist deviations with an energy cost per area of the form $(H - H_0)^2$, that is, different from the one giving rise to the Young–Laplace formula.

6.5.3 A glimpse of general relativity

Section 6.2.3 took some trouble to characterize a surface using invariantly defined local quantities (the scalar fields G and H). Only one of these was needed for our application.

But the concept of curvature enters physics in many other ways, and we’d be remiss not to mention a remarkable property of the other curvature G . For a 2D surface

embedded in ordinary 3D space, we defined curvature via a procedure involving points *outside* the surface (that is, via the deviation Δ between the surface and its tangent plane). However, the Gauss curvature can be reexpressed solely in terms of distance properties *within* the surface.¹² We need not even imagine *any* surrounding 3D space. This realization set in motion B. Riemann’s study of intrinsic curvature for spaces of dimension greater than two. Much later, that framework was just what Einstein needed to understand gravitation.

Riemann found that in higher dimensions there is an entire *tensor* of intrinsic curvatures generalizing Gauss’s simple scalar G . Einstein found that Riemann’s curvature tensor roughly plays the same role as $\nabla^2\phi_N$ in a field equation, and that it also controls the separation of two nearby freely falling bodies.

FURTHER READING

Mathematics of curvature: Dubrovin et al., 1992.

Young–Laplace formula: Butt & Kappl, 2018; Safran, 2003; Nelson, 2014, §7.2.2.

Don’t miss the hilarious yet profound video: Lloyd Trefethen, *Surface tension in fluid mechanics* (National Committee for Fluid Mechanics films, 1963)

<http://web.mit.edu/hml/ncfmf.html>.

It is also available at: <https://www.youtube.com/watch?v=DkEhPltiqmo>

https://www.youtube.com/watch?v=yiixl1f_HKw

<https://www.youtube.com/watch?v=5d6efCwkWs>.

The phenomenon discussed in this chapter is often called the “Taylor cone” after Taylor, 1964, p. 392.

John Fenn’s Nobel Lecture:

<https://www.nobelprize.org/prizes/chemistry/2002/fenn/lecture/>. From Gauss to Riemann: Spivak, 1979.

¹²Gauss called this fact his “Theorema egregium” (outstanding theorem).

T₂ Track 2

6.2.3'

We can rephrase the construction of Section 6.2.3 in a more elegant way, using ideas to be developed later in these notes.

The quadratic part of the deviation $\Delta^{[2]}$ defines a rank- $\binom{0}{2}$ tensor related to the “second fundamental form” of a 2-surface embedded in 3-space. The quadratic part of the distance-squared function ds^2 defines another rank- $\binom{0}{2}$ tensor called the metric (also called “first fundamental form”) of the surface. We can use the metric to convert the second fundamental form to a rank- $\binom{1}{1}$ tensor (by “raising an index”). The new tensor transforms by similarity, so its trace and determinant are scalars.

Something similar happened in our discussion in Section 3.1' (page 39). There we were working in *flat* 3D space, so we could just choose globally cartesian coordinates when defining the quadrupole tensor. We again face the issue that there is some freedom to choose different cartesian systems, but again this amounts to a similarity transformation acting on the components of \vec{Q}_E . So again its three eigenvalues are invariants, and hence they invariantly characterize different kinds of quadrupole (uniaxial versus biaxial).

T₂ Track 2

6.2.4'

Here we establish the formulas in Idea 6.5.

A 2-surface in 3-space is specified by a vector function $\vec{r}(u, v)$. The two parameters range over some fixed region of the uv plane. Let $\hat{n}(u, v)$ be a choice of normal vector to each point of the surface. The area of the surface can then be written as

$$\Sigma = \int dudv \left\| \frac{\partial \vec{r}}{\partial u} \times \frac{\partial \vec{r}}{\partial v} \right\|.$$

Abbreviate $\partial/\partial u$ by ∂_u and so on, and let J denote the square of the integrand above. Thus

$$J = \|\partial_u \vec{r}\|^2 \|\partial_v \vec{r}\|^2 - (\partial_u \vec{r} \cdot \partial_v \vec{r})^2.$$

A new surface is specified by an ordinary function ξ via $\tilde{\vec{r}}(u, v) = \vec{r}(u, v) + \hat{n}(u, v)\xi(u, v)$. Suppose that the normal displacement function ξ equals zero at the boundary of the chosen region in u, v . The first-order variation of the surface area is then

$$\delta A = \int dudv J^{-1/2} \left[\partial_u \vec{r} \cdot \partial_u (\xi \hat{n}) \|\partial_v \vec{r}\|^2 + \partial_v \vec{r} \cdot \partial_v (\xi \hat{n}) \|\partial_u \vec{r}\|^2 - (\partial_u \vec{r} \cdot \partial_v \vec{r}) (\partial_u \vec{r} \cdot \partial_v (\xi \hat{n}) + \partial_v \vec{r} \cdot \partial_u (\xi \hat{n})) \right]. \quad (6.9)$$

Now integrate by parts, using that $\xi = 0$ on the boundary, and write the result as $\delta A = \int dudv \xi q$. We wish to find a convenient expression for the function q at any point \mathbf{P} in terms of the surface shape near that point.

Our formula is invariant under translations and rotations of \vec{r} , so we may suppose that our 3D coordinates are centered on \mathbf{P} and that moreover the tangent plane to the surface is the xy plane. We can also shift the two parameters u, v to center them on \mathbf{P} and scale/rotate them to arrange that

$$\vec{r}(u, v) = u\hat{x} + v\hat{y} + \frac{1}{2}[u, v] \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \hat{z} + \mathcal{O}(3).$$

The last term represents terms of order three or greater in u , v . The constants B_{11} , $B_{12} = B_{21}$, B_{22} have the same meaning as in Equation 6.3.

Note that

$$\partial_u \vec{r} = \hat{x} + (B_{11}u + B_{12}v)\hat{z}, \quad \partial_v \vec{r} = \hat{y} + (B_{12}u + B_{22}v)\hat{z},$$

$$\|\partial_u \vec{r}\| = 1 + \mathcal{O}(2), \quad \|\partial_v \vec{r}\| = 1 + \mathcal{O}(2), \quad \text{and } J = 1 + \mathcal{O}(2).$$

The unit normal to the surface is

$$\hat{n}(u, v) = \frac{\partial_u \vec{r} \times \partial_v \vec{r}}{\|\partial_u \vec{r}\| \|\partial_v \vec{r}\|} = \hat{z} - \hat{x}(B_{11}u + B_{12}v) - \hat{y}(B_{12}u + B_{22}v) + \mathcal{O}(2).$$

We can now evaluate the integrand of Equation 6.9 at the chosen point:

$$\delta A = \int dudv \xi \hat{n} \cdot \left[-\partial_u (J^{-1/2} \partial_u \hat{r} \|\partial_v \hat{r}\|^2 J^{-1/2}) - \partial_v (J^{-1/2} \partial_v \hat{r} \|\partial_u \hat{r}\|^2 J^{-1/2}) + (\partial_u \partial_v \vec{r} + \partial_v \partial_u \vec{r}) (\partial_u \vec{r} \cdot \partial_v \vec{r}) \right].$$

That expression simplifies when we evaluate at \mathbf{P} : The integrand at that point is

$$q(\mathbf{P}) = \hat{n} \cdot (-\hat{z}(B_{11} + B_{22})).$$

Finally, we note that the local coordinates we have chosen have the property expressed in Equation 6.4 (page 71), and so the quantity $B_{11} + B_{22}$ equals twice the mean curvature at \mathbf{P} . Thus Equation 6.9 is equivalent to the first statement in Idea 6.5.

The second statement concerns the volume of a thin shell of perpendicular thickness $\xi(u, v)$. Multiply the area element by the thickness to get the volume.

PROBLEMS

6.1 *Variation of arclength and area*

a. A curve in a plane is specified by a vector function $\vec{r}(s)$, where s is arclength, $0 \leq s \leq L$. Let $\hat{n}(s)$ be a field of perpendicular vector all along the curve. A new curve is specified by an ordinary function ξ via $\vec{r}(s) = \vec{r}(s) + \hat{n}(s)\xi(s)$. The displacement function ξ equals zero at $s = 0$ and L .

The parameter s still runs from 0 to L , but it's no longer arclength for the new curve, so the new total length will no longer be L . Establish the formula in Idea 6.1a. [Hint: You'll need to use integration by parts.]

b. Establish the formula in Idea 6.1a.

6.2 *Mean curvature of a cone*

Show that the mean curvature of a cone with opening half-angle α is $H(r, \varphi) = (\cot \alpha)/(2r)$. Here r is distance from the cone's apex to the point of interest and φ is the angular position on each "latitude" line. [Hint: If you have difficulty, first draw a very wide cone, with α just slightly less than $\pi/2$. It's nearly a plane, so its curvature must be smaller for given r than that of a narrower cone. Make sure your derivation accounts for this.]

6.3 *A pointed remark*

Finish the derivation of the stability problem started in the main text. Set up spherical polar coordinates and consider a cone of electrically conductive fluid occupying the region of space with $\theta \geq \theta_0$, as in Figure 6.3 (page 69). Thus the half-opening angle of the cone is $\pi - \theta_0$. Take the electrostatic potential to have the form Equation 6.8, where $N(r) = r^{1/2}$, M is the Legendre function of order $1/2$, and C is an undetermined overall constant. You may assume the pressure drop is $\Delta p = 0$, but the interfacial tension is fixed to some given value T .

- a. Get a computer to find the only zero of the function M in the range $-1 < \cos \theta_0 < 1$, and in that way predict θ_0 and hence $\pi - \theta_0$.
- b. Evaluate the electrostatic potential throughout the plane $y = 0$ (or just the half-plane with $\varphi = 0$), display it as a contour plot, and comment.
- c. Evaluate the derivative $dM(\cos \theta)/d\theta$ at θ_0 numerically.
- d. Using your results in (a–b), write a formula for the electric field squared just outside the surface ($\theta \lesssim \theta_0$).
- e. Generalize the Young–Laplace formula Idea 6.6 appropriately by finding an expression for the electrostatic field energy density just outside the surface and setting it equal to $2TH$, where H is the mean curvature from Problem 6.2 and T is the interfacial tension of oil and water. Substitute the result you found in (c).
- f. Obtain a prediction for the constant C in terms of T and the relative permittivity ϵ/ϵ_0 of oil. Once we look up those values, for example for an oil–water interface, then we learn how many volts we'll need in an apparatus before we can expect to see a conical singularity.

CHAPTER 7

Charge Flux, Continuity Equation, Ohmic Conductors

We gradually start to look at non-static situations.

7.1 ONE DIMENSION

Imaging a long thin pipe with some conserved “stuff” inside. Maybe it’s air, and the “stuff” is mass.¹ Define a 1D density $\rho^{(1D)}(t, z)$ (stuff per meter). At any z_0 , also define the **1D flux** $j^{(1D)}$ as the net rate at which stuff crosses the point $z = z_0$, moving from smaller to larger z . For example, a particle crossing in the opposite direction makes a *negative* contribution to the 1D flux of mass.

Then conservation of stuff implies

$$-j^{(1D)}(z + dz) + j^{(1D)}(z) = \text{rate of pileup at } z = \frac{d}{dt}(\rho^{(1D)} dz), \text{ or}$$
$$\frac{\partial j^{(1D)}}{\partial z} + \frac{\partial \rho^{(1D)}}{\partial t} = 0. \quad (7.1)$$

Here is a pictorial way to understand Equation 7.1: Imagine a small range of space and time near (t, z) (dashed box in Figure 7.1a). Multiply Equation 7.1 by $\Delta t \Delta z$ to get

$$0 = \Delta t \left(\Delta z \frac{\partial j^{(1D)}}{\partial z} \right) + \Delta z \left(\Delta t \frac{\partial \rho^{(1D)}}{\partial t} \right). \quad (7.2)$$

Figure 7.1a shows three particle trajectories. Points 1 and 3' contribute to the first term of Equation 7.2, whereas 2, 3, 1', and 2' contribute to the second term. Because every trajectory that enters the dashed box must also leave it, these terms must sum to zero.

7.2 TWO OR MORE DIMENSIONS

From now on, we will be more interested in electric charge than in mass, so unless otherwise stated the symbol j will refer to **charge flux**. Figure 7.1 shows a world without z (two spatial dimensions). We now also generalize to allow particles to exchange charge, merge, or even explode as shown in the figure. In between such interactions, each particle’s trajectory is a curve in spacetime carrying a fixed number (its “charge”). Even in an interaction, this number is conserved locally (at each vertex

¹Yes, I know that mass isn’t really conserved! It’s just an illustration familiar from classical physics. Really we are interested in charge, which really *is* conserved, even in relativity.

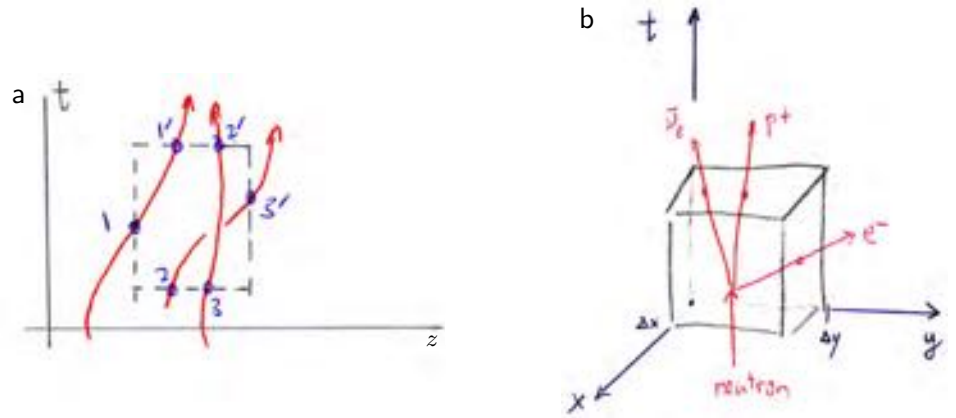


Figure 7.1: (a) Three trajectories in one spatial dimension. (b) A charge-conserving interaction in two spatial dimensions.

separately). For example, in the weak decay² shown, the incoming line has charge 0 while the outgoing lines have charges 0, e , and $-e$.

The net charge entering any closed surface, like the surface of the box shown, is therefore zero. In the neutron decay shown, we have

- 0 (neutron trajectory enters via bottom face of the box);
- $-e$ (proton and neutrino, total charge $+e$, exit via top face of the box);
- $-(-e)$ (electron exits via right face of the box).

Those quantities do sum to zero. For trajectories that don't branch inside the box, it's even simpler: Everything that enters the box must also exit, carrying its charge.³

Often it's reasonable to think of charge as a “river” of many particles, defining an essentially continuous flow. Charge density $\rho_q^{(2D)}(t, \vec{r})$ in 2D has units coul m^{-2} . **Charge flux** $\vec{j}^{(2D)}(t, \vec{r})$ in 2D has units $\text{coul m}^{-1} \text{s}^{-1}$. The quantity $\vec{j}_1^{(2D)}(t, x_0, y_0)$ is defined as the net charge per length per time crossing a short line segment of constant $x = x_0$ near position (x_0, y_0) at time t . Here again “net” means that a charge q passing from smaller to larger values of x contributes q , while charge passing the opposite way contributes $-q$.

What's new compared to one dimension is that now we get another component of flux, $\vec{j}_2^{(2D)}$, when we consider charge crossing a short line segment with constant y .

The total charge entering the infinitesimal spacetime box shown is

- $+\rho_q^{(2D)}(0, 0, 0)\Delta x\Delta y$ from the $t = 0$ (lower) face (plus terms of higher order in Δx and Δy);
- $-\rho_q^{(2D)}(\Delta t, 0, 0)\Delta x\Delta y$ from the $t = \Delta t$ (upper) face;
- $+\vec{j}_2^{(2D)}(0, 0, 0)\Delta x\Delta t$ from the $y = 0$ (left) face;
- $-\vec{j}_2^{(2D)}(0, 0, \Delta y)\Delta x\Delta t$ from the $y = \Delta y$ (right) face;

²By the way, the reaction shown is also an example of two other local conservation laws, those of lepton and nucleon numbers. Each has its own continuity equation analogous to the one discussed in these notes.

³And trajectories that *never* enter the box also never exit it.

- $+j_1^{(2D)}(0, 0, 0)\Delta y\Delta t$ from the $x = 0$ (rear) face;
- $-j_1^{(2D)}(0, \Delta x, 0)\Delta y\Delta t$ from the $x = \Delta x$ (front) face.

These contributions must always sum to zero. Grouping them in pairs and using a Taylor expansion gives

$$0 = \left(-\frac{\partial}{\partial t}\rho_q^{(2D)} - \frac{\partial}{\partial y}j_2^{(2D)} - \frac{\partial}{\partial x}j_1^{(2D)}\right)\Delta x\Delta y\Delta t. \quad (7.3)$$

(Higher order terms vanish when we take the limit of a small box.) The infinitesimal box may be located anywhere, so

$$0 = \frac{\partial}{\partial t}\rho_q^{(2D)} + \vec{\nabla} \cdot \vec{j}^{(2D)}. \quad \text{continuity equation} \quad (7.4)$$

We can do the whole derivation again, with any number of spatial dimensions (for example three). The charge density and flux have different units, but they still obey the same continuity equation.

7.3 REMARKS

1. Charge flux is sometimes called “current density,” though I prefer to reserve the word “density” to mean “per unit volume.” I use **flux** to mean “per transverse dimensions per time.” In 1D there are no transverse dimensions and $j^{(1D)}$ was just stuff per time. In 2D there is one dimension transverse to a given direction. In 3D there are two.
2. The result we have found is purely a kinematic identity, not an equation of motion. It is valid regardless of whether the particle trajectories obey any equation of motion. It merely expresses local conservation of charge (or any other scalar quantity); beyond that physical assumption, it’s just bookkeeping.
3. In a stationary situation, where charge density is unchanging (perhaps zero), the continuity relation guarantees that \vec{j} is divergence-free.

7.4 NONSTATIC SITUATIONS

Many materials are insulators: $\vec{j} = 0$. Some others are approximately **ohmic**: they develop currents via a dissipative law⁴

$$\vec{j} = \kappa\vec{E}. \quad \text{ohmic material} \quad (7.5)$$

The constant κ is a material parameter called **conductivity** of the material. I call this a “hypothesis” because we must *check* whether a medium is effectively ohmic for a given application. Copper, at ordinary frequencies, is approximately ohmic, as is salt water. I call Equation 7.5 “dissipative” because it relates \vec{j} , a quantity that changes

⁴Discovered by H. Cavendish, half a century before G. Ohm. Cavendish failed to publish this important result, and many others as well. Many books call it “Ohm’s law,” but so many exotic materials are not ohmic that it’s a bit silly to call it that. Instead we say that some materials, in some conditions, have ohmic behavior.

sign under time reversal, to \vec{E} , a quantity that doesn't. Thus, this formula breaks time-reversal invariance: It describes the irreversible conversion of electric energy into heat.

Equation 7.5 may not look like Ohm's "law" as it appeared in first-year physics: It's a local rule involving a material parameter. To make the connection, consider a thin wire of length h with cross-sectional area Σ . Total current I flows, leading to a charge flux $j = I/\Sigma$. Thus, $I = \Sigma\kappa E$. But the electric field within the wire leads to a potential drop as usual, $\Delta\psi = hE$. Thus

$$\Delta\psi = IR \quad \text{where} \quad R = h/(\Sigma\kappa).$$

The **resistance** R depends on the geometry (h and Σ) of the wire as well as on the material (κ). The SI unit for resistance is called ohm and abbreviated Ω . The SI unit for conductivity is then $\Omega^{-1}\text{m}^{-1}$. Another name for Ω^{-1} is the siemens, abbreviated S.⁵

Other quantities you may hear include **conductance**, defined as $1/R$, and **resistivity**, defined as $1/\kappa$.

Let's quantify the "frictional" (dissipative) character of the ohmic hypothesis. Some external agency must expend energy $(dq)\Delta\psi$ to push a lump of charge through our wire. Multiplying $\Delta\psi$ by the total rate of charge transport thus gives the power absorbed by the wire as $\mathcal{P} = (\Delta\psi)I = I^2R = (\Delta\psi)^2/R$. Again, that power must end up as heat, an effect called **Joule** or **ohmic heating**. Indeed, if you plug an appliance with an internal short circuit ($R \lesssim 1\Omega$) into the wall ($\Delta\psi$ fixed), you get a lot more heat than when you plug in a normal light bulb ($R \gg 1\Omega$).

7.5 QUASISTATIC SITUATIONS

We will be interested in situations where everything is changing slowly in time, for example, the millisecond time scales characteristic of nerve impulses. There is a useful simplification we can use in this case.

In static (zero-frequency) situations, Section 2.6 argued that charge will rearrange to erase any electric field inside a conductor. Even at nonzero frequency, we get the same conclusion for a perfect conductor. What about a non-static situation with a non-perfect conductor? Charge takes time to move around, because moving too fast incurs too much frictional resistance. Let's combine the continuity equation, the ohmic hypothesis, and the Gauss law to find

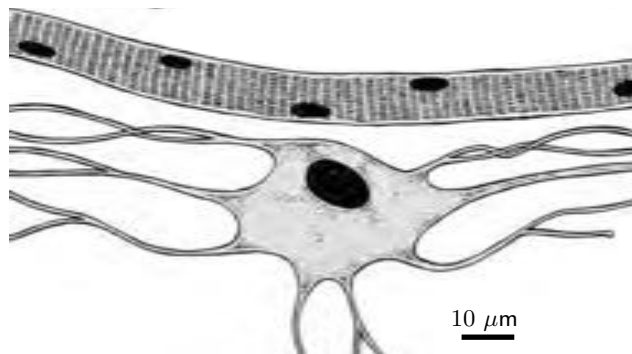
$$\frac{d}{dt}\rho_{\text{q}} = -\vec{\nabla} \cdot (\kappa\vec{E}) = -\kappa\rho_{\text{q}}/\epsilon.$$

We see that

$$\text{Any initial nonuniformity of net charge density gets exponentially suppressed over time scales longer than } \epsilon/\kappa. \quad (7.6)$$

⁵Don't confuse it with the sievert (Sv), a unit of ionizing radiation dose, nor with the svedberg (also abbreviated S), used to describe sedimentation rate.

Figure 7.2: Two typical animal cells drawn to scale. *Upper:* Human striated muscle cell (myocyte). *Dark blobs* are cell nuclei. *Lower:* Human neuron. The unbranched tube on *right* is the “output line” (**axon**), which may extend for up to a meter to communicate with another neuron, a muscle cell, or an endocrine cell. Other tubes represent “input lines” (**dendrites**), each of which communicate with other neurons (including sensory receptors).



Your Turn 7A

Check that the units work out.

Again we see that an ohmic material breaks time-reversal invariance: Fluctuations in charge density always shrink, never grow, in time.

For salt solution at concentration 100 mM, we can look up $\kappa \approx 0.1 \Omega^{-1} \text{m}^{-1}$. We also know that pure water is highly polarizable; indeed, $\epsilon \approx 80\epsilon_0$ at low frequency.⁶ So for frequencies below about 100 MHz, we can assume that salt water is everywhere locally neutral, and hence also that $\vec{\nabla} \cdot \vec{E} = 0$, just as in electrostatics! This simplification will help us in Section 7.6 and in later chapters. Just as in the case of a perfect conductor, however, surface charges can and do build up at boundaries.

7.6 ELECTROENCEPHALOGRAPH/ ELECTROCARDIOGRAM

7.6.1 EEG

Right now, in your brain vast numbers of nerve cells (**neurons**, Figure 7.2) are communicating with one another and with your muscles, sensory receptors, and even endocrine-secreting cells throughout your body. The mechanism by which these signals travel long distances, without diminution, is the subject of Chapters 10–11. Right now we will only study the effect of such signaling on nearby tissue.

All of your body’s cells are filled with a solution of salt and various other molecules. All are in turn bathed in another such solution. The interior and exterior fluids have equal hydrostatic pressure, which is why delicate structures like cells and their axons, bounded by fragile membranes, can exist. The two fluids also have well matched overall osmotic pressure, so that the cell’s interior volume can remain constant. But the concentrations of *particular* ions can be quite different inside and outside of the cell. Figure 7.3a shows some of these concentrations for a well-studied axon in the squid *Loligo*. The exterior sodium ions feel a big electrochemical gradient toward the interior, but are frustrated by the barrier membrane. The interior potassium ions get conflicting forces: The negative interior potential tends to keep them in, but is overbalanced by the high interior concentration. These nonequilibrium concentrations

⁶See Section 5.8 (page 61).

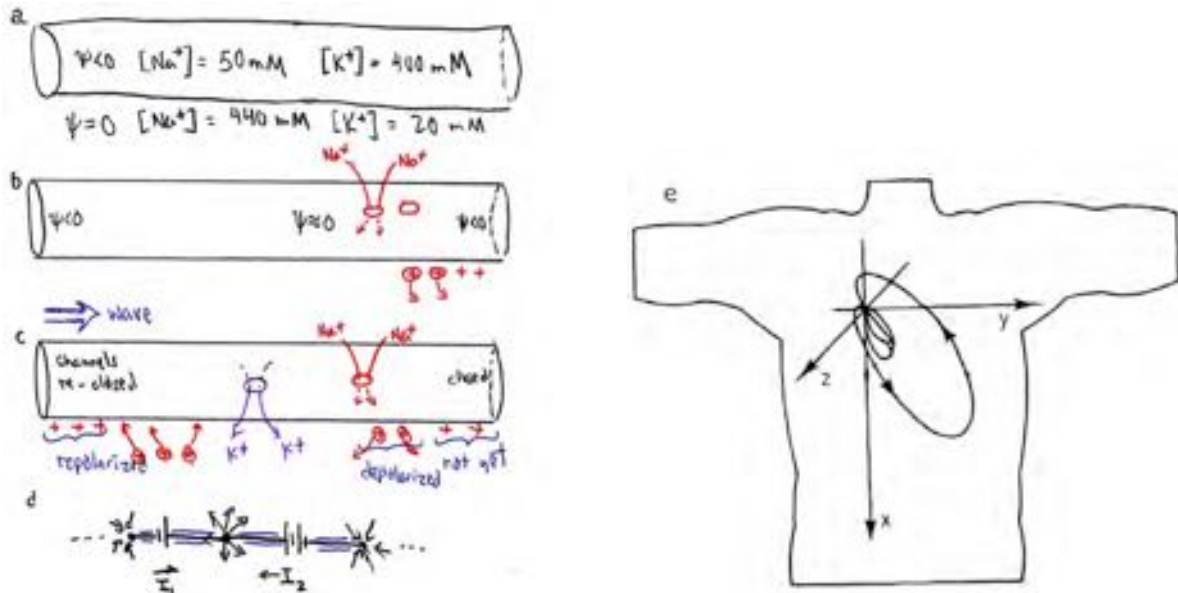


Figure 7.3: (a) A resting nerve axon. The cell membrane separates inner and outer fluids with strikingly different ion concentrations, of which two key players are shown. (b) A solitary, traveling wave of temporary depolarization progresses along the axon. At the wavefront, ion channels selective for one species (here sodium) open. This small ring-shaped region becomes a sink for sodium ions from the outer fluid. Other positive ions nearby that were previously attracted to the low-potential interior are now free to leave, effectively forming a source for positive ions in the region just in front of the wavefront, which has not yet released them. (c) Behind the leading edge, a second traveling wave follows in which different ion-selective channels open transiently. These allow the escape of potassium ions. (d) Idealization of the whole system as a set of short, insulated wires, parallel to the actual axon, that pull in charge at some points and emit it at others, forming a multipolar pattern of current in the outer fluid. (e) The total current dipole vector moves periodically, rotating and stretching with each heartbeat. The z axis comes out of the page.

form a continuously distributed source of free energy, constantly maintained by active transport of sodium out of, and potassium into, the cell.

The axon membrane is studded with doorways, **ion channels** that while normally closed, can open upon command, permitting the transport of specific ion types across the membrane. A nerve impulse involves the opening of ion channels specific for sodium in a small patch of membrane. As sodium ions rush into the long, narrow interior of the axon, others come from all directions in the outer fluid to replace them and maintain the overall neutrality of the solution. As a result of the ion motion, a region of the axon becomes **depolarized**: Its electrostatic potential rises from its resting negative value toward zero. The sodium channels soon close again, but by that time a nearby patch has opened, leading to a traveling wave of depolarization.⁷

The depolarization spreads to nearby membrane regions, releasing ions that had been clinging near the surface, attracted to the negative interior yet unable to enter (Figure 7.3b). Thus, a *source* of positive ions appears in the exterior solution, *displaced* from the sink. In addition, a later event in the nerve impulse involves the opening of

⁷More details appear in Chapter 11. Here we only regard the axon as a particular localized current source for the exterior world.

potassium-specific channels behind the sodium-channel opening (Figure 7.3c), creating a second displaced current source. Later still, all channels close and the whole system returns to its resting state after the impulse has passed.

In short, at any time t the exterior fluid sees a distribution $j^{(1D)}(t, x)$ of current sources and sinks localized at points x along the axon. This current spreads into the surrounding fluid following the quasistatic rule, Idea 7.6. That is, at any instant it obeys $\vec{\nabla} \cdot \vec{j} = 0$ with boundary conditions at the axon determined by the form of the nerve impulse. But this equation implies $\vec{\nabla} \cdot \vec{E} = 0$, which is just the Laplace equation. We therefore know that far from the axon, the currents will have a multipole expansion of the usual form.

Figure 7.3c looks complicated, but we can get its main qualitative feature by remembering charge neutrality. The axon's cross-sectional area Σ and its conductivity κ determine the internal axial current I_{ax} created by the varying potential ψ_{in} via

$$I_{\text{ax}} = -(\kappa\Sigma) \frac{d}{dx} \psi_{\text{in}}.$$

The 1D continuity relation would require that nonuniformity of this current leads to charge buildup, were it not for the possibility that charge passes through the membrane. Charge can either literally pass through, via ion channels, or effectively pass through by discharging the membrane capacitance (both mechanisms are shown in the figure). Either way, the axon maintains local neutrality by releasing charge to the exterior with total 1D radial charge flux $j_{\text{tot}}^{(1D)} = (\kappa\Sigma) \frac{d^2}{dx^2} \psi_{\text{in}}$.

The expression just found for $j_{\text{tot}}^{(1D)}$ is a total derivative, and the potential approaches a constant at $x = \pm\infty$, so the monopole moment of the current source equals zero. Moreover, the quantity

$$x j_{\text{r,tot}} = (\kappa\Sigma) \frac{d}{dx} \left(x \frac{d\psi_{\text{in}}}{dx} - \psi_{\text{in}} \right)$$

is *also* a total derivative. Because the potential approaches the *same* constant value at $x = \pm\infty$, we see that the dipole moment of the current source *also* equals zero.

Hence, the leading-order current distribution far from the axon is in general of *quadrupole* form (Figure 7.3d). Any one impulse will create extremely small distant currents and fields. However, the concerted firing of impulses on many parallel axons in the brain can create a macroscopically measurable effect. Electric fields set up by the internal current can in turn penetrate even an intervening electrical insulator (such as the skull and surrounding skin). In this way, at least major brain activities can be measured noninvasively simply by attaching external electrodes to the skin and measuring the electric potential, a procedure called **electroencephalography** (EEG).

7.6.2 EKG

Muscle cells also support traveling waves of depolarization much like those in nerve cells, with the important difference that a single wave spreads over the entire long cell for the duration of a contraction. Thus in this situation the dipole moment of the current source need not equal zero.

For mechanical reasons, muscle tissue consists of huge numbers of parallel fibers that all contract in unison, leading to a big net dipole moment of the current distribution. Again, exterior electrodes on the skin can easily pick up this signal, determining

not only the magnitude of the dipole (traditional **electrocardiogram**, or EKG) but also its spatial direction (vector electrocardiogram, Figure 7.3e). The time course of this net dipole vector is a valuable diagnostic of heart disease.

FURTHER READING

General: Pollack & Stump, 2002, chapt. 7.

Quasistatic approximation: Pollack & Stump, 2002, §7.6.

Neuroelectricity, EEG, EKG: Hobbie & Roth, 2015; Benedek & Villars, 2000.

Advanced: Malmivuo & Plonsey, 1995; Gratiy et al., 2017.

PROBLEMS

7.1 Reactance

A real capacitor's dielectric may not be a perfect insulator: Some current may “leak” across when a potential difference is applied. Here's a way to measure both the capacitance C and resistance R at once, by applying a time-varying current $I(t)$ and observing the resulting transmembrane potential $\psi(t)$.

- a. Write an expression for the total current into a membrane in terms of $\psi(t)$. The total current consists of the leakage plus the time change of the charge stored in the membrane's capacitance.
- b. Suppose we impose a known current $I = \bar{I} \cos(\omega t)$. Find the resulting $\psi(t)$, and show that it has both $\cos(\omega t)$ and $\sin(\omega t)$ terms; that is, it's not *in phase* with the current. Show how to deduce R and C from this measurement.

7.2 Bulk conductor, I

Consider two electrodes immersed in an infinite bath of poor conductor, such as salt water. The electrodes are insulated except for their ends, which are small metal spheres of radius R_0 . The conductor obeys an ohmic relation, and the zero-frequency (DC) conductivity of the medium is a constant, κ . The ends are separated by a distance $R \gg R_0$. Find the total DC resistance between the two electrodes as a function of R and comment on the (possibly surprising) form of your answer.

[*Hints:* Start by noticing that the units of conductivity are not the same as those of $1/(\text{resistance})$. Think about the possible forms of the desired formula for resistance as a function of κ , R , and R_0 , in the stated limit. Next begin the problem by guessing a form for the electrostatic potential in the medium that solves the relevant equations and is approximately constant over each electrode in the stated limit. From the potential you can find the current density everywhere, as well as the total potential drop.]

7.3 Electrosurgery

Patients undergoing electrosurgery sometimes suffer burns around the perimeter of the electrode. Consider a thin circular metal disk electrode of radius a and potential ψ_0 in contact with a medium of conductivity κ . The circuit is completed by another

electrode at some unspecified, distant, place; for example, you could imagine it as a sphere at potential 0 far from the disk electrode.

- a. Show that the normal component of current density at the surface of the electrode is given by

$$j_{\perp} = \frac{2\kappa\psi_0}{\pi(a^2 - r^2)^{1/2}}$$

where r is distance from the center of the disk (so it ranges from 0 to a). (The circuit is completed by some unspecified other electrode far away from the one in question.) This is a hard problem; we'll need to discuss the strategy in detail during the recitation.

- b. Even if you can't do (a), proceed assuming it. Find the total current coming out of the electrode.
 c. Thus find the electrode's resistance.
 d. Find the rate of heat production in the medium and explain why the patient gets burns near the edge of the electrode.

Method: Here is an outline of how to do (a). We will replace the disk by an ellipsoid, and at the end of the calculation take a limit where the ellipsoid gets squashed down to the desired disk.

- i.* To define the ellipsoid, let σ be some positive constant (the distance from the center to one focus of an ellipse). Set up cylindrical coordinates ρ, φ, z centered on the center, with \hat{z} the axis of symmetry. Now define $r_{\pm} = \sqrt{z^2 + (\rho \mp \sigma)^2}$ and

$$\xi = (r_+ + r_-)/2\sigma, \quad \eta = (r_- - r_+)/2\sigma.$$

Thus the surface $\xi = \xi_0$ is an ellipsoid for any constant ξ_0 . Our goal is to find the potential outside a conductor whose surface is this ellipsoid, given that the potential drop between the surface and infinity is ψ_0 . When we've done that, then the limit $\xi_0 \rightarrow 1$ will give us the case of a thin disk.

- ii.* We know that the potential solves Laplace's equation outside the conductor, so we're faced with a familiar class of problem: a boundary-value problem. If we use ξ, η , and φ as coordinates, the problem is

$$\nabla^2\psi = 0, \quad \psi \rightarrow 0 \text{ at infinity}, \quad \text{and } \psi(\xi_0, \eta, \varphi) = \psi_0.$$

The boundary conditions look nice in these coordinates. Let's show that the Laplace operator is separable.

- iii.* Express ρ and z in terms of ξ and η . [*Hint:* Express $\xi\eta$ and $(\xi^2 - 1)(1 - \eta^2)$ in terms of ρ and z , then think.]

- iv.* Thus express x, y, z in terms of ξ, η , and φ . Differentiate to find the vector $\vec{e}_{(\xi)} \equiv \partial\vec{r}/\partial\xi$, and similarly $\vec{e}_{(\eta)}$ and $\vec{e}_{(\varphi)}$. These three vectors have a very nice property similar to the one we found in class for plane polar coordinates — what is it?

- v.* Use (iv) to express the volume element d^3r in terms of $d\xi d\eta d\varphi$.

- vi.* Use (iv,v) to express the integral $\int d^3r \vec{\nabla}f \cdot \vec{\nabla}g$ in the coordinates ξ, η , and φ . Here f and g are any two functions, both independent of the azimuthal angle φ and vanishing at infinity.

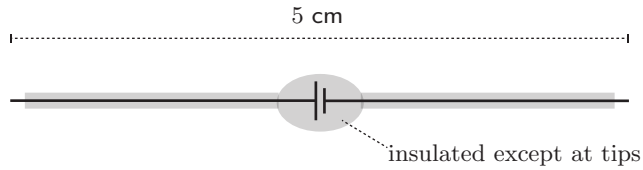


Figure 7.4:

vii. Use integration by parts to work out the Laplace operator $\nabla^2 g$ in these coordinates, for the case where g is independent of φ .

viii. We seek an exact solution $\psi = A(\xi)B(\eta)$ by separation of variables. Substitute this trial solution into the boundary conditions, and into your formula in (vii), and solve for ψ .

Other note: One might worry that the sharp edge of the disk could generate a singularity that gives pathological answer, like infinite (or zero) resistance. Indeed the formula above shows that there are large currents at the rim of the disk. But you'll find the total resistance is nice and finite.

7.4

Imagine a small current source (hearing-aid battery) with narrow wires sticking out. Everything is insulated except for the tips of the wires, which are separated by 5 cm. The whole thing is immersed in an infinite bath of isotropic conductor, for example seawater, and the current source supplies a steady total current $I = 1 \text{ mA}$ (Figure 7.4).

- What equation governs the steady electric potential throughout the seawater?
- Write down a solution to that equation appropriate to the problem by superposing two simpler solutions.
- The conductivity of seawater is $\kappa \approx 0.1 \Omega^{-1} \text{ m}^{-1}$. Use that fact, and the form of your answer to (b) up close to one electrode tip, to get the overall constant in front of your solution, and hence finish explicitly evaluating the steady electric potential throughout the seawater.

7.5 [Not ready yet.]

CHAPTER 8

Cell Membrane Capacitance

8.1 FRAMING

Every living cell needs a wrapper to maintain a distinct interior environment. We now know that this **plasma membrane** is a few nanometers thick, which is why nobody could see it prior to the invention of the electron microscope. Nevertheless, H. Fricke “saw” it (that is, deduced its existence and thickness) in 1925.

Actually, the existence of a molecular-scale membrane had been hypothesized prior to this. There was some precedent. Ben Franklin had long ago done measurements on the spreading of oil on an air-water interface; Rayleigh made these more precise and was brave enough to propose the interpretation, that oil could be spread to a monolayer just a few nanometers thick, but no further (without holes appearing).¹ Others realized that, even without an air-water interface, a *double* layer of such molecules could form, stably separating one aqueous medium from another one. There remained the problem of confirming this hypothesis and, if confirmed, characterizing the membrane in detail.

8.2 FRICKE EXPERIMENT

Living cells require a salty environment; in pure water, they burst from osmotic flow. And of course the interior of a cell is also salty (it contains small, mobile ions). But the membrane is largely impermeable to those ions, because of the high electrostatic cost to them of leaving a highly polarizable environment (water) and entering the nonpolar (oily) interior of the membrane. That is, the membrane amounts to a thin insulating layer between two conductors: a capacitor.

Fricke knew that such a thin membrane would have substantial capacitance, and that this might be relevant for electrophysiology, so he sought to measure the capacitance. Knowing that the lipid molecules constituting a membrane are similar to other oils then let him predict that the capacitance per unit area would be $\mathcal{C} = \epsilon/\delta$, where δ was the membrane thickness. Thus, knowing the permittivity ϵ and measuring \mathcal{C} would allow a determination of δ .

Naively, one could imagine stretching such a membrane all the way across a chamber, imposing a potential drop across it, and measuring how much charge flowed while establishing that drop (“charging the capacitor”). Incredibly, such an approach is possible today via **patch-clamp** measurements. But not in 1925. To get there 60 years ahead of when the measurement “ought” to have been possible, Fricke found a more clever approach.

¹Strutt, 1890.

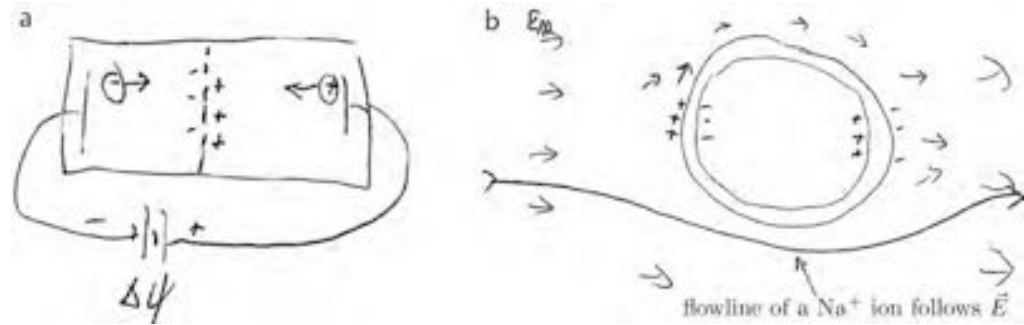


Figure 8.1: (a) Imagined experiment to measure membrane capacitance directly. (b) In Fricke's experiment, both electrodes were on the same side of the membrane (the cell exterior). The cell disrupted what would otherwise be a uniform flow of current (ions in solution). The membrane is modeled as a sphere of radius a .

Rather than having electrodes on either side of a membrane (Figure 8.1a), Fricke's experiment involved suspending many cells in salt water and passing alternating current through the chamber. I admit that I did not at once see how this would tell us anything about the membrane, but that is the point of these notes. The frequency of the current was around 100 kHz, so we may use the quasistatic approximation for our analysis.

We idealize the system as salt water on either side of a single spherical shell of radius a . (Later we will assume that if there are many cells, they are well separated in space.)

In a conducting medium, $\vec{j} = \kappa \vec{E}$. Because we assume that no current may cross the membrane, we must have $\vec{j}_\perp = 0$ at the inner and outer surfaces, and hence $\vec{E}_\perp = 0$ also. The system arranges this by having thin layers of excess charge pile up just outside the membrane as shown in Figure 8.1b. Elsewhere there is no net charge (good conductor), so $\vec{\nabla} \cdot \vec{E} = 0$. Thus, we may write $\vec{E} = -\vec{\nabla}\psi$ as usual, but with a jump in ψ as we cross the membrane.

We now have two decoupled electrostatics problems:

Inside cell

$\nabla^2\psi = 0$, subject to

$$\frac{\partial\psi}{\partial r} = 0 \quad \text{on the spherical surface } r = a.$$

There is only one spherical solution to the Laplace equation that is nonsingular at the origin, namely $\psi_{\text{in}} = \text{const.}$ ²

Outside cell

$\nabla^2\psi = 0$, subject to

$$\frac{\partial\psi}{\partial r} = 0 \quad \text{on the spherical surface } r = a \text{ and}$$

²See Section 4.3.

$$\psi \rightarrow -E_\infty z = -E_\infty r \cos \theta \quad \text{far away.}$$

There is one solution to the Laplace equation with the required behavior at infinity, namely $r \cos \theta$. In addition, many other solutions fall off at infinity, so adding them does not spoil the behavior at infinity. Of these, $r^{-2} \cos \theta$ has the same angular dependence as the first one, and so is a candidate to fix its behavior at $r = a$. (It doesn't matter that the second of these solutions is singular at $r = 0$, because we are only interested in the exterior region.)

Enforcing the boundary condition lets us find the unknown constant α multiplying the second solution:

$$\begin{aligned} 0 &= \left. \frac{\partial}{\partial r} \right|_a (-E_\infty r \cos \theta + \alpha r^{-2} \cos \theta) \\ 0 &= -E_\infty - 2\alpha a^{-3} \\ \psi_{\text{out}} &= -E_\infty \cos \theta \left(r + \frac{1}{2} \frac{a^3}{r^2} \right) + \text{const.} \end{aligned}$$

Match the solutions

By symmetry, ψ must not jump as we cross the membrane at the equatorial plane $\theta = \pi/2$. So $\psi_{\text{out}} = \psi_{\text{in}}$ there and we may take both = 0.

8.2.1

We solved the electrostatic problem, but we still must connect to what was experimentally measured. First notice that the potential jump across the membrane is $\Delta\psi(\theta) = -E_\infty \frac{3a}{2} \cos \theta$. Each surface area element is a capacitor charged to that potential, and hence stores energy

$$d\mathcal{E} = \frac{1}{2} (\Delta\psi)^2 dC \quad \text{where} \quad dC = \mathcal{C} d^2\Sigma.$$

The total stored energy is then

$$\begin{aligned} \mathcal{E} &= \int d\mathcal{E} = \int (\sin \theta d\theta d\varphi) \frac{1}{2} E_\infty^2 \cos^2 \theta \frac{9a^2}{4} \mathcal{C} \\ &= \frac{3\pi}{2} a^4 E_\infty^2 \mathcal{C}. \end{aligned}$$

For N well-separated cells in suspension, the total is N times this formula.

Fricke applied alternating voltage $\Delta\psi(t) = \bar{\psi} \cos(\omega t)$, and measured the resulting current. The current had the same frequency ω , so its form was $\bar{I} \cos(\omega t - \phi)$; Fricke therefore measured the dependence of \bar{I} and ϕ on $\bar{\psi}$ and ω at fixed, known values of N and a . We wish to see what our solution to the electrostatic problem predicts about this relationship, with the goal of finding the numerical value of the one remaining unknown parameter, the areal density of membrane capacitance \mathcal{C} .

8.2.2

The electric power entering the experimental chamber is

$$\mathcal{P} = \psi I = \bar{\psi} \bar{I} \cos \omega t \cos(\omega t - \phi)$$

$$= \bar{\psi} \bar{I} (\cos^2 \omega t \cos \phi + \cos \omega t \sin \omega t \sin \phi). \quad (8.1)$$

The first term is always nonnegative. This represents ohmic (resistive) dissipation of energy into heat. The second term averages to zero. This indicates an “elastic” element, constantly storing energy and giving it back. The storage mechanism is the charging and discharging of the membrane capacitance, so this term must equal

$$-\frac{d}{dt}(N\mathcal{E}) = -N \frac{3\pi}{2} a^4 \mathcal{C} \left(\frac{\bar{\psi}}{L}\right)^2 \frac{d}{dt} \cos^2 \omega t,$$

where L is the length of the chamber (distance between electrodes).

$$= N 3\pi a^4 \mathcal{C} \omega \cos \omega t \sin \omega t.$$

Compare that result to the second term of Equation 8.1 to find

$$\bar{\psi} \bar{I} \sin \phi = N 3\pi a^4 \mathcal{C} \omega \left(\frac{\bar{\psi}}{L}\right)^2.$$

Rearranging gives the desired result

$$\mathcal{C} = \frac{\bar{I} L^2 \sin \phi}{\bar{\psi} N 3\pi a^4 \omega}.$$

Fricke substituted his experimental values and found $C \approx 1 \mu\text{F}/\text{cm}^2$. The permittivity of oil is around $3\epsilon_0$, so he inferred a membrane thickness value $\delta \approx 3 \text{ nm}$, within a factor of two of the accepted value. Remarkably, that value is also similar to the one implied by measurements made by Franklin in 1773!

FURTHER READING

Sohn et al., 2000.

PROBLEMS

8.1 Measure cell membrane capacitance

In this problem you’ll find an experimentally practical way to measure the capacitance of a cell membrane.

Electrically speaking, a sea urchin egg is a thin spherical shell of insulator (the cell’s bounding membrane), surrounded by a medium-good conductor (sea water), and enclosing a medium-good conductor (also a salt solution). The apparatus consists of a suspension of such eggs in a chamber, which is a rectangular prism. Plates at either end set up a potential drop from one end of the chamber to the other, that is, from $z = 0$ to $z = L$.

An imperfect conductor can maintain a nonzero electric potential gradient according to the ohmic relation $\vec{j} = \kappa \vec{E}$, where κ is the conductivity. The component of the electric current perpendicular to the membrane must equal zero just inside and

just outside of it, because the membrane is assumed to be an insulator. Hence the component $\vec{E}_\perp = 0$ there as well.

Suppose that there is only one egg in the chamber. Set up spherical polar coordinates centered on the center of the egg, and let its radius be a . Then the electric potential approaches

$$\psi(\vec{r}) \rightarrow (\Delta\psi_{\text{tot}})(z/L + 1/2) = (\Delta\psi_{\text{tot}}/(2L))(r \cos \varphi)$$

far from the egg. In static, or quasistatic, conditions, we also know that $\vec{\nabla}^2\psi = \vec{\nabla} \cdot (-\vec{E}) = -(\vec{\nabla} \cdot \vec{j})/\kappa = 0$ everywhere.

- Inside the egg, we are looking for a solution to the Laplace equation that is everywhere nonsingular for $r < a$ and that satisfies $\hat{r} \cdot \vec{\nabla}\psi = 0$ at $r = a$. We can shift the solution by a constant to ensure $\psi(\vec{0}) = 0$. There is only one solution with all of those properties. What is it?
- Outside the egg, there are two linearly independent solutions to the Laplace equation with the required angular dependence. One of those solutions is $\psi(\vec{r}) = (\Delta\psi_{\text{tot}}/(2L))(r \cos \varphi)$, which of course also behaves the way we want at infinity. But by itself this solution does not satisfy the boundary condition at $r = a$. A second solution is a friend from the multipole expansion; it is singular at $r = 0$, but that's irrelevant, as we are only looking for solutions outside. Find the unique linear combination of these two solutions that does satisfy the boundary conditions both at $r = a$ and at $r = \infty$.
- We now know the potential inside and outside of the membrane. They are not equal at $r = a$: There is a potential jump

$$\Delta\psi(\theta, \varphi) = \psi_{\text{out}}(a, \theta, \varphi) - \psi_{\text{in}}(a, \theta, \varphi)$$

across the membrane. (Note that $\Delta\psi(\theta, \varphi)$ is the potential jump across the membrane at the indicated location; it is not the same as $\Delta\psi_{\text{tot}}$, which is the total potential drop across the entire experimental chamber.)

An insulating layer with a potential jump across it is a capacitor. Call the capacitance per unit area of the membrane \mathcal{C} ; this is the quantity that we wish to know. We therefore seek a formula connecting \mathcal{C} to something experimentally measurable. So each small area element $d\Sigma$ of the membrane has capacitance $dC = \mathcal{C}d\Sigma$. If a potential jump is maintained across the membrane at position θ, φ , this means that the patch of membrane is storing energy $d\mathcal{E} = \frac{1}{2}(\Delta\psi(\theta, \varphi))^2 dC$.

Integrate $d\mathcal{E}$ over the sphere to find the total capacitively stored energy in terms of $\Delta\psi_{\text{tot}}$, a , L , and \mathcal{C} .

- Now imagine applying a $\Delta\psi_{\text{tot}}$ that is alternating in time. In pure sea water, we'd just get an alternating current in the chamber that is always in phase with $\Delta\psi_{\text{tot}}$. With one (or more) eggs present, however, energy storage in the cell membrane will generate a phase lag between potential drop and current. Let $\Delta\psi_{\text{tot}}(t) = \psi_0 \cos(\omega t)$ for some frequency ω and amplitude ψ_0 , and write the resulting current as $I(t) = I_0 \cos(\omega t - \phi)$. By imposing $\Delta\psi_{\text{tot}}(t)$, and measuring $I(t)$, we can in principle find I_0 and ϕ .

The phase lag ϕ is a measurable manifestation of membrane capacitance. To get a prediction for it, notice that the power sent into the experimental chamber is

$\Delta\psi_{\text{tot}}(t)I(t)$. Express this in terms of ψ_0 , I_0 , ω , and ϕ . Your expression will have two terms. One term is always positive, and indicates electrical energy sent from the current source into the chamber, and converted to heat via resistive losses.

- e. The other term is more interesting, because its *time average is zero*. This is the result of a system element that is constantly *exchanging* energy with the external source. You can write another expression for the exchanged energy, as the time derivative of the stored energy that you found in part 3. Set these two expressions equal, to obtain a formula relating a , L , \mathcal{C} , ω , I_0 , ψ_0 , and ϕ . One of these quantities, \mathcal{C} , is the thing we wished to measure. All the others are either set by the experimenter (L , ω , ψ_0) or measured (a , I_0 , ϕ).

- f. If there are N eggs in the chamber, then the stored energy is N times as great as before. Modify your formula to include N , and explain in words how to measure \mathcal{C} .

Before we actually do an experiment, it's essential to make an estimate, to see if the supposedly measurable quantity really is measurable. (If not, then we say, "Oh Fricke" and redesign the experiment before we waste time doing it.)

The goal of the experiment was to measure \mathcal{C} , the membrane capacitance per unit area. But when designing the experiment, we turn things around and use an *estimate* for \mathcal{C} , in order to predict whether the observed phase lag ϕ between voltage and current will be large enough to measure (e.g. on an oscilloscope). Thus assume $\mathcal{C} \approx 1\mu\text{F cm}^{-2}$. Here are some other typical numbers I extracted from Fricke's original paper:

Cells of the sort studied by Fricke and Cole have radius $a \approx 3 \cdot 10^{-4}\text{cm}$.

The applied current had a frequency of 87000Hz, or angular frequency $\omega = 2\pi \times 87000\text{s}^{-1}$.

The overall resistance of the seawater in the chamber was $\psi_0/I_0 \approx 300\Omega$.

The number density of cells in the chamber is such that they occupy about 20% of the chamber volume.

The chamber dimensions are: cross-section $A \approx 15\text{cm}^2$, length $L \approx 7\text{cm}$.

- g. Use these numbers and our analysis to find the predicted phase lag angle ϕ in radians. (Ahem, make sure the units work out properly.) Does it seem likely to be measurable?

8.2 Fricke 2

Use a computer to visualize the electrostatic potential outside a spherical cell in conducting solution, with an applied \vec{E} field at infinity that is uniform along \hat{z} . That is, show $\psi(x, 0, z)$ as height above or below the xy plane. Then show the same function as a contour plot. Describe in words the relevant physical aspects of the solution. Finally, make a vector-field plot of the corresponding \vec{E} field.

CHAPTER 9

Electrostatics in Solution

Section 2.1 mentioned that it is often important to find condensed (implicit) descriptions of some of the actors in a complex system. Thus, we would like to follow mobile charges explicitly but not have to think about everything else. One example of this approach was our introduction of a modified permittivity to account for a dielectric medium. This chapter introduces another example, where we account for the incessant thermal bumping of uncharged actors, for example, water molecules, against the charges of interest via a Boltzmann distribution.

9.1

9.1.1 The Nernst relation sets the scale of membrane potentials

Many of the molecules floating in water carry a net electric charge, unlike the water molecules themselves. When table salt dissolves, for example, the individual sodium and chlorine atoms separate, but the chlorine atom grabs one extra electron from sodium, thereby becoming a negatively charged chloride ion, Cl^- , and leaving the sodium as a positive ion, Na^+ . Any electric field present in the solution will then exert forces on the individual ions, dragging them just as gravity drags colloidal particles to the bottom of a test tube.

Suppose first that we have a uniform-density solution of charged particles, each of charge q , in a region with electric field \vec{E} . For example, we could place two parallel, flat plates just outside the solution's container, a distance h apart, and connect them to a battery that maintains a constant electrostatic potential difference $\Delta\psi$ across them. We know from first-year physics that $E = \Delta\psi/h$ and each charged particle feels a force $q\vec{E}$, so it drifts with the net speed $v_{\text{drift}} = qE/\eta$, where η is a constant describing viscous friction.

Imagine a small net of area Σ stretched out perpendicular to the electric field (that is, parallel to the plates); see Figure 9.1. To find the flux of ions induced by the field, we ask how many ions get caught in the net each second. The average ion drifts a distance $v_{\text{drift}}dt$ in time dt , so, in this time, all the ions contained in a slab of volume $v_{\text{drift}}dt\Sigma$ get caught in the net. The number of ions caught equals this volume times the number density c_{ion} . The flux j_{ion} is then the number crossing per area per time, or $c_{\text{ion}}v_{\text{drift}}$. (Check to make sure this formula has the proper units.) Substituting the drift velocity gives $j_{\text{ion}} = qEc_{\text{ion}}/\eta$, the **electrophoretic flux** of ions.

Now suppose that the density of ions is *not* uniform. For this case, we add the driven (electrophoretic) flux just found to the diffusive (Fick's law) flux, obtaining

$$j_{\text{ion}}(x) = \frac{qE(x)c_{\text{ion}}(x)}{\eta} - D_{\text{ion}} \frac{dc_{\text{ion}}}{dx},$$

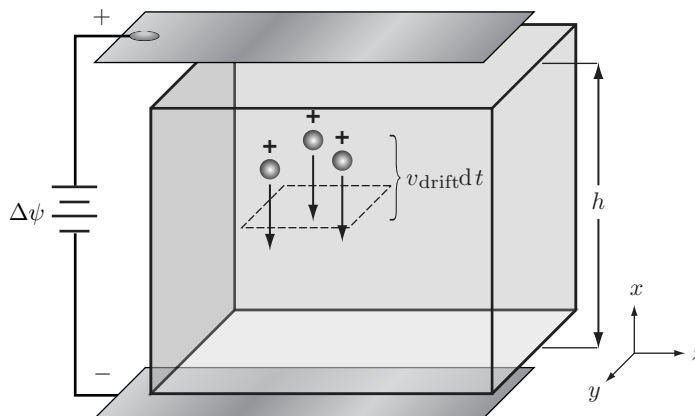


Figure 9.1: [Sketch.] **Origin of the Nernst relation** (Equation 9.3). An electric field pointing downward drives positively charged ions down. The system comes to equilibrium with a downward density gradient of positive ions and an upward gradient of negative ions. The number flux \vec{j}_{ion} for the ion species shown points downward with magnitude equal to the number density c_{ion} times v_{drift} . The corresponding contribution to charge flux is $q\vec{j}_{\text{ion}}$.

where D_{ion} is the diffusion constant for the ion species in question. We next rewrite the viscous friction coefficient in terms of D_{ion} , using the Einstein relation $\eta D_{\text{ion}} = k_{\text{B}}T$ to get¹

$$j_{\text{ion}} = D_{\text{ion}} \left(-\frac{dc_{\text{ion}}}{dx} + \frac{q}{k_{\text{B}}T} E c_{\text{ion}} \right). \quad \text{Nernst-Planck formula} \quad (9.1)$$

The Nernst-Planck formula helps us to answer a fundamental question: What electric field would be needed to get *zero* net flux, that is, to cancel the diffusive tendency to erase nonuniformity? To answer the question, we set $j_{\text{ion}} = 0$ in Equation 9.1. In a planar geometry, where everything is constant in the y, z directions, we get the condition

$$\frac{1}{c_{\text{ion}}} \frac{dc_{\text{ion}}}{dx} = \frac{q}{k_{\text{B}}T} E. \quad (\text{in equilibrium}) \quad (9.2)$$

The left side of this formula can be written as $\frac{d}{dx}(\ln c_{\text{ion}})$.

To use Equation 9.2, we now integrate both sides from the top plate to the bottom one (see Figure 9.1). The left side is $\int_0^h dx \frac{d}{dx} \ln c_{\text{ion}} = \ln(c_{\text{bot}}/c_{\text{top}})$, that is, the difference in $\ln c_{\text{ion}}$ from one plate to the other.² To understand the right side, we first note that qE is the force acting on a charged particle, so the particle's potential energy obeys $-dU/dx = qE$, or $U(x) = -qEx$. The electrostatic potential ψ is the potential energy per unit charge, so $\Delta\psi \equiv \psi_{\text{bot}} - \psi_{\text{top}} = -Eh$. Writing $\Delta(\ln c_{\text{ion}})$ for

¹[\mathcal{T}_2] In three-dimensional language, the Nernst-Planck formula becomes $\vec{j}_{\text{ion}} = D_{\text{ion}}(-\vec{\nabla}c_{\text{ion}} + (q/k_{\text{B}}T)\vec{E}c_{\text{ion}})$.

²Normally it is meaningless to speak of a nonlinear function like log applied to a quantity with units. However, a difference of two such logs can be written as the log of the dimensionless ratio, so we always get the same result regardless of what units we choose.

$\ln(c_{\text{bot}}/c_{\text{top}})$ then gives the condition for equilibrium:

$$\Delta(\ln c_{\text{ion}}) = -q\Delta\psi_{\text{eq}}/k_{\text{B}}T \quad \text{in equilibrium (Nernst relation).} \quad (9.3)$$

The subscript on $\Delta\psi_{\text{eq}}$ reminds us that this is the voltage needed to maintain a concentration jump *in equilibrium*.

Equation 9.3 predicts that positive charges will migrate toward the bottom of Figure 9.1. It makes sense: They're attracted to the negative plate. We have so far been ignoring the corresponding negative charges (for example, the chloride ions in table salt), but the same formula applies to them as well. Because they carry negative charge ($q < 0$), Equation 9.3 says they migrate toward the positive plate.


Substituting some real numbers into Equation 9.3 yields a suggestive result. Consider a singly charged ion like Na^+ , for which $q = e$. Suppose that we have a moderately big concentration jump, $c_{\text{bot}}/c_{\text{top}} = 10$. Using the fact that $(k_{\text{B}}T_{\text{r}}/e) \approx \frac{1}{40}$ volt, we find $\Delta\psi \approx +58$ mV. What's suggestive about this result is that many living cells, particularly nerve and muscle cells, really do maintain a potential difference across their membranes of a few tens of millivolts! We haven't proven that these are equilibrium (Nernst) potentials, and indeed they're not. But the observation *does* show that dimensional arguments successfully predict the scale of membrane potentials with almost no hard work at all.

Something interesting happened on the way from Equation 9.1 to Equation 9.3: When we consider equilibrium only, the value of D_{ion} drops out. That's reasonable: D_{ion} controls how *fast* things move in response to a field; its units involve time. But equilibrium is an eternal state; it can't depend on time. In fact, exponentiating the Nernst relation gives that $c_{\text{ion}}(x)$ is a constant times $e^{-q\psi(x)/k_{\text{B}}T}$. This result is an old friend: It says that the spatial distribution of ions follows the Boltzmann distribution. A charge q in an electric field has electrostatic potential energy $q\psi(x)$ at x ; its probability to be there is proportional to the exponential of minus its energy, measured in units of the thermal energy $k_{\text{B}}T$. Thus, a positive charge doesn't like to be in a region of large positive potential, and vice versa for negative charges. Our formulas are mutually consistent.

9.1.2 The electrical resistance of a solution reflects frictional dissipation

Suppose that we place the metal plates in Figure 9.1 *inside* the container of salt water, so that they become electrodes. Then the ions in solution migrate, but they don't accumulate: The positive ones get electrons from the $-$ electrode, whereas the negative ones hand their excess electrons over to the $+$ electrode. The resulting neutral atoms leave the solution; for example, they can electroplate onto the attracting electrode or bubble away as gas.³ Then, instead of establishing equilibrium, our system continuously *conducts* electricity, at a rate controlled by the steady-state ion fluxes.

The potential drop across our cell is $\Delta\psi = Eh$, where h is the separation of the plates. According to the Nernst–Planck formula (Equation 9.1), this time with uniform

³  Electroplating does not occur with a solution of table salt because sodium metal is so strongly reactive with water. Nevertheless, the following discussion is valid for the *alternating-current* conductivity of NaCl.

c_{ion} , the electric field is $E = (k_{\text{B}}T/(D_{\text{ion}}qc_{\text{ion}}))j_{\text{ion}}$. Recall that j_{ion} is the number of ions passing per area per time. To convert this expression to the total electric current I , note that each ion deposits charge q when it lands on a plate; thus, $I = qj_{\text{ion}}\Sigma$, where Σ is the plate area. Putting everything together gives

$$\Delta\psi = \left(\frac{k_{\text{B}}T}{D_{\text{ion}}q^2c_{\text{ion}}\Sigma} \right) I. \quad (9.4)$$

This is a familiar result: It says that our solution is ohmic, $\Delta\psi = IR$. Equation 9.4 gives the electrical resistance R of the cell as the constant of proportionality between voltage and current. To use this formula, we must remember that each type of ions contributes to the total current; for table salt, we need to add separately the contributions from Na^+ with $q = e$ and Cl^- with $q = -e$, or in other words, double the right-hand side of the formula.

The resistance depends not only on the solution but also on the geometry of the cell. It's customary to eliminate the geometry dependence by defining the electrical conductivity of the solution as $\kappa = h/(R\Sigma)$. Then our result is that each ion species contributes $D_{\text{ion}}q^2c_{\text{ion}}/k_{\text{B}}T$ to κ . It makes sense: Saltier water conducts better.

[\[T2\]](#) Section 9.1.2' (page 112) mentions other points about electrical conduction.

9.2 A REPULSIVE INTERLUDE

Until now, we have studied osmotic forces under the assumption that interactions between solute particles can be neglected. That may be reasonable for sugar, whose molecules are uncharged; but, as we'll see in a moment, electrostatic interactions between the objects contained in a cell can be immense. Accordingly, this section will introduce *mixed* forces, those that are partly entropic and partly energetic.

9.2.1 Electrostatic interactions are crucial for proper cell functioning

Biomembranes and other big objects (such as DNA) are often said to be “electrically charged.” The term can cause confusion. Doesn't matter have to be neutral? Let's recall why people said that in first-year physics.

Ex. Consider a raindrop of radius $R = 1 \text{ mm}$ suspended in air. How much work would be needed to remove just one electron from just 1% of the water molecules in the drop?

Solution: Removing an electron leaves some water molecules electrically charged. These charged water molecules migrate to the surface of the drop to get away from one another, thereby forming a shell of charge of radius R . The electrostatic potential energy of such a shell (also called its **Born self-energy**) is $\frac{1}{2}q\psi(R)$, or $q^2/(8\pi\epsilon_0R)$. The charge q on the drop equals the number density of water molecules, times the drop volume, times the charge on a proton, times 1%. Squaring gives

$$\left(\frac{q}{e}\right)^2 = \left(\frac{10^3 \text{ kg}}{\text{m}^3} \frac{6 \cdot 10^{23}}{0.018 \text{ kg}} \times \frac{4\pi}{3}(10^{-3} \text{ m})^3 \times 0.01\right)^2 = 1.9 \cdot 10^{36}.$$

Multiplying by $2.3 \cdot 10^{-28} \text{ J m}$ and dividing by $2R$ yields about $2 \cdot 10^{11} \text{ J}$.

Two hundred billion joules is a lot of energy—certainly it’s much bigger than $k_B T_r$! And indeed, macroscopic objects really are electrically neutral (they satisfy the condition of “bulk electroneutrality”). But things look different in the nanoworld.

Your Turn 9A

Repeat the calculation for a droplet of radius $R = 1 \mu\text{m}$ in water. Recall that the permittivity ϵ of water is about 80 times bigger than the one for air used in the Example. Then repeat for an $R = 1 \text{ nm}$ object in water.

Thus it *is* possible for thermal motion to separate a neutral molecule into charged fragments. For example, when we put an acidic macromolecule such as DNA in water, some of its loosely attached protons can wander away, leaving some of their electrons behind. In this case, the remaining macromolecule has a net negative charge: DNA becomes a negative **macroion**. This is the sense in which DNA is charged. The lost atoms are positively charged; they are called **counterions**, because their net charge counters (neutralizes) the macroion. Positive ions are also called **cations**, because they’d be attracted to a *cathode*; similarly, the remaining macroion is called **anionic**.

The counterions diffuse away because they were not bound by chemical (covalent) bonds in the first place and because by diffusing away, they increase their entropy.

The counterions, having left the macroion, now face a dilemma. If they stay too close to home, they won’t gain much entropy. But to travel far from home requires lots of energy, to pull away from the opposite charges left behind on the macroion. The counterions thus need to make a compromise between the competing imperatives to minimize energy and maximize entropy. This section will show that for a large flat macroion, the compromise chosen by the counterions is to remain hanging in a cloud near the macroion’s surface. After working Your Turn 9A, you won’t be surprised to find that the cloud can be a couple of nanometers thick. Viewed from *beyond* the counterion cloud, the macroion appears neutral. Thus, a second approaching macroion won’t feel any attraction or repulsion until it gets closer than about twice the cloud’s thickness. This behavior is quite different from the behavior of charges in a vacuum: In that case, the electric field doesn’t fall off with distance at all! In short,

*Electrostatic interactions are of long range in vacuum. But in solution, a screening effect reduces this interaction’s **effective** range, typically to a nanometer or so.* (9.5)

We’d like to understand the formation of the counterion cloud, which is often called the **diffuse charge layer**. Together with the charges left behind in the surface, it forms an **electric double layer** surrounding a charged macroion. The previous paragraph makes it clear that the forces on charged macroions have a mixed character: They are partly electrostatic and partly entropic. Certainly, if we could turn off thermal motion, the diffuse layer would collapse back onto the macroion, thereby leaving it neutral, and there’d be no force at all; we’ll see this in the formulas we obtain for the forces.

Before we proceed to calculate properties of the diffuse charge layer, two remarks may help set the biological context.

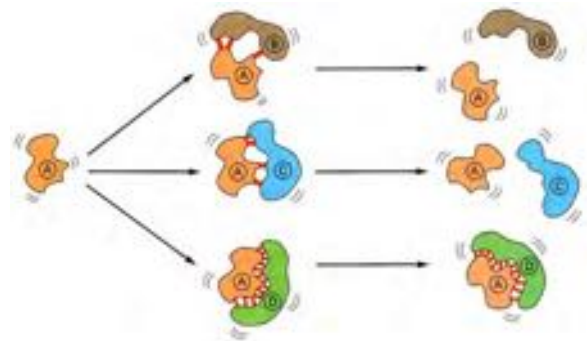


Figure 9.2:

First, the cells in your body contain a variety of macromolecules. A number of attractive forces are constantly trying to stick the macromolecules together, for example, the depletion force or the more complicated van der Waals force. It wouldn't be nice if they just acquiesced, clumping into a ball of sludge at the bottom of the cell, with the water on top. The same problem bedevils many industrial colloidal suspensions, for example, paint. One way Nature, and we its imitators, avoid this “clumping catastrophe” is to arrange for the colloidal particles to have the same sign of net charge. Indeed, most of the macromolecules in a cell are negatively charged and hence repel one another.

Second, the fact that electrostatic forces are effectively of short range in solution (summarized in Idea 9.5 above) matters crucially for cells, because it means that

- Macroions will not feel one another until they're nearby, but
- Once they *are* nearby, the *detailed surface pattern* of positive and negative residues on a protein can be felt by its neighbor, not just the overall charge (Figure 9.2).

This observation goes to the heart of how cells organize their myriad internal biochemical reactions. Although thousands of macromolecules may be wandering around any particular location in the cell, typically only those with precisely matching shapes and charge distributions will bind together. We can now see that the root of this amazing specificity is that

Even though each individual electrostatic interaction between matching charges is rather weak (relative to $k_B T_r$), still the combined effect of many such interactions can lead to strong binding of two molecules—if their shapes and orientations match precisely. (9.6)

Notice that it's not enough for two matching surfaces to come together; they must also be properly oriented before they can bind. We say that macromolecular binding is **stereospecific**.

Thus, understanding the very fact of molecular recognition, which is crucial for the operation of every cell process, requires that we first understand the counterion cloud around a charged surface.

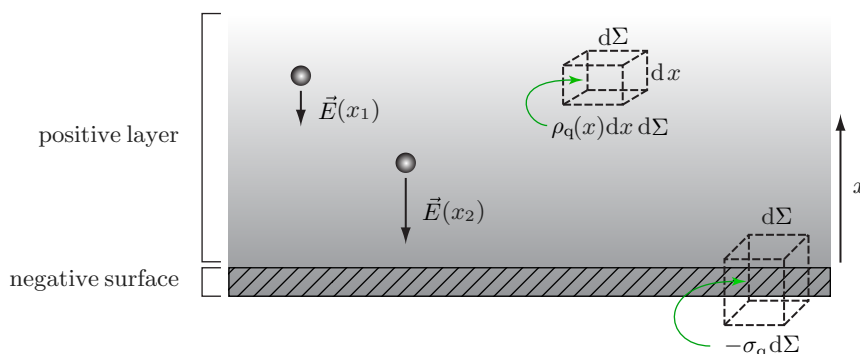


Figure 9.3: [Schematic.] **A planar distribution of charges.** A thin sheet of negative charge (*hatched, bottom*) lies next to a neutralizing positive layer of free counterions (*shaded, top*). The individual counterions are not shown; the shading represents their average density. The lower box encloses a piece of the surface; so it contains total charge $-\sigma_q d\Sigma$, where $d\Sigma$ is its cross-sectional area and $-\sigma_q$ is the surface charge density. The upper box encloses charge $\rho_q(x) dx d\Sigma$, where $\rho_q(x)$ is the charge density of counterions. The electric field $\vec{E}(x)$ at any point equals the electrostatic force on a small test particle at that point, divided by the particle's charge. For all positive x , the field points along the $-\hat{x}$ direction. The field at x_1 is weaker than that at x_2 , because the repelling layer of positive charge between x_1 and $x = 0$ is thicker than that between x_2 and $x = 0$. Moreover, there is less positive charge between x_1 and infinity pushing a test charge downward than between x_2 and infinity.

9.2.2 The Gauss law

Figure 9.3 represents a thin, negatively charged sheet with uniform surface charge density $-\sigma_q$, next to a spread-out layer of positive charge with volume charge density $\rho_q(x)$. Thus σ_q is a positive constant with units coul m^{-2} , whereas $\rho_q(x)$ is a positive function with units coul m^{-3} . Everything is constant in the \hat{y} and \hat{z} directions. We'll write E for the component of the electric field in the \hat{x} direction.

The electric field above the negative sheet is a vector pointing along the $-\hat{x}$ direction, so the function $E(x)$ is everywhere negative. Just above the sheet, the electric field is proportional to the surface charge density: Applying the Gauss law for a flat, charged surface gives

$$E|_{\text{surface}} = -\sigma_q/\epsilon. \quad (9.7)$$

Away from the surface, the Gauss law gives

$$\frac{dE}{dx} = \frac{\rho_q}{\epsilon}. \quad (9.8)$$

Section 9.2.3 will use this relation to find the electric field everywhere outside the surface.

9.2.3 A charged surface in pure water is surrounded by a neutralizing ion cloud

The mean field

Now we can return to the problem of ions in solution. A typical problem might be to consider a thin, flat, negatively charged surface with surface charge density $-2\sigma_q$ and water on both sides. For example, cell membranes are negatively charged. You might want to coax DNA to enter a cell (say, for gene therapy). Because both DNA and cell membranes are negatively charged, you'd need to know how much they repel.

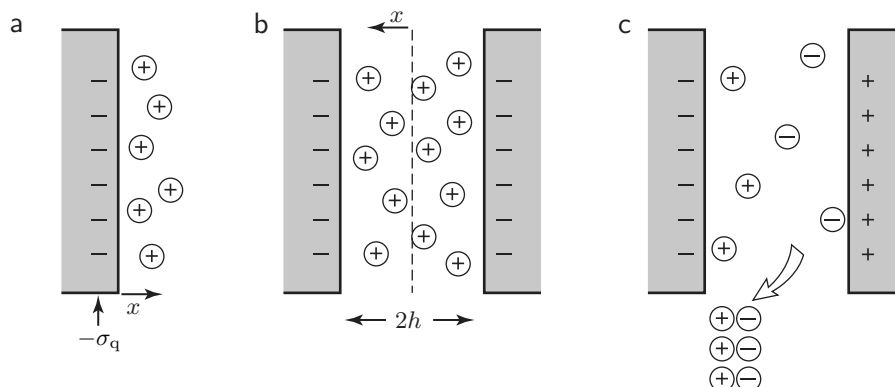


Figure 9.4: [Schematics.] **Behavior of counterions near surfaces.** (a) Counterion cloud outside a charged surface with surface charge density $-\sigma_q$. (b) When two similarly charged surfaces approach, their counterion clouds begin to get squeezed. (c) When two oppositely charged surfaces approach, their counterion clouds are liberated and entropy increases.

An equivalent, and slightly simpler, problem is that of a *solid* surface carrying charge density $-\sigma_q$, with water on just one side (Figure 9.4a). Also for simplicity, suppose that the loose positive counterions are **monovalent** (for example, sodium, Na^+). That is, each carries a single charge: $q_+ = e = 1.6 \cdot 10^{-19}$ coul. In a real cell, there will be additional ions of *both* charges from the surrounding salt solution. The negatively charged ones are called **coions** because they have the same charge as the surface. We will neglect the coions for now (see Section 9.2.3', page 112).

As soon as we try to find the electric field in the presence of mobile ions, an obstacle arises: We are not given the distribution of the ions, as we were in first-year physics, but instead must *find* it. Moreover, electric forces are of long range. The unknown distribution of ions will thus depend on each ion's interactions not only with its nearest neighbors but also with many other ions! How can we hope to model such a complex system?

Let's try to turn adversity to our advantage. If each ion interacts with many others, perhaps we can approach the problem by thinking of each ion as moving independently of the others' *detailed* locations but under the influence of an electric potential created by the *average* charge density of the others, or $\langle \rho_q \rangle$. We call this approximate electric potential $\psi(x)$ the **mean field** and this approach the **mean-field approximation**. The approach is reasonable if each ion feels many others; then the relative fluctuations in $\psi(x)$ about its average will be small. To make the notation less cumbersome, we will drop the averaging signs; from now on, ρ_q refers to the average density.

The Poisson–Boltzmann equation

We want $c_+(x)$, the number density of counterions. We are supposing that our surface is immersed in pure water; hence, far away from the surface, $c_+ \rightarrow 0$. The electrostatic potential energy of a counterion at x is $e\psi(x)$. We are treating the ions as moving independently of each other in a fixed potential $\psi(x)$, so the density of counterions, $c_+(x)$, is given by the Boltzmann distribution. Thus $c_+(x) = c_0 e^{-e\psi(x)/k_B T}$, where c_0 is a constant. We can add any constant we like to the potential because this change

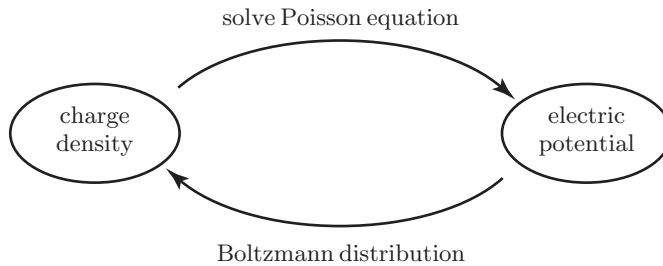


Figure 9.5: [Diagram.] **Strategy to find the mean-field solution.** Neither the Poisson equation nor the Boltzmann distribution alone can determine the charge distribution, but solving these two equations in two unknowns simultaneously does the job.

doesn't affect the electric field $E = -d\psi/dx$. It's convenient to choose the constant so that $\psi(0) = 0$. This choice gives $c_+(0) = c_0$; so the unknown constant c_0 is the concentration of counterions at the surface.

Unfortunately, we don't yet know $\psi(x)$. To find it, apply the Gauss law, taking ρ_q equal to the number density of counterions times e . The potential obeys the Poisson equation: $d^2\psi/dx^2 = -\rho_q/\epsilon$. Given the charge density, we can solve the Poisson equation for the electric potential. The charge density, in turn, is given by the Boltzmann distribution as $ec_+(x) = ec_0e^{-e\psi(x)/k_B T}$.

It may seem as though we have a chicken-and-egg problem (Figure 9.5): We need the average charge density ρ_q to solve the Poisson equation for the potential ψ . But we need ψ to find ρ_q from the Boltzmann distribution! Luckily, each of the arrows in Figure 9.5 represents an equation in two unknowns, namely, ρ_q and ψ . We just need to solve these two equations simultaneously to find the two unknowns.

Before proceeding, let's take a moment to tidy up our formulas. First, we combine the various constants into a length scale:

$$\ell_B \equiv \frac{e^2}{4\pi\epsilon k_B T}. \quad \text{Bjerrum length, in water} \quad (9.9)$$

ℓ_B tells us how close together we can push two like-charge ions, if we have energy $k_B T$ available. For monovalent ions in water at room temperature, $\ell_B = 0.71$ nm. Next, define the dimensionless rescaled potential $\bar{\psi}$:

$$\bar{\psi}(x) \equiv e\psi(x)/k_B T. \quad (9.10)$$

Your Turn 9B

Now combine the Poisson equation with the Boltzmann distribution to get

$$\frac{d^2\bar{\psi}}{dx^2} = -4\pi\ell_B c_0 e^{-\bar{\psi}}. \quad \text{Poisson-Boltzmann equation} \quad (9.11)$$

The payoff for introducing the abbreviations $\bar{\psi}$ and ℓ_B is that now Equation 9.11 is less cluttered, and we can verify at a glance that its dimensions work: Both d^2/dx^2 and $\ell_B c_0$ have units m^{-2} .

Like any differential equation, Equation 9.11 has, not one, but a whole *family* of solutions. To get a unique solution, we need to specify additional information, namely, some boundary conditions on the unknown function $\psi(x)$. For example, if you throw a rock upward, Newton's Law says that its height $z(t)$ obeys the equation $d^2z/dt^2 = -g$. But this equation won't tell us how high the rock will go! We also need to specify how hard you threw the rock, or more precisely, its speed and location when it left your hand at time zero. Similarly, we should not expect Equation 9.11 to specify the full solution because it doesn't mention the surface charge density. Instead, the equation has a family of solutions; we must choose the one corresponding to the given value of σ_q .

To see how σ_q enters the problem, we now apply the surface form of the Gauss law (Equation 9.7), which gives $-\frac{d\bar{\psi}}{dx}\Big|_{\text{surface}} = -\sigma_q/\epsilon$, or

$$\frac{d\bar{\psi}}{dx}\Big|_{\text{surface}} = 4\pi\ell_B \frac{\sigma_q}{e}. \quad (\text{when the allowed region is } x > 0) \quad (9.12)$$

When using this formula, remember that σ_q is a positive number; the surface has charge density $-\sigma_q$.

Ex. How does one remember the correct sign in this formula?

Solution: The electrostatic potential ψ goes down as we approach a negative object. Thus, approaching counterions feel their potential energy $e\psi$ decrease as they approach the surface, so they're attracted. If x is the distance from a negatively charged surface, then ψ will be decreasing as we approach it, or increasing as we leave: $d\psi/dx > 0$, so the sign is correct in Equation 9.12.

Solution of the Poisson–Boltzmann equation

We have reduced the problem of finding the counterion distribution outside a surface to solving Equation 9.11. This is a second-order differential equation, so we'll need to impose two boundary conditions to determine a unique solution. Furthermore, the equation itself contains a third constant c_0 , which requires a third condition to fix its value. The conditions are

- (i) The boundary condition at the surface (Equation 9.12),
- (ii) An analogous condition $d\bar{\psi}/dx = 0$ at infinity, because no charge is located there, and
- (iii) The convention that $\bar{\psi}(0) = 0$.

It's usually not easy to solve nonlinear differential equations like Equation 9.11. Still, in some special situations, we do get lucky. We need a function whose second derivative equals its exponential. We recall that the logarithm of a power of x has the property that both its derivative and its exponential are powers of x . We don't want $\bar{\psi}(x) = \ln x$, because that's divergent (equal to infinity) at the surface. Nevertheless,

a slight modification gives something promising: $\bar{\psi}(x) \stackrel{?}{=} B \ln(1 + (x/x_0))$. This trial solution has the feature that $\bar{\psi}(0) = 0$; it automatically satisfies conditions (ii–iii) regardless of what values we choose for B and x_0 .

We now check whether we can choose values for the constants B and x_0 in such a way that the proposed solution solves the Poisson–Boltzmann equation. Substituting $B \ln(1 + (x/x_0))$ into Equation 9.11, we indeed find that it works, provided we take $B = 2$ and $x_0 = 1/\sqrt{2\pi\ell_B c_0}$.

Next we must impose the boundary condition (Equation 9.12). Substituting into the trial solution, this condition says $2/x_0 = 4\pi\ell_B(\sigma_q/e)$. It may seem as though we have exhausted all our freedom to adjust the trial solution (when we chose values for B and x_0). But the Poisson–Boltzmann equation itself contains an unknown parameter, c_0 . You can check that taking this parameter to be $c_0 = 2\pi\ell_B(\sigma_q/e)^2$ ensures that our trial solution satisfies the boundary condition, and that then

$$\psi(x) = 2\frac{k_B T}{e} \ln(1 + (x/x_0)), \quad \text{where } x_0 = (2\pi\ell_B\sigma_q/e)^{-1}. \quad (9.13)$$

Your Turn 9C

Find the equilibrium concentration profile $c_+(x)$ away from the surface. Check your answer by calculating the total surface density of counterions, $\int_0^\infty dx c_+(x)$, and verifying that the whole system is electrically neutral.

The solution you just found is sometimes called the **Gouy–Chapman layer**; x_0 is called the Gouy–Chapman length. This solution is appropriate in the neighborhood of a flat, charged surface in pure water.⁴ Let’s extract some physical conclusions from the math.

First, we see from Your Turn 9C that, indeed, a diffuse layer forms, with thickness roughly x_0 . As argued physically in Section 9.2.1, the counterions are willing to pay some electrostatic potential energy in order to gain entropy. More precisely, the counterions pull some thermal energy from their environment to make this payment. They can do this because doing so lowers the entropic part of their free energy more than it raises the electrostatic part. If we could turn off thermal motion (that is, send $T \rightarrow 0$), the energy term would dominate and the layer would collapse. We see this mathematically from the observation that then the Bjerrum length would go to infinity and $x_0 \rightarrow 0$.

How much electrostatic energy must the counterions pay to dissociate from the planar surface? We can think of the layer as a planar sheet of charge hovering at a distance x_0 from the surface. When two sheets of charge are separated, we have a parallel-plate capacitor. Such a capacitor, with area Σ , stores electrostatic energy $\mathcal{E} = q_{\text{tot}}^2/(2C)$. Here q_{tot} is the total charge separated; for our case, it’s $\sigma_q\Sigma$. The capacitance of a parallel-plate capacitor is given by

$$C = \epsilon\Sigma/x_0. \quad (9.14)$$

⁴[\[T2\]](#) Or more realistically, a highly charged surface in a salt solution whose concentration is low enough; see Section 9.2.3’ (page 112).

Combining the preceding formulas gives an estimate for the density of stored electrostatic energy per unit area for an isolated surface in pure water:

$$\mathcal{E}/(\text{area}) \approx k_{\text{B}}T(\sigma_{\text{q}}/e). \quad (\text{electrostatic self-energy, no added salt}) \quad (9.15)$$

That makes sense: The environment is willing to give up about $k_{\text{B}}T$ of energy per counterion. This energy gets stored in forming the diffuse layer.

Is it a lot of energy? A fully dissociating bilayer membrane can have one unit of charge per lipid head group, or roughly $|\sigma_{\text{q}}/e| = 0.7 \text{ nm}^{-2}$. A spherical vesicle of radius $10 \mu\text{m}$ then carries stored free energy $\approx 4\pi(10 \mu\text{m})^2 \times (0.7/\text{nm}^2)k_{\text{B}}T_{\text{r}} \approx 10^9 k_{\text{B}}T_{\text{r}}$. It's a lot! We'll see how to harness this stored energy in Section 9.2.6.

9.2.4 Excess salt shrinks the electric double layer

For simplicity, the preceding calculations assumed that a dissociating surface was immersed in pure water. In real cells, however, the cytosol is an **electrolyte**, or salt solution. In this case, the density of counterions at infinity is not zero, and the counterions originally on the surface have less to gain entropically by escaping; so the diffuse charge layer will hug the surface more tightly than it does in Equation 9.13. That is,

$$\text{Increasing salt in the solution shrinks the diffuse layer.} \quad (9.16)$$

Problem 9.3 makes this expectation quantitative.

[\[T2\]](#) Section 9.2.3' (page 112) solves the Poisson–Boltzmann equation for a charged surface in a salt solution, arriving at the concept of the Debye screening length and making Equation 9.16 quantitative.

9.2.5 The repulsion of like-charged surfaces arises from compression of their ion clouds

Now that we know what it's like near a charged surface, we're ready to find the force between two such surfaces in solution. Figure 9.4b shows the geometry. One might be tempted to say, "Obviously, two negatively charged surfaces will repel." But wait: By symmetry, everything to the left of the central plane $x = 0$ (that is, the surface, together with its counterion cloud) is net electrically *neutral*, as is everything to the right. Thus, the electrostatic force that one side exerts on the other must equal zero! But electrostatic force is not the only kind of force in the problem. As the surfaces get closer than about twice their Gouy–Chapman length x_0 , their diffuse counterion clouds get squeezed; they resist that squeezing just as an ideal gas resists compression. Here are the details.

If we could turn off thermal motion, the mobile ions would collapse down to the surfaces, and there would be no net charge anywhere. That observation motivates us to look at *entropic* forces. Examining Figure 9.4b, we see that charged particles are required to be in the gap, by charge neutrality. That is, the concentration of a dissolved ion species is higher in the gap than in the bulk. In such a situation, we expect an *osmotic pressure* in the gap, proportional to the concentration difference times the absolute temperature. This fluid pressure is what physically pushes the two surfaces apart, not a literal electrostatic repulsion.

For simplicity, let's again suppose that the surrounding water has no added salt and, hence, no ions other than the counterions dissociated from the surface.⁵ This time we'll measure distance from the midplane between *two* surfaces, which are located at $x = \pm h$ (Figure 9.4b). We'll suppose that each surface has surface charge density $-\sigma_q$. We choose the constant in ψ so that $\psi(0) = 0$; hence the parameter $c_0 = c_+(0)$ is the unknown concentration of counterions at the midplane. $\psi(x)$ will then be symmetrical about the midplane, so our previous trial solution (Equation 9.13) won't work. Keeping the logarithm idea, though, this time we try $\bar{\psi}(x) = A \ln \cos(\beta x)$, where A and β are unknown constants. Certainly this trial solution is symmetrical and equals zero at the midplane, where $x = 0$.

The rest of the procedure is familiar. Substituting the trial solution into the Poisson–Boltzmann equation (Equation 9.11) gives $A = 2$ and $\beta = \sqrt{2\pi\ell_B c_0}$. The boundary condition at $x = -h$ is again Equation 9.12. Imposing the boundary conditions gives a condition fixing β :

$$4\pi\ell_B(\sigma_q/e) = 2\beta \tan(h\beta). \quad (9.17)$$

Given the surface charge density $-\sigma_q$, we solve Equation 9.17 for β as a function of the spacing $2h$; then the desired solution is

$$\bar{\psi}(x) = 2 \ln \cos(\beta x), \quad \text{or} \quad c_+(x) = c_0(\cos \beta x)^{-2}. \quad (9.18)$$

As expected, the charge density is greatest near the plates; the potential is maximum in the center.

By symmetry, the electric field at the midplane is zero, so an ion feels zero external force there. However, an ion that tries to diffuse out of the gap gets pulled back in, partially rectifying its Brownian motion and creating an high-pressure zone in the gap. The osmotic pressure difference equals $k_B T$ times the difference between c_0 and the concentration outside the gap (which is zero), so the repulsive force per unit area on the surfaces is given approximately by the osmotic pressure formula:⁶

$f/(\text{area}) = c_0 k_B T. \quad \text{repulsion of like-charged surfaces, no added salt}$

(9.19)

In this formula, $c_0 = \beta^2/(2\pi\ell_B)$ and $\beta(h, \sigma_q)$ is the solution of Equation 9.17. You can solve Equation 9.19 numerically (see Problem 9.4), but a graphical solution shows qualitatively that β increases as the plate separation decreases (Figure 9.6). Thus the repulsive pressure increases, too, as expected.

Note that the force just found is not simply proportional to the absolute temperature, because β has a complicated temperature dependence. This means that our pressure is not a purely entropic effect, but a mixed effect: The counterion layer reflects a *balance* between entropic and energetic imperatives. As remarked at the end of Section 9.2.3, the qualitative effect of adding salt to the solution is to tip this

⁵ [\[2\]](#) This is not as restrictive as it sounds. Even in the presence of salt, our result will be accurate if the surfaces are highly charged because in this case, the Gouy-Chapman length is less than the Debye screening length (see Section 9.2.3', page 112).

⁶ This formula is valid at low enough concentration c_0 .

Figure 9.6: [Mathematical functions.] **Graphical solution of Equation 9.17.** The sketch shows the function $2\pi\ell_B\sigma_q/(e\beta)$, as well as $\tan h\beta$ for two values of the plate separation $2h$. The value of β at the intersection of the rising and falling curves gives the desired solution. The figure shows that smaller plate separation gives a larger solution β_2 than does large separation (yielding β_1). Larger β in turn implies a larger ion concentration $c_0 = \beta^2/(2\pi\ell_B)$ at the midplane and larger repulsive pressure.

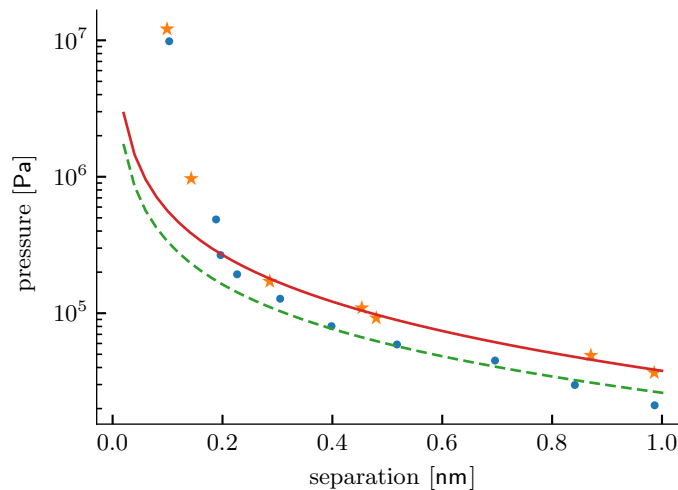
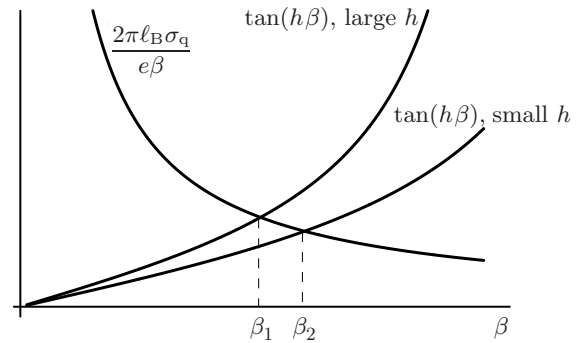


Figure 9.7: [Experimental data with fits.] **The repulsive pressure between two positively charged surfaces in water.** The surfaces were egg lecithin bilayers containing 5 mole% or 10 mole% phosphatidylglycerol (*circles* and *stars*, respectively). The curves show one-parameter fits of these data to the numerical solution of Equations 9.17 and 9.19. The fit parameter is the surface charge density σ_q . The *dashed line* shows the solution with one proton charge per 24 nm^2 ; the *solid line* corresponds to a higher charge density (see Problem 9.4). At separations below 2 nm, the surfaces begin to touch and other forces besides the electrostatic one appear. Beyond 2 nm, the purely electrostatic theory fits the data well, and the membrane with a larger density of charged lipids is found to have a larger effective charge density, as expected. [Data from Cowley et al., 1978; see Dataset 1.]

balance away from entropy, thereby shrinking the diffuse layers on the surfaces and *shortening the range* of the interaction.

This theory works (see Figure 9.7). You'll make a detailed comparison with experiment in Problem 9.4.

[T2] Section 9.2.5' (page 114) derives the electrostatic force directly as a derivative of the free energy.

9.2.6 Oppositely charged surfaces attract by counterion release

Now consider an encounter between surfaces of *opposite* charge (Figure 9.4c on page 104). Without working through the details, we can understand the attraction of such surfaces in solution qualitatively by using the ideas developed earlier. Again, as the surfaces approach from infinity, each presents a net charge density of *zero* to the other; there is no long-range force, unlike the constant attractive force between two such planar surfaces in air. Now, however, as the surfaces approach, they can shed counterion pairs without sacrificing the system's neutrality. The released counterions leave the gap altogether and hence gain entropy, thereby lowering the free energy and driving the surfaces together. If the charge densities are equal and opposite, the process proceeds until the surfaces are in tight contact, with no counterions left at all. In this case, there is no separation of charge, and no counterions remain in the gap. Thus, all the self-energy estimated in Equation 9.15 gets released. This energy is substantial: Electrostatic binding between surfaces of matching shape can be very strong.

FURTHER READING

Bioelectricity: Benedek & Villars, 2000; Nelson, 2014; Grodzinsky, 2011.

T₂ Track 2

9.1.2'

1. Frictional drag must generate *heat*. Indeed, it's well known that electrical resistance creates heat, for example, in your toaster. Each ion passed between the plates falls down a potential hill, losing potential energy $q \times \Delta\psi$. The total number of ions per time making the trip is I/q , so the power (energy per time) expended by the external battery is $\Delta\psi \times I$. Using Ohm's "law" gives the familiar formula: power = I^2R .

2. The conduction of electricity through a copper wire is also a diffusive transport process and also obeys an ohmic relation. But the charge carriers are electrons, not ions; and the nature of the collisions is quite different from that in salt solution. In fact, the electrons could pass perfectly freely through a perfect single crystal of copper; they only bounce off imperfections (or thermally induced distortions) in the crystal lattice. Figuring out this story required the invention of quantum theory.

T₂ Track 2

9.2.3'

The solution Equation 9.13 has a disturbing feature: The potential goes to infinity far from the surface! It's true that physical quantities like the electric field and concentration profile are well behaved (see Your Turn 9C), but still, this pathology hints that we have missed something. For one thing, no macromolecule is really an infinite plane. But a more important and interesting omission from our analysis is the fact that any real solution has at least some s ; the concentration c_∞ of salt in the surrounding water is never exactly zero.

Rather than introducing the unknown parameter c_0 and then going back to set it, this time we'll choose the constant in $\psi(x)$ so that $\psi \rightarrow 0$ far from the surface; then the Boltzmann distribution reads

$$c_+(x) = c_\infty e^{-e\psi(x)/k_B T} \quad \text{and} \quad c_-(x) = c_\infty e^{-(-e)\psi(x)/k_B T}$$

for the counterions and coions, respectively. The corresponding Poisson–Boltzmann equation is

$$\frac{d^2\bar{\psi}}{dx^2} = -\frac{1}{2}\lambda_D^{-2} [e^{-\bar{\psi}} - e^{\bar{\psi}}], \quad (9.20)$$

where again $\bar{\psi} = e\psi/k_B T$ and λ_D is defined as

$$\lambda_D \equiv (8\pi\ell_B c_\infty)^{-1/2}. \quad \text{Debye screening length} \quad (9.21)$$

In a solution of table salt, with $c = 0.1$ M, the screening length is about 1 nm.

The solutions to Equation 9.20 are not elementary functions (they're called elliptic functions), but once again, we get lucky for the case of an isolated surface.

Check that

$$\bar{\psi}(x) = -2 \ln \frac{1 + e^{-(x+x_*)/\lambda_D}}{1 - e^{-(x+x_*)/\lambda_D}} \quad (9.22)$$

solves the equation. In this formula, x_* is any constant. [Hint: It saves some writing to define a new variable, $\zeta \equiv e^{-(x+x_*)/\lambda_D}$, and rephrase the Poisson–Boltzmann equation in terms of ζ , not x .]

Before we can use Equation 9.22, we still need to impose the surface boundary condition. Equation 9.12 fixes x_* , via

$$e^{x_*/\lambda_D} = \frac{e}{2\pi\ell_B\lambda_D\sigma_q} \left(1 + \sqrt{1 + (2\pi\ell_B\lambda_D\sigma_q/e)^2}\right). \quad (9.23)$$

Suppose that we only want the answer at distances less than some fixed x_{\max} . Show that at low enough salt concentration (big enough λ_D), the solution Equation 9.22 becomes a constant plus our earlier result, Equation 9.13. How big must λ_D be?

We can now look at a more relevant limit for biology: This time, hold the salt concentration fixed and go out to large distances, where our earlier result (Equation 9.13) displayed its pathological behavior. For $x \gg \lambda_D$, Equation 9.22 reduces to

$$\bar{\psi} \rightarrow -(4e^{-x_*/\lambda_D})e^{-x/\lambda_D}. \quad (9.24)$$

That is,

The electric fields far outside a charged surface in an electrolyte are exponentially screened at distances greater than the Debye screening length λ_D . (9.25)

Idea 9.25 and Equation 9.21 confirm an earlier expectation: Increasing c_∞ decreases the screening length, shrinking the diffuse charge layer and hence shortening the effective range of the electrostatic interaction (Idea 9.16).

In the special case of weakly charged surfaces (σ_q is small), Equation 9.23 gives $e^{-x_*/\lambda_D} = \pi\ell_B\lambda_D\sigma_q/e$; so the potential simplifies to

$$\psi(x) = -\frac{\sigma_q\lambda_D}{\epsilon}e^{-x/\lambda_D}. \quad \text{potential outside a weakly charged surface} \quad (9.26)$$

The ratio of the actual prefactor in Equation 9.24 and the form appropriate for weakly charged surfaces is sometimes called **charge renormalization**: Any surface will, at great distances, look the same as a weakly charged surface, but with the “renormalized” charge density $\sigma_{q,R} = (4\epsilon/\lambda_D)e^{-x_*/\lambda_D}$. The true charge on the surface becomes apparent only when an incoming object penetrates into its strong-field region.

In the presence of added salt, the layer thickness no longer grows without limit as the layer charge gets smaller (as it did in the no-salt case, Equation 9.13); rather, it stops growing when it hits the Debye screening length. For weakly charged surfaces, then, the stored electrostatic energy is roughly that of a capacitor with gap spacing λ_D , not x_0 . Repeating the argument at the end of Section 9.2.3, we now find the stored energy per unit area to be

$$\mathcal{E}/(\text{area}) \approx k_B T \left(\frac{\sigma_q}{e}\right)^2 2\pi\lambda_D\ell_B. \quad \begin{array}{l} \text{(electrostatic energy with added salt,} \\ \text{weakly charged surface)} \end{array} \quad (9.27)$$

T₂ Track 2

9.2.5'

The crucial last step leading to Equation 9.19 may seem too slick. Can't we work out the force the same way we calculate any entropic force, by taking a derivative of the free energy? Absolutely. Let's compute the free energy \mathcal{F} of the system of counterions+surfaces, holding fixed the charge density $-\sigma_q$ on each surface but varying the separation $2h$ between the surfaces (see Figure 9.4b on page 104). Then the force between the surfaces will be $p\Sigma = -d\mathcal{F}/d(2h)$, where Σ is the surface area.

First we notice an important property of the Poisson–Boltzmann equation (Equation 9.11 (page 105)). Multiplying both sides by $d\bar{\psi}/dx$, we can rewrite the equation as

$$\frac{d}{dx} \left[\left(\frac{d\bar{\psi}}{dx} \right)^2 \right] = 8\pi\ell_B \frac{dc_+}{dx}.$$

Integrating this equation gives a simpler, first-order equation:

$$\left(\frac{d\bar{\psi}}{dx} \right)^2 = 8\pi\ell_B (c_+ - c_0). \quad (9.28)$$

To fix the constant of integration, we noted that the electric field is zero at the midplane, and $c_+(0) = c_0$ there.

Next we need the free energy density per unit area in the gap. The free energy density of an inhomogeneous ideal gas (or solution) is $c(\vec{r}) (U(\vec{r}) + k_B T \ln(c(\vec{r})/c_*)$). The free energy for our problem is the integral of this quantity, plus the electrostatic energy⁷ of the two negatively charged plates at $x = \pm h$:

$$\mathcal{F}/(k_B T \times \text{area}) = -\frac{1}{2} \frac{\sigma_q}{e} (\bar{\psi}(h) + \bar{\psi}(-h)) + \int_{-h}^h dx \left[c_+ \ln \frac{c_+}{c_*} + \frac{1}{2} c_+ \bar{\psi} \right].$$

In this formula, c_* is a constant whose value will drop out of our final answer.

We simplify our expression by first noting that $\ln(c_+/c_*) = \ln(c_0/c_*) - \bar{\psi}$, so the terms in square brackets are $c_+ \ln(c_0/c_*) - \frac{1}{2} c_+ \bar{\psi}$. The first of these terms is a constant times c_+ , so its integral is $2(\sigma_q/e) \ln(c_0/c_*)$. To simplify the second term, use the Poisson–Boltzmann equation to write $c_+ = -(4\pi\ell_B)^{-1} (d^2\bar{\psi}/dx^2)$. Next integrate by parts, obtaining

$$\mathcal{F}/(k_B T \times \text{area}) = 2 \frac{\sigma_q}{e} \left[\ln \frac{c_0}{c_*} - \frac{1}{2} \bar{\psi}(h) \right] + \frac{1}{8\pi\ell_B} \frac{d\bar{\psi}}{dx} \bar{\psi} \Big|_{-h}^h - \frac{1}{8\pi\ell_B} \int_{-h}^h dx \left(\frac{d\bar{\psi}}{dx} \right)^2.$$

We evaluate the boundary terms by using Equation 9.12 (page 106) at $x = -h$ and its analog on the other surface; they equal $-(\sigma_q/e) \bar{\psi}(h)$.

To do the remaining integral, recall Equation 9.28: it's $-\int_{-h}^h dx (c_+ - c_0)$, or $2(hc_0 - (\sigma_q/e))$. Combining these results gives

$$\mathcal{F}/(k_B T \times \text{area}) = 2hc_0 + 2 \frac{\sigma_q}{e} \left(\ln \frac{c_0}{c_*} - \bar{\psi}(h) - 1 \right) = \text{const} + 2hc_0 + 2 \frac{\sigma_q}{e} \ln \frac{c_+(h)}{c_*}.$$

The concentration at the wall can again be found from Equations 9.28 and 9.12: $c_+(h) = c_0 + (8\pi\ell_B)^{-1} (d\bar{\psi}/dx)^2 = c_0 + 2\pi\ell_B (\sigma_q/e)^2$.

⁷Notice that adding any constant to $\bar{\psi}$ leaves this formula unchanged, because the integral $\int c_+ dx = 2\sigma_q/e$ is a constant, by charge neutrality. To understand the reason for the factor $\frac{1}{2}$ in the first and last terms, think about two point charges q_1 and q_2 . Their potential energy at separation r is $q_1 q_2 / (4\pi\epsilon r)$ (plus a constant). This is *one half* of the sum $q_1 \psi_2(r_1) + q_2 \psi_1(r_2)$. (The same factor of $\frac{1}{2}$ also appeared in the electrostatic self-energy Example on page 100.)

A few abbreviations will make for shorter formulas. Let $\gamma = 2\pi\ell_B\sigma_q/e$ and $u = \beta h$, where $\beta = \sqrt{2\pi\ell_B c_0}$ as before. Then u and β depend on the gap spacing, whereas γ does not. With these abbreviations,

$$\mathcal{F}/(k_B T \times \text{area}) = 2hc_0 + \frac{\gamma}{\pi\ell_B} \ln \frac{c_0 + \gamma^2/(2\pi\ell_B)}{c_*}.$$

We want to compute the derivative of this expression with respect to the gap spacing, holding σ_q (and hence γ) fixed. We find

$$\frac{p}{k_B T} = -\frac{1}{k_B T} \frac{d(\mathcal{F}/(k_B T \times \text{area}))}{d(2h)} = -c_0 - \left(h + \frac{\gamma}{2\pi\ell_B c_0 + \gamma^2} \right) \frac{dc_0}{dh}.$$

In the last term, we need

$$\frac{dc_0}{dh} = \frac{d}{dh} \left(\frac{u^2}{h^2 2\pi\ell_B} \right) = \frac{u}{\pi\ell_B h^3} \left(h \frac{du}{dh} - u \right).$$

To find du/dh , we write the boundary condition (Equation 9.17 (page 109)) as $\gamma h = u \tan u$ and differentiate to find

$$\frac{du}{dh} = \frac{\gamma}{\tan u + u \sec^2 u} = \frac{\gamma u}{h\gamma + u^2 + (h\gamma)^2}.$$

This has gone far enough. In Problem 9.5, you'll finish the calculation to get a direct derivation of Equation 9.19. For a deeper derivation from thermodynamics, see Israelachvili, 1992, §12.7.

PROBLEMS

9.1 $\boxed{T_2}$ Debye–Hückel I

- Calculate the Debye screening length for a 100 mM solution of sodium chloride. That is, the concentration of Na^+ ions is 0.1 mole per liter.
- Calculate the Debye screening length for a salt solution whose ions are not necessarily monovalent (singly charged). Do this by writing the appropriate Poisson–Boltzmann equation, linearizing it, and collecting terms.
- Evaluate your answer for a 100 mM solution of magnesium chloride. That is, the concentration of Mg^{2+} ions is 0.1 mole per liter.

9.2 Debye–Hückel II

Context: The main text claimed that electrostatic interactions in solution have a number of features that make them well suited to implement the remarkable specificity of interactions between biomacromolecules. In this problem you explore the ranges of both the overall attraction due to total net charge, and also of the pattern-dependent part of the attraction.

Setup: Consider a surface that is the infinite xy plane. Suppose that the electric field inside the surface is everywhere zero, so that the potential gradient at the surface reflects the surface charge density. But unlike the discussion in class, suppose that the fixed charge distribution on the surface is a constant plus a “checkerboard” component, i.e. that

$$\left. \frac{\partial \psi}{\partial z} \right|_{z=0} = A + B \sin(kx) \sin(ky)$$

Suppose that the surface is immersed in a salt solution with Debye screening length λ . Suppose that A and B are both small enough to justify linearizing the Poisson–Boltzmann equation.

Do: Find $\psi(x, y, z)$. Comment on the z dependence of your solution in light of the above remarks.

9.3 Weak-charge limit

Section 9.2.3 considered an ionizable surface immersed in pure water. Thus, the surface dissociated into a negative plane and a cloud of positive counterions. Real cells, however, are bathed in a solution of salt, among other things; there is an external reservoir of *both* counterions and negative coions. Section 9.2.3' (page 112) gave a solution for this case, but the math was complicated; here is a simpler, approximate treatment.

Instead of solving Equation 9.20 exactly, consider the case where the surface’s charge density is small. Then the potential $\psi(0)$ at the surface will not be very different from the value at infinity, which we took to be zero. (More precisely, the dimensionless combination $\bar{\psi}$ is everywhere much smaller than 1.) Approximate the right-hand side of Equation 9.20 by the first two terms of its series expansion in powers of $\bar{\psi}$. The resulting approximate equation is easy to solve. Solve it, and give an interpretation to the quantity λ_D defined in Equation 9.21.

9.4 Charged surfaces

- Use some numerical software to solve Equation 9.17 for β as a function of plate separation $2h$ for fixed charge density σ_q . For concreteness, take σ_q to equal $e/(20 \text{ nm}^2)$. Now convert your answer into a force by using Equation 9.19 and compare your answer qualitatively with Figure 9.7.
- Obtain Dataset 1. Repeat (a) with other values of σ_q to find the one that best fits the upper set of points in the figure at separation greater than 2 nm . If this surface were fully dissociated, it would have one electron charge per 7 nm^2 . Is it fully dissociated?

9.5 Direct calculation of a surface force

Finish the derivation of Section 9.2.5' (page 114). The goal is to establish Equation 9.19.

9.6 Counterions in cylindrical geometry

Section 9.2.3 discussed the counterion distribution for a planar, charged surface. The text concluded that the counterions do not run away to infinity; that is, there is a nonzero concentration of ions near the surface.

One way to understand this result is to consider a single ion (of charge $e > 0$) near a surface with charge per unit area $\sigma_q < 0$. Suppose that the ion is initially confined to a distance a from the surface. If the ion is now allowed to explore a larger distance R from the surface, then the increase in its entropy is $k_B \ln(R/a)$. However, the electrostatic energy cost for the ion to travel out to a distance R is $e(R-a)\sigma_q/\epsilon$. The change in free energy is thus approximately $\Delta\mathcal{F} \approx e(R-a)\sigma_q/\epsilon - k_B T \ln(R/a)$, which increases as R gets very large. Therefore, to minimize the free energy, the ion does not run away to infinity but remains near the surface.

- Using a similar argument, determine whether or not the counterions will run away to infinity for an infinite-length charged cylinder of radius b and charge per unit length κ .
- Apply your result to the case of DNA, with two negatively charged phosphate groups for every basepair.

9.7 Counterion cloud

If you haven't done Problem 9.3, look at it before attempting this problem.

Consider a spherical macromolecule of charge $q = ze$ and radius a in a solution containing a monovalent salt, such as sodium chloride. As discussed in Problem 9.3, in the limit that the potential satisfies $|\psi(r)| \ll k_B T/e$, you may approximate the Poisson–Boltzmann equation in its linearized (“Debye–Hückel”) form. In spherical coordinates, the resulting equation is

$$\frac{1}{r} \frac{d^2(r\psi(r))}{dr^2} = \frac{1}{\lambda_D^2} \psi(r),$$

where λ_D is the Debye length.

- Justify the following boundary conditions:

$$\psi(r) \rightarrow 0 \text{ as } r \rightarrow \infty, \quad - \left. \frac{d\psi}{dr} \right|_{r=a} = E_r(\text{surface}) = \frac{q}{4\pi\epsilon a^2}.$$

- Find $\psi(r)$ in terms of λ_D , a , and q .

c. The charge density from salt ions is given by

$$\rho_q(r) = -\epsilon \frac{1}{r} \frac{d^2(r\psi(r))}{dr^2}.$$

Using your result for $\psi(r)$ from (b), show explicitly that the integral of this charge density is equal to $-q$.

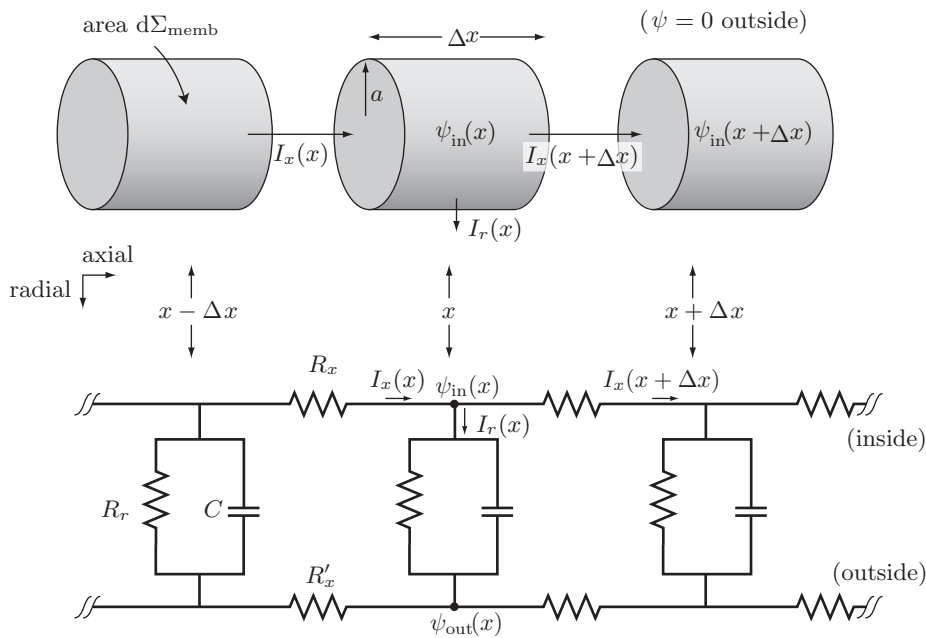
- d. Imagine placing the charge q on the surface of the spherical macromolecule by successive increments dq' , starting from $q = 0$. By integrating the work required to bring the charge up to q , find the total potential energy of the charged macromolecule and its neutralizing cloud.
- e. The solubility of proteins in dilute salt solution generally increases with increasing ionic strength of the solution. Use your result from (c) to explain this effect qualitatively.

9.8 [Not ready yet.]

9.9 [Not ready yet.]

CHAPTER 10

Cable Equation



10.1 CABLES AND THEIR EQUATIONS

10.1.1 The ill-fated transatlantic cable

By 1854, the first industrial revolution (steam power) had already transformed the world, and the second one (electric generation, motors, lights and related technology) was underway. But in at least one sense, the world remained unimaginably primitive: It still took weeks for any information to pass between Europe and America. The telegraph, by then a decade old, had eliminated communication barriers within continents, but between them, the only method of communication was by ship. In that year, a retired industrialist named Cyrus West Field decided to rectify this unsatisfactory situation. How hard could it be, he asked? One could simply string a cable across the narrowest part of the Atlantic ocean. With the growing economic significance of the United States, the first corporation to accomplish this simple task could reap enormous profits.

Field was ready to supply some of the needed capital investment, and he had the connections to bring in others like himself. But he also had the foresight to engage William Thomson, the future Lord Kelvin and already a noted expert on electricity.

Thomson took the assignment, but he saw some clouds on the horizon: Existing, but shorter, undersea cables in the Mediterranean were not behaving as expected. When electric current was poured in one end of such cables, a lot of it. . . disappeared. Equally bad, crisp on/off telegraph signals sent in one end arrived blurry at the other end (to the extent that they arrived at all).

Undersea cables had a “coaxial” structure. The one eventually laid across the Atlantic contained seven strands of a good conductor (copper) down the middle, surrounded by insulating substances (gutta-percha and tarred hemp), and then a third layer of iron strands, similar to those used in suspension-bridge cables. The iron was a poor conductor of return current; its main job was to supply strength, so that the entire cable could withstand undersea currents, as well as the stress from its own weight as it was reeled out from a giant spool on the ship initially laying it. The overall diameter was 1.8 cm. I actually possess a small chunk of the original cable:



Developing older ideas from Michael Faraday, Thomson realized that part of the transmission problem must be the *capacitance* of existing cables: Instead of passing all the way through the cable and out the other end, some charge could simply stop in the middle, paying a finite energy cost to create an electric field across the thin insulating layer. Eventually that charge could leak across the finite resistance of the insulating layer, never arriving at the other end at all. This loss mechanism was unexpected because for overland transmission cables it was negligible: There the standard design was a pair of wires separated by a meter of air, with negligible capacitance per unit length (and enormous leak resistance per length).

Thomson therefore recommended reengineering the cables with a much thicker insulation layer than had originally been planned. Unfortunately, the thin cable had already been ordered and paid for. Field took the time-honored approach of finding another engineer willing to reassure him that everything would be fine. The new chief engineer in turn pulled the elderly Faraday out of retirement for a public meeting to reassure the investors, after first misleading Faraday about some recent experimental results. Cable-laying began in 1857.

The first attempt ended in failure with the cable snapping in water too deep to retrieve the lost end. Another attempt the following year involved two ships. They planned to meet in the middle of the Atlantic, splice their respective cables together, then head for Ireland and Newfoundland respectively, paying out cable as they went.

The ships immediately encountered one the worst storms recorded in the North Atlantic. The ships were damaged; the cable snapped more than once and had to be

spliced; one ship was attacked by an angry whale. Nevertheless, ultimately an intact, if thin, cable at last stretched across the ocean. Wild celebrations ensued before the device had even been tested, including a torchlight procession that set fire to New York's City Hall.

Most of the initial telegraph traffic on the cable consisted of “Send more slowly,” “Repeat,” or simply “What?” It took sixteen hours to transmit the Queen's 99-word congratulation to the President, and thirty hours for the equally brief reply. Desperate to get a stronger signal, the lead engineer increased the voltage supplied to the cable, until the insulation broke down somewhere in the middle of the ocean, turning the entire cable into worthless undersea trash. The investors lost their money. A parliamentary inquiry was mounted to see who should be blamed. Eventually a rumor spread that the entire project had been a massive hoax. Not until 1866 (after another snapped-cable fiasco), did a successful cable, following Thomson's original advice, come into operation.

10.1.2 Setup

Thomson had understood both the loss and the spread of signals before the first cable was even attempted. He found his way through the physical problem by an approach that is routine today but astonishing in the mid-19th century: He set up the problem mathematically, then noticed that it involved the *same equation* as a problem that seemed physically to be completely different. The same equation must have the same solutions, so Thomson benefited at once from extensive work that had already been done on the other problem. Let's see how that worked.

We'll make some idealizations. Imagine a cable consisting of a solid cylindrical core of ohmic conductor (such as copper), surrounded by a sheath of partially insulating material, which in turn is surrounded by a perfect conductor. That last assumption is purely for mathematical convenience; relax it, the equations just get a bit longer.

Finally, we continue to work in the quasi-static regime, where we may neglect the back-reaction of any magnetic fields on electric fields and currents, and where we may assume that there is no net charge buildup on macroscopic length scales. That last statement is often called one of the “Kirchhoff laws,” but I would rather not refer to an approximation as a “law.”

To give things symbols, let a be the radius of the central core and κ its conductivity. Let g be the conductance per unit area of the sheath. It's positive and has units¹ $\Omega^{-1}\text{m}^{-2}$. Also let \mathcal{C} denote the capacitance per area.

If the system is isolated, it eventually comes to the boring state with potential everywhere equal to a constant. We are interested in transient solutions that have not yet arrived at that state, so we need to find and solve some equation.

¹Some authors use the synonym **siemens** (symbol S) for inverse ohm; an older synonym is the mho (symbol \mathcal{O}). We won't use either notation, instead writing Ω^{-1} . Note that *conductance* per area has units different from those of the *conductivity*, κ , of a bulk material: The latter has units $\text{m}^{-1}\Omega^{-1}$.

10.1.3 Discretize

The capacitance and the resistances are all continuously distributed along our cable. However, things will look more familiar if we imagine dividing the cable into slugs of length Δx and treating them as discrete elements (see the figure). This is not an approximation because later we'll take the limit $\Delta x \rightarrow 0$.

What *is* an approximation is that we'll suppose the central conductor to be good enough, and thin enough compared to longitudinal length scales we will consider, that potential may be taken to be uniform throughout each cross-section. Potential may of course jump across the sheath, and it may vary along the length of the conductor.

Each element has an axial resistance for the inner material, R_x . We are taking the corresponding axial resistance for the outer material to be $R'_x = 0$, so right away we learn that the exterior potential is a constant, which we may take to be $\psi_{\text{out}} = 0$. Another "radial" resistance, R_r , impedes current passage through the sheath. However, charge can instead approach the sheath and pile up against it, as long as an equal charge leaves the other side. In this way, no macroscopic net charge separation occurs, although there is some separation across the sheath. We are accounting for the electrostatic cost of this local separation via the capacitance. The combined effect of charge passage and charge pileup is symbolized in the figure by a resistor R_r and a capacitor C in parallel for each slug.

Elsewhere, currents must balance. Thus, for example, the three-way junctions at the top must each have zero net charge flowing into them:

$$I_x(t, x) - I_x(t, x + \Delta x) = \psi_{\text{in}}(t, x)/R_r + C \frac{\partial \psi_{\text{in}}}{\partial t}.$$

Similarly, charge entering each resistor on the top must all leave it: I_x is the same on both sides. Finally, the hypothesis of ohmic behavior in the core says

$$\psi_{\text{in}}(x - \Delta x) - \psi_{\text{in}}(x) = I_x(x)R_x.$$

Next, express the discrete element properties in terms of quantities introduced earlier:

$$C = \mathcal{C}\Sigma_{\text{memb}} \quad \text{and} \quad R_r = 1/(g\Sigma_{\text{memb}}), \quad \text{where} \quad \Sigma_{\text{memb}} = 2\pi a\Delta x$$

$$R_x = \Delta x/(\kappa\pi a^2).$$

10.1.4 Cable equation

You should combine the preceding formulas and take the continuum limit, obtaining

$$\kappa\pi a^2 \frac{\partial^2 \psi_{\text{in}}}{\partial x^2} = 2\pi a(g\psi_{\text{in}} + \mathcal{C} \frac{\partial \psi_{\text{in}}}{\partial t}). \quad (10.1)$$

Define the **space constant** and **time constant** as

$$\lambda_{\text{cable}} \equiv \sqrt{a\kappa/2g}, \quad \tau_{\text{cable}} \equiv \mathcal{C}/g. \quad (10.2)$$

(Check that these expressions have the units of length and of time, respectively.) These abbreviations yield

$$(\lambda_{\text{cable}})^2 \frac{\partial^2 \psi_{\text{in}}}{\partial x^2} - \tau_{\text{cable}} \frac{\partial \psi_{\text{in}}}{\partial t} = \psi_{\text{in}}. \quad \text{linear cable equation} \quad (10.3)$$

Change variables from ψ_{in} to $w(x, t) \equiv e^{t/\tau_{\text{cable}}}\psi_{\text{in}}(x, t)$. Then the linear cable equation becomes

$$\frac{(\lambda_{\text{cable}})^2}{\tau_{\text{cable}}}\frac{\partial^2 w}{\partial x^2} = \frac{\partial w}{\partial t}.$$

Kelvin's great insight was to recognize this as mathematically identical to the diffusion equation, or to Fourier's heat equation. The analog of the diffusion constant is $(\lambda_{\text{cable}})^2/\tau_{\text{cable}} = \kappa a/(2\mathcal{C})$, so we see that *a good cable must have small capacitance*.²

We already know some solutions to the diffusion equation. For example, the response of our cable to a localized injection of current is

$$\psi_{\text{in}}(t, x) = \text{const} \times e^{-t/\tau_{\text{cable}}}t^{-1/2}e^{-x^2/(4t(\lambda_{\text{cable}})^2/\tau_{\text{cable}})}. \quad \text{passive-spread solution} \quad (10.4)$$

You should confirm that this function does solve Equation 10.3. It's a gaussian profile at any instant of time, which initially widens out fast, then slows down, all the while dying off exponentially in time. *Try this:* Imagine sitting at a fixed location x_* and observing the time course of the potential disturbance. At what time does the disturbance reach its peak? How does the peak strength vary as a function of x_* ? Maybe you should also get a computer to draw $\psi_{\text{in}}(t, x_*)$ for various x_* .

In fact, *the linear cable equation has no traveling wave solutions*. *Try this:* substitute a trial solution of the form $\psi_{\text{in}}(t, x) = f(x - \vartheta t)$, into Equation 10.3, where ϑ is a constant, the speed of the proposed traveling wave. Is there any value of ϑ that yields a physical solution?

In short, our passive cable suffers from dispersion, even if there is no leak conductance (g dropped out of the expression for the diffusion constant).

10.2 NEURONS

10.2.1 A new phenomenon

People talk casually about the brain as a “computer” and its neurons as “wires,” but a little thought shows they must be very different from ordinary wires. A coaxial cable brings Internet into your apartment via signals that move at around $2 \cdot 10^8$ m/s. Your nerves carry signals that move at around 10–20 m/s. They are also poorly insulated and surrounded by a conductive medium!

A neuron has a long projection, its **axon**, that is a “cable” of the sort we are considering: It is a tube of conductor (salt water) surrounded by a partially insulating layer (cell membrane), which is surrounded by another conductor (salt water).³ So we may expect that electrochemical disturbances will also spread diffusively along an axon.

²That result led Kelvin to propose redesigning the cable with thicker insulation (smaller \mathcal{C}) and thicker central conductor (bigger a), but the Suits declared it was too late and too expensive to change the design.

³You may object that axons are filled with lots of other machinery, including microtubules. *Amazingly*, experiments have been done in which all those contents are emptied out of the axon and it is refilled with just a salt solution. All the phenomena we will discuss (passive spread and the action potential) behave identically with these gutted axons as they do in living cells.

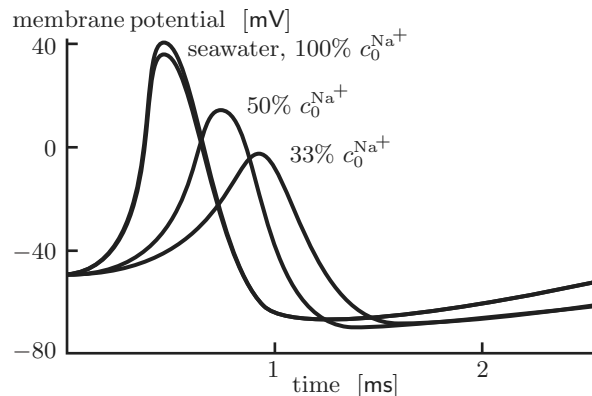


Figure 10.1: [Experimental data.] **The role of sodium in the conduction of an action potential.** One of the top traces was taken on a squid axon in normal seawater before exposure to low sodium. In the middle trace, external sodium was reduced to one-half that in seawater, and in the bottom trace, to one-third. (The other top trace was taken after normal seawater was restored to the exterior bath.) The data show that the peak of the action potential tracks the sodium Nernst potential across the membrane, an observation supporting the idea that the action potential is a sudden increase in the axon membrane's sodium conductance. [Data from Hodgkin & Katz, 1949.]

For some nerve cells, that's good enough (e.g. photoreceptors in the eye). They are short, and over a few micrometers diffusive spread is not a problem. Longer nerve axons also exhibit passive-spread behavior when stimulated with very small disturbances. But that wouldn't be very useful for, say, the axons that start in your spinal cord and end a meter away in your foot! In fact, above a threshold of stimulation, axons transmit a traveling impulse that moves unchanged in form, at constant speed. We just proved that is impossible, so we have some work to do.

One big difference, we will see, is that a resting neuron carries a distributed source of free energy that can continuously regenerate an impulse as it travels, counteracting dissipative (ohmic) loss. Indeed, ion imbalance is the one crucial thing needed to get action potentials in experiments where the interior of the axon is evacuated and replaced by salt solution; specifically, *excess exterior sodium ions* (Figure 10.1).

The other big difference will be a crucial *non-ohmic* behavior of the leakage conductance, which can continually *reshape* the impulse as it travels, counteracting dispersion. How these new elements conspire to generate a nonlinear traveling wave solution is a remarkable story. If you want a metaphor, think of a burning fuse in some Hollywood blockbuster: Stored chemical energy is released in a controlled way, leading to a flame front that self-regulates to move at constant speed.

10.2.2 Ions in and out of equilibrium

Salt water conducts electric current by the movement of *ions*, not electrons. One key difference with ordinary conduction in metals is that there are *several types of ions*, in contrast to just one charge carrier (electrons) in a metal. Each ion species ℓ has its own concentration c_ℓ . For electrons in a metal, the neutralizing atomic nuclei are fixed in space. So in the quasistatic approximation, charge neutrality implies that, although the electrons are mobile, their density cannot vary. In contrast, in aqueous

solution charge neutrality does not prohibit a change in one ion's concentration, as long as the other species make compensating changes.⁴

The membrane leakage conductances per area for each ion species, g_ℓ , can all have *different* values, because ions are passed through exquisitely engineered protein complexes embedded in the cell membrane; far from being just tubes, each of these **ion channels** is sculpted in a way that selects for a particular ion or class of ions. (However, we will make the approximation that ions of each species all have the same mobility in bulk solution, leading to an overall conductivity κ that doesn't care which species is moving.)

Thus, the net charge flow (current) through a channel due to species ℓ is the conductivity times the sum of two driving forces:

- . There is an electrostatic force proportional to the difference of electric potentials on either side of the membrane times the charge on species ℓ .
- . There is also a *thermodynamic* force, involving the difference of *concentrations*. Just like the air in a balloon, ions will “want” to escape from the side where their concentration is greater.

More precisely, ions of species ℓ will seek an equilibrium concentration c_ℓ which reflects a Boltzmann distribution of probability to be on either side:

$$\frac{c_{\ell,\text{in}}}{c_{\ell,\text{out}}} = \frac{c_0 e^{-q_\ell \psi_{\text{in}}/k_B T}}{c_0 e^{-q_\ell \psi_{\text{out}}/k_B T}} = e^{-q_\ell \Delta\psi/k_B T}. \quad (\text{equilibrium})$$

That is, equilibrium with given concentrations requires a potential drop called the **Nernst potential** for species ℓ :

$$\psi_\ell^{\text{Nernst}} = -\frac{k_B T}{q_\ell} \ln(c_{\ell,\text{in}}/c_{\ell,\text{out}}).$$

But beware: The Nernst potential may not be equal to the actual potential drop. If they disagree, that just means that species ℓ is out of equilibrium, and hence will flow. We may guess that, at least for small deviations from equilibrium, the resulting ion flow would give rise to a charge flux of the form

$$j_{r,\ell} = (\Delta\psi - \psi_\ell^{\text{Nernst}})g_\ell. \quad \text{ohmic conductance hypothesis} \quad (10.5)$$

This formula gives the radial charge flux contribution from species ℓ , with the sign convention that positive means net charge leaving the axon (radially outward). The potential drop is defined as $\Delta\psi = \psi_{\text{in}} - \psi_{\text{out}}$. The conductance per area g_ℓ involves the permeability of a channel, the density of channels in the membrane, and the square of the charge carried by species ℓ ; it is a positive quantity.

The two terms in Equation 10.5 mean that there can be net flow of ions *against* the electrostatic gradient, if the “pressure” term outweighs the “field” term

Here are some typical values for three ion species that are relevant in the squid “giant” axon (so called because it can be up to a millimeter in diameter—not because it comes from giant squid⁵):

⁴A similar remark applies in plasma physics, and indeed there are some phenomena in common between that situation and aqueous solution. For example, both exhibit charge screening.

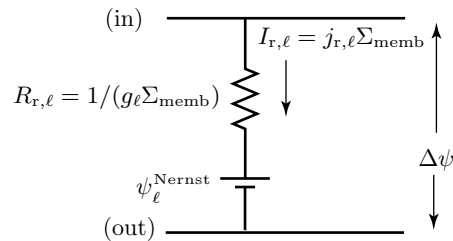
⁵Nor from a superconducting quantum interference device!

ion	charge q_ℓ	interior		exterior		Nernst potential $\psi_\ell^{\text{Nernst}}$, mV
		$c_{\ell,\text{in}}$, mM	relation	$c_{\ell,\text{out}}$, mM		
K^+	$+e$	400	$>$	20		-75
Na^+	$+e$	50	$<$	440		+54
Cl^-	$-e$	52	$<$	560		-59

The salient feature of this table is the middle line: There is no value of $\Delta\psi$ that even approximately satisfies all three of these ion species. In fact, resting neurons are polarized with $\Delta\psi$ negative. Sodium is way out of equilibrium under those conditions.

In its resting state, the neuron maintains these nonequilibrium concentrations by continuously pumping ions across its membrane, but we don't need to worry about that. Even when we shut down a living cell's metabolism, it still preserves the preceding values of ion concentrations for several minutes, because the interior and exterior are large reservoirs. During that time, the neuron's axon conducts action potentials and otherwise behaves electrically like a living cell. All the pumps do is to set up and maintain the conditions given in the table.

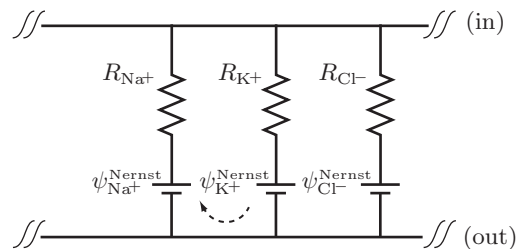
We can summarize the preceding discussion with a little circuit diagram representing the contribution of one species to the current through a patch of membrane:



Placing the resistor and battery symbols in series, as shown, encodes the fact that current is driven by the difference between actual $\Delta\psi$ and the Nernst potential for this species.

10.2.3 Linear cable equation for an axon

Let's see how the preceding considerations affect signal propagation along an axon. As far as charge is concerned, each ion species makes its own contribution to current, so we can simply represent the driving forces and conductances by three modules in parallel:⁶



⁶The circuit diagram also correctly represents the fact that all ion species share the same exterior and interior values of the electrostatic potential at any position x .

As before, we simplify by supposing the outer bulk resistance equals zero. Thus, again $\psi_{\text{out}} \equiv 0$ and $\Delta\psi = \psi_{\text{in}}$. The dashed arrow reminds us that, although the “resting” membrane transmits no net current, still individual ion species are flowing.

Our hope was that the latent energy source, symbolized by the battery symbols in the diagram, could regenerate a disturbance as it travels along the axon.

Your Turn 10A

To investigate, first show that the entire preceding diagram can be equivalently replaced by a *single* resistor/battery unit, and find formulas for the effective overall battery potential ψ^0 and radial resistance $R_{r,\text{tot}}$. Explain the sense in which “the ion species with the biggest conductance gets the biggest vote when determining the membrane potential.”

A typical magnitude for the overall conductance per area of a resting squid axon membrane is about $5 \text{ m}^{-2} \Omega^{-1}$.

With this insight, we see that the axon’s overall diagram is almost exactly the same as the one at the start of this document, just with the addition of a battery in each module. Thus, the needed modification to the linear cable equation just amounts to introducing ψ^0 :

$$\kappa\pi a^2 \frac{\partial^2 \psi_{\text{in}}}{\partial x^2} = 2\pi a (g_{\text{tot}}(\psi_{\text{in}} - \psi^0) + \mathcal{C} \frac{\partial \psi_{\text{in}}}{\partial t}).$$

We can then eliminate the battery altogether if we change variables to $v = \psi_{\text{in}} - \psi^0$:

$$(\lambda_{\text{cable}})^2 \frac{\partial^2 v}{\partial x^2} - \tau_{\text{cable}} \frac{\partial v}{\partial t} = v. \quad (10.6)$$

Here the space constant and time constant are defined as before.

Some numerical values are revealing: Taking illustrative values of $a = 0.5 \text{ mm}$, $g_{\text{tot}} \approx 5 \text{ m}^{-2} \Omega^{-1}$, $\mathcal{C} \approx 1 \mu\text{F cm}^{-2}$, and $\kappa \approx 3 \Omega^{-1} \text{ m}^{-1}$ yields

$$\lambda_{\text{cable}} \approx 12 \text{ mm} , \quad \tau_{\text{cable}} \approx 2 \text{ ms}. \quad (10.7)$$

We seem to have hit an impasse. All that stored electrochemical energy seems unable to affect nerve impulses—it dropped out of the equation, which has the same disappointing solutions as before! Indeed, experimentally that’s the *observed behavior* for weak disturbances. For example, when we inject a subthreshold charge into the axon, we do find passive spread, which in this context is also called “electrotonus.” For the more spectacular action potential, we must look for another physical idea.

10.2.4 Day breaks

The resting membrane potential in squid axon was found to be about -50 mV . This is not far from the Nernst potential of potassium ions given in the earlier table. That coincidence suggests one possible interpretation: The conductance for potassium ions is much bigger than that for sodium ions.

When an action potential travels along the membrane, the membrane potentially locally and temporarily shoots up to something more like $+40 \text{ mV}$. This is not so

different from the Nernst potential of *sodium*, suggesting another interpretation:

The conductance for sodium ions briefly overtakes that for potassium, and a resulting ion flow tries to establish that new value of membrane potential drop as a new steady state. (10.8)

In fact, Hodgkin and B. Katz had previously found that during an action potential, conductances change from their “resting” values:

$$g_{K^+} \approx 25g_{Na^+} \approx 2g_{Cl^-}. \quad (\text{resting}) \quad (10.9)$$

to

$$g_{K^+} \approx 0.05g_{Na^+} \approx 2g_{Cl^-}. \quad (\text{at the action potential peak}) \quad (10.10)$$

What could change the ion conductance of sodium in just the right way? Hodgkin and Huxley realized that even a few millivolts across a nanometer-thickness membrane amounts to a huge electric field, which could *tug on* charged residues in the proteins making up an ion channel. With the appropriate arrangement, a reversal in the direction of that tugging could open a channel that was normally closed! Hodgkin and Huxley therefore proposed that *the conductance of the membrane to specific ions is itself voltage-dependent*: We must use a *function* of potential $g_{Na^+}(\Delta\psi)$ in the cable equation. The result is now nonlinear in ψ . Interesting things can happen with nonlinearity.

In particular, suppose that depolarization (making $\Delta\psi$ less negative than usual) causes sodium channels to open. Then a localized electrical disturbance that depolarizes a patch of membrane lets sodium ions rush in, which *further depolarizes* that patch. The disturbance can then spread to a neighboring region, where the same sequence is repeated. Thus, the axon can be in a stable, steady state, yet poised to release energy. A disturbance at one end, like lighting a fuse, can lead to a propagating wave of depolarization.

Does it really work? See Chapter 11.

FURTHER READING

Undersea telegraph cables:

Bodanis, 2005.

https://en.wikipedia.org/wiki/Transatlantic_telegraph_cable,

archived as <https://perma.cc/QU4Y-YF6J>

https://en.wikipedia.org/wiki/Cable_theory

https://en.wikipedia.org/wiki/William_Thomson,_1st_Baron_Kelvin#Transatlantic_cable

https://en.wikipedia.org/wiki/Submarine_communications_cable#Bandwidth_problems

CHAPTER 11

Nerve Impulses

11.1

11.1.1 The time course of an action potential suggests the hypothesis of voltage gating

Chapter 10 foreshadowed what is about to come. We must abandon the ohmic hypothesis, which states that all membrane conductances are fixed (Equation 10.5), in favor of something more interesting: The temporary reversal of the sign of the membrane potential reflects a sudden increase in g_{Na^+} (Equation 10.10 instead of Equation 10.9), so g_{tot} temporarily becomes dominated by the sodium contribution instead of by potassium. This change drives the membrane potential away from the potassium Nernst potential and toward that of sodium, thus creating the temporary reversed polarization characteristic of the action potential.

In fact, the cable equation shows quite directly that the ohmic hypothesis breaks down during a nerve impulse. We know that the action potential is a traveling wave of fixed shape, moving at some speed ϑ . For such a traveling wave, the entire history $\psi(x, t)$ is completely known once we specify its speed and its time course at *one* point: We then have $\psi(x, t) = \tilde{\psi}(t - (x/\vartheta))$, where $\tilde{\psi}(t) \equiv \psi(0, t)$ is the time course of an action potential (Figure 11.1a). Hence,

$$\frac{\partial \psi}{\partial x} = -\frac{1}{\vartheta} \left. \frac{d\tilde{\psi}}{dt'} \right|_{t'=t-(x/\vartheta)},$$

by the chain rule of calculus. Rearranging the cable equation (Equation 10.1) then gives us the total membrane current j_r from the measured time course $\tilde{\psi}(t)$ of the membrane potential at a fixed position:

$$j_r = \frac{a\kappa}{2\vartheta^2} \frac{d^2\tilde{\psi}}{dt^2} - c \frac{d\tilde{\psi}}{dt}. \quad (11.1)$$

Applying Equation 11.1 to the measured time course of an action potential, sketched in Figure 11.1a, gives us the corresponding time course for the membrane current (Figure 11.1b). We can understand this result graphically, without any calculations. Note that the membrane current is particularly simple at the inflection points of panel (a) (the dashed lines labeled 1, 3, and 5): Here the first term of Equation 11.1 equals zero, and the sign of the current is opposite to that of the slope of $\tilde{\psi}(t)$. Similarly, at the extrema of panel (a) (the dashed lines labeled 2 and 4), we find that the *second* term of Equation 11.1 vanishes: Here the sign of the current is that of the *curvature* of $\tilde{\psi}(t)$, as shown in panel (b). With these hints, we can work out the sign of j_r at the points 0–6; joining the dots gives the curve sketched in panel (b).

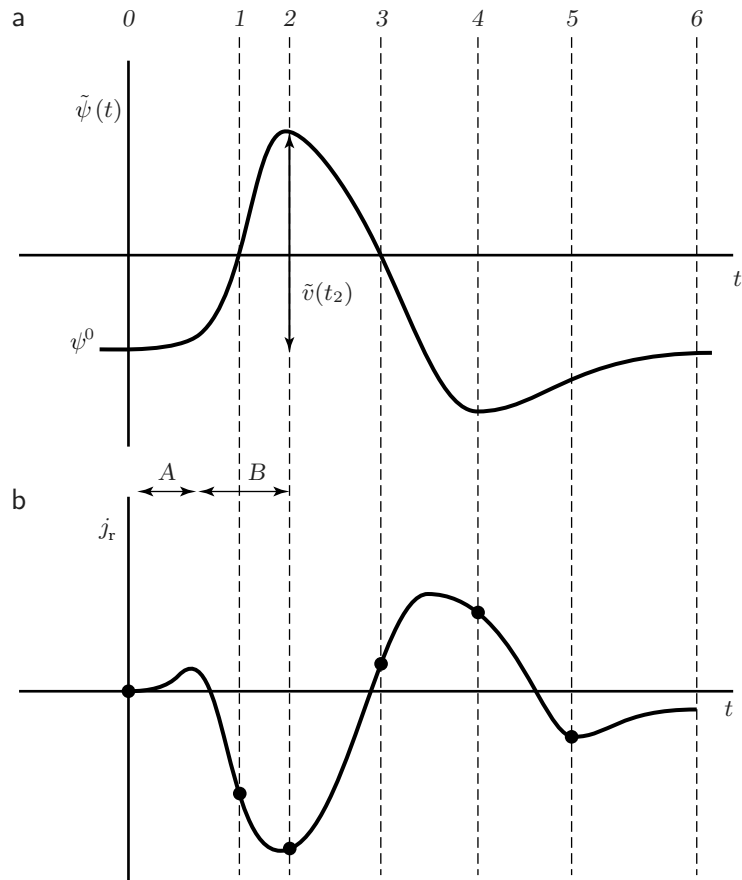


Figure 11.1: [Sketch graphs.] **The time course of an action potential.** (a) The sketch shows the membrane potential $\tilde{\psi}(t)$, measured at a fixed location $x = 0$. $\tilde{v}(t)$ refers to the difference between the membrane potential and its resting value ψ^0 . The *dashed lines* are six particular moments of time discussed in the text. (b) Reconstruction of the total membrane current from (a), using Equation 11.1. An ohmic stage A gives way to another stage B. In B, the membrane potential continues to rise but the current falls and then reverses; this is non-ohmic behavior. [Adapted from Benedek & Villars, 2000.]

Comparing the two panels of Figure 11.1 shows what is happening during the action potential. Initially (stage A), the membrane conductance is indeed ohmic: The cell's interior potential begins to rise above its resting value, thereby driving an outward current flux, as predicted from your calculation of the potential of three resistor–battery pairs (Your Turn 10A (page 127)). But when the membrane has depolarized by about 10 mV, something strange begins to happen (stage B): The potential continues to rise, but the net current falls.

Idea 10.8 made the key point needed for understanding the current reversal, in terms of a switch in the membrane's permeabilities to various ions. Net current flows across a membrane whenever the actual potential difference ψ deviates from the “target” value. But the target value itself depends on the membrane conductances. If these suddenly change from their resting values, so will the target potential; if the target switches from being more negative than ψ to more positive, then the membrane

current will change sign. Because the target value is dominated by the Nernst potential of the most permeant ion species, we can explain the current reversal by supposing that the membrane's permeability to sodium increases suddenly during the action potential.

So far, we have done little more than restate Idea 10.8. As outlined in Section 10.2.4, Hodgkin and Huxley noted that the increase does not begin until after the membrane has depolarized significantly (Figure 11.1, stage *B*), so they proposed that

Membrane depolarization itself is the trigger that causes the sodium conductance to increase. (11.2)

That is, they suggested that some collection of unknown molecular devices in the membrane allow the passage of sodium ions, with a conductance depending on the membrane potential. Idea 11.2 introduces an element of **positive feedback** into our picture: Depolarization begins to open the sodium gates, a process that increases the degree of depolarization. The increased depolarization opens still more sodium gates; and so on.

The simplest way to implement Idea 11.2 is to retain the ohmic hypothesis, but with the modification that each of the membrane's conductances may depend on ψ :

$$j_r = \sum_{\ell} (\psi - \psi_{\ell}^{\text{Nernst}}) g_{\ell}(\psi). \quad \text{simplified voltage-gating hypothesis} \quad (11.3)$$

In this formula, the conductances $g_{\ell}(\psi)$ are unknown (but positive) functions of the membrane potential. Equation 11.3 is our proposed replacement for the ohmic hypothesis.¹

The proposal Equation 11.3 certainly has a lot of content, even though we don't yet know the precise form of the conductance functions appearing in it. For example, it implies that the membrane's ion currents are still ohmic (linear in $\ln(c_{\text{out}}/c_{\text{in}})$) if we hold ψ fixed while changing the concentrations. However, the membrane current is now a *nonlinear* function of ψ , a crucial point for the following analysis.

Before proceeding to incorporate Equation 11.3 into the cable equation, let's place it in the context of this book's other concerns. We are accustomed to positive ions moving along the electric field, which then does work on them; they dissipate this work as heat as they drift against the viscous drag of the surrounding water. This migration has the net effect of reducing the electric field: Organized energy (stored in the field) has been degraded to disorganized (thermal) energy. But stage *B* of Figure 11.1b shows ions moving inward, that is, in a direction *opposite* to that of the potential drop. The energy needed to drive them can only have come from the thermal energy of their surroundings. Can thermal energy really turn back into organized (electrostatic) energy? Previous chapters have argued that such unintuitive energy transactions are possible, as long as they reduce the *free* energy of the system. And in fact, the axon started out with excess free energy, in the form of its nonequilibrium ion concentrations. The source of this stored free energy is the cell's metabolism, via the membrane's ion pumps.

¹The symbol $\Delta\psi$ appearing in Chapter 10 is abbreviated in this chapter as ψ .

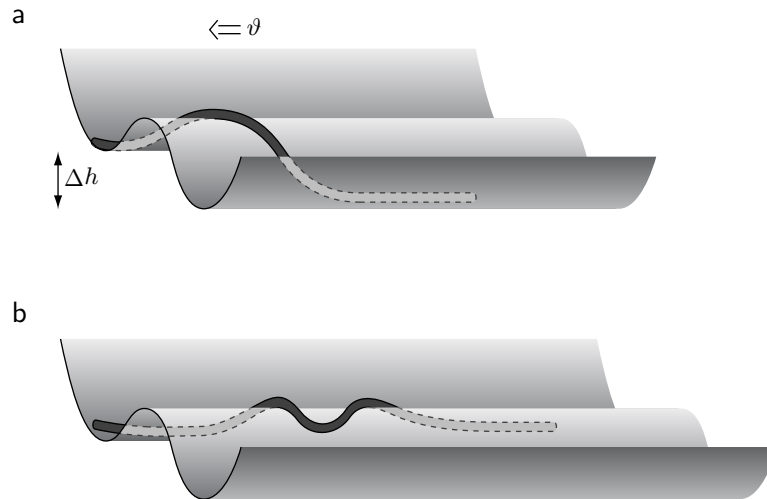


Figure 11.2: [Schematic.] **Mechanical analog of the action potential.** A heavy chain lies in a tilted channel, with two troughs at heights differing by Δh . (a) An isolated kink will move steadily to the left at a constant speed v : successive chain elements are lifted from the upper trough, slide over the crest, and fall into the lower trough. (b) A disturbance can create a pair of kinks if it is above threshold. The two kinks then travel away from each other.

Note that Equation 11.3 implies that the conductances track changes in potential instantaneously. Section 11.1.2 will show how this simplified conductance hypothesis already accounts for much of the phenomenology of the action potential.²

11.1.2 Voltage gating leads to a nonlinear cable equation with traveling wave solutions

We can now return to the apparent impasse reached in our discussion of the linear cable equation (Section 10.2.3): There seemed to be no way for the action potential to gain access to the free energy stored along the axon membrane by the ion pumps. The previous section motivated a proposal for how to get the required coupling, namely, the simplified voltage-gating hypothesis. However, it left unanswered the question of who *orchestrates* the orderly, sequential increases in sodium conductance as the action potential travels along the axon? The full answer to this question is mathematically rather complex. This section will implement a simplified version, in which we can actually solve an equation and see the outline of the full answer.

Consider first a mechanical analogy, a chain that progressively shifts from a higher to a lower groove (Figure 11.2a). This system can support a traveling wave of fixed speed and definite waveform. Now we must translate our ideas into the context of axons, and do the math.

The force needed to pull each successive segment of chain over its potential barrier comes from the *previous* segment of chain. Translating into the language of our axon,

²Nelson, 2014, §12.3.1 describes how Hodgkin and Huxley managed to measure the conductance functions and how they were forced to modify the simplified voltage-gating hypothesis somewhat.

this idea suggests that even though the resting state is a stable steady state of the membrane,

- Once one segment depolarizes, its depolarization spreads passively to the neighboring segment;
- Once the neighboring segment depolarizes by more than 10 mV, the positive feedback phenomenon described in the previous section sets in, triggering a massive depolarization; and
- The process repeats, spreading the depolarized region.

(11.4)

Let's begin by focusing only on the initial sodium influx. Thus we imagine only one voltage-gated ion species, say, Na^+ . We also suppose that the membrane's conductance for this ion, $g_{\text{Na}^+}(v)$, depends only on the momentary value³ of the potential disturbance $v \equiv \psi - \psi^0$.

A detailed model would use an experimentally measured form of the conductance per area $g_{\text{Na}^+}(v)$, as imagined in the dashed line of Figure 11.3a. We will instead use a mathematically simpler form (solid curve in the figure), namely, the function

$$g_{\text{Na}^+}(v) = g_{\text{Na}^+}^0 + Bv^2. \quad (11.5)$$

Here $g_{\text{Na}^+}^0$ represents the resting conductance per area; as usual, we lump this in with the other conductances and call the sum g_{tot}^0 . B is a positive constant. Equation 11.5 incorporates the key feature of increasing upon depolarization; moreover, it is always positive, as a conductance must be.

The total charge flux through the membrane is then the sum of the sodium contribution, plus ohmic terms from the other ions:

$$j_r = \left(\sum_{\ell} (\psi - \psi_i^{\text{Nernst}}) g_{\ell}^0 \right) + (\psi - \psi_{\text{Na}^+}^{\text{Nernst}}) Bv^2. \quad (11.6)$$

As in Your Turn 10A (page 127), the first term in Equation 11.6 can be rewritten as $g_{\text{tot}}^0 v$. Letting H denote the constant $\psi_{\text{Na}^+}^{\text{Nernst}} - \psi^0$, we can also rewrite the last term as $(v - H)Bv^2$, obtaining

$$j_r = v g_{\text{tot}}^0 + (v - H) Bv^2. \quad (11.7)$$

Figure 11.3b helps us understand graphically the behavior of our model. There are three important points on the curve of current versus depolarization, namely, the points where the membrane current j_r is zero. Equation 11.7 shows that these points are the roots of a cubic equation. We write them as $v = 0$, v_1 , and v_2 , where v_1 and v_2 equal $\frac{1}{2}(H \mp \sqrt{H^2 - 4g_{\text{tot}}^0/B})$, respectively. At small depolarization v , the sodium permeability stays small, so the last term of Equation 11.7 is negligible. In this case, a small positive v gives small positive (outward) current, as expected: We are in the ohmic regime (stage A of Figure 11.1). The outward flow of charge tends to reduce v back toward zero. A further increase of v , however, opens the voltage-gated sodium channels, eventually reducing j_r to zero, and then below zero as we pass the point v_1 . Now the net inward flow of charge tends to *increase* v , giving positive feedback—an

³These assumptions are not fully realistic; thus our simple model will not capture all the features of real action potentials. Nelson, 2014, §12.3.1 discusses an improved model.

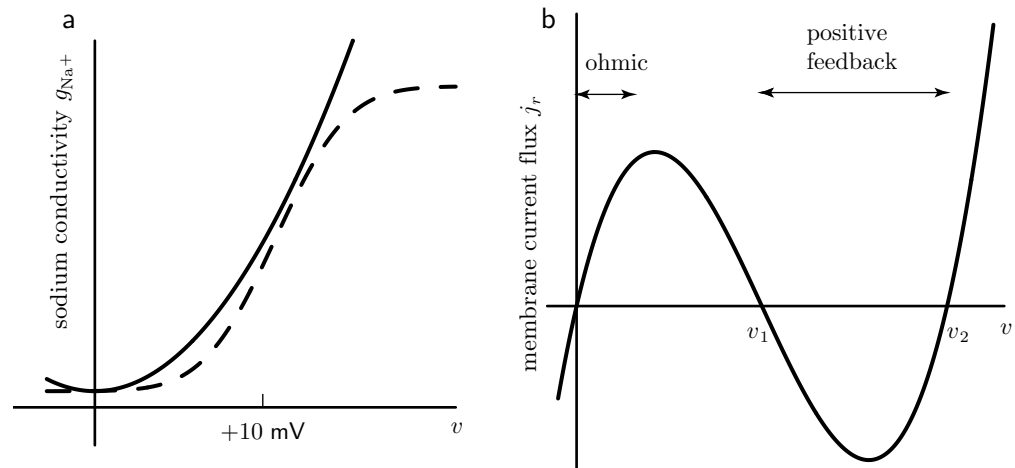


Figure 11.3: [Sketch graphs.] **Voltage-gating hypothesis.** (a) *Dashed curve:* The conductance g_{Na^+} of an axon membrane to sodium ions, showing an increase as the membrane potential increases from its resting value ($v = 0$). *Solid curve:* Simplified form for membrane sodium conductance (Equation 11.5). This form captures the relevant feature of the dashed curve, namely, that it increases as v increases and is positive. (Even the dashed line is not fully realistic: Real membrane conductances do not respond instantly to changes in membrane potential; rather they reflect the past history of v .) (b) Current-voltage relation resulting from the conductance model in (a) (Equation 11.7). The special values v_1 and v_2 are defined in the text.

avalanche. Instead of returning to zero, v drives toward the other root, v_2 . At still higher v , we once again get a positive (outward) current, as the large outward electric force on all the ions finally overcomes the entropic tendency for sodium to drift inward.

In short, our model displays threshold behavior: Small disturbances get driven back to $v = 0$, but above-threshold disturbances drive to the other stable fixed point v_2 . Our program is now to adapt the steps in Section 10.2.3.

Equation

We first substitute Equation 11.7 into the cable equation (Equation 10.1 (page 122)). Some algebra shows that $v_1 v_2 = g_{\text{tot}}^0 / B$, so the cable equation becomes

$$(\lambda_{\text{cable}})^2 \frac{\partial^2 v}{\partial x^2} - \tau \frac{\partial v}{\partial t} = \frac{v(v - v_1)(v - v_2)}{(v_1 v_2)}. \quad \text{nonlinear cable equation} \quad (11.8)$$

Unlike the linear cable equation, Equation 11.8 is not equivalent to a diffusion equation. In general, it's very difficult to solve nonlinear, many-variable differential equations like this one. But we can simplify things, because our main interest is in finding whether there are any traveling wave solutions to Equation 11.8. Following the discussion leading to Equation 11.1, we can represent a wave traveling at speed ϑ by a function $\tilde{v}(t)$ of *one* variable, via $v(x, t) = \tilde{v}(t - (x/\vartheta))$. Substituting into Equation 11.8 leads

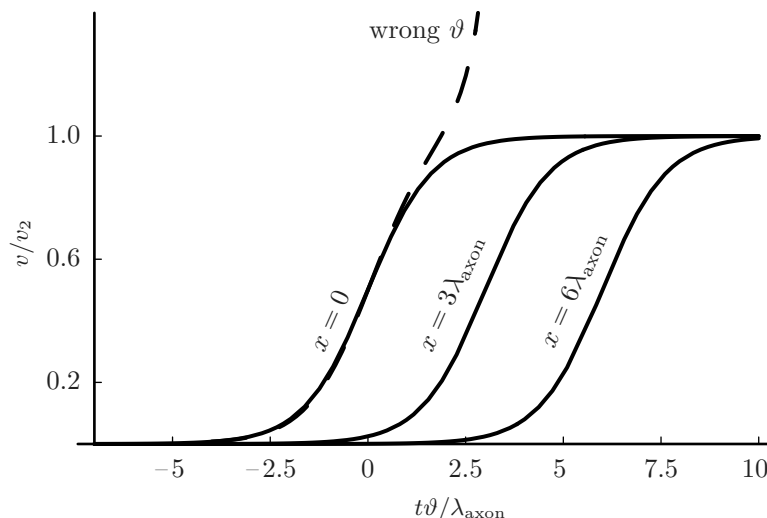


Figure 11.4: [Mathematical functions.] **Traveling wave solution to the nonlinear cable equation** (see Problem 11.1). The membrane potential relative to rest, $v(x, t)$, is shown as a function of time at three different fixed locations (*three solid curves*). Points at larger x see the wave go by at later times, so this wave is traveling in the $+\hat{x}$ direction. The parameter $s \equiv v_2/v_1$ has been taken equal to 3 for illustration. This simplified model qualitatively reproduces the leading edge of the action potential (Figure 11.1a). The *dashed line* shows a solution to Equation 11.9 with a value of the front velocity ϑ different from that in Equation 11.11; this solution is singular. Time is measured in units of $\lambda_{\text{cable}}/\vartheta$. The potential relative to resting is measured in units of v_2 (see text).

to an *ordinary* (one-variable) differential equation:

$$\left(\frac{\lambda_{\text{cable}}}{\vartheta}\right)^2 \frac{d^2 \tilde{v}}{dt^2} - \tau \frac{d\tilde{v}}{dt} = \frac{\tilde{v}(\tilde{v} - v_1)(\tilde{v} - v_2)}{v_1 v_2}. \quad (11.9)$$

We can tidy up the equation by defining the dimensionless quantities $\bar{v} \equiv \tilde{v}/v_2$, $y \equiv -\vartheta t/\lambda_{\text{cable}}$, $s \equiv v_2/v_1$, and $Q \equiv \tau\vartheta/\lambda_{\text{cable}}$, finding

$$\frac{d^2 \bar{v}}{dy^2} = -Q \frac{d\bar{v}}{dy} + s\bar{v}^3 - (1+s)\bar{v}^2 + \bar{v}. \quad (11.10)$$

Solution

You could enter Equation 11.10 into a computer-math package, substitute some reasonable values for the parameters Q and s , and look at its solutions. But it's tricky: The solutions are badly behaved (they blow up) unless you take Q to have one particular value (see Figure 11.4). This behavior is not surprising in the light of Figure 11.2: Our mechanical analog system selects one definite value for the pulse speed (and hence Q). You'll find in Problem 11.1 that choosing

$$\vartheta = \pm \frac{\lambda_{\text{cable}}}{\tau} \sqrt{\frac{2}{s} \left(\frac{s}{2} - 1\right)} \quad (11.11)$$

yields a traveling wave solution (the solid curves in Figure 11.4).

Interpretation

The hypothesis of voltage gating, embodied in the nonlinear cable equation, has led to the appearance of traveling waves of definite speed and shape. In particular, the amplitude of the traveling wave is fixed: It smoothly connects two of the values of v for which the membrane current is zero, namely, 0 and v_2 (Figure 11.3). We cannot excite such a wave with a very small disturbance. Clearly, for small enough v , the nonlinear cable equation is essentially the same as the linear one (Equation 10.3 (page 122)), whose solution we have already seen corresponds to passive, diffusive spreading (electrotonus), not an action potential. Thus

- a. *Voltage gating leads to a graded, diffusive response for stimuli below some threshold, but above-threshold, depolarizing stimuli yield a large, fixed-amplitude response.* (11.12)
- b. *The above-threshold response takes the form of a traveling wave of fixed shape and speed.*

Our model, a mathematical embodiment of Idea 11.4, has captured many of the key features of real nerve impulses. We didn't prove that the wave rapidly forgets the precise nature of its initial stimulus, remembering only whether it was above threshold or not, but such behavior should seem reasonable in the light of the mechanical analogy (Figure 11.2). We also get a quantitative prediction. The velocity ϑ is proportional to $\lambda_{\text{cable}}/\tau = \sqrt{a\kappa g_{\text{tot}}/(2\mathcal{C}^2)}$ times a factor independent of the axon's radius a . Thus the model predicts that if we examine a family of unmyelinated axons of the same general type, with the same ion concentrations, we should find that the pulse speed varies with axon radius as $\vartheta \propto \sqrt{a}$. This prediction is roughly borne out in experimental data. Moreover, the overall magnitude of the pulse speed is approximately $\lambda_{\text{cable}}/\tau$. For the squid giant axon, our estimates give this quantity as about $12 \text{ mm}/2 \text{ ms} = 6 \text{ m s}^{-1}$, a value within an order of magnitude of the measured action potential speed of about 20 m s^{-1} .

In the mechanical analogy, the wave speed is proportional to the density of stored energy divided by a friction constant. Examining our expression for ϑ , we notice that both κ and g_{tot} are inverse resistances, so $\sqrt{\kappa g_{\text{tot}}}$ is indeed an inverse "friction" constant. In addition, the formula $\mathcal{E}/\Sigma = \frac{1}{2}q^2/(\mathcal{C}\Sigma)$ for the electrostatic energy density stored in a charged membrane of area Σ shows that the stored energy is proportional to $1/\mathcal{C}$. Thus our formula for ϑ has essentially the structure expected from the mechanical analogy.

[\[T2\]](#) Section 11.1.2' (page 137) discusses how the nonlinear cable equation determines the speed of its traveling wave solution.

REFERENCES

Neurons:

The approximate treatment discussed here appears in Benedek & Villars, 2000; Nelson, 2014; Keener & Sneyd, 2009, §6.2; Bressloff, 2014, §2.2.

See also Phillips et al., 2012.

T₂ Track 2

11.1.2'

1. Problem 11.1 pulls an exact analytic solution out of a hat. The fact that any such solution exists may seem a miracle, a pathology of our particular simplification of the equations. To see that the behavior we found is actually generic, here is a physically inspired argument. Begin with Equation 11.8. We are interested in traveling wave solutions, representing the situation where the initial resting state ($v = 0$) is invaded by the excited state ($v = v_2$). Thus we explore trial solutions of the form $v = f(t - x/\vartheta)$ where $f(t) \rightarrow 0$ as $t \rightarrow -\infty$ etc. The wave velocity ϑ is not known yet.

To get a recognizable equation, first change variables:

$$y = -|\vartheta|t/\lambda_{\text{cable}}, \quad \frac{d}{dt} = -\frac{|\vartheta|}{\lambda} \frac{d}{dy}.$$

As a function of y , we now have $f(y) \rightarrow 0$ as $t \rightarrow +\infty$ etc. Now multiply both sides of Equation 11.8 by df/dy and rearrange to find

$$\frac{d}{dy} \left[\frac{1}{2} \left(\frac{df}{dy} \right)^2 + U(f) \right] = -Q \left(\frac{df}{dy} \right)^2$$

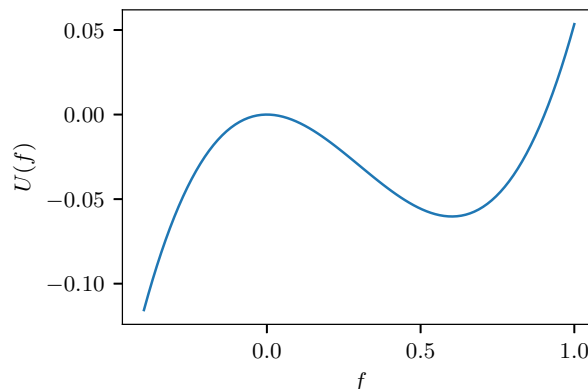
where

$$U(f) = -\frac{1}{v_1 v_2} \left[\frac{1}{4} f^4 - \frac{1}{3} (v_1 + v_2) f^3 \right] - \frac{1}{2} f^2, \quad \text{and} \quad Q = \tau |\vartheta| / \lambda_{\text{cable}}.$$

We know how to think about nonlinear ODEs of this sort. Think of a roller-coaster car, rolling on a potential energy landscape U . On the left side of this equation we have the time derivative of kinetic plus potential energy. On the right side we have frictional loss (in a world where roller coasters are immersed in a viscous fluid). Our roller coaster starts at “time” $y \rightarrow -\infty$ on top of the taller hill ($f = v_2$). After a long time (set by the size of an initial small perturbation), it rolls off the hill toward the left.

I took total resting membrane conductance to be $5 \Omega^{-1} \text{m}^{-2}$ and $H = 100 \text{ mV}$. To get a value for B , I noticed that the sodium conductance rises from negligible to about 13 times the resting total conductance as membrane potential rises from resting to about 40 mV greater than that. This gave $(B/g_{0,\text{tot}})(40 \text{ mV})^2 = 13$, and then $v_1 = 0.60$, $v_2 = 200 \text{ mV}$. The value of v_2 is much higher than the actual maximum of an action potential, but we only want the leading edge; we are neglecting the later potassium currents and sodium channel inactivation that later cut off the rise of potential.

The graph of the effective potential function $U(f)$ levels off at 0 , v_1 , and v_2 . The figure below shows a detail close to zero (the starting point is far to the right, at v_2):



The generic behavior that ensues is that the roller coaster ends up at v_1 , perhaps after some oscillations. But v_1 is an unstable solution for the nerve impulse: It's right in the middle of the negative-resistance regime. It may be stable in our roller-coaster model, but that explores only a very restricted domain (constant-speed traveling waves). Another generic behavior, if the friction is low enough, is that the roller coaster surmounts the hill on the left and flies off to $f \rightarrow -\infty$. That's definitely unphysical. But if the friction is just exactly right, we can get another behavior: The roller coaster ends up poised right at $f = 0$, approaching it asymptotically. That is the traveling wave solution we sought. We can always arrange for such a solution to exist by choosing the right value of friction Q , that is, the right wave propagation velocity ϑ . This argument is unchanged if we make small modifications to the assumed potential U , even if there are no longer any exact solutions.

2. Actually, *all* types of ion channels potentially have voltage-dependent conductance, not just sodium. We focused on sodium because it's responsible for the switch to the high-conductance state (leading edge of an action potential). Later, potassium channels open, and later still the sodium channels “inactivate”; both of those processes contribute to shutting down the conduction and returning the axon to its resting state.

Moreover, they don't respond instantaneously to changes in potential (each has its own kinetics). Still, we already see the germ of the idea, and neglecting these fine points gives us a tractable equation that displays the qualitative behavior we were trying to explain.

3. The word “later” reminds us that channels do not actually respond instantly to the current membrane potential. Introducing realistic kinetics leads to a much more complicated system, though with similar behavior at the leading edge. There is a useful intermediate theory, however, the “FitzHugh-Nagumo” system, in which the fastest ion channels (sodium) are assumed to respond instantly, and slower dynamics are merged into a single independent dynamical variable (Keener & Sneyd, 2009).

4. A nonlinear traveling wave is sometimes called a **soliton**.⁴ Here is another context for them. The cables that send Internet between cities are not wires at all, but optical fibers. They can be formulated with ultra-low loss (absorption), but they still suffer from optical dispersion (mushing-out of signals). Modern optical fibers have nonlinear optical effects that make them transmit those ones and zeros as solitons, preserving their shape for hundreds of kilometers.

⁴Some authors reserve this word for the special case of an “exactly integrable” system.

PROBLEMS**11.1** *Analytical solution for simplified action potential*

Show that the function $\bar{v}(y) = (1 + e^{\alpha y})^{-1}$ solves Equation 11.10 (page 135), if we take the parameter Q to be given by $\sqrt{2/s}(\frac{s}{2} - 1)$. Hence derive the speed of the action potential (Equation 11.11). α is another constant, which you are to find.

CHAPTER 12

Examples of 3-tensors in Physics

The image of the theorist that emerges is of a surprisingly unsophisticated individual who must anthropomorphize nature to understand it: I like symmetry and beauty; ergo, nature likes symmetry and beauty. It reminds one of the Parisian animal trainer who teaches his bear to respond to voice commands and concludes that bears speak French.

— *Dick Teresi*

12.1 FRAMING

We have informally introduced a mathematical object called a “tensor.” Chapter 3 introduced the quadrupole moment, a tensor in three-dimensional space or **3-tensor**. Chapter 6 constructed the metric and curvature tensors of a two-dimensional surface; they are **2-tensors**.

Ultimately our goal is to define and exploit 4-tensors. Before we go there, let’s see some more examples that may be familiar to you, at least implicitly, from previous work. Like the man who discovered he had been speaking prose all his life, you are already familiar with some tensors.

12.2 RANK ZERO; RANK ONE

A “rank-zero three-tensor” is just a fancy term for a single number. More precisely, it expresses a physical quantity that is the same in any coordinate system. Electric charge is an example. It doesn’t need any coordinate index (that is, it carries *zero* indices).

A “rank-one three-tensor” is just a fancy term for what we have been calling a 3-vector. Velocity is an example. It carries one coordinate index because we need three real numbers to express it in a given coordinate system.

Equally, we can think of a rank-one three-tensor as a function that eats a vector, emits a scalar, and is linear. For example, the projection $f(\vec{v}) = \vec{a} \cdot \vec{v}$ is such a function, where \vec{a}_i is a set of three constants. Either way, we need three numbers to specify an object in this class.

From now on, I’ll usually drop summation symbols on tensor indices, relying on the convention that a repeated index is to be summed unless otherwise noted.

12.3 RANK TWO

Three-tensors of rank two play two closely related roles in pre-Einstein physics:¹

- A tensor may express a vector-valued function of another vector that is linear (or the linearized approximation to a more general function).
- A tensor may express a scalar-valued function of a vector that is quadratic (or that is linear in each of two vector arguments).

12.3.1 Tensors as linear vector-valued functions of a vector

When an auto mechanic says your car's wheels need to be “balanced,” what does (s)he mean? Clearly it's desirable to ensure that the wheel's center of mass lies on the axle. Otherwise, spinning the wheel would require the CM to move in a circular orbit. Circular motion implies acceleration, which requires a force. So as the wheel spins, the axle is constantly subjected to sideways forces, which would wear out the bearings etc. if not corrected.

But there is more. Suppose that the CM lies on the axle, but the wheel is bent, so that its axis of symmetry does not coincide with the axle. Spinning the wheel about the axle, even at constant angular velocity, then generates *torque*, which is just as bad for the car as the forces mentioned previously. What is going on?

When we spin a rigid body about *any* axis with angular frequency ω , we can define the **angular velocity** $\vec{\omega}$ as the product of ω with a unit vector pointing along that axis, with sign chosen by the right-hand rule. Suppose that the body is subdivided into small masses m_ℓ located at positions \vec{r}_ℓ relative to a reference point fixed in the body. Then the resulting angular *momentum* \vec{L} has components that are linear functions of $\vec{\omega}$, and that therefore may be written² as $\vec{L}_i = \vec{J}_{ij}\vec{\omega}_j$, or simply $\vec{L} = \vec{J} \cdot \vec{\omega}$. The **moment of inertia tensor** is defined by the formula

$$\vec{J} = \sum_{\ell} m_{\ell} [(r_{(\ell)})^2 \mathbf{1} - \vec{r}_{(\ell)} \vec{r}_{(\ell)}]. \quad (12.1)$$

The first term above is the tensor whose entries are the unit matrix. The second term is called a **dyad product**, shorthand for the tensor with nine components

$$[\vec{r} \vec{r}]_{ij} = \begin{bmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{bmatrix}_{ij}.$$

When two vectors are juxtaposed with no dot or cross joining them, a dyad product is implied.³

¹Post Einstein, we have tensors in four or more dimensions, as well as tensor operators in quantum mechanics and tensor representations of internal symmetry groups in high energy physics. All are subject to similar analysis. There are also generalizations to handle intrinsic particle spin, called “spinors” (Section 33.4', page 376).

²Most authors omit the over-arrow when stating components, but I'll retain it to emphasize the tensor status of the object they describe.

³Some authors use the symbol \otimes to make dyad products explicit, and refer to it as the “tensor product.” Mathematicians sometimes call it “outer product.”

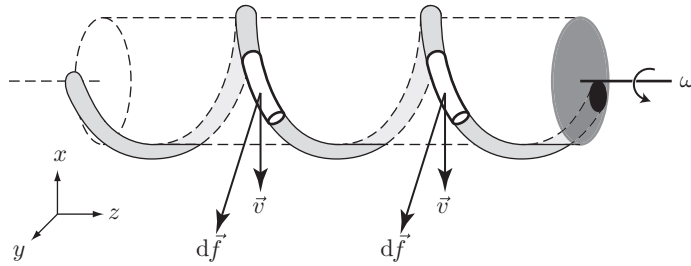


Figure 12.1: [Schematic.] **Principle of flagellar propulsion in bacteria.** A thin, rigid, helical rod is cranked about its helix axis at angular speed ω . For better visualization, a phantom cylinder has been sketched, with the rod lying on its surface. Two short segments of the rod have been singled out for study, both lying on the near side of the helix and separated by one turn. The rod is attached (black circle) to a disk and the disk is rotated, cranking the helix about its axis. The two short segments then move downward in the plane of the page (along $-\hat{x}$). Thus, $d\vec{f}$ lies in the xz plane, but tipped slightly to the left as shown. A net force with a negative z -component would be required to keep the helix spinning in place; without such a force, the helix will move to the right.

A tensor whose matrix of components is symmetric, for example $\overset{\leftrightarrow}{\mathcal{J}}$, will itself be called a **symmetric** tensor.

Note that although \vec{L} depends linearly on $\vec{\omega}$, it need not point parallel to it. Also note that, although both \vec{L} and $\vec{\omega}$ change sign if we switch to a left-handed coordinate system, nevertheless the *relation between them* is unaffected. Indeed, Equation 12.1 does not contain any Levi-Civita symbols (where did they go?).

Your Turn 12A

- Work out Equation 12.1 from the definition $\vec{L} = \overset{\leftrightarrow}{\mathcal{J}} \cdot \vec{\omega}$.
- Show that the matrix of components $\overset{\leftrightarrow}{\mathcal{J}}_{ij}$ is symmetric, though not traceless (unlike the quadrupole tensor).
- Work out the moment of inertia tensor of a cylinder with uniform mass density, using some convenient coordinate system. Make an Appropriate Comment about spinning it about an axis that passes through its center but does not coincide with the axis of symmetry.

Here are some more examples:

- When we pull a rigid body through a viscous fluid, the fluid exerts a retarding **drag force**. If the body is spherical, then the drag force points oppositely to the velocity, but more generally we have a linear relation $\vec{f} = \overset{\leftrightarrow}{\zeta} \cdot \vec{v}$, involving a **viscous drag tensor**. The fact that \vec{f} need not be parallel to \vec{v} is the secret to bacterial locomotion (Figure 12.1).
- An object suspended on an array of springs has an equilibrium position. If we apply a small force to the object, then it responds by finding a new mechanical equilibrium displaced by some \vec{r} , such that $\vec{f} = -\overset{\leftrightarrow}{K} \cdot \vec{r}$. Here the **spring constant tensor** $\overset{\leftrightarrow}{K}$ summarizes the spring system as far as its linear response is concerned. Conversely, $\vec{r} = -\overset{\leftrightarrow}{K}^{-1} \cdot \vec{f}$; the tensor whose components are the inverse matrix of $\overset{\leftrightarrow}{K}_{ij}$ is called the **compliance**.

3. Continuing (2), suppose that the object is charged; for example it could be part of a molecule. Then it responds to an applied electric field with a displacement, which in turn gives an **induced dipole moment**. If the force is applied by an external electric field, then in the linear regime

$$\vec{\mathcal{D}}_E = \vec{\alpha} \cdot \vec{E}$$

where $\vec{\alpha} = q^2 \vec{K}^{-1}$ is the **polarizability tensor**.

4. Some electrically conductive media are ohmic but anisotropic. Analogously to example #1, this means that $\vec{j} = \kappa \vec{E}$ is modified to read $\vec{j} = \vec{\kappa} \cdot \vec{E}$. More generally, any molecule or ion that moves diffusively has a **mobility tensor**, which will not be a scalar in an anisotropic medium.
5. The force exerted by a small element of fluid on its adjacent neighbor is proportional to the area $d\Sigma$ of the interface between them, but need not be directed perpendicular to that surface. Thus, $d\vec{f} = \vec{T} \cdot d\vec{\Sigma}$, where the **stress tensor** \vec{T} is another symmetric rank-2 tensor.⁴ Its trace divided by 3 is called the **pressure** of the fluid; other non-trace contributions represent **viscous stresses**.
6. The order parameter describing the state of a nematic liquid crystal can also be regarded as a traceless symmetric rank-two tensor.

12.3.2 Symmetric tensor as a scalar-valued, quadratic function of a vector

The length-squared function, $f(\vec{v}) = \|\vec{v}\|^2$, is a scalar-valued function that is quadratic in the components of \vec{v} . We'll call it the 3D **metric tensor**.⁵ Its components are given by the Kronecker symbol δ_{ij} .

Here are some more examples of this idea.

⁴If we regard an imaginary surface element as separating region 1 from region 2, then \vec{f} is the force that 1 exerts on 2 and $d\vec{\Sigma}$ is the normal directed from 1 to 2; thus a normal fluid's pressure is nonnegative. If we exchange the roles of 1 and 2, then both \vec{f} and $d\vec{\Sigma}$ change sign and the stress tensor is unchanged.

⁵This same tensor, regarded as a linear function on a vector, was called $\vec{\mathbf{I}}$ in Equation 12.1.

Your Turn 12B

- Show that the kinetic energy of a spinning rigid body is $\frac{1}{2}\vec{\omega} \cdot \vec{J} \cdot \vec{\omega}$.
- Show that the rate at which work is done pulling a rigid object through viscous fluid is $\vec{v} \cdot \vec{\zeta} \cdot \vec{v}$.
- Work out that the potential energy stored by the spring system is $\frac{1}{2}\vec{r} \cdot \vec{K} \cdot \vec{r}$. Similarly to the kinetic energy of a rigid body, we again see that a symmetric, second-rank 3-tensor can be used to specify a quadratic function of a vector.
- Convince yourself that the dissipated power density in a general ohmic material is $\vec{E} \cdot \vec{\kappa} \cdot \vec{E}$, another quadratic function of a vector. Think about how the units work in this formula.
- Chapter 6 introduced a quadratic function of small displacements describing how a curved 2D surface bends away from its tangent plane. (However, we extracted two *scalar* functions from this rank-2 symmetric 2-tensor, namely the mean and gaussian curvatures.)

A tensor that specifies a quadratic function must be symmetric, because any antisymmetric part would cancel in the expressions appearing above. So the moment of inertia, quadrupole, metric, and curvature tensors all have this property.

The electric quadrupole moment \vec{Q}_E also defines a contribution to the far potential that depends quadratically on \vec{r} (see the third term of the far potential, Equation 3.1, page 31). Also like the examples above, \vec{Q}_E has a coordinate representation as a 3×3 matrix, which changes when we change coordinates (or rotate the object) in the same way as any of the other tensors above. In addition, it is traceless.

12.3.3 Some linear vector functions, but not all, arise as the derivative of a quadratic scalar function

In ordinary calculus, any linear function can be written as the derivative of a quadratic function: $\alpha x = (\frac{1}{2}\alpha x^2)'$. Some vector-valued functions of a vector can similarly be written as the gradient of a quadratic function. For example, the Hooke-law force is the gradient of minus the potential energy. Unlike in one dimension, however, *not every* linear $\vec{f}(\vec{r})$ can be expressed in this way.

For example, consider again a rigid body. When we rotate it about the z axis, the position of each mass element ℓ changes from $\vec{r}_{(\ell)}$ to $\vec{r}_{(\ell)} + d\vec{r}_{(\ell)}$, where

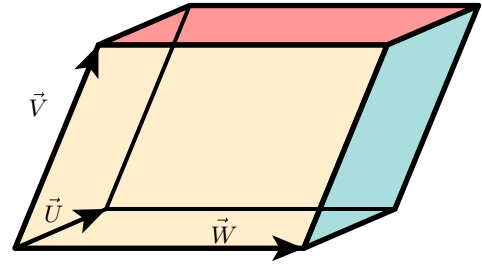
$$d\vec{r}_{(\ell)i} = d\vec{\Omega}_{ij}\vec{r}_{(\ell)j} \quad \text{with} \quad \vec{\Omega}_{ij} = d\theta \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{ij}.$$

This linear function of $\vec{r}_{(\ell)}$ is expressed by an *antisymmetric* matrix, whereas anything arising from a quadratic function would have to be expressed by a *symmetric* matrix.

12.4 RANK THREE

We can extend these ideas.

Figure 12.2: A **parallelepiped** is a six-sided solid with three pairs of parallel faces, each of which is a parallelogram.



12.4.1 A vector-valued bilinear function of vectors

Here is a recipe that you recall: Given two vectors, return zero if they are parallel (or if either is zero). Otherwise, find the vector \hat{n} perpendicular to the plane that they span and chosen using the right-hand rule. Let a be the area of the parallelogram with the two given vectors as edges, and define the cross product as $a\hat{n}$. This new vector is linear in each of the two that we began with, for example:

- If we double either vector, a doubles and \hat{n} is unchanged.
- If we replace either vector by its negative, a is unchanged but \hat{n} reverses.

The cross product returns another 3-vector, so we need an array of numbers with *three* indices to express it. Instead of imagining its components as a matrix (grid of cells addressed by row and column), imagine it as a *building* with “rooms” addressable by row, column, and floor, each containing a number. Those 27 numbers, the components of the Levi-Civita *tensor*, are given by the Levi-Civita *symbol* defined earlier (Figure 0.3), as we can see by substituting \hat{x} and \hat{y} into the definition. Chapter 13 will discuss this rank-3 tensor in more detail.

12.4.2 A scalar-valued trilinear function of vectors

Here is another way to look at the Levi-Civita tensor. Given three vectors \vec{U} , \vec{V} , and \vec{W} , construct the parallelepiped that has these vectors as three edges (Figure 12.2). Compute the volume v of this solid and multiply by $\sigma = -1$ if the three given vectors form a left-handed triad (otherwise $\sigma = +1$):

$$\varepsilon(\vec{U}, \vec{V}, \vec{W}) = v\sigma. \quad (12.2)$$

Exchanging any two of the three vectors reverses the sign of σ , so we say that ε is *totally antisymmetric*.⁶

To see that Equation 12.2 yields a function that is linear in all three of its vector arguments, note:

- If we double the length of any vector, v doubles and σ is unchanged.
- If we replace any vector by its negative, v is unchanged but σ is replaced by its negative.

⁶A totally *symmetric* rank-3 tensor would be *unchanged* under exchange of any of its inputs, just as in rank two.

In fact, our function is just $\vec{U} \cdot (\vec{V} \times \vec{W})$, similar to the relation between the two interpretations of rank-1 tensors in Section 12.2.

Again, substituting $\vec{U} = \hat{x}$, $\vec{V} = \hat{y}$, and $\vec{W} = \hat{z}$ shows that the 1,2,3 component of this tensor equals 1, which agrees with the 1,2,3 entry entry of the Levi-Civita symbol. We also see that permuting the three vectors leaves v unchanged but changes σ by the sign of the permutation, again like ε_{ijk} , so again we find that the Levi-Civita symbols are components of a totally antisymmetric, rank-3, 3-tensor.

Just as an antisymmetric matrix can be visualized in terms of reflections through the diagonal, so a totally antisymmetric 3-tensor has properties under reflections through planes (visible in Figure 0.3).

When there are more than 2 indices (rank higher than 2), it's too cumbersome to put any doodad above the symbol to indicate tensoriality. Also, in this situation we will rarely wish to drop the indices, so their presence suffices to announce that ε is a third-rank tensor.

T_2 Track 2

12.2' Tensor properties of probability density functions

[Not ready yet.]

PROBLEMS

12.1 *Rotationally invariant, symmetric 3-tensor*

Chapter 49 will argue that the propagation of light through a medium of randomly-oriented molecules involves the average of the polarizability tensor over rotations. Perhaps it seems reasonable to add, “That average must be a constant times the identity tensor.” Let’s prove this.

Note first that the rotational average must itself be a rotationally-invariant, symmetric 3-tensor. Call it \vec{A} ; then its matrix of components in some coordinate system must have the property that $S^t A S = A$ for any rotation matrix S . In particular, this property holds for any *infinitesimal* rotation. Recall that an infinitesimal rotation is given by $S = \mathbf{1} + \epsilon T + \mathcal{O}(\epsilon^2)$, where T is an antisymmetric matrix and $\mathbf{1}$ is the identity matrix.

Work out the consequences of invariance under such transformations (to order ϵ) and prove that A is a constant times $\mathbf{1}$.

12.2 *Liquid crystals*

Let’s illustrate the utility of tensor methods in another branch of physics.

Suppose that I tell you that some kind of matter (an “isotropic ferromagnet”) has states characterized by a spatially varying 3-vector field $\vec{v}(\vec{r})$ (the “order parameter”). The energy cost to be in one of these states is some analytic, local, rotationally invariant function of \vec{v} and its derivatives, integrated over space. Because it’s analytic, we can expand that function in Taylor series as a polynomial in the components of \vec{v} . Clearly the part of this function with no derivatives must involve only *even powers* of the components \vec{v}_i . This trivial fact has profound consequences for the phase-transition behavior of ferromagnets.

Now suppose that I tell you that some kind of matter (a “nematic liquid crystal”) has states characterized by a spatially varying, symmetric, traceless rank-2 tensor \vec{M} . The free energy cost to be in one of these states is some analytic, local, rotationally invariant function of \vec{M} and its derivatives, integrated over space. Because it’s analytic, we can expand that function in Taylor series as a polynomial in the components of \vec{M} . The part of this function with no derivatives must be at least quadratic in the components of M (explain why).

Now find all possible contributions to the free energy cost function (if any) that are quadratic or cubic in the components of \vec{M} . Your answer has profound consequences for the phase-transition behavior of nematic liquid crystals.

CHAPTER 13

Tensors from Heaven

13.1 FRAMING

The preceding chapter gave many examples of tensors in physics. A little thought shows that they fall into two main classes:

- Most of the examples were contingent; they describe properties of an object. If we rotate a mass distribution, its moment of inertia tensor in general changes (unless we rotate about a symmetry axis). Even total mass, which is rotationally invariant, changes if we consider a different object.
- Two of the examples were different: The 3D metric tensor is a property of *space itself*, not contingent on anything. And the Levi-Civita tensor is almost as universal; it depends only on a binary choice of which coordinate systems we will choose to call “right-handed.”

Let’s explore these last two tensors “from Heaven” a bit more. Along the way we will also examine how any tensor’s representation changes if, instead of changing the physical objects under consideration, we merely change our choice of coordinate system. This understanding will prove useful when we start to construct more elaborate things, and then again when we upgrade everything to four dimensions.

13.2 THE COMPONENTS OF A TENSOR TRANSFORM UPON CHANGE OF COORDINATES

13.2.1 Linear coordinate changes

Section 12.3.2 said that we may think about a spring constant tensor $\vec{\vec{K}}$ as a function that eats a displacement vector and spits out a number, the stored potential energy $\frac{1}{2}\Delta\vec{r}\cdot\vec{\vec{K}}\cdot\Delta\vec{r}$. This function is quadratic in the components of \vec{r} . It can be represented in any coordinate system by a matrix of ordinary numbers. We call those numbers the **components** of $\vec{\vec{K}}$ in the chosen coordinate system, and denote them by $\vec{\vec{K}}_{ij}$. It’s important that the nine *numbers* $\vec{\vec{K}}_{ij}$ depend not only on the physical *object* (system of springs), but also on a choice of *coordinate system* on space. That is, the same tensor can have different representations when referred to different coordinate systems.

Suppose that we define new coordinates by

$$\vec{r}'_a = S_{ai}\vec{r}_i. \quad (13.1)$$

Then the same spring potential energy function as before can also be written as

$\frac{1}{2}\Delta\vec{r}' \cdot \vec{K}' \cdot \Delta\vec{r}'$, where the new components are determined by

$$\vec{r}'_i \vec{K}'_{ij} \vec{r}'_j = \vec{r}'_a \vec{K}'_{ab} \vec{r}'_b = \vec{r} \cdot (\mathbf{S}^t \vec{K}' \mathbf{S}) \cdot \vec{r}.$$

This must hold for any spring displacement, so $\vec{K}' = \mathbf{S}^t \vec{K}' \mathbf{S}$, or

$$\vec{K}'_{ab} = S_{ai} S_{bj} \vec{K}_{ij}. \quad (13.2)$$

13.2.2 Cartesian coordinates and their transformations

In euclidean geometry there are always some special ways to associate numbers to points in space (that is, to choose a coordinate system¹). What's special about these "cartesian" coordinate systems is that the distance-squared between two points always takes the pythagorean form²

$$\|\Delta\vec{r}\|^2 = \sum_i \Delta\vec{r}_i \Delta\vec{r}_i \quad \text{in cartesian coordinates.} \quad (13.3)$$

Certainly we can find other coordinate systems for euclidean space in which the metric tensor *doesn't* have the simple form Equation 13.3, for example, polar coordinates. What makes euclidean space special is that at least one such set of "good" coordinates does exist (unlike, say, on the surface of a sphere).

If one cartesian coordinate system exists, then many other *equally good systems* will exist also. To see this, again define new coordinates via Equation 13.1, where now \mathbf{S} is specifically an **orthogonal matrix**, that is, one for which

$$\mathbf{S}\mathbf{S}^t = \mathbf{S}^t\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (13.4)$$

Then the new coordinates again have the property that the length-squared of a vector equals $\Delta\vec{r}'_a \Delta\vec{r}'_a$, which has the same form as Equation 13.3. For future use, note that Equation 13.4 implies

$(\det \mathbf{S})^2 = 1, \text{ and hence } \det \mathbf{S} = \pm 1 \text{ for an orthogonal matrix.} \quad (13.5)$

13.2.3 A reformulation

We defined tensors as operations involving vectors and showed that their components have certain transformation rules generalizing those of vectors. Alternatively, we could turn things around and instead define a general rank-2 tensor, such as a spring constant tensor \vec{K} , as a set of numbers that transform like the components of $\vec{r}\vec{r}$. Similar relations can be used to define a 3-tensor of any rank p : There will be p copies of the transformation matrix on the right-hand side of Equation 13.2.

¹Many physicists say "frame of reference" to mean specifically a coordinate system on space and time (four dimensions). For now, we consider three-dimensional space only.

²If you are worried about up- versus down-indices, we'll get to that fine point later. It's traditional to forget about this distinction when we work on euclidean space in cartesian coordinates, and always write coordinate indices as subscripts. If we use tensors on a non-euclidean space, or with curvilinear coordinates, the distinction becomes essential.

Let's look at the metric tensor from this new viewpoint. Instead of the geometric definition, we can say

Choose any cartesian coordinate system. Define $\overset{\leftrightarrow}{\mathbb{I}}$ to be that 2-tensor (13.6) whose components in this coordinate system are δ_{ij} .

The corresponding quadratic function defined by $\overset{\leftrightarrow}{\mathbb{I}}$ is then the usual length-squared.

The formulation Equation 13.6 may worry you: What if you and I start out with different cartesian coordinate systems? Will we agree on the meaning of $\overset{\leftrightarrow}{\mathbb{I}}$? To investigate, let's see how the components of your $\overset{\leftrightarrow}{g}$ look in my (primed) coordinate system:

$$\overset{\leftrightarrow}{\mathbb{I}}'_{ab} = S_{ai}S_{bj}\delta_{ij} = S_{aj}S_{bi} = [SS^t]_{ab} = \delta_{ab}, \quad (13.7)$$

the *same nine constants* as before. That is, it doesn't matter what coordinate system we started with, as long as it's cartesian: the components of the metric tensor are always the same. So the tensor we defined is not contingent on coordinates chosen; it is a property of space itself. Of course, in this case that conclusion is a tautology, not a surprise, because we explicitly restricted attention to those "good" coordinate systems in which it is true. However, Section 13.3.1 will use the same logic to get a more nontrivial result.

13.3 3D LEVI-CIVITA TENSOR

13.3.1 Coordinate version

Section 12.4 gave two geometric definitions of the Levi-Civita tensor, then noted that its components are given by the Levi-Civita symbol (that is, the constants ± 1 or zero). As before, one may worry: What if you and I choose different coordinate systems when defining it? As before, we must show that if we start in one cartesian system, then transform to any other, that the components are *numerically the same* as before. Then the tensor that they define won't actually depend on my original choice of coordinates—it will belong intrinsically to space itself. We know this must work out somehow, because we started with a geometric definition, but the details are interesting.

Even more interestingly, when mathematicians studied this problem they found that there were essentially *no more* new 3-tensors "from Heaven." You can build up higher-rank examples by sticking together some metric and Levi-Civita tensors (for example, $\delta_{ij}\delta_{k\ell}$), but that is all.

Suppose that we have a space that is euclidean, and that moreover we have agreed that one of the cartesian coordinate systems will be called "right-handed."³ We now define a 3-tensor by stating its components as in Section 0.2.2 (page 5):

$$\begin{aligned} \varepsilon_{ijk} &= 0 \text{ if any two of the indices are equal;} \\ \varepsilon_{ijk} &= +1 \text{ if } i, j, k \text{ are an even permutation of } 1, 2, 3; \\ \varepsilon_{ijk} &= -1 \text{ if } i, j, k \text{ are an odd permutation of } 1, 2, 3. \end{aligned} \quad (13.8)$$

³This has nothing to do with your hands, which side of your body your heart is on, nor the shape of your DNA. *Any* cartesian coordinate system may be singled out and given this special status.

Next, we must calculate the new components

$$\varepsilon'_{abc} = S_{ai}S_{bj}S_{ck}\varepsilon_{ijk}. \quad (13.9)$$

and show that they are *the same 27 numbers* as in Equation 13.8. First, note that

$$\varepsilon'_{112} = S_{1i}S_{1j}S_{2k}\varepsilon_{ijk}.$$

The sums over i and j involve something antisymmetric under exchange (that is, ε_{ijk}) times something symmetric under exchange (that is, $S_{1i}S_{1j}$). Altogether, the expression is therefore antisymmetric, so it gives zero when summed over i, j . Indeed, we get zero when *any* two indices of ε'_{abc} are equal, in agreement with Equation 13.8.

All that remains, then, is to check the case where i, j, k are all different. In fact, you can readily show that $\varepsilon'_{abc} = -\varepsilon'_{bac}$ and so on, as desired, so we only need to check a single permutation, for example, ε'_{123} . And of the 27 terms being summed in Equation 13.9, all but six are zero:

$$\begin{aligned} \varepsilon'_{123} &= S_{11}S_{22}S_{33} + S_{12}S_{23}S_{31} + S_{13}S_{21}S_{32} - S_{11}S_{23}S_{32} - S_{13}S_{22}S_{31} - S_{12}S_{21}S_{33} \\ &= \det S. \end{aligned}$$

But we know that $\det S = \pm 1$ for any orthogonal matrix (Equation 13.5). Moreover, any two right-handed coordinate systems are related by a rotation. Any rotation can be continuously obtained from the identity, whose determinant is $+1$. The determinant must always equal ± 1 , and it cannot change discontinuously,⁴ so it must be $+1$. Thus $\varepsilon'_{123} = +1$, completing the proof that all components are the same in any right-handed system.

13.3.2 Caveat

Had we used Equation 13.8 in conjunction with a *left*-handed system, then we would have defined a *different* tensor, equal to minus the Levi-Civita tensor. You can see that by reexpressing it in terms of a right-handed system, because in that calculation, $\det S = -1$.⁵ So the definition of the Levi-Civita tensor, as well as anything defined with its help (cross product, curl, vector representation of an area element $d^2\vec{\Sigma}$) requires that we make a convention about which is our “right” hand. For this reason, some authors refer to the “Levi-Civita *pseudotensor*.”⁶ I prefer the viewpoint that ε is a perfectly well defined 3-tensor, once we have made a choice for which coordinate systems we will call right-handed.

⁴After all, the determinant of a matrix is just a polynomial in the entries of that matrix.

⁵An orthogonal matrix with determinant -1 corresponds to a rotation combined with a *reflection* through a plane, or through a point, and therefore reverses the handedness of a coordinate system. Any two left-handed coordinate systems can be continuously connected by a family of rotations, and you can easily find examples where the determinant is -1 , so all must have this property.

⁶Similarly, a vector quantity that depends on a choice of handedness is sometimes disparaged by the prefix “pseudo.” Thus the three numbers we will use to represent the magnetic field and call \vec{B}_i are sometimes said to define a “pseudovector”; also the usual components of angular momentum, angular velocity, and torque are pseudovectors. There are even pseudoscalars, single quantities that change sign upon change of handedness, such as the field that when quantized represents the pion.

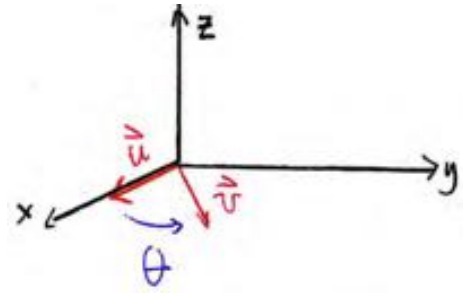


Figure 13.1:

13.4 CONNECT TO ELEMENTARY THINGS

Although the above reasoning is a model for more complicated things to come, it's also good to see how it connects to things you already know.

13.4.1 Dot product

Besides telling us how long a vector is, the metric tensor can tell us the angle between two vectors \vec{v} and \vec{w} . Define the **dot product** as $\frac{1}{2}(\|\vec{v} + \vec{w}\|^2 - \|\vec{v} - \vec{w}\|^2)$. It's a machine that eats *two* vectors and spits out a number that is separately linear in each one (it is “bilinear”). Although I expressed it invariantly, you can quickly see that in any cartesian coordinate system it's given by the usual formula $\vec{v}_i \delta_{ij} \vec{w}_j = v_i w_i$. The same derivation as the one above then assures us that we get the same answer regardless of *which* cartesian coordinate system we chose.

For example, choose a system with \hat{x} parallel to \vec{u} and \vec{v} lying in the xy plane (Figure 13.1). Let θ be the angle between \vec{u} and \vec{v} . Thus $\vec{u} = (1, 0, 0)$ and $\vec{v} = (v \cos \theta, v \sin \theta, 0)$. The sum $\vec{u}_i \vec{v}_i = uv \cos \theta$ as stated in the Prologue to these notes.

13.4.2 Cross product

Because we proved the coordinate-invariance of the Levi-Civita tensor, we know that we can compute $\varepsilon_{ijk} \vec{u}_j \vec{v}_k$ using any right-handed coordinate system we like. The three resulting numbers, interpreted as vector components in the same system, will then define a vector that does not depend on which system we chose. We will call that vector $\vec{u} \times \vec{v}$.

If \vec{v} is parallel to \vec{u} , for example $\vec{v} = \beta \vec{u}$, then the cross product becomes $\beta \varepsilon_{ijk} \vec{u}_j \vec{u}_k$. This is the sum (“contraction”) of something antisymmetric on jk times something symmetric on jk , so it's zero.

If \vec{v} and \vec{u} are not parallel, then as in Section 13.4.1 we may choose a right-handed, cartesian coordinate system with \hat{x} parallel to \vec{u} and \vec{v} in the xy plane. This time, we must be careful to specify that θ is the angle from \vec{u} to \vec{v} , and that θ is taken to be positive if that angle is counterclockwise when viewed along the z axis from positive toward negative values of z (Figure 13.1). Then again $\vec{u} = (u, 0, 0)$ and $\vec{v} = (v \cos \theta, v \sin \theta, 0)$ and

$$(\vec{u} \times \vec{v})_3 = \varepsilon_{31k} u \vec{v}_k = \varepsilon_{312} u (v \sin \theta) = uv \sin \theta, \quad (13.10)$$

as in the Prologue to these notes. (You should show that the other two components of $\vec{u} \times \vec{v}$ equal zero.)

13.5 USEFUL IDENTITIES

13.5.1 Swap dot and cross

The geometrical interpretation of $\vec{u} \cdot (\vec{v} \times \vec{w})$ as a volume makes it clear that this quantity equals $(\vec{u} \times \vec{v}) \cdot \vec{w}$. For practice, you should derive this algebraically by using the properties of the Levi-Civita symbol.

13.5.2 Triple cross product

First note that

$$\varepsilon_{ijk}\varepsilon_{ijk} = \sum_{\text{permutations}} (\pm 1)^2 = 6. \quad (13.11)$$

Next, try the same expression but don't set the last indices equal nor sum them: $\varepsilon_{ijk}\varepsilon_{ij\ell}$ is an invariant symmetric tensor of rank 2, so it must be a multiple of $\delta_{k\ell}$. To find the constant of proportionality, set $k = \ell$, sum over i , and compare to Equation 13.11. This gives

$$\varepsilon_{ijk}\varepsilon_{ij\ell} = 2\delta_{k\ell}. \quad (13.12)$$

Finally, try not setting the last *two* indices equal: $\varepsilon_{ijk}\varepsilon_{im\ell}$ is an invariant tensor of rank 4, and it's antisymmetric upon exchange of jk as well as $m\ell$. But it's symmetric if we swap jk with $m\ell$. Suppose $j = 1$, $k = 2$; then only one term of the sum over i is nonzero, namely $i = 3$. This in turn implies that m, ℓ must be either 12 or 21. For all those reasons, we must have

$$\varepsilon_{ijk}\varepsilon_{im\ell} = M(\delta_{jm}\delta_{k\ell} - \delta_{j\ell}\delta_{km}) \quad \text{for some constant } M. \quad (13.13)$$

To evaluate M , this time set $m = j$ and $\ell = k$, sum both, and again compare to Equation 13.11:

$$6 = \varepsilon_{ijk}\varepsilon_{ijk} = M(\delta_{jj}\delta_{kk} - \delta_{jk}\delta_{kj}) = M(3 \cdot 3 - \delta_{jj}) = 6M.$$

Thus $M = 1$ in Equation 13.13.

Your Turn 13A

Try using one of the three preceding identities to get a familiar formula for $\vec{u} \times (\vec{v} \times \vec{w})$.

13.6 PLUS ULTRA

It may seem that we have gone the long way round the barn to reconstruct something you already knew. But when calculations start to get complicated, the benefits of using ε to express cross products will become clear. Also, the approach used in these notes continues to work in any number of dimensions: For example, we will find it

useful to know that a metric space of dimension 4, with a choice of handedness, gets a rank-4 Levi-Civita tensor “from Heaven,” despite the fact that there is no concept of cross product. The argument is exactly the same as the one in Section 13.3.1.

Finally, understanding ε as a tensor will prove valuable as we seek to reformulate electrodynamics *without* any cross products, thereby making its inversion invariance obvious.

T₂ Track 2

13.3'

1. The appearance of the Levi-Civita tensor in the laws of Nature should bother you! Classical mechanics and electrodynamics are supposed to be *invariant* under spatial inversions, so why do we need any right-hand rule (or equivalently any choice of handedness) to formulate them? The answer is: We don't. Both classical mechanics and electrodynamics can be expressed completely without ever introducing cross products or pseudoquantities. In fact, doing this for electrodynamics is one of our goals in this course.⁷ The same can be said for the strong nuclear interaction, but *not* for weak interactions: For example, when a neutron decays, the outgoing neutrino always has the same helicity. There is a spin operator analogous to ε that appears in the weak interaction, that changes under spatial inversion, and that cannot be removed by redefining things.

2. We don't get a Levi-Civita tensor until we select a "handedness," that is, select one privileged class of cartesian coordinate systems that we call "right-handed." Mathematicians call this a choice an "orientation," but that term can lead to confusion and we won't use it. (Normally a physicist understands the words "change the orientation" to mean "rotate [an object]," not "reverse the handedness convention of space.")

3. What does "from Heaven" mean? Our constructions all relied on choosing cartesian coordinates. In fact, with some more work they can all be generalized to curvilinear coordinates on flat space, or even to curved space; for example, all we need in order to construct an invariant analog of $\vec{\mathbf{1}}$ is a local distance function. That's the first step to formulating electrodynamics on curved space(time), for example, to study diffraction effects in gravitational lensing.

Similarly, the Levi-Civita tensor can be defined on any space with a metric plus a distinction between left- and right-handed coordinate systems: The geometric construction of Section 12.4 works on any such space and does not require any coordinate choice.

T₂ Track 2

13.4'

The main text takes the following attitude:

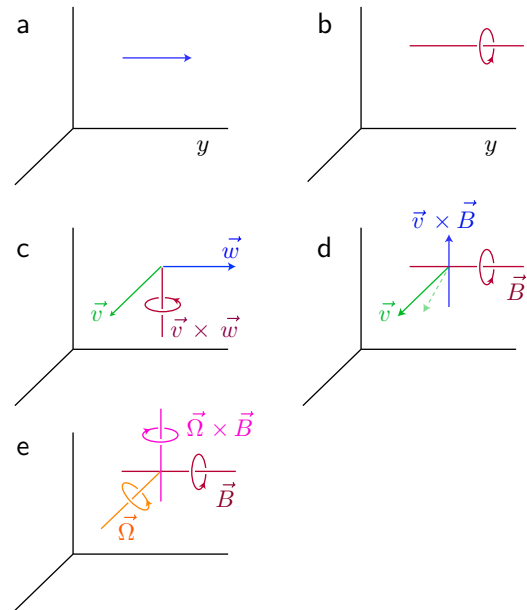
1. Vectors and tensors are real objects with concrete geometrical meaning independent of any choice of coordinate system (they "point").
2. The Levi-Civita tensor, and things constructed with its assistance, are ambiguous defined until we choose an orientation (choice of which hand is "right"). Once such an overall sign choice has been made, however, they become ordinary vectors and tensors.

Actually, however, in three dimensions there is an intriguing reinterpretation of "pseudo" objects that is just as intrinsic (independent of coordinate choice) as ordinary vectors and tensors. For this reason, some authors replace the deprecatory "pseudo" by the more neutral "twisted" to specify these objects.⁸ Thus, angular momentum and magnetic induction \vec{B} are twisted vectors, whereas velocity and force are ordinary vectors.

⁷Doing it for rigid-body dynamics is similarly rewarding, but outside the scope of these notes.

⁸Others speak of "tensor densities."

Figure 13.2: [Diagrams.] **Twisted vectors and their operations.** (a) An ordinary vector. The reflection $y \rightarrow -y$ turns it into minus itself. (b) A twisted vector. The reflection $y \rightarrow -y$ leaves it unchanged. (c) Cross product of vector with vector yields a twisted vector (see text). (d) Cross product of vector with twisted vector yields a vector (see text). (e) Cross product of twisted vector with twisted vector yields a twisted vector (see text).



To visualize an ordinary vector, we draw a line segment, choose one end, and draw an arrowhead on that end. To visualize a twisted vector, we again draw a line segment, but with no arrowhead. Instead, draw a directed *loop* encircling the segment. That loop can run in one of two ways, similar to the fact that we can draw the arrowhead on an ordinary vector in two ways. But contrast the objects in Figure 13.2a–b: One changes sign upon a particular reflection, whereas the other does not.

Of course, if we make a choice of which hand to call “right” then we can associate an ordinary vector to any twisted vector and vice versa. If we don’t make any such choice, we must keep these two categories distinct.

We can now define an intrinsic cross product that does not require any choice of right hand, as long as we keep track of the fact that it adds or removes “twistedness”:

- Given two ordinary vectors, return zero if they are parallel or antiparallel. Otherwise, the vectors determine a plane. Construct a line segment perpendicular to the plane with length $\|\vec{v}\| \|\vec{w}\| \sin \theta$. Imagine a rotation in the plane that turns from the first to the second vector. Instead of trying to put an arrowhead on the perpendicular segment, define the loop encircling it that turns from \vec{v} toward \vec{w} (Figure 13.2c). That choice of loop converts the segment into a twisted vector, which we call $\vec{v} \times \vec{w}$.
- Given a vector \vec{v} and twisted vector \vec{B} , return zero if they are parallel or antiparallel. Otherwise, proceed as above to draw a perpendicular line segment. This time, however, we place an arrowhead on one end of the segment, as follows: Rotate the arrow representing \vec{v} about the segment representing \vec{B} in the sense determined by the loop around it. This brings the arrowhead on \vec{v} closer to one end of the segment (dashed line in Figure 13.2d); place the arrowhead on the other end.⁹
- Given two twisted vectors \vec{B} and $\vec{\Omega}$, return zero if they are parallel or antiparallel.

⁹This construction also lets us associate an antisymmetric rank-2 tensor $\vec{\omega}$ to any twisted vector \vec{B} and vice versa: The tensor takes any vector \vec{v} and returns the vector $\vec{\omega} \cdot \vec{v} = \frac{1}{2} \vec{v} \times \vec{B}$, which is Equation 14.3 (page 161).

Otherwise, proceed as above to draw a perpendicular line segment. There will be a rotation in the plane spanned by the two twisted vectors that superimposes $\vec{\Omega}$'s loop onto that of \vec{B} . That rotation defines a direction for a loop about the perpendicular segment (Figure 13.2e), allowing us to define it as a twisted vector.

Higher rank twisted tensors can also be defined, but it's harder and less useful to make visualizable metaphors for them. For more details see Burke, 1985.

Because our goal is to move away from three dimensions, we will not pursue these constructions further. We regard the magnetic field as an ordinary vector defined with the help of some choice of right-hand convention, and similarly for cross products. Eventually we will eliminate "pseudo" quantities from our formulation of electrodynamics altogether.

PROBLEMS

13.1 *Dots and crosses*

Prove the identity $(\vec{A} \times \vec{B}) \cdot (\vec{C} \times \vec{D}) = (\vec{A} \cdot \vec{C})(\vec{B} \cdot \vec{D}) - (\vec{A} \cdot \vec{D})(\vec{B} \cdot \vec{C})$.

CHAPTER 14

Magnetostatics

Oersted received his PhD in 1799 in the medical faculty of Copenhagen; his topic dealt with Kant's philosophy. . . . [His] discovery, easy to reproduce, was the first direct demonstration of the connection between electricity (a current) and magnetism, and it was first done by accident at the end of a lecture demonstration. Interestingly, Oersted was apparently all "thumbs" in the lab, and all his experiments had to be carried out by his students and assistants.

— *R. M. Clegg*

You are quite right to say that it is inconceivable that for twenty years no one tried the action of the voltaic pile on a magnet. . . . Coulomb's hypothesis on the nature of magnetic action. . . rejected any idea of action between electricity and the so-called magnetic wires. This prohibition was such that when Arago spoke of [Oersted's] phenomena at the Institute, they were rejected. . . . Every one decided that they were impossible.

— *Ampère, to a friend*

We have already started thinking about charges in motion, but we have not yet considered the magnetic fields that they create. This was allowed because (a) we studied slow time variations, where the magnetic fields if any do not react back on the electric fields, and (b) we assumed that the charges were executing specified motions, so that any forces they might get from magnetic fields were unimportant. Nevertheless, magnetic fields do get generated by even slowly-moving charges, so let's start to study that.

14.1 A NEW FORCE AWAKENS

Imagine a steady current through a long, straight wire. There is no net charge anywhere to create any electric field. A test charge outside that wire will feel a kind of force that we have not yet encountered: It differs from the electrostatic force because

- The force is zero unless the test charge is *moving*; and
- The force is always perpendicular to the test charge's velocity.

Indeed, the force on a test charge is a linear function of the velocity. We now know what to call such a machine: It is a second-rank tensor on space:¹

$$(\vec{\text{force}})_i = 2q\vec{\omega}_{ij}\vec{v}_j. \quad (14.1)$$

Given a current distribution in the lab, we can operationally measure $\vec{\omega}$ by throwing a lot of charged test bodies and seeing how they accelerate.

Moreover, the second point above means that the current specifically creates an *antisymmetric* rank-two tensor $\vec{\omega}$. To see this, think about two velocities \vec{v} and \vec{u} . Then $\vec{\omega} \cdot (\vec{v} + \vec{u})$ must be perpendicular to $(\vec{v} + \vec{u})$:

$$0 = (\vec{v} + \vec{u}) \cdot \vec{\omega} \cdot (\vec{v} + \vec{u}) = \vec{v} \cdot \vec{\omega} \cdot \vec{v} + \vec{u} \cdot \vec{\omega} \cdot \vec{u} + \vec{v} \cdot \vec{\omega} \cdot \vec{u} + \vec{u} \cdot \vec{\omega} \cdot \vec{v}.$$

The first two terms are zero by assumption, so the last two must always sum to zero.

Up until now, well-meaning but misguided people have thought you weren't ready for tensors, so they have repackaged the magnetic field: Define the three quantities

$$\vec{B}_i = \varepsilon_{ijk}\vec{\omega}_{jk}. \quad (14.2)$$

In a sense, we lose nothing by this reformulation, because it is invertible: We can always recover $\vec{\omega}$ from \vec{B} .

Your Turn 14A

Show that

$$\vec{\omega}_{im} = \frac{1}{2}\varepsilon_{kim}\vec{B}_k. \quad (14.3)$$

(Where did the factor of 1/2 come from?)

Then \vec{B} looks superficially like a vector field and everyone is comfortable. But there is a terrible price to pay for this approach:

- Equation 14.2 requires us to choose a handedness on space. Using \vec{B} instead of $\vec{\omega}$ then introduces Levi-Civita tensors into our equations of physics, obscuring their inversion symmetry. For example, the force law Equation 14.1 becomes $\vec{f} = q\vec{v} \times \vec{B}$.
- \vec{B} also obscures the *Lorentz* invariance of electrodynamics, which is why it took a genius (Lorentz) to see that property, and *another* genius (Einstein) to see the implications. We will have to abandon \vec{B} later, in order to construct a formulation in which even fools like me can see the full invariance.

Despite those comments, of course we do need to be able to talk to people who use \vec{B} . So we'll need to be able to switch between *both* representations, by using Equations 14.2 and 14.3.

¹We will see later that this formula remains valid in relativistic situations, if we interpret the left side as the time derivative of particle momentum. Putting the factor of 2 in the definition Equation 14.1 is convenient because it makes another 2 elsewhere go away.

14.2 VECTOR POTENTIAL

14.2.1 Preliminary

Let's brush up on a point we'll need soon. Suppose that f is a function of \vec{r} . We can construct a function of *four* variables, $g(u, \vec{r})$, by evaluating f at the point $(u\vec{r})$. Make sure you understand how the chain rule implies that

$$\frac{\partial g}{\partial u} = \frac{\partial f}{\partial \vec{r}_i} \Big|_{u\vec{r}} \frac{\partial (u\vec{r}_i)}{\partial u} = \vec{r} \cdot \vec{\nabla} f \Big|_{u\vec{r}} \quad (14.4)$$

$$\frac{\partial g}{\partial \vec{r}_1} = \frac{\partial f}{\partial \vec{r}_i} \Big|_{u\vec{r}} \frac{\partial (u\vec{r}_i)}{\partial \vec{r}_1} = (\vec{\nabla}_i f \Big|_{u\vec{r}})(u\delta_{i1}) = u\vec{\nabla}_1 f \Big|_{u\vec{r}}, \quad (14.5)$$

and similar results for $\partial g/\partial \vec{r}_{2,3}$. Think about how the indices match on each side of these formulas.

14.2.2 No scalar potential this time

In electrostatics, Maxwell's equations boiled down to just one equation for a potential function. That was handy. It worked because we found a general solution to Faraday's law, $\vec{\nabla} \times \vec{E} = 0$, in terms of ψ , so we could just substitute $\vec{E} = -\vec{\nabla}\psi$ into the Gauss law and *forget* Faraday. Can we duplicate that victory?

At first it looks bad. The magnetic field is not curl-free: Ampère's law says $\vec{\nabla} \times \vec{B} \neq 0$. It's true that sometimes we want to solve for magnetic fields throughout a current-free region, and in such a case we may get some success by introducing a "magnetic scalar potential." But let's instead try to exploit the magnetic Gauss law, $\vec{\nabla} \cdot \vec{B} = 0$, because it's always true.

14.2.3 Revisit electrostatics

The magnetic Gauss law $\vec{\nabla} \cdot \vec{B} = 0$ looks pretty different from $\vec{\nabla} \times \vec{E} = 0$, but surprisingly there is a close analogy. To bring it out, let's return briefly to electrostatics. Previously I used Stokes's theorem and the equation $\vec{\nabla} \times \vec{E} = 0$ to conclude that the line integral of \vec{E} was well defined. Here is more self-contained construction, which also sets us up for a generalization we'll need.

Your Turn 14B

Use an appropriate Levi-Civita identity to show that the curl-free condition is equivalent to

$$\vec{\nabla}_i \vec{E}_k - \vec{\nabla}_k \vec{E}_i = 0 \text{ for any } i \text{ and } k \text{ (stationary case)}. \quad (14.6)$$

As in Chapter 2, choose some arbitrary reference point and center coordinates on it. Again choose a path from the reference point to a given observation point \vec{r} , for example, the straight line $u\vec{r}$ where u ranges from 0 to 1.

Now define

$$\psi(\vec{r}) = - \int_0^1 (\vec{r} du) \cdot \vec{E}(u\vec{r}). \quad (14.7)$$

In this expression, \vec{r} is held constant during the integration over u . Then the negative gradient is (see Equation 14.5)

$$\begin{aligned} -\frac{\partial\psi}{\partial\vec{r}_i} &= \int_0^1 du \left[\vec{E}_m(u\vec{r}) \frac{\partial\vec{r}_m}{\partial\vec{r}_i} + \vec{r}_m \frac{\partial\vec{E}_m}{\partial\vec{r}_k} \Big|_{u\vec{r}} \frac{\partial(u\vec{r}_k)}{\partial\vec{r}_i} \right] \\ &= \int_0^1 du \left[\vec{E}_i(u\vec{r}) + u\vec{r}_m \frac{\partial\vec{E}_m}{\partial\vec{r}_i} \Big|_{u\vec{r}} \right]. \end{aligned}$$

In the last term, we may replace $\partial\vec{E}_m/\partial\vec{r}_i$ by $\partial\vec{E}_i/\partial\vec{r}_m$, thanks to Equation 14.6. We can now use Equation 14.4 to find

$$-\vec{\nabla}_i\psi|_{\vec{r}} = \int_0^1 du \frac{d}{du} [u\vec{E}_i(u\vec{r})] = u\vec{E}_i(u\vec{r})|_0^1 = \vec{E}_i(\vec{r}).$$

Once again, we have established the potential representation for electrostatics. There is still the usual ambiguity in ψ : Adding a constant to ψ , for example, by choosing a different reference point, doesn't change its gradient.

The *payoff* for the potential formulation is that we have fewer and simpler equations to solve. In fact, Chapter 2 got a complete, general solution to electrostatics with a specified charge distribution. The *caveat* is that we'll need to rethink when we go beyond statics, because then $\vec{\nabla} \times \vec{E} \neq 0$.

14.2.4 Poincaré lemma

We'd like an integrability lemma like the one just given, but applicable to magnetism.

Your Turn 14C

Use Equation 14.2 to show that the magnetic Gauss law is equivalent to

$$\varepsilon_{imk} \vec{\nabla}_k \vec{\omega}_{im} = 0. \quad (14.8)$$

That is, when we take all the first derivatives of $\vec{\omega}_{im}$ and *antisymmetrize*, the result is always zero. This resembles Equation 14.6, albeit with an extra index.

Your Turn 14D

Show that, of the six nonzero terms in Equation 14.8, half are redundant; that is, it may be written as

$$\vec{\nabla}_k \vec{\omega}_{im} + (2 \text{ cyclic permutations}) = 0. \quad (14.9)$$

With this preparation, we're ready to generalize Section 14.2.3. Analogously to Equation 14.7, define

$$\vec{A}_i(\vec{r}) = 2 \int_0^1 (u\vec{r}_m du) \vec{\omega}_{mi}(u\vec{r}). \quad (14.10)$$

We now wish to compute the curl of this new vector field, or equivalently

$$\begin{aligned}\vec{\nabla}_k \vec{A}_i - \vec{\nabla}_i \vec{A}_k &= 2 \int_0^1 du u \left[\frac{\partial(\vec{r}_m)}{\partial \vec{r}_k} \vec{\omega}_{mi}(u\vec{r}) + \vec{r}_m \frac{\partial \vec{\omega}_{mi}}{\partial \vec{r}_n} \Big|_{u\vec{r}} \frac{\partial(u\vec{r}_n)}{\partial \vec{r}_k} \right] - (i \rightleftharpoons k) \\ &= 2 \int_0^1 du u \left[(\vec{\omega}_{ki} - \vec{\omega}_{ik}) + u\vec{r}_m \left(\frac{\partial \vec{\omega}_{mi}}{\partial \vec{r}_k} - \frac{\partial \vec{\omega}_{mk}}{\partial \vec{r}_i} \right) \Big|_{u\vec{r}} \right].\end{aligned}$$

The last two terms can be simplified by using Equation 14.9: They equal $-\vec{\nabla}_m \vec{\omega}_{ik} \Big|_{u\vec{r}}$. Thus, we get

$$\begin{aligned}&= 2 \int_0^1 du \left[2u\vec{\omega}_{ki}(u\vec{r}) - u^2 \vec{r}_m \frac{\partial \vec{\omega}_{ik}}{\partial \vec{r}_m} \Big|_{u\vec{r}} \right] \\ &= 2 \int_0^1 du \frac{\partial}{\partial u} \left[u^2 \vec{\omega}_{ki}(u\vec{r}) \right] = 2u^2 \vec{\omega}_{ki}(u\vec{r}) \Big|_0^1 = 2\vec{\omega}_{ki}(\vec{r}).\end{aligned}\quad (14.11)$$

Now tidy things up by recalling the formula for curl and Equation 14.2:

$$(\vec{\nabla} \times \vec{A})_m = \varepsilon_{mki} \vec{\nabla}_k \vec{A}_i = \frac{1}{2} \varepsilon_{mki} (\vec{\nabla}_k \vec{A}_i - \vec{\nabla}_i \vec{A}_k) = \varepsilon_{mki} \vec{\omega}_{ki} = \vec{B}_m.$$

Indeed, we have constructed a vector field \vec{A} whose curl equals \vec{B} .

Our *payoff* is that using potentials in magnetostatics will soon prove to be just as useful as it was in electrostatics. Also, the result we proved works in *any number* of dimensions:

Given any antisymmetric rank-two tensor with the property that its antisymmetrized first derivatives vanish (Equation 14.8), it is always possible to write it as the tensor of antisymmetrized derivatives of a vector field (Equation 14.11).

Poincaré lemma

(14.12)

Later, when we need this result in four dimensions, we won't need to prove this again.²

14.3 GAUGE INVARIANCE

We have found the general solution to the magnetic Gauss law, so we can just substitute $\vec{\nabla} \times \vec{A}$ for \vec{B} into Ampère's law and *forget* about Gauss.³ However, there is an ambiguity in this representation. After all, if we add the gradient of anything, $\vec{A} \rightarrow \vec{\tilde{A}} = \vec{A} + \vec{\nabla} \Xi$, then the curl of \vec{A} doesn't change. So \vec{B} doesn't fully determine its vector potential \vec{A} . This fact is known as **gauge invariance**, though maybe "redundancy" would have been a better term to use. The substitution $\vec{A} \rightarrow \vec{\tilde{A}}$ is called a **gauge transformation** of the vector potential. This is much more freedom than what we had in electrostatics, where adding a *constant* to ψ left \vec{E} unchanged.

²It even works for tensors of rank different from two. For example, for rank 1 it's what we proved in Section 14.2.3.

³Beware that all books use the symbol \vec{A} for magnetic vector potential. But all books also use the same letter A to denote a *different* quantity, the 4-vector potential. I will disambiguate by using \vec{A} for the former and \underline{A} for the latter.

Gauge invariance sounds like a nuisance, but it can be helpful. We can use that freedom to represent a magnetic field by a vector potential that additionally satisfies some other condition (**gauge fixing**). For example, we can always insist that \vec{A} obeys

$$\vec{\nabla} \cdot \vec{A} = 0. \quad \text{Coulomb gauge condition}$$

To see this, suppose that we represent a \vec{B} field by some vector potential that doesn't satisfy Coulomb gauge. If we then gauge transform it we get $\vec{\nabla} \cdot \vec{\tilde{A}} = \vec{\nabla} \cdot \vec{A} + \nabla^2 \Xi$. We just need to choose Ξ to be a function of position that solves the Poisson equation with source given by $-\vec{\nabla} \cdot \vec{A}$. With this choice of gauge transformation, we find that $\vec{\tilde{A}}$ is in Coulomb gauge.

14.4 BACK TO PHYSICS

14.4.1 Steady currents

To keep things simple, let's temporarily consider situations where matter is moving (net charge flux $\vec{j} \neq 0$) but neutral ($\rho_q(\vec{r}) = 0$). Also, we'll restrict to *steady* motion ($\partial \vec{j} / \partial t = 0$). Thus, our system will be invariant under time *translation* (it is **stationary**), though not under time *reversal* (it is not **static**).

This can only be an idealized, approximate situation. Really each electron or proton is pointlike, so as any one passes any point, the electric and magnetic fields pulse. We replace discrete charges by a continuous "river of charge," an approximation that certainly makes sense in a macroscopic apparatus. The overall river can flow steadily if we neglect its granular character in this way. (Don't worry; later we'll get to a fully dynamic formulation.)

Our approach in this course is to take the Maxwell equations as a physical hypothesis and explore their testable consequences. In the situations we have studied so far, their traditional form simplifies to just

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \rho_q / \epsilon_0 = 0 \text{ Gauss (no net charge)} \\ \vec{\nabla} \cdot \vec{B} &= 0 \quad \text{Gauss} \\ \vec{\nabla} \times \vec{B} &= \mu_0 \vec{j} \quad \text{Ampère (stationary case)} \\ \vec{\nabla} \times \vec{E} &= 0 \quad \text{Faraday (stationary case)}. \end{aligned} \quad (14.13)$$

These equations have falsifiable content because \vec{B} has an independent definition: We can measure it throughout space by looking at the motions of test charges, which feel a force $q\vec{v} \times \vec{B}$. Once \vec{B} is measured, we can check if it does or does not obey the above equations for a steady current distribution. Because we're also restricting to the case where $\rho_q = 0$, we also have $\vec{E} = 0$.

In the most basic situation, we can guess a trial solution to Equations 14.13 and adjust it until it works: Consider an infinite straight wire carrying steady current I , uniformly distributed across its cross-section. This situation has so much symmetry that we can try a trial solution where \vec{B} is everywhere pointing radially outward from the wire. That fails. But the next possibility, in which $\vec{B}(\vec{r}) = f(r)\hat{\phi}$, is also axially

symmetric and more promising. We integrate Ampère's law over a disk of radius R perpendicular to and centered on the wire:

$$\begin{aligned}\int d^2\vec{\Sigma} \cdot (\vec{\nabla} \times \vec{B}) &= \mu_0 \int d^2\vec{\Sigma} \cdot \vec{j} = \mu_0 I \\ &= \oint \vec{B} \cdot \vec{r} = 2\pi R f(R).\end{aligned}$$

So $\vec{B}(\vec{r}) = \hat{\phi}\mu_0 I/(2\pi R)$, the famous answer.

Other problems are harder than this one. We need a more systematic approach.

14.4.2 General solution

Section 14.3 showed that any magnetic field can be represented in terms of a divergence-free vector potential.

Your Turn 14E

To see the power of this observation, show that Ampère's law may be written as⁴

$$\nabla^2 \vec{A} = -\mu_0 \vec{j}. \quad \text{in Coulomb gauge} \quad (14.14)$$

That scary vector partial differential equation has magically separated into *three independent copies* of the Poisson equation. And we already know how to solve the Poisson equation, from electrostatics (Equation 2.5)! For each component of \vec{j} , compute⁵

$$\vec{A}_i(\vec{r}) = \mu_0 \int d^3r_* \frac{\vec{j}_i(\vec{r}_*)}{4\pi\|\vec{r} - \vec{r}_*\|}. \quad (14.15)$$

So we just *finished* magnetostatics: Given a steady current distribution, solve Equation 14.14 for the three components of \vec{A} . Then compute the curl to get \vec{B} .

14.4.3 Self-consistency

Before we accept Equation 14.15, we should check that it really is a potential in Coulomb gauge. If not, then the fact that it solves Equation 14.14 is irrelevant, because Equation 14.14 is *not Ampère's law* except in Coulomb gauge!

Your Turn 14F

Work out the divergence of the expression in Equation 14.15. [*Hint: Use the continuity equation to show that $\vec{\nabla} \cdot \vec{j}$ must always be zero in a steady situation.*]

⁴The notation $\nabla^2 \vec{A}$ means that we apply the Laplace operator to each component of \vec{A} and interpret the results as the components of a vector. This operation only makes sense in cartesian coordinates; a more elaborate form of the derivation is needed in curvilinear coordinates.

⁵This is P+S eq. 8.60.

14.4.4 Counting equations

The equations of electro- and magnetostatics (Equation 14.13) appear to be eight equations in the six unknown functions \vec{E}_i, \vec{B}_i , an issue first raised in Hanging Question #D (page 11). And yet, we have reformulated electrostatics as one equation in one unknown: the Poisson equation (Equation 2.3, page 22), and magnetostatics boiled down to three Poisson equations for the three components of \vec{A} (Equation 14.14). So at least in statics, our puzzle has disappeared: We have four equations in four unknown potential functions.

In fact, two of the eight Maxwell equations are *identities*; they do not constrain the fields and hence should not be included in our count. For example, taking the divergence of both sides of the Faraday law gives $0 = 0$ identically, regardless of what \vec{E} may be. Similarly, taking the divergence of both sides of Ampère's law (Equation 14.13) gives

$$\vec{\nabla} \cdot (\vec{\nabla} \times \vec{B}) = \vec{\nabla} \cdot \vec{j} \quad (\text{stationary case}).$$

The left side is identically zero regardless of what \vec{B} may be; the right side is also identically zero in magnetostatics (Your Turn 14F). Again, we end up with equal numbers of unknowns (the six field components) and equations (the remaining six Maxwell equations).

14.5 DOWN FROM THE MOUNTAIN: THE OERSTED PROBLEM

The previous sections got a bit abstract. Let's see how the story plays out in a familiar problem. As in Section 14.4.1, suppose that a thin wire carries current I . It stretches along the entire z axis. Thus, its charge flux is

$$\vec{j}(\vec{r}) = I\delta^{(2)}(\vec{r}_\perp)\hat{z}. \quad (14.16)$$

Here \vec{r}_\perp denotes the two-component vector $\begin{bmatrix} x \\ y \end{bmatrix}$. Each delta function contributes a dimension \mathbb{L}^{-1} , so this expression has dimensions appropriate for a charge flux.⁶ You already found the resulting magnetic field in Section 14.4.1.

Your Turn 14G

- Do it again, this time by solving Equation 14.14 with source given by Equation 14.16 and then computing the curl of your answer. Make sure it's what we already found in Section 14.4.1. [*Hint*: The Green-function solution given in Section 14.4.2 isn't the easiest way to do this problem, which has lots of useful symmetry. Instead, make a Good Guess, then check and adjust it.]
- Confirm that the vector potential you found really is in Coulomb gauge.

14.6 BIOT-SAVART LAW

⁶See Section 0.3.6.

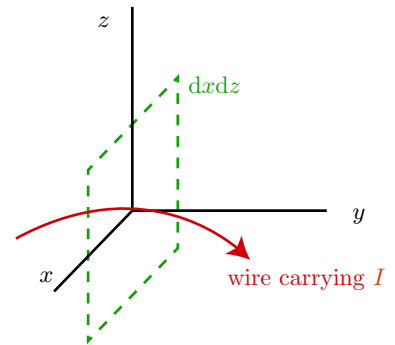


Figure 14.1:

14.6.1 General form

We can now find the magnetic field created by an arbitrary current distribution by computing the curl of Equation 14.15.

Your Turn 14H

Show that

$$\vec{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int d^3r_* \vec{j}(\vec{r}_*) \times \frac{\vec{r} - \vec{r}_*}{\|\vec{r} - \vec{r}_*\|^3}. \quad (14.17)$$

This is a generalization of the usual Biot–Savart formula to cover an arbitrary current distribution (not necessarily confined to a thin wire).

14.6.2 More general wires

Sometimes it is appropriate to consider a limiting case in which \vec{j} is everywhere zero except along a mathematical curve (a “thin wire”). We already considered the simplest case in Equation 14.16.

In a static situation, the total current I through any cross-section of that wire has everywhere a constant value, by the continuity equation. Suppose that the wire is described by a parameterized curve in space $\vec{\ell}(s)$. For example, we could choose s to be arclength along the curve. Then at any point s_0 the current is flowing parallel to the tangent vector, that is, to the unit tangent $d\vec{\ell}/ds|_{s_0}$.

Start by considering just one chunk of wire, of length ds and centered at s_0 (Figure 14.1). Choose a coordinate system centered on $\vec{\ell}(s_0)$, and rotated so that the tangent lies along \hat{y} . Chapter 7 explained how to find the y -component of the charge flux: Find the net charge piercing the element shown in the figure in time dt , and divide by $dxdzdt$. That charge equals $I dt$ if the element $dxdz$ includes the wire (the origin); otherwise it’s zero. Thus, our chunk has

$$\vec{j}_2(t, x, 0, z) = I \delta(x) \delta(z) \hat{y}. \quad (14.18)$$

Think about why this formula has the units appropriate for a charge flux.

We can now make our formula less dependent on a specific choice of coordinates. First, notice that the one chunk of wire we considered is also confined to a limited

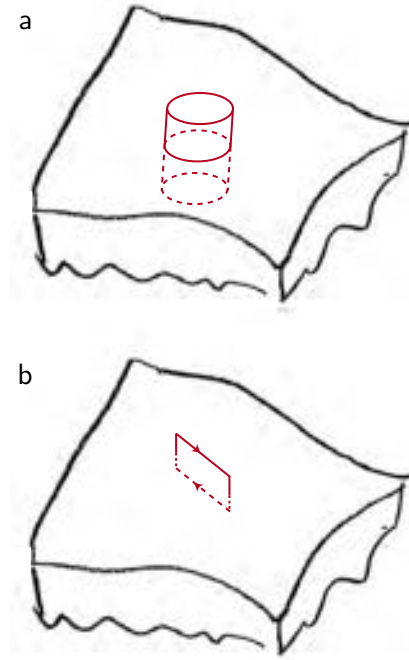


Figure 14.2: [Sketches.] **Boundary conditions near an interface.** (a) The short *red cylinder* has one end cap just outside a material and the other just inside. Integrating the magnetic Gauss law over it, and using the divergence theorem, shows that the component of \vec{B} perpendicular to the surface must be the same just inside and outside the material (Equation 14.20). (b) The shallow *red rectangle* has one of its longer edges just outside a material and the other just inside. Integrating Ampère’s law, and using Stokes’s theorem, shows that any component of \vec{B} parallel to the surface may jump if there is a surface current layer (Equation 14.21).

range $dy = (dy/ds)ds$. With that observation, we get the more general form

$$d\vec{j}(\vec{r}) = I\delta^{(3)}(\vec{r} - \vec{\ell}(s))\frac{d\vec{\ell}}{ds}ds. \quad \text{short segment of thin wire} \quad (14.19)$$

This formula has the same dimensions as Equation 14.18, but it’s no longer restricted to any special coordinate system, nor to one particular point on the wire. To get the charge flux for the entire wire, integrate Equation 14.19 over it.⁷

Your Turn 14I

Substitute Equation 14.19 into Equation 14.17 and recover the usual form of the Biot–Savart law.

14.7 BOUNDARY CONDITIONS

Regardless of whether we use the potential formalism, the magnetic Gauss law implies a jump condition for the magnetic field across a boundary, similar to the one in electrostatics⁸ but without any dependence on the behavior of charges or currents at the surface:

$$\Delta B_{\perp} = 0. \quad (14.20)$$

This fact can be especially useful if we know that the magnetic field is zero on one side. For example, superconductors exclude magnetic fields, so $B_{\perp} = 0$ just outside as well.

⁷Compare P+S eqn. 8.21.

⁸See Figure 14.2a and Section 5.10.

Similarly, integrating Ampère’s law around a loop near the surface gives a condition on the tangential component of \vec{B} (Figure 14.2b). We must allow for the possibility of currents confined to the surfaces of one or both of the media, so let $\vec{j}^{(2D)}$ denote the net 2D charge flux⁹ (with units A/m).

Your Turn 14J

Show that

$$\Delta \vec{B}_{\parallel} = \mu_0 \vec{j}^{(2D)} \times \hat{n}. \quad (14.21)$$

Here $\Delta \vec{B}_{\parallel} = (\vec{B}^{[2]} - \vec{B}^{[1]})_{\parallel}$ and \hat{n} is the unit normal vector pointing from region 1 to region 2.

14.8 MAGNETOENCEPHALOGRAPHY

[Not ready yet.]

14.9 PLUS ULTRA

We found the general solution to magnetostatics with specified, steady currents. But we actually got much more: We also found a simplified formulation of the equations that involves a potential (in this case a vector potential), and it *always works*, even in nonsteady situations (because $\vec{\nabla} \cdot \vec{B} = 0$ always). Soon we’ll find an even more powerful object that combines the vector potential with electric potential, and that, unlike our previous construction of ψ , remains valid beyond statics.

⁹Some authors call this quantity “surface current density.”

T₂ Track 2

14.9'a What about magnetic monopoles?

The magnetic monopoles predicted by grand unified theories, if observed, would seem to invalidate the discussion in Section 14.2.2: A point source of \vec{B} implies that $\vec{\nabla} \cdot \vec{B} = 0$ somewhere. Indeed, inside such hypothetical objects there is always a region in which classical electrodynamics breaks down altogether (other fields like the ones associated to the W and Z bosons have nonzero expectation values).¹⁰ Magnetic monopoles haven't (yet) been observed experimentally in free space.¹¹

Quite apart from such theoretical concerns, E. Parker realized that the observed filamentous structures in distant galaxies is evidence for large-scale magnetic fields, and that this observation in turn implies a severe bound on the hypothetical existence of magnetic monopoles. Just as free electric charges screen electric fields in a conductor, so also free magnetic charges (if they existed) would screen magnetic fields. The fact that such fields are observed (hence not screened) then implies a limit on the flux of free magnetic charges. For a review of magnetic monopoles and flux limits, see, e.g, J Preskill, *Annu. Rev. Nucl. Part. Sci.* **34**:46(1984).

14.9'b Against pseudovectors

The main text pointed out the conceptual benefits of formulating magnetic effects in terms of the tensor $\vec{\omega}$, not the traditional \vec{B} . In fact, *every* "pseudovector" quantity in classical physics, including angular velocity and angular momentum, can be eliminated in favor of tensor quantities, whereupon all the cross products appearing in rigid body dynamics etc. disappear and everything is manifestly inversion-invariant.

It is true that we can conveniently draw (visualize) a vector field. But nothing is really happening along \vec{B} : A moving particle does not feel any force in that direction, nor does it create a \vec{B} that points in that direction.

Is this distinction just Puritanical fussiness? First, I'll point out that you never see any physics formulas involving the sum $\vec{E} + c\vec{B}$, any more than you ever see people adding momentum to angular momentum (or temperature to velocity). Temperature and velocity have different tensor structures; it's not meaningful to add them, and the same for electric and magnetic fields.¹² It's a quirk of three dimensions that they happen to have the same numbers of components, but nevertheless they are different. Second, Section 14.2.4 showed a deep analogy that only becomes apparent when we *abandon* the *superficial* analogy obtained by representing magnetism by \vec{B} . Third, and most important, everybody does agree that \vec{B} has got to be scrapped when we unify electricity and magnetism and reformulate the theory relativistically. Our destination is a formulation in which invariance under Lorentz transformations is explicit; when we arrive there, we'll find that explicit invariance under inversions has come along for free.

Until that happy day, notice that in Equation 14.13, the Gauss law doesn't care about the ambiguous sign of \vec{B} . Ampère's law looks suspect because its right-hand side isn't zero, but the left side has *two* sign changes if we switch handedness conventions, so it, too, is secretly invariant.

¹⁰See Problem 16.2.

¹¹There may be collective excitations in condensed matter with this character.

¹²OK, I admit I have seen exotic articles that introduce $\vec{E} + i\vec{B}$. The formulas in those articles are not invariant under inversions.

14.9'c Totally antisymmetric tensors are so useful...

that mathematicians have a separate name for them: **differential forms** of rank p , or just “ p -forms” for short. Standard mathematical notation abbreviates by omitting the indices and overarrows; you must remember the tensor character of each symbol from its original definition. The totally antisymmetrized first derivatives of such a tensor form a similar object of rank $p + 1$, called the “exterior derivative” and denoted by the very concise symbol d . The exterior derivative operator has the property that $d^2 = 0$.

The Poincaré Lemma thus states that if $d\omega = 0$, then we may write $\omega = dA$ for some $(p-1)$ -form A . There is an important caveat: This result is valid only on a contractible region of space. (On a torus, for example, we would not be able to choose an unambiguous path to each \vec{r} in the way we did above, and different choices of path are not guaranteed to give answers that agree.) The study of exactly how the Poincaré lemma fails on a topologically nontrivial space is called **deRham cohomology**.

In this language:

- The existence of an electrostatic potential is the case $p = 1$. The Maxwell equation $dE = 0$ implies that we may write $E = -d\psi$.
- For magnetostatics we need the case $p = 2$: We found that the Gauss law for magnetic fields can be elegantly written as $d\omega = 0$,¹³ which implies that we may write $\omega = dA$. There’s an ambiguity: Because $d^2 = 0$, we may add any gradient $d\Xi$ to the vector potential A without altering dA . That’s gauge invariance.¹⁴
- Some exotic field theories derived from superstrings involve higher-rank antisymmetric tensor fields. They, too, are subject to the Poincaré lemma.

14.9'd What about angular momentum conservation?

Maybe you recall from first-year physics that the proof of angular momentum conservation, as presented even in the *Feynman Lectures*, involves the assumption that every force on any particle is directed along the line joining that particle to another one. That certainly does not seem to be the case with magnetic forces! What happens to angular momentum conservation? Chapter 34 will get back to this, but the spoiler is: It survives, once we correctly attribute angular momentum to the fields themselves.

¹³I leave it to you to find equally elegant forms of Ampère’s law and the Stokes theorem.

¹⁴For example, changing the base point used in Equation 14.10 results in a gauge-transformed \tilde{A} .

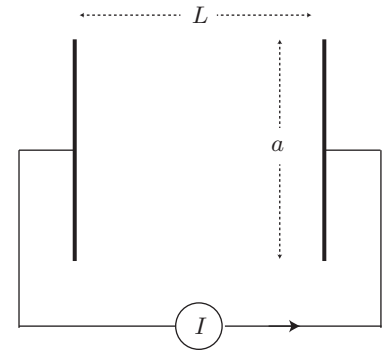
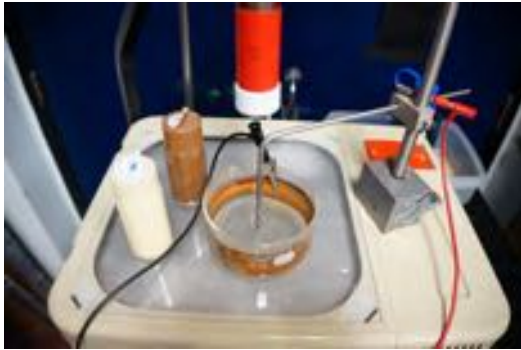


Figure 14.3: (a) An experiment to demonstrate magneto-electrophoresis. Pepper is sprinkled on the surface of a salt-water solution to visualize bulk flow. A central electrode sends direct current outward to a ring-shaped electrode at the rim of the dish. A magnet pole can be brought toward the surface (*red*). (b) Simplified 1D geometry for Problem 14.2.

PROBLEMS

14.1 *Jaws*

Let us explore a possible mechanism for sharks to navigate using Earth's magnetic field. Given that a shark can detect an *electric* field strength of $0.5 \mu\text{V}/\text{m}$, how fast would it have to swim through Earth's *magnetic* field to experience an equivalent force on a charged test particle? Can sharks really swim that fast?

14.2 *Salt and pepper*

Figure 14.3a shows a demo involving salt, pepper, and an overhead projector. It may seem remarkable that those tiny little ions could pull hard enough on the surrounding water to get it into bulk (macroscopic) motion. Let's make some estimates.

The demo setup shown has a circular geometry. But to simplify the math, in this problem instead imagine a rectangular geometry (Figure 14.3b): Current passes between two parallel plates separated horizontally by distance $L = 5 \text{ cm}$. The plates have width $w = 5 \text{ cm}$ and are immersed in a solution with depth $h = 1 \text{ cm}$. The water between the plates contains sodium and chloride ions, each at concentration (ions per volume) c . Each ion carries electric charge $\pm e$, where e is the charge on a proton (the pepper is irrelevant). The solution consisted of about one gram of NaCl dissolved in volume Lhw of water.

A total current of $I = 1 \text{ A}$ passes through the solution. In time dt , ions of each species migrate an average distance $v_{\pm}dt$ toward or away from the $+$ electrode. Thus, all $-$ charges originally in a layer with this thickness near the electrode arrive there and deposit negative charge; similarly, all $+$ charges originally in that layer move away and get replaced by new $+$ charges from the electrode. In all, net charge Idt leaves the $+$ electrode.

- Write a formula that connects v_{\pm} , I , and other quantities in the problem, and solve it for v_{\pm} . (Don't evaluate it numerically yet.)
- Now imagine adding a uniform magnetic field perpendicular to the plane of the

picture, with strength $B = 0.03 \text{ T}$. Write a formula for the resulting magnetic force on a single ion of each species. Then convert this to a formula for the total force per unit volume. (Still don't evaluate yet.)

- c. Multiply your result for (b) by the volume of the chamber to get the total force and evaluate it. Is it big enough to drive the slow but visible motion we observed?

14.3 Parity

- a. The Maxwell/Lorentz equations, in the traditional form (Sects. 1.1-1.2 of the Formula Sheet), are manifestly invariant under spatial rotations, because they involve constructions that we showed were invariant. They are also invariant under spatial inversions (parity), but this is not quite so obvious, because they involve the Levi-Civita tensor, which is not parity invariant. Find a reformulation of these equations that involves no ε_{ijk} factors, by re-expressing the magnetic field \vec{B} in terms of $\vec{\omega}_{ij} = \varepsilon_{ijk} \vec{B}_k$.
- b. Similar criticisms can be leveled at rigid body mechanics, which is also parity invariant, yet full of cross products. Construct second-rank tensors that are dual to the usual angular velocity, angular momentum, and torque. Consider a rigid body with some mass distribution $\rho_m(\vec{r})$. Write the law of motion for the rate of change of your angular momentum, without using any cross products.

14.4 [Not ready yet.]

14.5 Helmholtz coils [Not ready yet.]

14.6 3D magnetic field line plot

- a. Consider a circular loop of wire in the xy plane, of radius a and carrying a steady current. The magnetic field that it creates, when evaluated anywhere in the yz plane, itself lies in that plane. Hence the streamline that passes through any point in that plane remains confined to it. Learn how to get a computer to create 2D streamplots, and use it to show a representative collection of magnetic field lines in the yz plane.
- b. Learn how to get a computer to create 3D streamplots, and show a representative sample throughout space for the same system. Look at various viewing angles till you find one that is most informative.

[Hint for both parts: Replacing \vec{B} by $\vec{B}/\|\vec{B}\|$ will not change the field lines, but it will change the *parameterization* of the curves in space that you'll find. Specifically, this transformation will ensure that your streamlines are parameterized by arclength, which may help your computer to find them more readily.]

CHAPTER 15

Units in Electrodynamics

So far, we have been using SI units without much discussion. Some day you may need to read an article by (or even talk to) a person who uses some other system of units. Many books cheerfully suggest that everything will be fine if you just use a confusing table of conversions. It's better to know how to obtain those conversions for yourself.

15.1

There are several systems of units for electrodynamics, but today you mainly see just two: SI units and “gaussian units,” sometimes called the “CGS” system.¹ Of course, Nature doesn't care what units we use; the formulas we will write, like any correct physical formulas, must be valid in any system of units! So what's the problem?

One problem is that many students don't get the full benefit of dimensional analysis, in part because they've learned it in a haphazard way. Another problem, specific to electrodynamics, is that it turns out that the difference between the SI and gaussian systems actually involves *three independent* issues, only one of which is strictly about which base units to use. I put quotation marks around “gaussian units” above to remind us of this fact. Once we separate the issues and state them explicitly, it becomes fairly easy to see how to convert between different authors' work.

The three issues are (i) choice of base units, (ii) choice of whether to eliminate charge units, and (iii) choice of what physical quantity we use to represent the magnetic induction. (15.1)

15.2 UNITS IN MECHANICS

Just about every useful, correct, thing you've ever learned about units in mechanics can be systematized via a simple maxim:

*Almost any physical quantity should be regarded as the **product** of a pure number times one or more units.² A unit can be regarded as a symbol representing an unknown quantity, just as we use the letter x for an unknown number.*

¹Actually, gaussian units are just one of several systems using cm, g, and s. The others are much less often used, so frequently people say “CGS” to mean specifically gaussian. My goal here is to explain “gaussian units.”

²“Almost” because a few are dimensionless. Also, one quantity (temperature) is sometimes expressed with an offset that complicates its conversions.

Again: The units are part of the quantity. We carry these unit symbols along throughout our calculations. They behave just like any other multiplicative factor; for example, a unit can cancel if it appears in the numerator and denominator of an expression. Although they are unknowns, we do know certain relations among them; for example, we know that $1 \text{ inch} \approx 2.54 \text{ cm}$. Dividing both sides of this formula by the numeric part 2.54, we find $0.39 \text{ inch} \approx 1 \text{ cm}$, and so on.

In mechanics, all we really need to choose are **base units**, those for length, time, and mass. Two popular choices are {meter, kilogram, second} and {centimeter, gram, second}. This choice is the first of the three issues listed in (15.1) above. Other quantities, like force, can then be expressed as products of base units raised to various powers.

In newtonian physics, there is no universal speed, so units of length and time are independent, that is, both are arbitrary.³ In special relativity, there *is* a universal speed, so we could agree always to use unit pairs for length and time that are related by c , like meters and meters/ c (or lightyears and years). It's often inconvenient to do this because human-size units like meters and seconds are far from this relation. In other words, if we define the tick by $1 \text{ tick} = 1 \text{ m}/c$, then $\text{tick} \approx 3 \text{ ns}$, not a very human-size unit.

However, I made that point to introduce the issue (*ii*) mentioned in (15.1). First let me make explicit some options for how to proceed.

A. No elimination of units: A physical quantity like force, or a constant of Nature like Newton's gravitational constant, is the product of a number times some units. The numerical part changes if we change units. For example, $c \approx 3.0 \cdot 10^5 \text{ km/s} \approx 186\,000 \text{ mile/hour}$. We refer collectively to **m**, **cm**, **inches** . . . as different units for the same "dimension," which we denote generically by \mathbb{L} . Similarly time and mass have generic dimensions called \mathbb{T} and \mathbb{M} respectively. Any valid formula, like $f = ma$, involves an equality between quantities with the same dimensions, *and is valid in any set of units*. We can simplify formulas using the definitions $1 \text{ m} = 100 \text{ cm}$ etc., which are themselves valid formulas because both sides have the same dimensions.

Here is another, equivalent approach:

B1. Elimination of all units: Alternatively, we could choose once and for all a set of base units, and agree to express everything in terms of them. For example, if we choose SI base units then in place of acceleration a , we define $\bar{a} = a/(1 \text{ m/s}^2)$, which is the numerical part of a . Similarly $\bar{c} \approx 3 \cdot 10^8$. Now our formulas, expressed in terms of the barred quantities, contain no units at all. But those formulas are only valid if we consistently use the stated system.

Suppose we were asked to find a force. After we do our calculations, we wind up with a numerical value for \bar{f} . Knowing the meaning of force, we interpret this number as the actual force in newtons, because the combination of SI base units with the dimensions of force (MLT^{-2}) is kg m/s^2 , which is called newton. And we get the same final answer as in approach **A**—if we didn't make any errors along the way.

The virtue of approach **B** is that our formulas are compact. Moreover, if we choose

³Nor is there any natural scale of length (or of time). Perhaps we could declare the size of a hydrogen atom as our length unit, but it's not determined by newtonian physics, so this choice doesn't simplify anything.

our base unit of time to be m/c , not seconds, then we find $\bar{c} = 1$, so *we can drop all the factors of \bar{c} from our formulas*, abbreviating them still further. The disadvantage of approach **B** is that we forfeit the real benefit of dimensional analysis: Dimensional analysis expresses certain homogeneity (rescaling) properties of Nature, which appear as redundancies we can use to spot our errors. Eliminating units removes this helpful mechanism for checking our work.

We can compromise and take this process only partway:

B2. Elimination of some units: We agree to measure length in some unit \mathbb{X} and time in \mathbb{X}/c (for example, meters and “ticks,” where 1 tick = $1 m/c$); we don’t commit to any particular unit for mass. For each physical quantity X , we define \bar{X} to be X divided by as many powers of c as are needed to eliminate the time dimensions. Thus, all barred quantities have dimensions that are powers of \mathbb{L} and \mathbb{M} : We have only eliminated time units. We again have the virtue that $\bar{c} = 1$, so we needn’t write it. But we have also retained some of our error-checking abilities.

For example, we have force $\bar{f} = f/c^2$, mass $\bar{m} = m$, acceleration $\bar{a} = a/c^2$, and energy $\bar{\mathcal{E}} = \mathcal{E}/c^2$. Then some famous formulas become

$$\bar{f} = \bar{m}\bar{a}; \quad \bar{\mathcal{E}} = \bar{m}.$$

We’ll use approach **A** exclusively. Students often tacitly use **B**, e.g. when working exams, and so miss errors they would have caught by carrying the units explicitly. Generally **A** involves more writing, but it’s worth it.

Section 15.3.3 below outlines how “gaussian” authors use a variant of approach **B2**.

15.3 UNITS IN ELECTRODYNAMICS

Classical electrodynamics acknowledges no natural unit of electric charge.⁴ So we have to make an arbitrary choice of charge unit. There’s no human intuition for charge, so we don’t feel constrained to use “human sized” units. Some choices are nicer than others; we will explore two popular choices below.

Because charge units are arbitrary, we assign a new dimension symbol \mathbb{Q} to charge. Below we’ll discuss two popular choices called coulombs and statcoulombs.

⁴It’s a mysterious true fact that all isolated, fundamental charged particles do appear to have charges equal to an integer times the proton charge e . But making e our base unit would not be very convenient for lab work. Besides, both quarks and fractional quantum Hall effect excitations are also “fundamental,” and their charges are not integers times e .

15.3.1 SI units

Here are the Maxwell equations as stated in the Prologue:

$$\begin{aligned}\vec{\nabla} \cdot \vec{E} &= \rho_q / \epsilon_0 && \text{Gauss} \\ \vec{\nabla} \cdot \vec{B} &= 0 && \text{Gauss} \\ \vec{\nabla} \times \vec{E} + \dot{\vec{B}} &= 0 && \text{Faraday} \\ \vec{\nabla} \times \vec{B} - \mu_0 \epsilon_0 \dot{\vec{E}} &= \mu_0 \vec{j} && \text{Ampère.}\end{aligned}$$

[0.1–0.4, page 2]

and the Lorentz force law:

$$d\vec{p}/dt = q(\vec{E} + \vec{v} \times \vec{B}). \quad [0.5, \text{page 3}]$$

We had to introduce the two constants of Nature, ϵ_0 (the **permittivity of vacuum**) and μ_0 (**magnetic permeability of vacuum**), in order for the dimensions to work out in Eqns. 0.1–0.5.

The dimensions of the electric and magnetic fields follow from the Lorentz force law:

$$\vec{E} \sim \frac{\text{ML}}{\text{T}^2 \text{Q}} \quad \text{and} \quad \vec{B} \sim \vec{E}/c \sim \frac{\text{M}}{\text{TQ}}.$$

Here “ \sim ” means “has the same dimensions as.”⁵

The Gauss and Ampère laws then give the units of ϵ_0 and μ_0 :

$$\epsilon_0 \sim \frac{\text{Q}}{\text{L}^3} \frac{\text{T}^2 \text{Q}}{\text{M}} \sim \frac{\text{Q}^2 \text{T}^2}{\text{L}^3 \text{M}} \quad \mu_0 \sim \frac{\text{M}}{\text{TQ L}} \frac{\text{L}^2 \text{T}}{\text{Q}} \sim \frac{\text{ML}}{\text{Q}^2}.$$

Because these physical constants involve charge dimensions, their numerical parts will depend on what we choose for our unit of charge. We can use this freedom to arrange that the numerical part of *one of* the three quantities μ_0 , ϵ_0 , or the proton charge e has an exactly specified numerical part. Once we do that, then there’s no more freedom; the *other two* have numerical parts set by Nature, which we can only measure and quote to a certain number of significant figures.

The SI system uses **m**, **kg**, **s**, and **coul** as base units. Prior to 2019, the SI defined the **coulomb** by requiring that μ_0 have the exact value

$$\mu_0 = 4\pi \cdot 10^{-7} \frac{\text{m kg}}{\text{coul}^2} \quad (\text{obsolete}).$$

A more straightforward approach would be to declare that the proton charge is exactly one “attopenn.” Then the numerical values of ϵ_0 and μ_0 , expressed using “penns,” would both need to be measured in the lab. Indeed, after May 2019, the SI takes an approach equivalent to this one, though sadly not using that cool name: The coulomb is now defined as

$$1 \text{ coul} = e/1.602\,176\,634 \cdot 10^{-19} \quad \text{exact.}$$

⁵The symbol “ \approx ” means “is approximately equal to.”

That strange, but exact, multiple was chosen to make old and new definitions nearly equivalent in practice.⁶ Again, this means that μ_0 no longer has an exact value; instead, it is now a measured quantity, with experimental error:⁷

$$\mu_0 \approx (4\pi)(1.000\,000\,000\,82 \cdot 10^{-7}) \text{ m kg coul}^{-2}. \quad (15.2)$$

Thus, μ_0 and ϵ_0 are now on equivalent logical footing: Both are measured constants of Nature, not conventional values.

The **ampere** is still defined by $1 \text{ A} \equiv 1 \text{ coul/s}$, but using the revised coulomb. Physically, Equation 15.2 says that if two long, parallel wires each carry current 1 A and are separated by a meter of vacuum, then the force per length that they exert on each other is about $2 \cdot 10^{-7} \text{ N/m}$. That may seem like peculiar choice, but it implies a convenient magnitude for the ampere: It's approximately the current through a 100 W light bulb.⁸

Once we choose a unit of charge, there is no more freedom. The permittivity of vacuum ϵ_0 then has a measured value, namely $\epsilon_0 \approx 8.854187817 \cdot 10^{-12} \text{ coul}^2 \text{ N}^{-1} \text{ m}^{-2}$.

Note, however, that there's a combination of constants called c , with no dependence on the chosen unit of charge:

$$c \equiv (\mu_0 \epsilon_0)^{-1/2} \sim \frac{\mathbb{L}}{\mathbb{T}}. \quad (15.3)$$

In fact, the SI also defines the meter in a way that gives c the exact value

$$c = 299\,792\,458 \text{ m/s} \quad \textit{exact}.$$

15.3.2 Derived units

Starting from the base units coul, m, kg, and s, various useful combinations have been given names:⁹

⁶You may wonder why the proton is privileged among all the many fundamental particles. Remarkably, *every* known, isolable, fundamental particle has charge that is an integer multiple of e . Even quarks, which are not isolable, and quasiparticles in condensed matter are rational multiples of e . The Standard Model of particle physics offers no necessary reason for this numerical coincidence; explaining it was one of the original motivations behind grand unification, which to date has not been confirmed experimentally.

⁷The unit m kg coul^{-2} is also written H m^{-1} . Note that $\mu_0 = (4\pi\hbar/e^2c)\alpha$ where α is the “fine structure constant.” The factors in parentheses now all have exact conventional values, so the relative uncertainty in μ_0 is the same as that of α . See Davis, 2017.

⁸In the USA system of 110 volt mains. Also, the total charge delivered in a lightning strike is of order 1 coul.

⁹SI style police say to use V where I use volt and C where I use coul. I'd rather write these out than risk the confusion of a one-letter abbreviation. On the chalkboard, V could look like a volume; C could look like the speed of light.

Quantity	Symbol	Unit Name	Abbrev.	Alternative	Alternative
charge	q	coulomb	coul	C	
current	I	ampere	A	coul/s	
magnetic induction	\vec{B}	tesla	T	kg/(coul · s)	volt · s/m ²
magnetic flux	Φ_B	weber	Wb	kg · m ² /(coul · s)	volt · s
electric field	\vec{E}	—	—	kg · m/(coul · s ²)	volt/m
electric potential	ψ	volt	volt	V	J/coul
charge density	ρ_q	—	—	coul/m ³	
charge flux	\vec{j}	—	—	A/m ²	coul/m ² · s
inductance	L	henry	H	kg · m ² /coul ²	J/A ²
capacitance	C	farad	F	s ² · coul ² /(kg · m ²)	coul ² /J or coul/volt
electric dipole moment	\mathcal{D}_E	debye	debye	D	10 ⁻²¹ coul m ² s ⁻¹ /c
resistance	R	ohm	Ω	volt/A	J · s/coul ²
conductance	G	siemens=mho	S = \mathcal{U}	Ω^{-1}	

Hence ϵ_0 can also be written as $\approx 8.85 \cdot 10^{-12}$ F/m, and $\mu_0 \approx 4\pi \cdot 10^{-7}$ H/m = $4\pi \cdot 10^{-7}$ N/A².

15.3.3 Gaussian system

There are several CGS-based systems; here I describe the most common one, often called “gaussian units.”¹⁰ When someone says “I use gaussian units,” they are generally saying three distinct things. Let’s take them one at a time:

Base units

The gaussian system uses cm, g and s. The base unit of charge is called **statcoul**; it is defined by requiring that ϵ_0 (not e) have an exact numerical part:

$$\epsilon_0 = \frac{1}{4\pi} \frac{\text{statcoul}^2 \cdot \text{s}^2}{\text{g} \cdot \text{cm}^3}. \quad (15.4)$$

In this system, it’s μ_0 and e that have approximate, measured values. We determine μ_0 by using Equation 15.3:

$$\mu_0 = \frac{4\pi}{c^2} \frac{\text{g} \cdot \text{cm}^3}{\text{statcoul}^2 \cdot \text{s}^2}.$$

Combining Equations 15.2, 15.3, and 15.4 yields

$$1 \text{ statcoul} \approx (0.1 \text{ m/s})c^{-1} \text{ coul} \approx \frac{1}{3 \cdot 10^9} \text{ coul}. \quad (15.5)$$

We also measure ρ_q in statcoul/cm³ and \vec{j} in statcoul/(cm² · s), etc.

Another useful unit conversion involves electrostatic potential. The SI unit is volt = J/coul. The corresponding gaussian unit is **statvolt** = erg/statcoul.

Your Turn 15A

Find the relation between these.

¹⁰Maxwell and F. Jenkin had more to do with developing this system than Gauss (Maxwell & Jenkin, 1865).

Modified magnetic field

“Gaussian” authors also redefine the magnetic induction, introducing a physically different quantity that I’ll call

$$\check{B} \equiv c\vec{B}.$$

Confusingly, they call this quantity “the magnetic induction” and use the symbol \vec{B} for it! We won’t do that.

We can use \check{B} in any system of units; we’ll occasionally find it convenient to express magnetic fields in this way. For example, \check{B} has the same units as \vec{E} , so we can compare them directly.¹¹ “Gaussian” authors similarly define modified magnetic flux by $\check{\Phi}_B = c\Phi_B$ and magnetic moment by $\check{D}_M = \vec{D}_M/c$.

Maxwell’s equations then say:

$$\vec{\nabla} \cdot \vec{E} = \rho_q/\epsilon_0 \quad (15.6)$$

$$\nabla \cdot \check{B} = 0 \quad (15.7)$$

$$\nabla \times \vec{E} + \frac{1}{c} \frac{\partial \check{B}}{\partial t} = 0 \quad (15.8)$$

$$\nabla \times \check{B} - \frac{1}{c} \frac{\partial \vec{E}}{\partial t} = \frac{1}{\epsilon_0} \vec{j} \quad (15.9)$$

and the Lorentz force law says

$$\frac{\partial}{\partial t} \vec{p} = q \left(\vec{E} + \frac{\vec{v}}{c} \times \check{B} \right). \quad (15.10)$$

The equations are still valid in any system of units; in gaussian base units we have the numerical values $\epsilon_0 = \frac{1}{4\pi} \frac{\text{statcoul}^2 \cdot \text{s}^2}{\text{g} \cdot \text{cm}^3}$ and $c \approx 3 \cdot 10^{10} \text{ cm s}^{-1}$.

The electric and modified magnetic fields have the same dimensions, but it’s traditional to call the unit of \check{B} the **gauss**, and that of \vec{E} the **statvolt/cm**. In fact, these (and the **oersted**) are all the same as $\text{g cm}/(\text{s}^2 \text{ statcoul})$.

Your Turn 15B

Show by using Equation 15.5 that their SI equivalents are

$$1 \text{ gauss} \approx c \cdot 10^{-4} \text{ T}$$

$$1 \text{ statvolt/cm} \approx 3 \cdot 10^4 \text{ volt/m.}$$

That is, a field $\vec{B} = 1 \text{ T}$ corresponds to $\check{B} = 10^4 \text{ gauss}$.

Elimination of charge units

We could stop there. But “gaussian” authors take one more step, similar to the one outlined in Section 15.2. So far we have stayed with what I called approach **A**, but now we switch to:

¹¹A spatial region A , of uniform electric field $\vec{E}_{(A)}$, and no magnetic induction, will have the same energy density as a region B of uniform magnetic induction $\check{B}_{(B)}$, and no electric field, if $\|\vec{E}_{(A)}\| = \|\check{B}_{(B)}\|$.

B2: Elimination of charge units: For each physical quantity X , “gaussian” authors define \bar{X} to be X divided by as many powers of $(\text{statcoul})(\text{s})(\text{g cm}^3)^{-1/2}$ as are needed to eliminate the \mathbb{Q} dimensions.¹² Why this crazy choice? With this choice, $\bar{\epsilon}_0$ becomes *dimensionless*, with exact numerical value equal to $1/(4\pi)$.

Thus all barred quantities have dimensions that are powers of \mathbb{L} , \mathbb{M} , and \mathbb{T} : We have eliminated charge units. The vacuum Maxwell equations now take the elegant form

$$\begin{aligned}\vec{\nabla} \cdot \bar{\vec{E}} &= 4\pi\rho_q \\ \vec{\nabla} \cdot \bar{\vec{B}} &= 0 \\ \vec{\nabla} \times \bar{\vec{E}} + \frac{1}{c} \frac{\partial \bar{\vec{B}}}{\partial t} &= 0 && \text{ (“gaussian” units)} \\ \vec{\nabla} \times \bar{\vec{B}} - \frac{1}{c} \frac{\partial \bar{\vec{E}}}{\partial t} &= \frac{4\pi}{c} \bar{\vec{j}} \\ \frac{d\bar{\vec{p}}}{dt} &= \bar{q} \left(\bar{\vec{E}} + \frac{\vec{v}}{c} \times \bar{\vec{B}} \right).\end{aligned}$$

In this system, the constant $\bar{\epsilon}_0 = 1/4\pi$ not only has a nice numerical part; it has also been purged of all units. Then we get Coulomb’s Law in the ultra-simple form $\bar{\psi}(\vec{r}) = \bar{q}/\|\vec{r}\|$, etc. *The price we pay* is that the above equations are valid only in the gaussian system (unlike Equations 0.1–0.5, which are valid in any units).

“Gaussian” authors confuse us by omitting all the bars and checks! That explains a lot of bizarre-sounding assertions like “ $1\text{ F} = 9 \cdot 10^{11}\text{ cm}$,” which one sometimes hears. More precisely, this statement says that “a capacitance of $C = 1\text{ F}$ corresponds to the reduced quantity $\bar{C} = 9 \cdot 10^{11}\text{ cm}$.”

Your Turn 15C

Check the last claim. Then find analogous statements for an inductance of 1 H and a resistance of $1\ \Omega$.

15.3.4 One final confusion

“Gaussian units” eliminate the dimension \mathbb{Q} , but they still retain the familiar \mathbb{L} , \mathbb{T} , and \mathbb{M} .

Incredibly, it is commonplace for authors not to state any specific units. Instead they often just write something like “esu” for everything, which roughly means “whichever of those units is appropriate for this quantity in the system I’m using.” You’re supposed to supply the appropriate unit using context. It works if you never make any errors, and you are communicating with someone who uses the same unit system as you do, but I don’t recommend it!

¹²This step does not change the numerical part of X if we’ve expressed it in the base units cm , g , s , and statcoul . That’s because this factor’s numerical part equals one in that case.

15.4 REMARKS

I hope that the difference between units is starting to seem like, say, the difference between French and Spanish. You need to talk like the natives, wherever you're going, but they're always the same Maxwell equations in any language.

The “gaussian” unit system eliminates one of the two constants of Nature in Maxwell's equations: Instead of ϵ_0 and μ_0 , all we now have is c . Some people find this beautiful. I think that making fewer errors in my work is beautiful, so I don't advocate eliminating units.

Some say that gaussian units make the duality of the electric and magnetic field clearer. It's true, but in a trivial way. We will get the same benefit just by expressing Maxwell's equations in terms of \check{B} instead of \vec{B} (Equations 15.6–15.9), regardless of whether we measure \check{B} in gauss or in $\text{T} \cdot c$. Ultimately we'll construct a single, unified “Faraday tensor” out of the components of \vec{E} and \check{B} .

Finally, don't try looking on Amazon for a “statvoltmeter” or an “statammeter.”¹³ Using SI units in our math keeps us connected to the real world of experiments, where people use volts and amperes.

FURTHER READING

Redefinition of SI units:

https://en.wikipedia.org/wiki/2019_redefinition_of_SI_base_units (accessed 11/2018).

¹³Especially don't ask for an “abammeter.”

T₂ Track 2

15.2' Elimination of more units

In gravitational physics, many authors take elimination of units one step further, agreeing also to measure mass in units of $(1\text{ m})c^2/G_N$ (not kg). Here G_N is Newton's constant. The barred quantities are obtained by dividing physical quantities by enough powers of c and G_N to eliminate *both* \mathbb{M} and \mathbb{T} , leaving only \mathbb{L} . In this scheme, $\bar{G}_N = 1$ and $\bar{c} = 1$, so both can be dropped from formulas.

In high-energy physics, many authors choose instead to eliminate both \mathbb{L} and \mathbb{T} , leaving only \mathbb{M} , which they typically measure in GeV/c^2 . They set up barred quantities by demanding that $\bar{c} = 1$ and $\bar{\hbar} = 1$, leading to confusion when they try to talk to gravitational physicists.

PROBLEMS

15.1 Dimensional shortcut

Background: A magnet is dropped through a nonmagnetic, but conducting, tube. Friction with the tube's walls is negligible. But instead of increasing without bound, the magnet's velocity saturates at a surprisingly small value v_* . At this terminal velocity, the release of gravitational potential energy does not go into increasing the magnet's kinetic energy; instead, it all goes into ohmic heating of the tube, via induced "eddy" currents.

We could try to set up Maxwell's equations, but it would be a big pain. Instead, obtain an *estimate* for the terminal velocity as follows. Before you begin, note that:

- (i) The effect depends on the strength of the magnet, i.e. on its dipole moment $\vec{\mathcal{D}}_M$. Actually the dipole moment always enters into formulas multiplied by μ_0 , so let $X = \mu_0|\vec{\mathcal{D}}_M|$. In the limit $X \rightarrow 0$, there's no effect and (in vacuum) the falling magnet's velocity increases without limit, i.e. $v_* \rightarrow \infty$.
- (ii) The effect depends on the electrical conductivity κ of the material constituting tube. In the limit $\kappa \rightarrow 0$, there's no effect and again $v_* \rightarrow \infty$.
- (iii) The terminal velocity depends on the weight F of the magnet (a force). We expect that pulling harder on the magnet will let it achieve larger terminal velocity, by analogy to the case of pulling on an object immersed in a viscous fluid, i.e. v_* is an increasing function of F .
- (iv) The effect depends on the size scale of the apparatus, e.g. on the diameter L of the tube.¹⁴
- (v) I looked up some typical values: Button magnets like the ones used in demos have magnetic moment $\approx 0.3 \text{ A m}^2$ and mass $\approx 7 \text{ g}$. The conductivity of aluminum is $\kappa \approx 5 \cdot 10^7 \Omega^{-1} \text{ m}^{-1}$. A typical demo apparatus has diameter $L \approx 1 \text{ cm}$.

Now take these steps:

- a. Find a combination of the relevant constants X , κ , F , and L that has the dimensions of a velocity.
- b. Confirm that the formula you found in (a) has the expected behaviors listed in (ii–iii) above.
- c. Evaluate the formula for v_* numerically with the values given in (v) above.

15.2 Units: conductivity

- a. Infer the units of conductivity κ from the formula $\vec{j} = \kappa \vec{E}$. Infer the units of resistance from the formula $\Delta\psi = IR$.
- b. Use dimensional analysis to guess the relation between κ and R for a long wire of length L and cross section A .
- c. Substitute SI base units into the dimensions of R to find the definition of the SI unit of resistance (the ohm) in terms of base units.

15.3 Units: Polarizability

¹⁴You can neglect possible dependences on other dimensions, e.g. on the thickness of the tube's wall. (That is, pretend it's infinitely thick, a long straight hole bored into a big solid block of metal.)

- a. Show that the units N/coul and volt/m are the same. Then use that insight to give two interpretations of what we mean by the strength of the electric field.
- b. Explain the apparently paradoxical utterances of gaussian people, e.g. when they say say: “Electric polarizability is the ratio of the electric dipole moment of a molecule to the applied electric field. Its units are cm^3 .”

15.4 *Unit fun*

Explain the paradoxical-sounding utterances of gaussian people, when they say:

- a. “ $1 \Omega = ? \text{ s/cm}$.”
- b. “ $1 \text{ H} = ? \text{ s}^2/\text{cm}$.”

Also fill in the missing numbers (i.e., derive them).

15.5 *Units: Capacitance*

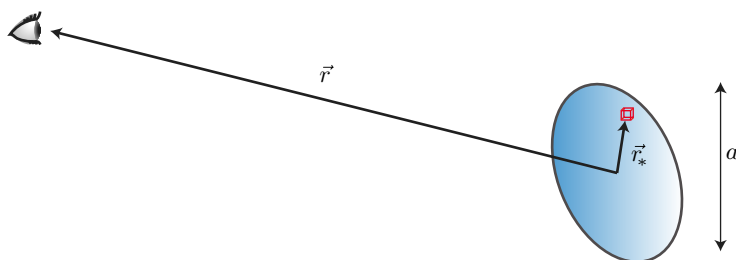
Explain the paradoxical-sounding utterances of gaussian people, who say:

- a. “ $1 \text{ farad} = 9 \cdot 10^{11} \text{ cm}$.”
- b. “Electric polarizability is the ratio of the electric dipole moment of a molecule to the applied electric field. Its units are cm^3 .”

CHAPTER 16

Magnetostatic Multipole Expansion

Analogously to electrostatics, we consider a stationary, localized distribution of electric current. Thus $\vec{j} = 0$ outside a region of size a , and we wish to know the fields far away, as an expansion in powers of a/r . Again place the origin of coordinates at some fixed point inside the source. We'll again exploit Taylor's theorem for $\|\vec{r} - \vec{r}_*\|^{-1}$, but there are some tricky tensor things to get right.



16.1 TENSOR PRELIMINARIES

First, recall that a stationary source must have $\vec{\nabla} \cdot \vec{j} = 0$.¹ So

$$0 = \int d^3r (\vec{r}_i)(\vec{\nabla} \cdot \vec{j}) = - \int d^3r \vec{j}_k (\vec{\nabla}_k \vec{r}_i) = - \int d^3r \vec{j}_i. \quad (16.1)$$

(The boundary term is zero because we assumed a localized source.) We conclude that each cartesian component of \vec{j} , when integrated over the source, yields zero.

Similarly,

$$\begin{aligned} 0 &= \int d^3r (\vec{r}_k \vec{r}_i) \vec{\nabla} \cdot \vec{j} = - \int d^3r \vec{j}_m \vec{\nabla}_m (\vec{r}_k \vec{r}_i) \\ &= \int d^3r (\delta_{mk} \vec{r}_i \vec{j}_m + \delta_{mi} \vec{r}_k \vec{j}_m) = \int d^3r (\vec{r}_i \vec{j}_k + \vec{r}_k \vec{j}_i). \end{aligned} \quad (16.2)$$

Define the **magnetic dipole moment tensor**

$$\vec{\vec{\Gamma}} = \int d^3r \vec{r} \vec{j}. \quad (16.3)$$

Equation 16.2 says that $\vec{\vec{\Gamma}}$ is a rank-2, antisymmetric 3-tensor.

From now on, we will change notation from \vec{r} to \vec{r}_* to refer to the location of a point inside the source. The notation \vec{r} will now refer to the position of an observer, as in the cartoon above.

¹Section 7.3 (page 83).

16.2 FAR FIELDS OF A STEADY, LOCALIZED CURRENT DISTRIBUTION

16.2.1 Magnetic dipole potential

Suppose that it makes sense to talk about a continuously distributed current source, maybe some interstellar plasma or the flow of ions outside a nerve cell. Section 14.4.2 (page 166) showed that each component of the vector potential obeys the Poisson equation. Applying a Taylor expansion as we did in electrostatics thus gives

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi r} \int d^3r_* \vec{j}(\vec{r}_*) \left(1 + \frac{\vec{r} \cdot \vec{r}_*}{r^2} + \dots\right). \quad (16.4)$$

In principle we're done! But some further simplification is useful.

Equation 16.1 says that the first term of Equation 16.4 is zero: There is no contribution at order r^{-1} , i.e. no “magnetic monopole” field.²

The definition Equation 16.3 lets us rephrase the second term:

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi r^3} \vec{r} \cdot \vec{\Gamma} + \dots$$

We have accomplished our usual goal of expanding the potential in a systematic power series and, at each order, separating a potential into a product of universal, standard functions of \vec{r} (here the three functions $\mu_0 \vec{r}/(4\pi r^3)$) multiplied by some constants characterizing the source (here $\vec{\Gamma}$). Next, we manipulate a bit to cast our result into a traditional form.

Recall that any antisymmetric 3-tensor can be rewritten in terms of a vector (as we did with \vec{B}). Thus we get relations analogous to Equations 14.3 and 14.2 (page 161):

$$\vec{\Gamma}_{in} = \varepsilon_{ink} \vec{D}_{M,k} \quad \text{where} \quad \vec{D}_M = \frac{1}{2} \int d^3r_* (\vec{r}_* \times \vec{j}(\vec{r}_*)). \quad (16.5)$$

The three constants \vec{D}_M are called the **magnetic dipole moment vector**. Then

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \frac{\vec{D}_M \times \hat{r}}{r^2} + \mathcal{O}(r^{-3}). \quad (16.6)$$

Thus the leading nonzero term of the vector potential outside a general local current distribution falls like r^{-2} , similarly to the electrostatic dipole potential in electrostatics.

16.2.2 A Familiar Example

Your Turn 16A

- To make sure you understand how it all works, consider a thin, circular loop of wire of radius a in the xy plane, centered on the origin of coordinates and carrying current I . Work out \vec{D}_M for this current distribution. *Hint:* Use Equation 14.19 to find the current and substitute into Equation 16.5.
- Also, compute the curl of Equation 16.6 to find the magnetic field far away from a current source to leading nontrivial order in a/r .

²See, however, Problem 16.2.

16.3 REMARKS

16.3.1 Higher moments

Naturally, there are higher magnetic multipole fields controlled by higher magnetic multipole moments. For example, consider a *pair* of circular wire loops, lying in parallel planes but shifted perpendicular to those planes and carrying opposite currents. The total magnetic dipole moment is zero, but there will nevertheless be magnetic fields outside this source. Those **magnetic quadrupole** fields fall off with distance faster than those of a magnetic dipole.

Your Turn 16B

Work out the far field by retaining the next-order term in Equation 16.4.

16.3.2 No base point ambiguity

Returning to Equation 16.5,

Your Turn 16C

Show that, had we chosen a different origin of coordinates shifted by some constant vector \vec{h} , we would have ended with the same values for \vec{D}_M .

(Similarly to the electrostatic case, higher moments may depend on the choice of basepoint; more precisely, only the first nonzero moment is unambiguously defined.)

16.4 FORCE AND TORQUE ON A MAGNETIC DIPOLE

16.4.1 Fixed dipole strength

Consider a current distribution that can translate or rotate in space but is otherwise rigid: All current elements are steady in time and maintain fixed spatial relations with each other. A macroscopic example could be to imagine a stiff loop of wire with a constant-current source. Also, some individual molecules can create a magnetic field because of persistent currents in their electron state.

This current distribution is immersed in an external static magnetic field \vec{B}^{ext} , which varies with a characteristic length scale much bigger than the size of the distribution itself. We have $\vec{\nabla} \times \vec{B}^{\text{ext}} = 0$ because whatever the source of the external field, it doesn't overlap the current distribution.

Choose an origin of coordinates somewhere inside the current distribution. Any internal forces must add up to zero. The Lorentz force law applied to each current element gives³

$$\vec{F} = \int d^3r_* \vec{j}(\vec{r}_*) \times \vec{B}^{\text{ext}}(\vec{r}_*).$$

³We'll revisit this argument in more detail later (Equation 34.5, page 385).

Similarly to Section 3.6, we now make a Taylor expansion of the external field near the reference point:

$$\vec{F}_i = \varepsilon_{ink} \left[\vec{B}_k^{(0)} \int d^3r_* \vec{j}_n(\vec{r}_*) + \left. \frac{\partial \vec{B}_k^{\text{ext}}}{\partial \vec{r}_m} \right|_0 \int d^3r_* \vec{r}_{*m} \vec{j}_n(\vec{r}_*) \right] + \dots$$

The first term on the right equals zero by Equation 16.1. The second involves the magnetic dipole moment, which we again express as in Equation 16.5:

$$\begin{aligned} \vec{F}_i &= \varepsilon_{ink} \left. \frac{\partial \vec{B}_k^{\text{ext}}}{\partial \vec{r}_m} \right|_0 \varepsilon_{mnl} \vec{\mathcal{D}}_{M,\ell} \\ &= \left. \frac{\partial \vec{B}_k^{\text{ext}}}{\partial \vec{r}_m} \right|_0 (\delta_{im} \delta_{kl} - \delta_{il} \delta_{mk}) \vec{\mathcal{D}}_{M,\ell} \\ &= \vec{\mathcal{D}}_{M,\ell} \vec{\nabla}_i \vec{B}_{\text{ext},\ell} - \vec{\mathcal{D}}_{M,i} \vec{\nabla} \cdot \vec{B}, \text{ or} \\ \vec{F} &= \vec{\nabla} (\vec{B}_{\text{ext}} \cdot \vec{\mathcal{D}}_M). \end{aligned} \quad (16.7)$$

As with an electric dipole in an electric field, so too a magnetic dipole feels no net force in a uniform magnetic field.

We can also work out the torque on this rigid current distribution:

$$\vec{\tau} = \int d^3r_* \vec{r}_* \times (\vec{j}(\vec{r}_*) \times \vec{B}^{\text{ext}}).$$

For example,

$$\vec{\tau}_3 = \int d^3r_* [\vec{j}_3(\vec{r}_*) (\vec{r}_* \cdot \vec{B}^{\text{ext}}) - \vec{B}^{\text{ext}} (\vec{r}_* \cdot \vec{j}(\vec{r}_*))].$$

First consider the terms without derivatives of the external field:

$$\vec{B}_i^{(0)} \int d^3r_* \vec{j}_3(\vec{r}_*) \vec{r}_{*i} - \vec{B}_3^{(0)} \int d^3r_* \vec{j}_i(\vec{r}_*) \vec{r}_{*i}.$$

The second term is zero because the magnetic dipole moment tensor is antisymmetric. The first term can be written in terms of the moment as

$$\begin{aligned} &= \vec{B}_i^{(0)} \frac{1}{2} \varepsilon_{3ik} \int d^3r_* (\vec{j} \times \vec{r}_*)_k \\ \vec{\tau} &= -(\vec{B}^{(0)} \times \vec{\mathcal{D}}_M). \end{aligned}$$

We conclude that a free magnetic dipole of fixed strength in an external field experiences a torque tending to align its moment with the \vec{B} field. Once it is aligned, Equation 16.7 shows that it also feels a force driving it to a region of higher $\|\vec{B}\|$. You'll explore a practical application of these results to manipulation of micrometer objects in Problem 16.3.

Note that a quantum-mechanical spin cannot freely “reorient,” due to spatial quantization. Thus a single neutron, which has a permanent magnetic dipole moment, (or a neutral atom such as silver) will migrate along *or against* the gradient of $\|\vec{B}\|$ depending on its spin state: The Stern–Gerlach effect (1922).⁴ Even particles currently thought to be fundamental, like the electron and muon, have permanent intrinsic magnetic dipole moments.

⁴Stern and Gerlach were astonished to find an even number of discrete spin states, not the odd number predicted from the theory of orbital angular momentum.

16.4.2 Diamagnetism, paramagnetism, ferromagnetism

Just as some molecules can “polarize” (develop an electric dipole moment) under the influence of an external electric field, so others are *magnetically* polarizable: They develop persistent internal currents under the influence of an external magnetic field, giving rise to a magnetic dipole moment. Bulk materials containing such molecules can develop a density of magnetic dipole moment throughout their volume.⁵ Also analogously to the electric case, a material can polarize simply by the alignment of preexisting, but initially disordered, intrinsic dipole moments.

The induced moment can be parallel to the applied field (**paramagnetism**), or antiparallel to it (**diamagnetism**). If there is a nonzero net magnetic dipole moment density even at zero applied field, we call the material **ferromagnetic**.

16.4.3 Magnetic levitation of objects at room temperature

See <https://www.youtube.com/watch?v=a8sCtLY-vZY>.

FURTHER READING

General: Zangwill, 2013, ch. 11.

Diamagnetic levitation: Berry & Geim, 1997.

PROBLEMS

16.1 *Cell sorting*

Magnetic cell sorting is a way to isolate cells of one particular type. Small particles (about 50 nm diameter spheres) are bound to an antibody that attaches specifically to the cell type of interest (e.g. a cancer cell). Cells are then mechanically separated by the difference in force applied to the target cells versus normal cells.

The magnetic particles are “superparamagnetic”; you may assume that this means that they respond to an external magnetic field \vec{B} by developing their own magnetic dipole moment $\vec{\mathcal{D}}_M = v\vec{B}/\mu_0$, where v is the volume of the particle.⁶

The cells are then placed in a magnetic field gradient, and the resulting force is used to manipulate the cell. What is the force if 100 of these particles are attached to a cell that is in a magnetic field of 1 T, with gradient 10 T/m?

16.2 *Magnetic monopole potential*

We found that a localized current distribution will not create any magnetic monopole field. Nevertheless, we can imagine a stationary magnetic field configuration for which \vec{B} points everywhere radially outward from some point in space, much like the electric field from a point charge. We hit an interesting problem when we seek a vector potential for this field.

⁵See Problem 16.1 and Problem 16.3.

⁶[\[T2\]](#)In more detail, generally $\vec{B} = \mu_0(\vec{H} + \vec{M})$ where $\vec{H} = \vec{M}/\chi_m$ and $\vec{M} = \vec{\mathcal{D}}_M/v$ (see Chapter 49). Superparamagnetic means the susceptibility $\chi_m \gg 1$, so $\vec{\mathcal{D}}_M = v\vec{B}/\mu_0$.

Let r , θ , φ be spherical polar coordinates.

- Find an expression for the gradient $\vec{\nabla}\varphi$. Find an expression for $\vec{\nabla}\theta$. Find an expression for the cross product $\vec{\nabla}\varphi \times \vec{\nabla}\theta$.
- Consider the time-independent magnetic vector potential given by

$$\vec{A} = g\hat{\varphi} \frac{\cos\theta}{r \sin\theta}. \quad (16.8)$$

Here g is an overall constant and $\hat{\varphi}$ is the unit vector in the azimuthal direction. Compute the magnetic field corresponding to this vector potential as follows. First reexpress \vec{A} as a scalar function times $\vec{\nabla}\varphi$ using your result in (a).

- Prove the identity $\vec{\nabla} \times (f \cdot \vec{V}) = (\vec{\nabla}f) \times \vec{V} + f \cdot \vec{\nabla} \times \vec{V}$ for any scalar function f and vector field \vec{V} .
- Use (a–c) to compute the curl of \vec{A} and interpret the result.
- Not surprisingly, the expression Equation 16.8 is singular at $r = 0$. But it's *also* bad all along the polar axis! Show that the two modified expressions

$$\vec{A}^{(\pm)} = g\hat{\varphi} \frac{\pm 1 + \cos\theta}{r \sin\theta}$$

differ from Equation 16.8 only by gauge transformations, and hence describe the same magnetic field.

- Show that one of the new vector potentials is nonsingular all along the axis $\theta = 0$, whereas the other one is nonsingular all along the axis $\theta = \pi$. Thus, we have a good vector potential representing a magnetic monopole everywhere except right at the origin, where there is a real singularity.

16.3 Magnetic tweezers

The following page shows some unpublished information I got from my friends in Paris about their magnetic tweezer setup. The first graph gives the magnetic moment per gram of their bead, as a function of applied magnetic field. The second graph shows the measured magnetic field as a function of the vertical distance z from the magnet pole.

- Apparently “emu” is some CGS unit for magnetic dipole moment. Figure out the appropriate unit and explain the cryptic notation “1 emu = 10^{-3} SI.”
- Look at the part of the first graph where it's approximately linear. Lay a ruler on the graph and find the slope. (Or if you prefer you may find it using the information given in the graph's inset.) Use this linear approximation from now on.
- Look at the part of the second graph for z between 2 and 4 mm. Lay a ruler along this semilog graph and approximate it by a straight line. That is, find the constants B_{\max} and z_0 mentioned at the top.
- Now get an approximate formula for the force on the bead as a function of z , using your results from (a–b). Write a suitable log scale for force along the left-hand side of the second graph, and draw the expected force-versus- z curve.

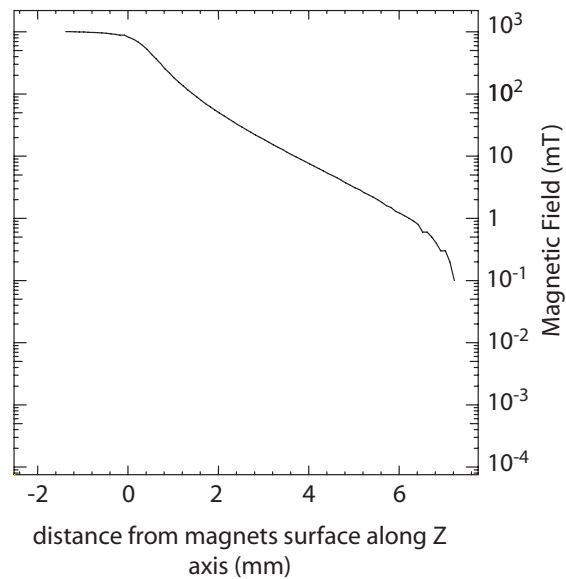
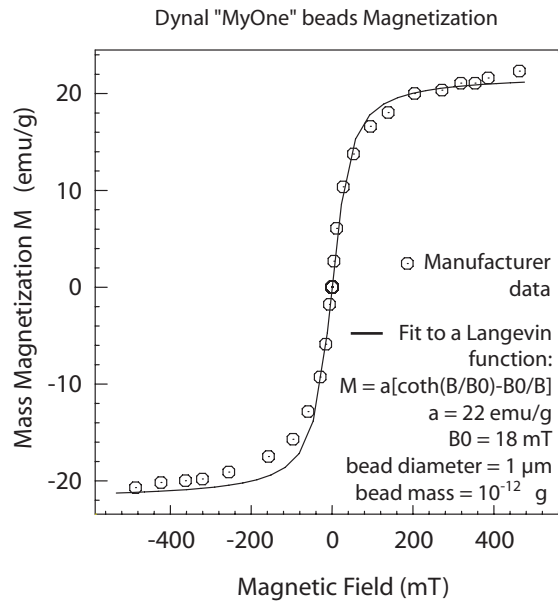
Force estimation and measurement in magnetic Tweezers:

Timothée Lionnet and Vincent Croquette, Ecole Normale Supérieure

The force can be estimated by approximating the magnetic field as an exponential:

$$B \sim B_{\max} \exp(-z/z_0)$$

$$1 \text{ emu} = 10^{-3} \text{ S.I.}$$



16.4 Levitation of single cells

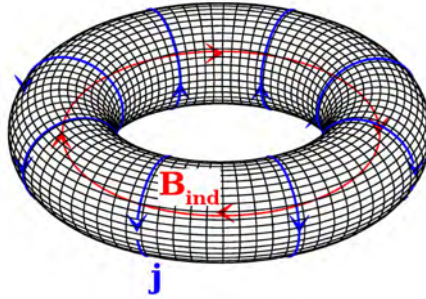


Figure 16.1: [From https://commons.wikimedia.org/wiki/File:Solenoid_currents_inducing_a_toroidal_magnetic_moment.tif]

[Not ready yet.]

16.5 *Ambidextrous 2*

Rederive the results of Section 16.4.1 without making use of the Levi-Civita tensor; that is, formulate them in terms of the magnetic field tensor $\vec{\omega}$ (Equation 14.3) and the magnetic dipole moment tensor $\vec{\Gamma}$ (Equation 16.2).

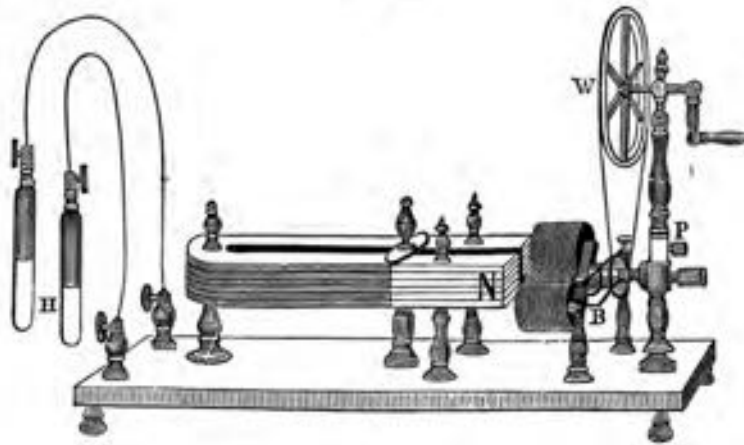
16.6 *Static toroidal moment*

Work out the far field from the steady current distribution shown in Figure 16.1. Comment in the light of your result to Your Turn 16B (page 189). Evaluate the relevant magnetic quadrupole tensor and show that it's not zero.

PART III

Dynamic

FIG. 228.



CHAPTER 17

Beyond Statics

17.1 STATICS REVIEW

17.1.1 Field equations

We have explored some equations whose solutions seem to describe the electric and magnetic fields set up by stationary charges and currents:

$$\vec{\nabla} \cdot \vec{E} = \rho_q / \epsilon_0 \quad \text{Gauss} \quad (17.1)$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{Gauss} \quad (17.2)$$

$$\vec{\nabla} \times \vec{E} = 0 \quad (\text{stationary case}) \quad (17.3)$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{j} \quad \text{Ampère (stationary case)}. \quad (17.4)$$

I belabored these equations because

- They have a lot of practical implications. You can understand nerve impulses, photocopiers, lightning rods, molecular recognition, and much more with these equations.
- We will need to be sure we are standing on firm ground when later we arrive at some conclusions that physicists found unpalatable, even revolutionary.

17.1.2 A worked example

To anchor all the abstractions that are to come, here is an old result from magneto-statics that you probably recall from first-year physics. Figure 17.1 represents a coil of wire wound in a helix of radius a around a long cylinder of length h . Such a coil is often called a **solenoid**. It consists of N loops. Steady current I is sent through the wire. Work through the next paragraphs to exercise those Stokes-theorem muscles.

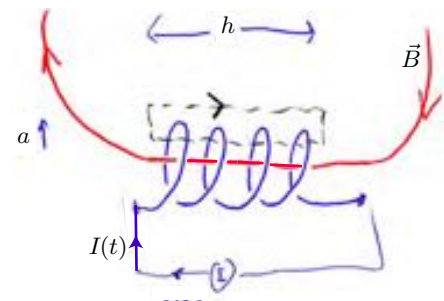


Figure 17.1:

Each loop makes a dipole field. The helicity of the coil is such that if $I > 0$, then in front of the page, current is moving upward; behind the page, current is moving downward. Deep inside the cylinder (far from the ends), symmetry suggests that \vec{B} will point axially as shown, though we still need to confirm the direction. To find its magnitude, consider the path shown as a dashed line. We can traverse that path in either direction;¹ a specific choice is shown. That choice determines a vector perpendicular to the rectangular surface bounded by that path via the right-hand rule: $\vec{\Sigma}$ points into the page.

Integrating Ampère’s law over the surface bounded by the path gives

$$\int d^2\vec{\Sigma} \cdot (\vec{\nabla} \times \vec{B}) = \int d^2\vec{\Sigma} \cdot (\mu_0\vec{j}). \quad (17.5)$$

Stokes’s theorem gives the left side as $\oint d\vec{\ell} \cdot \vec{B}$. The part of the path lying inside the cylinder contributes Bh , because \vec{B} is uniform along the coil² and points axially (B is its component in the leftward direction). The sides of the rectangular path are perpendicular to \vec{B} , so here $d\vec{\ell} \cdot \vec{B} = 0$. And $\vec{B} \approx 0$ outside the cylinder, because the field lines fan out once they exit the ends.

On the right side of Equation 17.5, each time the wire pierces the surface element we get a contribution I to the integral. (That’s because $\vec{\Sigma}$ and \vec{j} both point into the page at each such point.) Thus Equation 17.5 becomes

$$Bh = \mu_0NI, \quad \text{or} \quad B = \mu_0NI/h, \quad (17.6)$$

a familiar result. Inside the solenoid, \vec{B} is uniform; it does not depend on how far we are from the coil’s axis.

As an aside, our result can be reexpressed in terms of the total **magnetic flux**³ $\Phi_B = N\pi a^2B$ as

$$\Phi_B = LI, \quad (17.7)$$

where the **self-inductance** L in this case equals $\pi a^2\mu_0N^2/h$.

17.2 FARADAY LAW

In electrostatics, Equation 17.3 says that the electric field gives rise to a *conservative* force on charges, similarly to the newtonian gravitational force. We might guess that it would continue to hold in non-static situations, and indeed everyone did believe that before Michael Faraday. After all, newtonian gravity had that same form, even with all those planets whizzing around.⁴ Perhaps that prejudice was what prevented the Continental scientists from seeing what Faraday saw.⁵

¹Your Turn 0B (page 8).

²We are neglecting end effects.

³This traditional term is an exception to our general convention that a “flux” is the rate of transport of some conserved quantity (such as charge) per unit transverse area.

⁴Today we know that even newtonian gravity required modification; those moving planets actually generate tiny “gravitomagnetic” effects.

⁵See Chapter 35.

In the gravitational case, a roller-coaster that traverses a closed loop returns to its starting point with the same kinetic energy as it began (minus frictional losses), because the gravitational force on it is the gradient of a potential energy function. Michael Faraday observed, however, that plunging a magnet into a loop of wire generates an **electromotive force (EMF)**, which pushes on the electrons all around the loop. Rather than suppose that this EMF is something entirely new, we will regard it as simply a contribution to the electric field that is *not* conservative. Faraday found that its effects were proportional to the time rate of change of the magnetic field, which suggests the following modification to Equation 17.3:

$$\vec{\nabla}_i \vec{E}_j - \vec{\nabla}_j \vec{E}_i = -2 \frac{\partial}{\partial t} \vec{\omega}_{ij} \quad \text{Faraday.}$$

Here $\vec{\omega}$ is the antisymmetric tensor field representing magnetism.⁶ This formulation of Faraday’s proposal makes it clear that it’s rotationally invariant (it equates tensors of the same type) and also invariant under spatial inversion (it contains no Levi-Civita tensor). Also the units match on each side.⁷

It’s more conventional to contract both sides of the preceding formula with a Levi-Civita tensor, which yields⁸

$$\vec{\nabla} \times \vec{E} = -\frac{\partial}{\partial t} \vec{B}. \quad \text{Faraday} \quad (17.8)$$

Your Turn 17A

Suppose that we have a circular loop of wire. Integrate both sides of Equation 17.8 over a surface bounded by the loop, and show that the current induced by a time-dependent applied \vec{B} field flows in the direction that generates an opposing \vec{B} (“Lenz’s law”).

17.3

17.3.1 Back-EMF

We now return to the concrete situation considered Section 17.1.2, but this time suppose that we force a current $I(t)$ through the coil that varies slowly in time. Here “slowly” means too slowly for us to need to account for the time-derivative term in Ampère’s law (which is multiplied by a very small constant). Faraday’s law says that an electric field will result. To find it, we integrate both sides of Faraday’s law over a surface, but not the same surface as in Figure 17.1. Instead, our surface will be a disk transverse to the axis, bounded by the cylinder on which we wrapped the wire (Figure 17.2). Again we can choose any direction we like for the rim of that disk; to keep things simple, choose the same direction as that of current flow. So

$$\int d^2\vec{\Sigma} \cdot (\vec{\nabla} \times \vec{E}) = -\frac{d}{dt} \int d^2\vec{\Sigma} \cdot \vec{B}. \quad (17.9)$$

⁶Equation 14.3 (page 161).

⁷Why is the factor of -2 needed? Chapters 31–33 will argue that the form of this equation, including this factor, is ultimately dictated by Lorentz invariance.

⁸Recall Equation 14.2.

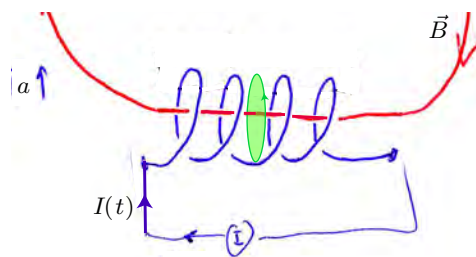


Figure 17.2: The *green object* is a disk-shaped surface of radius a , viewed edge-on, that is integrated over in Equation 17.9.

This time, Stokes's theorem gives the left side as $\oint d\vec{\ell} \cdot \vec{E}$. By axial symmetry, the integrand is constant, so we get $2\pi a \vec{E}_\varphi$, where \vec{E}_φ is the component in the direction of current flow.

The right-hand rule selects the perpendicular vector to our surface that points leftward. Equation 17.6 then lets us write the right side of Equation 17.9 as

$$-\frac{\pi a^2}{h} \mu_0 N \frac{dI}{dt}.$$

Setting this expression equal to the left side of Equation 17.9 gives

$$\vec{E}_\varphi = -\frac{\mu_0 N a}{2h} \frac{dI}{dt}.$$

The minus sign says that the induced electric field *opposes changes* in current.⁹ Thus, to increase I we must *do work* against an opposing electric field. Let's see how much work is needed.

The electric potential drop across the solenoid is the work per unit charge needed to push charge through it, that is, the force per charge times the distance:

$$\Delta\psi = \vec{E}_\varphi (2\pi a N) = -\frac{\mu_0 N^2 a^2 \pi}{h} \frac{dI}{dt} = -L \frac{dI}{dt}. \quad (17.10)$$

That is, the potential drop is linear in the current, like a resistor, but with a time derivative. Engineers sometimes call this contribution to the potential drop a **back-EMF**.

17.3.2 Cables, again

Chapter 10 introduced a model for the propagation of an electrical disturbance along a cable. As cables grew to transatlantic length, however, inadequacies of Kelvin's mathematical model became obvious. The problem was not just resistive loss. A sharp step function introduced at one end emerged blurred at the other end, limiting the speed of transmission. Eventually Heaviside and others realized that the problem was the neglect of self-induction in Kelvin's model. Incorporating self-induction raised the possibility of wavelike solutions, but those solutions were found to suffer from dispersion (Problem 17.3). The problem became more severe with the advent of the telephone, because audio signals cover a wider range of frequencies than was needed for telegraphy.

⁹Lenz's law again.

Astonishingly, Heaviside discovered that dispersion could be eliminated by *intentionally introducing leakage* between the conductors of a cable. Problem 17.4 has the details.¹⁰

17.3.3 Energy of magnetic fields

Imagine charge subdivided into packets Δq . The retarding force on one packet is $\vec{E}_\varphi \Delta q$. To get all the way through the wire, some external source must push this packet through a distance $2\pi wN$, doing work equal to the force times this distance. Moreover, packets arrive at rate $I/(\Delta q)$. Multiplying everything together, energy must be supplied at the rate

$$d\mathcal{P} = (dq\vec{E}_\varphi)(\vec{v}_\varphi) = (\rho_q^{(1D)} ds) \frac{\mu_0 Na}{2h} \frac{dI}{dt} \frac{I}{\rho_q^{(1D)}}.$$

Summing over all charges in the coil means replacing ds by $2\pi aN$, so the total power is

$$\mathcal{P} = \mu_0 \frac{\pi a^2 N^2}{h} \frac{d}{dt} \left(\frac{I^2}{2} \right).$$

To find the total energy cost to bring current up from zero to I , integrate the above formula over time. We can do that by just dropping the time derivatives:

$$\mathcal{E} = \mu_0 \frac{\pi a^2 N^2}{h} \frac{I^2}{2} = \frac{1}{2} LI^2. \quad (17.11)$$

In the last form, L is the self-inductance introduced in Equation 17.7.

It's important not to confuse the work \mathcal{E} with “frictional” loss (Joule heating due to ohmic resistance). Resistive losses occur even when current is held steady, and only in resistive media (not superconductors). In contrast, the energy cost \mathcal{E} just computed applies even to superconductors, but only when current is *changing*. Moreover, all the energy that we invest in increasing the current can be *recovered* if later we let the current decrease—the induced electric field also opposes *that* change, and can even be used to extract the same amount of useful work that we expended when we set up I . A superconducting coil *stores* energy; it doesn't *dissipate* it.

We get a very suggestive result if we reexpress our answer in terms of the magnetic field:

$$\mathcal{E} = \frac{\pi a^2 h}{2\mu_0} \|\vec{B}\|^2. \quad (17.12)$$

This formula has forgotten everything about the original geometry except the *volume* of the region with significant fields. It suggests that there's energy inside the cylinder, with *volume density* equal to a constant times the field strength squared.

Equation 17.12 is reminiscent of a result we got long ago for capacitors:¹¹ We found that the energy needed to charge up a capacitor is $\frac{1}{2}\epsilon_0 \|\vec{E}\|^2$ times the *volume* of the region with nonzero electric field. That result suggested that there's stored energy

¹⁰Not so astonishingly, given his personality, Heaviside neglected to patent his very practical discovery, so others made a fortune from it.

¹¹See Equation 5.2 (page 56).

between the capacitor plates, again with volume density given by a constant times field strength squared. That is:

Maxwell's equations for electrodynamics appear to be compatible with energy conservation if we attribute energy density to empty space. Specifically, we've found energy density $\frac{1}{2}(\epsilon_0 \vec{E}^2 + \vec{B}^2/\mu_0)$. (17.13)

How are these energies stored? Apparently not in any medium—our coil and capacitor each have nothing inside! A more precise version of this question is, “Can we prove a general statement of the conservation of energy, in which electric and magnetic fields themselves can store it?” We’ll pursue this in Chapter 34. Right now, we have just circumstantial evidence in special cases (a parallel-plate capacitor and a solenoid).

In Maxwell’s time, the answer seemed obvious. Paraphrasing what many believed:

“So-called vacuum, which you get by removing all the air from a vessel, is *still full of stuff*, the ‘æther.’ An electric field stretches that stuff, storing elastic energy. A magnetic field sets it in motion, storing kinetic energy.”

We’ll soon see that after Einstein, eventually *nobody* believed that proposition.¹² Then the question got more urgent: *What, then, carries the energy?* We’ll return to that story after we understand Einstein.

17.4 MAXWELL'S MODIFICATION TO AMPÈRE'S LAW

We have modified the static equation $\vec{\nabla} \times \vec{E} = 0$ in order to accommodate experimental reality (induction). Now we must modify Ampère’s law for a different reason.

17.4.1 A bold prediction

Hanging Question #D (page 11) raised the issue that we must solve eight Maxwell equations with just six fields \vec{E} and \vec{B} . In statics, Section 14.4.4 (page 167) argued that the field equations are secretly just six independent equations, by taking the divergences of the two curl equations.

Moving beyond statics, we now take the divergence of Faraday’s modified equation, Equation 17.8, and see that it’s still vacuous (always satisfied, doesn’t constrain the fields). But taking the divergence of Equation 17.4 and using the continuity equation now gives

$$0 = \mu_0 \vec{\nabla} \cdot \vec{j} = -\mu_0 \frac{\partial}{\partial t} \rho_q. \quad (17.14)$$

That’s just *false* in nonstatic situations, so we have a problem. However, notice that

(i) The bad right-hand side equals $-\mu_0 \epsilon_0 \vec{\nabla} \cdot \frac{\partial \vec{E}}{\partial t}$, so

¹²Like many overturned ideas, this one had a long half-life. Lenard, Lorentz, and Michelson reportedly never gave up on it.

- (ii) Modifying Ampère’s law could cure this inconsistency, replacing Equation 17.14 by an identity that’s always true and rescuing our 8 → 6 argument. The required modification is just¹³

$$\vec{\nabla} \times \vec{B} = \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} + \mu_0 \vec{j}. \quad \text{Ampère} \quad (17.15)$$

We have arrived at Maxwell’s famous modification of Ampère’s law. The exact reasoning that Maxwell used to motivate it is unclear, although it may have boiled down to assuming that the current associated with bound charge in a real dielectric medium¹⁴ was accompanied by a similar current from the æther that Maxwell and others believed filled empty space.

The argument given above based on mathematical consistency looks pretty trivial to us because we have the clean notation of vector calculus, and clean conceptions of quantities like charge density. What makes us call Maxwell a genius was his ability to see through the fog of the unclear notation and conceptual framework of his day. Indeed, there is no evidence that he was motivated by symmetry of the equations (Heaviside said that much later) nor by the requirement that they be consistent with the continuity of charge.

Nor was Maxwell following some definitive experimental result: the constant of proportionality $\mu_0 \epsilon_0 \approx 1.1 \cdot 10^{-17} \text{ m}^{-2} \text{ s}^2$ on the new term is extremely small, so no experiment existing in his day could directly confirm or refute it. One would need fields with extremely fast time dependence (large time derivative) to start seeing the effects of this hypothetical term on laboratory length scales.¹⁵ Today we say that the conclusive proof was Hertz’s verification of the predicted propagating electromagnetic waves.

We’re all done tinkering with the equations of electrodynamics. Equations 17.1–17.2, 17.8, and 17.15 are Maxwell’s equations in their final form.¹⁶

17.4.2 Boundary conditions

We can now revisit some conclusions we got in electro- and magnetostatics, concerning fields at interfaces. The results that rested on integrating Gauss laws are unmodified in dynamics, because the Gauss laws themselves are unmodified:

$$\Delta B_{\perp} = 0. \quad \text{always} \quad [14.20, \text{page } 169]$$

$$\hat{n} \cdot (\vec{E}_{\text{vac}} - \vec{E}_{(I)}) = -\hat{n} \cdot \vec{P}_{(I)} / \epsilon_0, \quad \text{dielectric/vacuum} \quad [5.13, \text{page } 63]$$

with a similar formula for a dielectric/dielectric interface.

¹³Note that the left hand side of Equation 17.15 can be expressed without any Levi-Civita tensors, if we use the antisymmetric tensor representation of the magnetic field. And the right side certainly doesn’t have them, so the whole thing is invariant under spatial inversions.

¹⁴See Section 49.2.1. Ironically we no longer believe in the æther, but the name “displacement current” that this analogy inspired has stuck.

¹⁵Maxwell wrote 1868: “This part of the theory... has not been verified by direct experiment. The experiment would be a very delicate and difficult one.”

¹⁶We will, however, build a useful alternate version of these equations to describe electromagnetism in media without having to handle every electron explicitly. That version is an approximation to the equations written here, which are more fundamental and universal.

Turning now to the results that rested on integrating the Faraday and Ampère laws, we find that they, too are unchanged! That’s because the time derivative terms are to be multiplied over an area that goes to zero in the limit of a narrow rectangle in Figure 5.3b or Figure 14.2b. Thus,

$$\Delta \vec{E}_{\parallel} = 0 \quad \text{and} \quad [5.15, \text{page } 63]$$

$$\Delta \vec{B}_{\parallel} = \mu_0 \vec{j}^{(2D)} \times \hat{n}, \quad [14.21, \text{page } 170]$$

where $\vec{j}^{(2D)}$ is the net 2D charge flux at the surface.

17.5 WAVE SOLUTIONS

Section 17.4 suggested that Maxwell’s modification to Ampère’s law might not be quantitatively important in experiments. But let’s keep an open mind, and look for solutions to the modified equations. They look a bit complex—lots of equations in lots of unknowns. Let’s try to eliminate \vec{B} , arriving at a smaller set of equations just involving \vec{E} . Also, let’s simplify by looking at empty space, a region with no charges nor currents. Certainly we know lots of *static* solutions applicable to that situation.

To do the elimination, consider taking the curl of both sides of the curl equations. In vacuum,

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{E}) = -\frac{\partial}{\partial t} \vec{\nabla} \times \vec{B}$$

or (using the Gauss law)

$$-\nabla^2 \vec{E} = -\mu_0 \epsilon_0 \frac{\partial^2}{\partial t^2} \vec{E}. \quad (17.16)$$

Maxwell noticed that this is an example of a *wave equation*. Substitute the trial solution

$$\vec{E}(t, \vec{r}) = \vec{E} \cos(kz - \omega t), \quad (17.17)$$

where \vec{E} is any real, constant vector, k is a real constant, and ω is a real positive constant. Such a vector field is called a **plane wave**, because there is a stack of planes (each parallel to the xy plane), on each of which it is constant. The wave moves at speed ω/k .

Substituting the trial solution into Equations 17.16 and 17.1 gives the condition for the trial solution to work:

$$k^2 = \mu_0 \epsilon_0 \omega^2 \quad \text{and} \quad \hat{z} \cdot \vec{E} = 0. \quad (17.18)$$

We call ω the **angular frequency** (dimensions \mathbb{T}^{-1} , SI unit rad/s) and k the **wavenumber** (dimensions \mathbb{L}^{-1} , SI unit rad/m). Note that “radian” is a dimensionless unit of angle (because it equals circumference divided by radius), and many authors omit it when stating numerical values of ω and k . But that risks confusion with the related quantities:

- .. **frequency** $\nu = \omega/(2\pi)$ (dimensions \mathbb{T}^{-1} , SI unit s^{-1} , also called Hz)
- .. **spectroscopic wavenumber** $k/(2\pi)$ (dimensions \mathbb{L}^{-1} , SI unit m^{-1}). Some authors call our k the “angular wavenumber” to avoid confusion with this quantity.

Additional descriptors include:

- .. **period** $T = 1/\nu$ (dimensions \mathbb{T} , SI unit s)
- .. **wavelength** $\lambda = 2\pi/k$ (dimensions \mathbb{L} , SI unit m).

The period is how long you have to wait at fixed position for the wavefront to repeat. The wavelength is how far you have to travel at a fixed instant of time for the wavefront to repeat.

Your Turn 17B

Confirm that Equations 17.17–17.18 really do yield a solution to all of Maxwell’s equations, not just the one combination Equation 17.16. You’ll need to find the appropriate $\vec{B}(t, \vec{r})$ first.

Combining the result you just found for \vec{B} with Equation 17.17 shows that the fields form a wave traveling at speed ω/k . Equation 17.18 gives that speed as $(\epsilon_0\mu_0)^{-1/2}$, independent of the amplitude $\|\vec{E}\|$, the polarization (direction of \vec{E}), or the direction of travel (sign of k). It’s a constant of Nature, which we’ll call c . Substituting numerical values shows that Maxwell’s modification leads to *wave solutions that travel at about three hundred million meters per second*. That rang a bell for Maxwell.¹⁷

17.6

We have anchored each ingredient in the Maxwell equations, including the sign of each term, using an observable electromagnetic phenomenon. The only exception is Maxwell’s term, which we introduced to salvage mathematical consistency. Yet even this term led to a testable prediction: Solutions that resemble the behavior of *light*. The obvious hypothesis is that *light is itself an electromagnetic phenomenon*. But there are a lot of aspects to light, which must all be checked to see if the equations correctly predict them. So we need to work on those, after introducing some helpful machinery in the following sections. First, however, a few remarks:

- There are many other interesting solutions besides plane waves, for example, *spherical waves* that spread from a point (Chapter 37).
- Because Maxwell’s equations are linear, we can get more solutions by superposing (adding) the fields of two solutions at each point of spacetime. So the rich world

¹⁷“Wilhelm Eduard Weber and Rudolf Kohlrausch demonstrated in 1856 that the ratio of electrostatic to electromagnetic units [today $(\epsilon_0\mu_0)^{-1/2}$] produced a number that matched the value of the then known speed of light. This finding led to Maxwell’s conjecture that light is an electromagnetic wave. Also, the first usage of the letter c to denote the speed of light was in an 1856 paper by Kohlrausch and Weber” [https://en.wikipedia.org/wiki/Wilhelm_Eduard_Weber]. Specifically they obtained $c = 3.10740 \cdot 10^8$ m/s, fairly close to Fizeau’s measured 3.14850. “Maxwell was impressed, as Kirchhoff had been before him, by the close agreement between the electric ratio and the velocity of light, and he did not hesitate [in 1862] to assert the identity of the two phenomena [Whittaker, 1951, p254].” (“He had worked out the formulae in the country, before seeing Weber’s result” [Campbell and Garnett, *Life of Maxwell* p244].) Later (1868), Maxwell and C Hockin made an improved measurement of $(\epsilon_0\mu_0)^{-1/2} \approx 2.88 \cdot 10^8$ m/s, and compared it to Foucault’s improved measurement of light speed 2.9836.

of interference phenomena observed with light and other EM radiation is all contained in the electromagnetic-wave theory of light.

- All kinds of wave phenomena display interference, e.g. sound, ripples on water, etc. We have the more specific prediction that there are *polarizations* of light corresponding to the directions transverse to the direction of propagation. Indeed, we noticed transverse polarization effects in the demo that generated microwaves via electric currents. And visible light was already well known in Maxwell's time to display two independent polarizations, a detailed agreement with the electromagnetic theory of light. In contrast, there is only one kind of sound wave in air or water (one "polarization"). Sound in a rigid solid like steel has a *three-dimensional* space of polarizations, because steel can elastically resist both compression (longitudinal) and shear (transverse) deformation. Thus light differs from all kinds of sound by having *no longitudinal polarization*.
- Notice that in Your Turn 17B, \vec{E} and \vec{B} are perpendicular to each other, and each is perpendicular to the direction of motion \hat{z} . Also notice that each varies sinusoidally with time and space, and they are *in phase* with each other. So at any instant, there are periodically-spaced planes where both equal *zero!* Normally we don't notice that, because waves rush around so fast that we can only perceive the time-averaged fields. But we can use superposition to create a standing wave, and it really does have points with zero field.

17.7 COMPLEX NOTATION

Our trial solution Equation 17.17 involved the cosine function, but there is nothing special about cosines.

Your Turn 17C

Use the chain rule to show that $\vec{E}(t, \vec{r}) = \vec{E} f(kz - \omega t)$ works for *any* function f , and constant vector \vec{E} , as long as Equation 17.18 is satisfied and $\vec{E} \perp \hat{z}$.

That is, no matter what waveform we choose, the speed is always $\approx 3 \cdot 10^8$ m/s.

Nevertheless, sines and cosines are a convenient basis, from which any waveform can be constructed by Fourier synthesis. An even more convenient basis is the complex exponentials; we will usually write waves in terms of the functions

$$\Phi_{\vec{k}, \omega}(t, \vec{r}) = e^{i(\vec{k} \cdot \vec{r} - \omega t)}. \quad (17.19)$$

Of course, \vec{E} and \vec{B} must still be real-valued vector fields, so in any formula involving $\Phi_{\vec{k}, \omega}$ we will eventually need to take the real part to get the physical fields. But in intermediate steps, the complex notation is often nicer. That's because sine and cosine exchange roles under differentiation, whereas the derivative of exponential is always still exponential:

$$\frac{\partial \Phi_{\vec{k}, \omega}}{\partial t} = -i\omega \Phi_{\vec{k}, \omega}, \quad \vec{\nabla}_j \Phi_{\vec{k}, \omega} = i\vec{k}_j \Phi_{\vec{k}, \omega}.$$

Let's use complex notation to redo what was done in the preceding section, and extend it in two ways. We'll write a trial solution of the form

$$\vec{E}(t, \vec{r}) = \frac{1}{2} \vec{E} \Phi_{\vec{k}, \omega}(t, \vec{r}) + \text{c.c.} \quad (17.20)$$

The notation "c.c." denotes the complex conjugate of whatever precedes it, and guarantees that the overall expression is real. The factor of one half says that specifically we are taking the real part of the first term. The notation \vec{E} refers to a constant vector, called the **complex amplitude** of the real vector field $\vec{E}(t, \vec{r})$.

The two extensions we are considering are that

- . The wavevector \vec{k} need not point along \hat{z} .
- . The polarization vector \vec{E} need not be real. Write it as $\vec{E}^{(R)} + i\vec{E}^{(I)}$.

We now impose the Maxwell equations one by one.

17.7.1 Electric Gauss law

Spatial gradients are easy to compute by the rule $\vec{\nabla} \rightarrow \pm i\vec{k}$, so Equations 17.19 and 17.20 give

$$0 = \frac{1}{2} i\vec{k} \cdot \vec{E} \Phi_{\vec{k}, \omega} + \text{c.c.} = \frac{1}{2} i\vec{k} \cdot [\vec{E}^{(R)} i \sin(\dots) + i\vec{E}^{(I)} \cos(\dots)] + (\text{two more terms}) + \text{c.c.}$$

The third and fourth terms on the right get clobbered by taking the real part. The ellipses denote $\vec{k} \cdot \vec{r} - \omega t$.

Equation 17.7.1 must hold at every point of space, at every time. The only way this could happen is if the coefficients of $\sin(\dots)$ and $\cos(\dots)$ *separately vanish*. So each of $\vec{k} \cdot \vec{E}^{(R)} = 0$ and $\vec{k} \cdot \vec{E}^{(I)} = 0$ must hold, or

$$\vec{k} \cdot \vec{E} = 0. \quad (17.21)$$

In short, when dealing with linear expressions in the fields, we don't need to think explicitly about the complex conjugate terms. From now on, we'll abbreviate logic like the foregoing by passing directly from an equation of the form $\frac{1}{2} \bar{b} \Phi_{\vec{k}, \omega} + \text{c.c.} = 0$ to conclude that $\bar{b} = 0$, where \bar{b} is some complex constant.¹⁸

For the special case where $\vec{E} = \hat{z}$, Equation 17.21 is the same transversality condition that we found earlier (Equation 17.18).

17.7.2 Faraday law

If \vec{E} is a plane wave, it seems a reasonable guess that \vec{B} will be too, so extend the trial solution:

$$\vec{B}(t, \vec{r}) = \frac{1}{2} \vec{B} \Phi_{\vec{k}, \omega}(t, \vec{r}) + \text{c.c.},$$

where \vec{B} are three more unknown complex constants. Note that we allow for the possibility that the magnetic field's variation may be shifted in phase relative to that

¹⁸Nonlinear expressions will require more care; see Section 17.10.

of the electric field: One advantage of the complex exponential notation is that such a shift can be represented as a multiplicative factor in the coefficients \vec{B} .

Again we may replace $\vec{\nabla}$ by $\pm i\vec{k}$, and now $\partial/\partial t$ by $\pm i\omega$. Thus, Faraday becomes

$$\frac{1}{2}i\vec{k} \times \vec{E}\Phi_{\vec{k},\omega} + \text{c.c.} = -(-i\omega)\vec{B}\Phi_{\vec{k},\omega} + \text{c.c.}$$

Solving gives

$$\vec{B} = (\vec{k}/\omega) \times \vec{E}.$$

We conclude that \vec{B} must be perpendicular to \vec{k} , and also to \vec{E} . Moreover, the spatial and temporal variation of \vec{B} match that of \vec{E} (no relative phase shift). We see this from the fact that \vec{B} is a *real* constant times \vec{E} . These results generalize what you found in Your Turn 17B.

17.7.3 Magnetic Gauss law

Similar logic as before reduces this equation to $\vec{k} \cdot \vec{B} = 0$, but we already know that from the Faraday law. Thus we get no additional restriction on our trial solution.

17.7.4 Ampère law

$$i\vec{k} \times \vec{B}\Phi_{\vec{k},\omega} + \text{c.c.} = c^{-2}(-i\omega)\vec{E}\Phi_{\vec{k},\omega} + \text{c.c.}$$

$$\vec{k} \times \left(\frac{\vec{k}}{\omega} \times \vec{E} \right) = -c^{-2}\omega\vec{E}.$$

Your Turn 17D

Simplify the triple cross product to get $ck = \omega$ as before (Equation 17.18).

17.7.5 Summary

There are plane-wave solutions in vacuum that move in any direction, with any frequency, and any polarization as long as it's perpendicular to the direction of propagation. All such solutions move at the same speed c . All have \vec{B} perpendicular to, but in phase with, \vec{E} . Each satisfies the **dispersion relation** $ck = \omega$.

17.8 POTENTIALS

17.8.1 Representation of \vec{E} and \vec{B}

We found simplified reformulations of electrostatics and magnetostatics in terms of potentials ψ and \vec{A} . Can we do something similar for the full Maxwell equations?

We still have $\vec{\nabla} \cdot \vec{B} = 0$, so we can still write $\vec{B} = \vec{\nabla} \times \vec{A}$ for some vector potential \vec{A} . However, we no longer have $\vec{\nabla} \times \vec{E} = 0$, so electrons feel a nonconservative force, unlike in statics. That is, there is no function whose gradient is minus the electric field. Nevertheless, there *is* a vector quantity whose curl equals zero, namely $\vec{E} + \frac{\partial \vec{A}}{\partial t}$, so we

can construct a “scalar potential” whose gradient equals that quantity. We’ll continue to call it $-\psi$, but keep in mind that ψ can no longer be interpreted as potential energy of a test body per unit charge. In short, we can always find potential functions such that

$$\vec{E} = -\vec{\nabla}\psi - \frac{\partial \vec{A}}{\partial t} \quad \text{and} \quad \vec{B} = \vec{\nabla} \times \vec{A}. \quad (17.22)$$

Equations 17.22 let us express six unknown fields (\vec{E} and \vec{B}) in terms of just *four* unknown potentials (\vec{A} and ψ), a significant simplification.¹⁹ We will soon see that further simplifications arise when we substitute this representation into Maxwell’s equations.

17.8.2 Gauge invariance

One key idea about potentials in the static case was gauge invariance.

Your Turn 17E

Show that the generalized transformation

$$\vec{A} \rightarrow \vec{A} + \vec{\nabla}\Xi, \quad \psi \rightarrow \psi - \frac{\partial \Xi}{\partial t} \quad \text{gauge transformation}$$

doesn’t change the electric or magnetic fields.

17.8.3 Coulomb gauge

Thus again, the potentials are ambiguous, and we can use that fact to insist on a subsidiary condition if doing so simplifies our equations. For the moment, we’ll again impose the Coulomb gauge choice

$$\vec{\nabla} \cdot \vec{A} = 0.$$

The proof that this is always possible is the same as it was in statics (Section 14.2.4), because we have not modified the gauge transformation formula for \vec{A} .

We can now substitute $\vec{E} = -\vec{\nabla}\psi - \frac{\partial \vec{A}}{\partial t}$ and $\vec{B} = \vec{\nabla} \times \vec{A}$ into the Maxwell equations and simplify by using the Coulomb gauge condition $\vec{\nabla} \cdot \vec{A} = 0$. As in statics, $\vec{\nabla} \cdot (\vec{\nabla} \times \vec{A}) = 0$ is now an identity, so we can forget the magnetic Gauss law. Also, Faraday’s law becomes an identity, so forget it too. We are left with four equations in the four unknowns \vec{A} and ψ :²⁰

$$\nabla^2 \psi = -\rho_{\text{q}}/\epsilon_0 \quad (\text{electric Gauss}), \quad \text{and}$$

$$\nabla^2 \vec{A} = -\mu_0 \vec{j} + \mu_0 \epsilon_0 \left(\vec{\nabla} \frac{\partial \psi}{\partial t} + \frac{\partial^2 \vec{A}}{\partial t^2} \right). \quad (\text{Ampère})$$

It’s tempting to say that we have just found another resolution of Hanging Question #D (page 11) (“eight equations in six unknowns”), but we must be a bit careful.

¹⁹This result addresses Hanging Question #F.

²⁰The derivation of this formula depended on our default choice of cartesian coordinates. The left-hand side would look more complicated in curvilinear coordinates.

The four equations just given are only correct if $\vec{\nabla} \cdot \vec{A} = 0$, which looks like a fifth equation constraining the potentials. However, when we take the divergence of the second equation, and substitute the first, we find that this combination is vacuously satisfied; it does not constrain the potentials. So effectively we do have four equations in four unknowns.

17.8.4 The case of zero charge density

We can simplify still more if we're studying a region with zero net charge density.²¹ (There can still be *currents*, however, as in a neutral wire.)

Suppose that we have described our fields with a vector potential in Coulomb gauge. Then we *still* have some freedom to apply a further gauge transformation: Transforming with any function Ξ that obeys $\nabla^2 \Xi = 0$ will not spoil the Coulomb gauge condition. Let's try

$$\Xi(t, \vec{r}) = \int_{t_0}^t dt_* \psi(t_*, \vec{r}).$$

Your Turn 17F

Show that:

- This choice of gauge transformation preserves Coulomb gauge, and
- This choice of gauge transformation eliminates the scalar potential altogether (transforms it to zero).

There can still be electric fields, of course—they are just represented by minus the time derivative of \vec{A} (Equation 17.22). In short, we found that in vacuum we can reduce still further from four unknown potential functions to just three.

Your Turn 17G

- Show that the electric Gauss law is now just an identity.
- Show that what remains is just three *independent* equations in three unknowns:

$$\nabla^2 \vec{A} = -\mu_0 \vec{j} + \mu_0 \epsilon_0 \frac{\partial^2 \vec{A}}{\partial t^2} \quad \text{in Coulomb gauge if } \rho_q = 0 \text{ and } \psi = 0. \quad (17.23)$$

Here, then is another resolution to Hanging Question #D, for the special case where net charge density is zero. Again we must be careful, but again the additional condition $\vec{\nabla} \cdot \vec{A} = 0$ is balanced by the fact that the divergence of Equation 17.23 is vacuously satisfied.

²¹Actually, Chapter 37 will achieve a similar simplification even with charges present, but we don't need that much power yet.

17.9 WAVES, AGAIN

We can use the representation of fields by potentials to explore plane wave solutions in vacuum more systematically. In Your Turn 17G you showed that all of Maxwell's equations in vacuum reduce to three copies of the scalar wave equation, supplemented by $\vec{\nabla} \cdot \vec{A} = 0$. From this, we can quickly recover the results in Section 17.5, and other results too. For example, the plane wave solutions moving along \hat{z} take the form

$$\vec{A}(t, \vec{r}) = \frac{1}{2} \vec{\zeta} \Phi_{\vec{k}, \omega}(t, \vec{r}) + \text{c.c.}, \quad (17.24)$$

where $\vec{k} = k\hat{z}$, the **polarization vector** $\vec{\zeta}$ is a constant vector in the xy plane, and $\Phi_{\vec{k}, \omega}$ is one of the family of complex functions in Equation 17.19.

More generally, we get plane wave solutions moving in *any* direction, as long as \vec{k} and ω obey the **dispersion relation**

$$\|\vec{k}\| = \omega/c \quad \text{where} \quad c = 1/\sqrt{\mu_0 \epsilon_0} \quad (17.25)$$

and $\vec{\zeta}$ is any vector perpendicular to \vec{k} .

We have simplified the Maxwell equations, and streamlined the derivation of plane waves, but it may seem that we have been *too* successful: For any choice of \vec{k} , Equation 17.24 seems to give *three* linearly independent solutions, whereas the analysis in either Section 17.5 or Section 17.7 gave only two (for the two directions perpendicular to \vec{k})! The resolution to this puzzle is that Equation 17.23 is only equivalent to the Maxwell equations in Coulomb gauge, and hence our trial solution will only work if $\vec{\zeta} \perp \vec{k}$.

Your Turn 17H

Work out the electric and magnetic fields arising from the solution Equation 17.24, and hence the relation between the polarization vector $\vec{\zeta}$ and the vector \vec{E} appearing in Equation 17.17. Show that as before, \vec{E} , \vec{B} , and \vec{k} are mutually perpendicular.

17.10 COMPLEX POLARIZATIONS

17.10.1 Linear, circular, elliptical

If $\vec{\zeta}$ is a vector with *real* components, then \vec{E} oscillates about the $\pm\vec{\zeta}$ direction; we say the plane wave is **linearly polarized**, because the tip of its \vec{E} vector oscillates back and forth on a line in the plane perpendicular to \vec{k} .

But there are other options. There's no mathematical problem with a complex polarization vector, just as in our earlier derivation Section 17.7. But this is a new and physically interesting wave.

Your Turn 17I

If you assumed that $\vec{\zeta}$ was real when you checked Equation 17.25, work through it again without this assumption. Specifically, work out $\vec{E} \cdot \vec{k}$ and $\vec{B} \cdot \vec{k}$.

Your Turn 17J

- a. Consider the wave with $\vec{k} = k\hat{z}$ and $\vec{\zeta} = \hat{x} + i\hat{y}$. If we sit at a fixed location in space, say the origin of coordinates, and watch $\vec{E}(t, \vec{0})$ as time goes by, what figure does its tip trace out? Why do you suppose this wave is said to be **circularly polarized**?
- b. Try it again with $\vec{\zeta} = \hat{x} + 2i\hat{y}$ and interpret such **elliptically polarized** solutions.

17.10.2 Circular polarization basis

Starting from a particular \vec{k} , choose a pair of real unit vectors $\hat{\zeta}_{(1)}, \hat{\zeta}_{(2)}$ perpendicular to \vec{k} and forming a right-handed triad with it. That is, $\hat{\zeta}_{(1)} \times \hat{\zeta}_{(2)} = \hat{k}$. Any polarization for the given \vec{k} can be written as a linear combination of these two basis vectors.

Alternatively, we can define

$$\hat{\zeta}_{(\pm)} = (\hat{\zeta}_{(1)} \pm i\hat{\zeta}_{(2)})/\sqrt{2}. \quad \text{helicity basis} \quad (17.26)$$

Any polarization vector $\vec{\zeta}$ can be written as a (complex) linear combination of $\hat{\zeta}_{(1,2)}$, or of $\hat{\zeta}_{(\pm)}$. If the polarization vector is purely along $\hat{\zeta}_{(+)}$, then the wave is said to be circularly polarized with **positive helicity**, and similarly for a pure $\hat{\zeta}_{(-)}$ wave (which is **negative helicity**).²²

17.11 SPHERICAL WAVES

You may ask, “What was the point of redoing everything with potentials? Section 17.5 already found plane waves directly in terms of \vec{E} and \vec{B} , and it wasn’t much easier in Section 17.9.” One answer is that the calculations get harder, and the benefit of the potential formulation becomes more important, when we study spherical waves (Chapter 37):²³ The algebra is much easier than dealing directly with \vec{E} , \vec{B} .

17.12 PLUS ULTRA

In electrostatics, the electric field could be regarded as a mathematical convenience—introducing it into the formulas was optional. We could, after all, just say that all charges exert forces on each other directly, following Coulomb’s law. Although we found a useful concept of electrostatic energy density in the space between capacitor plates, this interpretation, too, was physically optional—we could just say that the energy of a capacitor was the total potential energy of all the separated charges in each others’ force fields.

Waves change everything. We’ll see that shaking (accelerating) a charge generates these waves, and they in turn can shake other distant charges. Suppose that we shake

²²Beware that different authors disagree about the convention for which is positive and which negative.

²³Also see P+S §11.5.4.

a charge for a while, then stop. Suppose too that the nearest other charges are far away. Then there will be a period after the original charge has lost some energy, but before any other charge has gained energy. Hanging Question #H (page 24) already asked: *Where is the energy at that time?*

As mentioned in Section 17.3, Maxwell and his contemporaries believed that the so-called vacuum was actually filled with some substance, the stuff that jiggles when a wave goes by. The fields were just the state of motion and deformation of that stuff, and their stored energy was just its kinetic and deformation energy, just as when sound passes through steel. Einstein realized, however, that this stuff (the “luminiferous æther”) had to have contradictory physical properties. Eventually he concluded that it didn’t exist, or at least not as any material substance. Then the question comes back to us: *If vacuum is truly empty, what carries that energy?* Our answer will be, “It’s in the fields themselves. The propagation of waves is what transports energy through space.”

PROBLEMS

17.1 *Faraday*

A thin ring of copper spins freely in zero gravity, about an axis that includes one of its diameters. The ring’s radius is 0.1 m. Its initial angular velocity is ω_0 , a certain number of radians per second.

At time zero, we turn on a magnetic field \vec{B}_0 , with magnitude 0.02 T and directed perpendicular to the axis of rotation. The ring’s initial kinetic energy gets dissipated in resistive heating of the ring. Calculate the time needed for the angular frequency to decrease to ω_0/e , where e is the base of natural logarithms.

The electrical resistivity of cold-drawn copper is²⁴ $1.7 \cdot 10^{-8} \Omega \text{ m}$, and its mass density is 9000 kg/m^3 . You may assume that the slowdown is gradual, or

$$\frac{d}{dt} \ln \omega \ll \omega_0.$$

17.2 *Feeling the heat*

In this problem, you will develop a simple model for estimating radio-frequency (RF) energy absorption in a patient undergoing an MRI scan.

- a. The wavelength of an RF wave is bigger than a person, so suppose that a spatially uniform, but time-varying magnetic field $\vec{B}(t) = \hat{z}(B_{(0)} + \delta \bar{B} \cos \omega t)$ is applied. Apply Faraday’s law to a circular path in a plane perpendicular to \vec{B} to find the amplitude of the resulting electric field. Your answer depends on the radius R of the circular path; later we will set R to a value comparable to a human radius.
- b. Model the patient as a uniform conductor with electrical conductivity κ . Use the ohmic transport relation to find the average power dissipated in the conductor per

²⁴<http://www.matweb.com>. You may be more familiar with the conductivity, which is the reciprocal of resistivity.

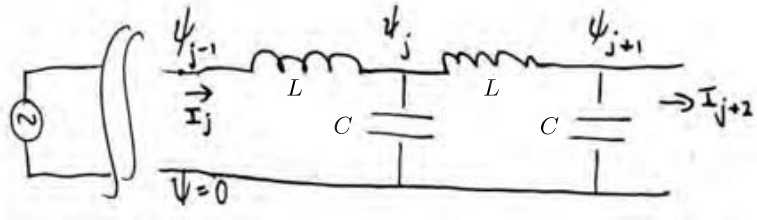


Figure 17.3:

- volume. Actually, the RF signal is not continuous; it consists of pulses of duration Δt which come once every repetition period T_R , so make the appropriate correction.
- It's customary to report the "specific absorbed rate," which is power per unit body mass. Find the SAR in terms of body mass density ρ_m and κ , R , ω , $\delta\bar{B}$, Δt , and T_R .
 - The pulse duration, field strength, and frequency are related by the requirement that the pulse rotate proton spins by an angle $\pi/2$. You can take as given that this requirement amounts to $\delta\bar{B} = 2\pi/(2\gamma \Delta t)$ and $\omega = \gamma B_{(0)}$, where γ is some constant and $B_{(0)}$ is the background magnetic field, a given number. Use this information to eliminate $\delta\bar{B}$ and ω from your formula for SAR.
 - Now substitute typical human values: $R \leq 0.17\text{ m}$, $\kappa \approx 0.3\ \Omega^{-1}\text{m}^{-1}$. And use typical instrument values $B_{(0)} \approx 0.5\ \text{T}$ and $T_R \approx 1\ \text{s}$. Also, the "gyromagnetic ratio" of a proton is $\gamma \approx 2.7 \cdot 10^8\ \text{Hz/T}$.
 - Safety requires that we not heat the patient too much! So demand that $\text{SAR} < 0.4\ \text{W/kg}$. Find the corresponding requirements on Δt and also on $\delta\bar{B}$.

17.3 Lumped-element transmission line

This problem explores a circuit that is sometimes useful for signal conditioning, for example, removing noise known to have a specific frequency. Recently, a filter like this was added to the MicroBooNE experiment's electronics at Fermilab.

The main text introduced a solenoid. More generally, any circuit element that obeys the linear relation Equation 17.10 for some constant L is called an **inductor**.²⁵ You can purchase devices that approach this idealized behavior (approximately, over some frequency range).

Consider a chain of discrete elements with circuit diagram shown in Figure 17.3. The chain contains inductors each with inductance L , and capacitors each with capacitance C . Write equations analogous to the ones we wrote for the cable equation but appropriate to this situation (no resistors). Unlike in the cable equation, however, we will not take any continuum limit.

- Show that the quantity LC has the dimensions s^2 .
- Following the analysis in Chapter 10, eliminate the currents I_j to get an infinite set of coupled, linear, ordinary differential equations in the variables $\{\psi_j\}$. They have constant coefficients, so we expect solutions of the form

$$\psi_j(t) = \frac{1}{2} \bar{\psi}_j e^{-i\omega t} + \text{c.c.} \quad (17.27)$$

²⁵In particular, an idealized inductor has negligible electrical resistance and capacitance.

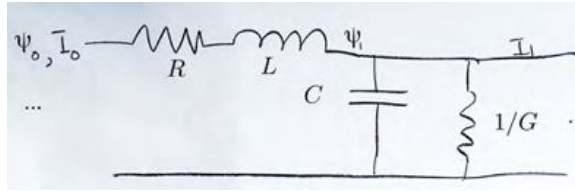


Figure 17.4:

- c. Substitute that trial solution to get an infinite set of coupled *algebraic* equations.
- d. It still looks hard, but your equations are invariant under shifting everything one step in space. So our experience with related systems suggests that we make the trial solution

$$\bar{\psi}_j = \psi_0 e^{ij k} \quad (17.28)$$

- where k is some constant. Substitute this trial solution into your algebraic equations and see what k must be, for a given angular frequency ω , in order to get a solution.
- e. For a certain range of angular frequency values there will be a real solution k to your condition, and a solution of the form Equations 17.27–17.28 describes a wave traveling along the chain to infinity. Outside that range, however, there will be no real solution; the transmission line has a **cutoff**. Find the allowed range of frequencies.

17.4 Realistic transmission line

Figure 17.4 shows another transmission line, but made more realistic by the addition of resistance R along the segment shown and leak conductance G . The figure shows the line as a series of lumped-element circuits, but actually we suppose that all four material properties R , L , C , and G are continuously distributed with densities r , ℓ , c , and g respectively. Thus you should initially consider a segment of length Δx , with $R = r\Delta x$ and so on; at an appropriate moment, take the limit $\Delta x \rightarrow 0$.

The line is infinitely long. We suppose that at some point an external agency imposes a harmonic potential $\psi(0; t) = \frac{1}{2}\bar{\psi}e^{-i\omega t} + \text{c.c.}$ We would like to find the solution everywhere. The problem is time-translation invariant, so a reasonable guess is again harmonic: $\psi(x; t) = \frac{1}{2}\bar{\psi}(x)e^{-i\omega t} + \text{c.c.}$

- a. Follow the strategy in Chapter 10 to write a second-order differential equation for $\bar{\psi}(x)$.
- b. The problem is also spatially translation invariant apart from the imposed boundary condition, so seek a solution of the form $\bar{\psi}(x) = e^{ikx}$ where k is a constant you are to find.
- c. The wavenumber k you found depends on the frequency (the cable's **dispersion relation**). Why would it be desirable for k to take the general form $k = \pm(\omega/v_{\text{cable}} + i\lambda)$, where v_{cable} and λ are constants?
- d. The desirable condition does not generally hold, but Heaviside found that it does hold if the material parameters r , ℓ , c , and g obey a certain relation. Find that condition.
- e. Some resistance R is unavoidable in any long cable. But it had previously seemed that any nonzero value of g would be a bad thing, to be avoided at all costs. Why

did Heaviside disagree?

17.5 Helicity basis

The helicity basis is defined in Equation 17.26, starting from a choice of two vectors $\hat{\zeta}_{(1)}, \hat{\zeta}_{(2)}$ perpendicular to each other and to \vec{k} and forming a right-handed triad with it.

- Show that if we choose a different pair of unit vectors $\hat{\zeta}'_{(1)}, \hat{\zeta}'_{(2)}$, which also make a right-handed, orthonormal triad with \vec{k} , then we get essentially the same helicity basis. That is, $\hat{\zeta}'_{(+)}$ is a constant times $\hat{\zeta}_{(+)}$ and similarly for $\hat{\zeta}'_{(-)}$.
- Also show that $\hat{\zeta}_{(\pm)} \cdot \hat{\zeta}_{(\pm)}^* = 1$ and $\hat{\zeta}_{(\pm)} \cdot \hat{\zeta}_{(\mp)}^* = 0$.
- Compute $\hat{k} \times \hat{\zeta}_{(\pm)}$ and express it in the helicity basis; show that the helicity basis vectors are eigenvectors of the operation “ $\hat{k} \times$.” (This operation generates infinitesimal rotation about \hat{k} .)

17.6 Microwave waveguide

We have been studying wave solutions in infinite empty space, but there are interesting solutions in confined regions as well. P+S devote an entire chapter to this subject, but for the special case discussed here you’re probably better off starting from scratch.

A waveguide consists of an infinitely long rectangular prism along the z axis, made of perfect conductor. It encloses the region $0 < x < a$, $0 < y < b$. Inside is vacuum, where the electric and magnetic fields must obey the wave equation as usual. All we need are solutions obeying boundary conditions appropriate to conductors.

Consider the following trial solution for the electric field in this region:

$$\vec{E}(t, \vec{r}) = \frac{1}{2} E_0(y) \hat{x} e^{-i(\omega t - kz)} + \text{c.c.}$$

Here $E_0(y)$ is a function of y only, which you are to find.

- Find a condition on the function $E_0(y)$ so that the electric field obeys the wave equation, as it must inside the waveguide. Find boundary conditions so that the electric field obeys $\vec{E}_{\parallel} = 0$ on the walls of the cavity. This is a familiar math problem with a series of solutions; find the simplest one (other than $E_0 = 0!$).
- Use Faraday’s law to find the magnetic field corresponding to your solution in (a). The magnetic field needs to satisfy $\vec{B}_{\perp} = 0$ on the walls of the cavity, so impose this.
- Once you have found your solution, examine its dispersion relation (relation between ω and k). In empty space, there are always solutions for any value of ω ; is that the case here? Show that all solutions of this form have “group velocity” smaller than c . (The group velocity is defined as $d\omega/dk$.)

17.7 Zeeman effect

Background: The “Zeeman effect” refers to the effect on atomic spectra of an applied magnetic field. Remarkably we can understand it (partially) without using quantum mechanics.

Problem: Consider a charged particle of mass m and charge q in an isotropic, 3D harmonic oscillator potential: $U(\vec{r}) = \frac{1}{2} k r^2$. The particle moves nonrelativistically; its oscillation is affected very little by the radiation it gives off. It has three independent normal modes of oscillation, all with the same frequency $\omega_0 = \sqrt{k/m}$.



Figure 17.5: Magnet used in Zeeman's experiments.

- Now we place this system in a static external magnetic field \vec{B} directed along the $+\hat{z}$ -axis. Find the frequencies of the resulting oscillation modes. You can suppose that \vec{B} is “small” in any relevant sense, and work to leading nontrivial order in it. [Hint: Treat oscillations in the xy plane together, but separately from those along z . Try to guess two trial solutions for xy motions that will still give solutions to Newton's $\vec{F} = m\vec{a}$, even when \vec{B} is turned on.]
- The frequencies you found in (a) correspond to three kinds of radiation the system can emit. We have not yet systematically worked out the radiation by a moving point charge. However, from the symmetries of the problem and what you do know about light, you should be able to make an educated guess about what kinds of polarizations will be emitted. So find the frequencies and corresponding polarizations of radiation seen by an observer located far away on the \hat{z} -axis.
- Explain how observation of this radiation can be used to determine the charge/mass ratio of the electron, including its *sign*, even if the value of the spring constant k is unknown.
- Evaluate your answer for the frequency shift numerically, assuming $\|\vec{B}\| = 2\text{ T}$. Compare to the frequency of visible light. Is it a big effect?

Comments: Zeeman did this experiment in 1896. Following a suggestion by HA Lorentz, he looked for, and found, the polarization effect discussed in the problem. Lorentz then analyzed the data and found the charge-mass ratio that they implied. Zeeman and Lorentz shared a Nobel Prize for this work. In the second iteration of this experiment, Zeeman obtained a charge/mass ratio for the electron within 10% of the modern value. Still later experiments showed that the effect is much more complicated than the classical picture discussed here. However, the qualitative conclusion about the *sign* of q/m is valid.

Some highly magnetized stars (**magnetars**) have much bigger B than what is attainable in the lab, so this effect gives a useful way to establish the value of B on a distant object.

CHAPTER 18

First Look at Energy and Momentum Transport by Waves

Section 17.12 suggested that “EM fields store energy; the field equations have traveling wave solutions; therefore such a wave *carries* energy.” We’ll eventually make a general framework for studying this claim, but first let’s do some simple calculations a concrete situation.

Sound and water waves carry energy: Sound can actuate those tiny bones in your inner ear; the tsunami brings the earthquake to your shores. So it’s not surprising that EM waves can also carry energy, though the details are significantly different from the other cases. We’ll see, however, that light also does some completely new things: It also transports linear and angular *momentum*.

18.1 LINEAR POLARIZATION

18.1.1 Energy transport

As in Chapter 17, we’ll make the useful abbreviation

$$\Phi_{\vec{k},\omega}(t, \vec{r}) = e^{i(\vec{k}\cdot\vec{r}-\omega t)}, \quad (18.1)$$

and consider a solution to Maxwell’s equations that propagates along the $+\hat{z}$ direction and is linearly polarized along \hat{x} :

$$\vec{E} = \frac{1}{2}i\omega\bar{\zeta}\hat{x}\Phi_{k\hat{z},\omega} + \text{c.c.}, \quad \vec{B} = \frac{1}{2}ik\bar{\zeta}\hat{y}\Phi_{k\hat{z},\omega} + \text{c.c.} \quad (18.2)$$

Here $\bar{\zeta}$ is a real scalar constant.

We suppose that this wave travels through empty space, then impinges on a charged particle with charge q . The particle is constrained to move only in the xy plane, that is, the plane $z = 0$, so we will denote its trajectory as $\vec{r}_\perp(t)$. We assume that within that plane, its motion is damped by viscous friction with coefficient η . That is, it feels a friction force $-\eta(d\vec{r}_\perp/dt)$.

In the limit of strong friction, we may neglect inertia in Newton’s law of motion:

$$0 = -\eta\frac{d\vec{r}_\perp}{dt} + q(\vec{E} + \frac{d\vec{r}_\perp}{dt} \times \vec{B})_\perp.$$

The second term on the right equals zero, because $d\vec{r}_\perp/dt$ and \vec{B} both lie in the xy plane, so their cross product has no component in that plane. Thus

$$\frac{d\vec{r}_\perp}{dt} = \frac{q\vec{E}}{\eta}.$$

We can now find the rate at which the field does work on the particle. Because the particle is constrained to move only in the xy plane, and we assumed $\bar{\zeta}$ is real,

$$\mathcal{P} = \vec{f}_\perp \cdot \frac{d\vec{r}_\perp}{dt} = q^2 \|\vec{E}\|^2 / \eta \quad (18.3)$$

$$= \frac{q^2 \omega^2 \bar{\zeta}}{4\eta} \|\hat{x}e^{-i\omega t} - \hat{x}e^{+i\omega t}\|^2 = \frac{q^2 \omega^2}{\eta} \bar{\zeta}^2 (\text{Im } e^{-i\omega t})^2. \quad (18.4)$$

This quantity is always greater than or equal to zero. Its time average is

$$\langle \mathcal{P} \rangle = \frac{q^2 \omega^2}{2\eta} \bar{\zeta}^2. \quad (18.5)$$

Your Turn 18A

- Check that the units in this formula (and every formula) make sense.
- Also, redo this derivation for the more general case in which η is not so huge, so that we must also account for the inertial term $m\ddot{x}$ in Newton's law. Check that the limits $m \rightarrow 0$ and $\eta \rightarrow \infty$ holding frequency fixed work the way you expect.

So far, our result is not very surprising: Like any wave, an EM wave carries energy proportional to its amplitude squared. The charged particle can extract some of that energy, much as a cork floating on water extracts kinetic energy from passing waves.

18.1.2 Momentum

Even though we assumed our particle was constrained to move only in the xy plane, still it can feel forces in every direction. You might expect that because force is a vector, unlike energy, such forces would average out to zero. Indeed the electric force, which is directed along \hat{x} , does follow that expectation. But a moving particle will also experience a magnetic force directed along \vec{k} :

$$\vec{f}_\parallel = q \left(\frac{d\vec{r}_\perp}{dt} \times \vec{B} \right)_\parallel = q \left(\frac{q\vec{E}}{\eta} \times \vec{B} \right)_\parallel. \quad (18.6)$$

We now substitute Equation 18.2:

$$\begin{aligned} \vec{f}_\parallel &= \frac{q^2 \omega k}{\eta} \frac{1}{4} \left((i\bar{\zeta}\hat{x}e^{-i\omega t} + \text{c.c.}) \times (i\hat{y}\bar{\zeta}e^{-i\omega t} + \text{c.c.}) \right)_\parallel \\ &= -\frac{q^2 \omega k \bar{\zeta}^2}{4\eta} (e^{-i\omega t} + \text{c.c.})^2 = \frac{q^2 \omega k \bar{\zeta}^2}{\eta} (\text{Im } e^{-i\omega t})^2. \end{aligned} \quad (18.7)$$

The time average is then

$$\langle \vec{f}_\parallel \rangle = \frac{q^2 \omega k \bar{\zeta}^2}{2\eta}. \quad (18.8)$$

Recall that force is the rate of momentum transfer. So the wave continually transfers momentum to the particle, or in other words the particle continually extracts momentum from the wave.

Your Turn 18B

As before, generalize the calculation to include the inertia term, and check the limits $m \rightarrow 0$ and $\eta \rightarrow \infty$ for reasonableness.

Our result has no counterpart with, say, sound waves: Sound in air involves pressure variation. It can *shake* things along its direction of propagation, but gives no *net* push. Even sound in, say, steel, which can have transverse polarizations, only shakes things. In contrast, we just found net momentum transport.

Our derivation still suffers from the same critique as in the preceding section: We see that the wave carries momentum, but we don't yet know how much. All we found was how much momentum one particular system can extract.

18.1.3 Some electromagnetic phenomena

Just knowing that light can transport momentum, and that the delivered momentum is in the direction of its propagation, already gives us a lot of physics.

- This “radiation pressure” phenomenon underlies the observation that a comet’s dust tail always streams away from the comet in the direction away from the Sun.
- At the earliest times after the Big Bang, radiation pressure dominates over the gas pressure of ordinary matter, so it is crucial for cosmology.
- It supplements ordinary gas pressure in stars, opposing gravitational collapse (until the nuclear fuel is exhausted).
- It detonates thermonuclear bombs.
- It allows exquisitely fine manipulation of micrometer-size objects via optical tweezers.
- One day it may even provide a tiny but inexhaustible source of impulse for “solar sail” spacecraft.¹

18.2 LIGHT CANNOT BE INTERPRETED AS A STREAM OF NEWTONIAN PARTICLES

Although we haven't found the absolute energy or momentum content of a wave, something interesting comes up if we divide the results of the two preceding sections:

$$\frac{\text{rate of energy extraction}}{\text{rate of momentum extraction}} = c. \quad (18.9)$$

Everything specific to our silly little imagined system (charge q , friction coefficient) *cancels out* of this universal ratio.

Your Turn 18C

Confirm that the particle mass m , which you added in Your Turns 18A–18B, also drops out.

¹See Problem 18.1.

So it's plausible that this result will have far greater generality, and will continue to apply to *all* the energy and momentum carried by a plane wave.

This result gains further significance in the quantum theory of light. That is a dual picture of light as a stream of particles, each with energy $\mathcal{E} = \hbar\omega$. Our charged particle intercepts and absorbs some of them at a rate r . Equation 18.9 then implies that each particle of light must also carry linear momentum $p = \hbar\omega/c$, or

$$\mathcal{E} = pc.$$

That result sounds paradoxical: Newtonian mechanics instead says that $\mathcal{E} = p^2/(2m) = pv/2$! It's no accident that when Albert Einstein was working on his light-quantum hypothesis, he was *also* working out special relativity. Chapter 30 will give his resolution to this apparent paradox.

18.3 OTHER POLARIZATIONS

We now drop the assumption that our wave be linearly polarized along \hat{x} . That is, let $\vec{\zeta}$ be any complex vector satisfying $\vec{\zeta} \cdot \vec{k} = 0$.

Your Turn 18D

- Start from Equation 18.3 and find the analog of Equation 18.5 in this situation.
- Start from Equation 18.6 and find the analog of Equation 18.8.
- Is Equation 18.9 still true in this more general situation?

Equations 18.4 and 18.7 showed that the power and force transmitted to a particle by a linearly polarized wave fluctuate (though they don't change sign).

Your Turn 18E

- Show that on the contrary, if the wave is circularly polarized then the power and force are both *constant* in time.
- Show that for elliptical polarization, we get something in between those extremes.

18.4 ANGULAR MOMENTUM

Section 17.10.1 showed that there are plane waves in which the electric and magnetic fields twirl around the axis of propagation, instead of shaking along a fixed direction. You can generalize the preceding thought-experiment to a situation in which the charged particle is only constrained to move in the xy plane. It may seem reasonable that then the particle will execute uniform circular motion, in a direction determined by the wave's helicity. That motion requires a *torque*, or in other words the transfer of *angular* momentum from the wave to the particle (which in turn is coupled by friction to the surrounding fluid that we imagined). You'll work out details in Problem 18.2.

FURTHER READING

Historical:

JH Poynting predicted the phenomenon of radiation pressure in 1884. P Lebedev, and independently E F Nichols and G Ferrie Hull, detected its effect on macroscopic objects and absorbing gases in 1901.

Poynting also predicted angular momentum of EM fields (Proc. Roy. Soc. Lond. **A82**:560(1909)). Experimental discovery: Beth, R. A. (1935). Direct Detection of the Angular Momentum of Light. Phys. Rev., 48, 471–471 <http://doi.org/10.1103/PhysRev.48.471> (Beth, 1935). Beth, R. A. (1936). Mechanical Detection and Measurement of the Angular Momentum of Light. Phys. Rev., 50, 115–125 <http://doi.org/10.1103/PhysRev.50.115> (Beth, 1936).

In the early 1970s, Arthur Ashkin showed that laser-induced optical forces could be used to alter the motion of microscopic particles and neutral atoms, work honored in 2018 with a Nobel Prize.

Modern:

Optical tweezers: Jones et al., 2015; Perkins, 2014.

T_2 Track 2

18.1.2' Pondermotive force

[Not ready yet.]

PROBLEMS

18.1 *Radiation pressure*

“Yuri Milner, a Russian physicist and billionaire investor, announced a plan to develop the technologies that interstellar flight would need. Mr. Milner is devoting himself to the challenges of deep space... He is going to spend \$100m on a “Breakthrough Starshot” research programme.” – *The Economist*, April 2016.

Sounds crazy, but maybe you could get a little slice of that \$100m. (Previously he pledged a *different* \$100m to SETI.)

Milner’s idea is to power a tiny spacecraft—with mass just *five grams*—by radiation pressure from a humungous laser based on Earth. The *Economist* makes it all clear by stating that “A gigawatt laser beam—roughly the power output of a large nuclear plant—provides a force equivalent to that required to lift a glass of beer.”

- a. Estimate the attainable force and see if the *Economist* got it right. If that last quote is not precisely phrased, for example, if it’s missing some other parameter describing the spacecraft or laser, choose some parameter value(s) that seem reasonable to you and that allow a precise statement.
- b. Milner’s plan involves illuminating a reflector on the tiny spacecraft for ten minutes. The spacecraft is launched from outside Earth’s atmosphere (no air resistance). With the acceleration corresponding to the force you found in (a), how fast would the spacecraft be flying at the end of ten minutes?

18.2 *Angular momentum transport*

Suppose that a plane, circularly polarized electromagnetic wave of angular frequency ω travels along the $+\hat{z}$ direction.

- a. Write the electric and magnetic fields analogous to Equations 18.2, again parameterized by a single real constant $\bar{\zeta}$ with appropriate dimensions.

The wave encounters a point charge q . Again, the charge is free to move in the xy plane. There is friction slowing it down, so its equation of motion is

$$m(d^2\vec{r}_\perp/dt^2) = -\eta(d\vec{r}_\perp/dt) + (\text{Lorentz force}).$$

Neglect any radiation by the charge, and again neglect the left-hand side of the above formula (pretend that it’s zero).

- b. Find the steady state solution to the equation of motion for the charge. Your formula will involve \bar{E} , q , and other constants.
- c. The EM field does work against friction. Let \mathcal{P} be the rate at which it does this work, averaged over a cycle. Find \mathcal{P} .
- d. The wave also pushes the charge in the xy plane, exerting a *torque* τ_z . Find the average of this torque over a cycle.
- e. The ratio $\langle\tau_z\rangle/\mathcal{P}$ has a remarkably simple form: Find it in terms of the parameters in the problem.
- f. Following Section 18.2, momentarily unlock the quantum part of your brain and reinterpret your answer (e) in terms of a stream of little packets, each carrying a lump of energy \mathcal{E}_* and a lump of angular momentum L_* . That is, interpret your

answers to (c,d) as saying that the charge absorbs some of these lumps; then make a statement about the relation between \mathcal{E}_* and L_* using your result in (e). Draw a conclusion about the intrinsic angular momentum carried by one packet.

CHAPTER 19

Ray Optics and the Eikonal

19.1 FRAMING

- Real-world problems are mathematically harder than the idealized problems we encounter in our first textbooks. Often, we need some sort of unfair advantage before we can make a dent in a real-world problem. Often such an advantage comes in the form of a limiting case; for example, some quantity may be numerically small in cases of interest. In this chapter, we'll study the propagation of light in media that, while not uniform, at least vary on length scales much bigger than the wavelength of the light under study. That is, we'll study situations in which $\lambda \|\vec{\nabla} n\|$ is very small. This circumstance arises in many practical problems, and lets us approach otherwise forbiddingly complex situations.
- In everyday life, light seems to travel along “rays” that are generally straight lines—except when the light gets reflected or refracted. No concept of “rays” appears in the Maxwell equations, however. What, then, is a “ray?” This chapter will explore that question, which turns out to be connected to the preceding point.

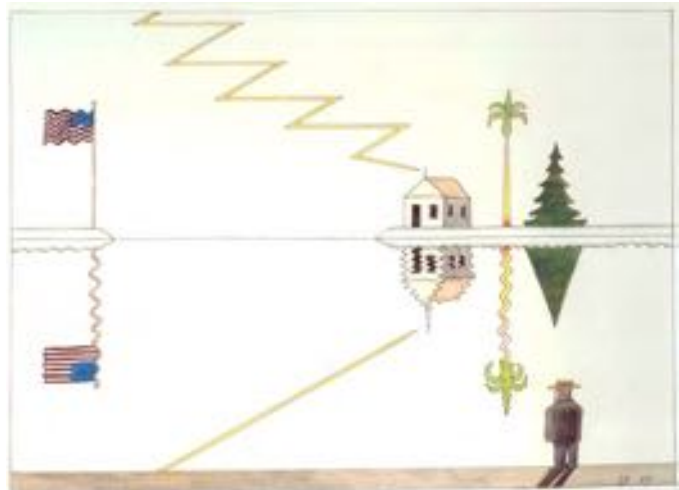
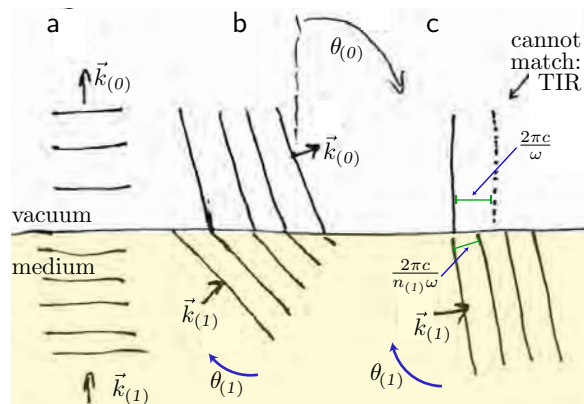


Figure 19.1: [Saul Steinberg.]

Figure 19.2: Refraction and total internal reflection. A plane wave passes through a planar interface. Lines represent the planes of constant phase for the incident- and transmitted-wave parts of the solution. (A reflected-wave part is also present but not shown.) *Left*, perpendicular incidence ($\theta_{(1)} = \theta_{(0)} = 0$). *Center*, the angle $\theta_{(1)}$ is nonzero but less than the critical value. *Right*, no solution is possible when the angle $\theta_{(1)}$ is too large. If the wave is coming from within the medium (traveling upward) as shown, then in this case it cannot escape and must be totally internally reflected.



19.2 UNIFORM MEDIUM

Consider a uniform, isotropic dielectric medium. Following Chapter 5, we will assume that the medium can be summarized simply by using an effective permittivity $\epsilon_{(1)} > \epsilon_0$ that is scalar in character.¹ Then the same analysis as in Chapter 17 shows that there will be transverse wave solutions with dispersion relation $\omega = (c/n_{(1)})\|\vec{k}\|$, where the **index of refraction** $n_{(1)} = \sqrt{\epsilon_{(1)}/\epsilon_0}$ is a constant larger than 1.

19.3 PIECEWISE-UNIFORM MEDIUM

19.3.1 Refraction law

Consider a sharp junction between an otherwise uniform dielectric medium 1 and vacuum. (Junctions between two media can be handled similarly.) We assume that the medium and its boundary are not changing in time (think about a chunk of glass). Then Maxwell’s equations are still linear partial differential equations with coefficients that are constant in time, so they will still have solutions with overall time dependence everywhere $\propto e^{-i\omega t}$.

The coefficients are not constant in *space*, however, due to the boundary, so we *can’t* expect solutions with a single overall $e^{i\vec{k}\cdot\vec{r}}$. Separately on each side, however, there are solutions of this form. So consider a trial solution with transverse plane waves on either side of the boundary, with wavevectors $\vec{k}_{(1)}$ and $\vec{k}_{(0)}$.

Figure 19.2a illustrates the situation when $\vec{k}_{(0)}$ is perpendicular to the interface. The horizontal lines represent the planes of constant phase, for example, where $\vec{E} = \vec{B} = 0$. These “wavefronts” are more widely spaced on the vacuum side because the two regions have the same frequency but different wave speed.

Figure 19.2b shows a more general situation. The vertical component of electric field may change discontinuously across the boundary, due to the possibility of bound charges there, but Faraday’s law shows that the horizontal components must be

¹Chapter 49 will justify this prescription in greater detail. We also assume that the magnetic polarizability is negligible, so $\mu = \mu_0$.

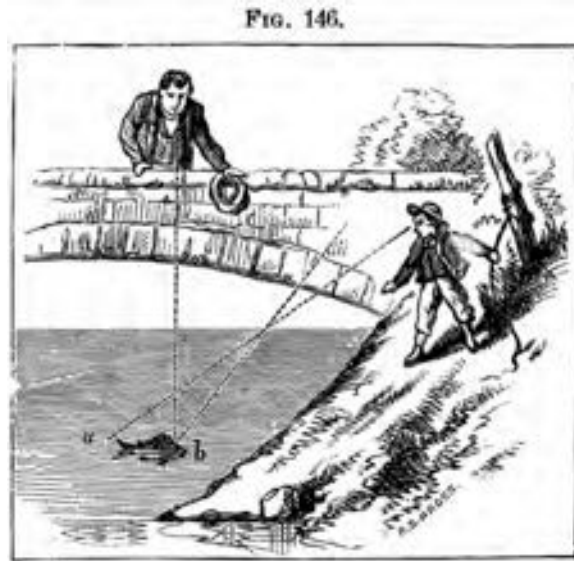


Figure 19.3: (a) The observer on the bridge gets an accurate impression of the location of the fish (head is at b). The observer on the bank, unconsciously assuming that light travels on straight lines, gets the inaccurate impression (the head seems to be at a). (b) Viewed from below, part of the air-water interface appears to be a mirror.

continuous. Then in particular the loci of zero horizontal \vec{E} field must match up across the boundary. The only way for this boundary condition to be consistent with different wavefront spacing is for the wavevector to *bend*, as shown.

Your Turn 19A

Because the frequency is the same on each side, we know from the dispersion relations that $\|\vec{k}_{(1)}\| = n\|\vec{k}_{(0)}\|$.

a. Convince yourself geometrically that the direction of the wavevector must change according to

$$\sin \theta_{(0)} = n \sin \theta_{(1)}. \quad (19.1)$$

b. How does this formula change for an interface between two dielectric media?

Figure 19.3a shows one familiar consequence of refraction.

19.3.2 Optical tweezers

Figure 19.4 shows how a spherical object with differing index of refraction from its surroundings will feel a net sideways force when it encounters a beam of light. This phenomenon is useful for manipulation of micrometer-scale objects (and of nanometer-scale objects that we may tether to them): The **optical tweezers** effect.

19.3.3 Spherical aberration

The law of refraction is also the basis for the focusing of light by a lens. Figure 19.5a shows a bundle of parallel light rays that impinge on a spherical dielectric object. If the

Figure 19.4: Generation of transverse force on a dielectric sphere by a beam of light, in the ray-optics regime. (a) The sphere is not centered in the beam. Two rays in the beam miss the sphere altogether. One ray is bent, undergoing a change in its momentum (a vector quantity). Newton's third law requires that the bead receives a continuous impulse toward the left. (b) The sphere is centered in the beam. The central ray is undeflected, and the ones flanking it make cancelling contributions to the net transverse impulse.

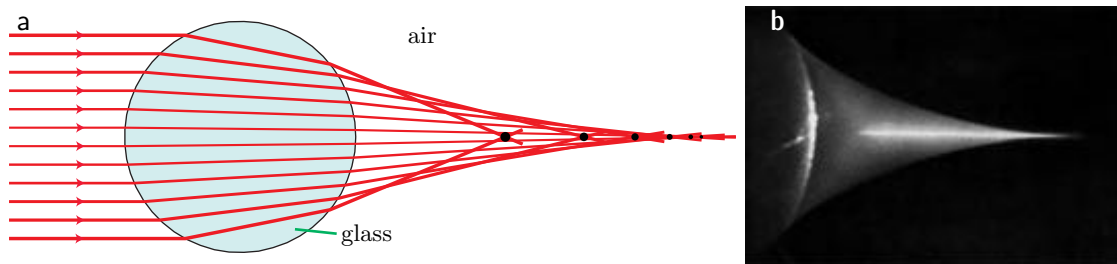
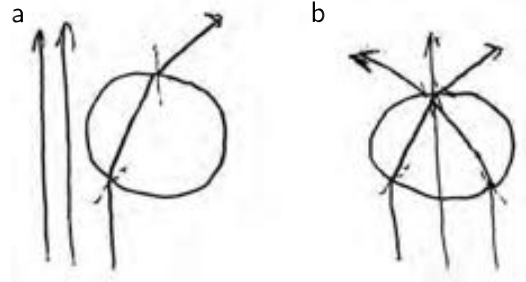


Figure 19.5: Spherical aberration. (a) [Ray diagram.] Parallel rays arriving at a lens with spherical surfaces, and passing close to its center (*thinner lines*), nearly coincide at a common focus (*smallest dot at far right*). However, rays initially farther from the axis (*heavier lines*) cross it in a spread-out array (*larger dots*). The rays shown were computed by using the law of refraction (Equation 19.1), for the case of a glass lens immersed in water. (b) [Photograph.] The spread-out focus is visible as the bright line in this photo. [(b) From Cagnet et al., 1962.]

object's diameter is much bigger than the wavelength of the light, then we may apply the law of refraction separately to each of the lines shown. The line passing through the center of the sphere is undeflected, but flanking rays are bent more and more, which tends to bring them to a common point, or **focus**. Unfortunately, however, the focusing is not perfect. The figure shows piecewise-straight lines that bend according to Equation 19.1, with index values appropriate for glass and water. The lines close to the center do arrive at a common point, but the ones farther out do not, a phenomenon called **spherical aberration** that limits the useful light-collecting region of microscope lenses.

19.3.4 Total internal reflection

Figure 19.2c shows geometrically that there may be no solution of the type described above, if the angle of incidence exceeds a critical value. In terms of your result from Your Turn 19A, we see that $\sin \theta_{(1)}$ must be smaller than $1/n$ because $\sin \theta_{(0)}$ cannot exceed 1. If a plane wave originates in the medium (directed toward the vacuum side) and this condition is violated, then there can be no transmitted plane wave. All incoming energy instead gets reflected back into the medium, a phenomenon called **total internal reflection** (TIR).

Your Turn 19B

- What if a plane wave originates on the vacuum side (\vec{k} directed toward the medium)?
- Imagine yourself submerged in a swimming pool. Looking straight upward, you see the sky. But beyond a certain angle, the surface above you looks like a mirror (Figure 19.3b). Why?

TIR is the basis for guiding light through fibers. As long as the fiber does not bend too sharply, an initially axially propagating wave will remain trapped inside it.² Such a fiber can carry vastly more data than a coaxial cable because the frequency of visible light is so much higher than the radio frequencies that the coax can carry. Also, a bundle of such fibers can carry each pixel of a complete input image faithfully to the same relative position at its output end, regardless of overall bends along the way. Such fiber-optic **endoscopes** are indispensable for noninvasive medical diagnosis.

TIR is also the basis for a form of fluorescence microscopy that achieves high signal-to-noise, called **TIRF microscopy**.

19.4 GRADIENT-INDEX MEDIUM**19.4.1 Streamlines**

We can now return to the framing questions (Section 19.1). First, however, think to a more concrete situation, a steady flow of water. At any point in a flow there is a local flux of mass, $\vec{j}_m(\vec{r})$. We can ask about the **streamlines** of this vector field. The streamlines are curves in space that are everywhere tangent to \vec{j}_m . No individual water molecule will literally follow a streamline, due to its random Brownian motion; nevertheless, the streamlines give a good impression of what is going on. A small but macroscopic tracer particle suspended in the water really will follow a streamline.

Similarly, in optics we can consider the streamlines of the *energy* flux \vec{j}_E . In a uniform, isotropic medium, we found plane wave solutions, for which \vec{k} is a constant.³ Because the energy flux is always parallel to \vec{k} , the streamlines of a plane wave are a family of straight, parallel lines. But this idea has wider usefulness than that one example.

19.4.2 Almost-plane waves

In a medium whose index changes, but only slowly compared to the wavelength of light, it seems reasonable to look for solutions to Maxwell's equations that *locally* resemble plane waves, but for which \vec{k}_{local} varies slowly over space. In a moment we will make that notion precise, and verify our expectation that such solutions exist. So

²This primitive description is appropriate for thick fibers. Modern fiber-optic lines are thin and function more like waveguides; their composition is also modulated across their cross-section; some even transmit light in the form of nonlinear traveling waves (solitons).

³For an anisotropic medium like calcite we must reconsider this statement.

it makes sense to define

Rays of light are the streamlines of the energy flux for a solution of this locally-plane wave form.

Just as a given chamber can have water flowing in various ways, so too a given optical system can have various locally-plane wave solutions, and hence various different families of rays passing through it.

With this terminology, in a piecewise-uniform medium, the law of refraction can be interpreted as saying that rays bend as they pass a boundary, a phenomenon that indeed corresponds to the behavior of a laser pointer’s beam when crossing from air to water (or the other way). So we will continue to use this viewpoint when we have a continuously varying index of refraction. For example:

- Radio waves that originally were sent away from Earth’s surface encounter the ionosphere, whose index of refraction can be smaller than 1. We’ll discuss the resulting refraction phenomenon in Section 19.4.6.
- The air close to a hot road surface has nonuniform temperature, and hence also density and hence also n , leading to mirage phenomena (see Section 19.5.1).
- Our own eye lenses have this property: Although they are transparent, the index varies continuously from a maximum at the center to a minimum at the surface (see Section 19.5.2).
- Perhaps most exotic, Einstein’s gravity theory predicts that even empty space will behave like an inhomogeneous medium if strong gravitational fields are present (Section 19.5.3).

We might expect some continuous version of the law of refraction to hold in situations like these. Let’s find it.

19.4.3 Eikonal equation

Solving vector PDEs without a lot of symmetry is in general difficult. But at least the situations just mentioned are all *steady*, that is, invariant under time translation, so we can again assume harmonic time dependence for our solutions. Moreover, all of the situations in the preceding list share a convenient aspect: The length scale L_0 over which the index varies is much greater than the wavelength of the light we wish to study, or in other words, $c/(L_0\omega) \ll 1$. In this regime, it’s reasonable to look for approximate solutions to Maxwell’s equations of **eikonal** form

$$\vec{A} = \frac{1}{2}e^{-i\omega t}\vec{\zeta}(\vec{r})e^{i\omega\beta(\vec{r})/c} + \text{c.c.} \quad (19.2)$$

In this expression, $\beta(\vec{r})$ is called the **eikonal function**, or simply “the eikonal.” For a plane wave it would be $\hat{k} \cdot \vec{r}$. The other unknown function, $\vec{\zeta}(\vec{r})$, allows for the possibility that the polarization is not constant throughout space. We assume that both β and $\vec{\zeta}$ vary slowly in space, with a length scale similar to L_0 .

We would like to see under what conditions the eikonal trial solution works, to leading order in the small parameter $c/(L_0\omega) \ll 1$. One way to describe this short-wavelength limit is to say that we are neglecting diffraction effects; this is the regime where we may hope that a “ray” concept will be useful. We will now develop

a formulation called **ray optics** that is useful for handling such situations, and that makes the notion of “ray” precise.

Close to any point \vec{r}_* , our trial solution Equation 19.2 thus resembles a plane wave with local wavevector $\vec{k}_{\text{local}} = \frac{\omega}{c} \vec{\nabla} \beta|_{\vec{r}_*}$. In particular, the energy flux $\vec{E} \times \vec{B}$ everywhere points along $\vec{\nabla} \beta$. So once we establish that a particular phase function $\beta(\vec{r})$ solves Maxwell’s equations, we can compute its gradient. The streamlines of the resulting vector field will be the rays that we seek.

First, we impose Coulomb gauge; then the rest of the Maxwell equations will take the simple form Equation 17.23. Thus, we require

$$0 = \frac{1}{2} e^{-i\omega t} (\vec{\nabla} \cdot \vec{\zeta} + \vec{\zeta} \cdot \frac{i\omega}{c} \vec{\nabla} \beta) e^{i\omega\beta/c} + \text{c.c.}$$

We may drop the first term, because the second dominates in the short-wavelength limit. Thus, not surprisingly, $0 = \vec{\zeta} \cdot \vec{k}_{\text{local}}$, just as we found for plane waves.

The wave equation now says

$$\frac{1}{2} e^{-i\omega t} \left(\vec{\nabla}_j \left((\vec{\nabla}_j \vec{\zeta}_i + \frac{i\omega}{c} (\vec{\nabla}_j \beta) \vec{\zeta}_i) e^{i\omega\beta/c} \right) \right) + \text{c.c.} = -\left(\frac{\omega}{c}\right)^2 \frac{1}{2} \vec{\zeta}_i e^{-i\omega t + i\omega\beta/c} + \text{c.c.}$$

Again drop the first term in parentheses on the left, because the other term dominates it.

$$\frac{i\omega}{c} (\vec{\nabla}_j \beta) \cdot (\vec{\nabla}_j \vec{\zeta}_i) + \frac{i\omega}{c} \vec{\zeta}_i \nabla^2 \beta + \left(\frac{i\omega}{c}\right)^2 \vec{\zeta}_i (\vec{\nabla} \beta)^2 = -\left(\frac{\omega}{c}\right)^2 \vec{\zeta}_i.$$

The last term on the left dominates the others, so we find that our trial solution works if

$$\|\vec{\nabla} \beta\|^2 = 1 \quad \text{eikonal equation in vacuum} \quad (19.3)$$

Some simple solutions to the eikonal equation include $\beta(\vec{r}) = \hat{k} \cdot \vec{r}$ (plane wave) or $= \|\vec{r}\|$ (spherical wave). In the former case, the rays are parallel straight lines; in the latter case, they are straight radial lines.

In principle, we’re now done with the vacuum case, but it may not be clear that we have made progress: We have approximated Maxwell’s equations, which are linear, with the new PDE Equation 19.3 which is *nonlinear*. But we do not always need all the information in the phase function β . Let’s now convert our equation into a direct characterization of the rays (streamlines of $\vec{\nabla} \beta$) themselves.⁴

19.4.4 Rays in vacuum

We are looking for a family of curves, each of which is everywhere tangent to \vec{k}_{local} . We can write a curve in parametric form as $\vec{\ell}(s)$, where s is arclength. That is, $d\vec{\ell}/ds$ is everywhere a unit vector. It’s everywhere parallel to $\vec{\nabla} \beta$, which itself is everywhere a unit vector, so

$$\frac{d\vec{\ell}}{ds} = \vec{\nabla} \beta|_{\vec{\ell}(s)} \quad \text{for all } s. \quad (19.4)$$

⁴Note that the polarization vector drops out of Equation 19.3, so we learn nothing about $\vec{\zeta}$ from this approach other than that it must everywhere be perpendicular to $\vec{\nabla} \beta$. To learn more, we would have to retain some of the subleading terms dropped earlier; instead we will concentrate on just the rays, and not their polarization behavior.

One way to characterize a curve is to state its **curvature**, that is, how its tangent vector deviates from being a constant. More precisely, we define the curvature vector as the derivative of the unit tangent to the curve with respect to arclength:

$$\frac{d^2\vec{\ell}}{ds^2} = \frac{d}{ds} \left(\vec{\nabla}\beta|_{\vec{\ell}(s)} \right).$$

The right-hand side of this formula is the derivative of a function as we walk along the curve. To evaluate that quantity, we can find the dot product of the gradient (that is, all partial derivatives) with the unit tangent:

$$\begin{aligned} \frac{d^2\vec{\ell}_i}{ds^2} &= \left(\frac{d\vec{\ell}}{ds} \cdot \vec{\nabla} \right) \vec{\nabla}_i\beta|_{\vec{\ell}} = (\vec{\nabla}_j\beta)(\vec{\nabla}_j(\vec{\nabla}_i\beta)) = (\vec{\nabla}_j\beta)(\vec{\nabla}_i(\vec{\nabla}_j\beta)) \\ &= \frac{1}{2}\vec{\nabla}_i\|\vec{\nabla}\beta\|^2 = 0. \end{aligned}$$

Note that the phase function β has disappeared from this expression; we don't need to solve the eikonal equation after all in order to find the rays. Instead, we conclude that *the curvature is zero*:

$$\textit{Light rays in vacuum are straight lines.} \quad (19.5)$$

That makes sense: Ray optics neglects diffraction, and when that approximation holds indeed objects cast sharp shadows. The two explicit families of solutions found earlier (straight parallel rays and straight radial rays) both obey this rule.

19.4.5 Rays in an inhomogeneous medium

We now consider the case in which the local speed of light, $c/n(\vec{r})$, is not constant in space. (The symbol c always refers to the speed of light in vacuum.)

Your Turn 19C

a. Show that generalizing our previous derivation (Equation 19.3) gives

$$\|\vec{\nabla}\beta\|^2 = n^2. \quad \text{eikonal equation in medium} \quad (19.6)$$

b. Show that therefore the analog to Equation 19.4 gives the tangent to a ray as

$$\frac{d\vec{\ell}}{ds} = \frac{\vec{\nabla}\beta}{n} \Big|_{\vec{\ell}(s)}. \quad (19.7)$$

Your Turn 19D

a. Next show

$$\frac{d}{ds} \left(n(\vec{\ell}) \frac{d\vec{\ell}}{ds} \right) = \vec{\nabla} n \Big|_{\vec{\ell}(s)} \quad \text{ray equation} \quad (19.8)$$

at every value of s .b. Check that your result from (a) is compatible with arclength parameterization. That is, show that $\|d\vec{\ell}/ds\|$ remains equal to one if it starts that way.

Once again, the ray equation makes no explicit mention of the eikonal function β . It tells us how light rays bend as they pass through a medium—a generalization of the law of refraction. When the ray-optics approximation is justified, this equation reduces Maxwell’s partial differential equations to an *ordinary* vector differential equation, a net simplification.

The ray equation is nonlinear, but it is simpler than the eikonal equation in a crucial respect: It is an *ordinary* differential equation, not a PDE. Much like Newton’s $\vec{f} = m\ddot{\vec{r}}$, we can start a ray trajectory at any point, with any initial direction of motion, and then step through the ray equation to find the subsequent path of that ray. Solving systems of ODEs numerically is a routine task.

19.4.6 An example

Suppose that $n(x)$ depends only on one variable, the “altitude.” This could be the case when radio waves travel upward and encounter the Earth’s ionosphere (over distances short enough to neglect Earth’s curvature).

Initially a ray makes an angle θ_0 with respect to the vertical. Further along on the ray, $\theta = \cos^{-1}(\hat{z} \cdot d\vec{\ell}/ds)$ may change. If at any point this angle increases to $\pi/2$, then the ray bounces (or “skips”) back downward.

Taking the dot product of Equation 19.8 with \hat{z} gives

$$\frac{d}{ds} (n \cos \theta) = \frac{dn}{dz} \Big|_{\vec{\ell}(s)}.$$

Multiply both sides by n :

$$n \frac{d}{ds} (n \cos \theta) = \frac{1}{2} \frac{dn^2}{dz}.$$

Next, note that when we move by arclength ds , altitude changes by $dz = ds \cos \theta$, so

$$n \cos \theta \frac{d}{dz} (n \cos \theta) = \frac{1}{2} \frac{d}{dz} (n^2 \cos^2 \theta) = \frac{1}{2} \frac{d}{dz} (n^2).$$

Thus, $n^2 \cos^2 \theta - n^2$ is a constant along the ray, a generalized law of refraction:

$$n \sin \theta = \text{constant.} \quad \text{if } n \text{ depends only on } z \quad (19.9)$$

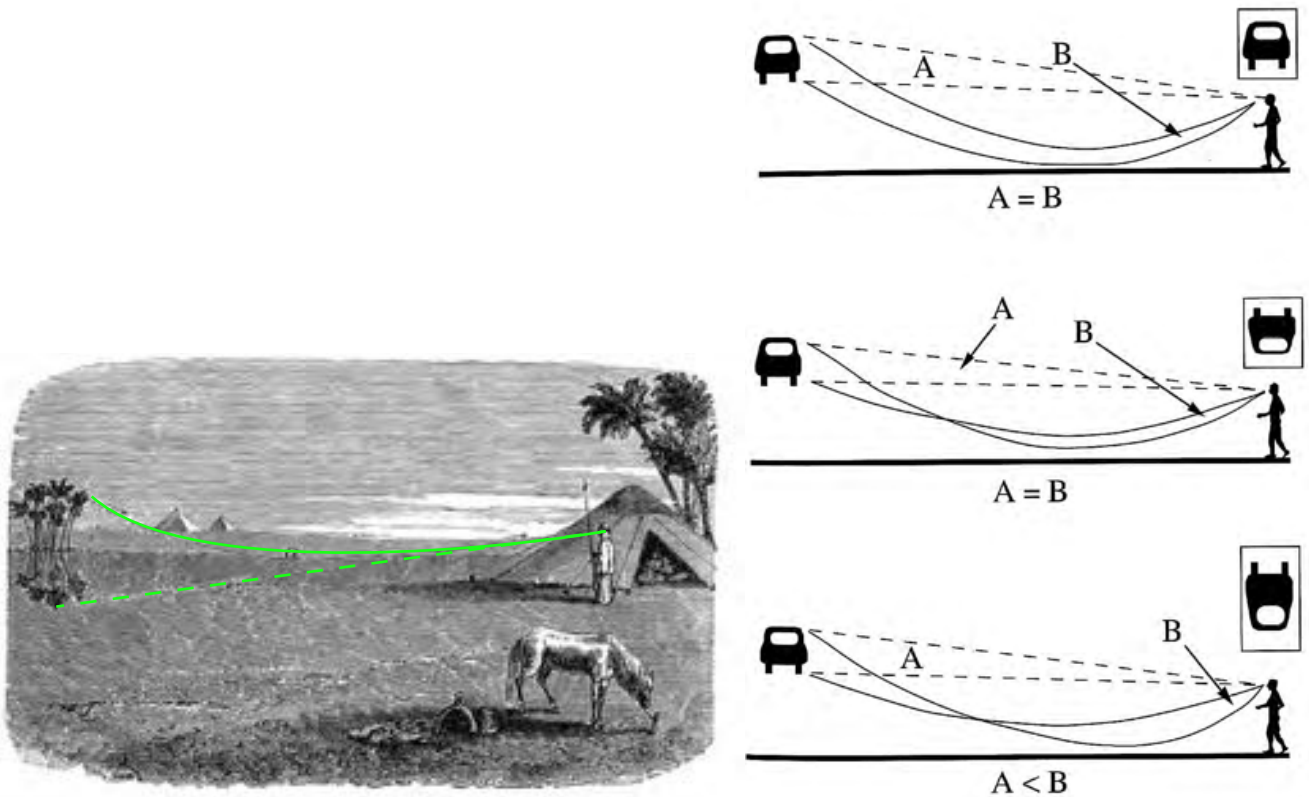


Figure 19.6: (a) Formation of a mirage. (b) Depending on atmospheric index profile, a mirage can appear upright, inverted, and/or stretched. [Lynch & Livingston, 2001, Fig 2.24C p57.]

In the special case where n changes suddenly at a planar boundary, this result reduces to the usual law of refraction. More generally, given a profile for n it tells us which initial angles, if any, will give rays that bounce back down to Earth.⁵

19.5 MORE APPLICATIONS

Here are several situations in which light travels through a medium whose index varies slowly, that is, whose length scale of variation is large compared to the light's wavelength.

19.5.1 Mirage

The formation of mirages is another example of this situation (Figure 19.6). On a long, flat stretch of highway, solar heating creates a layer of air near $z = 0$ that is hotter than elsewhere. That air is less dense than the cooler upper layers. Thus, it can happen that, when we direct our gaze downward (toward the road) we'll see

⁵In Problem 19.3/19.4, you'll show how a similar framework can describe the phenomenon of mirages.

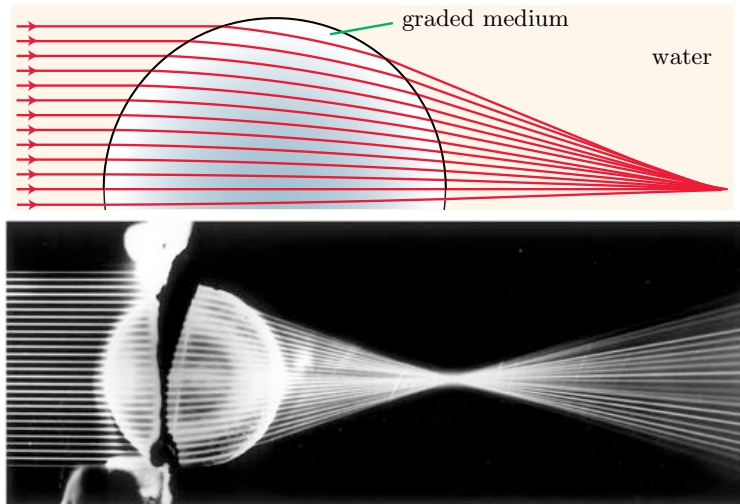


Figure 19.7: [Ray diagram.] **Correction for spherical aberration by a continuously graded index of refraction.** (a) A set of parallel incoming rays is shown, with their computed trajectories upon entering the medium. In this case, the rays curve inside the lens, because its index of refraction is greater in the center than at the periphery. The extra bending has been arranged to make all the rays nearly meet at a common focus. (Problem 19.5 describes the index function that was used to make this diagram.) (b) Actual light rays traversing the eyelens of an octopus. [From Jagger & Sands, 1999.]

light originating from the sky that has traveled on the curved path in the figure. It is easy to misinterpret that light as a reflection from (nonexistent) water on the road, particularly because it tends to shimmer (due to air convection currents). You probably know from experience that this illusion only appears in the distance, not up close. You'll work out this and other details in Problems 19.3–19.4.

19.5.2 A spherical lens with minimal spherical aberration

see Figure 19.7.

19.5.3 Gravitational lensing

Einstein's theory of gravitation postulates that space and time can deviate from the cartesian (flat) geometry assumed throughout this course, and that this deviation is responsible for the familiar effects of gravitation. Moreover, because light (and everything else) inhabits spacetime, it, too will be affected by gravitational fields. Of special interest is the fate of a ray that travels through empty space far from any mass, then passes close to a massive object, and finally emerges back into empty space. This ray will travel on straight lines before and after the flyby, but those two lines may not be parallel, because of the transit through a non-cartesian region during the encounter.

Einstein realized that, although the mathematics of curved spacetime gets complicated, his final expression for the bending of a light ray was mathematically identical to that of a ray passing through ordinary spacetime with a *refracting medium* having

effective index given by

$$n_{\text{eff}} \approx 1 - 2\phi_{\text{N}}/c^2 + \dots \quad (19.10)$$

Here ϕ_{N} is the newtonian gravitational potential far from the mass, and the ellipsis represents terms of higher order in ϕ_{N}/c^2 .

Your Turn 19E

- Equation 19.10 may be unfamiliar to you, so check that the units make sense.
- In the neighborhood of a point mass, the formula becomes $n_{\text{eff}} \approx 1 + r_*/r$. Look up the mass of our Sun and find a formula for r_* in terms of M/M_{sun} .

19.6 PLUS ULTRA

Erwin Schrödinger was well trained in optics and acoustics. He reasoned that:

- Einstein and deBroglie say that particles correspond to waves.
- Bohr says that in the atomic world, where the length scale is comparable to the deBroglie wavelength, the wave idea explains the observed quantization of energy, analogously to the quantization of harmonics in an organ pipe.
- It is true that newtonian mechanics seems to rule the macro world.
- But this sounds familiar: Maybe we need to seek a wave equation (not the usual one, but some equation with wavy solutions) whose geometric-optics limit gives trajectories that solve Newton's laws (not the law of refraction).

It was already known that, remarkably, newtonian mechanics could be formulated in a way resembling the rays associated to our eikonal equation. So Schrödinger did not have to look far. His crazy idea needed some interpretation, to be sure. But it worked out OK. In fact, this could be the biggest successful lateral-thinking jump in scientific history.

FURTHER READING

Total internal reflection: Nelson, 2017.

Eikonal approximation: Landau & Lifshitz, 1979, chapt. 7; Elmore & Heald, 1969, §9.2.

Ray equation: Elmore & Heald, 1969, §9.2.

Mirage: Richey et al., 2006.

Gravitational lensing: Basic: Schutz, 1990, chapt. 11 Advanced: Straumann, 2013, chapt. 5, Nye, 1999. Optical tweezers: Perkins, 2014;

<https://www.cell.com/biophysj/collections/optical-tweezers>; Smith, 1999; Bechhoefer & Wilson, 2002; van Mameren et al., 2011; Lang, MJ; Block, SM. 2003. Resource letter: LBOT-1: Laser-based optical tweezers. American Journal of Physics 71 (3): 201-215; Jones et al., 2015; Appleyard et al., 2007.

Bustamante et al., 2000; Mehta et al., 1999; Svoboda & Block, 1994; Sheetz, 1998;

A. Ashkin, "Optical trapping and manipulation of neutral particles using lasers,"

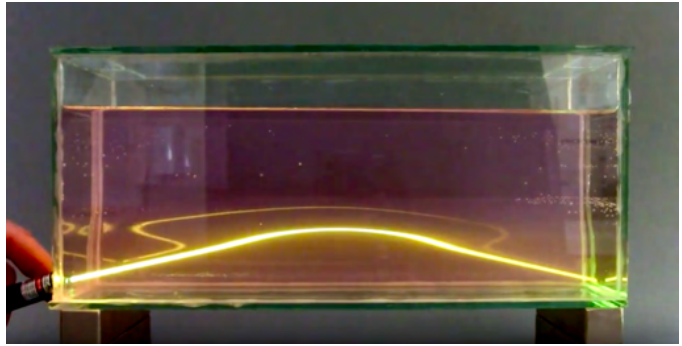


Figure 19.8:

PNAS 94 (1997) 4853–60.

<https://www.cell.com/biophysj/collections/optical-tweezers>;

<https://www.nobelprize.org/uploads/2018/10/advanced-physicsprize2018.pdf>;

Novaya Zemlya Effect: Corliss, 1984, p151

PROBLEMS

19.1 *Waves in conductive medium*

An electromagnetic plane wave propagates through vacuum, then enters a medium. The medium is not polarizable ($\epsilon = \epsilon_0$, $\mu = \mu_0$). However, it is electrically conductive, obeying an ohmic relation with conductivity κ :

$$\vec{j} = \kappa \vec{E}.$$

Assume the medium is everywhere electrically neutral.

- Find the dispersion relation for plane waves of angular frequency ω traveling through such a medium, and interpret it physically.
- The wave is initially traveling along a direction perpendicular to the planar surface of the medium, which extends to infinity beyond that surface. Find a solution to Maxwell's equations that accounts for the free current density set up in the medium, and that includes the incoming wave, a transmitted wave, and possibly a reflected wave as well.

19.2 *Poor wandering one*

Figure 19.8 shows light shone from a laser pointer into a tank of—mostly—water. The surface of the water is near the top of the tank. What do you think might cause the light to take this bizarre, wandering path?

19.3 *Mirage 1*

Section 19.4.2 mentioned the problem of light passing through a layer of air that is heated at the bottom, leading to a temperature gradient, hence a density gradient, hence a gradient in the index of refraction. This is the special case of a “gradient-index” material whose index depends only on height z .

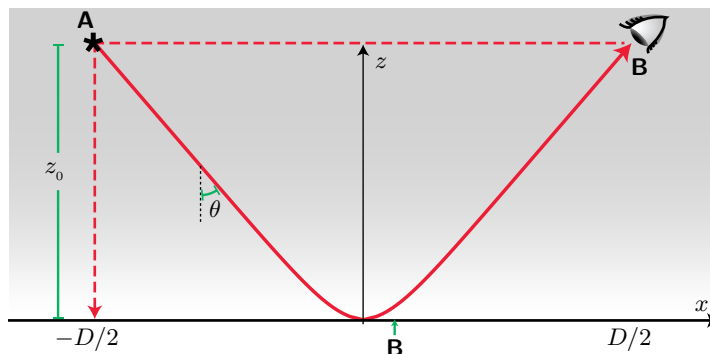


Figure 19.9: [Mathematical function.] **Curved stationary-phase path in an inhomogeneous medium.** The x and z axes are not drawn to the same scale. *Dashed lines* show the simple solutions mentioned in Problem 19.4. An observer who assumes straight-line propagation will interpret light from **A** as having come from **B**.

Section 19.4.6 worked out a general formula for the angle θ that a ray makes with the z -axis (Equation 19.9).

Suppose that the index has the functional form

$$n(z) = n_0(1 - ae^{-z/L})$$

where $a \ll 1$ and $n_0 > (1 - a)^{-1}$; hence $n(z) > 1$ for all positive values of the height z . We are interested in incident rays that are almost horizontal (grazing incidence, $\theta_0 = (\pi/2) + \epsilon$ where ϵ is small and positive).

We want to know whether the ray will hit the ground ($z = 0$) or will on the contrary bounce back. That is, we want to know if θ decreases to $\pi/2$ before z crosses zero. Show that this will happen if ϵ is small enough, and find out how small is enough.

19.4 *Mirage 2*

Section 19.5.1 mentioned the problem of light passing through a layer of air that is heated at the bottom, leading to a temperature gradient, hence a density gradient, hence a gradient in the index of refraction. This is the special case of a “gradient-index” material whose index depends only on height z .

Section 19.4.6 worked out a general formula for the angle θ that a ray’s trajectory makes with the z -axis (Equation 19.9). This condition has two unsurprising solutions: One is a straight, horizontal line: $z = z_0$, $\theta(x) = \pi/2$. The other is a straight, vertical line: $\theta(x) = 0$. But there can also be solutions that are *curved*.

Suppose that the density profile $n(z)$ is strictly increasing as z increases, and that θ starts out tilted downward ($0 < \theta < \pi/2$). Then θ can increase as z decreases, potentially even leveling off ($\theta \rightarrow \pi/2$), as shown in Figure 19.9.

Suppose that light is emitted by a source at height z_0 , and detected somewhere else, also at height z_0 but a distance D away. We can characterize a curve in the xz plane by its height function, $z = h(x)$, where $h(\pm D/2) = z_0$. We wish to find functions $h(x)$ that give solutions to Equation 19.9 subject to these boundary conditions.

- To be specific, suppose that $n(z) = n_\infty(1 - \alpha e^{-z/L})$, where n_∞ is the index of air at 30°C, $n_\infty(1 - \alpha)$ is the index of air at 50°C, $L = 20$ cm, and your eyes are $z_0 = 2$ m off the ground. Look up the values of the two indices of refraction for

visible light in Appendix C. Use Equation 19.9 to see how close θ_0 must be to $\pi/2$ in order for the ray's trajectory to level off before hitting the ground. Then estimate how far away the mirage will appear to be.

- b. Reformulate Equation 19.9 as a differential equation determining the entire curve; that is, an equation involving dh/dx . Solve it analytically or numerically for the situation discussed above. If any simplifying approximations are valid, go ahead and use them. Use the smallest value of θ_0 for which you found that a mirage would be possible, and use a computer to make a graph showing your solution. (Use different scales for the x and z axes, to show the shape of your solution clearly.)

19.5 Gradient-index lens

Use the ray-optics approximation for this problem. If you haven't done Problems 19.3 and 19.4 yet, do them first. Those problems asked you to find light ray trajectories in a nonuniform medium whose index of refraction depends on only one Cartesian coordinate, the height. In the present problem, you'll generalize your results to a nonuniform medium (a "gradient-index lens") whose index of refraction depends only on *radius*, that is, the distance r to the center of the lens. Section 19.5.2 mentioned that this situation holds for the eye lenses of animals, and claimed that such nonuniformity can eliminate much of the aberration created by a uniform spherical lens (compare Figure 19.5a to Figure 19.7).

In this problem, you can scale all lengths by the radius a of the sphere, that is, work in terms of $\bar{r} = r/a$ and so on. Let $n_c = n(0)$ be the index at the center, $n_p = n(1)$ its value at the periphery, and $K = n_p/n_c - 1$. Fish eyes have $n_c \approx 1.52$, $n_p \approx 1.38$, and

$$n(\bar{r}) \approx n_c(1 + K(0.82\bar{r}^2 + 0.30\bar{r}^6 - 0.12\bar{r}^8)),$$

and are immersed in media with $n_w \approx 1.33$ on both sides. It will be convenient to define $g(\bar{r}) = n^{-1}(dn/d\bar{r})$.

- a. Choose coordinates centered on the lens center, and a plane passing through that origin, say the xy plane. Write out both components of the ray equation (Equation 19.8, page 234), which determines the stationary-phase paths $\vec{\ell}(s)$. It's a pair of coupled, second-order ordinary differential equations in the two Cartesian coordinates of a path lying in the chosen plane, $\ell_x(s)$ and $\ell_y(s)$. Parameterize the curve by arclength s , that is

$$ds = \|d\vec{\ell}/d\xi\| d\xi. \quad (19.11)$$

- b. Now generate a picture similar to Figure 19.7, by constructing a series of solutions to the ray equation. Each ray initially starts outside the lens, traveling parallel to the x axis. Find the x and y values at which the incoming ray enters the lens, and the angle it makes relative to the perpendicular (the angle of incidence).
- c. Use the law of refraction to find the tangent vector to the ray just *after* it enters the lens.
- d. Use your results in (b,c) to get the required four initial conditions for the ray equation, then use a computer to solve it numerically.
- e. Follow your solution to find the value \bar{s}_{exit} at which \bar{r} once again reaches the value 1.



Figure 19.10: The graceful arcs at the center of this image from the Hubble Space Telescope are actually the distorted light of distant galaxies, twisted to form an Einstein ring by the gravitational influence of the closer galaxy cluster SDSS J0146-0929. Credit: ESA/Hubble and NASA; Acknowledgment: Judy Schmidt [<https://www.nasa.gov/image-feature/goddard/2018/hubble-finds-an-ei>]

- f. The tangent vector $d\vec{\ell}/ds|_{s_{\text{exit}}}$ then tells you the angle of incidence as the ray crosses the lens→water interface. Use the law of refraction again to find its angle after it leaves the lens.
- g. After leaving the lens, the ray is once again straight. Find the point where it hits the x axis, then have your computer draw all three segments (straight, curved, straight). Repeat for each ray that you wish to trace.

19.6 Gravitational lens

Use the ray-optics approximation for this problem. Section 19.4.6 considered light ray trajectories in a nonuniform medium whose index of refraction depends on only one cartesian coordinate, the height. In the present problem, you'll generalize your results to a nonuniform “medium” (a static gravitational field) whose “index of refraction” depends only on *radius*, that is, the distance r to the location of a point mass (Equation 19.10).

Choose coordinates centered on the lens center, and a plane passing through that origin, say the xy plane.

- a. Write out both components of the ray equation (Equation 19.8, page 234), which determines the stationary-phase paths $\vec{\ell}(s)$. It's a pair of coupled, second-order ordinary differential equations in the two cartesian coordinates $\vec{\ell}_x(s)$ and $\vec{\ell}_y(s)$ of a curve (ray) lying in the chosen plane parameterized by arclength s .

It's convenient to scale all lengths by the radius r_* that you found in Your Turn 19E, that is, to work in terms of $\bar{s} = s/r_*$ and so on. It will also be convenient to define $g(\bar{r}) = n^{-1}(dn/d\bar{r})$.

- b. Show that $g(\bar{r}) = -1/(\bar{r}^2(1 + 1/\bar{r}))$.
- c. Consider a series of rays that each start at $\bar{x}_0 = -10$, traveling parallel to the x axis at various y values. The initial position and direction of each ray amounts to the

four initial conditions needed in order to solve the ray equation. Use a computer to solve it numerically for several values of \bar{y}_0 . Because Equation 19.10 is only valid for weak gravitational fields, only examine values of \bar{y}_0 that are greater than (say) 5.

- d. Now generate a picture analogous to Figure 19.7, by having your computer draw your solutions.
- e. Your trajectories are distinguished by their y_0 values. For each, find the value x_* at which the trajectory hits the symmetry axis $y = 0$ and graph x_* as a function of y_0 .
- f. Your trajectories become straight lines far from the point mass, and in particular when they hit the symmetry axis. So you can find the angle of approach θ_* at that intersection from your numerical result in (c). This gives an apparent angular location in the sky. By the problem's axial symmetry, the background star appears as a ring with this angular radius: the **Einstein ring** (Figure 19.10). Make a graph of θ_* as a function of y_0 .
- g. Finally, combine your two previous results to graph θ_* as a function of x_* , that is, apparent angular width of the Einstein ring as a function of rescaled distance from observer to the lensing object, for a background star at infinity.

[Note: There is a more elegant way to handle trajectories in a spherically-symmetric field. However, the method advocated in this problem remains useful in an arbitrary gravitational potential, not just the field near a point mass.]

CHAPTER 22

Partial Polarization

22.1 LIGHT IS USUALLY CHAOTIC

See also Zangwill §16.4; Landau and Lifshitz 4th ed §50.

We found plane-wave solutions to Maxwell's equations. Each such solution had a single, definite wavevector \vec{k} , and hence a definite frequency: They described **monochromatic light**, such as might be obtained from a laser. Each also had a single, definite polarization vector. Monochromatic solutions have no start nor end; they have infinite extent in space and time. For all of these reasons, they are caricatures of light from real sources. For example, even a laser got switched on at some finite time in the past, so it has finite duration and hence some spread in frequency. And even if we send light through a colored filter, that filter has some finite spread in its transmission function. Moreover, natural light is usually unpolarized (like sunlight), or partially polarized (like the blue sky). In these notes we'll see how to characterize partial polarization more precisely.

A single atom, making a transition between definite states, gives off a pulse of light of finite duration. Even if that pulse has a definite polarization vector, the superposed light from zillions of independent atoms (for example, in the Sun) will be a jumble of many polarizations. We will model the light in narrow frequency range, traveling in one direction \hat{z} , as an incoherent superposition. Evaluating at just one point of space, such a superposition looks like

$$\vec{E}(t) = \frac{1}{2} \vec{\bar{E}}(t) e^{-i\omega t} + \text{c.c.} \quad (22.1)$$

In this expression, $\vec{\bar{E}}(t)$ is the sum of the profiles of a bunch of pulses. It varies more slowly than the mean frequency ω . Each pulse may have a complex phase relative to the others, and each may be polarized in a different way (but always perpendicular to \hat{k}).

In practice, optical instruments in millimeter wavelength and shorter don't measure the detailed time dependence of the electric field.¹ They just measure the *energy flux* delivered by a light source, averaged over a time that's long compared to the frequency, and that includes many of the contributions in Equation 22.1. We may place various filters between the source and detector, to restrict to various polarization or frequency ranges, but ultimately what's measured are energy fluxes of the filtered lights. Moreover, in most optics applications the filters we might use always perform *linear* operations. For example, an ideal color filter multiplies $\vec{\bar{E}}(t)$ by a scalar fraction

¹Radiotelescopes do measure just this, so they pick up more detailed information about the waves they detect than optical instruments like bolometers or photoelectric cells.

that depends on ω ; a polarizer multiplies it by a matrix that doesn't depend (much) on frequency, and so on.

The preceding logic implies that, in optics, anything we can really measure can be extracted from the eight time-averaged quantities²

$$\langle \vec{E}_i e^{-i\omega t} \vec{E}_j e^{-i\omega t} \rangle \quad \text{and} \quad \langle \vec{E}_i e^{-i\omega t} \vec{E}_j^* e^{+i\omega t} \rangle \quad \text{where} \quad i, j = 1, 2$$

(and the complex conjugates of these quantities). Of these eight quantities, the first four average to essentially zero because of their fast time variation.

The remaining four constitute a 2×2 hermitian matrix:

$$\vec{J}_{ij} = \langle \vec{E}_i \vec{E}_j^* \rangle.$$

The most general such matrix can be written in terms of four real quantities; a traditional choice involves the four **Stokes parameters**:

$$\vec{J} = \frac{1}{2} \begin{bmatrix} s_0 + s_1 & s_2 - is_3 \\ s_2 + is_3 & s_0 - s_1 \end{bmatrix}.$$

The Stokes parameters describe light for the purposes of detectors such as those used in optics experiments.³ Note that

$$\det \vec{J} = (s_0^2 - s_1^2 - s_2^2 - s_3^2)/4. \quad (22.2)$$

22.2 FULLY POLARIZED CASE

Note that the average of a product is not in general the same as the product of the corresponding averages. So although \vec{J} is the average of a dyad product, still it need not itself be expressible as such a dyad. If, however, the light in question is truly monochromatic, then \vec{E} is a constant in time, we may drop the averages, and so we do have a dyad.

For such a wave traveling along \hat{z} , you should substitute the general polarization vector $\vec{E} = A\hat{x} + Be^{i\delta}\hat{y}$ into the definition of \vec{J} and see how the Stokes parameters look in terms of A , B , and δ .⁴ Quite generally, the determinant of a dyad product equals zero; confirm that your answer has that property. Thus, s_1 , s_2 , and s_3 always sit on a sphere of radius s_0 (see Equation 22.2). Comment on what parts of this **Poincaré sphere** correspond to linearly polarized light, and what parts to circular polarization.

Warning: Although we speak of the Stokes parameters s_1 , s_2 , and s_3 as lying on a 3-sphere, they do not constitute a “vector” in the sense of pointing somewhere in real space. That is, they do not define a rank-1 tensor.⁵ The Poincaré sphere is an abstract, but sometimes useful, representation of \vec{J} , a complex, rank-2, 2D tensor.

²The magnetic field of a plane wave just tracks the electric field, so we would learn nothing new by considering terms with \vec{B} .

³Landau and Lifshitz factor out the overall normalization and define Stokes parameters as $\xi_1 = s_2/s_0$, $\xi_2 = s_3/s_0$, $\xi_3 = s_1/s_0$.

⁴The tensor \vec{J} discards any overall phase, so we don't need to give A and B separate phases.

⁵Nor do the full set of four Stokes parameters constitute a 4-vector!

22.3 UNPOLARIZED CASE

The opposite extreme situation is called **unpolarized light**. In this case, \vec{E}_i wanders randomly in spatial direction (though always perpendicular to \hat{z}) and also in complex phase. The matrix \vec{J} must then be isotropic (rotationally invariant in the xy plane), and hence a constant times the 2×2 identity matrix. Thus, $s_1 = s_2 = s_3 = 0$; unpolarized light sits at the *center* of the Poincaré sphere.

It's worthwhile to confirm this conclusion more explicitly. Suppose that $\vec{E}(t)$ consists of a series of N pulses, each linearly polarized in a direction that's uniformly distributed over the circle perpendicular to \hat{z} , and each with a random overall phase, again uniformly distributed. For simplicity, assume that each pulse has the same amplitude A . Then

$$\vec{J}_{11} = \langle A \cos \theta(t) e^{i\phi(t)} A \cos \theta(t) e^{-i\phi(t)} \rangle = \frac{1}{2} A^2, \quad (22.3)$$

$$\vec{J}_{12} = \langle A \cos \theta(t) e^{i\phi(t)} A \sin \theta(t) e^{-i\phi(t)} \rangle = 0, \quad (22.4)$$

and so on. Thus,

$$\vec{J} = \frac{A^2}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (22.5)$$

as expected.

22.4 PARTIAL POLARIZATION

These limits motivate us to define the **degree of polarization** as $(s_1^2 + s_2^2 + s_3^2)/s_0^2$. It ranges from zero (unpolarized) to one (fully polarized).

22.5 HOW TO MEASURE THE STOKES PARAMETERS

It's straightforward to measure s_0 , because it's a constant times the total intensity (energy flux) of the light.

To see how to measure the others (and indeed, why they are needed), let's first think about the sorts of filters that we could apply to a light source. Idealize a polarizing filter as performing a *linear projection* on the electric field, that is, the linear operation $\vec{E} \rightarrow \hat{\zeta}(\hat{\zeta}^* \cdot \vec{E})$. Then the corresponding transformation on the polarization tensor \vec{J} is

$$\vec{J} \rightarrow \langle \hat{\zeta}(\hat{\zeta}^* \cdot \vec{E})(\vec{E}^* \cdot \hat{\zeta}) \hat{\zeta}^* \rangle = (\hat{\zeta} \hat{\zeta}^*) \cdot \vec{J} \cdot (\hat{\zeta} \hat{\zeta}^*).$$

Consider the case of a linear polarizer, that is, $\hat{\zeta} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, acting on unpolarized light. Interpret the new polarization tensor. Repeat for a circular polarizer.

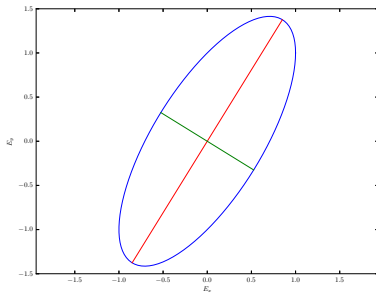
Think about how applying such an operation to an *arbitrary* \vec{J} , then finding the intensity of the resulting light, lets us deduce the various matrix elements of \vec{J} , and hence the Stokes parameters.

PROBLEMS

22.1 Stokes

For a given direction of wave propagation, I defined a 2D tensor $\vec{J}_{ij} = \langle \vec{E}_i \vec{E}_j^* \rangle$ to describe the polarization state of a superposition of plane waves. A special case is a pure (fully-polarized) plane wave, $\vec{E} = \frac{1}{2} \vec{E} e^{-i\omega(t-z/c)} + \text{c.c.}$ I then proposed to repackage the information in \vec{J} as four real quantities s_α .

- a. Suppose that we have fully polarized light traveling along the z axis, with $s_\alpha = 3, -1, 2, -2$ respectively (times an overall constant). Find a formula for $\vec{E}(t)$ at the origin of coordinates $\vec{r} = \vec{0}$. Confirm that the tip of the electric field vector sweeps out an ellipse in the xy plane, and describe that ellipse. That is, give its semimajor and semiminor axes, and the angle that the semimajor axis makes with the x axis.
- b. Repeat with $s_\alpha = 25, 0, 24, 7$.



CHAPTER 23

First Look at Radiation

The economy of science requires of us that we should avoid roundabout ways when a straight path is possible. If with the aid of our electric waves we can directly exhibit the phenomena of light, we shall need no theory as interpreter; the experiments themselves will clearly demonstrate the relationship between the two things. As a matter of fact such experiments can be performed.

— *Heinrich Hertz*

23.1 THE WAVE EQUATION

We formulated the Maxwell equations in terms of potentials, then specialized to the situation where the vector potential satisfied $\vec{\nabla} \cdot \vec{A} = 0$ (Coulomb gauge). In cartesian coordinates we found

$$\nabla^2 \psi = -\rho_q / \epsilon_0 \quad (23.1)$$

$$\nabla^2 \vec{A} - c^{-2} \left(\frac{\partial^2}{\partial t^2} \vec{A} + \vec{\nabla} \frac{\partial}{\partial t} \psi \right) = -\mu_0 \vec{j}. \quad (23.2)$$

To keep things simple, we will for now assume that the charge density is everywhere zero. In Your Turn 17F you showed that in this case, we may take $\psi = 0$.

However, there can be regions in space where the charge flux $\vec{j} \neq 0$. The continuity equation requires that $\vec{\nabla} \cdot \vec{j} = 0$, but this can be satisfied, for example, by having current in a wire that is uniform all along the wire's length. Equation 23.2 reduces to three decoupled copies of the **inhomogeneous wave equation**,

$$\nabla^2 \vec{A} - c^{-2} \frac{\partial^2}{\partial t^2} \vec{A} = -\mu_0 \vec{j}. \quad \text{Coulomb gauge, no net charge} \quad (23.3)$$

In empty space, we found some simple solutions to this equation: the plane waves. But of course empty space can also contain *no* radiation (fields everywhere zero). We'd like to see how, in the presence of accelerating charges, waves are *obligatory*.

23.2 RECALL MAGNETOSTATICS

We already encountered the special case of Equation 23.3 in which the charge flux \vec{j} is time independent. In that case, we had three independent (decoupled) copies of the

Poisson equation, each of which had the same solution as in electrostatics:¹

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \int d^3r_* \frac{\vec{j}(\vec{r}_*)}{\|\vec{r} - \vec{r}_*\|}. \quad \text{static case} \quad (23.4)$$

We called this expression the Green function solution to the Poisson equation. As usual, we call \vec{r} the “field point” and \vec{r}_* the “source point.” Also define $\vec{R} = \vec{r} - \vec{r}_*$, and as usual denote its length by R (no arrow). Then the function $G(\vec{r}, \vec{r}_*) = (4\pi R)^{-1}$ is called the Green function of the Laplace operator.

Today we’d like to find a similar solution for the time-dependent case.

23.3 A PHYSICALLY MOTIVATED GUESS

We might expect that the fields at a spatial position \vec{r} would again be determined by currents at \vec{r}_* , with a $1/R$ falloff. But we also expect that signals will travel from source point to field point at the finite speed c . So a simple guess for the generalization of Equation 23.4 is that each component of \vec{A} is given by

$$\vec{A}(t, \vec{r}) \stackrel{?}{=} \frac{\mu_0}{4\pi} \int d^3r_* \frac{1}{R} \vec{j}(t - R/c, \vec{r}_*). \quad \text{Coulomb gauge} \quad (23.5)$$

In words, we are again proposing that the vector potential at time t gets contributions from each source point. In the case of stationary currents, \vec{j} is time-independent and our guess reduces to the known answer for that case. For time-dependent currents, our guess says we must *look back in time* to the moment $t - R/c$ when that source region’s current could have influenced our observer’s field point \vec{r} at time t .

Your Turn 23A

Before proceeding, verify that the proposed solution Equation 23.5 really obeys the Coulomb gauge condition $\vec{\nabla} \cdot \vec{A} = 0$. [*Hint*: Adapt the approach used in magnetostatics (Section 14.4.3, page 166).]

The form of our solution suggests part of the answer to Hanging Question #H (page 24): The fields observed at some time t have nothing to do with the source *at that time*. We may have turned off the apparatus; a radiating star may have died out; an electron/positron pair may have annihilated by the time radiation gets to our apparatus. *Once formed, radiation proceeds autonomously* through space. It reflects only the behavior of currents at the *retarded time*.

23.4 CHECK THE GUESS

We now apply the **wave operator** $\square = \nabla^2 - c^{-2}\partial^2/\partial t^2$ to our proposed solution, to see whether we recover $-\mu_0\vec{j}$ (Equation 23.3).²

The wave operator involves derivatives with respect to the field point and observation time (the unstarred coordinates), so all time derivatives below will denote $\partial/\partial t$. Also $\vec{\nabla}_i = \partial/\partial r_i$.

¹See Equation 2.5.

² \square is also called the D’Alembert operator, or “d’Alembertian.”

23.4.1 Preliminary results

Your Turn 23B

Show that (or review why)

$$\vec{\nabla} R = \hat{R}; \quad \vec{\nabla} \cdot \vec{R} = 3; \quad \vec{\nabla}(R^{-p}) = -pR^{-(p+1)}\hat{R}; \quad \nabla^2(R^{-1}) = -4\pi\delta^{(3)}(\vec{R}).$$

To save writing, let ϕ denote any component of $4\pi\vec{A}/\mu_0$, and \mathcal{J} the corresponding component of \vec{j} . So our proposed Green function solution Equation 23.5 says

$$\phi(t, \vec{r}) = \int d^3r_* \frac{1}{R} \mathcal{J}(t - R/c, \vec{r}_*), \quad (23.6)$$

and we wish to show

$$\nabla^2 \phi - c^{-2} \frac{\partial^2}{\partial t^2} \phi \stackrel{?}{=} -4\pi\mathcal{J}. \quad (23.7)$$

The gradient of Equation 23.6 is

$$\vec{\nabla} \phi = \int d^3r_* \left[(\vec{\nabla}(R^{-1})) \mathcal{J}(t - R/c, \vec{r}_*) - \frac{1}{cR} (\vec{\nabla} R) \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} \right].$$

Here the subscript “ret” means to evaluate at the **retarded time** $t - R/c$ (after taking any indicated derivatives).

Taking another derivative gives

$$\begin{aligned} \nabla^2 \phi(t, \vec{r}) &= \int d^3r_* \left[(\nabla^2 R^{-1}) \mathcal{J}(t - R/c, \vec{r}_*) - c^{-1} (\vec{\nabla} R^{-1}) \cdot (\vec{\nabla} R) \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} - c^{-1} \vec{\nabla} \cdot (R^{-1} \hat{R} \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}}) \right] \\ &= \int d^3r_* \left[-4\pi\delta^{(3)}(\vec{R}) \mathcal{J}(t - R/c, \vec{r}_*) - c^{-1} (-R^{-2} \hat{R}) \cdot \hat{R} \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} - c^{-1} \vec{\nabla} \cdot (R^{-2} \vec{R} \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}}) \right] \end{aligned}$$

The three delta functions eliminate the integral over \vec{r}_* and set $\vec{r}_* = \vec{r}$, so continuing,

$$\begin{aligned} &= -4\pi\mathcal{J}(t, \vec{r}) + \underbrace{\int d^3r_* \left[(cR^2)^{-1} \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} + c^{-1} 2R^{-3} \hat{R} \cdot \vec{R} \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} - (cR^2)^{-1} 3 \frac{\partial \mathcal{J}}{\partial t} \Big|_{\text{ret}} \right]}_{\text{cancel}} \\ &\quad + (cR)^{-2} \vec{R} \cdot (\vec{\nabla} R) \frac{\partial^2 \mathcal{J}}{\partial t^2} \Big|_{\text{ret}}. \end{aligned}$$

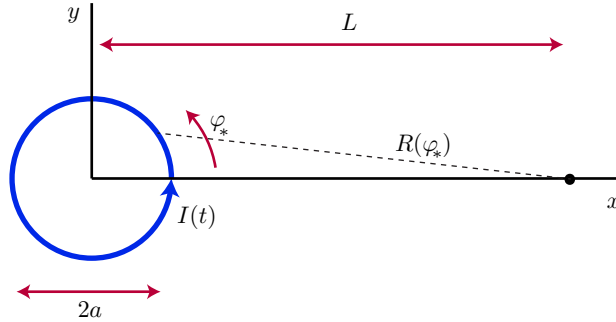
The three terms in the brace cancel.

Bringing it all together, we have shown that Equation 23.6 solves Equation 23.7 for any \mathcal{J} . Reinstating the vector character of \vec{A} and multiplying by $\mu_0/(4\pi)$ proves Equation 23.5, the Green function solution to the Coulomb-gauge vector potential created by a specified current distribution with net charge everywhere zero.

23.5 OUR FIRST ANTENNA

23.5.1 Far fields

Consider a circular loop of wire in the xy plane, centered on the origin, with radius a . A sinusoidal current $I(t) = \bar{I} \cos(\omega t)$ runs around this loop. In short, our antenna is an *oscillating magnetic dipole*. Charge never piles up anywhere, so $\rho_q = 0$ and we may use the formula developed in the preceding section.



Our sign convention for the current is that at time zero, current flows counterclockwise around the loop as viewed from above (from $z > 0$).

We know the fields far from a *static* magnetic dipole: $\vec{E} = 0$ and \vec{B} falls with distance like $1/r^2$. Now we want to explore what changes when the current alternates.

We imagine sitting somewhere far away along the $+x$ axis, at position $\vec{r} = (L, 0, 0)$.³ We ask: What are the electromagnetic fields there, to leading nontrivial order in powers of $1/L$?

We parameterize the wire loop by azimuthal angle φ_* , which runs from zero (closest point to our observer) to 2π (same point). At any point on the loop, the current points in the direction $\pm\hat{\varphi}_*$. So Equation 23.5 gives⁴

$$\vec{A}(t, \vec{r}) = \frac{\mu_0 \bar{I}}{4\pi} \int_0^{2\pi} (a d\varphi_*) R^{-1} \left[\frac{1}{2} e^{-i\omega(t-R/c)} \hat{\varphi}_* + \text{c.c.} \right].$$

In the preceding formula, $R = \sqrt{(L - a \cos \varphi_*)^2 + a^2 \sin^2 \varphi_*}$ and $\hat{\varphi}$ is the unit tangent vector to the loop at angular position φ_* .

Our answer can be simplified a lot because we are only interested in the leading-order behavior in powers of $1/L$ (the **far fields**). Thus $R^{-1} = L^{-1} + \dots$, where the ellipsis contains only terms that we agreed to drop. The leading term is independent of φ_* , so it comes outside the integral, along with the time dependence factor:

$$\vec{A} = \frac{\mu_0 \bar{I}}{4\pi} \frac{a}{L} \frac{1}{2} e^{-i\omega t} \int_0^{2\pi} d\varphi_* (-\hat{x} \sin \varphi_* + \hat{y} \cos \varphi_*) \exp \left[i \frac{\omega}{c} L \left(1 - \frac{a}{L} \cos \varphi_* + \dots \right) \right] + \text{c.c.}$$

We must be careful with the exponential. Here the first subleading term may *not* be dropped. Even though it is smaller than the leading term, nevertheless it is not small in an absolute sense, because the L factors cancel.

Your Turn 23C

Check that the terms even higher than this one *do* fall off as $L \rightarrow \infty$. (That is why we abbreviated them by an ellipsis.)

We therefore find

$$\vec{A} \rightarrow \frac{\mu_0 \bar{I}}{4\pi} \frac{a}{L} \frac{1}{2} e^{-i\omega(t-L/c)} \int_0^{2\pi} d\varphi_* (-\hat{x} \sin \varphi_* + \hat{y} \cos \varphi_*) \exp[-i(\omega a/c) \cos \varphi_*] + \text{c.c.} \quad (23.8)$$

³By rotational symmetry, we get a similar result when we go far away in any direction in the xy plane. Later we'll study this situation more generally and get the fields everywhere.

⁴Equation 14.19 gives the charge flux in the thin-wire approximation.

The term that points along \hat{x} integrates to zero by a symmetry argument: It is an odd function of φ_* , which may be integrated over the symmetric range $(-\pi, \pi)$. The \hat{y} term need not be zero, however. We conclude that the vector potential far away from the loop has a contribution that, at nonzero frequency, falls slowly with distance, as L^{-1} .

Note that the \hat{y} term of Equation 23.8 would *also* integrate to zero in the static case ($\omega = 0$); more generally, however, it does not vanish.

Your Turn 23D

- Suppose that ω is small but nonzero; use a Taylor expansion of the exponential to get an approximate answer for the integral.
- Suppose that ω is large; use stationary-phase approximation to get an answer in this limiting case.
- Ask Wolfram Alpha about `Integrate[Cos[t]*E^(-I*p*cos[t]),{t,-Pi,Pi}]`. Graph the answer and look at the limits for large and small $p = \omega a/c$.

*Although there are **no net charges** anywhere, we nevertheless found an electric field, in contrast to the case of a static magnetic dipole. Moreover, the field falls off slowly with distance, as $1/L$, in contrast to even a static **electric dipole**.*

Because we used restricted Coulomb gauge, the scalar potential is zero. Thus, the electric field is simply $-\frac{\partial}{\partial t}\vec{A}$. The time derivative just introduces a factor of $-(-i\omega)$, so

$$\vec{E} \rightarrow (\text{const})\frac{1}{L}e^{-i\omega(t-L/c)}\hat{y} + \text{c.c.} \quad \text{on } x \text{ axis}$$

We also get a prediction that the outgoing wave observed at this point is nearly a *plane wave* traveling along $+\hat{x}$ and *linearly polarized* along \hat{y} . Thus, it is polarized transversely to the direction of propagation.

What about the magnetic field, given by the curl of \vec{A} ? We might naively imagine that it must fall as L^{-2} , but think about the factor $e^{i\omega L/c}$ in Equation 23.8. When we move in the \hat{x} direction, this factor has a derivative that introduces a factor of $i\omega/c$, and *not* any additional L^{-1} . Thus, the leading far-field behavior of \vec{B} is

$$\vec{B} \rightarrow (\text{const}')\frac{1}{L}e^{-i\omega(t-L/c)}\hat{z} + \text{c.c.},$$

a *slower falloff* with distance than in the case of a static magnetic dipole. The magnetic field is also perpendicular to \hat{x} , and also to the electric field, as in any plane wave. In fact it points along the direction of the magnetic dipole moment whose oscillation gave rise to the wave.

The magnetic field also falls off slowly with distance, as $1/L$, in contrast to a static magnetic dipole. Together, the fields again form an approximately plane wave moving toward the observer. For an observer in the xy plane the magnetic field points along $\pm\hat{z}$.

Your Turn 23E

Keep track of factors that were dropped in the preceding formulas and confirm two other key features:

For an oscillating magnetic dipole source, the fields are also proportional to the amplitude of the oscillating dipole moment (here $\pi a^2 \bar{I}$), and to the frequency squared.

23.5.2 Energy loss

The slow field falloffs in \vec{E} and \vec{B} are the hallmark of *radiation*. They imply that energy is being continually sent out to infinity, if the frequency $\omega \neq 0$. To see this, recall from Chapter 18 that a test charge can extract power proportional to $\|\vec{E}\|^2$. Although the direction of \vec{E} oscillates, its mean-square value is nonzero. Imagine a shell of such receivers at distance L from the source. The area of that shell increases as L^2 , whereas the energy we can extract falls like $\|\vec{E}\|^2$, that is, as L^{-2} . So the total energy sent out from the source is *independent of L* . In other words, our antenna sends energy out all the way to infinity: It *radiates*, as a candle radiates light.

23.5.3 Directionality

See Problem 23.1.

23.6 PLUS ULTRA

This is the end of Part Three of these notes. In a sense, we could stop here: We know most of what's needed to understand the second Industrial Revolution.⁵ We have also found an unexpected electromagnetic phenomenon (waves), including specific details (about polarization). We have seen one way for waves to be generated, and how they deliver energy, momentum, and even angular momentum across empty space.

But there is a lot more work to do! Many antennas of interest don't have zero net charge (for example, the one in our microwave generator demo), so we'll need a more general formalism. Also, today we only found the fields along one axis, and so on. However, every complicating thing that we'll do later is just a variation on the fairly simple calculation in Section 23.5.1.

More importantly, although the derivation we gave today was straightforward, there was *too much magic*. We should develop a more sophisticated formalism, and accompanying physical intuitions, that will make it clear that Equation 23.5 is correct, without all the messy verification. The first step to a deeper understanding of this and other magic is to uncover an important aspect of Maxwell's equations that has been hiding in plain sight ever since we added Maxwell's correction to Ampère's law.

⁵That's the one involving electrical technology. The next revolution (semiconductors) involved quantum mechanics.

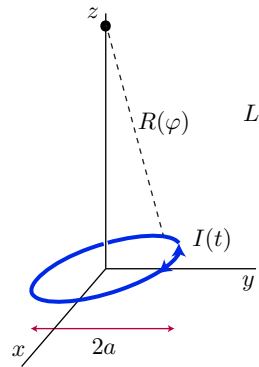
FURTHER READING

An alternative to the derivation in this chapter appears in Pollack & Stump, 2002, §15.1.1.

PROBLEMS

23.1 *Directionality of antenna*

A circular loop of wire, carrying an oscillating current, lies in the xy plane:



I obtained a formula for the potentials:

$$\psi(t, \vec{r}) = 0; \quad \vec{A}(t, \vec{r}) = \frac{\mu_0 \bar{I}}{4\pi} \int_0^{2\pi} (a d\varphi_*) R^{-1} \left[\frac{1}{2} e^{-i\omega(t-R/c)} \hat{\varphi} + \text{c.c.} \right].$$

In this formula, \vec{r} is position of the field observation. The angle φ_* specifies an element of the loop located at $\vec{r}_* = a\hat{r}$. The unit vectors \hat{r} and $\hat{\varphi}$ are evaluated on the loop at φ_* . The distance $R(\varphi_*) = \|\vec{r} - \vec{r}_*\|$. The current in the loop is everywhere $I(t) = \bar{I} \cos \omega t$.

The main text examined the far fields at points along the x or y axes. Instead, now find the vector potential, this time for an observer located along the z axis. Then characterize the far electric and magnetic fields in words and contrast with their far-field behavior when viewed at points along the x axis.

23.2

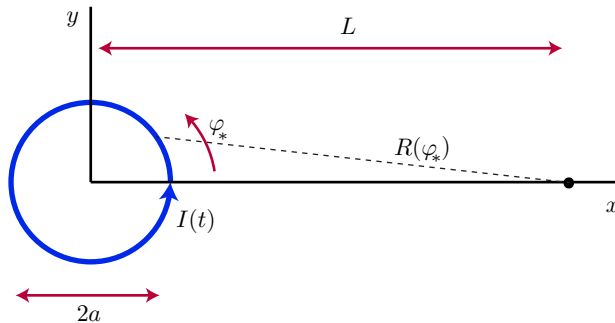
Repeat the analysis of Section 23.5 for an antenna that is a *square* loop of wire with side a . In particular evaluate the far fields for the limiting case of low frequency and compare to the result in Your Turn 23Da (page 264). Can you make a statement that covers both cases?

23.3 From far to near fields

Background: This chapter derived an exact expression for the vector potential outside an arbitrary current distribution, for the situation with zero charge density everywhere. Section 23.5 (page 262) specialized to the case of an oscillating current confined to a loop of wire. Then we made a “far field” approximation: The observer was assumed to be far away, so we discarded $\mathcal{O}(L^{-2})$ terms. Your Turn 23Da made the additional approximation of long wavelength (low frequency, nonrelativistic source motion). That was useful for specialized situations. In this problem, you’ll get your assistant to compute the fields *without* either of these approximations.

We may guess that close to the source, at each instant of time the magnetic field looks like the field around a *static* dipole. Thus, each magnetic field line wraps around the wire (it’s *linked* with the current loop). Farther from the source, however, the field lines must detach from the source and move outward on closed paths that don’t pass through the current loop. We’d like to see how and where this detachment occurs.⁶

Problem: Again consider a circular loop of wire of radius a in the xy plane, carrying a prescribed, harmonically oscillating current $I(t) = \frac{1}{2}[\bar{I}e^{-i\omega t} + \text{c.c.}]$:



Thus \bar{I} is one half of the peak-to-peak current amplitude. You are to find and plot the magnetic field $\vec{B}(t, \vec{r})$ everywhere, at various times. This “merely” involves numerically evaluating a formula obtained in Section 23.5.1:

$$\vec{A}(t, \vec{r}) = \frac{\mu_0 \bar{I}}{4\pi} \int_0^{2\pi} (a d\varphi_*) R^{-1} \left[\frac{1}{2} e^{-i\omega(t-R/c)} \hat{\varphi} + \text{c.c.} \right].$$

More precisely, you’ll work out the curl of this expression, and *then* evaluate it numerically. In this formula $R = ((L - a \cos \varphi_*)^2 + a^2 \sin^2 \varphi_*)^{1/2}$ and $\hat{\varphi}$ is the unit tangent vector to the loop at angular position φ_* .

Actually, it’s enough to examine $\vec{B}(t, \vec{r})$ only for \vec{r} in the xz plane, and indeed to look only at $x > 0$, because of the azimuthal symmetry. But unlike in the main text, don’t restrict to \vec{r} just along the x axis.

Because $\vec{\nabla} \cdot \vec{B} = 0$, none of the field lines can terminate. Nevertheless we’ll find that some are attached to the source whereas others are not.

Steps:

Measure all lengths in units of a . (Or equivalently, measure lengths in meters and take

⁶Heinrich Hertz made similar sketches (Wiedemann’s Ann. **36**, 1 (1889)).

$a = 1$ m.) Measure time in units of a/c . The numerical value of c is 1 in these units (that is, in units of $a/(a/c)$).

- a. Write the general expression to be evaluated, specialized for the situation in the problem. This expression involves an integral, which you'll eventually have to do numerically, but not yet. Using this unevaluated expression, show that one of the three cartesian components of \vec{B} equals zero throughout the xz plane.

That's convenient: it means that every integral curve (streamline) of \vec{B} that starts in the xz plane will remain completely in that plane. These curves are Faraday's magnetic "field lines."

- b. Set $\omega = 0$ and check your analytic results by comparing to a case that you know, that is, the far fields that we computed when we did the magnetic multipole expansion in magnetostatics.⁷ Then numerically evaluate your complete result for \vec{B} on a grid of points with, $y = 0$ and say, $0 < x < 5a$ and $-5a < z < +5a$. Get your software to show the streamlines of this vector field.⁸ I found I got nice results if I told the software specifically to make me streamlines that pass through the points $(x_{0i}, 0)$, that is, points along the x axis at an evenly-spaced series of values $\{x_{0i}\}$. (Just make sure none of your choices is $(1, 0)$, because the fields are singular exactly on the wire.)

Note that overall factors like $\mu_0 \vec{I}/(4\pi)$ drop out when all we ask for is the streamlines. Your computer will probably choose different scales for the x and z axes in your plot, because it wants to make you happy. But that's not what you want in this instance, so figure out how to override that default behavior.

- c. Some or all of your integral curves will have the property that they *link* (pass through) the current loop:⁹ We say they are "attached to the source." Find which ones have this property and comment.
- d. Repeat, this time taking $\omega = 2\pi c/(3a)$ (not zero) and time $t = 0$. This time, we expect the far fields to be waves with wavelength $3a$. Comment on the behavior you observe both close to and far from the origin; on the z axis versus on the equatorial plane; etc. If some of the integral curves (field lines) are not linked with the source loop, estimate the locus separating the attached lines from the detached ones.
- e. Also compute the longitudinal part of \hat{B} , that is, $\hat{r} \cdot \vec{B}$ at time zero, and plot it in some way that shows how it falls off with distance from the loop. If there's something surprising about your answer, explain it; otherwise explain why it's not so surprising. [Remark: You may get more visually informative plots if you just graph this quantity along a couple of straight rays from the origin, that is, lines $x = \eta z$ for a couple of values of the slope η . Maybe you should also make a relevant comparison between $\hat{r} \cdot \vec{B}$ and $\|\vec{B}\|$.]

⁷See Your Turn 16A (page 188).

⁸You learned how to get a computer to draw streamline plots in Problem 3.9. For example, Python has a function `plt.streamline` that accomplishes this. Then Problem 14.6 discussed the fields created by a stationary current loop of finite size.

⁹You are plotting a slice, the field in the xz plane, so the current loop just looks like the two points $(\pm a, 0)$, one of which is outside the range you are plotting. Indicate the other one in your plot by a dot. A curve in the xz plane "links" the current loop if it encircles that point.

- f. Show some initiative. Suppose these are figures in a paper you're trying to publish—figure out some improvements in presentation, informative labels, etc.¹⁰ If you think that the range from 0 to $5a$ doesn't show the physics optimally, choose some better range. Play.
- g. A picture may be worth a thousand words, and N pictures may be worth N thousand words, but still a *movie* of those N pictures would be better still. After all, we are studying a spacetime phenomenon. So get your computer to make video frames for many moments throughout a period $2\pi/\omega$, then stitch them together into a video. The *Student's Guide to Python* §8.3.2 tells one simple way to do this, and that's good enough. However, if you want to upload to YouTube or other social media, then you'll need the more elaborate recipe involving `ffmpeg`, also described in that section of the *Guide*.

Hint: You'll get a smoother movie if you choose initial points appropriately. At time t , ask your software for streamlines that pass through $(x_{0i} + ct, 0)$, where $\{x_{0i}\}$ are the points you used at time zero.

23.4 *Twist it up*

First do problem Problem 23.3 parts a–d. But then consider a current source consisting of *two* circular loops of wire. One lies in the xy plane and again carries sinusoidal current with frequency ω the same as in part (d). The other lies in the xz plane and carries sinusoidal current with the same frequency and amplitude, but shifted in phase by $1/4$ cycle relative to the first one. In this situation we may *not* restrict everything to the xz plane.

- Write a superposition of two formulas each similar to the one you used in Problem 23.3 part (d).
- Choose a moment of time at which the current in the xz loop equals zero (and hence the current in the xy loop is maximum). Write a function that can evaluate \vec{B} anywhere in space at the one instant of time you chose.
- Make a three-dimensional streamplot of some representative magnetic field lines that pass through a collection of starting points lying along the $+x$ axis. Rotate your plot to gain some perspective. Print one or two good-looking views, but describe in words how they look as three-dimensional curves, and how they interpolate between what you expected at short and at long distances.

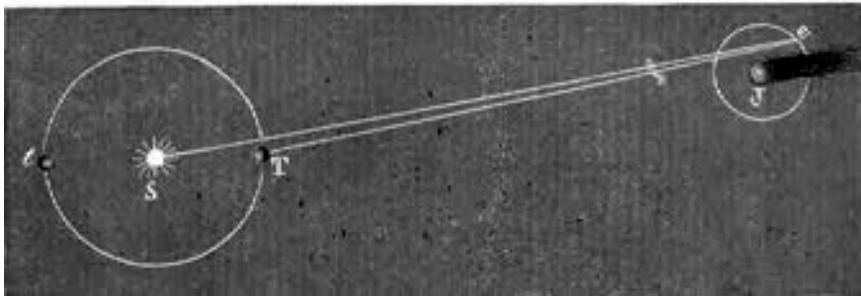
Optional: If you think this would be better as a movie... nobody's stopping you.

¹⁰Life is short, so if you wish you can hand-write your labels on a hard copy instead of figuring out how to get your computer to set them.

PART IV

Relativity: Low Tech

FIG. 136.



“When, in a relativistic discussion, I try to make things clearer by a spacetime diagram, the other participants look at it with polite detachment and, after a pause of embarrassment as if some childish indecency had been exhibited, resume the debate in their own terms.” – J L Synge, 1960

CHAPTER 25

Galilean Relativity

False views, if supported by some evidence, do little harm, as every one takes a salutary pleasure in proving their falseness.

— Charles Darwin

This chapter’s goal is to rephrase some familiar ideas in a useful way.

25.1 PRINCIPLE OF RELATIVITY

Galileo believed that the Earth moved around the sun, while also spinning on its axis. Many found this proposition absurd. If the Earth moves, why doesn’t it *feel* like we’re moving? Why aren’t we thrown off? Galileo patiently constructed arguments about how you can play ping-pong on a ship moving uniformly on a calm sea and never notice that the ship is moving. While he didn’t have it completely straight, his successors (Huygens and Newton) eventually elevated this idea to the status of a fundamental principle, which we now call the **Principle of Relativity**:¹

No experiment done within an isolated system can determine whether or how fast that system is moving. More precisely, if we put all our apparatus in a box and measure time and space via instruments anchored to that box, then the results of any experiment will be the same regardless of whether that box is at rest or moving in a straight line at uniform speed.

Einstein didn’t introduce the P of R. Nor did he overthrow it: We still believe it to be experimentally correct. What Einstein said was that newtonian physics *implements* the principle in a way that is demonstrably wrong. Before we get into that, these notes will review the newtonian situation.

[T2](#) Section 25.1’ (page 280) discusses the notion of “isolated system.”

25.2 A SIMPLE SYSTEM

Let’s see how the P of R plays out in a concrete situation. Consider two point masses of equal mass m joined by a spring with equilibrium length L and spring constant k , floating freely in outer space (or moving in 1D along a frictionless air track). Newtonian mechanics says that their motions are solutions to the equations

$$\frac{d^2x}{dt^2}(1) = -\frac{k}{m}(x_{(1)} - (x_{(2)} - L)) \quad \frac{d^2x}{dt^2}(2) = -\frac{k}{m}((x_{(2)} - L) - x_{(1)}). \quad (25.1)$$

¹Henri Poincaré seems to have introduced this phrase, centuries later. A “principle” is not a firm starting point that you can use to prove other things. Nor is it itself a provable proposition. Think of a “principle” as a *generator of interesting hypotheses*.

Although these are familiar equations, let's unpack their content a bit.²

In classical physics, the fundamental entities are **events**. An event is specified by a location in space and a moment in time. A **trajectory** is a continuous string of events. We think of events as points in a four-dimensional space, called **spacetime**, and trajectories as curves in spacetime. To do analytical work, we must uniquely assign *four numbers* to each event; that is, we must impose a choice of coordinate system on spacetime. In this language, Equations 25.1 are shorthand for saying that:

It is possible to label events (points in spacetime), in such a way that every allowed motion of this system corresponds to a pair of curves in spacetime whose coordinate representations are solutions to Equation 25.1.

The following sections review a key fact about newtonian mechanics in this context:

*Newton's laws of motion have a mathematical property called **galilean invariance**, which guarantees that the physics they predict will obey the Principle of Relativity.*

Our ultimate goal is to investigate the same claim about Maxwell's equations and show it's *not* valid. However, we'll find a *different*, true, property that again guarantees the P of R. First we'll see how it works in newtonian physics, in two equivalent formulations.

25.3 ACTIVE VIEWPOINT: SYMMETRY

Let's recall some solutions to our equations:

$$x_{(1)}(t) = C \cos(\omega t) \quad x_{(2)}(t) = L - C \cos(\omega t).$$

Here C is any constant and $\omega = \sqrt{2k/m}$. Of course, starting from one such solution we can manufacture many others by adding any constant A to both $x_{(1)}$ and $x_{(2)}$:

$$\tilde{x}_{(1)}(t) = x_{(1)}(t) + A \quad \tilde{x}_{(2)}(t) = x_{(2)}(t) + A. \quad (25.2)$$

I'll call such transformations **active**, because the new solution is a *physically different motion* from the original. The operation in Equation 25.2 transforms any solution of the equations of motion into *another* solution (and nonsolutions to nonsolutions). We call it a **symmetry** of the equations.

That is, symmetries *permute* the solutions of a system of equations among themselves. In addition to the overall translation described by Equation 25.2, any isolated, 1D newtonian system also has symmetry under shifts of *time* by any constant. (There are also discrete symmetries involving reflections in space and in time.)

25.4 PASSIVE VIEWPOINT: INVARIANCE

There is an equivalent viewpoint to the one in the preceding section that is often more convenient. Instead of cataloging all the *solutions* and looking for transformations that permute them, we focus on a property of the *equations* themselves.

²See also Section 1.5 (page 17).

To see how it works, start with any trajectory and reexpress the *same trajectory* in a new set of coordinates:

$$x' = x - A \quad t' = t - B. \quad (25.3)$$

Because we are not physically changing the trajectory, this transformation is called **passive**: it just changes the *representation* of a trajectory. Equation 25.2 shifted any trajectory to the right by A , whereas Equation 25.3 shifts the coordinate *axes* to the right by A .

We now change variables in the equations of motion to see how they look when expressed in terms of the new coordinates: The usual rules of calculus give $d/dt = d/dt'$. Everywhere else, we just substitute $x' + A$ wherever we see x :

$$\frac{d^2}{dt'^2}(x' + A) = -\frac{k}{m}(x' + A - (x'_{(2)} + A - L)).$$

Cleaning up, we see that the *form* of the equation of motion, after expressing it in the new variables, is the same as it was in the old variables (Equation 25.1), including the numerical values of constants, k , L , and m . We say that the original equations have an **invariance** under the passive transformation Equation 25.3.

The active and passive viewpoints are equivalent: To every active symmetry there's a corresponding passive invariance and vice versa.

25.5 ROTATIONS

Continuing with the passive viewpoint, we now upgrade to a world with two spatial dimensions. If we set up cartesian axes, we can label every point in the plane by two numbers $\begin{pmatrix} x \\ y \end{pmatrix}$. Then the *same* point viewed from a *rotated* point of view will be labeled by two different numbers $\begin{pmatrix} x' \\ y' \end{pmatrix}$. We can find the new coordinates by using trigonometry, and the fact that the new coordinate axes are rotated by some angle α relative to the old ones. There's a simple formula expressing this:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \quad (25.4)$$

To think about this conceptually, imagine digging up all the streets in Manhattan and laying down a new grid of streets rotated counterclockwise relative to the old one by α . Then if the Empire State Building is at a point P , it will still be at the same point P after the new grid is laid down, but the *coordinates* of that point (which street and which avenue) will no longer be the same as they were before.

Now, certainly there are many other coordinate systems we could use to label points in the plane, besides the two cartesian systems just described. For example, we could use axes that are not at right angles. But there is something special about a cartesian system: The distance between two points P_1 and P_2 is given by the simple formula $d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. If we describe the points using the rotated coordinate system, the formula has exactly the same form:³ $d = \sqrt{(x'_1 - x'_2)^2 + (y'_1 - y'_2)^2}$.

³See Section 13.2.

Generic coordinate transformations don't have this property. For example, if we define new coordinates via a "dilatation" transform, $\vec{r}' = 2\vec{r}$, the form of the distance function is not quite the same. In short,

In euclidean geometry, one class of coordinate systems is special (the cartesian systems). Within that class, however, any system is just as good as any other one.

When we upgrade the equations of motion for two balls on a spring from 1D to 2D or 3D, they involve the spring potential energy $U = \frac{1}{2}k\|\vec{r}_{(1)} - \vec{r}_{(2)}\|^2$. Because the distance function takes the same form when expressed in terms of a rotated coordinate system, the equations of motion will have the same property: They are rotation invariant.⁴ In contrast, when expressed in terms of dilated coordinates the equations of motion take a new form that look similar but that have a different value of k/m : Newtonian physics does not have any invariance under dilatations.

25.6 GALILEAN GROUP

25.6.1 Some coordinate systems on spacetime are preferred

In math, the assignment of a coordinate system to a space is pretty flexible. Certainly there are lots of choices we could make on our four-dimensional spacetime. But in most of these choices, the equations of physics look pretty weird. We already saw one example (dilatation). Similarly, most time-dependent transformations, such as $\vec{r}' = \vec{r} + \vec{a}t^2/2$, introduce new "fictitious forces."⁵ That is, the equations are again not form-invariant when reexpressed in terms of this \vec{r}' .

Turning that observation around, we can ask which coordinate systems *do* leave the form of Newton's laws invariant. In other words, we can let *physics* select the good systems. We will call them **G-inertial**, in honor of Galileo. Translations like Equation 25.3 and rotations like Equation 25.4 are invariances of newtonian physics, and hence they take one such G-inertial coordinate system to another.

A lot of confusion arises over the use of phrases like "frame of reference" (and "observer," which sounds like it gives an essential role to human consciousness). I will instead usually refer to a "coordinate system," which may or may not have the property that the equations of motion take their usual form. If they do, then the coordinate system is "G-inertial" (or simply "good"). A human observer always has the *option* of setting up a G-inertial coordinate system to describe what she measures,⁶ and many accounts of relativity implicitly assume this, but actually doing so may be an elaborate and subtle procedure in practice. Also, beware that the good coordinates for newtonian physics differ from those in Einstein physics, yet most authors refer to both indiscriminately as "inertial." When necessary, I will disambiguate with the prefix "G-" (galilean) (and later "E-" for Einstein).

⁴Exactly the same reasoning establishes the rotation invariance of two masses bound by *gravitational* force.

⁵The "Coriolis force" is another example.

⁶We must qualify this statement to say "in the absence of gravitational effects." Near a gravitating body, an observer can only set up a *locally approximately* inertial system.

25.6.2 Boosts

Returning to one dimension, there's another important class of symmetry transformations, called **galilean boosts**. In active form we may write them as

$$\tilde{x} = x + v_*t, \quad \tilde{t} = t. \quad (25.5)$$

You can readily show that applying this transformation to any solution of the particular equations of motion Equation 25.1 yields another solution.

Your Turn 25A

Show that the corresponding passive coordinate transformations:

$$x' = x - v_*t, \quad t' = t \quad (25.6)$$

are invariances of the equations of motion. That is, show that reexpressing Equation 25.1 in terms of the new variables yields equations of identical form.

Equation 25.6 describes a new coordinate system whose axes are moving to the right at speed v_* relative to the original. The minus sign indicates that these moving axes can overtake an object moving to the right.

25.6.3 Matrix notation

It will sometimes be convenient to express Equation 25.6 in matrix form:

$$\begin{bmatrix} t' \\ x' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -v_* & 1 \end{bmatrix} \begin{bmatrix} t \\ x \end{bmatrix}. \quad \text{galilean boost} \quad (25.7)$$

Your Turn 25B

Show that if we make a second transformation of this sort, to t'' , x'' , then we just get the product of two matrices, which is again a galilean boost, this time by $v_{*1} + v_{*2}$, that is, the matrix $\begin{bmatrix} 1 & 0 \\ -(v_{*1}+v_{*2}) & 1 \end{bmatrix}$.

That **galilean velocity addition formula** agrees with our everyday experience with baseballs and water waves.

25.6.4 Group structure

All together, in one space dimension newtonian physics has a 3-parameter family of continuous symmetries/invariances (space translation, time translation, boost). We call that family the **galilean group**. Its elements are **galilean transformations**. In the passive viewpoint, they connect the various G-inertial coordinate systems to one another.

Suppose that we define a primed coordinate system by applying a galilean boost, and a translation, to the original frame. Next, define a *double*-primed frame by applying

a second galilean boost, and another translation, to the primed frame. Still working in one dimension,

$$x'' = x' - v_{*1}t' - A_1 = (x - v_{*2}t - A_2) - v_{*1}t - A_1, \quad t'' = t' = t. \quad (25.8)$$

We see that the overall effect is again the combination of a boost (with speed $v_{*1} + v_{*2}$ as you found in Your Turn 25B) and a shift (by $A_1 + A_2$).

In three space dimensions, the galilean group includes a 10-parameter family of invariances (3 space translations, 1 time translation, 3 space rotations, 3 boosts).

Your Turn 25C

- Generalize Equation 25.8 to include time shifts also.
- Show that if we apply any two of these transformations in succession, the result is a single transformation that is also in this family.
- Show that any such transformation has an inverse, which is again in the family.

Mathematicians call a set of transformations with those properties is **group**, hence the name “galilean group.”

25.6.5 Physical significance of invariance

By now, certain questions may be bothering you:

- Why are we spending so much time with balls on springs? Even within newtonian physics, that’s a specialized, and idealized, system.
- A coordinate system is just an arbitrary labeling scheme for points of spacetime. So what has all this formalism got to do with physics?

The answer to the first question is that

All of newtonian physics has the overarching mathematical property of galilean invariance that transcends details of particular springs, clocks, planets, etc. (25.9)

Your Turn 25D

For example, confirm that in newtonian gravity, the equations of motion for two point masses attracting each other also have full galilean invariance.

Idea 25.9 partly explains why in physics we get so much mileage out of studying systems that are obviously absurdly oversimplified, for example, linear springs, spherical planets, and other nonexistent things. Often we are just working out the consequences of invariances that continue to apply to realistic versions of those things. For a simple example of why this principle is significant, notice that invariance under spatial translations means *there is no distinguished special central point* in space.

For question #2 above, note that some of the invariances (the boosts) connect two coordinate systems that are in uniform, straight-line motion relative to each other.

Because any set of newtonian equations of motion is invariant under such transformations, then *those two frames of reference are indistinguishable* by any experiment confined to the system under study. You can do all the experiments you like, and always find the same equations of motion in each such frame. Nothing you can measure says that one such frame is at “absolute rest” or indeed “better” in any way than another. In short:

Newtonian physics hardwires the Principle of Relativity by using equations of motion that are invariant under galilean boosts. (25.10)

Let me try one more lapidary phrase:

Physicists study invariance because it strips away details and lays bare the structural essentials of a dynamical theory. (25.11)

We can now see why Idea 25.9 is so important: If *part* of physics had galilean invariance, but another part did not, then we could devise an experiment using the second part to determine which frames are at absolute rest. Even if two parts of physics have slightly *different* boost invariances, we could say that “absolute rest” was the coordinate system in which both simultaneously took their simplest forms. Only if *all* of physics has the same boost invariance can we say that absolute rest is completely unobservable—the Principle of Relativity.

Many physical problems involving relativity become clearer when seen from this high-altitude viewpoint: Often, their solution boils down to

- *There’s an inertial coordinate system where I know what’s going on.*
- *But I want to know what’s going on in some other inertial system (perhaps one that I set up in my lab).*
- *So I can use the appropriate transformation to go from the first to the second.*

**Relativity
Strategy**

(25.12)

Applying this strategy to every situation is not always the fastest route to solve a particular problem. But in the long run it’s a unified, sure-footed way to cut through the fog.

We will soon see that Einstein *retained* most of Ideas 25.9–25.10 and merely tweaked some details of how the transformations work (Chapters 28–29). Once we discover the right transformations, we’ll see some examples of the Relativity Strategy at work.

25.6.6 Light cannot be interpreted as a stream of newtonian particles, part 2

Section 18.2 (page 220) argued that Newton’s model of light as a stream of particles was incompatible with the alternative model implied by Maxwell’s theory. Here is a more direct, experimental objection to the newtonian model.

Suppose that we have a catapult that, when at rest, can fire a projectile in any direction with initial speed v_{*1} . Imagine mounting that catapult on a train car, bringing it up to speed v_{*2} directed along \hat{x} , and firing the projectile in the $+\hat{x}$ direction.

Intuitively we might expect that on the ground, we'll observe the projectile moving with velocity $(v_{*1} + v_{*2})\hat{x}$.

Let's obtain the result just stated as a consequence of galilean invariance, using Equation 25.12. We know that there's a G-inertial frame in which the catapult appears to be at rest. *Whatever* mechanism is inside the catapult, we are assuming it to be galilean-invariant, so the speed of the projectile from the moving catapult, viewed in the moving frame, must again equal v_{*1} . Apply your result in Your Turn 25B to find the speed as seen in the ground-based coordinate system.

This result seemed to be bad news for Newton's model of light as tiny particles emitted from a source. Consider a binary pulsar, that is, a neutron star orbiting a companion and emitting x rays, which are periodically eclipsed by the companion. If that radiation consisted of a stream of newtonian particles, then those particles would move faster when the pulsar was approaching us, and slower when it was receding. When the difference in (velocity)⁻¹ got multiplied by the distance to Earth, it would amount to a big change in arrival times. Some times we might even see a *double* image, because the pulsar would emit faster light, then move, then emit slower light from the new position and both would arrive simultaneously at Earth! No such phenomena are observed, so light can't be a newtonian particle.⁷

25.7 1905 AND ALL THAT

The preceding argument seems to favor a wave model of light over Newton's particle conception. But Chapter 27 will expose problems with the classical wave model as well. Chapter 30 will show how Einstein evaded both problems, paving the way for today's dual particle/wave picture of light. We'll see that Einstein's contribution was to say that

Electrodynamics, mechanics, and the rest of physics do hardwire in the P of R by using equations of motion that are invariant under a kind of boost transformations, but they're not quite the galilean transformations described above.

The correct invariance principle, and hence the correct equations of motion, were missed for centuries because, for mechanical objects moving relative to each other much more slowly than $3 \cdot 10^8$ m/s, the difference from galilean invariance is quantitatively small. For objects (or waves) that move at or near that large speed, however, the distinction becomes important.

⁷More realistically, the newtonian hypothesis predicts an irregularity in the apparent timing of the eclipse that was not observed [K. Brecher, PRL **39**:1051 (1977)]. Also, light emitted in the forward direction by a rapidly moving pion travels at c , not at $\approx 2c$ [T. Alvager et al., Physics Lett. **12**:260 (1964)].

T₂ Track 2

25.1'

Can a system be truly isolated? You could put it in a Faraday cage to screen out cosmic microwave background radiation. Then you wouldn't be able to detect the tiny anisotropy that arises because we are moving relative to the cosmic microwave background radiation (see Problem 29.4). In principle, there must be analogous *gravitational* radiation, which cannot be so screened, plus relic neutrinos etc., so a truly isolated system may be an unattainable idealization in practice. However, such radiation has not yet been observed experimentally.

PROBLEMS

25.1 *Thump*

Newton imagined light as a stream of tiny material particles obeying the same sort of laws as ordinary matter. Benjamin Franklin objected to this model; in 1752 he wrote in a letter “I must own I am much in the *dark* about light. . . . Must not the smallest particle conceivable, have with such a motion, a force exceeding that of a [cannonball]?” Suppose that a tiny particle, weighing just a picogram, could be brought up to the speed of light. Evaluate the newtonian kinetic energy formula, $\frac{1}{2}mv^2$, for this particle, and comment on Franklin’s assertion.

CHAPTER 26

Springs and Strings

We continue our little newtonian holiday. This really is a course on electrodynamics, but more generally it's a course about *where theories come from*. It's good to see abstract things first in a concrete setting.

In the preceding chapter, we started with a vague Principle (of Relativity), but then it turned into precise algebra and calculus (an invariance property). That's a very appealing progression, but in this chapter we'll see that we need to be a bit careful applying it. The payoff is that we'll get a framework that we can apply to field theories, including eventually relativistic ones including electrodynamics. Indeed, historically physicists' obsession with symmetry began with electrodynamics.

We'll also develop some framework relevant to other themes of this course, involving energy and momentum transport by waves.

26.1 EQUATION OF MOTION

Imagine a coil spring, initially straight and in its zero-tension state, with linear mass density $\rho_m^{(1D)}$ ($\sim \text{kg/m}$).

To analyze this system's motions, we temporarily break it down into finite elements with equilibrium separation Δx , each with mass $\Delta m = \rho_m^{(1D)} \Delta x$ and spring constant $\kappa/\Delta x$. Here κ is a material parameter describing the spring (the stretch modulus, with units of force).

Consider the mass element in Figure 26.1 whose equilibrium position is $x = 0$. Displace it in x by distance $u(0)$. The two springs flanking this element exert restoring forces on it. More generally, the element gets force $-(\kappa/\Delta x)(u(0) - u(-\Delta x))$ from its neighbor to the left, and $+(\kappa/\Delta x)(u(\Delta x) - u(0))$ from the right. For small enough

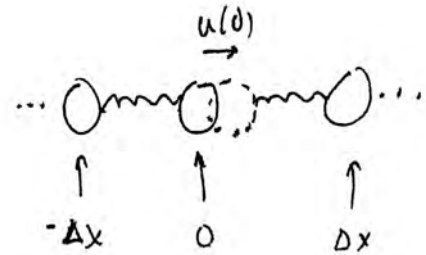


Figure 26.1: The mass initially at $x = 0$ has been displaced from equilibrium to the *dashed* position.

Δx we can apply a Taylor expansion to find

$$\Delta m \frac{\partial^2 u}{\partial t^2} \Big|_{x=0} = \frac{\kappa}{\Delta x} \frac{\partial^2 u}{\partial x^2} \Big|_{x=0} (\Delta x)^2 + \dots, \quad (26.1)$$

where the ellipsis denotes terms that are higher order in Δx . We now take the continuum limit $\Delta x \rightarrow 0$, or equivalently consider only distortions $u(x)$ that vary on length scales $\gg \Delta x$.

Any other mass element has the same dynamics, so u obeys the **wave equation**

$$\frac{\partial^2 u}{\partial t^2} - c_s^2 \frac{\partial^2 u}{\partial x^2} = 0 \quad \text{where } c_s^2 = \kappa / \rho_m^{(1D)}. \quad (26.2)$$

Because this is a partial differential equation, we call it a field theory in one space and one time dimension. It's a warmup for Maxwell's equations.

Solutions to the wave equation include the familiar harmonic ones moving at the “sound” speed c_s :

$$u_{\pm}(t, x) = \bar{u} \cos(\omega(\pm x/c_s - t)). \quad (26.3)$$

Here the angular frequency ω can have any value.

26.2 TRANSVERSE WAVES

You can repeat all the above analysis for disturbances in which a string under tension F_0 is plucked transverse to the x axis. This time, the displacement (height) $u(t, x)$ gives rise to a net transverse component of the tension proportional to $F_0(\partial u/\partial x)$, and so on. Again you get Equation 26.2 but with $c_s = \sqrt{F_0/\rho_m^{(1D)}}$.

26.3 INVARIANCE LOST

As in the preceding chapter, we will examine galilean invariance from both the active (Section 25.3, page 273) and passive (Section 25.4) viewpoints.

26.3.1 Active viewpoint

An active transformation replaces a spring configuration u by a different one, \tilde{u} . At any event Q , the new $\tilde{u}(Q)$ equals the value of u at a corresponding event P , related to Q via

$$\begin{bmatrix} t_Q \\ x_Q \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ v_* & 1 \end{bmatrix} \begin{bmatrix} t_P \\ x_P \end{bmatrix}.$$

Thus,

$$\tilde{u}_{\pm}(t_Q, x_Q) = \bar{u} \cos(\omega(1 \pm (v_*/c_s))(\pm \frac{x_Q}{c_s \pm v_*} - t_Q)). \quad (26.4)$$

Dropping the subscript Q , we see that the new functions don't belong to our original family of solutions (Equation 26.3)!

Your Turn 26A

Show that indeed, \tilde{u}_{\pm} moves at speed $\pm c_s + v_*$.

In contrast, every wave solution to Equation 26.2 moves at speed $\pm c_s$.

26.3.2 Passive viewpoint

Let's rederive the preceding result by focusing on the wave equation itself, rephrasing it in terms of the new variables

$$x' = x - v_*t, \quad t' = t. \quad [25.6, \text{ page } 276]$$

Thus, we define

$$u'(t', x') = u(t', x' + v_*t)$$

and find that u' obeys

$$\left[\left(\frac{\partial x'}{\partial x} \frac{\partial}{\partial x'} \right)^2 - c_s^{-2} \left(\frac{\partial x'}{\partial t} \frac{\partial}{\partial x'} + \frac{\partial t'}{\partial t} \frac{\partial}{\partial t'} \right)^2 \right] u' = 0.$$

Simplifying yields

$$\left[\frac{\partial^2}{\partial x'^2} - c_s^{-2} \left(-v_* \frac{\partial}{\partial x'} + \frac{\partial}{\partial t'} \right)^2 \right] u' = 0.$$

The original equation, Equation 26.2, expressed in the new variables, doesn't have the original form. Thus, the wave equation has neither active symmetry, nor passive invariance, under galilean boosts.

26.4 INVARIANCE REGAINED

In short, the wave equation is not galilean invariant. Is this a crisis in Physics? No, of course not—this is a newtonian system, and newtonian dynamics does have galilean invariance. The problem is that we have neglected a relevant dynamical variable: Before we plucked that string, it could have been in motion with respect to the observer, and hence with respect to any coordinate system in which the observer appears to be at rest. We did not account for this possibility.

That is, the wave equation that we wrote down is *incomplete*: It only applies to the *special case* where the initial state of the string is at rest with respect to the coordinate system. If that situation holds for the coordinate system t, x , then it *won't* hold for the boosted t', x' coordinates, so we *shouldn't* (and didn't) find the same form for the equation of motion.

Let's start over and formulate a more general situation, a spring initially in uniform motion at arbitrary speed v_m (the medium's speed) and again subject to transverse displacement. Let $u(t, x)$ be the displacement of whichever spring segment is located at x at time t . Note that observing a fixed coordinate position x_0 at two different times is *not* the same as following one particular spring segment.

Consider the spring segment that is located at x_0 at time $t_0 = 0$. Imagine painting that one segment red and applying Newton's Second Law to it. At later time Δt , the red segment has moved to $x = x_0 + v_m \Delta t$. Hence, its transverse velocity $v_y(t_0, x_0)$ is the limit of

$$v_y(t_0, x_0) = \frac{1}{\Delta t} [u(t_0 + \Delta t, x_0 + v_m \Delta t) - u(t_0, x_0)] = \left(\frac{\partial}{\partial t} + v_m \frac{\partial}{\partial x} \right) u \Big|_{t_0, x_0}.$$

The net transverse force on this segment is still $F_0\Delta x(\partial^2 u/\partial x^2)$ as before, so during time Δt its transverse momentum p_y changes by $F_0\Delta x(\partial^2 u/\partial x^2)\Delta t$. That is,

$$p_y(t_0 + \Delta t, x_0 + v_m\Delta t) - p_y(t_0, x_0) = F_0\Delta x(\partial^2 u/\partial x^2)\Delta t, \quad \text{or}$$

$$\left[\left(\frac{\partial}{\partial t} + v_m \frac{\partial}{\partial x} \right)^2 - c_s^2 \frac{\partial^2}{\partial x^2} \right] u = 0. \quad (26.5)$$

We just found the generalized wave equation for a spring that is moving w.r.t. the coordinate system at speed v_m . When $v_m = 0$, it reduces to the familiar form Equation 26.1.

Your Turn 26B

Show that:

- Equation 26.5 has traveling wave solutions, but they move at speed $v_m \pm c_s$. (Indeed, if a distant bell is rung you'll hear it slightly sooner if there is a wind blowing toward you than you would in still air.)
- A traveling-wave solution to this equation, viewed in a boosted coordinate frame, belongs to the same family of solutions (though with a different v'_m).
- In particular, an observer who flies alongside the spring at speed $v_m = c_s$ will see some waves that appear *static* (find one).
- Equation 26.5 is invariant under galilean transformations, once we understand that both $u(t, x)$ and v_m must transform.

Thus, galilean transformations *really are* invariances of the spring system, once we include all relevant dynamical variables and attribute appropriate transformations to them. That is, our error in Section 26.3 lay in mistakenly setting the *scope* of the system too narrow (treating v_m as a fixed constant of the system, rather than as a dynamical variable subject to transformation).

26.5 CONNECTION TO ELECTROMAGNETISM

Chapter 17 showed that Maxwell's equations imply the wave equation, and Section 26.3 showed that the wave equation lacks galilean invariance. Everyone already knew this prior to 1905. Everyone assumed that the cure would be along the lines described in Section 26.4: "Maxwell's equations are incomplete, valid only in the special case of a coordinate frame at rest with respect to the æther. After we generalize them to account for æther wind, then their full galilean invariance will appear." One thing that bothered Einstein was that, despite great efforts, nobody had succeeded in finding the right generalization that was mathematically consistent and also consistent with experiments. We'll see soon where he went with that line of thought, but first we pause to think about the transport of energy and momentum in the familiar setting of springs.

26.6 ENERGY AND MOMENTUM

26.6.1 Continuity equations

For future use, let's see how energy and momentum are locally conserved in the vibrating string. In this section, we will choose a spacetime coordinate system in which the string is at rest ($v_m = 0$). We also continue to look at transverse waves.

We now seek continuity equations for energy and momentum, analogous to the one we found for *charge* (Section 7.2):

$$\text{kinetic energy} = \int dx \frac{1}{2} \rho_m^{(1D)} (\partial u / \partial t)^2; \quad \text{potential energy} = \text{const.} + F_0 \int dx \frac{1}{2} (\partial u / \partial x)^2. \quad (26.6)$$

One way to get the second formula is to imagine that an external agent is pulling the string horizontally with force F_0 . When curved, the string's end-to-end distance shortens by a distance $L_{\text{tot}} - \int_0^{L_{\text{tot}}} (dx / \cos \theta(x))$, where θ is the angle relative to straight.¹ Shortening does work against whatever external mechanism is supplying the tension force. Making small-angle approximations gives the work done against the outside force when the string is curved, per unit length, as $\frac{1}{2} F_0 (\partial u / \partial x)^2$.

Thus, in the continuum limit the total linear density of energy (J/m) at t, x is

$$\rho_{\mathcal{E}}^{(1D)}(t, x) = \frac{1}{2} \rho_m^{(1D)} ((\partial u / \partial t)^2 + c_s^2 (\partial u / \partial x)^2).$$

If you pluck just one mass, you'll create some localized energy, which then spreads. That energy cannot just vanish somewhere and pop up elsewhere! Instead, energy *flows* with a 1-dimensional flux $j_{\mathcal{E}}^{(1D)}$ (units J/s). To find that flux, note that the rate at which energy gets transported from any mass element to the one at its right is the rate at which work is done on the right side by the left side. This is the product of velocity (which is transverse) times the transverse component of force: $j_{\mathcal{E}}^{(1D)} = -F_0 (\partial u / \partial x) (\partial u / \partial t)$.

Your Turn 26C

Prove that for any solution of the wave equation,

$$\frac{\partial \rho_{\mathcal{E}}^{(1D)}}{\partial t} + \frac{\partial j_{\mathcal{E}}^{(1D)}}{\partial x} = 0. \quad \text{continuity equation for energy}$$

Similarly to the continuity equation for charge (Chapter 7), your result expresses the fact that energy is a locally conserved quantity: In order to change energy density at a point (first term on the left side), there must be an imbalance in the fluxes on either side of that point (second term on the left side).

Your Turn 26D

Now repeat the analysis to find the density and flux of transverse *momentum* and prove an appropriate continuity equation relating them.

¹We are assuming an inextensible string, so its contour length does not change.

26.6.2 The case of harmonic waves

For the solutions given in Equation 26.3, the energy density is

$$\rho_{\mathcal{E}}^{(1D)} = \frac{1}{2}\rho_m^{(1D)}\bar{u}^2(\omega^2 + c_s^2(\omega/c_s)^2)\sin^2(\omega t - (\omega/c_s)x). \quad (26.7)$$

Thus, the kinetic and potential energy terms are in phase. They're both nonnegative, but both drop to zero twice per cycle, at $t = \omega x/c_s + n\pi$ for integer n . At these “dead spots,” even the energy flux is zero, because

$$j_{\mathcal{E}}^{(1D)} = -\rho_m^{(1D)}c_s^2\left(-\frac{\omega}{c_s}\right)\omega\bar{u}^2\sin^2(\omega t - (\omega/c_s)x)$$

falls to zero at the same places as Equation 26.7. How can energy flow to the right if there are spots where its flux is zero? Well, first of all the flux is nowhere *negative*. And at a node where energy density is zero, the *gradient* of flux is nonzero. The continuity equation says that energy arriving from the left of that point begins to pile up there. So that point stops being a point of zero energy density, and so on.

26.7 PLUS ULTRA

The preceding section started with expressions for energy density and flux that were nearly obvious, then showed that they obey a continuity relation. Later we will wish to understand the energy density and flux of electromagnetic fields, which are *not* so obvious in form. To find the right expressions, we'll work backward, guided by the expectation that these quantities, too, must obey continuity relations.

PROBLEMS

26.1 *Slinky*

Consider a stretched spring of mass density $\rho_m^{(1D)}$ ($\sim \text{kg/m}$) and spring constant κ ($\sim \text{N}$). Rederive the results of Section 26.6 for the case of longitudinal (compression) waves.

CHAPTER 27

Einstein's Version of Relativity: Overview

Failure to appreciate the role of the structure of Indo-European languages in affecting perception has repeatedly led western science into error. The “luminiferous æther” of classical physics was created for the express purpose of standing as a subject of the verb “to wave.”

— Garrett Hardin

Here is an overview of what we're going to cover, stated without any equations or diagrams. Some of these ideas won't be clear, however, until embodied in equations and diagrams. That comes later.

27.1 THE ÆTHER HYPOTHESIS

The Principle of Relativity seems experimentally valid for any system that can be isolated from the rest of the world. Newtonian physics has an overarching mathematical property (galilean invariance) that transcends details of particular springs, clocks, etc. and that guarantees that any system will obey the P of R. One way to expose that property is to see how the equations change their form when expressed in a different coordinate system on spacetime, identify a subfamily of systems among which the form does *not* change, and observe that some of those good systems are in uniform, straight-line motion relative to the others.

Maxwell found some equations that agreed with experiments in a particular (lab) coordinate system, and were mathematically self consistent. The equations predicted a new phenomenon (EM radiation), and Hertz firmed up the evidence that this radiation was the same as light.

Neither Maxwell nor anyone else at that time believed that the equations were fully general: At best, they were regarded as correct *in a coordinate system at rest relative to an omnipresent medium* called the **luminiferous æther**. People believed this because of a general sense that waves could only move through a medium. (How do you have ripples, without the pond?) Tacitly the words “material medium” implied a substance that itself had states of motion, like air, water, or a string. Obviously the state of motion of the medium would have to enter the fully general equations of electromagnetism, as it does for the equations of sound, water waves, or string vibrations (Section 26.4).

But the æther had to have some pretty weird properties. It had to be completely unaffected by any vacuum pump ever invented, because light travels just fine through vacuum. It had to be present throughout the space between planets, yet exert no frictional drag on them. It must be rigid, like steel and unlike air, because it supports

transverse waves. It had to be incompressible, because if not, there would also be a longitudinal polarization of light, as there is for waves in air or steel (compression waves). Yet the planets move through it effortlessly.

Another problem with the æther was that it did *no other job* than the one for which it was introduced (transmitting light). In contrast, air transmits sound, but it also has other measurable attributes giving rise to other phenomena, for example, its mass density, temperature, pressure, viscosity, and so on; moreover, these attributes can be *changed* by experimental interventions.

Why were people so desperate to cling to this crazy idea? We can look back and say, a bit more clearly than was said at the time, that people also expected that all laws of Nature must be form-invariant under rotations, translations, *and galilean boosts*. Maxwell's equations as stated do not have the last of these properties, but it was assumed that after generalizing them to include the possible motion of the æther, they would.

Einstein found too many logical problems with this position, not least his and others' inability to find a self-consistent set of equations as candidates to generalize Maxwell's.¹ Even setting aside this formal objection, modifying the wave equation to account for æther motion did not produce any theory consistent with all experiments. For example:

- When an object moves through an incompressible fluid, it sets the fluid into motion. Lab-based experiments could not detect any consequences of such motion.
- And if the Earth itself dragged along the æther, then the observed “aberration of starlight” wouldn't happen (Chapter 29).
- But if somehow Earth didn't drag the æther, then there would be an “æther wind,” and the Michelson–Morley and Fizeau experiments wouldn't have given the results that they did (Chapter 28).

27.2 THE NO-ÆTHER HYPOTHESIS

So Einstein entertained the bizarre suggestion that Maxwell's equations were actually *correct and complete as written*.

- To the objection that they lacked galilean invariance, he said, perhaps experiments don't demand such invariance after all; perhaps the equations have some *other* invariance. Perhaps the different invariance that Lorentz had already found was in fact exact and good enough to satisfy the demands of experiment, for example, the Principle of Relativity.
- To the objection that replacing galilean invariance with Lorentz invariance had bizarre consequences, Einstein asked, are those consequences actually ruled out by experiment? For example, is there really any feasible *method* to measure absolute

¹Although the wave equations for sound and light are formally similar, they have quite different origins. If you propose a modification to the electromagnetic wave equation, you can't stop there: You must also propose an acceptable modification to the full set of Maxwell's equations that gives rise to your proposed new wave equation and agrees with experiments. This is what Einstein and others could not do.

simultaneity? If not, then it's not so disturbing if theory predicts that different observers disagree about the simultaneity of two events not located at the same point in space.

- To the objection that Newton's laws are incompatible with Lorentz invariance, Einstein said, maybe we need to reexamine the experimental status of Newton's laws.

I did not say above that "The æther does not exist." It is not really very scientific to claim the nonexistence of a poorly defined thing. Indeed, one sometimes hears somebody smugly pronounce that the quantum vacuum "is" the æther. Einstein would not object. His proposal merely amounts to saying that the vacuum—the state you can approach experimentally by using better and better vacuum pumps, or by going into interstellar space—is *unique*. Its properties (like the values of μ_0 and ϵ_0) are *constants*. It has no further state variables beyond \vec{E} and \vec{B} that need to appear in Maxwell's equations, and in particular no states of motion. (More precisely, it is Lorentz-invariant.) If you want to say it's filled with an "æther" of virtual particles and antiparticles, fine, but it's not the material substance that Maxwell and his contemporaries had in mind.

In other words, Einstein convinced himself that there's no logical *need* for the æther. Maxwell equations don't need it. It's only our brains, trying to make inappropriate analogies to experience, that *want* it. We can't intuitively imagine the EM field, nor the vacuum which it disturbs. The birth of modern physics came when Einstein said, "That's OK—I don't *need* to imagine it intuitively."

Einstein also thought about an observer who flies over a water surface at the speed of wave propagation. Looking down, that observer sees waves that appear to be *standing still* (Your Turn 26B). But there are no static wave solutions to the Maxwell equations, nor did Einstein see any way to modify the equations to admit such solutions. Instead of attempting any such modification, Einstein's clarified a mathematical property (a new invariance) *already hiding* in Maxwell's equations.² Then he proposed that all the rest of physics had this same invariance, for example, the mechanisms inside clocks. All his "thought experiments" were mainly attempts to see if his proposal was obviously ruled out by existing knowledge. Over and over, he found that potential objections (paradoxes) were based on assuming some procedure that could not in fact be implemented experimentally (for example, knowing the reading on a distant clock instantaneously).

Then Einstein asked if his proposal made any characteristic, quantitative predictions that were testable. We'll never know how much he really knew about Michelson–Morley; what he explicitly stated was that he relied on the aberration of starlight, and the Fizeau experiment, as sufficient to show he was on the right track. Not coincidentally, both of these concerned . . . electromagnetic phenomena. So we'll discuss them in detail in the following chapters.

²Lorentz had already established this in 1904, but even today it is hard to grasp that from what he wrote. By the way, Einstein actually called his ideas the "theory of invariants." The phrase "theory of relativity" was coined by somebody else, and Einstein only adopted it reluctantly some time later.

27.3 WHERE WE ARE HEADING

Anyone can open Einstein's 1905 paper, copy out the transformations of the fields (updating the awful notation), substitute into Maxwell's equations, and show they are indeed an exact invariance. Afterwards, we are still stumped—how could any human have figured that out?³ Instead we will take a longer route, following Minkowski and others: We will build a system of thought and notation in which the invariance of Maxwell's equations, and other relativistic field theories, becomes *obvious at a glance*. That way, even mortals like us can create *new* relativistic field theories, for example the ones needed to describe the strong and weak nuclear forces.⁴

³The whole thing reminds me of public-key encryption.

⁴It still required some more of Einstein's personal genius to adapt the ideas to gravitation. And even Einstein needed the benefits of tensor notation before he could succeed.

T₂ Track 2

27.1'

The main text asserted that the vacuum has no user-adjustable properties. Like any bedrock postulate in science, this one is more subtle, and more subject to fine interpretation, than it looks.

The empty space outside the pole of a magnet in vacuum has a “property,” the static magnetic field, because that property is attached to specific points in space in that region. Physics in that region of space is not isotropic and hence not Lorentz-invariant. So it’s more precise to say that only a region of vacuum that is far from or shielded from any matter is universal, including its ability to *carry* EM fields (or rocks etc), should they be introduced. When charged matter is present, we attribute their effects to a *deviation* from field-free vacuum (the EM field) whose dynamics is invariant under a group of transformations, and so on with other kinds of interaction (strong, electroweak). It’s always the same group, a circumstance whose universality we attribute to the field-free vacuum itself being invariant.

Remarkably, Einstein abandoned even this more limited statement a few years later when he formulated general relativity. He found that it proved fruitful to attribute *gravitation* directly to . . . user-modifiable properties of spacetime. Moreover, there is no such thing as “shielding” a region from gravitational fields and no region in space “far enough” from gravitation to be unaffected by it; indeed, the whole expansion of the Universe is controlled by gravitation.

Remarkably, however, the statements made in this chapter are still accepted today, in the following sense. Far enough from any gravitating bodies, Einstein’s general theory predicts the existence of special coordinate systems (“locally inertial” or “freely falling”), in which gravitational effects are approximately absent and all the *rest* of physics, including electrodynamics, has the properties discussed in this chapter. For example, the locally inertial systems are all related to each other by ordinary Lorentz transformations; those transformations are invariances of the all non-gravitational dynamics; and so on. Section 33.10 will return to this train of thought. Ultimately it led to a combined theory of gravitation and other interactions that, although still not integrated fully with quantum mechanics, nevertheless has been successfully extrapolated to make predictions about physics even close to massive objects.

T₂ Track 2

27.2' Poincaré's work

“History has not been kind to [Poincaré]’s contributions. In his *Science and Hypothesis*, first published in 1902, Poincaré boldly declares:

1. There is no absolute space, and we only conceive of relative motion; and yet in most cases mechanical facts are enunciated as if there is an absolute space to which they can be referred.
2. There is no absolute time. When we say that two periods are equal, the statement has no meaning and can only acquire a meaning by a convention.
3. Not only have we no direct intuition of the equality of two periods, but we have not even direct intuition of the simultaneity of two events occurring in two different

places.

4. Finally, is not our Euclidean geometry in itself only a kind of convention of language?

These ideas are at the heart of relativity, and it is difficult to believe they did not have a profound effect upon Einstein's thinking. Poincaré was also the first to use the term 'principle of relativity,' which is also stated forthrightly in *Science and Hypothesis*. In a famous 1904 speech at the International Congress of Arts and Sciences in St. Louis, Poincaré even glimpses a new theory in which 'the velocity of light becomes an impassable limit.' But the mathematician did more than make oracular pronouncements; he wrote a pair of technical papers on Lorentz's theory, and in the longer one, completed just before Einstein's own, he has nearly everything his shadowy rival does, and in some respects more. In that paper, Poincaré shows, as Lorentz did, that Maxwell's equations are invariant if the Lorentz transformation is correct; he anticipates Minkowski's combining of space and time, and he virtually derives $E = mc^2$. What Einstein did in those fateful weeks that Poincaré did not was to show that the whole thing results from just the two postulates: the principle of relativity and the constancy of the speed of light." – Rothman, 2003, p73

CHAPTER 28

Provisional Lorentz Transformations and the Fizeau Experiment

Oh, that Einstein, always cutting lectures—I really would not have believed him capable of it.

— *Einstein's former teacher Minkowski, upon reading the relativity paper.*

We've seen that the wave theory of light has scored some successes, giving a detailed account of polarization phenomena (Chapter 17), the transport of energy and momentum (Chapter 18), and so on. But there is still a puzzle, which eventually led Einstein to some disturbing insights into space and time.

28.1 REVIEW

28.1.1 Galilean

Chapter 25 argued that newtonian physics implements the Principle of Relativity by having an invariance under galilean transformations. One way to express this is to say that if we have a system of particles, and a solution to the equations of motion given by some functions $\vec{r}_{(1)}(t), \vec{r}_{(2)}(t), \dots$, then the modified functions

$$\tilde{\vec{r}}_{(1)}(t) = \vec{r}_{(1)}(t) + \vec{v}_* t, \quad \tilde{\vec{r}}_{(2)}(t) = \vec{r}_{(2)}(t) + \vec{v}_* t, \dots$$

will also solve the same equations. Here \vec{v}_* is one overall constant vector.

Equivalently, we can think of relabeling all the events in spacetime according to

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -v_*/c & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix}, \quad [25.7, \text{page 276}]$$

or a similar formula in three spatial dimensions. Chapter 25 showed that if we take some equations of newtonian physics (two masses joined by a spring¹) and reexpress them in terms of the primed coordinates, the new versions have the same algebraic form as the old ones. Section 26.3 also showed that the wave equation does *not* have this property, but Section 26.4 gave a resolution of that puzzle appropriate for vibrating strings, sound waves, and water waves: The wave equation must be generalized to account for possible motion of the medium relative to the observer.

Finally, Section 25.6.4 found a velocity addition formula, which can be stated in

¹Or two masses with their newtonian gravitational attraction, etc.

a rather longwinded way as

If we have two coordinate systems on spacetime, in each of which the newtonian equations of motion take their usual simple form, then those systems differ by a galilean transformation. If that transformation is just a boost by velocity \vec{v}_ , and a wave or particle in one system is moving at constant velocity \vec{v}_0 , then in the other system it will be observed to be moving at constant velocity $\vec{v}_0 - \vec{v}_*$.* (28.1)

It's necessary to be so pedantic, because as we'll see, the post-1905 version of this claim is weirdly different from the version just stated.

28.1.2 The problem in a nutshell

Chapter 25 noted that light from distant objects was known to come to us at a velocity independent of the source's motion. The same can be said of sound or water waves: Imagine running your finger just above the surface of a ripple tank and periodically dipping it into the water. Each ripple you cause moves outward at a fixed speed independent of how fast your finger is moving. That is, as long as the observer is at rest relative to the medium, waves in a material substance move at a constant speed independent of the motion of the source. Thus, the wave model of light seemed to explain why the a binary pulsar never appears doubled (Section 25.6.6).

But the apparent speed of a wave on water or air certainly *does* depend on the motion of the *observer*. In contrast, the speed of light *also* was found to be unchanged by uniform, straight-line observer motion. After all, the Earth is hurtling through space, yet the physics we see in a closed lab does not depend on orientation relative to that motion.² This looks bad for the wave model of light. Einstein was alluding to this problem when he mentioned the prior “failure of attempts to detect a motion of the earth relative to the ‘light medium’.”³

Let's pause to dispose of a red herring. Certainly there are bizarre coordinate systems we could choose in which a particular ray of light seems to move at a speed other than c . Simply take $\vec{r}' = 2\vec{r}$, and leave time unchanged; in the primed coordinate system, light travels at speed $2c$.⁴ What makes this observation physically irrelevant is that in the primed system, the equations of physics take nonstandard forms. We would know right away that something was wrong in the new system, for example, because atoms would have different apparent sizes than in our usual coordinate system. *Even the good coordinate systems* (those in which the equations take their usual form) will disagree about wave speed if there is a material medium, but no such effect is observed for flashes of light.

It seemed to everyone that the observation in the preceding section was not a serious problem. “Obviously” the solution was that Maxwell's equations were incomplete, just like the wave equation we found for a vibrating string: They were only valid in those inertial coordinate systems that were at rest with respect to the “light medium”

²The restriction to “closed lab” is needed because we *do* see effects of our motion when we look at distant stars (Chapter 29).

³Einstein was never clear whether he was thinking specifically about the Michelson-Morley experiment, but there were many such experiments at the time that all came out null.

⁴Similarly, applying a galilean boost changes the apparent speed of light.

(or “æther”). But Einstein also noted that previous attempts to modify Maxwell’s equations in the way we did for the vibrating string had been unsuccessful.

It’s not enough just to say blithely, “So there’s no æther.” After all,

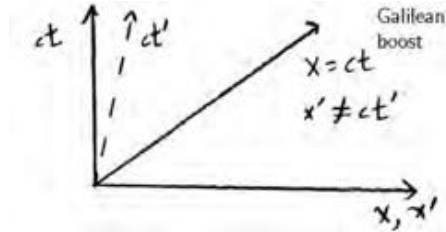
- Eliminating the medium would also eliminate our rescue of galilean invariance (Section 26.4).
- Galilean invariance is what guaranteed the Principle of Relativity, which is experimentally validated.

In this chapter and the next, we’ll see how Einstein reconciled Maxwell with the P of R at the level of a single (scalar) wave equation, temporarily neglecting all the delicious complexity brought by the vector character of electromagnetic fields.⁵ As always, we’ll look to some key experiments for guidance.

28.2 GRAPHICAL EXPLORATIONS

If we suppose the Maxwell equations to be complete and correct as written, then what invariances *do* they have? Maybe they have some *non*-galilean invariance that nevertheless connects coordinate systems that (i) are in uniform, straight-line motion relative to each other, yet (ii) also agree on the physical prediction that the speed of light should be the constant $c \approx 3 \cdot 10^8$ m/s determined by electrostatics (ϵ_0) and magnetostatics (μ_0). That sounds like a contradiction, but in Einstein’s words these two requirements are “only seemingly incompatible.” In fact, W. Voigt proved that the scalar wave equation was invariant under what we now call “Lorentz” transformations in 1887, just 14 years after Maxwell. Lorentz and Einstein took this result seriously, and crucially, they extended it from the scalar wave equation to the full Maxwell equations and then to all of Physics.

We can think graphically about the galilean transformation (Equation 25.7) as introducing a new set of coordinate axes on the xt plane. Actually, it’s easier to think about the quantities x and ct , because these have the same units, and because then a trajectory traveling at speed c is a line at 45° to the horizontal. Such a trajectory is drawn as a solid diagonal line in the figure below:



The original x and ct axes are also shown as solid lines. The new x' axis is the same as the x axis, because it’s the locus of events with $t' = 0$ and $t' = t$. However, the new ct' axis is bent over (dashed line). We see graphically that the trajectory shown bisects the right angle between x and ct axes, but *doesn’t* bisect the acute angle between x' and ct' axes: It changes the apparent speed of light.

⁵This oversimplification will be remedied in Chapters 31–32.

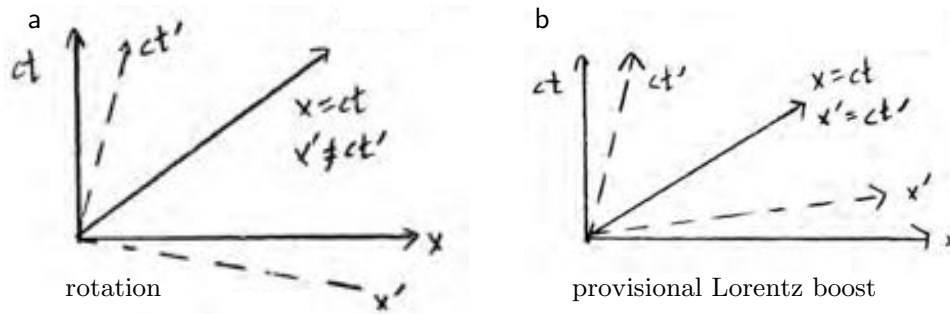


Figure 28.1:

We have experience with another sort of linear transformation in the plane: a *rotation* of the axes. Figure 28.1a shows this option. But this transformation, too, alters the apparent slope of the trajectory shown; again, the trajectory does not bisect the angle between x' and ct' axes. But there is another possibility (panel (b) of the figure): If we bend *both* axes by *opposite* angles, then the diagonal line continues to bisect the angle between x' and ct' axes.

Your Turn 28A

Think about the other allowed light trajectory in 1D, which moves at speed $-c$. It bisects the angle between the $-x$ and ct axes. Convince yourself that it also bisects the angle between the $-x'$ and ct' axes.

28.3 THE WAVE EQUATION IS INVARIANT UNDER PROVISIONAL LORENTZ TRANSFORMATIONS

Figure 28.1 represents the following linear transformation of coordinates:

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix}. \quad \text{provisional Lorentz boost transformation} \quad (28.2)$$

Here β and γ are constants and $\gamma > 0$. Equation 28.2 says “provisional” because, although all transformations of this form are invariances of the vacuum wave equation, we’ll see that not all are invariances of the *rest* of physics (nor even of the full Maxwell equations).⁶

Equation 28.2 has a feature that bothered many people: $t' \neq t$. To many, it seemed necessary that all good coordinate systems would agree on *one correct, universal choice for time*. To Einstein, that necessity was not obvious.

⁶Chapter 29 will argue that the true Lorentz transformations are the special case with $\gamma = (1 - \beta^2)^{-1/2}$, but we don’t need that fact yet.

28.3.1 Active viewpoint

To see if Equation 28.2 is at least promising, consider a harmonic wave solution to the wave equation: $\phi_{\pm}(t, x) = \cos(\frac{\omega}{c}(-ct \pm x))$. Following Section 26.3.1 (page 283) we apply an active transformation, that is, construct different functions $\tilde{\phi}_{\pm}$ defined by

$$\tilde{\phi}_{\pm}(t, x) = \cos\left(\frac{\omega}{c}(-\gamma(ct - \beta x) \pm \gamma(-\beta ct + x))\right),$$

which can be written as

$$\cos\left(\frac{\tilde{\omega}}{c}(-ct \pm x)\right), \quad \text{where } \tilde{\omega} = \omega\gamma(1 \pm \beta).$$

These functions are again solutions to the wave equation, traveling at speed $\pm c$! It's true that each has a different frequency from the original, but we expected that—there should be a Doppler shift. Because any solution to the wave equation can be expanded in Fourier series, we have established symmetry under active transformations.

28.3.2 Passive viewpoint

Encouraged by that result on a particular solution, we now switch to the passive viewpoint, that is, we focus on the wave equation itself, not its solutions:

Your Turn 28B

Reexpress the wave equation in terms of primed coordinates. (Follow the same approach as in Chapter 26, but with the new transforms Equation 28.2 instead of galilean boosts.) Show that the wave equation maintains its original form after this passive transformation.

Indeed, the wave equation is invariant under a family of transformations that take a coordinate system and boost it into uniform straight-line motion relative to the original one. Hence, the wave equation is still compatible with the Principle of Relativity—just not in the way people had expected.

28.4 VELOCITY ADDITION

Let's apply the Relativity Strategy (Idea 25.12, page 278) to a wave or particle that in one good coordinate system appears to be in uniform, straight-line motion with velocity $\pm v$. We can describe that motion parametrically by giving ct and x as functions of a parameter ξ , then see what the motion looks like in the equivalent boosted coordinate system:

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} \xi \\ \pm v\xi/c \end{bmatrix} = \begin{bmatrix} \gamma\xi(1 \mp v\beta/c) \\ \gamma\xi(-\beta \pm v/c) \end{bmatrix}. \quad (28.3)$$

Thus, in the primed coordinate system the trajectory again has uniform velocity, given by

$$v' = c \frac{\gamma\xi(-\beta \pm v/c)}{\gamma\xi(1 \mp v\beta/c)} = c \frac{-\beta \pm v/c}{1 \mp v\beta/c}. \quad (28.4)$$

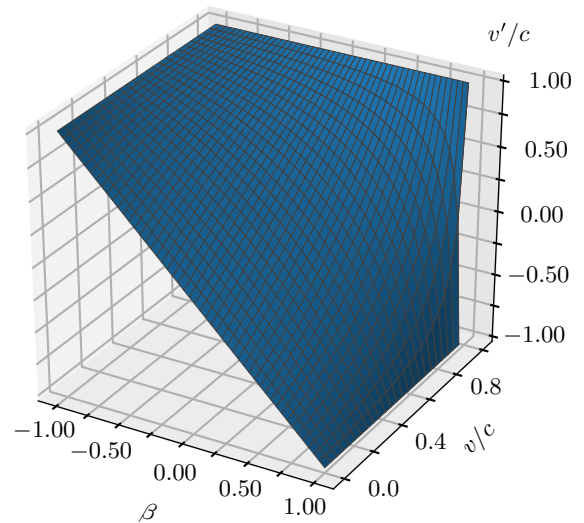


Figure 28.2: The velocity addition formula (Equation 28.4) never gives a result larger than c .

That's pretty weird. It sure doesn't look like the galilean formula $v' = -\beta c \pm v$. But suppose that $\beta \ll 1$ and $v \ll c$; in this case we can forget the denominator, and we do recover galilean behavior. *In the everyday world of things moving much more slowly than light, Einstein's kinematics resemble the galilean behavior.* This is the world in which we got our intuitions over millions of years of evolution: Throw a spear while running forward, and the spear's velocity will be the sum of your arm velocity and how fast you're running (better able to bring down that gazelle).

In the opposite, less familiar, regime where $v \rightarrow \pm c$, our formula boils down to $v' \rightarrow \pm c$. *A trajectory that moves at speed c in (t, x) has the same property in (t', x') .* As we saw with pictures in Section 28.2, Lorentz invariance reconciles our desire to connect coordinate systems in relative motion (and hence hardwire the P of R), with the universality of the speed of light required by the Maxwell equations.⁷

Finally, you should think about the limit $\beta \rightarrow 1$, holding v fixed to some value less than c . Figure 28.2 shows this and every other case graphically.

Because every provisional Lorentz transformation preserves the form of the wave equation, the combined effect of two such transformations in succession will have the same property.

Your Turn 28C

- Suppose that a boost with (γ_1, β_1) is followed by another with (γ_2, β_2) . Show that the combined transformation is again of the form Equation 28.2.
- Find the inverse of the transformation Equation 28.2 and show that it, too, is a provisional Lorentz transformation.

Thus, our provisional Lorentz transformations form a group, analogous to but distinct

⁷This observation eliminates an objection we made to the particle picture of light in Chapter 25: Regardless of how a pulsar may be moving relative to us, light leaving it always travels toward us at speed c . Although this course will focus on the wave picture, the fact that both viewpoints are experimentally tenable underpins the dual nature of light revealed in quantum field theory.

from the galilean group (Section 25.6.4). Just as in newtonian physics, we can promote everything to three space dimensions, again obtaining a group of invariances. Later, when we finish specifying the relation between γ and β in Section 29.3, this group will be called the **Lorentz group**.

28.5 A NONNULL, FALSIFIABLE PREDICTION

Is this just philosophy? We should think about some real experiment. Later in his life, Einstein said that *just two* experimental observations were all he needed to be convinced he was on the right track. They were the aberration of starlight and an experiment first done by Fizeau.⁸ We'll discuss the second of these now, and the first in Chapter 29.

You were probably taught that the famous Michelson–Morley experiment falsified the æther/galilean hypothesis. Einstein was never quite clear on whether he knew and was influenced by that experiment, but in any case, one problem with it is that it was a *null* result; the result was *zero* dependence of light speed on apparatus velocity, whereas the æther/galilean theory predicted a nonzero result. Null experiments are subject to the criticism that zero is a very special value. There may be various explanations for why you got zero (maybe your sensitivity wasn't as good as you thought).

It's more convincing when two theories make two quantitative, different, *nonzero* predictions for an experimentally observable quantity, and an experiment excludes one but not the other. Fizeau's experiment had that character. Before doing it, Fizeau first measured the speed of light in air, finding near-agreement with Rømer's old astronomical measurement.⁹ Then he measured the speed of light *in water*, finding it to be c/n where $n \approx 4/3$. That was a comforting result: Huygens had shown that a slowdown of light in water was just what was needed to explain the law of refraction in the wave theory of light. But crucially, Fizeau proceeded to study the propagation of light in *flowing water* at various velocities, both along and against the direction of a light beam.¹⁰ Unlike in vacuum, he found that the motion of the water can slow down or speed up the light, depending on its motion.

Let's apply the Relativity Strategy to this problem (Idea 25.12, page 278):

- We set up a “good” coordinate system in our lab, that is, one in which the laws of Nature have their simplest form. Then we measure the speed of light in still water. Call the result $v = \pm c/n$, where c is the vacuum speed and n is a constant larger than 1. Thus, the trajectory of a flash of light can be written in parametric form as

$$\begin{bmatrix} ct \\ x \end{bmatrix} = \begin{bmatrix} \xi \\ \pm \xi/n \end{bmatrix}. \quad \text{first light flash, lab frame}$$

⁸Fizeau's experiment was first done in 1859 (the version Einstein knew), then redone with greater precision by Michelson and Morley in 1896 (a little-known result distinct from the famous MM experiment). A version with still higher precision was then done by Zeeman. Figure 28.3 below shows MM's data. Many more tests of relativity came only after 1905, so were not available to Einstein.

⁹Light travels a tiny bit slower in air than in interplanetary space.

¹⁰Today one uses a chunk of quartz on the rim of a rapidly spinning disk, to eliminate turbulence that occurs in water.

Also, the trajectory of an element of water can trivially be written as $\begin{bmatrix} ct \\ x \end{bmatrix} = \begin{bmatrix} \xi \\ \text{const} \end{bmatrix}$.

- Both the hypothesis of galilean invariance, and the hypothesis of provisional Lorentz invariance, then assert that there will be other coordinate systems, moving relative to the lab system, in which the relevant laws of Nature have the same form. The same equations must have the same solutions, so we predict that in the new frame there will be a *different*, but allowed, light trajectory described by the same function of ξ :

$$\begin{bmatrix} c\tilde{t}' \\ \tilde{x}' \end{bmatrix} = \begin{bmatrix} \xi \\ \pm\xi/n \end{bmatrix}. \quad \text{new light flash, boosted frame}$$

Also, for this new solution the trajectory of an element of water is

$$\begin{bmatrix} c\tilde{t}' \\ \tilde{x}' \end{bmatrix} = \begin{bmatrix} \xi \\ \text{const} \end{bmatrix}. \quad \text{water, boosted frame}$$

- If the equations of electrodynamics were invariant under galilean transformations, we could then obtain the speed of the light flash in flowing water, as measured in the lab coordinate system, by performing a galilean boost back to the lab frame, obtaining $\tilde{v} = v_* \pm c/n$.¹¹
- If, on the contrary, the equations of electrodynamics are invariant under provisional Lorentz transformations, then we must instead use the inverse of Equation 28.2. Applying this procedure to a water element gives that the water is moving at speed v_* . Of greater interest, the light flash's trajectory

$$\begin{aligned} \begin{bmatrix} c\tilde{t} \\ \tilde{x} \end{bmatrix} &= \frac{1}{\gamma(1 - (v_*/c)^2)} \begin{bmatrix} 1 & v_*/c \\ v_*/c & 1 \end{bmatrix} \begin{bmatrix} \xi \\ \pm\xi/n \end{bmatrix} \\ &= \frac{1}{\gamma(1 - (v_*/c)^2)} \begin{bmatrix} \xi(1 \pm \omega/(cn)) \\ \xi(v_*/c \pm 1/n) \end{bmatrix}. \quad \text{new light flash, lab frame} \end{aligned}$$

The velocity is then $d\tilde{x}/d\tilde{t} = c(v_*/c \pm 1/n)/(1 \pm v_*/(cn))$.¹²

In experiments, we can never get the water flowing anywhere near the speed of light. So $v_*/c \ll 1$, and we can make a simplified approximate formula:¹³

$$v' \approx (v_* \pm c/n)(1 \mp v_*/(cn)) \approx \pm(c/n + v_*(\pm 1 - n^{-2})). \quad (28.5)$$

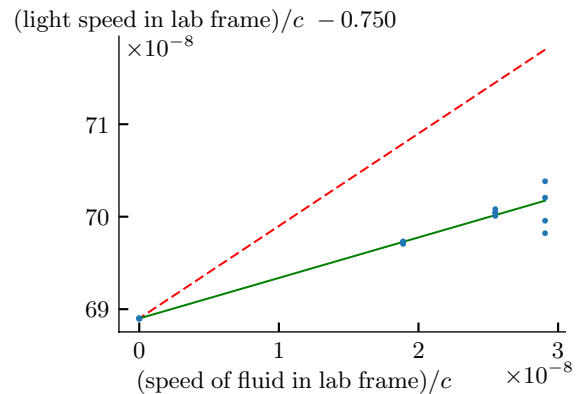
At last, we have a testable prediction. The hypothesis that the full equations of electromagnetism have galilean invariance predicted $v' = \pm(c/n \pm v_*)$, which *differs from Equation 28.5*. If we plot v' (speed of light in water, measured in the lab's coordinate system) versus v_*/c , then the two competing theories make different predictions for the *slope* of the data.

¹¹After Fizeau's experiment was done, æther theorists tried to wriggle out of this failed prediction with a theory that we now regard as laughably contrived. But it's best not to laugh—we'll see which of today's theories also look comical in the future.

¹²Alternatively, Equation 28.4 gives this result directly.

¹³The first published derivation of this formula seems to be by Max von Laue (1907).

Figure 28.3: Results from Michelson and Morley’s lesser-known experiment. The leftmost dot shows light velocity in water at rest with respect to the lab. Other dots show the data from a total of 12 trials spanning three different nonzero fluid velocities. The *solid line* shows the prediction of Equation 28.5. For comparison, the *dashed line* shows the prediction based on the hypothesis of galilean invariance. [Data from Michelson & Morley, 1886.]



That is, both theories make firm, nonnull predictions, with no fudge factors (no fit parameters).¹⁴ That is, both are highly *falsifiable*, if you’ve got enough accuracy to measure the effect at all. Figure 28.3 shows the data from Michelson and Morley’s version of Fizeau’s experiment. To get enough accuracy, Fizeau and successors all used interferometry. Even so, the figure shows significant scatter. But the data certainly rule out the prediction of the galilean invariance theory (slope 1), and they don’t rule out Einstein’s prediction.

Michelson and Morley actually measured the *differences* in light speed between propagation with and against the water flow. In the figure, these differences have been divided by two and added to the speed c/n at zero flow. Because the graph shows only a small range of values for $(\text{water speed})/c$, the Einstein prediction appears to be nearly a straight line. At water speed approaching c , the curve would level off (Figure 28.2, page 299), but we’re nowhere near that regime.

28.6 POSTSCRIPT

Science is a way of trying not to fool yourself. The first principle is that you must not fool yourself, and you are the easiest person to fool.

— *Richard Feynman*

1. You may be asking, “Why try so hard to convince me that the theory is true? Just tell me the facts, so I can get on with becoming a scientist!”

Actually,

Some of today’s accepted theories are wrong, but we don’t know which ones.

We all need finely-honed critical skills. Studying past revolutions is useful to be ready for future revolutions.

¹⁴The index of refraction, n , is independently measured from experiments on refraction. So it’s a parameter, but not a fit parameter. Our derivation neglects the effect of dispersion in the medium; see Lerche, 1977.

In fact, science is a system of tools to prove that your wonderful new theory, which you love so much, is *not true*. Discovering that unfortunate fact is the first step to letting go and finding your next wonderful new theory, which *may* be true. When you find it, its truth may still not be clear to the world. It is instructive to see how classic theories gained the assent of a world that initially was opposed to them, by *surviving tests specifically designed to falsify them*.

2. Our provisional Lorentz boost (Equation 28.2) has the disturbing feature that $t' \neq t$. A lot of people got philosophically confused: “How can *time itself* change?” I won’t put words into Einstein’s mouth, but speaking for myself, I’d reply,

- I said nothing about *time itself*. I don’t know what *time itself* means. I have no apparatus to measure *time itself*.
- I do have various kinds of *clocks*. They are physical devices with periodic behavior. As such, they are subject to the postulate under investigation, which is that the equations governing them are invariant under Lorentz transformations.
- I know ways to attach sets of four numbers to events.¹⁵ Some of these coordinate systems are “good” in the sense that in them, physics is described by simpler equations than in the others (and always by equations of the same form).
- Einstein’s hypothesis is that the “good” coordinate systems are related to each other by transformations that include some of the ones given in Equation 28.2.
- It is true that these transformations imply that different, equally good, coordinate systems will disagree about whether two distant events are simultaneous ($t_{(1)} = t_{(2)}$ does not imply $t'_{(1)} = t'_{(2)}$). But what experimental result does that contradict? (Einstein couldn’t find any.)
- We have seen that the hypothesis of invariance under these transformations implies a testable, and verified, prediction for a nontrivial phenomenon, the “dragging of light” by a moving medium. We’ll add more phenomena to this list later.

3. There’s a remarkable feature of the derivation above: Nowhere did we find it necessary to describe the *mechanism* for the slowing of light in water. That is, details of the *dynamics* did not enter, apart from the hypothesis that whatever the slowdown mechanism is, it (like the rest of physics) is invariant under Lorentz transformations.¹⁶

I hope you’ll agree that the kinematic approach followed above is much simpler than solving Maxwell’s equations for light moving through a medium of water molecules!

4. Other scientists came close to relativity before Einstein. Today we regard their work as unreadable, because they got bogged down in detailed dynamical hypotheses. Einstein got simple results because he focused on the *kinematics*, specifically on one hypothesis about the *invariances* of the dynamics.

5. Although Lorentz invariance looks promising, we are far from being done. We wish to prove that the *full Maxwell equations* also have exact Lorentz invariance. Rather than attempt that head-on, we will first construct a new kind of tensor language

¹⁵One way to set up such a coordinate system is to use an array of identical clocks and synchronize them using light flashes.

¹⁶Later, when we complete our specification of the Lorentz transformations, you can confirm that our derivation continues to hold.

in Chapters 31–32. The new language seems elaborate at first, but it makes many derivations of this sort very straightforward.

FURTHER READING

Zhang, 1997

Fizeau, 1859.

Michelson & Morley, 1886.

Shankland, 1964.

A critical review of this class of experiments is presented in Lerche, 1977.

Also see Galison, 2003.

PROBLEMS

28.1

Confirm that the provisional Lorentz transformation (Equation 28.2, page 297) really implements the sketch Figure 28.1b by finding the angles by which the ct and x axes are bent.

CHAPTER 29

Aberration of Starlight and Doppler Effect

And then, beside the Thames at Kew,
the house of Samuel Molyneux
supplied the firm foundations needed.
James Bradley, Samuel's friend, succeeded
in tracking Hooke's draconic star ...
The trouble was, it moved too far,
too fast, and in the wrong direction!
Despite the most minute inspection,
Bradley found nothing to suggest
his telescope was not at rest;
the star was shifting in the sky,
though maybe God alone knew why!

— James Muirden

29.1 FRAMING

You showed in Your Turn 28B that a family of transformations leave the 1D wave equation invariant. Some of these were unsurprising (translations and reflections in space and time), but a two-parameter family I called “provisional Lorentz boosts” were more interesting (Equation 28.2, page 297), because they relate two coordinate systems in uniform relative motion, and hence are candidates for implementing the Principle of Relativity. We saw that every coordinate system in the family we are considering agrees about whether or not a trajectory is moving at speed c , by construction.

So it's *not true* that Einstein said “everything is relative”: The property of moving at speed $3 \cdot 10^8$ m/s *is* absolute (all coordinate systems in an objectively “good” class agree about it). Certain *other* properties will prove to be relative. For example, different “good” systems will disagree about whether two events are simultaneous.¹

Chapter 28 stressed the value of predicting a testable, non-null effect that differs from newtonian physics; the present chapter will make more predictions of this type. First, however, we'll refine our “provisional” form of our proposed transformations to get their final form.

29.2 NO DILATATION INVARIANCE

The wave equation in vacuum is just one combination of the Maxwell equations. We'll now see that some of the “provisional Lorentz” boosts are *not* invariances of all of

¹Section 40.2 will discuss this statement in detail.

electrodynamics. So we need to throw some of them out. But we must do so carefully: The ones we keep must form a **subgroup**, that is, the composition of two successive transformations in that subset must also be in it. (Also, the inverse of any one of them must be in the chosen subgroup.)

To see that some of the transformations we found are spurious, consider a situation that's not just fields in empty space, specifically the Coulomb repulsion of two charged particles:²

$$m \frac{\partial^2 \vec{r}_{(1)}}{\partial t^2} = \frac{q^2 \vec{R}}{4\pi\epsilon_0 R^3}, \text{ etc., where } \vec{R} = \vec{r}_{(1)} - \vec{r}_{(2)}. \quad (29.1)$$

Here two point charges labeled by $\alpha = 1$ or 2 are assumed to each have the same charge q and mass m .

Now consider the provisional Lorentz boost with $\gamma \neq 1$ but $\beta = 0$, that is, $\vec{r}' = \gamma\vec{r}$ and $t' = \gamma t$. Rephrasing Equation 29.1 in terms of t' and \vec{r}' , we find that in the new coordinates it does not have the same form as initially—there's a factor of γ that fails to cancel.³

Actually, we needn't have worked so hard. If dilatations were an invariance of the laws of Nature, then there would be hydrogen atoms of any size! In the active viewpoint, just apply a dilatation to whatever laws dictate the usual solution, and find new solutions stretched by an arbitrary amount.⁴

There are several attitudes we could now take:

- We could just try saying, “The charges and/or masses of the particles must also change under such transformations.” But if the world had such an invariance, then there'd be a whole family of different electrons with continuously varying charges and/or masses. Nobody has seen them.⁵
- Or we could try saying, “There is some new dynamical entity, implicitly set equal to a fixed value in the Maxwell equations, which should instead be allowed to transform.” Maybe its transformation rule under dilatations could be arranged to be exactly what's needed to make Coulomb's law invariant. Actually, many authors have tried theories with such “dilaton” fields, and correspondingly “spontaneously broken dilatation invariance,” but none is widely accepted yet.
- Anyway, this course is dedicated to exploring the hypothesis that Maxwell's equations are already correct and complete as written. We just noted that those equations do not have dilatation symmetry. Should we therefore restrict to just those provisional Lorentz boosts with $\gamma = 1$? Unfortunately, you can quickly show that doing two of those transformations in succession does *not* amount to

²We combine the equations in this way so that we don't have to worry about how \vec{E} transforms; it's been eliminated. Although this formula will later need relativistic corrections, it's certainly valid for slowly-moving particles.

³It's true that the new equation has the same form except for the value of $q^2/(m\epsilon_0)$, but that's not good enough to declare that it's invariant. Note that an equation of the same form also describes two uncharged particles attracting each other gravitationally, so newtonian gravity, too, lacks dilatation invariance.

⁴Atomic sizes involve quantum mechanics, but even in classical electrodynamics Chapter 46 will show that an electron's ability to scatter radiation involves the “classical electron radius,” a length scale with a fixed value for electrons (distinct from its value for muons etc.).

⁵Also, each atomic species would come in a continuous range of sizes, contrary to observation.

any single boost with $\gamma = 1$. That is, the $\gamma = 1$ transformations do not “close into a group.”

- But Einstein already knew that there was a different subset, which really do close into a group, and are still sufficient to bake in in the Principle of Relativity. We will discover them in the next section.

Einstein then proved that these transformations were also exact invariances of the Maxwell equations. We’re not ready to do that,⁶ but nevertheless we’ll be able to show that the hypothesis that physics is invariant under them makes experimentally testable predictions, e.g. for the aberration of starlight and the two kinds of Doppler effect. Those predictions agree quantitatively with experiment, which will give us the courage to later push through the proof that they are invariances of the full Maxwell equations.

29.3 LORENTZ TRANSFORMATIONS IN ONE SPACE DIMENSION

Again, our task is to find a subset of provisional Lorentz transformations that forms a group, excludes the bogus dilatations, but still includes boosts. If we succeed, then we can explore the physical hypothesis that this reduced set of transformations are all invariances of Nature.

29.3.1 A subgroup that excludes dilatations

One way to specify a 1-parameter subset of the provisional 1D Lorentz boosts is to require that γ is not independent of β , but instead is a scalar function of it. We wish to do this in such a way that the subset closes into a group. We will guess a trial solution, then confirm it. Then we’ll see a deeper meaning for that solution.

The isotropy of space, and the fact that $\vec{\beta}$ is a vector in 3D, lead us to expect that γ won’t depend on which direction $\vec{\beta}$ points. We also expect that a boost by $\vec{\beta}$, followed by a boost by $-\vec{\beta}$, should amount to no boost at all (think about jogging backwards at speed v_* inside a train car that itself is moving at $+v_*$ relative to Earth). Thus, we require

$$\gamma \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix} \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This fixes $\gamma = (1 - \beta^2)^{-1/2}$, or⁷

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = (1 - \beta^2)^{-1/2} \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix}. \quad \text{Lorentz boost, 1D} \quad (29.2)$$

In particular, if $\beta = 0$ then $\gamma = 1$, and so pure dilatations are not part of this subset of transformations, as desired. From now on, whenever I write γ I’ll mean this particular function of β .

⁶See Chapters 31-33.

⁷Lorentz actually showed in 1904 that these transformations were invariances of the full Maxwell equations. However, Lorentz viewed this invariance as mathematical curiosity about the Maxwell equations—not an invariance of all of physics—and certainly not as justification to eliminate the æther.

For very small β , the transformations Equation 29.2 reduce to

$$t' = t - (v_*/c^2)x \approx t, \quad x' \approx x - v_*t \quad \text{where } v_* = \beta c.$$

These look just like galilean boosts. That's why Einstein's correction to the t' formula was missed for hundreds of years, during which Newton's laws made accurate predictions about terrestrial and celestial mechanics.

To see the significance of the Lorentz boosts, consider what happens when we reexpress the wave operator in terms of transformed coordinates:

$$\left[-\frac{\partial^2}{\partial(ct)^2} + \frac{\partial^2}{\partial x^2} \right] u \quad \text{becomes} \quad \gamma^2(1 - \beta^2) \left[-\frac{\partial^2}{\partial(ct')^2} + \frac{\partial^2}{\partial x'^2} \right] u. \quad (29.3)$$

We see that among the provisional Lorentz boosts, the subgroup of true invariances are those that leave the wave operator completely invariant—not just a multiple of itself. The following chapters will show that indeed these transformations are invariances of the full Maxwell equations. That is, *the coordinate systems in which Maxwell take the simplest form are related by Equation 29.2, which is physically different from the situation in newtonian physics.* In honor of Einstein, we'll call any of the “good” systems an **E-inertial coordinate system** to distinguish them from the corresponding notion in newtonian physics.⁸

29.3.2 Rapidity parameter

The preceding section characterized true Lorentz transformations as those that leave something (the form of the wave operator) unchanged. Two such transformations in succession will *also* have that property, so right away we see that the Lorentz transformations must close into a group.

It's algebraically messy to prove that statement directly, but there is a remarkable reformulation that makes it easy. Begin with an analogy to ordinary rotations. Why are rotations given by matrices that, in two dimensions, have the form $\begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix}$? One way to characterize such matrices S is by the statement $S^t S = \mathbf{1}$ (the identity matrix). But equivalently, rotations are those linear maps of coordinates that leave the algebraic form of the pythagorean formula invariant:⁹ $x^2 + y^2 = (x')^2 + (y')^2$. Chapter 13 used this property to show that the Laplace operator is invariant under rotations. Rotations close into a group: For example, in 2D α_1 followed by α_2 is equivalent to $\alpha_1 + \alpha_2$.

The wave equation involves something analogous but a bit different:

Your Turn 29A

Show that, in one spatial dimension, the Lorentz boosts are linear maps that preserve the form of the quantity

$$\Delta\tau^2 = c^{-2}((c\Delta t)^2 - (\Delta x)^2), \quad (29.4)$$

which we'll call the **invariant interval** between two events. [*Hint:* The proof is very similar to the proof of Equation 29.3.¹⁰]

⁸The latter were called galilean, or “G-inertial” systems in Section 25.6.1.

⁹More precisely, the linear maps that leave the pythagorean formula form-invariant consist of the rotations *and reflections* in x and/or y .

Because Equation 29.4 looks similar to the rotation case (except for the minus sign), we may hope that the appropriate symmetries will also look similar. Indeed, you should show that

$$\begin{bmatrix} \cosh \Upsilon & -\sinh \Upsilon \\ -\sinh \Upsilon & \cosh \Upsilon \end{bmatrix} \quad (29.5)$$

does the job, for any Υ . Some authors call Υ the **rapidity parameter**.

Your Turn 29B

- Confirm that any transformation of the form Equation 29.5 is a special case of the provisional Lorentz boosts, with $\gamma = \cosh \Upsilon$ and $\beta = \tanh \Upsilon$...
- ...and that conversely any Lorentz boost can be written in the form Equation 29.5.
- Section 29.3.1 argued that because the transformations Equation 29.5 can be characterized as those that leave something invariant (namely the wave operator, Equation 29.3), they must close into a group. Now confirm this expectation directly: Use a trig identity and Equation 29.5 to show that a boost with Υ_1 , followed by one with Υ_2 , is equivalent to a single boost with $\Upsilon_{\text{tot}} = \Upsilon_1 + \Upsilon_2$.
- Finally, confirm that this combination rule amounts to the same thing as a boost by the velocity v' obtained from the formula we found earlier, Equation 28.4 (page 298).

Thus, once again, we have found a 1-parameter subset of the provisional 1D Lorentz boosts that closes into a group, excludes dilatations, but includes boosts. Compared to Section 29.3.1, however, the derivation just given has the advantage of revealing a *geometric* interpretation: Lorentz transformations are the analogs of rigid rotations in a weird new kind of geometry. Either way, we now have a candidate physical hypothesis about the invariances of Nature and can get to work testing it.

29.4 A TYPICAL PARADOX AND ITS RESOLUTION

People made many objections to Einstein's theory, and still do. Out of many we could explore, here is one:

"I can determine whether I am at rest, as follows: I place my apparatus in a cart that can move at any uniform velocity v with respect to absolute rest. The cart is rigid: Its length, measured in its own rest frame, is always 0.5 meter. At some moment I prepare two identical alarm clocks at the center of the cart (as seen in its rest frame), and synchronize them. One of them is then taken toward the rear of the cart at uniform velocity $-u$ w.r.t. the cart, while the other is taken toward the front at uniform velocity $+u$ w.r.t. the cart. After equal times have elapsed on each clock, each emits a flash of light. If those flashes are both observed to arrive simultaneously at the center of the cart, then we know $u = 0$ (the cart is at absolute rest). If $u \neq 0$ then the detector will see them arrive non-simultaneously at the center, because it is advancing on the light coming from the front but retreating from the light coming from the rear."

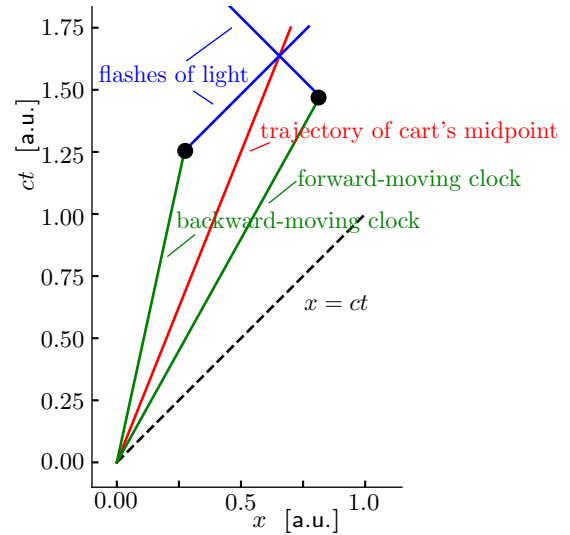


Figure 29.1: Thought experiment described in Section 29.4. The alarm clocks emit flashes of light at the events shown as *dots*. The coordinate axes refer to an E-inertial system in which the lab is at rest.

To evaluate this claim, Figure 29.1 shows an accurate spacetime diagram with $v = 0.4c$ and $u = 0.2c$. The red line is the trajectory of the center of the cart. The green lines are the trajectories of the two clocks on their ways to the ends of the cart. Their slopes are fixed by the velocity addition formula, and their end points are fixed by the time dilation and length contraction formulas.¹¹ Although the time to alarm is the same in each clock's rest frame (because the clocks are identical), they differ in the lab's coordinate system. The blue lines are the trajectories of light flashes emitted when the alarms go off. Their slopes are $\pm 45^\circ$ in any E-inertial frame, for example, the lab.

We see that, contrary to the claim in the indented story above, the flashes always coincide at the center of the cart, regardless of the value of u . Therefore we cannot use that observed coincidence to claim that u has any special value. *If we use the rules consistently, we avoid paradoxes.*

29.5 LORENTZ TRANSFORMATIONS IN THREE SPACE DIMENSIONS

Another advantage of the geometric picture is that it makes it clear how to introduce the other two space dimensions: Any transformation that looks like Equation 29.2 in a 2×2 block that includes ct and one spatial direction, and is the identity matrix in the other two directions, will preserve the form of the $(3 + 1)D$ invariant interval defined by extending Equation 29.4:

$$\Delta\tau = c^{-1} \sqrt{(c\Delta t)^2 - \|\Delta\vec{r}\|^2}, \quad (29.6)$$

¹⁰A more complete statement is that the linear maps that leave the formula Equation 29.4 form-invariant consist of Lorentz boosts *and reflections* in x and/or t .

¹¹These formulas in turn came from applying the Relativity Strategy (Equation 25.12, page 278).

and hence of the wave operator $\nabla^2 - (\partial^2/\partial(ct)^2)$, and hence of the wave equation itself. Thus, there are three independent kinds of Lorentz boosts, just as in the galilean case. Along with the three kinds of rotations, they amount to a six-parameter group of transformations called the full **Lorentz group**. Including the four space and time translations gives a ten-dimensional invariance group called the “Poincaré group.”

We can now put all our ideas together and formulate a successor to Idea 28.1 (page 295):

If we have two coordinate systems on spacetime, in each of which electromagnetic fields obey the wave equation, then those systems differ by a Lorentz transformation. If that transformation is just a boost by velocity $c\beta\hat{x}$, and a wave or particle in one system is moving at constant velocity $c\alpha\hat{x}$, then in the other system it will be observed to be moving at constant velocity $\hat{x}c(-\beta + \alpha)/(1 - \alpha\beta)$. (29.7)

If the two velocities are not parallel, then the formula is not as simple. However, the next section shows that in at least one important case it is still straightforward.

Your Turn 29C

Show that only one of the following matrices is a Lorentz transformation:

$$\begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad \gamma \begin{bmatrix} 1 & -\beta & 0 & 0 \\ -\beta & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

29.6 MORE KEY EXPERIMENTS: ABERRATION OF STARLIGHT AND DOPPLER SHIFT

29.6.1 Bending of light-speed trajectories

We are now ready to discuss the second of the two experimental observations that Einstein said convinced him: the aberration of starlight. Each time we look at the night sky, the stars’ positions relative to each other are always *almost* the same, but not quite. Even when we correct for refraction in our atmosphere, there are some apparent relative shifts, which are periodic with period one year. That is, the stars all crowd very slightly toward the direction of our orbital motion around the Sun. At its maximum, the displacement is just 20 arcsec.

Incredibly, this tiny effect was already observed in the late 1600s by astronomers searching for something completely different (stellar parallax in order to confirm the heliocentric model of the Solar System). As outlined in the epigraph to this chapter, Bradley was mystified to find there were indeed tiny annual variations in the relative positions of stars, but with the wrong magnitude and sign to be explained by parallax. Bradley even found an æther-based explanation for this phenomenon. Einstein realized that any alternative theory would need to address it again.

To get a prediction, we again apply the Relativity Strategy (Equation 25.12, page

278). Consider a trajectory specified in parametric form by

$$\begin{bmatrix} ct \\ x \\ y \end{bmatrix} = \begin{bmatrix} \xi \\ \xi \\ 0 \end{bmatrix}.$$

This formula specifies a chain of events depending on a parameter ξ , that is, a curve in spacetime. It could describe the progress of a flash of light (a wave packet, or one crest of a wave train) moving at speed c along the x axis.¹² Applying a Lorentz boost transformation along \hat{y} yields the same trajectory as viewed in another E-inertial coordinate system:

$$\begin{bmatrix} ct' \\ x' \\ y' \end{bmatrix} = \begin{bmatrix} \gamma & 0 & -\gamma\beta \\ 0 & 1 & 0 \\ -\gamma\beta & 0 & \gamma \end{bmatrix} \begin{bmatrix} \xi \\ \xi \\ 0 \end{bmatrix} = \begin{bmatrix} \gamma\xi \\ \xi \\ -\gamma\beta\xi \end{bmatrix}. \quad (29.8)$$

Your Turn 29D

- Show that the new trajectory's speed is $\sqrt{(\Delta x')^2 + (\Delta y')^2}/(\Delta t') = \sqrt{(\Delta\xi)^2 + \gamma^2\beta^2(\Delta\xi^2)}/(\gamma\Delta\xi/c)$.
- Confirm that this equals c , as it must.
- But the new trajectory is no longer directed along \hat{x}' . Show that instead, it makes an angle θ with the \hat{x}' axis, where $\tan\theta = \Delta y'/\Delta x' = -\gamma\beta$.

We could do a similar calculation for any initial angle between the trajectory and the boost direction (above you did the case where that angle is 90°). The new angle depends both on that original angle, and on β , so *the relative positions of the stars are different* according to the boosted (Earth-bound) observer. The effect is small because Earth's velocity change over six months is much smaller than c , but it's measurable.

Thus, Einstein's proposal for the invariances of physics makes an absolute prediction for the aberration, with no fudge factors (no parameter at all other than c). It either succeeds or fails—it's *falsifiable*. And, as he pointed out in his very first paper, it works, without any special pleading, no extra ad hoc hypotheses about how the æther wind is blowing, etc.

29.6.2 Effect on frequency

Now you try the derivation again. But instead of transforming a *trajectory*, this time transform a plane-wave solution to the wave equation.

Your Turn 29E

Show that again the apparent direction of \vec{k} changes, but also find the change in its *magnitude*. Specialize your result to two famous cases:

- \vec{k} is parallel to the boost (“longitudinal **Doppler shift**”).
- \vec{k} is perpendicular to the boost as in Your Turn 29D (“transverse Doppler shift”).

¹²We'll suppress the z coordinate to shorten the formulas. It's there, but it's not doing anything interesting.

Note that newtonian physics also predicts a longitudinal Doppler shift, but with a different magnitude from Einstein’s prediction.¹³ And newtonian kinematics predicts *zero* transverse frequency shift, unlike your answer to Your Turn 29E—a testable prediction.

Quantitative confirmation that the Doppler effect follows the relativistic formula, and excludes the galilean formula, had to wait for the Ives–Stilwell experiment (1938). Much more accurate experiments have been done right into the 21st century.¹⁴

29.7 AN ENORMOUS GENERALIZATION

29.7.1 Lorentz invariance must apply to all of physics

Let’s step back. Section 28.1.2 offered the paradox that the wave equation implied by Maxwell’s equations doesn’t have galilean invariance, so it was not clear that they are compatible with the Principle of Relativity. But we have now seen that the wave equation, with no modifications or additions, *is* invariant under a family of passive transformations that relate coordinate systems moving at constant velocity with respect to one another. We still need to do some work to upgrade this result to a corresponding statement about the full Maxwell equations, but looking ahead, we can state Einstein’s proposed resolution to the problem of Section 28.1.2 by saying¹⁵

Maxwell’s equations hardwire in the Principle of Relativity by using equations of motion that are invariant under Lorentz transformations—not galilean transformations.

Einstein took an extraordinary additional step. Up till now, Lorentz invariance may have seemed to be a peculiarity of electrodynamics, which we could safely ignore if, say, we were only interested in the motions of planets. But suppose that Maxwell’s equations *and* newtonian mechanics were both correct as written. That is, suppose that there is even one coordinate system in which both of those systems’ equations of motion correctly describe physics. Applying a galilean boost to that system would then spoil the form of Maxwell. Applying a Lorentz boost to it would spoil the form of Newton. In fact, there would be *no other coordinate system* in uniform, straight-line motion relative to the original one in which all equations of motion have the same form. So in such a world we could define “absolute rest” as that original coordinate system—contradicting the Principle of Relativity:

*If we want to hardwire in the P of R via an invariance, then that invariance must apply to **all** of physics—even to phenomena not yet discovered.*

That’s quite a leap. We can’t have it both ways. Einstein’s hypothesis is that

*Although newtonian physics has looked good for hundreds of years, actually it hasn’t been tested for objects moving at speeds near c , whereas Maxwell has. **So it’s Newton that has to be changed, not Maxwell.***

¹³Compare Equation 26.4, page 283.

¹⁴Some experiments were based on an ultrasensitive measure of wave frequency (Mössbauer effect). Other experiments used single atom emitters moving at high speeds.

¹⁵Compare our galilean statement (Equation 25.10, page 278).

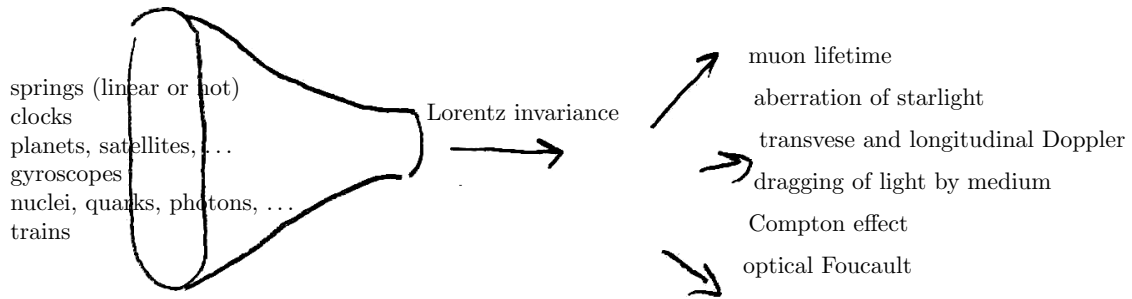


Figure 29.2:

Or, paralleling Idea 25.9:

Physics has an overarching mathematical property that transcends details of particular springs, clocks, planets, etc. That property is that the specific equations for any situation always have a family of preferred coordinate systems, which are related to each other by Poincaré transformations. (29.9)

(Recall that the Poincaré group contains Lorentz transformations along with translations.)

29.7.2 Muon lifetime, CMBR dipole, and more

The hypothesis of universal Lorentz invariance now gives us *many* nontrivial physical predictions, all of which start by saying (Figure 29.2) “Suppose that the dynamical laws governing [some process] are invariant under Lorentz transformations. . . .” From there, we can apply the Relativity Strategy (Equation 25.12, page 278). For example, we’ve seen how to understand Fizeau’s experiment, the aberration of starlight, and both kinds of Doppler shift, by using that approach.¹⁶

Note that when we hypothesize that “all laws of physics are invariant under Lorentz transformations,” we mean *all*, including quantum physics. Here are more examples:

- Suppose that, whatever process makes the muon disintegrate, that process is invariant under Lorentz transformations. We capture some muons, bringing them to rest with respect to our lab, and find that each one’s lifetime is $2.2 \mu\text{s}$. Then we can predict that a muon moving rapidly relative to the lab’s E-inertial coordinate system will *also* live $2.2 \mu\text{s}$ in an E-inertial coordinate system in which the muon is at rest (called a **rest frame**). Transforming this duration into the laboratory coordinate system via Equation 29.2 (page 307) shows that a fast-moving muon appears, in the lab, to live longer before disintegrating than does a muon at rest, as observed. Specifically we predict a lab lifetime $\gamma(2.2 \mu\text{s})$, during which the muon travels $\gamma\beta c(2.2 \mu\text{s})$, farther than it would have gone under the hypothesis

¹⁶For Fizeau: “Suppose that, whatever interactions slow light down in water, they are invariant under Lorentz transformations. . . .” For aberration and Doppler: “Suppose that, whatever dynamics are responsible for the propagation of light in vacuum, they can be expressed in terms of equations invariant under Lorentz transformations. . . .”

of galilean invariance.

[T2] Section 29.7.2' (page 318) discusses this experiment in more detail.

- Suppose that, whatever nuclear physics is responsible for an excited nucleus of iron to give off a gamma photon by recoilless emission,¹⁷ that process is Lorentz invariant. Then a second iron nucleus that could resonantly absorb such a photon will not do so if it's in motion relative to the first one, because in *its* rest frame the photon is Doppler shifted, and hence off resonance.
- The Doppler shift formula also lets us deduce the motion of distant galaxies relative to us:¹⁸ We suppose that, whatever atomic physics is responsible for making hot gas give off light with a pattern of spectral lines, that process is invariant under Lorentz transformations. Then a hydrogen atom moving rapidly relative to us will have the same spectral lines as one in our lab, if it's measured in the E-inertial coordinate system in which that atom is at rest. Transforming that outgoing wave to our lab's E-inertial coordinate system gives its apparent frequency when we observe it with a spectrometer.

The Doppler effect also predicts that the apparent temperature of the cosmic microwave background radiation appears slightly higher in one direction of the sky, and slightly cooler in the opposite direction (the **dipole anisotropy**). This was observed shortly after the discovery of the cosmic microwave background radiation.¹⁹ This anisotropy must be subtracted from observations if we want to see the far smaller, and more cosmologically interesting, anisotropy that arises from early Universe fluctuations.

- Strong and weak nuclear forces, which are not electrodynamic in origin, lead to particle reactions that conserve energy and momentum. But we'll soon see that, in order for energy and momentum to be conserved in every E-inertial coordinate system, we must modify their newtonian definitions, in ways that have experimentally testable consequences in nuclear and high-energy physics.

The incredible power of relativity lies in the fact that these apparently unrelated phenomena, and many others, are all quantitatively explained with *one* idea, (29.9). The existence of laws of this sweeping generality is a miracle, the *basic epistemological miracle of physics*. It's what gives physical law a different character from the rules governing other branches of science.

Again: The revolutionary aspect of Einstein's logic was not just the factual content of his proposal, but also the *method*: Till then, the general approach had been to propose a law of Nature, then test it. Instead Einstein went straight to the next higher level, writing a transformation principle that's proposed to be an invariance of *all* laws of Nature, *whatever they may turn out to be*.

¹⁷You'll examine this phenomenon in Problem 30.2.

¹⁸In 1868, William and Margaret Huggins detected a Doppler shift in the spectrum of Sirius, the birth of this indispensable astronomical method.

¹⁹You'll explore the CMBR dipole anisotropy in Problem 29.4. Prediction: Peebles and Wilkinson, Phys Rev **174**(1968)2168. Observation: Smoot, Gorenstein, Muller Phys Rev Lett **39**(1977)898.

29.8 WHAT'S NEXT

1. We now have a proposal for a set of transformations that:

- . Are invariances of the wave equation; and
- . Form a group.

But the wave equation we have studied assumed a *scalar* field, whereas we know that the electric and magnetic fields are not scalars. Not only do the components of \vec{E} transform among themselves under rotation; that nagging experiment with the magnet and coil seems to imply that \vec{E} *mixes with* \vec{B} under a boost (Hanging Question #A, page 11). So we need to augment our Lorentz transformations on spacetime by making a proposal for what exactly happens to the components of \vec{E} and \vec{B} under them. Only then will we have a firm proposal for what transformations are supposed to leave the Maxwell equations invariant. Then we can do the math to see if it's true—after first inventing some powerful notation to help us (“high-tech relativity”), based on the close relation of Equation 29.5 to rotations.

2. Our logic may still feel a bit ad hoc, but here we were still just feeling our way trying to guess the right hypothesis. Now that we've got it, and it looks promising, we are in a position to develop a more streamlined formulation in Chapters 31–32.

3. First, however Chapter 30 will explore more generic (kinematic) consequences of Lorentz invariance, and their experimental signatures.

T₂ Track 2

29.3'a Light-cone coordinates

Here's a more elegant derivation of Lorentz transformations than the one in the main text.

Suppress y, z for the moment, and consider only ct, x . It is helpful to define **light-cone coordinates**

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix} \quad \text{so} \quad \begin{bmatrix} ct \\ x \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \quad (29.10)$$

Then the general solution to the wave equation takes the elegant form $f(u) + g(v)$ for any two functions f, g . The function f describes a waveform moving to the left; g is a waveform moving to the right.

The linear transformations $u' = Au, v' = Bv$ change a solution to $f'(u') + g'(v')$ where $f'(u') = f(Au)$ etc., which has the same functional form as before. So any such transformation is an invariance of the solution space of the wave equation; that is, waves traveling left or right at velocity $\pm c$ in the original coordinates are again traveling left or right at velocity $\pm c$ in the new system.

In light-cone coordinates, the operation appearing in the wave equation (the wave operator, or D'Alembertian) has the simple form $\partial^2/\partial u\partial v$. In terms of the transformed coordinates, this is $(AB)(\partial^2/\partial u'\partial v')$. Such transformations are therefore invariances of the wave equation. They include dilatations with $A = B \neq 1$; those are invariances of the vacuum wave equation, but not of the rest of physics. We can eliminate them, and get the expected 1-parameter family of boosts, if we restrict to the case where $A = B^{-1}$. That family of transformations are precisely the Lorentz boosts.

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \frac{1}{2} \begin{bmatrix} A+A^{-1} & A-A^{-1} \\ A-A^{-1} & A+A^{-1} \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix}. \quad (29.11)$$

This can be placed in its more famous form by letting $\gamma = (A + A^{-1})/2$ and $\beta = (A^{-1} - A)/(A^{-1} + A)$, yielding Equations 29.2 or 29.5.

29.3'b Invariant interval

Light-cone coordinates also make it easy to see that the quantity $-2(\Delta u)(\Delta v)/c^2 = (\Delta t)^2 - (\Delta x/c)^2$ is invariant under Lorentz transformations (it acquires a factor of $A/A = 1$). The square root of this quantity, $\Delta\tau$, has dimensions of time; we call it the invariant interval between two events. If those events can be joined by a trajectory moving at $\pm c$, the interval equals zero because either $\Delta u = 0$ or $\Delta v = 0$; if they can be joined by a trajectory moving slower than c then the invariant interval is real and positive.

29.3'c Velocity addition

It's also easy to find the combined effect of two Lorentz boosts by using light-cone coordinates. Please convince yourself that the combined operation is itself a Lorentz boost with $A_{\text{tot}} = A_1 A_2$. To interpret this result, invert the relations between A and (β, γ) to find

$$A = \gamma(1 + \beta) \quad \text{or} \quad A^{-1} = \gamma(1 - \beta).$$

Thus, $A_{\text{tot}} = \gamma_1(1 + \beta_1)\gamma_2(1 + \beta_2)$ gives

$$\beta_{\text{tot}} = \frac{(1 + \beta_1)(1 + \beta_2) - (1 - \beta_1)(1 - \beta_2)}{(1 + \beta_1)(1 + \beta_2) + (1 - \beta_1)(1 - \beta_2)} = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}.$$

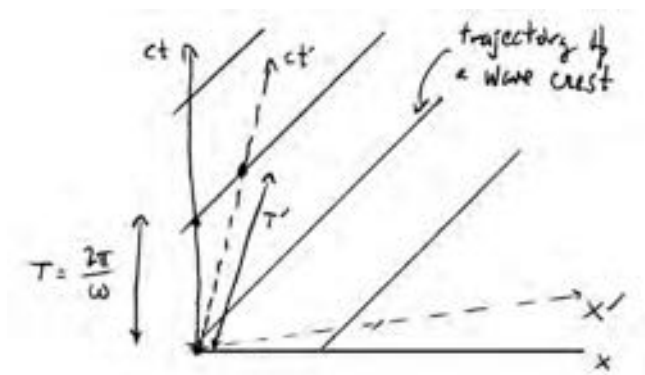
29.3'd Relation to rapidity

Equation 29.11 is the same as Equation 29.5 (page 309) with $\Upsilon = \ln A$. This is helpful, because in light-cone coordinates the composition law is simply $A_{\text{tot}} = A_1 A_2$ (show that). So $\Upsilon_{\text{tot}} = \ln(A_1 A_2) = \Upsilon_1 + \Upsilon_2$, which agrees with your result in Your Turn 29A (page 308).

T₂ Track 2

29.6.2'

There is a more geometrical (less algebraic) way to think about the longitudinal Doppler shift:



The diagram above shows the loci of a chain of wavefronts, each moving along \hat{x} at speed $+c$ and separated in time t by period T . The dashed lines are coordinate axes for an E-inertial coordinate system moving with respect to the unprimed system. The period T' of the same wave observed in this system depends on the intersection of the t' axis with a wavefront, as shown. You can work out the relation between T' and T , and again recover the longitudinal Doppler formula.

T₂ Track 2

29.7.2'

The muon had yet not been discovered in 1905, so this result was not available to Einstein. We now call the relevant physical law “the weak interaction,” part of the more general “electroweak theory.”

Actually, the muon lifetime is a *random variable*. It has an exponential distribution with expectation $2.2 \mu\text{s}$. It is this expectation that gets transformed when the muon is moving

relative to the lab. This sounds like an annoying extra complication, but actually it's what makes possible a measurement of muon lifetime when we don't know the exact creation times of individual cosmic-ray muons. We actually measure the probability per unit time of disintegration for a sample of muons in flight, and compare it to the corresponding quantity for a sample of muons that have been captured, and hence slowed down, by atomic nuclei.

PROBLEMS

29.1 *Cart before the horse*

Figure 29.1 showed a particular case of the thought experiment described in Section 29.4. Maybe the result shown was accidental. Make a similarly accurate figure showing the case in which the cart's velocity relative to the lab is $v = 0.2c$ and the clocks move apart from its center at speeds $u = \pm 0.4c$. For concreteness, suppose that the cart's total length is 0.5 m. [Hint: It may be easiest to get a computer to make this figure. If you do that, make sure to use equal scaling for the x and ct axes.]

29.2 *Length contraction*

A long, straight, thin wire carries current I . A negative point charge $-q$ moves relative to the atoms making up the wire; its relative motion is parallel to the wire's axis in the opposite direction to the flow of current. The test charge is located a distance r from the axis of the wire.

Calculate the force on the point charge, in each of two different coordinate frames S ("lab frame") and S' ("rest frame"), where:

- In S , the positive charges in the wire are at rest; the negative charges each move at average velocity v_0 ; and the point charge also moves at v_0 . The wire is electrically neutral in this frame;
- In S' , the positive charges in the wire move with velocity $-v_0$; the negative charges and the point charge are at rest.
- Explain the relation between the two force vectors you found.

29.3 *Optical Foucault pendulum*

A lab that is anchored to Earth's surface sets up a non-inertial coordinate system, due to Earth's rotation. We can detect this small acceleration without looking at the stars, for example, by setting up a Foucault pendulum. In this problem you will explore an optical analog, which is the basis of an important technology.

Imagine a flat table with mirrors, such that light will traverse a roughly square path in vacuum and return to its starting point. More precisely, the light path is a trapezoid: One edge is oriented North–South and has length L in its rest frame. The next edge (called b) is oriented East–West and has length L in its rest frame. The third edge is oriented North–South and has length L in its rest frame. The last edge (called a) is oriented East–West and has length *slightly longer than* L in its rest frame, because lines of latitude on Earth are not of equal length.

You will be working out the round-trip transit time for light in the rotating apparatus, and specifically the *difference* in transit time depending on whether the light goes round clockwise or counterclockwise (when viewed on a line directed toward the center of Earth.) The apparatus is much smaller than Earth: $L = 1$ m. It is located at north latitude α , that is, the polar angle measured from the north pole is $\theta = \pi/2 - \alpha$.

- You know the angular frequency ω of Earth's rotation (and which way it is rotating). From that you can make a dimensionless parameter $\epsilon = \omega R_{\text{earth}}/c$. Evaluate this numerically.

There would be no difference in transit times if Earth were not rotating. But we may hope that there will be an effect at order ϵ . So work the following steps keeping only first-order contributions. (If the answer is zero, you can go back and look at higher-order terms.) You can also use the fact that $L \ll R_{\text{earth}}$.

Let unprimed variables ct and \vec{r} refer to an inertial (hence nonrotating) coordinate system in which the center of Earth is at rest. The key facts about rotation are that (i) edges a and b move at *different speeds* relative to the unprimed system, because they are located at slightly different polar angles $\theta_a > \theta_b$, and that (ii) each is directed nearly parallel to its velocity. (Actual lines of latitude and longitude are curved, and so do not coincide perfectly with the straight edges of the apparatus, but this difference is unimportant in the problem.)

You can forget about the other two edges, which are oriented perpendicular to their velocities.

You know the length of each edge in its own rest frame. Begin by studying a light beam that proceeds in a clockwise direction. Thus, it starts at the southeast corner, traverses a heading West, reflects off a mirror, proceeds North, and reflects again. Then it traverses b heading East, reflects one more time, and proceeds South to its starting point.

- Find the transit times in the unprimed frame for edges a and b and add them. [*Hint:* The Relativity Strategy may be helpful (Idea 25.12).]
- Repeat for a light beam circulating counterclockwise.
- Subtract the two preceding results and express your answer in terms of θ , L , ω , R_{earth} , and constants of Nature. Although you have computed time in the unprimed frame, explain why the round-trip transit time difference will have the same value according to a clock fixed to the instrument.
- Evaluate your answer for an apparatus located at north latitude $\alpha = \pi/4$. Which transit time is faster: the clockwise or the counterclockwise route?
- Compare your answer to the period of visible light. Is this a measurable effect?

29.4 CMBR anisotropy

Let's think of the cosmic microwave background as a classical EM field consisting of a superposition of many plane waves. We assume that there's a reference frame (ct, \vec{r}) in which it's nearly isotropic. That is, in this frame the waves have random phases and polarizations, and wavenumbers drawn from the probability distribution

$$\wp(\vec{k})d^3\vec{k} = C f_0(\|\vec{k}\|/\tau)d^3\vec{k}.$$

In this formula τ is a constant related to the temperature of the radiation (about $(2.7\text{K})k_{\text{B}}/(\hbar c)$). C is a normalization constant. $f_0(x) = (e^x - 1)^{-1}$ is the familiar Planck function. We want to know what this EM field looks like in our reference frame (ct', \vec{r}') , which is moving at speed V_0 along the $-\hat{z}$ direction relative to the original frame. So find the probability distribution $\wp'(\vec{k}')$ of \vec{k}' vectors. (You can forget about polarization.) Show that the distribution in the primed frame, restricted to any particular direction \hat{k}' , again has the Planck form, but with a direction-dependent effective temperature $\tau_{\text{eff}}(\hat{k}')$, which you are to find.

29.5 Lorentz I

In class I gave you the famous formula for the transformations of z and t corresponding to a boost in the \hat{z} direction. That formula contained the mysterious γ factor. Define the more general formula

$$\begin{bmatrix} ct' \\ z' \end{bmatrix} = \Lambda \begin{bmatrix} ct \\ z \end{bmatrix} \quad \text{where} \quad \Lambda = f(\beta) \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix}. \quad (29.12)$$

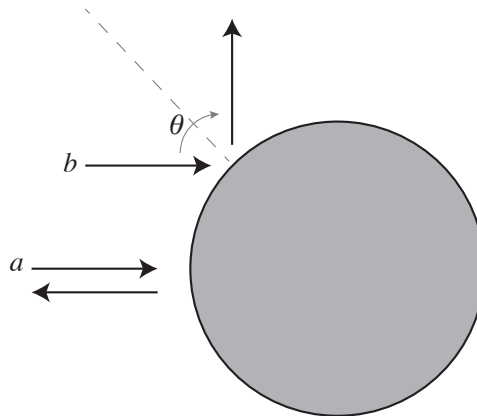
I claimed that only by making the choice $f(\beta) = (1 - \beta^2)^{-1/2}$ can we get the required composition law, i.e. that $\Lambda_1 \Lambda_2$ is again a matrix of the form Equation 29.12 for some β_{tot} .

- Show the converse of this claim, that is, assume $f(\beta) = (1 - \beta^2)^{-1/2}$, work out the matrix product for two boosts β_1 and β_2 , and show that the product is of the form (*) for some β_{tot} (find it).
- Let's rederive it in another, more insightful, way. We know that any 2D rotation matrix can be written $\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$. Try replacing the angle θ by an *imaginary number* $i\chi$, so that the trig functions turn into the corresponding hyperbolic functions ($\cosh \chi$ etc.). Make any other necessary changes, and see if you can get a family of matrices Λ with the required property to be Lorentz transformations.
- The form you found in (b) makes it much easier to work out the composition property than was the case in (a). Work it out: Find χ_{tot} given χ_1 and χ_2 , etc.
- The family of transformations you found in (c) are the same as the ones in (a). Show this explicitly by finding the correspondence between χ and β .

29.6 Disco discovery

In this problem, as always in this course, use classical (not quantum) physics.

An electromagnetic plane wave has wavelength λ and moves along the positive \hat{z} direction when measured in one E-inertial coordinate system (the “lab frame”). The wave is reflected by a spherical mirror, which is moving relative to the lab frame, also in the $+\hat{z}$ direction, but with velocity v .



- Some of the light will be reflected directly backward, along the $-\hat{z}$ direction. Find its wavelength $\tilde{\lambda}$ as measured in the lab frame, in terms of λ , v , and physical constants. [*Hint*: Apply the Relativity Strategy (Equation 25.12, page 278). There's

another E-inertial frame (not the lab frame) in which you certainly know the relation between incident and reflected frequencies. Convert that knowledge into a statement relating the wavelengths as seen in the lab.]

- b. Generalize your answer to the case where the scattered light is observed in an arbitrary direction, making an angle θ with the \hat{z} axis in the lab frame. [*Hint:* Same hint as in (a). You may find it easier to express your answer in terms of the scattering angle as seen in the other frame, then express that angle in terms of θ .]

[*Notes:* (a) If you know the Compton formula, and it disagrees with your answer, don't worry. Historically this disagreement led to the acceptance of Einstein's light-quantum theory—a modification to classical electrodynamics. In the domain of classical EM (coherent states of many photons, e.g. bouncing radio off a satellite) your result is accurate.

(b) Bouncing a radar beam off a speeding car and measuring the beat frequency between outgoing and returning signals is another real-world application.]

29.7 *Velocity addition*

[Not ready yet.]

29.8 *Rapidity*

CHAPTER 30

Relativistic Momentum and Energy of Particles

There is another outstanding kinematic consequence of the hypothesis that all of physics, not just electrodynamics, is Lorentz invariant. It concerns energy and momentum. However, the experiments confirming it (and later ending WW2), came long after Einstein's initial discovery, which was based on . . . electromagnetic phenomena.

30.1 CONSERVATION LAWS LOST

Section 29.7 mentioned that Lorentz invariance is all-or-nothing: We can't have some of physics invariant under Lorentz transformations while some other part is invariant under galilean transformations. Accordingly, let's think beyond Maxwell's equations, to consider any sort of interaction that could be called a "collision" among "particles." For our purposes, a "particle" is a region of space containing something that is initially isolated from the rest of the world (no relevant interactions). We imagine a collection of these, all initially mutually noninteracting, which come together and interact during a finite time interval (a "collision"), and suppose that eventually some other collection of "particles" emerge that are again noninteracting. Thus in some contexts it may even be appropriate to treat an entire galaxy as a "particle," or a planet, . . . , on down to atomic nuclei and beyond.

Suppose that two particles with masses $m_{1,2}$ and velocities $\vec{v}_{(1,2)}$ are initially noninteracting, then a "collision" occurs, and two other particles with $m_{3,4}$ and $\vec{v}_{(3,4)}$ emerge, eventually separating so that they are again noninteracting.

In first-year physics, we start with Newton's laws and prove that

$$\vec{p}_{(1)}^N + \vec{p}_{(2)}^N = \vec{p}_{(3)}^N + \vec{p}_{(4)}^N, \quad \text{where } \vec{p}_{(\ell)}^N = m_{\ell} \vec{v}_{(\ell)} \text{ (newtonian)}. \quad (30.1)$$

The quantity $\vec{p}_{(\ell)}$ is called the **newtonian momentum** of particle ℓ .

But even if we didn't yet know Newton's laws, and had merely guessed the conservation law Equation 30.1, we could nevertheless state confidently that it is consistent with the rotational invariance of the world. That's because under rotations the components of velocity (and hence those of \vec{p}) transform in a simple way, as a 3-vector. Moreover, mass is rotation-invariant (scalar), so the $m_{\ell} \vec{v}_{(\ell)}$ are also 3-vectors:

$$\vec{p}_{(\ell)}^N = S \vec{p}_{(\ell)}^{N'}. \quad (30.2)$$

When we express each term of Equation 30.1 in terms of a rotated coordinate system, then, the matrix S is a common factor:

$$S(\vec{p}_{(1)}^{N'} + \vec{p}_{(2)}^{N'} - \vec{p}_{(3)}^{N'} - \vec{p}_{(4)}^{N'}) = 0. \quad (30.3)$$

Multiplying both sides of this equation by S^{-1} gives an equation of the same form as Equation 30.1, so the newtonian conservation law is invariant under rotations.

Your Turn 30A

In newtonian physics, mass can be exchanged but *total* mass is conserved in any collision:

$$m_1 + m_2 = m_3 + m_4 \quad (\text{newtonian}). \quad (30.4)$$

From this, show directly (without appeal to Newton's laws) that Equation 30.1 is also invariant under galilean boosts.

In short,

Even if we didn't know Newton's Laws, or the details of what's inside our "particles," we could nevertheless say that Equation 30.1 is at least compatible with the overarching principle of invariance under the galilean group.

However, we cannot adapt the simple argument in Equation 30.3 to show that Equation 30.1 is consistent with Lorentz invariance, because \vec{v}' is a complicated, non-linear function of \vec{v} (Equation 28.4, page 298). Indeed, given a set of four momenta $\vec{p}_{(\ell)}^N$ that obey Equation 30.1, then their values in another E-inertial frame will *not* in general obey it. So Equation 30.1 cannot be a valid law of Nature in the Lorentz-invariant world that we are exploring. Nor can Newton's laws be valid, because Equation 30.1 is a consequence of them.

There is another famous conservation law in first-year physics:¹

$$\mathcal{E}_{(1)}^N + \mathcal{E}_{(2)}^N = \mathcal{E}_{(3)}^N + \mathcal{E}_{(4)}^N, \quad \text{where } \mathcal{E}_{(\ell)}^N = \frac{1}{2}m_\ell\|\vec{v}_{(\ell)}\|^2 \quad (\text{newtonian}). \quad (30.5)$$

This formula is rotation invariant by an even easier argument than before: Each term is *separately* invariant. It is also straightforward to check that Equation 30.5 is galilean invariant.

However, Equation 30.5 also turns out not to be Lorentz invariant. Therefore it, too, cannot be a valid law of Nature in any Lorentz-invariant world.

So are energy and momentum not conserved?

30.2 CONSERVATION LAWS RECOVERED

30.2.1 An unexpected analogy

Once again, Einstein realized that *there is some freedom in how we interpret the conservation laws*. Maybe $\vec{p}^N = m\vec{v}$ and $\mathcal{E}^N = \frac{1}{2}m\|\vec{v}\|^2$ are not the right formulas, and some other formula *would* give conserved quantities.

But where should we look for such formulas? Einstein's reply was so radically different from his contemporaries' that it really deserves to be called **Einstein thinking**. Faced with this sort of question, the obvious approach seems to be to guess or deduce the right equations of motion, then prove a theorem about a mathematical property they possess.² By 1905, this approach had led to a lot of unreadable papers, and

¹Newton himself didn't use conservation of energy. Émilie du Châtelet and others seem to have been responsible for that insight..

²[\[2\]](#) For example, we might guess the correct lagrangian function, then apply Noether's theorem to it.

moreover, scientists didn't even realize how hopeless it was, because many phenomena now described by particle physics hadn't even been discovered.

Einstein stood the traditional approach on its head:

- Start with a proposal for a symmetry of physics, in this case Lorentz.
- Discard hypotheses incompatible with the proposed symmetries, in this case conservation of newtonian momentum and energy.
- Find replacement hypotheses that are compatible, *without* attempting yet to deduce them from any equations of motion.
- Seek experimentally falsifiable consequences of the proposal.
- If the proposal survives enough nontrivial challenges, use it as a guide to *find* the right equations of motion.

To get started on this program, recall again the root of the problem: Velocity is $d\vec{r}/dt$, and both the numerator and denominator of this expression transform under Lorentz boosts (unlike the case with galilean boosts). If only we could replace the denominator by something that *didn't* transform, then we'd be in a simple situation like that for rotations: t is invariant under rotations, so d/dt doesn't alter the rotational properties of \vec{r} , so velocity transforms linearly, leading us to Equation 30.2.

The following argument is an outstanding example of lateral thinking. First, note that the invariant interval between two events in spacetime (Equation 29.6, page 310) is invariant under Lorentz transformations: $\Delta\tau = \Delta\tau'$. Thinking of a particle's trajectory as a chain of events in spacetime, the invariant interval $d\tau$ between any two neighboring events is always real, because particle trajectories cannot move faster than speed c . In fact, we'll see that an ordinary material particle cannot ever reach speed c , so $d\tau > 0$ for any two distinct points. That means that we can integrate $d\tau$ along the trajectory to obtain a parameter for the trajectory, called **proper time** τ . That is, we may *consider the time and the spatial position along the particle's trajectory both to be functions of τ* .³ We now define

$$\check{p} = m \frac{dx}{d\tau}, \quad \text{relativistic momentum} \quad (30.6)$$

which is a function along the trajectory. In this formula, m is a constant with dimensions of mass, an invariant property of the particle. We'll call it "the mass" of the particle.⁴

We also introduce an analogous quantity

$$K = m \frac{d(ct)}{d\tau}. \quad (30.7)$$

The point of these definitions is that then the pair

$$\begin{bmatrix} K \\ \check{p} \end{bmatrix} = m \frac{d}{d\tau} \begin{bmatrix} ct \\ x \end{bmatrix}$$

³See also Problem 30.1. We have previously used ξ to denote a generic parameter along a trajectory; τ is specifically proper time. For the trajectory describing a light flash, however, $d\tau = 0$, so we must use some other parameterization, for example the one used in Section 29.6.1.

⁴Old books introduce the term "rest mass." That quantity is now simply called "mass," because the alternative concept "relativistic mass" is no longer deemed worthy of any name at all.

has the same simple transformation under Lorentz boosts as do ct and x :

$$\begin{bmatrix} K \\ \check{p} \end{bmatrix} = m \frac{d}{d\tau} \Lambda \begin{bmatrix} ct' \\ x' \end{bmatrix} = \Lambda \left(m \frac{d}{d\tau} \begin{bmatrix} ct' \\ x' \end{bmatrix} \right) = \Lambda \left(m \frac{d}{d\tau'} \begin{bmatrix} ct' \\ x' \end{bmatrix} \right) = \Lambda \begin{bmatrix} K' \\ \check{p}' \end{bmatrix}. \quad (30.8)$$

Here Λ is a 2×2 Lorentz transformation matrix and we used the fact that $d\tau' = d\tau$. Note that we are allowed to pull the Lorentz transformation matrix outside the derivative because its entries are constants. Even if the particle is itself accelerating, nevertheless we are boosting to a coordinate system with some *constant* velocity βc relative to the original one. That is, (K, \check{p}) form a pair with a simple, linear transformation rule. In fact, it's the *same rule* as the one for ct, x .

We now propose two new conservation laws:

$$\check{p}_{(1)} + \check{p}_{(2)} - \check{p}_{(3)} - \check{p}_{(4)} = 0 \quad \text{and} \quad (30.9)$$

$$K_{(1)} + K_{(2)} - K_{(3)} - K_{(4)} = 0, \quad (30.10)$$

which differ from the discredited newtonian versions. Exactly as in the discussion of rotation, we know at once that Equations 30.9–30.10 are automatically Lorentz invariant. Proof: Equation 30.8 is analogous to Equation 30.2, and we can repeat the argument based on Equation 30.3.

That's remarkable: We still haven't postulated any detailed dynamical laws for collisions (possibly involving nuclear forces etc.), and yet we *still found the corrected form of the momentum* that leads to an acceptable conservation law. Indeed, for a slowly moving particle \check{p} becomes equal to Newton's momentum. To see this, note that

$$d\tau = \sqrt{dt^2 - (vdt/c)^2} = dt \sqrt{1 - (v/c)^2} = \gamma^{-1} dt, \quad (30.11)$$

and $\gamma \rightarrow 1$ for a slowly moving particle. Thus $m(dx/d\tau) \rightarrow m(dx/dt) = p^N$.

What about the new quantity K ? To identify its meaning, note that Equation 30.11 gives $K = mc\gamma \approx mc(1 + \frac{1}{2}(v/c)^2 + \dots)$. So Equation 30.10 multiplied by c says

$$(m_1 + m_2 - m_3 - m_4)c^2 + \mathcal{E}_1^N + \mathcal{E}_2^N - \mathcal{E}_3^N - \mathcal{E}_4^N + \dots = 0. \quad (30.12)$$

Equation 30.12 is indeed compatible with the newtonian equations Equations 30.4 (which says the first four terms sum to zero) and 30.1 (which says that the next four also sum to zero).

More generally, we define

$$\boxed{\check{\mathcal{E}} = cK = mc \frac{d(ct)}{d\tau}. \quad \text{relativistic energy}} \quad (30.13)$$

How can we dare to change the meaning of “momentum” and “energy?” The newtonian quantities are just not *useful*, because they cannot be conserved quantities in any Lorentz-invariant world. We found different quantities that *could* be conserved, and named them after the things they resemble. In fact, from now on we'll follow other authors and drop the checks: p , and its 3D generalization $\vec{p} = m \frac{d\vec{r}}{d\tau}$, will henceforth refer only to the relativistic formula, and \mathcal{E} will always mean $mc \frac{d(ct)}{d\tau}$. There won't be any ambiguity, because from now on we won't use the newtonian quantities at all. Reinstating the other spatial components gives our proposed conservation law as an equality of 4D vectors (Chapters 31–32 will christen such quantities **four-vectors**):

$$\begin{bmatrix} \mathcal{E}_{(1)}/c \\ \vec{p}_{(1)} \end{bmatrix} + \begin{bmatrix} \mathcal{E}_{(2)}/c \\ \vec{p}_{(2)} \end{bmatrix} - \begin{bmatrix} \mathcal{E}_{(3)}/c \\ \vec{p}_{(3)} \end{bmatrix} - \begin{bmatrix} \mathcal{E}_{(4)}/c \\ \vec{p}_{(4)} \end{bmatrix} = 0. \quad (30.14)$$

30.2.2 What has/has not been shown

We have shown that proposed conservation laws involving two replacements for newtonian formulas, Equations 30.6 and 30.13, are at least compatible with the physical hypothesis that all of physics is Lorentz invariant. We would eventually like these formulas to emerge from some complete theory, but in 1905 it was too early for that, and remained so for at least another 70 years.⁵ Instead, following “Einstein thinking,” we will shelve that project and instead look for direct experimental tests of the proposed conservation laws, Equations 30.14.

Later chapters will develop the dynamical details in the context of electrodynamics. Specifically, we will look for appropriate formulas for the energy and momentum of *fields*, then prove a conservation theorem about the total energy and momentum of particles *and fields* starting from Maxwell’s equations and the Lorentz force law.

30.2.3 A geopolitical consequence

Newtonian physics proves the conservation of energy and assumes separate conservation of mass. But we only obtained a single combined law, Equation 30.12 in the newtonian limit. Although this is consistent with the separate conservation of mass and energy, Einstein realized that there was *no fundamental reason why masses must be unchanged*, nor even for total mass to be conserved, in collisions. He concluded that a **mass defect** (change in total mass) must, according to Equation 30.12, appear as nonconservation of kinetic energy in a collision reaction.⁶ He immediately grasped that even a fraction of a percent change in mass could account for the enormous energies that seemed to come from nowhere in radioactive decay.⁷ Experiments performed decades later, with the first particle accelerators, confirmed this prediction quantitatively.⁸

That is definitely a practical result. Eventually, everybody realized that if you could slowly release the energy equivalent of a gram of matter, you’d get 10^{14} J, plenty to run a big city for a long time. Everybody also realized that if you could do the same conversion in a few microseconds, you could burn that same city to the ground.

Nobody knew at the time how either of these transformations could be done in practice. But within a few decades the outlines began to form. All three belligerents in the second World War embarked on urgent crash programs to develop such weapons, with the explicit aim of using them on each other. An entire world vanished forever

⁵Remarkably, the resulting Standard Model’s interactions all look a *lot* like electrodynamics.

⁶The first complete, general derivation appears to be due to Max von Laue in 1911.

⁷Einstein was up to date: Rutherford/Barnes and Soddy/Ramsey had measured the energy of a single decay of radon just two years earlier (1903), finding it to be over a million times the energy released when hydrogen and oxygen combine to form a molecule of water. Just two years later, Einstein wrote: “Bodies whose energy content is variable to a high degree, for example radium salts,” may perhaps be used to test his prediction about the mass-energy equivalence. Then in a laconic, eerily prescient remark in 1907, he wrote “It is possible that radioactive processes may become known in which a considerably larger percentage of the mass of the initial atom is converted into radiations. . . than is the case for radium.”

⁸Cockcroft and Walton, 1932: ${}^7\text{Li} + p \rightarrow 2\alpha + 14\text{ MeV}$. The masses of all the participating particles were measured and mass defect times c^2 was found to agree with the observed increase in total kinetic energy. For a modern measurement with precision $4 \cdot 10^{-7}$, see S Rainville et al., *Nature* **438**:1096(2005).

on 16 July, 1945.

30.3 PARTICLES WITH SPEED AT OR NEAR c

Suppose that a particle's speed approaches c , that is, suppose $\beta \rightarrow 1$. In this limit, we expect Newton's formulas to be badly inaccurate. In this situation, Equations 30.6, 30.7 and 30.13 give

$$\frac{p}{\mathcal{E}} = \frac{m dx/d\tau}{cm d(ct)/d\tau} = c^{-2} \frac{dx}{dt} \rightarrow \frac{1}{c},$$

or

$$\mathcal{E} \approx pc. \quad (30.15)$$

This is precisely the relation that we found earlier for energy and momentum *fluxes* of a classical plane wave solution! So a dual, quantum-mechanical interpretation of light seems possible after all: The “missing” factor of 1/2 that we noticed earlier is actually just as it should be.⁹ What was wrong was the expectation that newtonian formulas should apply to things moving at speed c .

You may object that as $\beta \rightarrow 1$, our formula for $\gamma \rightarrow \infty$, and hence also the momentum becomes infinite! Indeed, there is no way to push an ordinary particle (one with nonzero mass) up to speed c . However, we can imagine a limit in which $\beta \rightarrow 1$ and $m \rightarrow 0$ in just such a way that $p \rightarrow$ constant:

The only way for a particle to move at speed c is for it to be massless. The only way for a massless particle to have nonzero energy and momentum (and hence to exist at all) is for it to be moving at c . We can take the limit in various ways, so any values of p and \mathcal{E} are allowed, as long as $\mathcal{E} = pc$.

So that's another viewpoint on why light always moves at a universal speed. The dual particle and wave viewpoints are compatible, at least insofar as kinematics is concerned. It's no accident that Einstein's light-quantum and relativity papers both appeared the same year.

Another objection to the particle viewpoint was that “If light has a dual character as a stream of particles, then light from a binary pulsar would move faster, and hence arrive earlier at Earth, when the pulsar is approaching us than when it is receding.”¹⁰ We have already disposed of that objection, however: According to the velocity addition formula, a massless particle emitted at speed c in the rest frame of the moon *also* moves at speed c in the E-inertial frame in which Earth is at rest (our lab frame).

The newtonian conservation laws allow us to predict the results of collisions among, say, two balls that collide and stick. Similarly, our proposed Lorentz-invariant conservation laws allow us to make a falsifiable, quantitative prediction for the result when, say, an x ray photon collides with an electron at rest. The successful test of this **Compton scattering** process lent credence not only to the photon hypothesis, but also to relativity itself.

⁹See Section 18.2 (page 220).

¹⁰Section 25.6.6 introduced this puzzle.

30.4 PLUS ULTRA

This concludes our study of “low-tech relativity.” Although the structure is logically satisfying, I have tried to make it clear that Einstein’s version of relativity is justified because it makes predictions for real experiments (not just thought-experiments). Those predictions were confirmed, and differed from the corresponding newtonian predictions.

We are starting to see something remarkable: The four coordinates (ct and spatial position \vec{r}) undergo a peculiar kind of linear transformation, a little like rotations. And now we see that \mathcal{E}/c and \vec{p} undergo *the same* peculiar but linear transformation (Equation 30.8). This observation suggests that there may be a tensor formalism describing such quantities, and other more elaborate ones. Just as 3-tensor notation helped us to classify quantities and formulate rotationally-invariant laws of Nature, so we will find that **4-tensor** notation will help us to deal systematically with the consequences of the hypothesis that Nature is Lorentz-invariant. Briefly, we will set up a parallel between:

3D euclidean geometry: Cartesian coordinates are the ones in which the pythagorean formula takes its usual form. All cartesian coordinate systems are related to each other by euclidean motions (translations and rotations, plus reflections). Three-tensors have definite, linear transformations under rotations. Every physical quantity in newtonian physics belongs to (is a component of) some 3-tensor. Any law of physics that sets a 3-tensor equal to zero, such as Equation 30.1, is automatically rotation-invariant.

and

4D spacetime geometry: E-inertial coordinate systems are the ones in which the invariant interval has its usual form. All E-inertial coordinate systems are related by Poincaré transformations (translations, rotations, and Lorentz boosts, plus reflections). Four-tensors have definite, linear transformations under Lorentz transformations. Every physical quantity in true (Lorentz-invariant) physics belongs to (is a component of) some 4-tensor. Any law of physics that sets a 4-tensor equal to zero, such as Equation 30.14, is automatically Lorentz-invariant.

The second of these viewpoints will prove extraordinarily helpful as we get to work proving that the full Maxwell equations are Lorentz-invariant, and it will also have practical benefits for solving harder problems than the ones we’ve done so far.

PROBLEMS

30.1 *Proper time*

Section 30.2.1 claimed that the trajectory of any material particle (that is, not a photon) admits a convenient parameterization by an invariant quantity called proper time. This claim is supposed to hold even in the full three spatial dimensions, and even for particles that are not free, that is, particles that are being accelerated by some force. You can establish it as follows.

Suppose that we are given a trajectory specified by four functions $t(\xi)$ and $\vec{r}(\xi)$. The parameterization is arbitrary, except that time t is strictly increasing as a function of ξ . To be physical, the trajectory must always be moving with speed less than c , or in other words $\|\mathrm{d}\vec{r}/\mathrm{d}\xi\|^2 < (c\mathrm{d}t/\mathrm{d}\xi)^2$ everywhere. Show how to obtain a new parameter τ (an increasing function of ξ) that gives the property

$$-(c\mathrm{d}t)^2 + \|\mathrm{d}\vec{r}\|^2 = -c^2(\mathrm{d}\tau)^2.$$

30.2 *Recoilless emission*

In this problem you'll study a phenomenon that is the basis for many extremely accurate measurements, because it involves emission and absorption spectra that are fantastically narrow. In its most common implementation, the unstable nuclide ^{60}Co decays in two steps to an excited state of ^{57}Fe , which then drops to the ground state emitting a photon with energy 14.4 keV. The half-life of this decay is long, so the natural width of the spectral line, set by the Uncertainty Relation, is very narrow: The fractional width $\Delta E/E \approx 3 \cdot 10^{-13}$. Conversely, the absorption line for ^{57}Fe to get excited by an incoming photon is equally narrow.

An isolated nucleus will give off a photon with reduced energy, because some of ΔE must go to the kinetic energy of the recoil of the ^{57}Fe nucleus. Remarkably, however, for cobalt atoms in a crystal lattice there is a significant probability that the final state will involve bulk motion of the *entire crystal*, not just the one nucleus that decayed. The mass of the entire crystal is essentially infinite, so the kinetic energy of the final state is essentially unchanged; no energy is lost to recoil, and the outgoing photon gets the entire 14.4 keV. This is "recoilless emission."

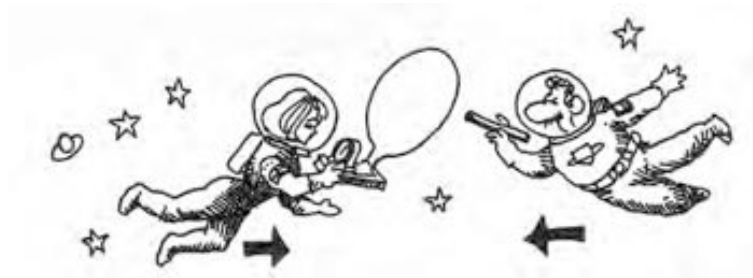
- The mass of an ^{57}Fe nucleus is 56.9 Da. Find the recoil kinetic energy if the iron nucleus is isolated. A convenient definition of the dalton is $1 \text{ Da} = 931.5 \text{ MeV}/c^2$.
- What is the corresponding fractional reduction of the energy of the photon, as compared with the energy of a photon emitted under recoilless conditions?
- Could the photon emitted by a free nucleus be reabsorbed by another nucleus?

30.3 *Constant force*

A particle of charge q and mass m , initially at rest, is released in a region of uniform electric field directed along the \hat{x} axis. Find the subsequent motion $x(t)$. Be sure to check that in the nonrelativistic limit your solution has the expected form.

PART V

Relativity: High Tech



“The traditional conceptions of electricities that attract and repel each other, and that are endowed with actions-at-a-distance as with almost spiritual properties—with these we are all familiar, and in a way we are fond of them.... Electric and magnetic attractions followed the same law as gravitational attraction; no wonder men thought the simple assumption of action-at-a-distance sufficient to explain these phenomena.... [T]hings changed in [our 19th] century, when the reactions between electric currents and magnets became known.... It became necessary to increase the number of actions-at-a-distance, and to improve their form. Thus the conception [of action-at-a-distance] gradually lost its simplicity and physical probability....

“Faraday undoubtedly heard it said that when a body was electrified something was introduced into it; but he saw that the changes that occurred only made themselves felt outside and not inside the body. Faraday had learned that forces simply acted across space; but he saw that an important part was played by the particular kind of matter filling the space across which the forces were supposed to act. Faraday read that electricities certainly existed, whereas there was much contention as to the forces exerted by them; but he saw that the effects of these forces were clearly displayed, whereas he could perceive nothing of the electricities themselves. And so he formed a quite different, opposite conception of the matter. To him the electric and magnetic forces became the actually present, tangible realities; to him electricity and magnetism were the things whose existence might be disputed. The lines of force, as he called the forces independently considered, stood before the eye of his intellect as states of space, as tensions, vortices, currents, whatever they might be—this he himself was unable to determine—but there they were, acting on each other, pushing and pulling bodies about, spreading themselves around and carrying the disturbance from point to point.” – Heinrich Hertz, 1889

CHAPTER 31

Four-vectors and -tensors

In the fall of 1943 [Julian Boyd asked] Einstein to give the manuscript of the June paper to the Book and Authors War Bond Committee as a contribution to the sale of war bonds. Einstein replied that he had discarded the original manuscript after its publication but added that he was prepared to write out a copy of its text in his own hand.... Helen Dukas told me how the copy of the June paper was produced. She would sit next to Einstein and dictate the text to him. At one point, Einstein laid down his pen, turned to Helen and asked her whether he had really said what she had just dictated to him. When assured that he had, Einstein said, ‘Das hatte ich einfacher sagen konnen.’ [‘I could have said that more simply.’]

— *Abraham Pais*

This chapter begins developing what one might call “high-tech relativity.” All your life, profs have been withholding this vital information from you on the dubious premise that you’re “not ready yet.” Now you’re ready.

This chapter will rediscover some results already seen in the preceding Parts III–IV. Why repeat?

- The high-tech approach unifies various ideas that may have seemed disconnected previously. Before we press on to new results, it is important to see how compactly we can regenerate the old ones.
- But the high-tech approach is abstract. Physical intuition was better served by seeing first what could be seen from the older viewpoint, and by building that viewpoint based on a few key experiments.

31.1 HOW TO AVOID READING THIS CHAPTER AND THE NEXT ONE

We are studying the system of Maxwell’s equations for fields, plus the Lorentz force law for charged particles. We have seen that these equations correctly describe many phenomena.

We abstracted Lorentz invariance from just a *simplified subset* of these equations (the scalar wave equation). We got some experimentally testable predictions (Fizeau experiment, aberration of starlight, mass–energy equivalence.) But so far we neglected the vector character of the fields, and hence also polarization of light. We now want to build a bridge between the equations and the hypothesis of Lorentz invariance. To do this, we’ll construct a grammar of Lorentz-invariant constructs, that we can then stick

together (following some grammatical Rules) to build equations that are guaranteed to be Lorentz invariant. Then we'll see if the Maxwell equations can be expressed in that way.

I expect you to read and work through this chapter. However, nothing stops us from considering the *hypothetical* student who wants the plot spoilers up front.

Up and down indices

This chapter develops the modifications to tensor analysis needed to make relativistic invariance obvious at a glance in equations of motion, just as ordinary vector/tensor notation makes rotational invariance obvious at a glance. A key complication is that we will need to keep track of *two kinds of coordinate index*, which will be called “up” and “down” indices. Why, when all your life one kind has been sufficient?

The answer will turn out to be that derivatives ($d/d(ct)$ and $\vec{\nabla}$) transform differently from coordinates (ct and \vec{r}). In euclidean 3-space, if we use cartesian coordinates, then we can forget about the distinction. In the *non*-euclidean space that we'll develop for relativity, we do have to keep track of it.¹

Luckily, we'll find a set of notational Rules that will make it unnecessary to think much about this complication. Once we've justified the Rules, we'll see they are easy to follow. You could, hypothetically, just jump to Section 33.4.

Chicken and egg

We have accumulated some evidence that a new group of transformations may be symmetries of electrodynamics, and indeed of all of Physics. But now we seem to face a chicken-and-egg problem: How can we prove that the Maxwell equations are invariant under these transformations, when we don't know how the \vec{E} and \vec{B} fields should transform? The thought-experiment about the coil and magnet has suggested that under a boost transformation the components of electric and magnetic fields should *mix* (Hanging Question #A). It sounds complicated. Once we make the right guess we can confirm it by mathematical operations... but how do we make the right guess?

Thinking back, the structure of electrodynamics as I have presented it is that we took the Lorentz force law as a starting point; it gave an operational meaning to \vec{E} and \vec{B} . Once those vector fields were defined, then the Maxwell equations make falsifiable predictions about their relations to each other and to charges and currents. So Section 32.2 will again begin with the Lorentz force law, asking:

1. Can it be formulated (perhaps with modifications that are small in the world of slowly-moving objects) in a way that is Lorentz-invariant?
2. If so, what does that say about the transformation properties of \vec{E} and \vec{B} ?
3. Are the Maxwell equations really invariant under those transformations?

The plot spoiler is that the answers are:

1. Yes. All we need to do is substitute relativistic momentum for newtonian momen-

¹Even in euclidean space, if we use curvilinear coordinates, the distinction matters, which is why we have done all our tensor analysis in cartesian coordinates. (In the curved spacetime of general relativity, it matters even more.)

tum.

2. *The electric and magnetic fields together form a single 4-tensor field.* When we transform to a new inertial coordinate system,² the components of \vec{E} and \vec{B} scramble among themselves, just as the components of the quadrupole tensor in electrostatics mix under rotations. We are going to make this analogy precise.
3. Yes. No tweaks will be needed.

You could, hypothetically, jump to Equations 32.2 and 32.5 to see how it works.

31.2 3D PRELUDE

First let's review some material introduced in Chapters 12–13.

31.2.1 Rotations

Here are some things we've already discussed. The components of a 3-vector \vec{r} , referred to a particular cartesian coordinate system, are three numbers \vec{r}_i , $i = 1, 2, 3$. These numbers *represent* the vector, which is itself a geometrical object.

When we change to another cartesian coordinate system, the same vector is *represented* by three different numbers \vec{r}'_a , where³

$$\vec{r}'_a = S_{ai}\vec{r}_i \quad (\text{and } t = t'). \quad (31.1)$$

The matrix S is a set of nine constants. Any 3-component quantity that transforms in this way will be called a **3-vector**, or **3-tensor of rank 1**. Again, prime denotes a new coordinate system. For extra clarity, I will often use coordinate indices i, j, \dots from the middle of the alphabet for one coordinate system, but a, b, \dots from the start of the alphabet for the alternative coordinate system.

The matrix S is not arbitrary: The fact that both coordinate systems are cartesian implies that the pythagorean formula has the same form in each:⁴

$$\|\vec{r}'\|^2 = \vec{r}'_a \vec{r}'_a = S_{ai}\vec{r}_i S_{aj}\vec{r}_j. \quad (31.2)$$

It will sometimes be convenient to use the mathematician's matrix notation. We write vectors and matrices with square brackets, omit explicit indices, and imply summations with the usual rules of matrix multiplication. Thus Equation 31.2 becomes

$$[\vec{r}']^t [\vec{r}'] = [\vec{r}]^t [S^t S] [\vec{r}] \quad (31.3)$$

This notation is very compact, but you have to be careful about the order in which you write things.

²Now that we have left newtonian physics far behind, we will abbreviate "E-inertial" as just "inertial."

³ S is set in sans-serif to remind us it's a matrix. But it doesn't get any arrow because it's not a tensor: Instead of having a tensorial transformation rule under change of coordinates, it *specifies* a change of coordinates.

⁴The pythagorean formula *doesn't* have this form in curvilinear coordinates, but we will stick to representing tensors in cartesian coordinates.

The expression in Equation 31.3 will equal $\vec{r}_i \vec{r}_i = \|\vec{r}\|^2$, for any \vec{r} , only if \mathbf{S} has the property⁵

$$\boxed{[\mathbf{S}^t \mathbf{S}] = \mathbf{1}. \quad \text{for a rotation matrix}} \quad (31.4)$$

Both sides of Equation 31.4 are *symmetric* matrices, so it amounts to *six* independent constraints on the nine entries of \mathbf{S} . Therefore we expect a family of solutions with $9 - 6 = 3$ parameters—for example, the Euler angles used to specify a rotation.⁶ Note that if \mathbf{S} and \mathbf{T} both satisfy Equation 31.4, then so does the product \mathbf{ST} (and also \mathbf{S}^{-1}): We say that rotations close into a **group**.⁷

31.2.2 3-vectors

Other sets of three numbers have the same transformation rule when viewed in a rotated coordinate system as \vec{r} . For example, the time derivatives $d\vec{r}/dt$ and $d^2\vec{r}/dt^2$ are also 3-vectors, because rotations don't affect time. Now consider Newton's law for a harmonic oscillator with viscous friction:

$$m(d^2\vec{r}/dt^2) = -k\vec{r} - \zeta(d\vec{r}/dt). \quad (31.5)$$

Let's multiply everything from the left by \mathbf{S} :

$$\mathbf{S} \cdot m(d^2\vec{r}/dt^2) = -\mathbf{S} \cdot (k\vec{r} + \zeta(d\vec{r}/dt)) = 0.$$

We can push the constant matrix \mathbf{S} inside the derivatives:

$$m(d^2\vec{r}'/dt^2) = -k\vec{r}' - \zeta(d\vec{r}'/dt).$$

This shows that Equation 31.5, reexpressed in the primed coordinate system, retains its original form: It's invariant under rotations.

A bit more precisely, we got rotational invariance under the assumption that t , m , k , and ζ were all unaffected by the rotation: They are **scalars**, also called **3-tensors of rank zero**. Of these, m , k , and ζ are scalar *constants*, whereas t is a scalar variable.

31.2.3 3-tensor transformation rule

Next, consider an anisotropic, but still linear, system of springs (this time without friction). There is a coordinate system for which every allowed motion is a solution to the equation

$$m(d^2\vec{r}/dt^2) = -\vec{K} \cdot \vec{r}. \quad (31.6)$$

Here \vec{K} is a 3×3 matrix of constants. Multiply everything from the left by a rotation matrix:

$$\mathbf{S} \cdot m(d^2\vec{r}/dt^2) = -\mathbf{S} \cdot \vec{K} \cdot \vec{r} = 0$$

⁵Mathematicians call such matrices **orthogonal**, and call the group of all such 3×3 matrices $O(3)$.

⁶There are additional solutions to Equation 31.4 corresponding to reflections, which we won't study. Here we're interested in rotations, which have the additional property $\det \mathbf{S} = +1$. Because the determinant of an orthogonal matrix must always equal ± 1 , this additional restriction doesn't reduce the number of parameters in the space of solutions; it's still three. Mathematicians call this subgroup of $O(3)$ the "special orthogonal" matrices, or $SO(3)$.

⁷Note also that if $\det \mathbf{S} = \det \mathbf{T} = +1$ then $\det(\mathbf{ST}) = +1$.

$$m[d^2\vec{r}'/dt^2] = -[\mathbf{S}\vec{K}\mathbf{S}^{-1}][\vec{r}'].$$

The new version has the same form as the original equation, albeit with a modified spring constant matrix:

$$\vec{K}'_{ab} = S_{ai}(S^{-1t})_{bj}\vec{K}_{ij}.$$

We have reverted to explicit-index notation, so that we can write the factors in any order we please. This formula simplifies when we recall that $\mathbf{S}^{-1} = \mathbf{S}^t$ for a rotation matrix (Equation 31.4):

$$\vec{K}'_{ab} = S_{ai}S_{bj}\vec{K}_{ij}. \quad (31.7)$$

Any 9-component quantity with this transformation behavior is called a **3-tensor of rank 2**. We say that one copy of \mathbf{S} “acts on” each index of \vec{K} . In this case, we have discovered a “spring tensor” governing the restoring force.

The dyad product $\vec{r}\vec{r}$ is another example of a 3-tensor of rank 2, because each factor separately contributes an \mathbf{S} .⁸ More generally, we can define 3-tensors of any rank p : They are represented by collections of 3^p numbers, with a transformation law involving p copies of \mathbf{S} .

Back to the spring system, we can say more if we know more details. For example, suppose that our mass is suspended between three springs stretched along the original x , y , and z axes respectively. Then

$$\vec{K} = A_{(1)}\hat{x}\hat{x} + B_{(2)}\hat{y}\hat{y} + C_{(3)}\hat{z}\hat{z},$$

which indeed is a 3-tensor, because each of its terms is separately a 3-tensor.

31.2.4 Symmetric and antisymmetric 3-tensors

A spring constant tensor has the property that $\vec{K}_{ij} = \vec{K}_{ji}$, or in matrix language $[\vec{K}] = [\vec{K}]^t$. The quadrupole moment tensor from Chapter 3 also has this “symmetric” property.

Your Turn 31A

Show that if a tensor is symmetric in one cartesian coordinate system, the same will be true after transformation via Equation 31.7.

Thus the property of being symmetric is *itself* a rotationally-invariant property of a tensor, and hence something that we may legitimately specify without spoiling rotational invariance.

Similar remarks apply to *antisymmetric* tensors, for example, the magnetic field tensor $[\vec{\omega}]$ or the magnetic dipole moment tensor $\vec{\Gamma}$.⁹

Even if a tensor \vec{T} is not symmetric (or antisymmetric), nevertheless its transpose represents a new tensor of the same rank. That tensor can be added/subtracted from the original version to produce the “symmetric/antisymmetric part” of \vec{T} , with components

$$\vec{T}^{[S]} = \frac{1}{2}(\vec{T} + \vec{T}^t), \quad \vec{T}^{[A]} = \frac{1}{2}(\vec{T} - \vec{T}^t) \quad (31.8)$$

respectively. Then $\vec{T} = \vec{T}^{[S]} + \vec{T}^{[A]}$.

⁸Thus the electric quadrupole tensor, and the moment of inertia tensor, are physical quantities specified by 3-tensors of rank 2.

⁹See Sections 14.1 (page 160) and 16.1 (page 187).

31.3 OTHER ROTATIONALLY INVARIANT SYSTEMS IN MECHANICS

31.3.1 Gravitation

Here is another example, mentioned in Section 25.5 (page 274): To study celestial mechanics, we combine Newton’s Second Law with his law of gravitation for a mass M that is anchored at the origin:

$$m(d^2\vec{r}/dt^2) = -\frac{GMm}{r^3}\vec{r}. \quad (31.9)$$

To analyze this equation’s symmetry, begin with the denominator, which involves the invariant function $r = \sqrt{\|\vec{r}\|^2}$ studied in Section 31.2. So the right-hand side of Equation 31.9 is a scalar constant $-GMm$, times a scalar function r^{-3} , times the 3-vector \vec{r} . All together, it’s therefore a 3-vector. Setting it equal to the left side then yields a rotationally-invariant equation, just as in the isotropic harmonic oscillator.

Equation 31.9 assumes that the Sun is fixed in space.

Your Turn 31B

- Write the more general form in which two gravitating bodies (“Sun” and “Jupiter”) are both free in space, and show that the equation is still rotation-invariant.
- Show that expanding the scope of the system in this way (acknowledging that \vec{r}_{Sun} is a dynamical variable) also restores explicit translation invariance. This was hidden in Equation 31.9.

31.3.2 Angular momentum

You can apply the same reasoning to formulas like the rotational Newton law $d\vec{L}/dt = \vec{\tau}$:

$$\frac{d}{dt} \left[\sum_{\ell} \vec{r}_{(\ell)} \times (m_{\ell} \vec{v}_{(\ell)}) \right] = \sum_{\ell} \vec{r}_{(\ell)} \times \vec{f}_{(\ell)}^{\text{ext}}.$$

Both sides transform like pseudovectors because the cross product is defined using the Levi-Civita symbol (which is a tensor despite having constant entries¹⁰), and index contractions (which we already found to be rotation invariant).

31.3.3 Field equations in 3D

We can also discuss field equations in this language, for example, Newton’s gravitational field equation:¹¹

$$\vec{\nabla}^2 \phi_N = 4\pi G \rho_m. \quad (31.10)$$

First notice that the chain rule from calculus gives

$$\vec{\nabla}_i \equiv \frac{\partial}{\partial \vec{r}_i} = \frac{\partial \vec{r}'_a}{\partial \vec{r}_i} \frac{\partial}{\partial \vec{r}'_a} = S_{ai} \frac{\partial}{\partial \vec{r}'_a} = (S^t)_{ia} \vec{\nabla}'_a, \text{ or} \quad (31.11)$$

¹⁰It is a “tensor from Heaven” (Chapter 13).

¹¹See Chapter 1.

$$\vec{\nabla}'_a = S_{ai} \vec{\nabla}_i. \quad (31.12)$$

We again used the fact that $S^{-1t} = S$ for a rotation matrix.

Equation 31.12 is of the same form as Equation 31.1: $\vec{\nabla}$ itself transforms as a vector. More precisely, the gradient of a scalar *function* (like temperature), is a vector *field* (telling us locally which direction to go if we seek higher temperature). *This is the step that will fail in 4D*, requiring us to introduce two kinds of index.¹²

From Equation 31.12, you can prove that $\vec{\nabla}^2 = (\vec{\nabla}')^2$, and hence that Equation 31.10 is rotationally invariant if we take G to be a scalar constant, and the mass density ρ_m and the gravitational potential ϕ_N to be scalar fields.

Your Turn 31C

To practice the concepts, show that, if \vec{V} is a vector field, then

- $\vec{\nabla}\vec{V}$ is a rank-two tensor field; and
- $\vec{\nabla} \cdot \vec{V}$ is a scalar field, that is, an ordinary function.
- What can we say about the tensor whose components are $\vec{\nabla}_i \vec{V}_j + \vec{\nabla}_j \vec{V}_i$?

31.4 SUMMARY: THE RULES IN 3D

It's time to announce something that is generally implicit in physics books, yet crucial to the general comprehensibility of Physics. I'll call it the **Tensor Principle**:

*Physical quantities all seem to arrange themselves into 3-tensors (or 3-tensor fields), in some cases constrained by symmetry or antisymmetry.*¹³

Thus mass and charge are rank-0 tensors (scalars); velocity is a rank-1 tensor (vector); moment of inertia is a symmetric rank-2 tensor, and so on. Temperature is a rank-0 tensor field; \vec{E} is a rank-1 field; and so on. It may have seemed that “pseudo” quantities such as magnetic field, torque, and so on were exceptions, but we saw how they can be repackaged as true tensors; for example, Section 14.1 (page 160) reexpressed magnetic field as an antisymmetric rank-2 tensor field $\vec{\omega}$.

If everything is a tensor, then we should learn the Way of the Tensor. Here are some Rules that you have been implicitly using all your life. It will help us to generalize them if we take a moment to state them out loud. Some were proved earlier in this chapter; others are easy (but worthwhile) to prove now:

- A 3-tensor of rank p can be represented in a particular cartesian coordinate system by a collection of 3^p numbers, with a transformation law involving p copies of S , each “acting on” an index.
- A 3-tensor field of rank p is the same idea, but each entry is a function of \vec{r} .
- Permuting the indices of a tensor yields another tensor of the same rank.
- The sums of corresponding components of two tensors with the same rank yield the components of a new tensor of that same rank.

¹²See Section 33.2.1 (page 361). This step fails even in 3D euclidean space, if we use curvilinear coordinates.

¹³Mathematicians refer to the sort of tensors we are discussing as “linear representations of the group $O(3)$.” There is also a more general concept of tensors suitable for curved (non-euclidean) spaces.

- e. The derivative operator $\vec{\nabla}$ increases the rank of a tensor field increased by 1.
- f. The collection of all products of the components of a rank- p and a rank- p' tensor itself constitutes a rank- $(p+p')$ tensor. For example, the dyad product $\vec{r}\vec{r}$ is rank 2.
- g. The contraction (dot) is an invariant operation that converts a tensor, or tensor field, to another one with rank decreased by 2.
- h. A physics equation of the form $A_{i_1, i_2, \dots} = 0$, where A is a tensor, is rotationally invariant. Hence, the same is true for an equation of the form $A = B$, where both A and B are tensors (or tensor fields) of the same rank. Examples include Equations 31.5, 31.6, and 31.9.
- i. The volume element d^3r transforms to d^3r' under rotations because the Jacobian matrix has determinant 1. Thus, we may convert a tensor field to a constant tensor of the same rank by integrating over all space.

Note that galilean invariance is not as simple as rotations. Diagnosing whether an equation has this important invariance is not just a matter of glancing at its index structure. We won't need to deal with this, however, because we're pursuing the hypothesis that the world is not galilean invariant after all.

31.5 FOUR DIMENSIONS

We want to construct an upgraded tensor analysis in which the inertial coordinate systems in Einstein's version of relativity play a role analogous to the cartesian coordinate systems in 3D. That is, we want a formulation of physics in which invariance under the Lorentz transformations, which take us from one E-inertial coordinate system to another, is an obvious property of the equations of motion. The Lorentz transformations modify both the space and time coordinates describing events. So we introduce a new kind of object that, in a particular inertial coordinate system, is represented by four numbers:

$$\underline{X}^\mu = \begin{bmatrix} ct \\ \vec{r} \end{bmatrix}^\mu = \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}^\mu. \quad (31.13)$$

Here μ is an index that runs over the four values $0, \dots, 3$. Note the conventions:

- . *Time is regarded as coordinate number zero*, or more precisely, $\underline{X}^0 = ct$.
- . *The index indicating which coordinate we're discussing is placed in the upper position*, not lower as we always do in three dimensions. Thus, \underline{X}^1 is the quantity we've been calling x or \vec{r}_1 up till now, etc.¹⁴ (Lower indices will be given a different meaning below.)
- . Instead of overarrows, we'll flag 4-tensor quantities with an underscore.

¹⁴How do we avoid confusion between a vector component index and an exponent? Often context helps with this. But sadly, sometimes even experts do get confused. Unfortunately, I am powerless to change the world's notation, nor should I teach in some personal me-only notation system. "If the Lord Almighty had consulted me before embarking on creation I should have recommended something simpler" (Alphonso X "Alphonso the Wise," 1221–1284).

As in Section 31.2.3, we will sometimes write $[\underline{X}]$ as an abbreviation for \underline{X}^μ (that is, we suppress the explicit index μ) and use the rules of matrix multiplication to imply summations. As in 3D, we regard $[\underline{X}]$ as a column vector: $[\underline{X}]^t$ is the corresponding row vector.

31.5.1 Lorentz transformations and the invariant interval

We are exploring certain linear transformations on the coordinates representing an event (that is, a point in spacetime):

$$\underline{X}'^\alpha = \Lambda^\alpha_\mu \underline{X}^\mu. \quad (31.14)$$

As in 3D, summation over repeated indices (here μ) is implied. As with 3D rotations, the entries of Λ are all *constants*, and so may be pushed past derivatives.¹⁵

It's convenient to introduce an abbreviation: The **metric** $g_{\mu\nu}$ is the matrix of constants¹⁶

$$[g] = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (31.15)$$

Analogously to the condition for a rotation (Equation 31.4), let's consider those special matrices Λ with the property that

$$\boxed{[\Lambda^t g \Lambda] = [g]. \quad \text{defining property of Lorentz transformation}} \quad (31.16)$$

This property is slightly different from Equation 31.4 (page 337), because $[g]$ is not the identity matrix. Please confirm that the Lorentz transformations we found in Chapter 29 obey Equation 31.16¹⁷, for example,

$$\text{Boost along } \hat{x}: [\Lambda] = \begin{bmatrix} \gamma & -\gamma\beta & & \\ -\gamma\beta & \gamma & & \\ & & 1 & \\ & & & 1 \end{bmatrix}; \quad \text{Rotation about } \hat{z}: [\Lambda] = \begin{bmatrix} 1 & & & \\ & \cos\theta & \sin\theta & \\ & -\sin\theta & \cos\theta & \\ & & & 1 \end{bmatrix}. \quad (31.17)$$

The components of \underline{X} , or any other set of four quantities that transform under Lorentz transformation in the same way, are said to constitute a **4-vector**.

Lorentz transformations form a **group**:

Your Turn 31D

Show that, if Λ_1 and Λ_2 both satisfy Equation 31.16, then so does the product $\Lambda_1\Lambda_2$ (and also the inverse $(\Lambda_1)^{-1}$).

¹⁵Also as in 3D, this statement assumes that we work in cartesian coordinates; curvilinear coordinates introduce more complexities. Those complexities must be faced in general relativity, where *no* globally inertial coordinate systems exist, but we won't need this. Also as in 3D, Λ has no underscore because it's not a tensor: Instead of having a tensorial transformation rule under change of coordinates, it *specifies* a change of coordinates.

¹⁶ $[g]$ is the same set of numerical values in any inertial coordinate system. It may not be obvious that g defined in this way is a tensor, but Section 33.3.3 will show that that is true, just as in three dimensions the collection of nine constants δ_{ij} is a "tensor from heaven" (Section 13.2.3).

¹⁷Later, we'll connect our original method of discovering Lorentz transformations to Equation 31.16 (Section 33.2.2).

In fact, any Lorentz transformation can be generated from matrix products of the boosts and rotations (plus reflections). This should not be too surprising: Equation 31.16 sets two symmetric 4×4 matrices equal, so it's ten independent constraints on the sixteen numbers $[\Lambda]$. So we expect a six-parameter family of solutions (because $16 - 10 = 6$). If we consider only transformations very close to the identity, there are indeed a total of six things we can do to a coordinate system: Rotate it (three Euler angles) or boost it (three components of velocity).

Your Turn 31E

For practice, and for future use, check that Equation 31.16 implies the identities:

$$\begin{aligned} [\Lambda^t \underline{g} \Lambda \underline{g}] &= \mathbf{1}, & [\Lambda^t \underline{g}] &= [\Lambda \underline{g}]^{-1}, & [\underline{g} \Lambda] &= [\Lambda^{-1t} \underline{g}] \\ [\underline{g} \Lambda^t \underline{g} \Lambda] &= \mathbf{1}, & [\underline{g} \Lambda^t] &= [\underline{g} \Lambda]^{-1}, & [\underline{g} \Lambda^{-1t}] &= [\Lambda \underline{g}]. \end{aligned} \quad (31.18)$$

[Hint: First notice that $[\underline{g}]^2 = \mathbf{1}$ and $[\underline{g}]^t = [\underline{g}]$.]

Lorentz transformations are nearly as simple as the rotations in Section 31.2.2. For example, Section 30.2.1 found a quantity related to time that really is a scalar. Consider a particle trajectory as a curve in spacetime. For any two nearby points on that curve, define the **invariant interval** as¹⁸

$$\Delta\tau = c^{-1} \sqrt{-(\Delta X)^\mu \underline{g}_{\mu\nu} (\Delta X)^\nu}. \quad (31.19)$$

To show that the invariant interval really is form-invariant under Lorentz transformations, write¹⁹

$$c\Delta\tau' = \sqrt{-[\Lambda \Delta X]^t [\underline{g} \Lambda] [\Delta X]} = \sqrt{-[\Delta X]^t [\Lambda^t \underline{g} \Lambda] [\Delta X]} = \sqrt{-[\Delta X]^t [\underline{g}] [\Delta X]} = c\Delta\tau. \quad (31.20)$$

The invariant interval has units of time. It equals the time that elapses between two events in an E-inertial coordinate system in which both occur at the same position $\vec{r} = \vec{0}$. Chapter 30 called its integral along a trajectory the **proper time**, which is apt²⁰ because that coordinate system would also be the rest frame of an inertial observer who runs from one event to the other and carries a clock to measure the time between the two events.

31.5.2 Other invariant quantities

The idea of invariant interval is so useful that we generalize it. If \underline{Y} is *any* 4-vector (not necessarily a displacement in spacetime), we define the notation $\|\underline{Y}\|^2$ by the formula

$$\|\underline{Y}\|^2 = \underline{Y}^\mu \underline{g}_{\mu\nu} \underline{Y}^\nu. \quad (31.21)$$

This quantity equals $(\underline{Y}')^\alpha \underline{g}_{\alpha\beta} (\underline{Y}')^\beta$; we say it's a **4-scalar**. (The proof is the same as in Equation 31.20.)

¹⁸We already encountered this quantity in Equation 29.6 (page 310).

¹⁹This step is analogous to Equation 31.2. This analogy is the reason that \underline{g} is again called the “metric.”

²⁰In French or German, “propre” or “eigen” respectively can mean “one’s own.”

Similarly, for any two 4-vectors the **invariant inner product** is defined as $\underline{Y}^\mu g_{\mu\nu} \underline{Z}^\nu$. It's also a 4-scalar, analogous to the scalar product ("dot product") in 3D.

A big difference with ordinary geometry, however, is that we can have $\|\underline{Y}\|^2 = 0$ even if \underline{Y} itself is not zero. Any 4-vector with this property is called **lightlike**, because any two points on a light ray's trajectory have such a separation.²¹ If $\Delta\underline{X}$ is the spacetime separation between two events, then we call the three cases $\|\Delta\underline{X}\|^2 < 0$, $= 0$, and > 0 by the names **timelike**, **lightlike**, and **spacelike separation**, respectively. A material particle always moves slower than c , so it will always move to a new spacetime point that is separated by a timelike displacement vector from its original point.

31.5.3 Four-velocity

We can describe the trajectory of a material particle as a parametric curve in spacetime by using proper time as the parameter: $\underline{X}^\mu(\tau)$.²² Because the invariant interval is a 4-scalar (Equation 31.20), the operation $d/d\tau$ does not alter the transformation properties of whatever it hits. Thus, the quantities

$$\underline{U}^\mu = \frac{dX^\mu}{d\tau} \quad (31.22)$$

also form a 4-vector, called the trajectory's **4-velocity** at whatever point we evaluate the derivative. One way to evaluate it is to write the curve with an arbitrary parameter ξ , then compute $\underline{U} = (d\underline{X}/d\xi)/(d\tau/d\xi)$.

Your Turn 31F

Show that the 4-velocity always obeys the identity

$$\|\underline{U}(\tau)\|^2 \equiv -c^2. \quad (31.23)$$

Here is an example: Consider a particle in uniform straight-line motion with speed $v = \beta c$ directed along \hat{x} :

$$[\underline{X}(\xi)] = \begin{bmatrix} \xi \\ \beta\xi \\ 0 \\ 0 \end{bmatrix}; \quad \frac{d}{d\xi}[\underline{X}] = \begin{bmatrix} 1 \\ \beta \\ 0 \\ 0 \end{bmatrix}.$$

Equation 31.19 gives $d\tau = c^{-1}\sqrt{1-\beta^2}d\xi = (c\gamma)^{-1}d\xi$, where²³ $\gamma = (1-\beta^2)^{-1/2}$, and so

$$[\underline{U}] = (d\underline{X}/d\xi)/(d\tau/d\xi) = \begin{bmatrix} c\gamma \\ \beta c\gamma \\ 0 \\ 0 \end{bmatrix}. \quad (31.24)$$

Your Turn 31G

Confirm that Equation 31.23 holds, starting from Equation 31.24.

²¹Some authors use the synonym **null separation** for lightlike.

²²We can't use this strategy for the trajectory of a light pulse, because $d\tau \equiv 0$ everywhere along a lightlike curve.

²³We previously obtained this in Equation 30.11 (page 327).

31.5.4 Summary and first payoff

This material has been pretty abstract. But unlike a lot of subjects, where “in theory it’s easy but not in practice,” in this case it’s the other way round! For many purposes, all you need to remember is

Position \underline{X} has an upper index, and hence so does its derivative \underline{U} . The constant matrix \underline{g} as we have used it so far has two lower indices. Keep calm and only contract upper with lower indices. If you feel an urge to contract upper with upper, you may be missing a \underline{g} matrix. (31.25)

For example, if you forget the $[\underline{g}]$ factor in Equation 31.19, the rule (31.25) will quickly alert you.

Here is another example. When we discussed plane waves in Chapter 29, we found ourselves manipulating the phase expression $-\omega t + \vec{k} \cdot \vec{r}$. Notice that this expression can be compactly written as $\underline{k}^\mu \underline{g}_{\mu\nu} \underline{X}^\nu$, where the **4-wavevector** is defined as

$$\underline{k}^\mu = \begin{bmatrix} \omega/c \\ \vec{k} \end{bmatrix}^\mu = \begin{bmatrix} \omega/c \\ k_x \\ k_y \\ k_z \end{bmatrix}^\mu. \quad (31.26)$$

The virtue of this reformulation is that it tells how \underline{k} must transform. The invariance of the inner product says that

$$\underline{k}'^\alpha \underline{g}_{\alpha\beta} \underline{X}'^\beta = \underline{k}^\mu \underline{g}_{\mu\nu} \underline{X}^\nu, \text{ where } \underline{k}'^\alpha = \Lambda^\alpha_\mu \underline{k}^\mu. \quad (31.27)$$

Thus, the same wave, viewed in the new coordinate system, has a phase function of the same form (that is, linear) but with modified 4-wavevector, and \underline{k} *transforms as a 4-vector*.

Your Turn 31H

- Show that this compact statement contains our previous results about the aberration of starlight and both kinds of Doppler shift (Section 29.6).
- Also show that the fact that light travels at speed c can also be expressed by the formula $\|\underline{k}\|^2 = 0$.

Besides being pretty, that last formula is manifestly Lorentz invariant, as it must be—we designed Lorentz transformations precisely to maintain the speed of light in every inertial coordinate system.

31.6 MOMENTUM AND ENERGY REVISITED

With the framework we have developed, we can elegantly restate our earlier proposal for relativistic energy and momentum²⁴ as

$$\underline{p} = m\underline{U}. \quad \text{four-momentum} \quad (31.28)$$

²⁴See Section 30.2.1 (page 325).

Thus, \underline{p}^0 is a particle’s energy/ c and \vec{p} is its momentum. The mass m is a 4-scalar, a single number characterizing the particle. Because \underline{U} transforms as a 4-vector, and the mass is a 4-scalar, therefore the proposed formula for four-momentum is also a 4-vector. That is, *unlike Newton’s formula, it has a linear transformation law under Lorentz boosts*. \underline{p} has the same units as the newtonian momentum; indeed, combining Equations 31.24 and 31.28 gives that $\underline{p}^0 = mc\gamma$ and $\underline{p}^i = m\gamma\vec{v}_i$.

With this definition, Einstein’s proposed conservation law says

$$\sum_{\ell} \underline{p}^{(\ell,\text{in})} = \sum_{\ell} \underline{p}^{(\ell,\text{out})}. \tag{31.29}$$

Certainly if that formula is true in any one inertial coordinate system, it will take the same form in any other one, by an argument like the one we applied to Equation 31.5: Both sides transform as 4-vectors.

In short, the distinction between energy and momentum has now melted away (apart from a factor of c). They are parts of a single 4-vector.

31.6.1 Aside on quantum mechanics

We also saw earlier that frequency and wavevector can be combined into a quantity that transforms as a 4-vector (Equation 31.27). So when de Broglie proposed that a particle (for example an electron) has a dual nature as a wave, and related those two viewpoints by

$\underline{p} = \hbar\underline{k}, \quad \text{Einstein/de Broglie relations}$

(31.30)

he didn’t need to check that this proposed law of Nature is Lorentz invariant: Viewed in another inertial coordinate system, it says $\underline{p}' = \hbar\underline{k}'$, with the same numerical value of the constant of Nature \hbar .

The 0-component of Equation 31.30 is Einstein’s relation, $\mathcal{E}/c = \hbar\omega/c$, whereas the other components are de Broglie’s $\vec{p} = \hbar\vec{k}$. In fact, this logic is what led de Broglie to his (then outlandish) prediction that electrons should have wavelike properties with wavelength given by $2\pi\hbar/\|\vec{p}\|$, just as photons do. De Broglie’s insight is all the more impressive because at that time there was no known candidate for a relativistic wave equation for electrons. It’s another example of “Einstein thinking.”²⁵

31.6.2 $\mathcal{E} = mc^2$

Equations 31.28 and 31.23 imply a relationship between the momentum, energy, and mass of any particle:

$$\|\underline{p}\|^2 = -(mc)^2 \quad \text{or} \quad -(\underline{p}^0)^2 + \underline{p}^i \underline{p}^i = -(mc)^2. \tag{31.31}$$

Our identifications of \underline{p}^0 as a particle’s total \mathcal{E}/c , and the spatial components \underline{p}^i as its momentum, \vec{p}_i , yield the relation

$\mathcal{E}^2 = (\|\vec{p}\|c)^2 + (mc^2)^2.$

(31.32)

²⁵See Section 30.2.1 (page 325).

For a particle at rest, this gives the famous and dangerous result discussed in Section 30.2.3.

For a particle moving slowly, so that $pc \ll mc^2$, we can use a Taylor expansion to get $\mathcal{E} \approx mc^2 + \frac{p^2}{2m} + \dots$, approximately a constant plus the newtonian formula, recovering Equation 30.12 (page 327).

31.6.3 Massless particles

There is another interesting limiting case. For a particle moving *fast*, so that $pc \gg mc^2$, we recover $\mathcal{E} \approx pc$ (Equation 30.15, page 329). For the case $m = 0$, this relation is true regardless of the value of momentum:

$$\mathcal{E} = \|\vec{p}\|c. \quad \text{massless particle} \quad (31.33)$$

You may still be bothered, however: *How can a “real thing” have no mass?* Maybe the following thought experiment will help. Imagine a box whose interior walls are perfect mirrors. Initially there’s no light inside. The box will have some resistance to acceleration (inertia), which we describe by a mass m_{box} . Now I fill the box with lots of light, but otherwise change nothing. The light carries energy, but its net momentum is zero. The relation Equation 31.32 with $p = 0$ and $\mathcal{E}_{\text{tot}} > m_{\text{box}}c^2$ implies that the mass of the light-filled box is greater than the empty box. So the light contributes mass, even though it consists of particles that, taken individually, obey $\mathcal{E} = pc$.

31.6.4 Particle creation and destruction

Prior to 1897, those scientists who believed in the atomic theory of matter (by no means everyone) had a vision in which everything was constructed from about a hundred species of little, hard marbles that had not been created nor destroyed, only rearranged, since the Creation. The birth of atomic and then nuclear physics shook that edifice to its foundations, only to replace it by something rather similar: Atoms had constituents (electrons and nuclei), and the nuclei themselves had constituents (protons and eventually neutrons), but *those* particles were deemed to be little, hard marbles that had not been created nor destroyed, only rearranged, since the Creation.

Just as Einstein had found no scientific necessity for the masses of atomic nuclei to be unchanged in a collision, however, so too there proved to be no reason why their *numbers and types* should not change. If the incoming participants in a collision have sufficient energy, then more participants can exit than entered, created from nothing but that energy. The barrier is especially small to create *massless* particles. Indeed, everybody knew that an excited hydrogen atom can give off light without ceasing to be a hydrogen atom, but initially that process had seemed difficult to imagine from a light-particle point of view. The idea of creation ex nihilo solved that puzzle, and then the much more perplexing puzzle of where the electrons emitted in nuclear beta decay were located prior to the reaction.²⁶ Later, as particle accelerators became available, creation ex nihilo was observed even for massive particles, first electrons and

²⁶Enrico Fermi broke this impasse in 1933, proposing that the electron or positron *did not exist* prior to emission from the nucleus. This article was also the first to use quantized spin-1/2 fields in particle physics, predating Heisenberg by several months.

then everything else. Even without constructing an accelerator, we can see showers of cosmic rays created in the upper atmosphere from a single energetic incoming particle.

Conversely, an electron and positron can mutually *annihilate*, the key process underlying positron emission tomography (PET). The energy equivalent of their combined masses emerges as light.

FURTHER READING

Note that many authors use a different convention that takes \underline{g} to be *minus* the matrix in Equation 31.14. This convention leads to correct results if it is applied consistently. Be sure you know which convention is in force before you take formulas from a book or article.

An older tradition, now deprecated, treats time as an *imaginary* quantity. This desperate, unphysical attempt to make the metric look euclidean leads to endless confusion with quantum mechanics, where complex variables enter legitimately.

PROBLEMS

31.1 *Time for the stars*

Suppose that you receive an invitation to a birthday party on a planet of a distant star. The star is located along the \underline{X}^1 axis of an inertial frame \underline{X}^μ in which Earth is at rest.

You get in your spaceship and accelerate along the \hat{x} direction. Your trajectory is a curve in spacetime. Take a minute to sketch how you think this curve should look in the x - ct plane (and also the trajectory corresponding to the friends and loved ones you left at home.)

Your trajectory can be written in parametric form: $\underline{\Gamma}(\tau)$, where τ is the time you perceive on the ship.²⁷ Section 31.5.3 (page 344) defined four-velocity as $\underline{U} = d\underline{\Gamma}/d\tau$. In the following, we'll use a dot to indicate $d/d\tau$. It will be convenient to define the dimensionless variable $w = \underline{U}^1/c$ and substitute cw for \underline{U}^1 . Equation 31.23 gave a relation that also lets us express \underline{U}^0 in terms of w .

To travel without too much discomfort, you adjust the rockets so that you feel pushed against the rear wall of your ship with a constant force just 1.5 times your normal Earth weight.²⁸ Now translate that requirement into a differential equation for $\dot{\underline{U}}(\tau)$, as follows.

Consider one moment τ_* along your journey. There is an inertial frame \underline{X}'^α in which you are momentarily at rest at τ_* . This is the frame obtained by boosting the unprimed frame by β_*c where

$$\beta_* = \underline{U}^1(\tau_*)/\underline{U}^0(\tau_*).$$

²⁷Of course, you won't see the Sun rise and set, but you could measure τ by the growth of your fingernails, or the number of heartbeats, or a clock you carry with you.

²⁸Don't worry about how the ship is propelled, fuel requirements, etc!

In it, your velocity at τ_* equals zero, and hence your velocity *near* τ_* is increasing from slightly negative to slightly positive.

Even if we don't know the relativistic modification of Newton's law, we do know that physics should reduce to newtonian form when things are moving slowly. So we know that the acceleration at τ_* , measured in the primed frame, should equal your weight on Earth, times 1.5, divided by your mass. Call that quantity $a_0 = 1.5(10 \text{ m/s}^2)$. Thus, we demand of the trajectory that

$$\frac{d}{d\tau} \left[\frac{\underline{U}^1}{c^{-1}\underline{U}^0} \right] \Big|_{t'_*} = a_0. \quad (31.34)$$

Now apply the Relativity Strategy,²⁹ that is, translate Equation 31.34 to the Earth-bound inertial coordinate system. Remember that (i) the Lorentz boost connecting the primed and unprimed systems depends on τ_* , but not on τ (it's not an accelerating system). (ii) Factors like \underline{U}^μ that do depend on τ may be evaluated at τ_* , but not until after the derivatives have been evaluated.

- a. Express Equation 31.34 in terms of the one unknown function $w(\tau)$ and its derivative(s). Specifically show that

$$\dot{w}|_* = \frac{a_0}{c} \sqrt{1 + w_*^2}. \quad (31.35)$$

- b. Equations 31.34 and 31.35 must hold at every τ_* along the acceleration part of the trip; that is, it is a differential equation. Solve it for $w(\tau)$ with appropriate initial condition.
- c. Integrate your answer to (b) to find the actual trajectory $\underline{\Gamma}(\tau)$.

Of course, you don't want to arrive at your destination and crash into it! You must also *decelerate*. So after proper time τ_{mid} , you reverse the engines and accelerate along the $-\hat{x}$ direction, again maintaining a constant force of 1.5 times your normal Earth weight, this time from the front wall of the spaceship, until you come to rest.

- d. Revise your sketch to show the entire journey.
- e. Suppose that your total elapsed time is $2\tau_{\text{mid}} = 1$ year. Find the total distance $\Delta \underline{X}^1$ you've traveled from Earth after carrying out both steps of the outbound journey. Express your answers in light-years.
- f. You spend a couple of hours at the party, then reverse your trip to come home. Thus, upon your return you have aged two years. How much have your friends aged since you last saw them?
- g. Convinced that Earth will soon be rendered uninhabitable by its inhabitants, you organize expeditions to scout other planets, then return home and report. Each spaceship takes a trip like the one above, but this time the round-trip duration is such that the crew ages by 30 years (not 2 years). How big a chunk of our galaxy can you explore in this way? When should we, who stayed behind, expect the scouts to return home to us?
- h. Following (g), take the total distance $\Delta \underline{X}^1$ to the destination and divide by 15 years, obtaining a quantity with dimensions of speed. Make an Insightful Comment about

²⁹Idea 25.12 (page 278).

your answer, then find and calculate some other, more meaningful, quantity with the same dimensions.

CHAPTER 32

The Faraday Tensor

To imagine a language is to imagine a form of life.

—Ludwig Wittgenstein

32.1 A NEW WAY OF THINKING

Prior to Einstein, physicists thought of Physics as a search for the right equations of motion. When they attempted to marry the mechanics of charged particles with EM fields, they got bogged down. Einstein and his successors realized that *invariances* of Nature should be the primary drivers; once the right invariance principle was found, dynamics could then follow along.

To get started, let's propose an upgraded Tensor Principle:¹

Physical quantities all arrange themselves into 4-tensors (or 4-tensor fields), in some cases constrained by symmetry or antisymmetry.

4D Tensor
Principle

(32.1)

If we restrict to rotations only, then every 4-tensor falls into blocks that are themselves 3-tensors; thus Idea 32.1 includes and extends our earlier 3D principle.

So far our evidence in favor of Idea 32.1 is that indeed we found that some quantities obey it:

- The mass m of a point particle is a single, Lorentz-invariant quantity—a 4-scalar. Later we'll also refer to m as a “4-tensor of rank- $\binom{0}{0}$,” because it has no indices of any type.
- The speed of light c is a single, Lorentz-invariant constant of Nature—also a 4-scalar.
- The invariant interval $d\tau^2$ between neighboring events is a 4-scalar as well.
- The time and location of an event have been fused into \underline{X}^μ , which we have called a 4-vector. Later we'll also refer to it as a **4-tensor of rank** $\binom{1}{0}$ because it has one index in the upper position and none in the lower position.
- The frequency and wavenumber of a plane wave have been fused into \underline{k} , which we saw indeed transforms the same way as \underline{X} and hence is also a 4-vector.²
- The energy and momentum of a point particle have been fused into \underline{p} , which again is a 4-vector.³

¹Compare Section 31.4 (page 340). [\[T2\]](#) Quantum mechanics amends this slightly to allow an additional class of quantities called “spinors” (Section 33.4', page 376).

²See Equation 31.27 (page 345).

³See Equation 31.28 (page 345).

The next section will explore whether the electric and magnetic fields also follow the Tensor Principle. First let's review how we have already begun to see that some laws of Nature can usefully be written as relations among 4-tensors:

Wave equation

For plane waves, you showed in Your Turn 31H (page 345) that the wave equation boils down to $\|\underline{k}\|^2 = 0$, a manifestly invariant condition on \underline{k} . (Section 33.2.2 will return to the wave equation itself.)

Momentum conservation

Chapter 30 gave us our first taste of “Einstein thinking”:

- We still expect four conservation laws, even if they're not exactly Newton's.
- What could they be? Instead of trying to tinker with Newton's formulas, start from scratch. The statement that a four-vector quantity is the same before and after a collision is an invariant statement.
- What could that four-vector be? Newton says that both energy and momentum are proportional to an invariant constant, m , intrinsic to the body in question. And $\underline{p} = m\underline{U}$ is a four-vector related to velocity.
- The four quantities $\sum_{\ell} \underline{p}_{(\ell)}^{\mu}$ look like Newton's momentum and (a constant plus) energy, in the case of slowly moving bodies whose masses do not change.
- So that revised conservation equation is a strong candidate for a law of Nature. We then found some experimental confirmation.

Next steps

“Einstein thinking” proved to be powerful, and quickly came to dominate in the search for other new laws. Next, we'll apply it to rediscover the Lorentz force law.

32.2 LORENTZ FORCE LAW

32.2.1 The Faraday tensor

Let's abstract some of structural features of the Lorentz force law, try to guess a reformulation in terms of 4-tensors, and then compare to the pre-Einstein version. For fields that are constant in space and time, it has the general structure (Equation 0.5, page 3):

$$\text{force} = (\text{time rate of change of momentum}) = q(\text{linear function of velocity}),$$

where q is a constant of proportionality intrinsic to a test body.

We can write a formula of this sort involving 4-tensors:

$$\boxed{\frac{d\underline{p}}{d\tau} = q\underline{F}(\underline{U}(\tau)). \quad \text{Lorentz force law, 4-vector}} \quad (32.2)$$

Here q is a 4-scalar constant characterizing a test particle and $\underline{p}(\tau)$ is its 4-momentum at proper time τ . \underline{F} is a linear function that takes a 4-vector and spits out a different 4-vector. We know that in three dimensions such a machine is specified by a rank-two tensor: For example,⁴ the anisotropic spring system studied in Section 31.2.3 had a restoring force given by $-\vec{K} \cdot \vec{r}$. That observation motivates us to define a **4-tensor of rank $\binom{2}{0}$** as a collection of 16 numbers $\underline{F}^{\mu\nu}$ that transforms analogously to Equation 31.7 (page 338):

$$\underline{F}'^{\alpha\beta} = \Lambda^\alpha_\mu \Lambda^\beta_\nu \underline{F}^{\mu\nu}. \quad (32.3)$$

Such an object can be used to specify a linear function via

$$\underline{U} \rightarrow \underline{F}(\underline{U}) \quad \text{where} \quad \underline{F}(\underline{U})^\mu = \underline{F}^{\mu\nu} g_{\nu\lambda} \underline{U}^\lambda. \quad (32.4)$$

We will call $\underline{F}^{\mu\nu}$ the **Faraday tensor**, although Faraday himself never thought of it this way. More precisely, the components $\underline{F}^{\mu\nu}(\underline{X})$ are a collection of functions of space and time, which are to be evaluated along the particle's trajectory in Equation 32.2. That is, \underline{F} is a 4-tensor *field*.

Your Turn 32A

Show that including the \underline{g} factor in Equation 32.4 guarantees that $\underline{F}(\underline{U})$ will transform as a 4-vector, just as $\underline{X}^\mu \underline{g}_{\mu\nu} \underline{Y}^\nu$ transforms as a 4-scalar. Multiplying by the 4-scalar q and setting the result equal to the 4-vector $d\underline{p}/d\tau$ thus gives an invariant equation of motion (Equation 32.2).

At first, Equation 32.2 may not seem promising as a reformulation of the Lorentz force law, however. We were seeking a 4-tensor to accommodate the electric and magnetic fields, which have a total of six components, but the object \underline{F} appearing in Equation 32.2 seems to have $4 \times 4 = 16$ entries!

To make progress, note that \underline{F} is not entirely free. Equation 32.2 says that it specifies a change in \underline{U} , but \underline{U} cannot change in an arbitrary way: Section 31.5.3 pointed out that always $\|\underline{U}\|^2 = -c^2$, a constant. Thus

$$\frac{d}{d\tau} (\underline{U}^\mu \underline{g}_{\mu\nu} \underline{U}^\nu) = 0.$$

Using the product rule gives $2\underline{U}^\mu \underline{g}_{\mu\nu} \frac{d\underline{U}^\nu}{d\tau} = 0$. Equations 32.2 and 32.4 then imply

$$(\underline{U}^\mu \underline{g}_{\mu\nu}) \underline{F}^{\nu\lambda} (g_{\lambda\xi} \underline{U}^\xi) = 0 \quad \text{for any } \underline{U}.$$

That is, \underline{F} must always give us zero when contracted on each of its indices with the same thing. To guarantee that, we must demand that⁵ \underline{F} is an *antisymmetric* 4-tensor of rank $\binom{2}{0}$. This extra condition is itself Lorentz-invariant.⁶

⁴Other examples we studied included electric polarizability and the moment of inertia, which are 3-tensors defining linear, vector-valued functions of 3-vectors (Section 12.3.1, page 141).

⁵The logic is the same as when we interpreted the magnetic field as an antisymmetric 3-tensor (Section 14.1, page 160), because that machine eats a particle's velocity and always yields a force perpendicular to \vec{v} .

⁶The logic is the same as when we argued that the condition that a 3-tensor is antisymmetric is rotation-invariant (Section 31.2.4, page 338). Section 33.3.2 will argue that for 4-tensors, it only makes sense to impose this condition on indices that are all in the same position (in this case, up).

An antisymmetric 4×4 matrix has just *six* independent entries—just what we need to hold the electric and magnetic fields.

32.2.2 Relate to traditional form

We can give those six entries any names we like. Here are some suggestive names for them:

$$\underline{F}^{\mu\nu} = \left[\begin{array}{c|c} 0 & \vec{E}^t/c \\ \hline -\vec{E}/c & 2\vec{\omega} \end{array} \right]^{\mu\nu} = \frac{1}{c} \left[\begin{array}{cccc} 0 & \vec{E}_x & \vec{E}_y & \vec{E}_z \\ -\vec{E}_x & 0 & \vec{B}_z & -\vec{B}_y \\ -\vec{E}_y & -\vec{B}_z & 0 & \vec{B}_x \\ -\vec{E}_z & \vec{B}_y & -\vec{B}_x & 0 \end{array} \right]^{\mu\nu}. \quad (32.5)$$

Here the magnetic field tensor $\vec{\omega}$ is defined by Equations 14.2 or 14.3 (page 161) and $\vec{B} = c\vec{\omega}$. Equation 32.5 can be summarized by⁷

$$\underline{F}^{0i} = -\underline{F}^{i0} = \vec{E}_i/c \quad \text{and} \quad \underline{F}^{ij} = \varepsilon_{ijk} \vec{B}_k, \quad i, j, k = 1, 2, 3. \quad \text{Faraday tensor} \quad (32.6)$$

With these names, the 1-component of the proposed reformulation of the Lorentz force law (Equation 32.2) says

$$\frac{d}{d\tau} p^1 = q(\underline{F}^{10} g_{00} U^0 + \underline{F}^{12} g_{22} U^2 + \underline{F}^{13} g_{33} U^3).$$

Use Equations 30.11, 31.24, and 32.5 to find

$$\gamma \frac{d}{dt} p^1 = q(-(\vec{E}_1/c)(-1)(c\gamma) + \vec{B}_3(+1)\gamma \vec{v}_2 - \vec{B}_2(+1)\gamma \vec{v}_3).$$

Canceling the γ factors shows that this is just the 1-component of the Lorentz force law in its traditional form (Equation 0.5, page 3), modified only by using the relativistic formula for momentum. The other two spatial components work similarly.

In short,

The Lorentz force law, formulated using relativistic momentum, can be compactly stated in 4-tensor form as Equation 32.2. The electric and magnetic fields enter as the components of an antisymmetric rank- $\binom{2}{0}$ 4-tensor via Equation 32.5 or 32.6.

Your Turn 32B

Work out the 0-component of Equation 32.2 in terms of \vec{E} and \vec{B} , and interpret it.

32.2.3 Summary

Like any equation of physics, Equation 32.2 is packed with implicit meaning—a framework established in the preceding chapters. Let's pause to say some of those things explicitly one more time.

⁷The identifications in Equation 32.6 are only valid in a *right-handed*, inertial coordinate system. That restriction is the drawback to describing Nature using \vec{B} . Equation 32.2 doesn't suffer from this restriction.

We imagine some apparatus, with coils, charged plates, spark gaps, whatever, that creates some conditions in a region of vacuum (possibly time-dependent). We imagine interrogating those conditions by shooting in charged test particles and observing their trajectories in some coordinate system. Equation 32.2 claims that those trajectories are always solutions to a set of ordinary differential equations. More precisely, it claims that we can find:

- a coordinate system t, \vec{r} independent of what kind of test particles we use, or their initial conditions, or the apparatus,
- two fixed numbers m, q characterizing each test particle, independent of what apparatus we choose and the initial conditions on the test particle trajectory, and
- six functions $\underline{F}^{\mu\nu}$ on spacetime, depending on the apparatus and coordinate choice but independent of the test particle type or initial conditions,

such that every physical trajectory, in every apparatus, is a solution of Equation 32.2. Although there are many ways to make these choices, there are even more possible apparatuses, trajectories, and test particle types, so the claim has falsifiable content, while at the same time also telling us in principle how to *measure* the Faraday tensor.

What gives us the right to just declare that $\underline{F}^{01} = \vec{E}_1/c$ and so on? Remember, *names are arbitrary*. We could give all six entries different letters of the alphabet if we wished (as indeed Einstein did). Equation 32.5 just assigns names that clarify the connection to our previous form of the Lorentz force law. What's important is that we *consistently* use the same names everywhere (for example, rename \vec{E}_1 as $c\underline{F}^{01}$ both in the Lorentz force law and in the Maxwell equations).

Note that every entry of the Faraday tensor participates in Equation 32.2 in the same way. The asymmetry that bothered us between electric and magnetic fields (Hanging Question #C) was more a matter of unfortunate language than real physics.

32.2.4 On beauty

Any physicist will tell you that Equation 32.2 is “beautiful.” What is beauty?

Opinions differ, but I'd suggest that it's the combination of *surprise* and *inevitability*. We asked for an invariant force law with a general structure, and there was *only one reasonable choice*.

Soon we'll extend this observation to the Maxwell equations themselves. Those ad hoc-looking features (like the minus sign that's hard to remember) aren't real; they are just artifacts of awkward traditional notation. In *good* notation, not only is the Lorentz invariance manifest; also the structure of the equations is will turn out to be rigidly dictated, with no ad hoc features.

“Beauty” also can involve getting something for nothing, because physicists are so cheap (we prefer to say “parsimonious”). Without consciously trying, we wrote a formula (Equation 32.2) that is automatically also invariant under spatial inversions! That is, if you observe the world with a left-handed coordinate frame, and deduce the six functions $\underline{F}^{\mu\nu}$, and I observe with a right-handed frame, and deduce $\underline{F}'^{\alpha\beta}$, then our \underline{F} 's will be related by the inversion matrix and we will both agree that particle motion is described by Equation 32.2, with the same value of q . We need never introduce “pseudo-tensor” quantities like \vec{B} .

T2 Section 32.2.4' (page 359) muses more on beauty in physics.

32.2.5 Better than beauty: an experimental consequence

We have drifted far out into Theoryland. Are there any Electromagnetic Phenomena that can ratify our proposed modification of the Lorentz force law?

We've seen that the manifestly-invariant formula Equation 32.2 reduces to the Lorentz force law as we have been using it, with the *one key modification* that we must use Einstein's formula for momentum on its left side. We can test this modification: When charged particles orbit in a uniform magnetic field ("cyclotron motion"), the naive form of the Lorentz force law predicts that the orbital period will be independent of energy. The corrected form predicts deviations from this behavior as the particles' speed approaches c . Not only is this effect seen experimentally; it also imposes an important practical limitation on the design of cyclotron accelerators.

Your Turn 32C

Work out the correction.

32.3 TRANSFORMATION OF THE FARADAY TENSOR

32.3.1 Electric and magnetic fields mix under Lorentz boosts

It is fun to play with tensors, and nice to have beautiful equations. But finding and confirming the right Lorentz force law has additional benefits. Because Equation 32.2 sets one 4-vector equal to another one, we know that there will be other coordinate systems, related to the first one by Lorentz transformations, in terms of which the same set of trajectories that solve it will *also* solve an equation of the same form, apart from the very specific transformation⁸ of F given by Equation 32.3.

Translating into \vec{E} and \vec{B} language via the dictionary Equation 32.5 or 32.6 then gives another falsifiable prediction about electromagnetism.

Let's just work out one example situation. Suppose that in one coordinate system $\vec{B}_3 \neq 0$ but all other components of \vec{E} and \vec{B} are zero. Suppose also that the primed coordinate system is moving at speed βc relative to the unprimed one, along \hat{x} . Then the components of F' will be given by the matrix product $[\Lambda F \Lambda^t]$, or

$$\begin{bmatrix} \gamma & -\gamma\beta & & \\ -\gamma\beta & \gamma & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \vec{B}_3 & 0 \\ 0 & -\vec{B}_3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \gamma & -\gamma\beta & & \\ -\gamma\beta & \gamma & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\gamma\beta\vec{B}_3 & 0 \\ 0 & 0 & \gamma\vec{B}_3 & 0 \\ \gamma\beta\vec{B}_3 & -\gamma\vec{B}_3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (32.7)$$

The final expression is again antisymmetric, as it must be. Comparing to Equation 32.6, we read off the primed fields:

$$\vec{B}'_3 = \gamma\vec{B}_3; \quad \vec{E}'_2 = -\gamma\beta c\vec{B}_3. \quad (32.8)$$

The second of these formulas illustrates the mixing of electric and magnetic fields upon Lorentz boosts anticipated in Hanging Question #A.⁹ That is, \vec{E} and \vec{B} have

⁸Again, our logic just follows Section 31.2.3 (page 337).

⁹See Section 0.4.1 (page 9).

no separate meaning. They are just bits of some bigger, unified object, the Faraday tensor. Thus, the situation with $\vec{E} = 0$ is *not* a Lorentz-invariant property; it was true in our original frame but not in the boosted one.¹⁰

In his first relativity paper, Einstein somehow managed to find the right transformations in the ugly, mysterious form Equation 32.8, and show that they were exact invariances of the Lorentz force law and Maxwell's equations, all without the benefit of 4-tensor notation. Today, we view them as consequences of the *beautifully* simple Equation 32.3, reexpressed in the awkward, but traditional, symbols. The reformulation of relativity using tensor methods was initiated by Minkowski and developed by many others.

Was it worth the effort? One reply is that most of us would not have been able to see through the algebra to the happy ending had we tried to guess the right transformation law, and prove the invariance, in the old 3D notation. The lucidity we get from 4-tensor notation was also crucial when it was time to invent the more elaborate parts of the Standard Model (general relativity, Dirac spinors, Yang-Mills theory) and beyond (supersymmetry...). Even in electrodynamics, we'll need that clarity in the following chapter to establish the full invariance of Maxwell's equations, and later find the radiation Green function and prove the local conservation of field energy and momentum.

32.3.2 A charge in uniform, straight-line motion

Let's apply what we have learned to find the fields created by a point charge q moving uniformly relative to the lab with velocity $c\beta\hat{x}$. Rather than solve the Maxwell equations with a tricky moving boundary condition, we can apply the Relativity Strategy:¹¹ First solve them in the inertial coordinate system that is itself moving at $c\beta\hat{x}$ w.r.t. the lab. In this system, the problem is easy: A point charge q is at rest.¹² There is no magnetic field, and the electric field is given by Coulomb's law.

For brevity, let's restrict to the xy plane and suppress the z direction from our notation.

Your Turn 32D

Apply the appropriate Lorentz transformation to find that then

$$\vec{E}_x = \frac{\gamma(x - \beta ct)}{(\gamma^2(x - \beta ct)^2 + y^2)^{3/2}} \frac{q}{4\pi\epsilon_0} \quad (32.9)$$

$$\vec{E}_y = \frac{\gamma y}{(\gamma^2(x - \beta ct)^2 + y^2)^{3/2}} \frac{q}{4\pi\epsilon_0}. \quad (32.10)$$

These are complicated formulas,¹³ but note first the ratio

$$\frac{\vec{E}_x(t, \vec{r})}{\vec{E}_y(t, \vec{r})} = \frac{x - \beta ct}{y} = \frac{x - x_*(t)}{y - y_*} \quad \text{where} \quad \begin{bmatrix} x_*(t) \\ y_* \end{bmatrix} = \begin{bmatrix} \beta ct \\ 0 \end{bmatrix}.$$

¹⁰Unless of course *all* fields equal zero.

¹¹Idea 32.3 (page 353).

¹²Remember that charge is a 4-scalar quantity.

¹³P+S §12.5 give a similar derivation.

This ratio determines the direction that \vec{E} points. Thus, at any moment, \vec{E} points along the line of sight from the particle's position *at that time* toward the observer (if $q > 0$).

Think about how remarkable that result is. When we look at a distant charge, we are actually seeing it in the past, due to the finite speed of light. And yet, the electric field at the observer is seen to be directed at the particle's position *at the time of observation*, even though simultaneity between that point and the observation is relative! The reason this can occur is that the electric field vector from the charge's retarded position, which is all that the observer can see, gets *bent* by the Lorentz boost in exactly such a way as to point in the direction from the charge's *current* position at the time of observation.

The magnitude of the electric field is also noteworthy:

$$\|\vec{E}\| = r^{-2}\gamma(1 + (\gamma^2 - 1)\cos^2\theta)^{-3/2}\frac{q}{4\pi\epsilon_0}.$$

This is isotropic when the velocity is small, but peaked around $\theta = \pi/2$ (the equatorial plane) in a way that gets more pronounced the closer β gets to 1. In short,¹⁴

At any time t , $\vec{E}(t, \vec{r})$ points radially outward from the particle's position at that time to the observation point \vec{r} . Its magnitude is nonuniform: It reflects a bunching of field energy into the plane transverse to the particle's velocity. $\|\vec{E}\|$ also falls off as distance to that position squared.

Your Turn 32E

Do a similar calculation to find the \vec{B} field, and describe it in words.

32.4 PLUS ULTRA

It is hard to overstate the importance of symmetry analysis in physics. *All three* of the physical interactions that today are considered to be both fundamental and accepted (electroweak, strong nuclear, and the general theory of relativity) are relativistic field theories that were invented as offshoots of electromagnetism, starting with proposed extensions of its *invariance* properties. (The same is true of all the crazy speculative theories that may one day supplant the Standard Model.) In each case, appropriate tensor analysis had to be created or generalized to assist in writing a field theory whose symmetry was manifest.

[\[T\]](#) Section 32.4' (page 359) gives some more hints about the Standard Model.

¹⁴You'll display the field graphically in Problem 32.4.

T₂ Track 2

32.2.4' More on beauty

Is an idea likely to be true *because* it seems beautiful? Surely not—to think so would be to anthropomorphize Nature. Rather, the role of beauty may simply be that a scientist who is moved by a beautiful idea will follow it to the ends of the Earth, without being overwhelmed by the many red herrings that seem to say the idea contradicts some aspect of reality, nor by the myriad distractions of everyday life.

Why did evolution install this imperative in our brains? Certainly humans are programmed to figure things out, and to make connections; the pleasure we get from using these skills may be reinforcement for a behavior that enhanced our survival in difficult times. We habituate, so we need novelty to keep getting that reinforcement. In science, this means that the most powerful jolts come from unexpected connections that nevertheless carry conviction—the quality called “surprising yet inevitable” earlier. We call that beauty, both in art and in science.

T₂ Track 2

32.4'a Bigger symmetry groups

Invoking the Tensor Principle was another sweeping generalization that we owe to Einstein, Minkowski, and others in that generation. The equations governing strong and electroweak interactions have additional “internal” symmetries under other groups (called SU(3) and SU(2)×U(1) respectively), and *all fundamental particles* are described by quantizing fields that are tensors jointly under the Lorentz group and these additional groups. The tensor structures associated to the extra transformations are called “multiplets”; for example, each flavor of quark consists of a “color triplet” under SU(3); the up- and down-quark color triplets in turn form an “electroweak doublet,” and so on. Leptons such as electron, muon, and tau (and their neutrinos) are all color singlets but some form electroweak multiplets.

Successfully quantizing these field theories required a method that preserves the symmetry. After many false steps, such methods were found, though they still only work if an “anomaly cancellation” condition holds. general relativity has proven to be yet more subtle.

32.4'b Supersymmetry

Finally, there is also an even bigger set of proposed spacetime symmetries, called “supersymmetry,” which includes the Lorentz transformations as a subgroup. Invariance under supersymmetry may or may not be a property of the physical world. Some of the parameters that specify a supersymmetry transformation (analogously to the Euler angles) are not ordinary real numbers, but rather are *anticommuting* variables!

We have seen that the parameters of a symmetry transformation may themselves transform, for example under rotations. The anticommuting parameters of supersymmetric transformations themselves transform as spinors under rotations and other Lorentz transformations.

PROBLEMS

32.1 *Too much of a good thing?*

Section 32.2.1 proposed the equation of motion

$$\frac{dp^\mu}{d\tau} = qF^\mu{}_\nu U^\nu$$

as a manifestly invariant form of the Lorentz force law. But, this is four equations, whereas the Lorentz force law as we initially stated it has only three components. Give a physical interpretation for the “extra” component of the above equation, and explain why we don’t really have to solve four independent equations in three unknown functions $\vec{r}(t)$ defining the particle trajectory.

32.2 *It adds up*

A particle of charge q and mass m , initially at rest, is released in a region of uniform \vec{E} directed along the \hat{x} axis. Find the subsequent motion. Be sure to check that in the nonrelativistic limit your solution has the expected form.

32.3 *Cyclotron motion*

A proton is released into a region of uniform magnetic field (that is, \vec{B} is a constant vector field). Its initial velocity is directed perpendicular to the field. Find the orbital period of the resulting circular motion, in terms of the radius r of the proton’s orbit, its mass m and charge q , and the field strength $|\vec{B}|$. Comment on the small- and large- r limits of your answer (at fixed $|\vec{B}|$).

32.4 *Uniformly moving charge*

Get a computer to evaluate Equations 32.9–32.10 and display it as a vector field plot. That is, find the electric field vector in the xy plane for a charged particle in uniform motion along the x axis. Luckily, \vec{E} lies in the xy plane, so a two-dimensional plot is adequate. Also, you lose no generality if you assume $t = 0$ (why?).

- Show the case $\beta = 0.1$ and comment.
- Show the case $\beta = 0.9$ and comment.

Make sure your computer uses the same scale for the x and y axes, and also that it displays the lengths of each arrow accurately (without rescaling). If you use Python, the *Student’s Guide* §6.9 discusses vector field plots.

32.5 *Induced charge*

A rigid, conducting sphere of radius R moves with constant velocity \vec{v} through a uniform magnetic field \vec{B} . Assume $v \ll c$ and find the surface charge density induced on the sphere to lowest order in v/c .

CHAPTER 33

Manifestly Invariant Form of Maxwell

You boil it in sawdust, you salt it with glue, you condense it
with locusts and tape,
Still keeping the principal object in view: To preserve its
symmetrical shape.

— Lewis Carroll

33.1 FRAMING

The whole discussion in the preceding chapter may seem like outrageous wish-fulfillment—a scam. If we get to assume any transformation law we like for the electric and magnetic fields, then surely we can always arrange for our equations to be invariant?

Surely not. We have completely specified the transformation of our fields just by studying the Lorentz force law. There is no further freedom. Now we must cross our fingers and hope that the Maxwell equations will also be invariant *under the same field transformations*.

This chapter begins by studying fields only, that is, no charges or currents. Then we will construct the notion of charge flux 4-vector, and add it as a source term in our invariant form of Maxwell’s equations.

33.2 FIELD EQUATIONS IN 4D

So far, many of our constructions have closely paralleled the three-dimensional situation. Now one key difference will emerge.

33.2.1 Transformation of spacetime derivatives

Let’s use the abbreviation $\underline{\partial}_\mu$ to mean $\partial/\partial\underline{X}^\mu$. Then proceeding as in Equation 31.11 (page 339) gives

$$\underline{\partial}_\mu = \frac{\partial}{\partial\underline{X}^\mu} = \frac{\partial\underline{X}'^\alpha}{\partial\underline{X}^\mu} \frac{\partial}{\partial\underline{X}'^\alpha} = \Lambda^\alpha{}_\mu \underline{\partial}'_\alpha = [\Lambda^t]_\mu{}^\alpha \underline{\partial}'_\alpha, \quad \text{or} \quad (33.1)$$

$$\underline{\partial}'_\alpha = [\Lambda^{-1t}]_\alpha{}^\mu \underline{\partial}_\mu. \quad (33.2)$$

For example, applying both sides of Equation 33.2 to a scalar field tells us that the 4-gradient $\underline{\partial}_\mu\phi$ of a scalar function is a set of four functions with the transformation rule Equation 33.2. The new wrinkle is that this rule is *different from* the one we started with ($\underline{X}'^\alpha = \Lambda^\alpha{}_\mu \underline{X}^\mu$).¹ This issue did not arise in three dimensions, because

¹Equation 31.14 (page 342).

for rotation matrices $S^{-1t} = S$, whereas $\Lambda^{-1t} \neq \Lambda$ in general.

Thus there are *two fundamental tensor types* in relativity: the ones previously called 4-vectors (or 4-tensors of rank $\binom{1}{0}$), and new ones that transform like Equation 33.2, which are called **4-covectors** (or 4-tensors of rank $\binom{0}{1}$). The rank notation is motivated by the fact that the 4-gradient $\underline{\partial}_\mu \phi$ has one index in the *lower* position.

This doubling of index types causes us surprisingly little trouble in practice, however. Suppose that \underline{W}_μ is a collection of four numbers that constitute a 4-covector. Define $\underline{g}^{\mu\nu}$ to be a 4×4 matrix of constants that is the inverse of the matrix $\underline{g}_{\mu\nu}$.² We now show that the four quantities $\underline{g}^{\mu\nu} \underline{W}_\nu$ amount to a 4-vector. That is, there is a standard way to interconvert between 4-vectors and 4-covectors; if we like, we can do all of our work using only 4-vectors.

To prove the claimed result, substitute the transformation of the 4-covector into the quantities being considered:

$$\underline{g}^{\alpha\beta} \underline{W}'_\beta = [\underline{g} \underline{W}']^\alpha = [\underline{g} \Lambda^{-1t} \underline{W}]^\alpha.$$

Next, use one of the identities in Equation 31.18 (page 343) to rewrite this as

$$= [\Lambda \underline{g} \underline{W}]^\alpha = \Lambda^\alpha_\mu [\underline{g} \underline{W}]^\mu.$$

Thus as claimed $\underline{g}^{\mu\nu} \underline{W}_\nu$ transforms as a 4-vector! It's traditional to name these four new quantities \underline{W}^μ , to emphasize that:

- They are very closely related to \underline{W}_μ , and so deserve to be called by the same letter of the alphabet, but
- Unlike \underline{W}_μ , they transform like \underline{X}^μ (or any other quantity with one upper index).

The process of constructing a 4-vector from a 4-covector by contraction with \underline{g} is called **index raising**.³ Because $[\underline{g}]^2 = \mathbf{1}$, we can invert this operation by *another* multiplication by \underline{g} :

$$\underline{W}_\mu = \underline{g}_{\mu\nu} \underline{W}^\nu. \quad \text{index lowering}$$

It's not hard to find an invariant product for two covectors: Simply convert each to a 4-vector and use the usual product:

$$(\underline{g}^{\mu\nu} \underline{W}_\nu) \underline{g}_{\mu\lambda} (\underline{g}^{\lambda\sigma} \underline{V}_\sigma) = [\underline{W}^t \underline{g} \underline{g} \underline{V}] = \underline{W}_\nu \underline{g}^{\nu\sigma} \underline{V}_\sigma.$$

It's easier still to find the invariant product of a covector and a vector:

$$(\underline{g}^{\mu\nu} \underline{W}_\nu) \underline{g}_{\mu\lambda} \underline{U}^\lambda = [\underline{W}]^t [\underline{g} \underline{g}] [\underline{U}] = \underline{W}_\nu \underline{U}^\nu.$$

No \underline{g} factor at all is needed in this case.

33.2.2 The wave operator

The ideas in the previous section make it straightforward to find a manifestly invariant derivative operator that, when applied to a scalar function, yields another scalar

²In fact, these are two names for the *same* matrix, because $[\underline{g}]^{-1} = [\underline{g}]$, but we nevertheless use different notation for the two different uses, in part because they won't be the same in general relativity, nor even in special relativity with curvilinear coordinates.

³A mathematician might call this operation "taking the adjoint with respect to the inner product \underline{g} ."

function. To define it, it's first convenient to define $\underline{\partial}^\mu$ by raising the index on $\underline{\partial}_\mu$. Then we can construct the Lorentz-invariant operator

$$\square = \underline{\partial}^\mu \underline{\partial}_\mu.$$

It's called the **wave operator**, D'Alembert operator, or **dalembertian**.

Your Turn 33A

Show that \square is the same wave operator that we have been writing all along (Section 23.4, page 261), and whose invariances led us to discover the Lorentz transformations in the first place.

But now we can take another step. If we apply the wave operator to *any* rank tensor, the result is again a tensor of the same rank. Setting that to zero yields a Lorentz-invariant field equation. That observation immediately suggests the candidate equation

$$\square \underline{F}^{\mu\nu} \stackrel{?}{=} 0$$

for electrodynamics! Could it really be that simple? Well, no: The Maxwell equations are only first-order in derivatives. We'll soon find something almost as simple, and correct.

33.3 GENERAL 4-TENSORS

33.3.1 Rank

We can now define a **4-tensor of rank** $\binom{p}{q}$ as a set of 4^{p+q} numbers with p upper and q lower indices, transforming with p copies of Λ and q copies of Λ^{-1t} . Extending the list we started in Section 32.1,

- . The gradient of a scalar function has rank $\binom{0}{1}$;
- . The Faraday tensor has rank $\binom{2}{0}$;
- . The quantities $\underline{F}^{\mu\nu} \underline{g}_{\nu\lambda}$ constitute a 4-tensor of rank $\binom{1}{1}$; and so on.

33.3.2 Symmetry

Let $A^{\mu_1 \dots \mu_p}{}_{\nu_1 \dots \nu_q}$ be a 4-tensor of rank $\binom{p}{q}$.

Your Turn 33B

Show that:

- a. If the components of a tensor \underline{A} are antisymmetric under permutation of some or all of its upper indices in one inertial coordinate system, then \underline{A} will have that same property in any other such system (and similarly for lower indices).⁴ Similarly, if the components are symmetric under permutations, that property, too, is invariant.
- b. Also show that the operation of antisymmetrizing (or symmetrizing) a tensor on some or all of its upper (or lower) indices is invariantly defined.

But beware: There is no invariant sense to (anti)symmetry between an upper and a lower index. We must lower one index, or raise the other, before we can speak invariantly of (anti)symmetry.

33.3.3 The metric is itself a tensor

You now have all the tools to show that the metric is a “tensor from Heaven,” that is, numerically the same when viewed in any inertial coordinate system.⁵

Your Turn 33C

- a. The metric as we first introduced it, $\underline{g}_{\mu\nu}$, has two lower indices. Prove that this matrix indeed gives the components of a constant 4-tensor of rank $\binom{0}{2}$, as implied by the notation. [*Hint*: Use an identity from Equation 31.18 (page 343).]
- b. Section 33.2.1 defined the related symbol $\underline{g}^{\mu\nu}$ as the inverse matrix to $\underline{g}_{\mu\nu}$ (that is, numerically equal to it). Prove that this matrix indeed gives the components of a constant 4-tensor of rank $\binom{2}{0}$, as implied by the notation.

33.4 SUMMARY: THE RULES IN 4D

This is getting scary. What saves us from total confusion is that a few Rules make it unnecessary to think much about these intricate transformations. These Rules correspond to the ones in Section 31.4 (page 340), and are almost as easy to use.

We are exploring the hypothesis that electrodynamics is invariant under Lorentz transformations. To generate Lorentz-invariant equations as candidate laws of Nature, we organize all the dynamical variables into 4-tensors of suitable rank,⁶ where:

- a'. A 4-tensor of rank $\binom{p}{q}$ can be represented in a particular inertial coordinate system by a collection of 4^{p+q} numbers, indexed by p upper and q lower indices, with transformation law appropriate for each index's position.
- b'. A 4-tensor field is the same idea, but each entry is a function of \underline{X} .
- c'. Permuting a set of indices of a tensor, all in the same position (all up or down) yields another tensor of the same rank.
- d'. The sums of corresponding components of two tensors with the same rank yield the components of a new tensor of that same rank.
- e'. The derivative operator $\underline{\partial}$ increases the rank of a tensor field by $\binom{0}{1}$ (see Section 33.2.1).
- f'. The collection of all products of the components of a rank- $\binom{p}{q}$ and a rank- $\binom{p'}{q'}$ tensor itself constitutes a rank- $\binom{p+p'}{q+q'}$ tensor.

⁴In particular, the statement that a tensor is totally antisymmetric is a Lorentz-invariant property, as we saw in an example already (Equation 32.7, page 356).

⁵See Chapter 13.

⁶See Idea 32.1. Mathematicians refer to the sort of tensors we are discussing as “linear representations of the group $O(3,1)$.” They also have a more general concept of tensors suitable for curved (non-euclidean) spacetimes, which is handy when we wish to study electrically charged black holes or the gravitational bending and redshifting of light as it passes by a massive object.

- g'1. Only contract indices in up/down pairs. Such a contraction is invariant; that is, the result is again a tensor, with reduced rank $\binom{p-1}{q-1}$.
- g'2. Whenever we are tempted to contract two upper indices, we instead introduce a factor of the metric. Contracting one upper index with $g_{\mu\nu}$ is an invariant operation that changes the rank from $\binom{p}{q}$ to $\binom{p-1}{q+1}$. Then we contract the resulting new lower index with the other upper index, bringing the rank down to $\binom{p-2}{q}$ as desired.
- g'3. Contracting one lower index with $g^{\mu\nu}$ again yields a new tensor, with rank changed to $\binom{p+1}{q-1}$. Because $[g]^2 = \mathbb{1}$, index raising and lowering are each others' inverse operations.
- h'. A physics equation of the form $A = B$, where both A and B are tensors (or tensor fields) of the same rank, is then guaranteed to be Lorentz invariant.
- i'. The volume element d^4X transforms to d^4X' under Lorentz transformations because the Jacobian matrix has determinant ± 1 .⁷ Thus, we may convert a tensor field to a constant tensor of the same rank by integrating over all spacetime.

With these Rules, 4-tensor manipulations become so automated that most physicists don't consciously distinguish between, say, $\underline{F}^\mu{}_\nu$ and $\underline{F}_{\mu\nu}$; both are called \underline{F} , and only index placement is used to tell them apart. Either one is called the Faraday tensor. If you've got one, but you want the other, then you convert by index raising or lowering operations. But beware: If you plan to use index-free (matrix) notation, you need to state which of these quantities you mean, because they are numerically different. Matrix notation is extremely concise, but for that very reason we will generally avoid it, now that we have established our "grammar" of invariant constructions.

33.5 VACUUM MAXWELL EQUATIONS

We wish to establish that the Maxwell equations have the property of form invariance under Lorentz transformations. But they look pretty complicated; they have some apparently ad hoc minus signs; we found that \vec{E} and \vec{B} have complicated transformation rules under Lorentz transformations. To see through the derivation, let's start from scratch.

Chapters 31–32 explained what "from scratch" could mean, via a new way of thinking, driven by invariance properties. Let's apply that "Einstein thinking" to the Maxwell equations:

- Abstract away from Maxwell's version the structural features: The desired equations are first-order in space and time derivatives. They involve an antisymmetric, rank- $\binom{2}{0}$ tensor field \underline{F} . In addition, half of them involve charges and currents. There are also two scalar constants ϵ_0 and μ_0 , or equivalently μ_0 and $c = (\epsilon_0\mu_0)^{-1/2}$.
- What could the equations be? If they take the form (tensor field) = 0, then the Rules say they'll be automatically invariant (Section 33.4, page 364).

⁷Take the determinant of both sides of Equation 31.16 (page 342).

- Once we have guessed candidate equations that meet the criteria, we can ask how they look when phrased in terms of the old-school \vec{E} and \vec{B} fields. If they coincide with the Maxwell equations as we've been writing them, then we've proved that electrodynamics is Lorentz-invariant.

We could implement the first bullet with the candidate equation

$$\partial_\nu \underline{F}^{\mu\sigma} \stackrel{?}{=} 0, \quad (33.3)$$

but that can't be right. For one thing, it's $4 \times 6 = 24$ equations, because $\mu\sigma$ is an antisymmetric pair, but *we only wanted eight* equations. Worse, we know all about the solutions to those equations: They say that all six components of \underline{F} are *constants*. Too many equations have too impoverished a set of solutions.

But maybe we could reduce the equations without spoiling their Lorentz-invariance. One possibility is to *contract* indices:

$$\partial_\nu \underline{F}^{\mu\nu} \stackrel{?}{=} 0. \quad (\text{in vacuum}) \quad (33.4)$$

The Rules say this formula is still Lorentz-invariant, but now it's just *four* equations, because there's one loose index.

Your Turn 33D

- Rephrase Equation 33.4 in terms of the traditional \vec{E} and \vec{B} using Equation 32.5 (page 354) and confirm that indeed it's *precisely* the electric Gauss law and Ampère's law—there is no need to tweak those equations, which were secretly Lorentz-invariant all along.
- There are three ways to contract two indices in Equation 33.3, and so far we've only considered one. What about the other two ways?

A second reduction of the candidate equation is possible, but a bit more subtle:

Your Turn 33E

Show that taking the totally antisymmetric part of Equation 33.3 gives

$$\partial_\mu \underline{F}_{\nu\lambda} + \partial_\nu \underline{F}_{\lambda\mu} + \partial_\lambda \underline{F}_{\mu\nu} = 0. \quad (33.5)$$

The Rules say that the left side of Equation 33.5 is a tensor, so this equation is Lorentz invariant,⁸ and hence a candidate for a law of Nature.

Equation 33.5 may appear to be $4^3 = 64$ equations, because it has three loose indices. Really, however, most of these equations are vacuous or redundant, because a totally antisymmetric 4-tensor of rank $\binom{0}{3}$ has only four independent components.

⁸Note that before we can invariantly antisymmetrize a tensor, we must push all the indices into matching position, either by raising the lower one or (as done above) by lowering the upper ones.

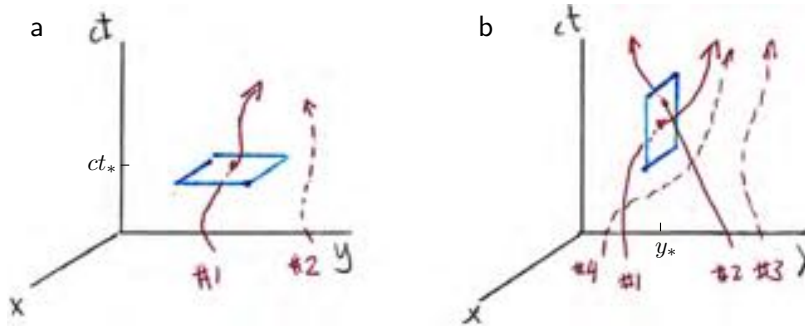


Figure 33.1: Unified construction of (a) charge density and (b) charge flux, an extension of the one in Figure 7.1 (page 82)b. Dashed lines indicate charged particle trajectories that make no contribution because they don't pass through the selected windows. Thus in (a), trajectory #2 may eventually pass through the spatial region shown, but not at time t_* . Similarly, in (b), trajectory #4 passes through the selected range of ct and x , but not at y_* .

Your Turn 33F

Confirm that last claim in general, then write down all four independent components of Equation 33.5. You'll need the expressions obtained by index lowering the identifications we found in Equation 32.5 (page 354):

$$\underline{F}_{\mu\nu} = \begin{bmatrix} 0 & -\vec{E}_1/c & -\vec{E}_2/c & -\vec{E}_3/c \\ \vec{E}_1/c & 0 & \vec{B}_3 & -\vec{B}_2 \\ \vec{E}_2/c & -\vec{B}_3 & 0 & \vec{B}_1 \\ \vec{E}_3/c & \vec{B}_2 & -\vec{B}_1 & 0 \end{bmatrix}_{\mu\nu}. \quad (33.6)$$

Once again, you'll find *precisely* the magnetic Gauss law and Faraday's law—so they, too, were secretly Lorentz-invariant all along.

33.6 THE CHARGE FLUX 4-VECTOR

To complete our job, we need to upgrade Equation 33.4 to include charges and currents. (We already know that Equation 33.5 is complete, because the magnetic Gauss law and Faraday's law don't involve charges nor currents.)

33.6.1 A geometrical formulation

This section repeats the discussion in Chapter 7. For artistic reasons, Figure 33.1 only shows two space dimensions x, y , but z is understood to be present.

We suppose space to contain a swarm of charged particles with trajectories $\underline{X}^\mu = \underline{\Gamma}_{(\ell)}^\mu(\tau)$ for $\mu = 0, \dots, 3$. Each trajectory is a curve in spacetime, parameterized by proper time τ and labeled by a constant q_ℓ (its charge). We choose an inertial coordinate system on spacetime.

To define charge density at some point \underline{X}_* (an “event”), set up a small spatial volume element $\Delta^3 \underline{X}_\perp$, that is,

$$\begin{aligned} ct &= \underline{X}_*^0 = \text{constant}, \\ \underline{X}_*^1 &< x < \underline{X}_*^1 + \Delta \underline{X}^1, \\ \underline{X}_*^2 &< y < \underline{X}_*^2 + \Delta \underline{X}^2, \\ \underline{X}_*^3 &< z < \underline{X}_*^3 + \Delta \underline{X}^3. \end{aligned}$$

In Figure 33.1a, the blue rectangle shown represents $\Delta^3 \underline{X}_\perp$. Now we add up all the charges on lines crossing this element from past to future, divide by volume $\Delta^3 \underline{X}_\perp$, multiply by c , and call the result $\underline{J}^0(\underline{X}_*)$. For example, trajectory #1 contributes $cq_1/\Delta^3 \underline{X}_\perp$, whereas trajectory #2, which misses the volume element, contributes nothing.

Note that the quantity \underline{J}^0 just defined has units coul/(s m²). In fact, \underline{J}^0 is the quantity we’ve previously called $c\rho_q$.

Next, define charge flux at \underline{X}_* by setting up a new small volume element (Figure 33.1b), again called $\Delta^3 \underline{X}_\perp$:

$$\begin{aligned} \underline{X}_*^0 &< ct < \underline{X}_*^0 + \Delta \underline{X}^0, \\ \underline{X}_*^1 &< x < \underline{X}_*^1 + \Delta \underline{X}^1, \\ y &= \underline{X}_*^2 = \text{constant}, \\ \underline{X}_*^3 &< z < \underline{X}_*^3 + \Delta \underline{X}^3. \end{aligned} \tag{33.7}$$

Add up all the charges on trajectories crossing this element from smaller to larger values of y , and subtract all the charges on trajectories crossing it in the opposite sense. Again divide by $\Delta^3 \underline{X}_\perp$, multiply by c , and call the result $\underline{J}^2(\underline{X}_*)$. Thus in the sketch trajectory #1 contributes $cq_1/\Delta^3 \underline{X}_\perp$, #2 contributes $-cq_2/\Delta^3 \underline{X}_\perp$, and #3–4 contribute nothing.

Define the other two components \underline{J}^1 and \underline{J}^3 similarly. Thus, all four components of \underline{J} have the same units. In fact, \underline{J}^i are the three quantities called the charge flux⁹ \vec{j}_i in Section 7.2 (page 81). The advantage of the present formulation is that it treats all four components in the same way. In any inertial frame,

$$\begin{aligned} \underline{J}^\mu &= \text{net amount of charge crossing the surface } \underline{X}^\mu = \text{constant}, \\ &\text{from smaller to larger } \underline{X}^\mu, \text{ per } d^3 X_\perp, \text{ times } c. \end{aligned} \tag{33.8}$$

33.6.2 \underline{J} is a 4-vector

The Tensor Principle claims that all physical quantities can be packaged into 4-tensors. Does \underline{J}^μ defined in the preceding section fit?

Chapter 7 considered a small hypercube and showed that, because charge is locally conserved, we must have

$$\frac{\partial}{\partial t} \rho_q + \vec{\nabla} \cdot \vec{j} = 0. \tag{7.4, page 83}$$

⁹And that some authors instead call the “current density.”

We now can recognize that this continuity equation can be written more elegantly as

$$\frac{\partial \underline{J}^\mu}{\partial \underline{X}^\mu} = 0, \quad (33.9)$$

or more concisely still as

$$\partial_\mu \underline{J}^\mu = 0. \quad (33.10)$$

Our derivation of Equation 7.4 was valid in *any* coordinate system, so in particular the form of Equation 33.10 is the same in any inertial system. We also know that $\underline{\partial}_\mu$ form a covector. Thus the four quantities

$$\underline{J}^\mu(\underline{X}) = \left[\begin{array}{c} c\rho_q(t, \vec{r}) \\ \vec{j}(t, \vec{r}) \end{array} \right]^\mu \quad (33.11)$$

must themselves transform as a rank- $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ field: the **charge flux 4-vector** field.¹⁰

33.7 COMPLETE, INVARIANT MAXWELL EQUATIONS

We are now ready to add charges and currents to Equation 33.4. Once again, there's really *no freedom!* The left side of Equation 33.4 is a 4-vector, so we must set it equal to a 4-vector. We have seen that charges and currents constitute a 4-vector. All we need is a constant of proportionality to make the units work out:

$$\partial_\nu \underline{F}^{\mu\nu} = \mu_0 \underline{J}^\mu \quad \text{and} \quad \partial_\mu \underline{F}_{\nu\lambda} + \partial_\nu \underline{F}_{\lambda\mu} + \partial_\lambda \underline{F}_{\mu\nu} = 0. \quad \text{Maxwell equations} \quad (33.12)$$

Your Turn 33G

Extend Your Turn 33F to confirm that the Equation 33.12 really gives the full Maxwell equations as we have been using them.

The eight beautiful¹¹ new equations, Equations 33.4–33.5, have turned out to be exactly the Maxwell equations we have been using all semester! But their complete Lorentz invariance (and that of the Lorentz force law) is now obvious. Along the way, we have also addressed Hanging Question #B: The form of the equations isn't arbitrary after all, but rather is dictated by general principles. Moreover, no Levi-Civita tensor appears in Equations 33.12; thus, they are also manifestly invariant under inversions, unlike the traditional formulation in terms of \vec{E} and \vec{B} .¹²

[T2] Section 33.7' (page 376) discusses the proper counting of these equations and Hanging Question #D.

¹⁰See also Section 33.9.3 for an explicit proof.

¹¹"Surprising yet inevitable."

¹²This addresses Hanging Question #E. Nor is any choice of right hand buried in the recipe that converted particle trajectories into \underline{J}^μ (Equation 33.8), nor in the one that let us operationally define (measure) \underline{F} (the Lorentz force law, Equation 32.2).

33.8 FOUR-VECTOR POTENTIAL

33.8.1

The second of Equation 33.12, together with the Poincaré Lemma,¹³ implies that we can always write the Faraday tensor in terms of a **four-vector potential**:¹⁴

$$F^{\mu\nu} = \partial^\mu \underline{A}^\nu - \partial^\nu \underline{A}^\mu. \quad (33.13)$$

The known transformation properties of $\underline{\partial}$ and \underline{F} imply that \underline{A} is indeed a four-vector field.

Your Turn 33H

Work out the corresponding \vec{E} and \vec{B} , and show that Equation 33.13 reproduces Equation 17.22 (page 208) if we identify

$$\underline{A}^\mu = \begin{bmatrix} \psi/c \\ \vec{A} \end{bmatrix}^\mu.$$

Thus, the potentials we found long ago also adhere to the 4D Tensor Principle. SI units for the 4-vector potential are $[\underline{A}] \sim \text{kg m}/(\text{coul s})$.

Gauge invariance is the observation that the Faraday tensor doesn't change when we replace \underline{A}^μ by

$$\tilde{\underline{A}}^\mu = \underline{A}^\mu + \partial^\mu \Xi. \quad (33.14)$$

Your Turn 33I

- Prove that last statement starting from Equation 33.13.
- Show that when we substitute Equation 33.13 into Maxwell's equations, one set is vacuous (always automatically satisfied).
- Show that the remaining Maxwell equations become

$$-\square \underline{A}^\nu + \partial_\mu \partial^\nu \underline{A}^\mu = \mu_0 \underline{J}^\nu. \quad (33.15)$$

Your result establishes that the Maxwell equations can be written as four equations in four unknown functions, even though they started as eight equations in six unknowns.

[\[T\]](#) Section 33.8.1' (page 376) discusses the counting in more detail.

33.8.2 Particle in uniform motion revisited

For a first look at the benefits of using potentials, we can return to the problem posed in Section 32.3.2. Again restrict to the xy plane and suppress the z direction from our notation. Also over the next few lines we'll temporarily drop the tiresome $q/(4\pi\epsilon_0)$

¹³See Chapter 14.

¹⁴Chapter 17 already derived this in a less compact way.

factor. Denote the moving frame with a prime. Then the 4-vector potential seen in the moving coordinate system is just that of a point charge at rest:

$$\underline{A}'^{\alpha} = \left[\frac{1}{(cr')} \right]^{\alpha} \begin{bmatrix} 1 \\ \vec{0} \end{bmatrix}.$$

So

$$\underline{A}^{\mu} = (\Lambda^{-1} \underline{A}')^{\mu} = \frac{1}{c} \left[\begin{bmatrix} \gamma & \beta\gamma & & \\ \beta\gamma & \gamma & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} f(t, x, y) \\ 0 \\ 0 \\ 0 \end{bmatrix} \right]^{\mu} = \frac{1}{c} \begin{bmatrix} \gamma f \\ \gamma \beta f \\ 0 \\ 0 \end{bmatrix}, \quad (33.16)$$

where $f(t, x, y) = 1/r' = (\gamma^2(x - \beta ct)^2 + y^2)^{-1/2}$.

We can now compute the Faraday tensor as usual. For example,

$$\underline{F}^{01} = c^{-1} \vec{E}_x = \underline{\partial}^0 \underline{A}^1 - \underline{\partial}^1 \underline{A}^0 = -\frac{\partial}{\partial ct} (c^{-1} \gamma \beta f) - \frac{\partial}{\partial x} (c^{-1} \gamma f) = \frac{\gamma f^3}{c} (x - \beta ct).$$

Reinstating the dropped factor $q/(4\pi\epsilon_0)$ gives again the results found in Your Turn 32D and Your Turn 32E (page 358). However, sometimes \underline{A} is all that's needed, and we see it was easier to obtain than the electric and magnetic fields.

33.9 MORE ABOUT \underline{J}

The geometric definition of the charge flux 4-vector in Section 33.6.1 is useful for some purposes, for example, to see why it obeys the continuity equation. However, for other purposes it's good to know that another formulation is equivalent to the geometric one.

33.9.1 A property of the delta function

First we need to review a key fact about the delta function.¹⁵ Think of it as a bump, $\delta(x; \sigma) = (2\pi\sigma)^{-1/2} e^{-x^2/(2\sigma^2)}$ with σ very small. So

$$\int_{-\epsilon}^{\epsilon} dx \delta(x; \sigma) \rightarrow 1$$

if we hold ϵ fixed to any positive value and take $\sigma \rightarrow 0$.

Now define a new function $f(x; \sigma) = \delta(2x; \sigma)$ and compute the integral, changing variables to $y = 2x$:

$$\int_{-\epsilon}^{\epsilon} dx f(x; \sigma) = \int_{-2\epsilon}^{2\epsilon} \frac{dy}{2} (2\pi\sigma)^{-1/2} e^{-y^2/(2\sigma^2)} \rightarrow \frac{1}{2}.$$

Again the limit is taken holding ϵ fixed to any positive value and $\sigma \rightarrow 0$. In the same limit, the integral would have been zero had we chosen any range not centered on $x = 0$.

Thus f has the same properties as those defining $\frac{1}{2}\delta(x)$. More generally,

$$\delta(ax) = \frac{1}{a} \delta(x) \quad \text{for positive constant } a. \quad (33.17)$$

¹⁵This was introduced in Section 0.3.6 (page 9).

Next define $g(x; \sigma) = \delta(-2x; \sigma)$. Its graph is the same as that of f , so it has the same integral:

$$\delta(ax) = \frac{1}{|a|} \delta(x) \quad \text{for any constant } a.$$

More generally, if $h(x)$ is any smooth function that vanishes at an isolated point x_* , then¹⁶

$$\delta(h(x)) = \left| \frac{dh}{dx} \Big|_{x_*} \right|^{-1} \delta(x - x_*). \quad (33.18)$$

(Equation 33.17 corresponds to $h(x) = \pm ax$.) If $h(x) = 0$ at several points, then we get the sum of one term for each such point.

33.9.2 A useful alternative formulation of \underline{J}

Here is another set of quantities that may also seem reasonable as a candidate for the current. We will propose it, then show that it's the same as \underline{J} .

Define four functions on spacetime by putting bumps all along each trajectory $\underline{\Gamma}_{(\ell)}$:

$$\underline{J}_{\text{alt}}^\mu(\underline{X}) = \sum_{\ell} \int_{-\infty}^{\infty} cd\tau q_{\ell} \underline{U}_{(\ell)}^\mu(\tau) \delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau)). \quad (33.19)$$

We now want to show that $\underline{J}_{\text{alt}}$ is equal to the \underline{J} defined above. (At least the units match those of \underline{J} .)

Consider any component of Equation 33.19, for example $\mu = 2$, and any starting point \underline{X} . Thus, we wish to show $\underline{J}_{\text{alt}}^2(\underline{X}) = \underline{J}^2(\underline{X})$. Let \underline{X}_{\perp} denote just the 0, 1, and 3 components (all except the direction 2 that we chose to investigate). As in Equation 33.7, let $\Delta^3 \underline{X}_{\perp}$ be a small region about \underline{X} obtained by varying everything except \underline{X}^2 . We will now integrate $\underline{J}_{\text{alt}}^2$ and \underline{J}^2 over this region and show that the answers are the same. Because the region was arbitrary, that result will suffice to show that $\underline{J}_{\text{alt}} = \underline{J}$.

Thus we wish to simplify

$$\int_{\Delta^3 \underline{X}_{\perp}} d(ct) dx dz \underline{J}_{\text{alt}}^2 = \sum_{\ell} \int_{\Delta^3 \underline{X}_{\perp}} d(ct) dx dz \underbrace{\int cd\tau q_{\ell} \underline{U}_{(\ell)}^2(\tau) \delta(\underline{X}^2 - \underline{\Gamma}_{(\ell)}^2(\tau))}_{\delta^{(3)}(\underline{X}_{\perp} - \underline{\Gamma}_{(\ell)\perp}(\tau))} \delta^{(3)}(\underline{X}_{\perp} - \underline{\Gamma}_{(\ell)\perp}(\tau)). \quad (33.20)$$

The things in the brace don't depend on t , x , or z , so we may bring them to the front:

$$= \sum_{\ell} \int \underbrace{cd\tau q_{\ell} \underline{U}_{(\ell)}^2(\tau) \delta(\underline{X}^2 - \underline{\Gamma}_{(\ell)}^2(\tau))}_{\delta^{(3)}(\underline{X}_{\perp} - \underline{\Gamma}_{(\ell)\perp}(\tau))} \int_{\Delta^3 \underline{X}_{\perp}} d(ct) dx dz \delta^{(3)}(\underline{X}_{\perp} - \underline{\Gamma}_{(\ell)\perp}(\tau)). \quad (33.21)$$

The part of this expression in the second brace just gives 1 if particle $\# \ell$'s transverse coordinates fall anywhere inside $\Delta^3 \underline{X}_{\perp}$ at proper time τ ; otherwise it's zero. That is, as a function of τ it's a kind of step function.

Now turn to the rest of Equation 33.21. If trajectory $\# \ell$ is ever inside the range $\Delta^3 \underline{X}_{\perp}$ and crosses the fixed y that we are considering, then let τ_* be the proper time

¹⁶See also P+S §3.5.1.

when that crossing occurs.¹⁷ Equation 33.18 gives the integrand in the first brace as

$$cq_\ell \frac{d\Gamma_{(\ell)}^2}{d\tau} \left| \frac{d\Gamma_{(\ell)}^2}{d\tau} \right|^{-1} \delta(\tau - \tau_*) = \pm cq_\ell \delta(\tau - \tau_*).$$

We get the plus sign if the trajectory crosses from smaller to larger y , or the minus sign in the contrary case.

Putting it all together, the only trajectories that make nonzero contributions to Equation 33.20 are those that actually pass through $\Delta^3 \underline{X}_\perp$. We may thus restrict the sum to only those trajectories, which we denote by \sum'_ℓ , and so Equation 33.20 becomes

$$\int_{\Delta^3 \underline{X}_\perp} d(ct) dx dz \underline{J}_{\text{alt}}^2 = c \sum'_\ell (\pm q_\ell). \quad (33.22)$$

At last we can see that Equation 33.22 is the same property that we used to define the current \underline{J}^2 in Equation 33.8. Repeating the argument for the other three components yields that $\underline{J}_{\text{alt}} = \underline{J}$.

33.9.3 Another proof that \underline{J} is a 4-vector

Before proceeding, let's pause to show that $\delta^{(4)}(\underline{X})$ is a 4-scalar. Suppose that \underline{G}^α are a set of functions of \underline{X} that define a new set of coordinates, and that they all vanish at a point \underline{X}_* . We can generalize the result Equation 33.18 to say that

$$\delta^{(4)}(\underline{G}^\alpha(\underline{X})) = \left| \det \frac{\partial \underline{G}^\alpha}{\partial \underline{X}^\nu} \right|^{-1} \delta^{(4)}(\underline{X}^\mu - \underline{X}_*^\mu). \quad (33.23)$$

For a Lorentz transformation, \underline{G} is a set of four *linear* functions, so the derivatives appearing in Equation 33.23 are a constant matrix, which we have called Λ^α_ν . The determinant of that matrix is ± 1 because $[\Lambda^t g \Lambda] = [g]$, so Equation 33.23 says $\delta^{(4)}(\underline{X})$ is a 4-scalar.

Now we can use our reformulation of the current (Equation 33.19) to show that \underline{J} is a 4-vector. Indeed, in that equation $d\tau$ is a 4-scalar, the q_ℓ are all 4-scalars, we just showed that the delta function is a 4-scalar, and \underline{U} is a 4-vector (it is the derivative of the 4-vector \underline{X} with respect to the invariant τ).

33.10 A DIZZYING VISTA

Einstein famously said, “Leave elegance to the tailor.” Should we care that Equations 33.12 are so beautiful?

I'd say: When it's time to quantize electrodynamics, the covariant form of Maxwell's equations is the indispensable *starting point*. Recast them in a (Lorentz-invariant) Lagrangian formulation, write the appropriate path integral, and you're on your way. It can be done, sort of, without 4-tensor notation, but it's almost impossible to do it right without the simplicity we've now gained.

¹⁷For a small enough region $\Delta^3 \underline{X}_\perp$ there will be at most a single crossing. In Figure 33.1b (page 367), trajectory #4 passes through $\Delta^3 \underline{X}_\perp$, but it's not there when it crosses the chosen y value. Trajectory #3 never visits the chosen y at all.

Moreover, the train of thought begun in the last few chapters led Einstein in seven more years to unravel a seemingly unrelated puzzle. It's a fantastic detective story: A formal observation about the structure of *electromagnetism* led Einstein to a hypothesis, with testable quantitative predictions, about the nature of *gravitation*.

Einstein began by asking himself, what is it that makes some coordinate systems (the inertial ones) particularly good? Why aren't all systems equally good?

Our discussion of waves on a vibrating string gives a hint. Faced with a dynamical equation (for the string's transverse displacement) with less symmetry than expected (no galilean invariance), we realized that some additional dynamical variable (the velocity of the string) is hiding in the equation, implicitly set to some particular value (zero). Explicitly acknowledging this implicit physical object, and realizing that its value, too, will change under coordinate transformations, restored the full galilean invariance to the string's wave equation.

Should we try the same thing with the Maxwell equations? What is the hidden dynamical variable? Einstein argued it's *not* the velocity of any luminiferous æther. Rather, Section 31.5.1 characterized the "good" coordinate systems as those in which the interval—a *metric* function on spacetime—looks nice. Thus, to make progress we should start asking

- . What is the origin of the invariant interval function? Is it really a fixed property of spacetime, or could g itself be a dynamical object? (If so, then we'll need to propose some new dynamical law for the metric itself!)
- . Do Maxwell's equations become fully coordinate-invariant if we promote the metric tensor to a fully dynamical object, with an appropriate transformation law?

Remarkably, Einstein found that *again* there is essentially only one acceptable equation of motion that a metric tensor could have.¹⁸ He then asked, what new physical phenomena are predicted if we introduce this new dynamical variable?

The big clue was a fact from the geometry of curved surfaces: Any metric looks equivalent to any other one, if we only look to first order in excursions about a point. Einstein asked, "is there any physical property of spacetime that also has this property?" His answer was: "Yes, the gravitational field does." Once again, "Einstein thinking" suggested that the unique equation of motion dictated by general principles like invariance should then describe *all* gravitational phenomena, including even those not yet imagined (such as dark energy), and once again, this vision was borne out.¹⁹

33.11 PLUS ULTRA

As the world's first relativistic field theory, electrodynamics is also the indispensable intellectual substrate for creating a relativistic theory of electron spin, and from there onward to supersymmetry. One could argue that Einstein thinking merely ratified

¹⁸Here "essentially" means there's actually a two-parameter family of equations. One parameter is Newton's constant, as expected. The other one is the "cosmological constant." Despite some initial missteps, we now see that this parameter, too, corresponds to physical phenomena that are observed.

¹⁹Although these notes cannot go further, this thread was the insight needed for Hanging Question #G.

Maxwell's equations. But this sort of thinking was then critical for Dirac to even *propose* the right wave equation for particles with spin $1/2$.

$\boxed{T_2}$ Section 33.11' (page 377) outlines the relativistic treatment of spin.

FURTHER READING

$\boxed{T_2}$ Spinors: Nonrelativistic: Landau & Lifshitz, 1977.

Relativistic: Wess & Bagger, 1992, Appendix A; Berestetskii et al., 1971; Streater & Wightman, 1964; Sternberg, 1994, §1.2.

T₂ Track 2

33.4' Inversions

One of our goals is to eliminate the Levi-Civita tensor from all of classical physics (Hanging Question #E). Chapter 14 advocated rephrasing electrodynamics by replacing \vec{B} by the antisymmetric rank-3 tensor $\vec{\omega}$, and indeed we see that the spatial block of Equation 32.5 does just that. Then our manifestly-invariant form of the Lorentz force law, Equation 32.2, is also manifestly invariant under inversions of space or time because inversions satisfy the condition $[\Lambda^t \underline{g} \Lambda] = [\underline{g}]$. We'll see soon that once Maxwell's equations are formulated in terms of \underline{F} , they, too, will make spatial inversion invariance manifest.

T₂ Track 2

33.7' Degeneracy of Maxwell equations

So in all there are eight distinct equations, just like the usual form of the Maxwell equations. Previously we worried that the Maxwell equations are overdetermined, being eight equations in six unknown functions,²⁰ but we found that the system of equations is singular: Two of the eight equations are tautologies, vacuously satisfied regardless of what the fields and particles are doing. To see this again, more invariantly,

. Take the 4-divergence of the first set of equations and recall that $\partial_\mu \underline{J}^\mu = 0$ identically. So one combination of these four equations is vacuous.

. Apply $\underline{\varepsilon}^{\mu\nu\lambda\kappa} \partial_\kappa$ to the second set of equations and recall that partial derivatives commute. Here $\underline{\varepsilon}$ is the 4D analog of the Levi-Civita tensor. Again, you find that one combination of these four equations is vacuous.

A further reduction is possible if we use potentials (Section 33.8.1, page 370).

T₂ Track 2

33.8.1' Counting equations, again

The main text arrived at four equations, Equation 33.15 (page 370). However, one degree of freedom in \underline{A} drops out of the equations, due to their gauge invariance. What rescues the equations from being overdetermined is that one combination is vacuously satisfied, as we see by taking the 4-divergence of both sides and using the continuity equation for \underline{J} .

²⁰Hanging Question #D.

T₂ Track 2

33.9' Geometric status of the charge flux

If you know a little differential geometry, then we can give a more general formulation of charge and charge flux, one that does not require any choice of coordinate system at all. The geometric formulation in Section 33.6.1 specified an inertial coordinate system, so that the volume of $\Delta^3 \underline{X}_\perp$ was defined. But the same approach can be used to define a machine, called $*\underline{J}$, that eats any small solid element $\Delta^3 \underline{X}$ and spits out a number (net charge crossing it). It is linear in the volume. Such a machine is called a **3-form** (that is, a totally antisymmetric, rank $\binom{0}{3}$ 4-tensor). More precisely, to obtain a 3-form we must first choose an orientation on spacetime. Given a point in spacetime and an ordered set of three vectors, we get an oriented 3-volume element (Figure 12.2). If a particle trajectory pierces this element, then its velocity at that intersection completes the triad to give an orientation on spacetime. We count that particle's charge positively if that orientation matches the one we chose, or negatively if not. Summing over particles thus gives a number that depends multilinearly and antisymmetrically on the three vectors: A 3-form depending on a choice of orientation.²¹

Maxwell's equations then say

$$d\underline{F} = 0, \quad d * \underline{F} = \mu_0 (*\underline{J}),$$

where $*$ on the left denotes the Hodge dual operation, $*\underline{J}$ is the 3-form just defined, and d is the exterior derivative. In fact, $*\underline{J}$ really is related to our \underline{J} by the Hodge dual operation. That operation, however, requires the use of a metric tensor, so $*\underline{J}$ is more directly related to charged particle trajectories than \underline{J} .²² Still, at least we can say that on flat Minkowski spacetime, the 4-vector field \underline{J} is the same regardless which inertial coordinate system we used to define it, because the metric has the same form in any inertial system. That's the result we got in a more concrete way in Section 33.6.2.

T₂ Track 2

33.11' Spinors

One of my life's strongest emotional experiences related to science occurred when for the first time I understood Dirac's equation.

— *Abraham Pais*

Previous sections have discussed the linear representations of the rotation group $\text{SO}(3)$, which we've called "3-tensors." The key theorem says that, up to equivalence, there is just one

²¹We don't even need to assume flat spacetime, because we made no use of any metric tensor; given the trajectories of particles, their charges, and a choice of orientation, we directly get the charge flux 3-form $*\underline{J}$.

²²The difference matters in curved spacetime. The Hodge dual operation also depends on a choice of orientation on 4-space, so the net effect is that \underline{J} does *not*.

real, fundamental representation (3-vector); all other nontrivial representations are obtained by tensor products of this one (3-tensors of rank p), possibly (anti)symmetrized.

In the relativistic case, we have been discussing linear representations of the Lorentz group $SO(3,1)$, which we've called "4-tensors." This time we found *two* distinct fundamental representations (called rank $\binom{1}{0}$ and $\binom{0}{1}$); all other nontrivial representations are obtained by tensor products of these two (and are called rank $\binom{p}{q}$ 4-tensors), possibly (anti)symmetrized. In notation, the two fundamental representations are distinguished by up/down placement of the corresponding indices. Although they are distinct representations, they are "equivalent": We can convert between them by index raising and lowering.

We now outline extensions to these ideas introduced by quantum mechanics.

3D

In quantum mechanics, the existence of a symmetry group G only implies that the Hilbert space of states gives a *projective* representation of G , or equivalently a true representation of an extended form of G .²³ In nonrelativistic quantum mechanics, the relevant covering group is $SU(2)$.²⁴ To see this, first note that any real 3-vector \vec{v} corresponds to a traceless hermitian matrix via $\vec{v} \leftrightarrow M = \vec{\sigma} \cdot \vec{v}$, where $\vec{\sigma}_i$ are the Pauli matrices. Moreover,

$$\det M = -\|\vec{v}\|^2.$$

Let U be any special unitary matrix and dagger represent hermitian conjugate. Then $[U^\dagger M U]$ is traceless and hermitian with the same determinant as M , so it corresponds to a new vector that's a rotation of \vec{v} . The correspondence we have set up between $SU(2)$ and $SO(3)$ preserves the product structures of the groups. But this rotation is the same as the one determined by $-U$, so the correspondence is 2-to-1.²⁵

The key theorem says that all representations of the covering group are obtained as totally symmetric tensor products of one fundamental representation. The ordinary 3-tensors appear as the even-numbered entries on this list (they give ordinary representations of $SO(3)$). The odd-numbered entries are *new* (not encountered in classical physics): They are generically called "3-spinor representations."

4D

In *relativistic* quantum mechanics, we instead need projective representations of the Lorentz group. Its covering group turns out to be the group of 2×2 complex matrices with determinant one (also called $SL(2, \mathbb{C})$).²⁶ To see this, first note that any real 4-vector \underline{X} corresponds to a hermitian matrix via $\underline{X} \leftrightarrow M = \underline{X}^0 \mathbf{1} + \underline{X}^i \vec{\sigma}_i$. Moreover,

$$\det M = -\|\underline{X}\|^2.$$

Let W be any complex matrix with determinant equal to 1. Then $[W M W^\dagger]$ is also hermitian with the same determinant as M , so it corresponds to a new 4-vector that's a Lorentz transformation of \underline{X} . The correspondence we have set up between $SL(2, \mathbb{C})$ and $O(3,1)$ preserves the product structures of the groups. But the Lorentz transformation corresponding to W is the same as the one determined by $-W$, so the correspondence is 2-to-1.

²³In this case the appropriate extension is called the "covering group" Bargmann, 1954.

²⁴The special unitary 2×2 matrix group $SU(2)$, which double-covers the rotation group $SO(3)$, is also called $Spin(3)$ in this context.

²⁵Note that spatial *inversions* cannot be obtained in this way and must be treated separately.

²⁶Again, inversions must be treated separately. The group of special linear 2×2 complex matrices is also called $Spin(3,1)$ in this context.

This time, the key theorem says that there are *two* inequivalent fundamental representations of $SL(2, \mathbb{C})$. Any other irreducible representation is obtained as the totally symmetric tensor product of m copies of the first representation, combined with the totally symmetric tensor product of n copies of the second one. In notation, they are often distinguished by placing a dot over each two-valued index corresponding to one of the fundamental representations (and no dot for the other one):

$$\psi'_\alpha = W_\alpha{}^\beta \psi_\beta; \quad \eta'_\dot{\alpha} = [W^{\dagger\dot{\alpha}}]_{\dot{\alpha}}{}^{\dot{\beta}} \eta_{\dot{\beta}}.$$

The distinction between the two transformation rules just given is not superficial like the one between up and down indices on ordinary 4-tensors: The two representations are not equivalent because there is no standard conversion from one type of index to the other.²⁷

The representations with $n + m$ an even integer correspond to ordinary 4-tensors. The others are new (not encountered in classical physics): They are generically called “4-spinor representations.” For example,

- A chiral neutrino has $(n, m) = (1, 0)$.
- An electron can be split into a $(1, 0)$ and a $(0, 1)$.
- Four-vectors appear as the case $(n, m) = (1, 1)$. The sum $n + m$ is even, so this is an ordinary representation of Lorentz.
- An antisymmetric rank-2 tensor (such as the Faraday tensor \underline{F}) can be split into a positive-helicity part, with $(n = 2, m = 0)$, plus a negative-helicity part with $(n = 0, m = 2)$. (These names arise because the Faraday tensor of a plane wave with circular polarization will belong to one or the other of these types, depending on its helicity.)

There are two “spinors from Heaven,” that is, constant matrices that are unchanged by the transformations we have found: For example, $\epsilon_{\alpha\beta} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}_{\alpha\beta}$ has the property that

$$W_\gamma{}^\alpha W_\sigma{}^\beta \epsilon_{\alpha\beta} = \epsilon_{\gamma\sigma},$$

and similarly for $\epsilon_{\dot{\alpha}\dot{\beta}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}_{\dot{\alpha}\dot{\beta}}$, because $\det[W] = 1$. Moreover, following a construction we made for ordinary tensors, we can also define $\tilde{\epsilon}^{\alpha\beta} = [\epsilon^{-1}]^{\alpha\beta}$ and show that it, too is an invariant constant spinor:

$$[W^\dagger]_\gamma{}^\alpha [W^\dagger]_\sigma{}^\beta \tilde{\epsilon}^{\gamma\sigma} = \tilde{\epsilon}^{\alpha\beta},$$

with a similar result for the dotted version.

We can now apply “Einstein thinking” to construct invariant equations of motion as candidates for field equations for spinors, much as we did in the main text for the Faraday tensor. Here is one:

$$\boxed{[\underline{\sigma}_\mu]_{\alpha\dot{\beta}} \tilde{\epsilon}^{\dot{\beta}\dot{\sigma}} \partial^\mu \eta_{\dot{\sigma}} = 0, \quad \text{Weyl equation}} \quad (33.24)$$

where $[\underline{\sigma}_0]_{\alpha\dot{\beta}}$ is the unit matrix and $[\underline{\sigma}_i]_{\alpha\dot{\beta}}$ are again Pauli matrices. Indeed, quantizing a spinor field that obeys Equation 33.24 yields states describing massless chiral particles of spin 1/2. A little more tinkering yields the **Dirac equation** for a pair of chiral fermions with mass (for example, an electron).

²⁷However, if we restrict to *unitary* matrices (that is, to 3D rotations), then $[\underline{\sigma}_2 \eta]$ does transform the same way as $[\psi]$, so the two corresponding representations of $SO(3)$ are equivalent. Thus in 3D there is only one fundamental representation, as stated earlier.

PROBLEMS

33.1 *Uniformly moving charge revisited*

A charged point particle moves in a straight line with constant speed v . The charge creates electric and magnetic fields. Find a manifestly covariant expression for the Faraday tensor. That is, your formula should be an antisymmetric rank $\binom{2}{0}$ tensor constructed out of scalars and the four-vectors \underline{U} and \underline{X} using The Rules. Here \underline{X} is displacement from the particle to the observer. Check that your result is equivalent to the ones in Section 32.3.2. [*Hints:*

- (i) Sometimes it's easier to start by finding the 4-vector potential, as in Section 33.8.2.
- (ii) Again, your result must reduce to Coulomb's law if the particle is at rest in the chosen inertial coordinate system.
- (iii) The combinations

$$\underline{K}^{\mu\nu} = (\underline{U}^\mu \underline{X}^\nu - (\mu \rightleftharpoons \nu)) \quad \text{and} \quad \|\underline{K}\|^2 = \underline{K}^{\mu\nu} \underline{K}_{\mu\nu}$$

are useful intermediate building blocks for your answer, because the latter is equal to something useful when computed in the rest frame of the particle.]

CHAPTER 34

Energy and Momentum of Fields

Initially, Einstein was not impressed [by Minkowski's geometrical formulation] and regarded the transcriptions of his theory into tensor form as “überflüssige Gelehrsamkeit” [superfluous erudition]. However, in 1912 he adopted tensor methods and in 1916 acknowledged his indebtedness to Minkowski for having greatly facilitated the transition from special to general relativity.

— *Abraham Pais*

34.1 FRAMING

Every physical quantity carries dimensions, which help us to see its role and to formulate reasonable candidate laws. Now we have seen that every physical quantity also has a tensor character, another meta-property that helps us to see its role and to formulate reasonable candidate laws. This idea goes way beyond theories like gravitation and electrodynamics—when you study liquid crystals, fluctuating fluid membranes, etc., it's everywhere.

34.2 WHAT NEEDS TO BE SHOWN AND WHY

So far we've just reformulated old laws, but now it's time for something more ambitious. We no longer believe that space is filled with gears, pulleys, rubber bands, etc. that carry the EM fields, so we can't write down any functions for energy and momentum based on intuitions gleaned from mechanics. Instead we hope to prove a theorem about our system of equations stating that certain quantities are locally conserved and include familiar bits corresponding to energy and momentum of point particles. But to get started we need a good guess for what those quantities might be.

Chapter 5 computed the work that must be done to charge a capacitor. That energy isn't lost—you can get your investment back. Where is that energy in the meantime? We got a hint: It's proportional to the volume occupied by electric field. Maybe it's in the *empty space* between the capacitor plates.

Similarly, Chapter 17 computed the work that must be done to set up a current in a coil of wire. If the wire is superconducting, then the energy is not lost—you can get your investment back. Where is that energy in the meantime? We found that it, too, is proportional to the volume. Maybe it, too, is in the *empty space* inside the coil. That is, our hypothesis is that the vacuum itself can store energy in static electric and magnetic fields. We need to make that more general and precise.

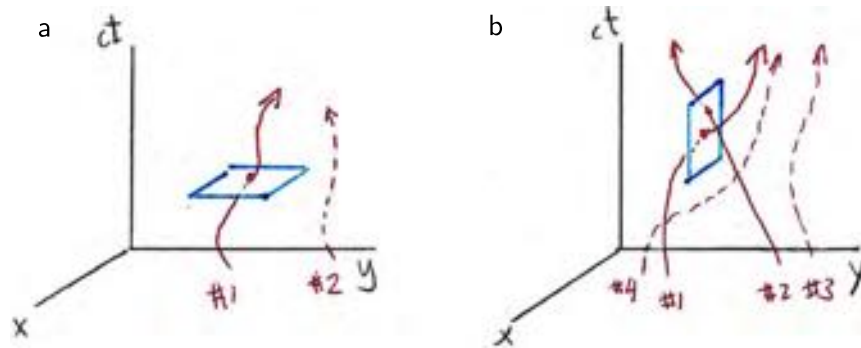


Figure 34.1:

Chapter 18 also studied energy and momentum *fluxes* in nonstatic situations, specifically plane waves. Here again we found them to be quadratic in the field amplitudes, although we didn't yet get the constant of proportionality: We just found how much of the energy and momentum could be extracted by a particular charged test body.

In short, it's been an ad hoc approach until now. Now that we have unified \vec{E} and \vec{B} , now that we have unified energy and momentum, it's time for one big result that covers all these electromagnetic phenomena at once. To get it, we'll generalize the discussion of waves on a string (Chapter 26). We found formulas for energy flux and density, and momentum flux and density. (They were quadratic in the amplitude.) Then you proved continuity equations expressing local conservation of energy and momentum. We'll now attempt the same thing with EM fields.

Using "Einstein thinking," the strategy will be: Find an expression that's the sum of the particles' $\underline{p}_{(\ell)}^\mu$, plus a quadratic function of fields with appropriate tensor properties. Requiring that this expression must also obey a continuity equation will uniquely determine it. Then the field term, whatever it turns out to be, will deserve to be called the "energy and momentum of the fields," and its continuity equation will be the local conservation law that we wanted to prove. We'll see that indeed, energy and momentum can slosh locally back and forth between fields and particles, while staying conserved overall.

Certainly the tensor structure will be more complex than in the string/spring metaphor. That's one reason why we invented our big language.

34.3 CONTINUITY EQUATION FOR ENERGY AND MOMENTUM IN THE ABSENCE OF LONG-RANGE FORCES

First consider a swarm of particles with no external forces and no mutual long-range forces. Between collisions, each trajectory $\underline{\Gamma}_{(\ell)}(\tau)$ is therefore a straight line, which we parameterize by proper time. Let's suppose that each collision locally conserves energy and momentum, much as we assumed in Chapter 7 that collisions locally conserve electric charge. Analogously to the charge flux 4-vector \underline{J} , define the **energy-**

momentum flux tensor¹ by a recipe analogous to Equation 33.8 (page 368):

$$\begin{aligned} \underline{T}^{\mu\nu} &= \text{net amount of } \underline{p}^\nu \text{ crossing the surface } \underline{X}^\mu = \text{constant,} \\ &\text{from smaller to larger } \underline{X}^\mu, \text{ per } d^3X_\perp, \text{ times } c. \end{aligned} \quad (34.1)$$

Your Turn 34A

Using Figure 34.1, convince yourself that

$$\begin{aligned} \underline{T}^{00} &= c \text{ times the density of (energy}/c) \\ \underline{T}^{i0} &= \text{flux of (energy}/c) \\ \underline{T}^{0k} &= c \text{ times density of the } k \text{ component of momentum} \\ \underline{T}^{ik} &= \text{flux along } i \text{ direction of the } k \text{ component of momentum.} \end{aligned}$$

We can call the part of the energy-momentum flux tensor carried by particles $\underline{T}_{\text{part}}$, and write an equivalent formula like the one used for \underline{J} in Equation 33.19 (page 372): Just replace the charge on particle ℓ by the 4-momentum on particle ℓ at proper time τ :

$$\underline{T}^{\mu\nu}(\underline{X}) = \sum_{\ell} \int_{-\infty}^{\infty} c d\tau \underline{p}_{(\ell)}^\nu(\tau) \underline{U}_{(\ell)}^\mu(\tau) \delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau)). \quad (34.2)$$

Your Turn 34B

Convince yourself that \underline{T} is a symmetric, rank- $\binom{2}{0}$ tensor. Then show that the energy-momentum flux tensor obeys

$$\frac{\partial}{\partial \underline{X}^\mu} \underline{T}_{\text{part}}^{\mu\nu} = 0. \quad \text{no long-range forces} \quad (34.3)$$

That is, if no long-range forces act then $\underline{T}_{\text{part}}$ obeys four continuity equations, expressing the local conservation of each component of the 4-momentum.

34.4 INTERACTIONS SEEM TO SPOIL LOCAL CONSERVATION

34.4.1 Long-range forces

Of course, if some external force acts on our particles, then we *don't* expect their energy or momentum to be conserved: A falling body accelerates (gains momentum). Even *mutual* forces, if they act at long range, would destroy local conservation: Two distant plus charges, initially at rest, start to accelerate away from each other, so equal and opposite amounts of momentum seem to appear at two distant locations.²

¹Often abbreviated “energy-momentum tensor.” Some authors call it the “stress-energy tensor.”

²Also, each gets *not*-opposite amounts of kinetic energy, again seemingly from nowhere.

Sections 2.4.1 and 17.12 argued that introducing an entity called the “electromagnetic field” is needed in order to rescue locality. To deliver on this promise, we need to attribute local energy and momentum to fields as well as to particles. Then the repulsion of two particles involves each one getting momentum locally from the *field* nearby, and so on. It’s not obvious that this can be done! Let’s begin by getting quantitative about the preceding paragraph.

Adapting our proof of the continuity equation (Chapter 7), again draw a small four-dimensional box (hypercube) and ask how much net momentum enters it by particles crossing its faces. As with electric charge, that net change will equal $(-c^{-1}\partial_\mu T_{\text{part}}^{\mu\nu})(\Delta^4 X)$.³ Unlike that case, however, this quantity won’t equal zero, because each particle’s momentum and energy *change* during its passage through the box. Thus, upon exit from the box each particle carries out a different momentum and energy from what it brought in, even if it didn’t collide with any other particle:⁴

$$\Delta_{\text{box}} \underline{p}^\nu = \text{net } \underline{p}^\mu \text{ into 4-box} = - \sum'_\ell \int_{\tau_{\text{in},\ell}}^{\tau_{\text{out},\ell}} d\tau \frac{dp'_\ell}{d\tau}. \quad (34.4)$$

In this formula, we only include those trajectories that actually enter the box; \sum'_ℓ denotes the restricted sum. Moreover, we only include the part of each particle’s trajectory that is actually spent inside the box. That explains the limits on the τ integral. Finally, we only need to include the contributions to $dp'_\ell/d\tau$ arising from electromagnetic forces on the particles. Although there can also be collisions inside the box involving short-range forces, these locally conserve 4-momentum and so cancel in Equation 34.4.

34.4.2 Nonconservation of particle energy and momentum

We now use the Lorentz force law to relate the last factor in Equation 34.4 to the fields. The formula is cumbersome, however, because of the restricted sum and integral. To make it easier to work with, we now make the unobvious step of multiplying by one, using the identity $1 = \int d^4 X \delta^{(4)}(\underline{X} - \underline{X}_*)$ for any point \underline{X}_* in spacetime. For each term ℓ and each value of τ , make the choice $\underline{X}_* = \underline{\Gamma}_{(\ell)}(\tau)$. Then we move the integration over \underline{X} all the way to the left (do it last):

$$\Delta_{\text{box}} \underline{p}^\nu = - \int d^4 X \sum'_\ell \int_{\tau_{\text{in},\ell}}^{\tau_{\text{out},\ell}} d\tau \frac{dp'_\ell}{d\tau} \delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau)).$$

This looks like it’s making our formula more complicated, but now note what happens if we restrict the \underline{X} integral to just our little box (hypercube). Then the delta function automatically selects only the trajectories that pass through the box, so we don’t need to restrict the sum. And the delta function also automatically selects only those τ values for which a trajectory lies inside the box, so we don’t need to restrict the τ integral either. Using that insight, and the Lorentz force law, gives

$$\Delta_{\text{box}} \underline{p}^\nu = - \int_{\text{box}} d^4 X \sum'_\ell \int_{-\infty}^{\infty} d\tau q_\ell \underbrace{F^{\nu\lambda}(\underline{\Gamma}_{(\ell)}(\tau))}_{\text{Lorentz force}} U_{(\ell),\lambda}(\tau) \delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau)).$$

³See Equation 7.3.

⁴To understand the minus sign, note that if a particle gains momentum during its sojourn in the box, then it transports more out when it exits than it had upon entry.

Use the delta-function to reexpress the factor in the brace as $\underline{F}^{\nu\lambda}(\underline{X})$, and then push it to the left of the τ integral. What remains is just the electric charge flux four-vector:

$$\Delta_{\text{box}} \underline{p}^\nu = -c^{-1} \int_{\text{box}} d^4 X \underline{F}^{\nu\lambda}(\underline{X}) \underline{J}_\lambda(\underline{X}). \quad (34.5)$$

We have now expressed the net change of momentum in the box in terms of electromagnetic fields and the charge flux 4-vector.

For a small enough box, we may approximate the integral as $\Delta^4 X$ times the integrand. But Section 34.4.1 argued that this change is also $c^{-1} \Delta^4 X$ times minus the 4-divergence of \underline{T} , or

$$\partial_\mu \underline{T}_{\text{part}}^{\mu\nu} = \underline{F}^{\nu\lambda} \underline{J}_\lambda. \quad (34.6)$$

This formula makes precise what was argued qualitatively before: The energy-momentum flux tensor of *particles only* does not obey a continuity equation, if long-range forces are present.

34.5 ACCOUNTING FOR FIELD CONTRIBUTIONS RESTORES LOCAL CONSERVATION OF ENERGY AND MOMENTUM

Rather than give up, we are hoping to find *another contribution* to the total energy-momentum flux tensor of the world, attributing 4-momentum to *fields*, with the properties that:

- $\underline{T}_{\text{field}}^{\mu\nu}$ is a symmetric 4-tensor given by a local expression in the fields; and
- $\partial_\mu (\underline{T}_{\text{part}}^{\mu\nu} + \underline{T}_{\text{field}}^{\mu\nu}) = 0$.

That is, we want to find a contribution to the energy-momentum flux tensor depending only on fields and with the property that the *total* $\underline{T}^{\mu\nu}$ obeys a continuity equation. Once we prove it, that continuity equation will be a Lorentz-invariant formulation of the local conservation of total energy and momentum (Poynting's theorem, which was independently codiscovered by Heaviside).

Equation 34.6 shows what we need:⁵

$$\partial_\mu \underline{T}_{\text{field}}^{\mu\nu} = -\underline{F}^{\nu\lambda} \underline{J}_\lambda. \quad (34.7)$$

But we can't prove this until we guess the correct formula for $\underline{T}_{\text{field}}^{\mu\nu}$!

To get past this impasse, let's apply "Einstein thinking." What sorts of symmetric, rank-two tensors can we build from the Faraday tensor? We already have some anecdotal evidence that stored electrostatic energy is a *quadratic* function of electric field, with no derivatives ($\propto \vec{E}^2$). And stored magnetic energy is also a quadratic function of magnetic field, with no derivatives ($\propto \vec{B}^2$). Can we write any such expression that is a symmetric, rank- $\binom{2}{0}$ tensor?

In fact, we can write just *two* such expressions. Rather than choose one or the other, we must keep our options open and suppose that the tensor we are seeking is some linear combination of them both:

$$\underline{T}_{\text{field}}^{\mu\nu} = \alpha \underline{F}^{\mu\sigma} \underline{F}_\sigma{}^\nu + \beta g^{\mu\nu} \underline{F}_{\sigma\lambda} \underline{F}^{\sigma\lambda}. \quad \text{provisional formula} \quad (34.8)$$

⁵Compare P+S equation 12.112.

Indeed, the expression above is tensor of the right sort that's quadratic in fields and has no derivatives. We don't know the values of α and β yet, but already we've made a huge simplification: Just those *two numbers* is all the freedom we have to construct a suitable tensor.

We now take the 4-divergence of our provisional formula:

$$\partial_\mu (\alpha \underline{F}^{\mu\sigma} \underline{F}_\sigma{}^\nu + \beta g^{\mu\nu} \underline{F}_{\sigma\lambda} \underline{F}^{\sigma\lambda}).$$

Use the fact that the fields obey Maxwell's equations, specifically the first of Equations 33.12 (page 369):

$$= \alpha \underbrace{(-\mu_0 \underline{J}^\sigma \underline{F}_\sigma{}^\nu + \underline{F}^{\mu\sigma} \partial_\mu \underline{F}_\sigma{}^\nu)} + \beta \underbrace{g^{\mu\nu} 2(\partial_\mu \underline{F}_{\sigma\lambda}) \underline{F}^{\sigma\lambda}}. \quad (34.9)$$

The first term (first brace) is just what we want! Simply choose the value $\alpha = -\mu_0^{-1}$ and we get Equation 34.7.

We are left with the unwanted other terms (second brace). Can we choose a value of β such that these terms cancel each other identically? That is, can we ensure that

$$0 \stackrel{?}{=} \frac{\alpha}{\beta} \underline{F}^{\mu\sigma} \partial_\mu \underline{F}_{\sigma\nu} + 2 \underline{F}^{\sigma\lambda} \underbrace{\partial_\nu \underline{F}_{\sigma\lambda}}? \quad (34.10)$$

It's not as crazy as it sounds, because so far we have only used half of the Maxwell equations to obtain Equation 34.9. The other half indeed say that something involving first derivatives of \underline{F} equals zero:⁶ The quantity enclosed by the brace in Equation 34.10 equals

$$\underbrace{-\partial_\sigma \underline{F}_{\lambda\nu} - \partial_\lambda \underline{F}_{\nu\sigma}}.$$

In Equation 34.10, this tensor is contracted on $\sigma\lambda$ with something antisymmetric, so we may replace its first term (in the brace) by $+\partial_\lambda \underline{F}_{\sigma\nu}$. Then Equation 34.10 becomes

$$0 \stackrel{?}{=} \frac{\alpha}{\beta} \underline{F}^{\mu\sigma} \partial_\mu \underline{F}_{\sigma\nu} + 2 \underline{F}^{\sigma\lambda} (\partial_\lambda \underline{F}_{\sigma\nu} - \partial_\lambda \underline{F}_{\nu\sigma}) \quad (34.11)$$

$$= \left(\frac{\alpha}{\beta} - 4\right) \underline{F}^{\mu\sigma} \partial_\mu \underline{F}_{\sigma\nu}. \quad (34.12)$$

This will be identically true if we choose $\beta = \alpha/4$.

Substituting the values we found for α, β into Equation 34.8, we conclude that

$$\underline{T}_{\text{field}}^{\mu\nu} = -(\mu_0)^{-1} (\underline{F}^{\mu\sigma} \underline{F}_\sigma{}^\nu + \frac{1}{4} g^{\mu\nu} \underline{F}_{\sigma\lambda} \underline{F}^{\sigma\lambda}). \quad \begin{array}{l} \text{energy-momentum flux tensor} \\ \text{of the electromagnetic field} \end{array}$$

(34.13)

This choice meets all the criteria listed at the start of this section.

⁶See the second of Equation 33.12.

Your Turn 34C

- a. Confirm that its 00 component (energy density), when written in terms of \vec{E} and \vec{B} , has the form that you expect from Sections 5.3 (page 55) and 17.3 (page 198).
- b. Then show that its $i0$ components (flux of energy, or density of momentum) also have a form anticipated in Section 18.3 (page 221).
- c. The ij components may be new to you; they are interesting, too, so work them out and interpret in terms of radiation pressure (Section 18.1.3, page 220).

Note that the traditional formulas for energy density and Poynting vector need *no corrections* to account for relativity.

34.6 WHAT HAS BEEN ACCOMPLISHED

At a single stroke, we have established the local conservation not only of energy, but also of all three components of momentum.⁷ It is true that the final formulas for energy density, energy flux, and momentum flux agree with what we found informally in earlier chapters (Your Turn 34C), but

- Our earlier explorations *assumed* energy conservation. Now we have *proved* it as a property of Maxwell's equations and the Lorentz force law.
- Previously we didn't show that our expressions had the appropriate Lorentz transformation properties. Now it's obvious because we followed the Rules.
- Previously we only got expressions for energy and momentum flux in plane waves, and we didn't find the correct prefactor. Now we have complete and general formulas.
- Finally, the same derivation will also give us an analogous theorem when we later add media in Chapter 52.

It may seem that we have cheated! After all, we just cooked up a quantity precisely so that it would give $\partial_\mu T_{\text{tot}}^{\mu\nu} = 0$, so what has been proved? But it was highly nontrivial that this works at all. The only cookery allowed was the choice of *two constants*, α and β , but the theorem we proved was that *four functions* of space and time are everywhere zero.

It may also seem magical that our highly constrained guess, Equation 34.8, could be adjusted to satisfy the continuity equation. Chapter 39 will rediscover the energy and momentum conservation laws as consequences of the translational invariance of the Lagrange function giving rise to Maxwell's equations.

⁷This was Hanging Question #H.

T₂ Track 2

34.5'

Look in a freshman physics book. In mechanics they prove conservation of linear momentum. Later they prove conservation of angular momentum, but with an extra assumption that all forces between particles are directed along the line between them. There is no footnote saying “Um, that’s false for magnetic interaction.” A hundred pages later, they introduce the magnetic force; there is no footnote saying, “Um, that invalidates our proof of conservation of angular momentum.” We must do better than that.

Define the rank- $\binom{3}{0}$ tensor

$$\underline{M}^{\mu\nu\lambda} = \underline{X}^\nu \underline{T}^{\lambda\mu} - \underline{X}^\lambda \underline{T}^{\nu\mu}.$$

You can readily show that $\partial_\mu \underline{M}^{\mu\nu\lambda} = 0$. Thus, we find six densities by taking $\mu = 0$, leading to six conserved quantities

$$\underline{L}^{\nu\lambda} = \int d^3r \underline{M}^{0\nu\lambda}(t, \vec{r}).$$

The spatial bits of this tensor, \underline{L}^{ij} , are the relativistic version of the angular momentum, and we have just shown that they are conserved when we include both particle and field contributions to \underline{T} .⁸

⁸See Weinberg, 1972, page 46.

PROBLEMS

34.1 *Boom 2008*

The Large Hadron Collider project at CERN suffered something of a setback in October 2008, when, during a test of one of the quadrupole magnets, which focus the beams down to tiny size at the collision regions, the magnet failed catastrophically. The resulting “event” lifted a 20-ton magnet off its mountings, filled a tunnel with helium gas, and forced an evacuation (Figure 34.2).

The problem is that a big superconducting magnet stores a lot of magnetic field energy. If any bit of that magnet stops being superconducting, then suddenly the huge electric current generates a lot of heat. Eventually *all* the stored magnetic field energy ends up as heat. Let’s look at rough numbers. Suppose that the magnet maintains a field of 7 T in a channel of length 3 m and cross-section of area $(56 \text{ mm})^2$.

- a. Find the total magnetic energy in joules.
- b. The magnet is normally kept superconducting by a reservoir of liquid helium. The heat of vaporization of liquid helium is 83 J/mole (you can neglect the additional energy needed to bring He gas up to room temperature). If all the energy in (a) goes to vaporizing helium (and there’s an unlimited supply in the reservoir), how many moles of He gas do we get?
- c. Suppose all that helium gas exits the system via pressure-release valves, then comes up to room temperature. A mole of any ideal gas occupies about 24 liters at room temperature. What volume of helium gas would then flood the underground tunnel near the magnet?

34.2 *Magnetic stress*

Consider the attraction between two bar magnets placed end-to-end with one’s N pole separated from the other’s S pole by a narrow gap. You can ignore fringe fields in this problem, and assume that \vec{B} is uniform in the gap and points in the \hat{x} direction.

- a. For this pure magnetic field, show that $\underline{T}^{\mu\nu}$ takes the form $u\underline{M}^{\mu\nu}$, where $\underline{M}^{\mu\nu}$ is a constant 4×4 matrix and u is the energy density of the field.
- b. Use the continuity equation for momentum to show that the force on each magnet



Figure 34.2:

(total rate of transport of momentum) equals $u\Sigma$, where Σ is the area of the pole faces. (Or, if this is not necessarily true, use the idea behind the equation to describe when it will be true.) Then use Equation 34.13 for $\underline{T}_{\text{field}}^{\mu\nu}$ to evaluate this force.

- c. The total energy in the field is dominated by the contribution from the high-field space between the magnet poles, so it's $u\Sigma a$, where a is the distance between poles. Give a second derivation, based on energy conservation, for the force of attraction between the magnets.

34.3 Boom II

Suppose that a superconducting magnet is a cylinder of length 1 m with circular cross-section of radius 0.5 m. A current maintains a uniform, static magnetic field of 2 T inside the cylinder (and zero outside). Find the total electromagnetic field energy in this situation, in joules.

[*Culture:* If suddenly the magnet stops being superconducting, the current will rapidly crash to zero. Then all that energy must end up . . . somewhere. Magnets do explode in the lab. This Electromagnetic Phenomenon furnished the dramatic climax to the otherwise numbingly stupid film *The Man with the Golden Gun*.]

34.4 Angular momentum of fields

Background: EM waves can also carry *angular* momentum. You may use the following fact: The density of angular momentum J_z , computed using the origin as reference point, is $\hat{z} \cdot \frac{1}{\mu_0} [\vec{r} \times (\vec{E} \times \vec{B})]$. As usual we will suppose that the fields are harmonically varying in time and consider only the time average of our answers.

Do:

- Suppose we have two oscillating dipoles of strength p_0 at the origin, pointing at right angles to each other and both in the xy plane. The dipoles oscillate at the same frequency ω but 90° out of phase. Compute the density of the z component of angular momentum far away from the origin, to leading order in powers of $1/r$. Because everything moves radially outward, the radial component of the flux of J_z is then your answer divided by c .
- A sphere of large radius surrounds the dipoles and absorbs all the radiation. Before you compute anything: Will the sphere absorb any net angular momentum J_z ? Why/why not? Now do the calculation using (a), to get the rate at which J_z is transferred to the sphere.
- Also find the power absorbed by the sphere.
- Divide your answers to (b,c) and comment.

34.5 Impulse from changing field

Two opposite walls of a rigid, nonconducting, rectangular box are uniformly charged with surface charge densities σ and $-\sigma$ respectively. The positively charged wall occupies the region $0 < x < a$, $0 < y < b$ of the plane $z = c$. The other wall occupies the corresponding region of the plane $z = 0$. Inside the box there is a uniform magnetic field $\vec{B} = B_0 \hat{y}$. Assume that c is much smaller than either a or b .

- Use the Lorentz force law to find the impulse experienced by the box (that is, momentum delivered to it) if the magnetic field is suddenly switched off.
- Find the initial momentum of the electromagnetic field in the box. Make an In-sightful Comment.

34.6 *Fine point — energy and momentum of fields*

At one point in a derivation, I stated that

$$2\underline{F}^\sigma{}_\lambda \underline{\partial}_\sigma \underline{F}^\lambda{}_\mu$$

could be replaced by

$$-2\underline{F}^\sigma{}_\lambda \underline{\partial}^\lambda \underline{F}_{\sigma\mu},$$

where \underline{F} is the Faraday tensor. Why is this substitution justified?

CHAPTER 35

Faraday's Field Lines

In the treatises on physics published in England, there is always one element which greatly astonishes the French student: that element, which nearly invariably accompanies the exposition of a theory, is the model. . . . Here is [Lodge's book]. . . . In it there are nothing but strings which move around pulleys, which roll around drums, which go through pearl beads, which carry weights; and tubes which pump water while others swell and contract; toothed wheels which are geared to one another and engage hooks. We thought we were entering the tranquil and neatly ordered abode of reason, but we find ourselves in a factory.

— Pierre Duhem

35.1 FRAMING

Starting in 1821, Michael Faraday drew a lot of diagrams like the ones in Figure 35.1, and similar ones involving magnets. He found that he could get a consistent picture of both electric and magnetic forces by imaging invisible “lines of force” sprouting out of charges and magnet poles. The magnitude of the field increased as the lines were compressed laterally. The lines of force were under tension, like stretched rubber bands, yet repelled nearby lines with a transverse pressure-like force. This transverse pressure made the lines want to avoid each other, so they spread as they left a point charge; then the connection to density gave rise to the $1/r^2$ law.

It sounds crazy! Even decades later, the Continental *philosophes* were particularly severe on Faraday and his successors (e.g. the quote above). And yet Faraday, with

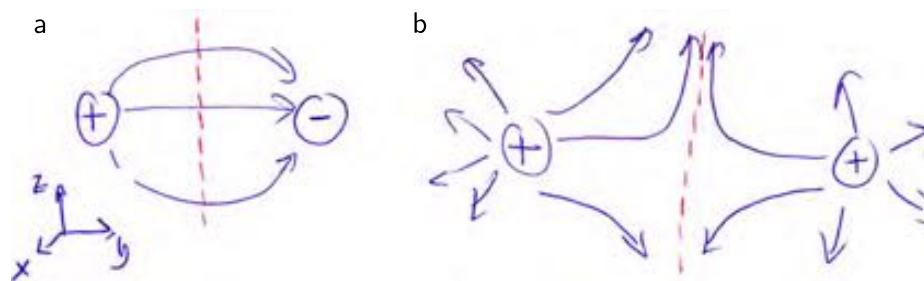


Figure 35.1: (a) Electric “lines of force” set up by two opposite point charges. The magnetic “lines of force” set up by two opposite pole tips look the same. The figure is antisymmetric upon reflection through the central plane (*dashed line*). (b) Two identical point charges or magnetic pole tips. This time the figure is symmetric upon reflection.

practically no formal education and certainly no math, used his intuitive picture to make a discovery that had eluded everyone else: the law of induction. Maybe his viewpoint belongs in our toolkit alongside the others.

35.2 FIELD LINES

So far in this course we have expressed electromagnetic phenomena using *vector fields*, not “lines of force.” But it’s easy to make a connection: The *streamlines* of a vector field define curves in space, and they do resemble the curves Faraday drew for the two situations in the figure. (Today they are often called **field lines**.) As in the figure, they spray out of a point charge, or the pole tip of a magnet. They then spread apart, indeed as if by mutual repulsion.

To get more precise, let’s warm up with a more tactile system: an incompressible fluid flowing steadily through a pipe. There is a vector field (the local velocity near each point inside the pipe), whose streamlines are literally the paths taken by individual molecules (maybe averaged over thermal motion). Suppose that the flow encounters a constriction in the pipe. Then individual flowlines must converge. We know from daily life that the fluid must also speed up as it passes through the constriction; even though if we sit at any given point we see a time-independent (steady) fluid velocity there, still a speck of dust being swept along will be moving faster at the constriction.

Indeed, if $\vec{V}(\vec{r})$ is the velocity field and $\rho_m = \text{const}$ is the density of the incompressible fluid, then the flux of mass is $\rho_m \vec{V}$ is the flux of mass, and the continuity equation for mass says

$$\vec{\nabla} \cdot (\rho_m \vec{V}) = -\dot{\rho}_m = 0.$$

That is, $\vec{\nabla} \cdot \vec{V} = 0$: Incompressible flow has divergence-free velocity. We know from Maxwell’s equations that the magnetic field everywhere has this property, and the electric field has it in empty space.

Next, write \vec{V} in terms of its magnitude and direction: $\vec{V} = f(\vec{r})\hat{n}(\vec{r})$. The divergence-free property implies

$$\begin{aligned} \hat{n} \cdot \vec{\nabla} f &= -f \vec{\nabla} \cdot \hat{n} \\ \hat{n} \cdot \left(\frac{\vec{\nabla} f}{f} \right) &= -\vec{\nabla} \cdot \hat{n}. \end{aligned} \quad (35.1)$$

The left side of this equation is the relative rate of change of the magnitude of velocity as we move along a streamline. The intuition cited above leads us to expect that this should reflect changes in the *transverse density* of a set of neighboring streamlines, so let’s see if the right hand side has any such interpretation.

Consider a simple situation, in which the constriction is just in one direction (y). Then \hat{n} lies always in the xy plane, as shown in Figure 35.2. In the middle of this small box $\partial \hat{n} / \partial x = 0$ but $\partial \hat{n}_y / \partial y < 0$. Thus $\vec{\nabla} \cdot \hat{n}$, as we expect for a converging flow. We ask what is happening to the transverse density of streamlines. If N lines enter at the left, spread over area BC , then they exit crammed into the smaller area $(B - 2dx \tan \theta)C$, where $\tan \theta \approx \theta \approx -\hat{n}_y$ evaluated at the top of the box. But $\hat{n}_y = 0$ at the center of the box, so by a Taylor expansion $\hat{n}_y(\text{top}) \approx \frac{1}{2} B \frac{\partial \hat{n}_y}{\partial y}$.

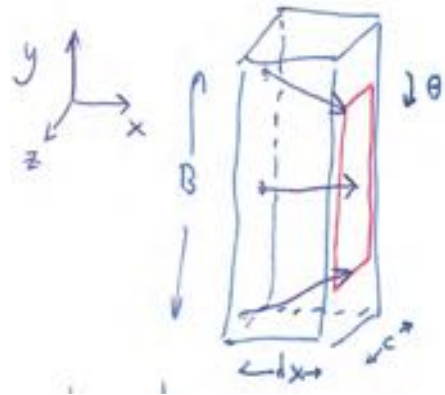


Figure 35.2:

There's a similar shrinkage at the bottom, so the area of the rectangle containing the streamlines decreases from BC to $B(1 + dx\vec{\nabla} \cdot \hat{n})C$. Then the transverse density of streamlines increases from $N/(BC)$ to $N(1 + dx\vec{\nabla} \cdot \hat{n})^{-1}/(BC)$. Its relative change is then $-dx\vec{\nabla} \cdot \hat{n}$, which is the right-hand side of Equation 35.1.

The relative rate of change is the logarithmic derivative. If two functions have the same logarithmic derivative everywhere, then one of them is a constant times the other. We have therefore established that the magnitude of velocity in an incompressible fluid is a constant times the transverse density of streamlines. (The constant is arbitrary because we could start with any number of streamlines.)

The same result holds for magnetic fields, and for electric fields in vacuum. (Electric charges act like sources or sinks of fluid.) Michael Faraday is smiling.

35.3 ELECTRIC AND MAGNETIC FORCES

The streamlines of \vec{E} thus contain all the information needed to reconstruct the direction and magnitude of the electric field, and similarly for \vec{B} . Drawing in the lines for two opposite point charges, we see maximum density right at the charges, high density between them, and zero density out at infinity, as we should expect (Figure 35.1a). Moreover, bringing the two charges closer reduces the volume over which the lines are closely packed, and increases the volume in which the lines are sparse. That reduces the integral of \vec{E}^2 , that is, the stored electrostatic field energy, so the opposite charges attract, as if the lines were real rubber bands under tension.

For two identical charges (Figure 35.1a), pushing them together increases the crowding at the central plane and increases energy, so the charges repel—as if the lines were real with a transverse pressure.

35.4 FORCES VIA THE STRESS 3-TENSOR

Your study of physics has probably made it clear that often there is both an “energy” approach to a problem and also a different-seeming “force” approach. In any given

problem one of those may be easier, so it's good to understand both. So let's now look at electric and magnetic forces predicted by our formula for the stress 3-tensor, \vec{T}_{ij} .

Let

$$\vec{R}_{ij} = \underline{F}^{0i} \underline{F}_0^j + \underline{T}^{ik} \underline{F}_k^j \quad \text{and} \quad S = \underline{F}^{0k} \underline{F}_k^0$$

Recall $\underline{F}^{0i} = \vec{E}_i/c$ and $\underline{F}^{ij} = \varepsilon_{ijk} \vec{B}_k$. Thus

$$\begin{aligned} \vec{R}_{ij} &= \frac{1}{c^2} \vec{E}_i \vec{E}_j + (\delta_{im} \delta_{lj} - \delta_{ij} \delta_{lm}) \vec{B}_l \vec{B}_m \\ &= \frac{1}{c^2} \vec{E}_i \vec{E}_j - \delta_{ij} \vec{B}^2 + \vec{B}_i \vec{B}_j. \end{aligned}$$

Similar steps give $S = -(\vec{E}/c)^2$.

Your Turn 35A

Use these results to show that

$$\vec{T}_{ij} = -\epsilon_0 \vec{E}_i \vec{E}_j + \frac{1}{2} \epsilon_0 \delta_{ij} \vec{E}^2 - \frac{1}{\mu_0} \vec{B}_i \vec{B}_j + \frac{1}{2\mu_0} \vec{B}^2 \delta_{ij}. \quad (35.2)$$

Along the midplane in Figure 35.1a, the electric field points along \hat{y} , by symmetry. Everything to the left of the midplane transfers momentum to everything to the right with flux of \vec{p}_2 equal to

$$\vec{T}_{22} = \epsilon_0 (-(\vec{E}_2)^2 + \frac{1}{2} \vec{E}^2).$$

That flux density is strictly negative, so when integrated over the plane it predicts a force on the right charge that is directed to the left, that is, attraction. Michael Faraday is smiling: This is his rubber-band tension at work.

Along the midplane in Figure 35.1b, the electric field is always perpendicular to \hat{y} . Thus

$$\vec{T}_{22} = \epsilon_0 (-(\vec{E}_2)^2 + \frac{1}{2} \vec{E}^2),$$

which is strictly positive. This time we predict repulsion. Michael Faraday is smiling: This is his transverse pressure at work.

35.4.1 Magnetic forces

The pictures look the same. And the magnetic terms of Equation 35.2 have the same forms as the electric terms. So we get the same results, and again Faraday is smiling.

35.5 MAGNETIC INDUCTION

Faraday took his field lines seriously, as objects with some sort of reality. That helped him to suggest that whenever a wire “cut across” magnetic field lines, something physical would happen—its charge carriers would feel a force. Such “cutting across” could happen when a wire was dragged through a static \vec{B} field (as in a dynamo), or when a motionless wire was subjected to a growing or shrinking \vec{B} (as in a transformer). Those statements eventually evolved into the magnetic part of the Lorentz force law and the field equation today called Faraday's law, respectively.

PROBLEMS

35.1 *Push comes to shove*

- a. Take the expression we found for the energy–momentum flux tensor:

$$\underline{T}_{\text{field}}^{\mu\nu} = -\mu_0^{-1} \left(\underline{F}^{\mu\lambda} \underline{F}_{\lambda}{}^{\nu} + \frac{1}{4} g^{\mu\nu} (\underline{F}^{\lambda\sigma} \underline{F}_{\lambda\sigma}) \right).$$

Consider a region where the magnetic field is zero. Write out the component T_{zz} (part of the “stress tensor”) in terms of the electric field.

- b. Suppose two identical point charges on the z axis are brought close together. We know they will repel. Draw a picture of the fields near those two poles. Then use your result in (a) to rederive this qualitative conclusion. [*Hint:* Put the charges at $z = \pm a$ and think about what crosses the xy plane.]

CHAPTER 36

Plane Waves in 4D Language

Organize, systematize, consolidate, integrate. Let's see how some more of our earlier results reemerge in our new language.

36.1 LORENZ GAUGE

36.1.1 It's useful

Section 33.8.1 introduced the 4-vector potential via

$$\underline{F}^{\mu\nu} = \underline{\partial}^\mu \underline{A}^\nu - \underline{\partial}^\nu \underline{A}^\mu, \quad [33.13, \text{page 370}]$$

which cast Maxwell's equations as

$$-\underline{\partial}_\mu \underline{\partial}^\mu \underline{A}^\nu + \underline{\partial}_\mu \underline{\partial}^\nu \underline{A}^\mu = \mu_0 \underline{J}^\nu \quad [33.15, \text{page 370}]$$

with gauge invariance under

$$\underline{A}^\mu \rightarrow \tilde{\underline{A}}^\mu = \underline{A}^\mu + \underline{\partial}^\mu \Xi. \quad [33.14, \text{page 370}]$$

We could use this freedom to insist on Coulomb gauge as before. But it's nicer to insist on a Lorentz-invariant condition,¹

$$\underline{\partial}_\mu \underline{A}^\mu = 0. \quad \text{Lorenz gauge} \quad (36.1)$$

Your Turn 36A

Show that in Lorenz gauge, Equation 33.15 become four *decoupled* copies of the wave equation: $\square \underline{A} = -\mu_0 \underline{J}$, or

$$c^{-2} \frac{\partial^2}{\partial t^2} \psi - \nabla^2 \psi = \rho_q / \epsilon_0 \quad \text{and} \quad c^{-2} \frac{\partial^2}{\partial t^2} \vec{A} - \nabla^2 \vec{A} = \mu_0 \vec{j}. \quad \text{Lorenz gauge} \quad (36.2)$$

Unlike our discussion in restricted Coulomb gauge,² Equations 36.2 are valid regardless of whether the charge density is zero or not. They are decoupled, but remember that the Lorenz gauge condition is a constraint linking the four variables ψ and \vec{A} .

¹Named in honor of Ludvig Valentin Lorenz. It's a Lorentz-invariant condition, but not named for Henrik Lorentz.

²See Section 23.1 (page 260).

36.1.2 It's permitted

Can we really insist on Lorenz gauge? Suppose that we had a vector potential not obeying Equation 36.1; that is, $\partial_\mu \underline{A}^\mu = f$ is some arbitrary function. Now apply a gauge transformation $\underline{A}^\mu \rightarrow \underline{A}^\mu + \partial^\mu \Xi$. Then $f \rightarrow f + \square \Xi$. But we have already found the solution to $\square \Xi = -f$ via its Green function in Chapter 23. So an appropriate Ξ exists to bring any 4-vector potential into Lorenz gauge. The whole argument is an upgrade of one we made in magnetostatics (Section 14.3).

36.2 PLANE WAVES

The scalar wave equation has plane-wave solutions of the form

$$\Phi(\underline{X}) = \frac{1}{2}(\exp(i\underline{k}_\mu \underline{X}^\mu) + \text{c.c.}),$$

characterized by a 4-vector $\underline{k}^\mu = \left[\frac{\omega/c}{\vec{k}} \right]^\mu$ (the **4-wavevector**). Such a solution solves the scalar wave equation if $\|\underline{k}\|^2 = 0$ (“ \underline{k} is a **null 4-vector**”). Recall that this is just the condition that the wave moves at speed c .

Similarly to the scalar wave equation, the Maxwell equations in Lorenz gauge have plane wave solutions characterized by a null wavevector \underline{k} . Unlike the scalar field case, each wave also has a polarization 4-vector $\underline{\zeta}$:

$$\underline{A}^\mu(\underline{X}) = \frac{1}{2}\underline{\zeta}^\mu \exp(i\underline{k}_\nu \underline{X}^\nu) + \text{c.c.}$$

This 4-vector field will be in Lorenz gauge if $\underline{k}_\mu \underline{\zeta}^\mu = 0$.

Gauge invariance also implies that we may add any multiple of \underline{k} to $\underline{\zeta}$ without changing the field strengths; this gauge transformation does not spoil Lorenz gauge, because $\underline{k}_\mu \underline{k}^\mu = 0$. We can use this freedom to require that also $\underline{\zeta}^0 = 0$. With that choice,

$$\underline{\zeta}^\mu = \begin{bmatrix} 0 \\ P \\ Q \\ 0 \end{bmatrix}^\mu.$$

Your Turn 36B

- Work out the Faraday tensor and show that the electric field is parallel to $\vec{\zeta}$, and thus perpendicular to \vec{k} .
- Show that the magnetic field is perpendicular both to \vec{k} and to $\vec{\zeta}$.
- Also confirm that your formula for \underline{F} has the expected units.
- Suppose that we had not used our freedom to set $\underline{\zeta}^0 = 0$. That is, suppose that $\underline{\zeta}^\mu = \begin{bmatrix} S \\ P \\ Q \\ S \end{bmatrix}$. What happens when you compute the Faraday tensor this time?

One way to express what you found in (d) is to note that the Faraday tensor contains the *projection* of $\vec{\zeta}$ onto the plane perpendicular to \vec{k} .

36.3 ENERGY AND MOMENTUM

We found the electromagnetic part of the energy-momentum tensor:

$$\underline{T}_{\text{field}}^{\mu\nu} = -\mu_0^{-1} \left(\underline{F}^{\mu\lambda} \underline{F}_{\lambda}{}^{\nu} + \frac{1}{4} g^{\mu\nu} (\underline{F}^{\lambda\sigma} \underline{F}_{\lambda\sigma}) \right). \quad [34.13, \text{page 386}]$$

Your Turn 36C

Now show that for the Lorenz-gauge plane wave, the time-averaged energy-momentum flux tensor is

$$\langle \underline{T}_{\text{field}}^{\mu\nu} \rangle = \frac{1}{2\mu_0} \underline{k}^{\mu} \underline{k}^{\nu} \|\underline{\zeta}\|^2 = \frac{1}{2\mu_0} \underline{k}^{\mu} \underline{k}^{\nu} (|P|^2 + |Q|^2). \quad (36.3)$$

This compact formula contains the energy and momentum densities, and the energy and momentum fluxes, of plane electromagnetic radiation. Our previous derivations of those quantities were less compelling, and anyway did not give us the overall constant of proportionality.

You should confirm that Equation 36.3 has units appropriate for energy density. Note that the two polarizations contribute *independently* to the energy and momentum (no cross-terms). This implies that they cannot interfere with each other; each polarization can only display interference phenomena with itself.

Your Turn 36D

How would Equation 36.3 change if we had instead used a circular polarization basis?

The preceding expression is appropriate for a pure plane wave. For a chaotic mixture of many different plane waves, with uniformly distributed polarizations and directions, the off-diagonal contributions to $\langle \underline{T}_{\text{field}}^{\mu\nu} \rangle$ will average to zero. The energy density, however, is a 3-scalar and will not be zero: $\rho_{\mathcal{E}} = \langle \underline{T}_{\text{field}}^{00} \rangle$. Also, a symmetric rank-two 3-tensor such as $\vec{k} \vec{k}$ (which appears in \underline{T}^{ij}) need not average to zero. For example, the identity tensor $\vec{\mathbf{1}}$ is unchanged by rotations (it's a “tensor from Heaven”). Also, rotation does not affect the trace of a 3-tensor, so the rotational average of $\vec{k} \vec{k}$ must be $\frac{1}{3} \|\vec{k}\|^2 \vec{\mathbf{1}}$.

The diagonal elements of the stress tensor give the *pressure*,³ so we get the simple conclusion that

$$p = \frac{1}{3} \rho_{\mathcal{E}}. \quad \text{equation of state for isotropic EM radiation} \quad (36.4)$$

As mentioned in Section 18.1.3 (page 220), radiation pressure dominates over the gas pressure of ordinary matter in the early Universe, so Equation 36.4 is crucial for cosmology.

³See Section 12.3.1 (page 141).

T₂ Track 2

36.1'a Gravity waves

A nearly identical derivation can be done in the theory of gravitational radiation. This time, there are *ten* components to the polarization tensor (it's a symmetric, rank-2 tensor). Remarkably, a combination of the equation of motion, a suitable Lorentz gauge, and removal of residual gauge artifacts *again* reduces the true number of independent polarizations to just two.

36.1'b Spin versus polarization

You may ask, "If the quantum analog of light is a spin-one particle, then how can there be only two polarizations? After all, other spin-one states (e.g. the p-orbitals of a hydrogen atom, or a triplet bound state of two spin- $\frac{1}{2}$ particles) have *three* angular momentum states!"

This is interesting. You can always take a hydrogen atom, or a positronium "atom" in its triplet state, and view it in its rest frame. Then the usual analysis indeed guarantees three states. But a photon has *no rest frame*. There is thus no guarantee that the third polarization must be present, and it's not.

In contrast, a fundamental particle with spin 1 and *nonzero* mass, for example a W or Z boson, will indeed have a third polarization state.

Similarly, the spin-2 graviton has only two polarizations, not the five we might have expected based on nonrelativistic quantum mechanics. This is again possible only because the graviton is massless.

Even weirder things can happen in a theory that contains massless particles but that is not invariant under spatial inversions. When neutrinos were thought to be massless, theories were written in which only left-handed neutrinos (and only right-handed antineutrinos) existed! There is no Lorentz boost that changes the helicity of a particle moving at speed c , so invariance under the connected part of the Lorentz group does not require that the existence of one helicity entails the existence of the other.

PROBLEMS

36.1 Interference versus polarization

A plane wave of light, monochromatic with frequency ω , gets split into two beams of equal amplitude. The beams travel different distances in vacuum, then recombine at a screen. Let Δ be the difference in path lengths. The incident beam is linearly polarized in some direction $\vec{\zeta}$. Along the way, some optical element may rotate the polarization to $\vec{\zeta}'$ (without changing anything else). If that element is removed, then $\vec{\zeta}' = \vec{\zeta}$.

- Write an expression for the electric field of the recombined beam. Don't worry about overall factors, but do include the dependence on $\vec{\zeta}$, $\vec{\zeta}'$, Δ , and time.
- Write an expression for the time-average of the energy flux in the recombined beam, including its dependence on $\vec{\zeta}$, $\vec{\zeta}'$, Δ . (You may neglect any overall constant.) [Hint: You may prefer to find the energy *density* of the beam instead, then state and use its relation to the energy *flux*.]
- What Electrodynamical Phenomenon that we studied does your answer to (b) help explain?

36.2 Plane waves in Lorenz gauge

In Lorenz gauge, we studied the plane wave with vector potential $\vec{A}(t, \vec{r}) = \frac{1}{2}\vec{\zeta}e^{i(kz - \omega t)}$ +

c.c. We found that its electric and magnetic fields were proportional to

$$\vec{\zeta} - \hat{z}(\hat{z} \cdot \vec{\zeta}) \quad \text{and} \quad \vec{\zeta} \times \hat{z},$$

respectively.

- What is the significance of these results for the paradox that the formula for \vec{A} appears to predict three independent polarizations of light?
- How might we have resolved that paradox without even bothering to compute \vec{E} and \vec{B} , by invoking gauge invariance?

36.3 Waves in 4d notation and $\underline{T}^{\mu\nu}$

- Write down an expression for the 4-vector potential corresponding to a plane wave propagating along $+\hat{z}$, in Lorenz gauge with frequency ω .
- Your answer involves a polarization 4-vector $\underline{\zeta}$. Write down an expression for the most general such $\underline{\zeta}^\mu$. Your answer will involve three independent, arbitrary constants.
- You have found three linearly independent solutions to the wave equation. But we know light has only two independent polarizations! Resolve this discrepancy by calculating the Faraday tensor $\underline{F}^{\mu\nu}$ for this wave and making an Insightful Comment.
- Use your answer to (c) to work out the time-averaged energy-momentum flux tensor for your wave. Your answer will be expressed in terms of ω , the constants you introduced in (b), and some physical constants. Express in words the meaning of each nonzero component of your formula for $\underline{T}_{\text{field}}^{\mu\nu}$ in this situation. Make another Insightful Comment about the roles of the two polarizations in your answer.

[Hints: Remember, you're working in Lorenz gauge; that simplifies the math. Recall the formula is

$$\underline{T}_{\text{field}}^{\mu\nu} = -\mu_0^{-1} \left(\underline{E}^{\mu\lambda} \underline{E}_{\lambda}{}^{\nu} + \frac{1}{4} \underline{g}^{\mu\nu} (\underline{E}^{\lambda\sigma} \underline{E}_{\lambda\sigma}) \right)$$

Stick to 4-dimensional notation; don't bother to reexpress things in terms of \vec{E} and \vec{B} .]

36.4 CMBR polarization

The cosmic microwave background radiation fills all of space. In class I mentioned that even if the CMBR were perfectly isotropic (the same in every direction) when viewed in one inertial frame, nevertheless in another inertial frame it would appear anisotropic, slightly hotter in one direction than in the antipodal direction.⁴

We now ask a different, more detailed question. Suppose that in one inertial frame (the "CMBR frame") the radiation is both isotropic and also *unpolarized*. Will it then appear partially polarized in another inertial frame? To answer this physical question in the context of classical electrodynamics, take the following steps.

Suppose we are moving at velocity $\beta c \hat{z}$ relative to the CMBR frame. Clearly, if we look out in directions $\pm \hat{z}$ we won't detect any apparent polarization, by azimuthal symmetry of the problem. So let's consider looking out in one of the perpendicular directions, say $-\hat{y}$. Now we wonder if there will be some apparent preference for the polarization along \hat{z} relative to \hat{x} , or vice versa.

- Write down the 4-vector potential corresponding to a plane wave of angular frequency ω , moving along $+\hat{y}$. Express the answer using a wave 4-vector \underline{k} and a polarization 4-vector $\underline{\zeta}$. Use the usual complex exponential representation, and assume that $\underline{\zeta}$ is real (linear polarization). It will be convenient to work in Lorenz gauge, i.e. to require $\partial_{\mu} \underline{A}^{\mu} = 0$. What conditions must \underline{k} and $\underline{\zeta}$ obey in order to have a solution to the vacuum Maxwell equations?
- Now apply a Lorentz boost to a frame moving relative to the original frame at speed (βc) in the $+\hat{z}$ direction. Confirm that, when viewed in the new coordinate frame, the wave still obeys the conditions you found in (a). Find the frequency as observed in this new frame. (What is the name for your result?) Find the direction of the wavevector in this new frame. (What is the name for your result?)
- Find the electric field in your wave solution in the original frame of reference. Show that it's unchanged if you replace $\underline{\zeta}$ by $\underline{\zeta} + \xi \underline{k}$ for any constant ξ . Using this freedom, we can simplify the problem by also requiring that $\underline{\zeta}^0 = 0$. Write the most general polarization 4-vector $\underline{\zeta}^{\mu}$ obeying all these requirements. Express it in terms of an amplitude b and the angle ψ that the electric field makes with the \hat{x} -axis.
- Take your boosted polarization vector from (b). Confirm that its electric field, viewed in the new frame, is still transverse. Use the trick in (c) to find an equivalent polarization vector with the convenient property $\underline{\zeta}'^0 = 0$. Express your polarization vector in terms of a new amplitude \tilde{b} and the angle $\tilde{\psi}$ that its electric field makes with the x' -axis. That is, find \tilde{b} and $\tilde{\psi}$ as functions of the original wave's parameters (ω , b , and ψ), and β .

⁴In fact this kinematic effect is much larger than the tiny true anisotropy of the radiation, and must be removed from the data before the true anisotropy is reported.

e. Suppose that Earth is bombarded by cosmic microwave background radiation that, in one inertial frame, is isotropic and unpolarized. That is, the radiation coming from any direction in the sky is a superposition of randomly polarized plane waves, whose polarization angles ψ are uniformly distributed. Find the corresponding distribution of polarization angles $\tilde{\psi}$ and comment.

CHAPTER 37

A Simple Spherical Wave

37.1 FRAMING

Plane waves are nice, but we are never literally going to encounter a wave with infinite, planar wavefronts. On the other hand, we do frequently encounter small sources of light (even a single fluorescent molecule) that we view from far away. Our intuition with mechanical waves leads us to expect some sort of expanding ripple—a **spherical wave** solution to the Maxwell equations.

37.2 SPHERICAL WAVE

37.2.1 Exact solution

We know about spherical waves in acoustics, where the wave function is a scalar. Let's therefore write the simplest possible generalization to a vector potential as a trial solution, and see whether it can be adjusted to work. Our trial solution is just a constant vector times the scalar spherical wave solution:

$$\vec{A}(t, \vec{r}) \stackrel{?}{=} \frac{1}{2} \vec{\xi} \frac{1}{kr} e^{-i\omega t \pm ikr} + \text{c.c.} \quad (37.1)$$

Here k is a scalar, r is distance from the origin, $\vec{\xi}$ is a constant vector, and as usual $\omega = ck$. The prefactor $1/k$ is a conventional choice designed to give $\vec{\xi}$ the same units as the polarization of a plane wave. The upper sign corresponds to outgoing spherical wavefronts; the lower sign to incoming.

Your Turn 37A

- a. The scalar potential ψ is not independent of \vec{A} ; find it by making a similar trial solution

$$A^0(t, \vec{r}) \stackrel{?}{=} \frac{1}{2} \alpha(r) e^{-i\omega t \pm ikr} + \text{c.c.}$$

and imposing the Lorenz gauge condition. Here $\alpha(r)$ is an unknown function that you are to find. The insight is that it may not be a constant, nor even a constant divided by r , but you can still find it.

- b. Confirm that each of the three functions in Equation 37.1 indeed solves the wave equation.
c. Also check that your answer to (a) has this property.

The second result mentioned is not a surprise—sound waves from a point source also have this same form for the air pressure as a function of position and time. What

may be surprising, however, is how the wave energy is distributed. Equation 37.1 has spherical wavefronts. Its amplitude $\|\vec{\xi}\|/(kr)$ is also independent of direction. We might suppose, then, that the wave sends energy isotropically in every direction. Let's calculate.

37.2.2 Far fields

We could now compute exact expressions for the electric and magnetic fields of the spherical wave. But first, consider what we see when we move very far away from the origin along some direction \hat{n} . Out there (near the position $L\hat{n}$), the wavefronts aren't curved very much, and the solution resembles a plane wave¹ with wavevector $\vec{k}_{\text{pw}} = k\hat{n}$ and polarization vector $\vec{\zeta}_{\text{pw},i} = \vec{\xi}_i/(kL)$. We can therefore apply the formulas in Chapter 36.

Your Turn 37B

- Work out the details, including ζ_{pw}^0 .
- Then find the electric and magnetic fields in terms of L , \hat{n} , and $\vec{\xi}$.
- Consider the case where $\vec{\xi}$ is real. Choose spherical polar coordinates with $\vec{\xi}$ pointing along the polar axis. How do the amplitudes of the far fields depend on the angle between \hat{n} and $\vec{\xi}$?

Perhaps surprisingly, the fields (and therefore the energy flux) are *not at all isotropic*. It is true that the wavefronts (loci where $\vec{A} = 0$) are nice concentric spheres. But the amplitudes of the effective plane waves in various directions do depend on angle. They are all maximal in the directions perpendicular to $\vec{\xi}$, and *zero* when we view the wave from far away long the directions $\pm\hat{\xi}$. This pattern of energy flux is sometimes called the **dipole doughnut** pattern.²

The far fields have another crucial property: Both \vec{E} and \vec{B} fall off with distance as $1/L$. So the energy density, and hence also the energy flux, fall off with distance as $1/L^2$. As mentioned in Section 23.5.2 (page 265), this implies that the *total* energy output passing through a sphere of radius L approaches a *constant* as $L \rightarrow \infty$. Whatever creates an exact outgoing spherical wave therefore constantly sends energy all the way out to infinity. (We'll soon see that an oscillating electric dipole can create such a wave.)

37.2.3 Near field

The opposite limit is interesting too. Instead of expanding for large r at fixed ω , sit at a *fixed* distance from the origin and consider the limit $\omega \rightarrow 0$, that is, keep only the leading behavior in powers of ω . You'll find that in this "near field" regime, \vec{E} dominates \vec{B} , and moreover \vec{E} has a very familiar form. The exact spherical wave solution considered in this section *interpolates* between this near-field form, which resembles the dipole field of electrostatics, and the plane-wavy far fields.

¹There are corrections that are higher order in powers of $1/L$.

²A 3D contour map depiction of $\sin^2\theta$ resembles a toroidal pastry, at least when you are hungry.

37.3 A CIRCULARLY POLARIZED SPHERICAL WAVE?

It's also instructive to work out the case of complex $\vec{\xi}$, for example $\hat{x} + i\hat{y}$.

Your Turn 37C

Use the same strategy as Section 37.2: What kind of plane wave does the solution look like when we stand far from the origin along some direction \hat{n} ? Is there any direction in which this wave is circularly polarized? Is there any direction in which it's *linearly* polarized? Can you explain your answers intuitively?

37.4 OTHER KINDS OF SPHERICAL WAVES

The solution considered here is just the simplest of a class of spherical waves. We'll encounter the others when we study radiation systematically. We'll also explain why the solution considered here is generally the dominant part of the radiation given off by oscillating charges.

37.5 INTERFERENCE

Just as with sound, we can imagine a set of point sources of spherical waves, all vibrating in sync. For example, an incoming plane wave could hit an ordered array of atoms, and set them all in synchronized motion; each will then re-radiate some spherical wave. The total fields that land on a distant projection screen can then form a **diffraction pattern**.

Unlike sound, however, the fact that light has two transverse polarizations complicates matters. There is no way that the crests of a wave traveling along \hat{z} and polarized along \hat{x} can cancel the troughs of another wave traveling in the same direction but polarized along \hat{y} , nor will crests combine with crests in the familiar way. Instead, when light from multiple sources lands on a screen, the illumination on each point of the screen involves the *vector sums* of the \vec{E} and \vec{B} waveforms.

37.6 SUMMARY

The plane wave solutions are exact and simple in either Coulomb gauge (Section 17.9) or Lorenz gauge (Chapter 36). They carry energy and momentum. For any \vec{k} there is a two-dimensional vector space of plane waves differing by polarization.³

The exact spherical wave solutions are simpler in Lorenz gauge than in Coulomb gauge. They carry energy and momentum from a point source out to infinity. For any k , there is a three-dimensional vector space of spherical waves (plus many more not considered here). Their wavefronts are spheres (hence the name), but they beam out energy in a “dipole doughnut” pattern that is maximal in the directions perpendicular to $\vec{\xi}$.

³See Sections 17.9 (page 210) and 36.2 (page 398).

PROBLEMS

37.1 *Exact spherical wave solution*

We discussed solutions to Maxwell's equations of the form

$$\vec{A}(t, \vec{r}) = \frac{1}{2r} \vec{\zeta} e^{-i\omega(t-r/c)} + \text{c.c.}$$

Here $\vec{\zeta}$ is a constant vector, and $r = \|\vec{r}\|$.

- a. To finish specifying the solution, I needed to tell you the scalar potential ψ , but I didn't. Instead I said that it was determined by the Lorenz gauge condition, $\vec{\nabla} \cdot \vec{A} = -\dot{\psi}/c^2$. Find an exact formula for this scalar potential.
- b. Consider the case for which $\vec{\zeta} = \hat{x} + i\hat{y}$, where $i = \sqrt{-1}$. Far from the origin, the solution looks like a plane wave. Obtain the limiting forms of the electric field for the cases where:
 - We stand far away along the $+x$ -axis;
 - We stand far away along the $+y$ -axis; and
 - We stand far away along the $+z$ -axis.
- c. Comment on the physical meaning of each result. Which of these directions is getting the largest energy flux, and why?

37.2 *Angular momentum of fields II*

Background: EM waves can also carry *angular* momentum. You found the density of field momentum in Your Turn 34C. So the density of *angular* momentum \vec{J}_3 , computed using the origin as reference point, is $(\mu_0 c^2)^{-1} [\vec{r} \times (\vec{E} \times \vec{B})]_3$. As usual, we will consider only the time average of \vec{J}_3 .

- a. Confirm that the formula given has the appropriate units to be the density of angular momentum.
- b. Consider the outgoing, exact spherical wave solution (Equation 37.1), with complex polarization⁴ $\vec{\xi} = C(\hat{x} + i\hat{y})$. Here C is an overall constant with appropriate units. Work out the electric and magnetic fields far from the origin, to leading order in an expansion in powers of $1/r$.
- c. Use your result in (b) to work out the density of the z component of angular momentum far away from the origin, to leading order in powers of $1/r$. What goes wrong?
- d. Go back to (b) and keep also the first *subleading* terms in the expansion. Then redo (c) retaining those terms.
- e. Because everything moves radially outward at speed c , the radial component of the flux of \vec{J}_3 is your answer to (d), multiplied by c to convert units into a flux. Suppose that a sphere of large radius R surrounds the origin and absorbs all the radiation. Before you compute anything: Do you expect physically that the whole sphere will gain any net angular momentum \vec{J}_3 ? Why/why not?

⁴We'll see later that this solution could represent the radiation given off by a rotating electric dipole at the origin, in electric dipole approximation.

- f. Now integrate the flux of \vec{J}_3 over the surface of the big sphere to get the rate at which angular momentum is transferred to the sphere.
- g. Also find the power absorbed by the sphere.
- h. Divide your answers to (f,g) and comment.

CHAPTER 39

Variational Formulation

x

—y

39.1 FRAMING

Our derivation of $\underline{T}^{\mu\nu}$ in Chapter 34 may have seemed magical—we desired a result (locally conserved energy and momentum), stated some constraints (Lorentz invariant tensor of the appropriate rank, quadratic function of fields), and the only candidate expression worked. But conservation laws should not be magical; they should be general consequences of *symmetries*.

Stepping back a bit, we may notice some habits of highly successful physical theories:

- They are Lorentz invariant.
- They are specified by differential equations, either in time (for particle mechanics) or in spacetime (for fields). That is, they are *local*; for example, they don't involve products of field values at two distant points.
- They generally admit a *variational* formulation; for example, Newton's law arises as the condition for an action functional to be extremal, and a similar result holds for relativistic mechanics as we review below.

We'll now see how these same themes play out in electrodynamics, following Joseph Larmor (1900) and others. Then we'll see how a variational formulation establishes conservation laws corresponding to continuous invariances of a field theory, a result known as Emmy Noether's theorem.

39.2 NEWTONIAN MECHANICS

Given any particle trajectory, we compute its **action** by evaluating the **action functional**,¹ which is the time integral of kinetic minus potential energy. For one-dimensional motion,

$$S[x(t)] = \int dt \mathcal{L}(x, \frac{dx}{dt}).$$

Here the notation $S[x(t)]$ means that S depends on an entire trajectory $x(t)$. The **lagrangian density** \mathcal{L} is an ordinary function of two variables:

$$\mathcal{L}(x, \frac{dx}{dt}) = \text{KE} - \text{PE} = \frac{m}{2} \left(\frac{dx}{dt} \right)^2 - U(x).$$

¹Many authors shorten “lagrangian density” to “lagrangian.”

Let's characterize those trajectories for which S is extremal over the space of all trajectories with fixed values at two time points:

$$\delta S = 0 = \int_0^T dt \delta \left[\mathcal{L} \left(x, \frac{dx}{dt} \right) \right] = \int_0^T dt \left(m \frac{dx}{dt} \frac{d\delta x}{dt} - \delta x \frac{dU}{dx} \right).$$

We now integrate by parts. Because we consider only variations that hold the endpoint values fixed, $\delta x = 0$ at time 0 and T , so the boundary terms equal zero:

$$0 = - \int_0^T dt (\delta x) \left[- \frac{d^2 x}{dt^2} - \frac{dU}{dx} \right].$$

The only way this first-order variation could equal zero for any variation $\delta x(t)$ is if the terms in square brackets cancel at each time:

$$\frac{d^2 x}{dt^2} = - \frac{dU}{dx}.$$

But that last formula is Newton's law.

Generalizing to many variables, we see we can reexpress Newtonian mechanics as a statement about the variation of the action functional. If moreover the action functional has some invariance, for example under translations or rotations, then that fact is also reflected in the resulting equations of motion.

39.3 FIELD EQUATIONS

39.3.1 Scalar field

We now upgrade the variational formulation to accommodate fields. We consider functionals of fields that are local:

$$S[\text{traj}] = \int d^4 X \mathcal{L}(\underline{A}, \underline{\partial A}).$$

For a scalar field, the lagrangian density \mathcal{L} is an ordinary function of five variables (the field and its four space and time derivatives at every point). More generally, \mathcal{L} is a local function of five variables for each component of the field A , for example, the 4-vector potential in electrodynamics

We will also require that \mathcal{L} be a 4-scalar function of the fields. Because $d^4 X$ is scalar (Section ?? (page ??)), therefore S will be Lorentz-invariant. Also, field theories such as electrodynamics are invariant under translations in space or time.

Adapting the preceding discussion shows that the generic variational equation is

$$\partial_\mu \left(\frac{\delta \mathcal{L}}{\delta (\partial_\mu \phi)} \right) - \frac{\delta \mathcal{L}}{\delta \phi} = 0 \quad \text{Euler-Lagrange equation} \quad (39.1)$$

The first term on the left denotes the result when we

- Vary \mathcal{L} with respect to one of its derivative variables, then

- Substitute values of ϕ and $\underline{\partial}_\mu\phi$, obtaining a function of \underline{X} , and
- Take a derivative with respect to \underline{X} .

(The second term denotes the variation with respect to the undifferentiated ϕ , again followed by substituting values of ϕ and $\underline{\partial}_\mu\phi$.)

A specific choice for the lagrangian density of a scalar field could be

$$\mathcal{L}[\phi] = \frac{1}{2}(\underline{\partial}^\mu\phi\underline{\partial}_\mu\phi - \lambda^{-2}\phi^2). \quad (39.2)$$

Equation 39.1 then says $-\underline{\partial}^\mu\underline{\partial}_\nu\phi - \lambda^{-2}\phi = 0$, in this context also called the **Yukawa equation** for its role in an early theory of nuclear forces.² Its solutions correspond to static interactions that fall exponentially with distance as $\exp(-L/\lambda)$.

39.3.2 Maxwell/Lorentz system

Beyond translations in space or time, electrodynamics is also gauge invariant. Can we find an action functional meeting all of the symmetry requirements, and whose variational equation recovers the Maxwell equations?

The Maxwell equations are linear in the vector potential, and second-order in its derivatives (Equation ??, page ??). So \mathcal{L} must be a *quadratic* function of \underline{A} , with at most two derivatives. There are very few such functions that are also gauge- and translation invariant:

- $\underline{F}^{\mu\nu}\underline{F}^{\lambda\sigma}\varepsilon_{\mu\nu\lambda\sigma}$: This term can be rewritten as $2\underline{\partial}^\mu(\underline{A}^\nu\underline{F}^{\lambda\sigma}\varepsilon_{\mu\nu\lambda\sigma})$, that is, as a total 4-divergence. Therefore its integral S is a boundary term, by the divergence theorem, and hence makes *no contribution* to the local variation of S .
- The expressions $(\underline{\partial}_\mu\underline{A}^\mu)^2$ and $\underline{A}_\mu\square\underline{A}^\mu$ are not gauge invariant. The expression $\underline{A}_\mu\underline{A}^\mu$ is not gauge invariant, and moreover contains no derivatives.
- $\underline{F}^{\mu\nu}\underline{F}_{\mu\nu}$ is the only remaining option.

We now find the first order variation of our candidate action functional and set it to zero:

$$0 = 2 \int d^4X \underline{F}_{\mu\nu}\delta(\underline{F}^{\mu\nu}) = 2 \int d^4X \underline{F}_{\mu\nu}(\underline{\partial}^\mu\delta\underline{A}^\nu - \underline{\partial}^\nu\delta\underline{A}^\mu) \quad (39.3)$$

$$= 4 \int d^4X \underline{F}_{\mu\nu}\underline{\partial}^\mu\delta\underline{A}^\nu \quad (39.4)$$

$$= -4 \int d^4X (\underline{\partial}^\mu\underline{F}_{\mu\nu})\delta\underline{A}^\nu. \quad (39.5)$$

For this quantity to vanish regardless of $\delta\underline{A}$, we must have that $\underline{\partial}^\mu\underline{F}_{\mu\nu} = 0$. This is indeed Maxwell's equations in vacuum.

39.3.3 Fields plus particles

Suppose that charged particles are present and executing prescribed motions; that is, we don't inquire yet into the equations of motion for particles. We can construct the charge flux 4-vector \underline{J} and add it $\underline{A}_\mu\underline{J}^\mu$ our lagrangian density, because:

²When quantized, the field ϕ is associated to particle states that could represent pions.

- This term is Lorentz- and translation invariant.
- \underline{J} is gauge invariant, so under gauge transformation we have (Equation ??, page ??)

$$\int d^4X \underline{J}_\mu \underline{A}^\mu \rightsquigarrow \int d^4X (\underline{J}_\mu \underline{A}^\mu + \underline{J}^\mu \partial_\mu \phi).$$

Integrating by parts shows that the second term equals zero by the continuity equation.

Combining the pure-field term from Section 39.3.2 with the particle term just found and choosing appropriate constants gives finally

$$\mathcal{L}[\underline{A}, \partial \underline{A}] = -\frac{1}{c} \left(\frac{1}{4\mu_0} \underline{F}_{\mu\nu} \underline{F}^{\mu\nu} + \underline{A}_\mu \underline{J}^\mu \right). \tag{39.6}$$

The corresponding variational equation is the Maxwell equation with charges and currents (Equation ??, page ??). The overall factor of $1/c$ gives our action functional the traditional units (Js).

Until now, we have assumed that particle motions were given. However, we can extend the theory to include equations of motion for the particles as well as the fields by adding a kinetic energy term for each one via

$$S_{\text{part},\ell}[\text{trajectory } \ell] = \int d\tau \frac{1}{2} (m_\ell)^{-1} \|\underline{p}_\ell\|^2.$$

Your Turn 39A

Show that adding the preceding terms to the action functional, and using Equation ?? (page ??), leads to a variational equation that is precisely the Lorentz force law for particle ℓ .

39.4 NOETHER THEOREM

39.4.1 Scalar field example

To warm up, let’s again begin with a simpler system, consisting of a single scalar field ϕ . We now explore the consequences of a continuous symmetry, that is, a field transformation that leaves the equations of motion form-invariant and that changes fields by an infinitesimal amount. Accordingly, consider a general transformation under which

$$\phi(\underline{X}) \rightsquigarrow \tilde{\phi}(\underline{X}) = \phi(\underline{X}) + \epsilon D[\phi](\underline{X}) + \dots \tag{39.7}$$

Here the ellipsis denotes terms of higher order in the bookkeeping parameter ϵ ; from now on we will drop such terms without comment. D is a local expression in fields and their derivatives, which is then evaluated at each spacetime point \underline{X} . We suppose that the expression just given leaves S invariant for any trajectory $\phi(\underline{X})$, then ask for consequences in the situation where ϕ also obeys the variational equation associated to its action functional.

Here are some concrete examples:

- A translation (shift of \underline{X} by a constant 4-vector $\epsilon \underline{b}$) corresponds to the local functional $D[\phi] = \underline{b}^\mu \partial_\mu \phi$, as we see by Taylor expanding ϕ .
- Next, consider a set of *two* scalar fields, each with its own lagrangian density of the form Equation 39.2. Then

$$D \begin{bmatrix} \phi^{(1)} \\ \phi^{(2)} \end{bmatrix} = \begin{bmatrix} \phi^{(2)} \\ -\phi^{(1)} \end{bmatrix}$$

generates an infinitesimal rotation in the *internal space* of ϕ 's components (not in physical space).

39.4.2 Consequences of invariance

We cannot assume that the lagrangian density is unchanged by an invariance, but we at least know that its change, if any, must be a total derivative (so that its integral will be a pure boundary term and hence zero for any localized variation). Thus, for each infinitesimal invariance of the system we have

$$\mathcal{L} \rightsquigarrow \mathcal{L}[\tilde{\phi}, \underline{\partial} \tilde{\phi}] = \mathcal{L}[\phi, \underline{\partial} \phi] + \epsilon \underline{\partial}_\mu \mathcal{I}^\mu[\phi, \underline{\partial} \phi]. \quad (39.8)$$

Here \mathcal{I}^μ is some local functional of fields and their derivatives that we can find from the chosen lagrangian density and invariance under consideration. Continuing the two examples in the preceding section,

- For internal rotations, XXXX.
- For translation by \underline{b} , we have $\mathcal{I}^\mu = \frac{1}{2} \partial_\nu \underline{b}^\nu \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} \lambda^{-2} \underline{b}^\nu \phi^2$.

We will now find a 4-vector field associated to our assumed invariance that obeys a continuity relation, and hence defines a conserved “charge.” To do this, first substitute Equation 39.7 into Equation 39.8:

$$\mathcal{L}[\phi, \underline{\partial} \phi] + \epsilon D[\dots] \underline{\partial}_\mu \left(\frac{\delta \mathcal{L}}{\delta \phi} \right) + \epsilon (\underline{\partial}_\mu D[\dots]) \frac{\delta \mathcal{L}}{\delta (\underline{\partial}_\mu \phi)} + \mathcal{O}(\epsilon^2) = \mathcal{L}[\phi, \underline{\partial} \phi] + \epsilon \underline{\partial}_\mu \mathcal{I}^\mu + \mathcal{O}(\epsilon^2).$$

Comparing the sides of this equation shows that the 4-vector quantity

$$\underline{J}^\mu = \frac{\delta \mathcal{L}}{\delta (\underline{\partial}_\mu \phi)} D[\phi, \underline{\partial} \phi] - \mathcal{I}^\mu$$

obeys $\underline{\partial}_\mu \underline{J}^\mu = 0$ as promised.

Returning to our two examples,

- XXXX
- Turning now to translation by \underline{b} ,

$$\underline{J}^\lambda = \underline{\partial}^\lambda \phi \underline{b}^\mu \underline{\partial}_\mu \phi - \frac{1}{2} \underline{b}^\lambda \partial^\mu \phi \partial_\mu \phi + \frac{1}{2} \lambda^{-2} \underline{b}^\lambda \phi^2.$$

We can summarize all four of the associated continuity equations as

$$\underline{\partial}_\mu \underline{T}^{\mu\nu} = 0 \quad \text{where} \quad \underline{T}^{\mu\nu} = \underline{\partial}^\mu \phi \underline{\partial}^\nu \phi - \frac{1}{2} g^{\mu\nu} (\underline{\partial}_\sigma \phi \underline{\partial}^\sigma \phi - \lambda^{-2} \phi^2).$$

In fact, the symmetric tensor \underline{T} just defined is the energy-momentum flux tensor of the scalar field theory under consideration.

39.4.3 Electrodynamics

Let's upgrade these ideas to electrodynamics, with the lagrangian density Equation 39.6.

[Not ready yet.]

39.5 *PLUS ULTRA*

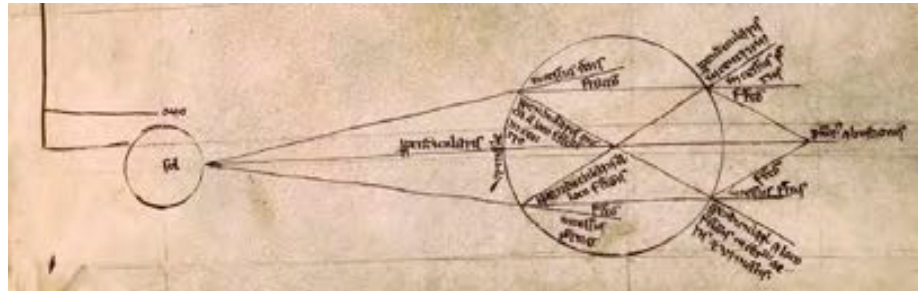
Why are all known fundamental physical theories expressible as variational principles? I don't know, but it's relevant that the *quantum* version of any such a theory can be straightforwardly constructed by a path integral: Simply divide the action by \hbar (which has units of action), multiply by $\sqrt{-1}$, and exponentiate to obtain a phase. Integrating that phase over all trajectories yields quantum amplitudes.

FURTHER READING

Chen et al., 2019; Peskin & Schroeder, 1995, ch. 2.

PART VI

Radiation



Optical studies from Roger Bacon's *De multiplicatione specierum*. The diagram shows light being refracted by a spherical glass container full of water.

CHAPTER 40

Radiation Green Function

40.1 FRAMING

- We found a spherical wave solution in Chapter 37, but we have not yet seen how such a wave may be created.
- We saw how certain waves may be generated in Chapter 23, but with some limitations on what kind of source.
- The antenna considered in Section 23.5 did generate a spherical wave, but not the same one as what we found in Chapter 37! We need a more general understanding of spherical waves.
- Chapter 23 found a solution to the wave equation, but by the unsatisfying method of “lucky guess.” Let’s use “Einstein thinking” to recover that result more straightforwardly, and to give the generalizations requested in the preceding points.

40.2 TIME ORDERING AND CAUSALITY

Is the upper-left corner of this page higher or lower than the upper-right corner? Obviously there’s no absolute answer to that question. The higher corner can be made lower by rotating the page. On the other hand, if you stub your toe in the night, and a dog barks on the next block, there doesn’t seem to be any doubt about which happened first. Or is there?

We now believe that newtonian physics is an imaginary world, not an accurate description of Nature. But its galilean transformations do have a nice property (Figure 40.1a): Any two G -inertial reference frames¹ will agree that event \mathbf{R} is simultaneous with \mathbf{P} , whereas \mathbf{S} precedes \mathbf{P} and \mathbf{Q} follows \mathbf{P} . Geometrically, this is a matter of whether you’re above or below the x axis, and *all G -inertial frames have the same x axis.*

Turning to Lorentz transformations, which do seem to be an invariance of Nature, we find a surprise: An observer moving to the right (figure panel (b)) will disagree with the original observer, saying that \mathbf{R} precedes \mathbf{P} (it lies below the x' axis)!²

Similarly, a leftward-moving observer (panel (c)) would say that \mathbf{R} and even \mathbf{S} happen later than \mathbf{P} . Interestingly, however, all E -inertial observers agree that \mathbf{T} is later than \mathbf{P} , and \mathbf{U} is earlier. That’s because these points lie beyond the wavy lines at $\pm 45^\circ$ to the axis, and we can never bend the x' axis past those lines.

¹Recall that a G -inertial frame is one in which the newtonian equations of physics take their usual form.

²An even faster-moving observer would also say that \mathbf{Q} precedes \mathbf{P} !!

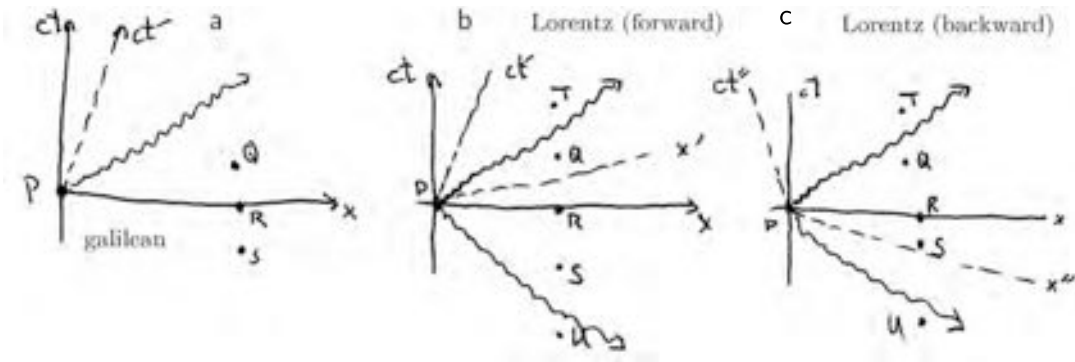


Figure 40.1: In each panel the wavy line depicts a light trajectory.

Algebraically, we say that the temporal ordering of **P** and **Q** is unambiguous if and only if $|t_Q - t_P| > \|\vec{r}_Q - \vec{r}_P\|/c$. We can restate this using the invariant interval:³ The temporal ordering of two events **P** and **Q** is unambiguous if $\Delta\tau^2$ is positive, that is, if $\|\Delta\underline{X}_{PQ}\|^2 \leq 0$. Section 31.5.2 introduced the terms timelike separation if $\|\Delta\underline{X}_{PQ}\|^2 < 0$, lightlike if it's exactly zero, or spacelike if it's positive. Temporal ordering is ambiguous (dependent on which initial frame we choose) if the separation is spacelike.

The **relativity of simultaneity** just discovered may seem to be a disaster for physics. How can we claim that anything “caused” anything else, if we don't know which happened first?! But it's not a complete disaster: When two events have timelike or lightlike separation, then we *do* know for sure which was first. So we can get out of difficulty if we insist that

If two events are spacelike separated, then neither one may be said to have caused, or even influenced, the other.

This makes sense when we notice that, in order for two such events to influence each other, one would have to send a signal to the other moving faster than the speed of light in vacuum.⁴ Really all we're asserting, then, is that no signal (information, causal agent) can move faster than light. This prohibition is consistent with the relativistic velocity addition formula, which always yields a new velocity $\leq c$. Now we see that the speed limit is also *necessary* to avoid a physically nonsensical confusion about causality.

40.3 RETARDED GREEN FUNCTION

³Note that the notation $\|\vec{r}\|^2$ denotes the ordinary length-squared of a 3-vector, whereas $\|\Delta\underline{X}\|^2$ denotes the invariant interval.

⁴What about quantum entanglement? Luckily that's not part of this course, but every discussion I've seen ends up, after a lot of analysis, concluding that there's still no way to transmit useful information faster than c .

40.3.1 Scalar wave equation

Section 36.1 obtained a version of Maxwell's equations valid in Lorenz gauge:

$$\square \underline{A}^\mu = -\mu_0 \underline{J}^\mu. \quad (40.1)$$

This is four decoupled copies of a single equation, so to simplify the notation we'll first solve the scalar inhomogeneous wave equation:

$$\square \phi = -\mathcal{J} \quad (40.2)$$

and later add the 4-vector index and factor of μ_0 . Chapter 23 found a solution to the inhomogeneous wave equation, but (a) we had to make an unobvious guess, and (b) Equation 23.5 didn't look exactly like a Green function solution. Let's use "Einstein thinking" to do better.

The wave equation is linear and translation-invariant, so we expect that the solution can be written in terms of a Green function:

$$\phi(\underline{X}) = \int d^4 X_* D_r(\underline{X} - \underline{X}_*) \mathcal{J}(\underline{X}_*). \quad (40.3)$$

We now use invariance to constrain the possible form of the function D_r , show that there is only one reasonable choice, then confirm that with that choice, the formula Equation 40.3 solves Equation 40.2 for any source function \mathcal{J} .

The constraints are that

- D_r must be a Lorentz-invariant, scalar function of the 4-vector $\Delta \underline{X} = \underline{X} - \underline{X}_*$.
- It must have dimensions $(\text{length})^{-2}$, by Equation 40.2.
- It must vanish when $\Delta X^0 < 0$, because the behavior of charges in the future cannot affect the values of fields in the past.

We can satisfy all these conditions with a function of this form:

$$D_r(\Delta \underline{X}) = \frac{1}{2\pi} \delta(\|\Delta \underline{X}\|^2) \Theta(\Delta X^0). \quad \text{radiation Green function} \quad (40.4)$$

Taking the factors in turn,

- Soon we'll see why the prefactor must be $1/(2\pi)$.
- The delta function is motivated by the idea that electromagnetic influences always travel at speed c .⁵ Two points can be joined by a path traversed at speed c only if they are lightlike-separated.
- The last factor is a "Heaviside step function," and it enforces causality. Together with the delta function, it says that fields at \underline{X} can only be influenced by sources lying in the "past light cone of \underline{X} ."
- The delta function has dimensions⁶ inverse to $(\text{length})^2$. The step function is dimensionless. So our proposal has the desired units.

⁵See Section 23.3 (page 261).

⁶See Section 0.3.6 (page 9).

The argument of the delta function is a constant times the invariant interval, so this whole factor is Lorentz invariant. The step function looks bad at first, because Lorentz transformations can affect the temporal ordering of two events: $\Delta \underline{X}'^0$ may not have the same sign as $\Delta \underline{X}^0$. However, this problem can only arise for spacelike-separated events, that is, a pair of events with invariant interval less than zero. The delta function tells us that such events cannot contribute anything to the proposed Green function. Only *lightlike*-separated events contribute, and the temporal ordering of any such pair of events is unambiguous.

In short, D_r is a 4-scalar function. The other ingredient in Equation 40.3 is d^4X , which we saw in Section 33.9.3 is also Lorentz-invariant. Thus, Equation 40.3 is a Lorentz-invariant recipe to obtain ϕ from \mathcal{J} , as desired.

Our trial solution has all the qualitative properties we expect it should have. Now we need to confirm that it really solves the wave equation. But once that's done, *everything about radiation will follow from Equation 40.4.*

40.3.2 Reformulate and confirm

Our proposed Green function is simple, and seems promising. After admiring it, we now rephrase it in a way that obscures Lorentz invariance but will facilitate checking that it does solve the wave equation.

We want to substitute our guess Equation 40.4 into Equation 40.3 and ultimately confirm that Equation 40.2 is valid. After the substitution, we've got four integrals and one delta function. We will now use the delta function to eliminate one of the integrals, specifically the one over $\underline{X}_*^0 = ct_*$.

Recall from Section 33.9.1 how delta functions transform:

$$\delta(f(t_*)) = \sum_{\ell} |f'(t_{*,\ell})|^{-1} \delta(t_* - t_{*,\ell}),$$

where $t_{*,\ell}$ are all the values of t_* at which $f(t_{*,\ell}) = 0$. For our application,

$$f(t_*) = -c^2(t - t_*)^2 + R^2 \quad \text{where } R = \|\vec{r} - \vec{r}_*\|.$$

The quantities t and R are constants for purposes of evaluating the integral over t_* .

There are two solutions to $f = 0$: $t_{*-} = (ct - R)/c$ and $t_{*+} = (ct + R)/c$. Of these, however, the second is acausal and so cannot contribute (the step function eliminates it). Turning to the first,

$$\left. \frac{df}{dt_*} \right|_{t_{*-}} = 2c^2(t - t_{*-}) = 2c^2(t - t + R/c) = 2cR.$$

Thus,

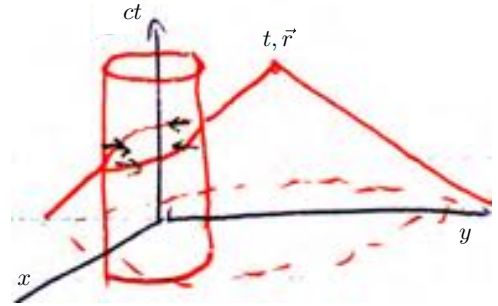
$$\delta(\|\Delta \underline{X}\|^2) \Theta(\Delta \underline{X}^0) = \frac{1}{2cR} \delta(t_* - t + R/c).$$

That result lets us easily do the t_* integral in Equation 40.3. The three remaining integrals become

$$\phi(t, \vec{r}) = \frac{1}{4\pi} \int d^3r_* \frac{1}{R} \mathcal{J}(t - R/c, \vec{r}_*). \quad [23.5, \text{page } 261]$$

We have already confirmed that ϕ defined by this formula solves the wave equation in Chapter 23. This time, however, we found it without having to make such a lucky guess, by “Einstein thinking” (insisting on Lorentz invariance and causality).

Figure 40.2: The cylinder represents a circular loop of current fixed in the xy plane. The past light cone of an observer at time t and position \vec{r} intersects the region of nonzero \vec{j} in the ellipse shown. Only points on that ellipse contribute to the potentials observed at t, \vec{r} .



40.4 REMARKS

40.4.1 Upgrade to 4-vector fields

Equation 23.5 is pretty simple: For each location inside the source, it tells us to look back in time to the moment when charges and currents at that location could have influenced the fields at (t, \vec{r}) , then introduce a factor of $1/(4\pi R)$. It's easy to upgrade this result to electrodynamics, because Equation 40.1 is four decoupled copies of the wave equation; just use the scalar solution four times with $\mathcal{J} = \mu_0 \underline{J}^\mu$:

$$\underline{A}^\mu(\underline{X}) = \mu_0 \int d^3 r_* \frac{1}{4\pi \|\vec{r} - \vec{r}_*\|} \underline{J}^\mu(\underline{X}^0 - \|\vec{r} - \vec{r}_*\|, \vec{r}_*). \quad \text{Lorenz gauge} \quad (40.5)$$

This result looks a bit like the one we found in Coulomb gauge (Chapter 23). Unlike that result, however, this one assigns a nonzero value to the scalar potential. It is also valid even when the charge density is not everywhere zero.

Our recipe gets especially simple for a point charge sitting at rest at the origin, because $\vec{J} = 0$ while ρ_q is time-independent. So our solution reproduces the static potential of a point charge.

More generally, we have shown that Equation 40.5 with Equation 40.4 gives the fields created (caused by) a general distribution of charges and currents. Other names for this **causal Green function** are **retarded Green function** or **retarded propagator**. The names refer to the fact that the formula “looks back in time.” For example, Figure 40.2 shows a current loop.

40.4.2 Check self-consistency

We're not quite done. Equation 40.1 is *not equivalent to Maxwell* unless \underline{A} is in Lorenz gauge. Does our solution have that property?

To find out, we must compute

$$\partial_\mu \underline{A}^\mu = \int d^4 X_* \underline{J}^\mu(\underline{X}_*) \frac{\partial}{\partial \underline{X}^\mu} D_r(\underline{X} - \underline{X}_*).$$

Next, note that

$$\frac{\partial}{\partial \underline{X}^\mu} D_r(\underline{X} - \underline{X}_*) = -\frac{\partial}{\partial \underline{X}_*^\mu} D_r(\underline{X} - \underline{X}_*).$$

After that substitution, we can integrate by parts to find⁷

$$\partial_{\mu} \underline{A}^{\mu} = \int d^4 X_* D_r(\underline{X} - \underline{X}_*) \frac{\partial}{\partial \underline{X}_*^{\mu}} \underline{J}^{\mu}(\underline{X}_*).$$

The right side of this expression is zero, by the continuity equation that any 4-current distribution must obey.

You may be dissatisfied: “The Green function is the response to a blip, but an isolated blip cannot obey the continuity equation!” The logic is that

- . The Green function is indeed a solution to the wave equation, Equation 40.1, for an isolated blip source.
- . If we assemble a lot of blips together into a \underline{J} field that obeys the continuity equation, then we just showed that our solution will also be in Lorenz gauge;
- . and therefore, it will also solve the Maxwell equations.

⁷The reasoning here is similar to something we used in magnetostatics, Section 14.4.3 (page 166), and again later in our first look at radiation, Your Turn 23A (page 261).

T₂ Track 2

40.3'a Alternative derivation in Fourier space

David Chow notes:

At times it will be necessary to move between covariant and non-covariant forms, but the motivations for each will be explained. The EM fields $F^{\alpha\beta}$ arising from the 4-current $J^\alpha(x)$ satisfy the inhomogeneous Maxwell equations

$$\begin{aligned}\partial_\alpha F^{\alpha\beta} &= \frac{4\pi}{c} J^\beta \quad \text{which we can write in terms of the potentials,} \\ \square A^\beta - \underbrace{\partial^\beta (\partial_\alpha A^\alpha)}_{=0} &= \frac{4\pi}{c} J^\beta\end{aligned}$$

Here $\square \equiv \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}$ and the Faraday tensor $F^{\alpha\beta} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}$ as usual. The braced term vanishes by choice of the Lorentz condition, so we may solve this by finding a Green's function satisfying

$$\square_x D(x, x') = \delta^{(4)}(x - x').$$

But the Green's function can only depend on the difference $z^\alpha = x^\alpha - x'^\alpha$ due to the lack of boundaries, so

$$\square_z D(z) = \delta^{(4)}(z). \quad (40.6)$$

Applying a Fourier transform to eqn 40.6, $D(z) = \frac{1}{(2\pi)^4} \int d^4k \tilde{D}(k) e^{-ik \cdot z}$, and recalling that $\delta^{(4)}(z) = \frac{1}{(2\pi)^4} \int d^4k e^{-ik \cdot z}$, hitting eqn 40.6 with two derivatives from the dalembertian brings down two factors of $-ik$. Therefore we can solve for $\tilde{D}(k)$ to be

$$\tilde{D}(k) = -\frac{1}{k \cdot k}, \text{ and so } D(z) = -\frac{1}{(2\pi)^4} \int d^4k \frac{e^{-ik \cdot z}}{k \cdot k}.$$

Notice that we have a singularity to deal with. Let's deal with the k_0 part first:

$$D(z) = -\frac{1}{(2\pi)^4} \int d^3k e^{i\mathbf{k} \cdot \mathbf{z}} \int_{-\infty}^{\infty} dk_0 \frac{e^{-ik_0 z_0}}{k_0^2 - \kappa^2}, \text{ where } \kappa \equiv |\mathbf{k}|.$$

Treating k_0 as complex so that we have simple poles at $k_0 = \pm\kappa$, we have two options as to the contours we may use to evaluate the integral (Figure 40.3). Let's name them r and a , with r the one in the positive imaginary half-plane and a in the negative. They will be closed with semicircles out at $\pm i\infty$. We need to enclose the poles and also have the term $e^{-ik_0 z_0}$ kill the contribution of the semicircle, so the sign of z_0 decides which contour we may use. First let's take contour r , for which $z_0 > 0$ so that it's closed in the lower half plane. Then

$$\oint_r dk_0 \frac{e^{-ik_0 z_0}}{k_0^2 - \kappa^2} = -2\pi i \text{Res} \left(\frac{e^{-ik_0 z_0}}{k_0^2 - \kappa^2} \right) = -\frac{2\pi}{\kappa} \sin(\kappa z_0);$$

thus

$$D_r(z) = \frac{\theta(z_0)}{(2\pi)^3} \int d^3k e^{i\mathbf{k} \cdot \mathbf{z}} \frac{\sin(\kappa z_0)}{\kappa},$$

where θ the Heaviside function enforces the condition on z_0 . Integrating over angles gives

$$\begin{aligned}D_r(z) &= \frac{\theta(z_0)}{2\pi^2 R} \int_0^\infty d\kappa \sin(\kappa R) \sin(\kappa z_0), \text{ where } R \equiv |z| \\ &= \frac{\theta(z_0)}{8\pi^2 R} \int_{-\infty}^\infty d\kappa \left(e^{i(z_0 - R)\kappa} - \underbrace{e^{i(z_0 + R)\kappa}}_{=0} \right).\end{aligned}$$

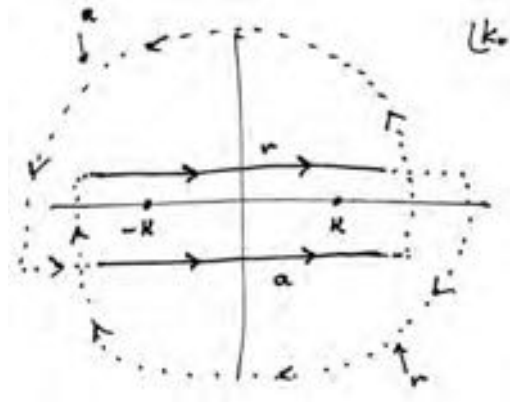


Figure 40.3:

Notice these are delta functions, so the second one vanishes because $z_0 > 0$ and $R > 0$. Thus $D_r(x - x') = \frac{\theta(x_0 - x'_0)}{4\pi R} \delta(x_0 - x'_0 - R)$. This is the (noncovariant) retarded Green's function; a nearly identical calculation gives $D_a(x - x') = \frac{\theta(-(x_0 - x'_0))}{4\pi R} \delta(x_0 - x'_0 + R)$. To put these into covariant form we'll use the identity

$$\begin{aligned} \delta[(x - x')^2] &= \delta[(x_0 - x'_0)^2 - |\mathbf{x} - \mathbf{x}'|^2] \\ &= \delta[(x_0 - x'_0 - R)(x_0 - x'_0 + R)] \\ &= \frac{1}{2R} [\delta(x_0 - x'_0 - R) + \delta(x_0 - x'_0 + R)]; \end{aligned}$$

here the θ functions always select the term you want in the last line, so we can write

$$\begin{aligned} D_r(x - x') &= \frac{1}{2\pi} \theta(x_0 - x'_0) \delta[(x - x')^2] \\ D_a(x - x') &= \frac{1}{2\pi} \theta(x'_0 - x_0) \delta[(x - x')^2]. \end{aligned}$$

You may be concerned that the θ functions are not invariant; however, when constrained in this way by the delta functions, you will find that they are. Therefore, the solutions we sought are

$$\begin{aligned} A^\alpha(x) &= A_{\text{in}}^\alpha + \frac{4\pi}{c} \int d^4x' D_r(x - x') J^\alpha(x'), \text{ or,} \\ A^\alpha(x) &= A_{\text{out}}^\alpha + \frac{4\pi}{c} \int d^4x' D_a(x - x') J^\alpha(x'), \end{aligned}$$

with A_{in}^α and A_{out}^α solutions to the homogeneous (sourceless) wave equations.

T2 Track 2

40.4'

When we expand Einstein's gravitational equation for small fluctuations about flat space, the result is a second-order PDE. It has a complicated tensor structure, until we impose a

gauge condition analogous to Lorenz gauge. Then it becomes ten decoupled copies of the same old wave equation that we just solved! Thus, *everything about (weak) gravitational radiation also follows from Equation 40.4.*

CHAPTER 41

J. J. Thomson's Pictorial Explanation of Radiation

41.1 FRAMING

The Gauss law implies that a static positive point charge creates an electric field that is directed radially outward and falls as r^{-2} . That behavior is quite different from radiation. For example, the energy density of such a field configuration falls as r^{-4} , too fast to transport any energy to infinity. But there are many other solutions to the Maxwell equations. In particular, when a charge is in motion, then it's no longer a spherically symmetric source, so we need not expect a spherically symmetric solution. Indeed, Sections 32.3.2 and 33.8.2 found bunching of the field into the equatorial plane.

This chapter will extend the discussion to accelerating charges by abstracting just one qualitative fact from Chapter 40: Disturbances in the field propagate at the fixed, finite speed c . From just those two ideas, and Michael Faraday's field-line concept, J. J. Thomson built a pictorial explanation that gives most of the qualitative features of the electric field arising in radiation. Adding Faraday's law of induction will then let us understand the magnetic field as well. Later chapters will work through the analytic details, but it's good to have this intuition first.

41.2 ELECTRIC FIELDS FROM AN ACCELERATING CHARGE

The lower panel of Figure 41.1 shows the trajectory in spacetime of a particle that is motionless from time $-\infty$ till time zero, then accelerates along \hat{x} , then decelerates to rest. At some time t_f after that last step, we ask what the fields look like throughout space.

- A very distant observer, at \mathcal{O}_1 , has not yet learned that the particle is no longer stationary at P , so it sees radial \vec{E} pointing outward from P toward \mathcal{O}_1 . A ring of such observers, all at the same distance, see uniformly spaced field lines with transverse density $1/r_{\text{OP}}^2$ (outermost arrows in the upper part of the figure).
- At the other extreme, a very nearby observer, at \mathcal{O}_3 , sees the up-to-date information, that is, radial \vec{E} pointing from Q .
- In between, an observer, at \mathcal{O}_2 , sees radial \vec{E} pointing from R because that's where the charge *would have been* at time t_f , had it not decelerated, and this observer has not yet had a chance to learn that the charge has decelerated.¹

¹This has nothing to do with the mental state of the observer. The causal structure of the theory is such that no instrument can, at the point in space and moment in time, distinguish the trajectory from one that is in eternal, uniform straight-line motion, and Section 32.3.2 showed that the field created in that situation is as described here.

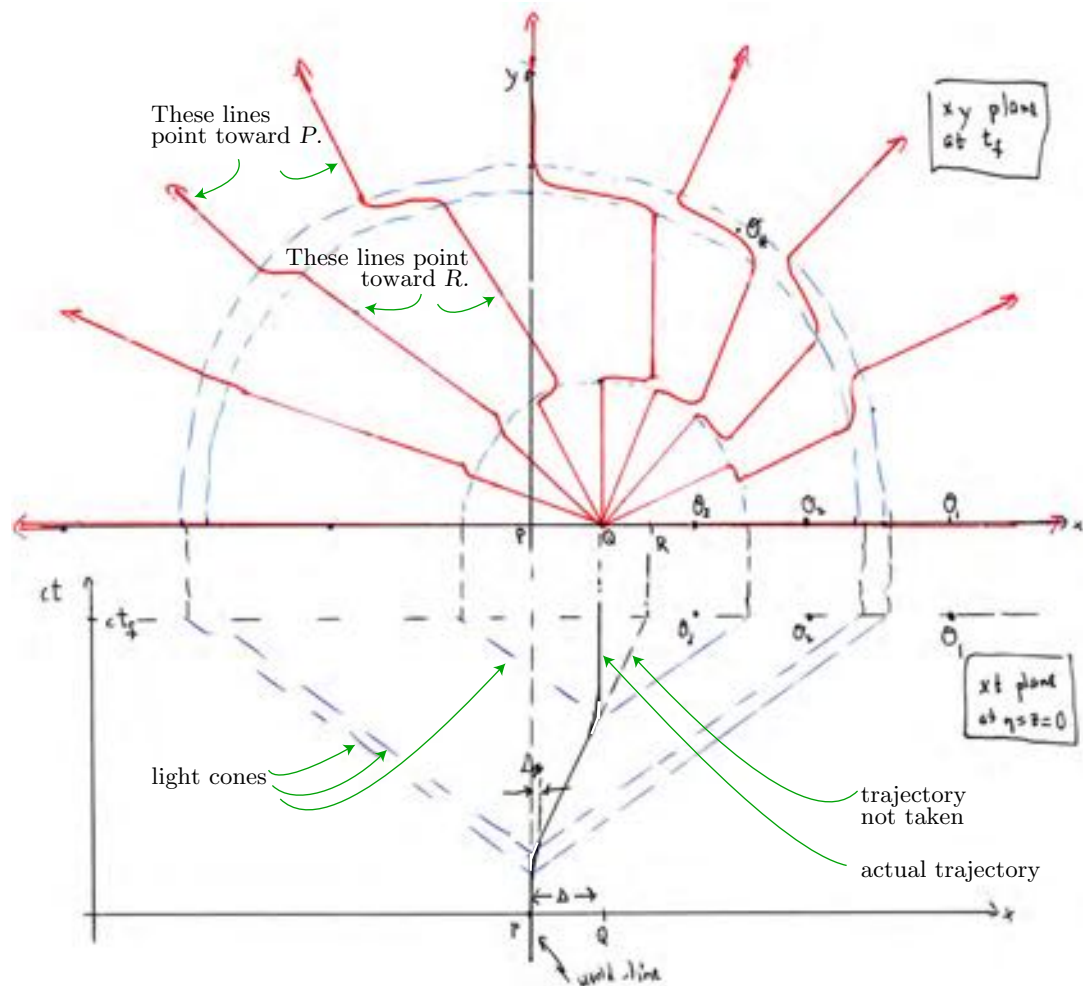


Figure 41.1: *Bottom:* A charged particle trajectory in the xt plane. *Top:* Snapshot of the corresponding field lines at one time, in the upper xy half-plane. The full 3D picture would be a figure of revolution about the \hat{x} axis. The bunching of field lines in the intermediate region, expected when the intermediate velocity is close to c , is not shown.

We now connect up the three regions whose fields we just described. We know that $\vec{\nabla} \cdot \vec{E} = 0$, so *field lines cannot terminate* anywhere except on the charges. Thus, in the two joining regions the field lines must look as they are drawn in the figure:

- An observer at \mathcal{O}_* , for example, sees a pulse of \vec{E} directed *transversely* to her line of sight to P (and $\vec{B} = \text{out of page}$). These kinks lie on a spherical shell whose radius expands outward in time at speed c .
- There is an opposite kink associated to the deceleration, on another spherical shell that is also expanding outward.
- The kinks are most pronounced at 90° to the direction of acceleration (on the $\pm y$ axis); there is no kink along the direction of acceleration (on the $\pm x$ axis). Specifically, the kink is directed along $\hat{r} \times (\hat{r} \times \vec{a})$ where \vec{a} is the acceleration.

Next let's ask about the strength of the transverse fields, for example in the first kink region (corresponding to the initial acceleration). We learned in Chapter 35 that $\|\vec{E}\|$ is proportional to the transverse density of the field lines, which in turn is the total length of all the lines in a volume, divided by that volume. And the *stretching* needed to accommodate the kink without breaking any line crowds more length into the thin shell than there would otherwise be (without any acceleration)!

Consider the situation at \mathcal{O}_* , a particular angle θ from the \hat{x} axis. The charge accelerates from velocity 0 to v , so the kink joins a line pointing toward P to one pointing toward R , a distance $\Delta = vt_f$ to the right in the \hat{x} direction (Figure 41.1), or $\Delta \sin \theta$ in the direction transverse to the field line. The acceleration occurs over a time interval v/a , so the thickness of the shell between the dashed lines is $\Delta_* = cv/a$.

Imagine drawing a total of N lines emerging from the charge. We wish to find the total length of all the field lines passing through a shell of thickness $c\Delta_*$ and cross-sectional area $d\Sigma$. A total of $Nd\Sigma/(4\pi r^2)$ lines enter, bend sideways, travel a distance $vt_f \sin \theta$, bend again, and emerge. Thus

$$\frac{\text{total length of lines}}{\text{volume}} = \frac{Nd\Sigma vt_f \sin \theta / (4\pi r^2)}{d\Sigma \Delta_*}.$$

The radius of the sphere is ct_f , so we find that $\|\vec{E}\|$ is proportional to the acceleration, to $\sin \theta$, and to $1/r$. These are the key features of radiation from an accelerated charge:

- The electric field is transverse to the line of sight from observer to source.
- The electric field is mainly in the equatorial plane $\theta = \pi/2$.
- The electric field falls with distance as r^{-1} , not r^{-2} .
- The electric field is proportional to the magnitude of the acceleration.

41.3 MAGNETIC FIELDS

The magnetic Gauss law does not give \vec{B} any sources or sinks. Hence, \vec{B} field lines must all be closed loops. Why then should they exist at all? The answer comes from the Faraday law.

We consider the same trajectory as before, but focus on only the final deceleration. The top panel of Figure 41.3 shows the electric field lines at a time t_1 . We argued a fixed point S in the xy plane, an observer will initially see a small, radial electric field ($\propto r^{-2}$), then around t_1 a pulse of \vec{E}_x ($\propto r^{-1}$), and then back to small field at later time t_2 .

An observer outside the expanding shell has not yet learned about the deceleration, so she sees the magnetic field of a particle in uniform, straight-line motion, which falls as r^{-2} . An observer inside the shell sees the magnetic field of a charge at rest, which is zero. But at the leading (outer) boundary of the shell (point T), the observer sees a rising \vec{E}_x . The figure shows a small rectangular surface area surrounding T , coming out of the page in the yz plane. Two of the four edges of this rectangle straddle the boundary of the shell. Integrating Faraday's law over this surface element shows that either \vec{B}_y or \vec{B}_z must be nonzero there.

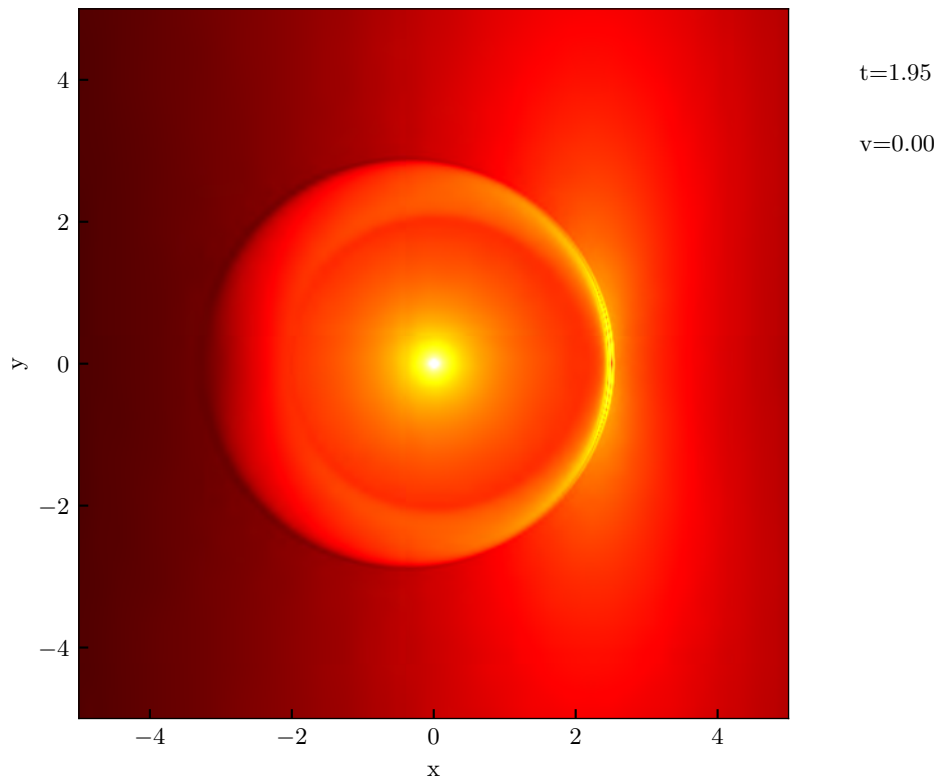


Figure 41.2:

The field lines must form a figure of revolution about the x axis, by axial symmetry, and they must also be closed curves by the Gauss law. A radial (y) component of \vec{B} would require field lines to extend to infinity, and hence not close. Also, $\vec{B} = 0$ on the outermost edge of the rectangle. But an azimuthal (z) component is allowed on the inner edge of the rectangle. So at T we have \vec{B} pointing into the page, with field lines forming rings in planes parallel to the xz plane.

A similar argument applies at point U on the trailing boundary of the shell. Here \vec{E}_x is *falling* over time, but only on the *inner* edge of the rectangle. So again we find \vec{B} pointing into the page.

Throughout the shell we have $\|\vec{B}\| \propto r^{-1}$ because \vec{E} has that behavior. Thus, we find that

$$\vec{E} \times \vec{B} \text{ is directed radially outward and falls as } r^{-2}.$$

FURTHER READING

Freeman et al., 2019, §3.3

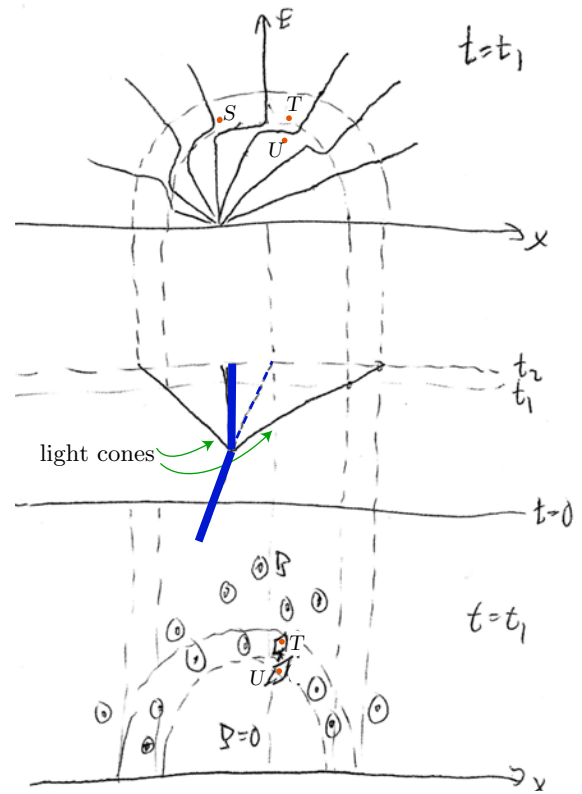


Figure 41.3: *Middle:* The decelerating part of the trajectory in Figure 41.1 (*heavy blue line*). *Bottom:* Determination of magnetic fields. The small rectangular paths surrounding points T and U are parallel to the yz plane, that is, they extend out of the page.

PROBLEMS

41.1 *Relativistic Bremsstrahlung*

Consider a charge that is motionless for a long time, then gets rapidly accelerated to uniform straight-line motion at speed $V \ll c$, then gets rapidly decelerated back to rest. The lower panel of Figure 41.1 shows the worldline of the charge in the xt plane. The upper panel depicts a snapshot of the electric field lines at a time after the particle has returned to rest, in the xy plane.

Now you sketch two similar pictures for the case where V is *not* much smaller than c . Your pictures will be similar to mine; discuss the differences physically.

41.2 *Bremsstrahlung graphic*

If you haven't done Problem 33.1 yet, do it first as a warmup.

A positively charged particle is initially in uniform motion along the x axis at speed $0.9c$. At time zero, it abruptly comes to a halt. An observer later maps out the electric field at time $t_0 > 0$ in the xy plane.

Close to the particle, the observer sees the usual $1/r^2$ field. Section 32.3.2 argued that far from the particle, the observer sees a field that is crowded into the yz plane and centered on the point where the particle would have been located, if it had continued to move. Section 41.2 argued further that on the boundary between these regions, there is a pulse of radiation (bremsstrahlung). Verify these claims numerically, as follows.

- Express all lengths as dimensionless quantities times ct_0 . Find the region in the xy plane where the observer will see the 4-vector potential of a charge at rest.
- Make a grid of points at which to evaluate the 4-vector potential. The grid should be fine enough to get reasonably accurate estimates of derivatives by numerical differentiation.²
- Evaluate the 4-vector potential at each of the grid points satisfying the condition in (a).
- Use ideas from Problem 33.1 to evaluate the 4-vector potential at every grid point *not* satisfying that condition.
- Repeat (a–d) for later time $(1.001)t_0$. Subtract from your previous answer and divide by 0.001 to estimate the time derivative of \vec{A} throughout the xy plane.
- Do whatever else you need to do to find the electric field at time t_0 .
- Make a graphical depiction of the magnitude $\|\vec{E}(t_0, x, y, 0)\|$. If the range of values attained is too large to display properly, compress it by taking a logarithm before making the plot.

[Remarks: Luckily, \vec{E} lies in the xy plane, so a two-dimensional plot is adequate.

Unluckily, the 4-vector potential field is discontinuous, so you won't get an accurate result by numerically differentiating it. However, you do get the right qualitative behavior. This problem is a pathology related to the unrealistic assumption that the charge stops instantly (that is, infinite deceleration).

Make your plot cover a range of xy values large enough (and also small enough) to show the interesting features. Make sure your computer uses the same scale for the x and y axes.

If you wish, you can compare your result to the more complicated formulas in Your Turn 32D, but that's not the approach you are to use in this problem.

If you use Python, the *Student's Guide* §8.2 discusses heatmaps. Or you may prefer a contour plot or surface plot. Use your judgement about what is clearest. Why didn't I need to tell you the values of t_0 and q ?

- Repeat for speed $0.1c$ and comment.

²Python users will find useful information in the Kinder & Nelson, 2018, §6.4.1, or in the builtin help for `np.meshgrid`.

CHAPTER 42

Electric Dipole Radiation

42.1 FRAMING

This chapter will show in a special situation that (a) charges will emit electromagnetic radiation when accelerated, (b) the radiation is polarized transversely to the line of sight, and (c) its energy flux falls with distance like $1/r$. The special situation, which is frequently realized in practice, is a limit in which the source size is much smaller than the outgoing wavelength. Unlike Chapter 23, this time, we make no restriction that charge density is everywhere zero.

42.2 FAR FIELD APPROXIMATION

Suppose that some charges executing prescribed motions are confined to a region of size a centered on the origin of coordinates. So their locations all obey $\|\vec{r}_*\| < a$. We observe fields at \vec{r} with $\|\vec{r}\| \gg a$. We'll keep only the leading terms in the power series in a/r . That's called the **far-field approximation**.

Familiar steps: Let $\vec{R} = \vec{r} - \vec{r}_*$. Please review why

$$R = r - \hat{r} \cdot \vec{r}_* + \dots \quad (42.1)$$

$$R^{-1} = r^{-1} \left(1 + \frac{\hat{r} \cdot \vec{r}_*}{r} + \dots \right). \quad (42.2)$$

In each case, we have kept the first two orders of a power series in a/r ; the ellipses denote terms of higher order.

Our general, Green-function solution gives the vector potential in Lorenz gauge as¹

$$\underline{A}^\mu(t, \vec{r}) = \frac{\mu_0}{4\pi} \int d^3r_* \frac{1}{R} \underline{J}^\mu(t - R/c, \vec{r}_*). \quad [40.5, \text{page 425}]$$

We need to be careful with our approximation. In the $1/R$ factor, the second and higher terms in Equation 42.2 can be dropped—they make subleading contributions to \underline{A} . But in the argument to \underline{J} , we must *keep* the subleading term of Equation 42.1 because, although it is smaller than the leading term,

- It tends to a constant, not zero, as $r \rightarrow \infty$, and
- When we take \underline{J} to vary harmonically in the next section, this additive constant will turn into a *multiplicative* constant that cannot be dropped.

¹We used relativity to obtain this formula. However, in these notes we will consider a nonrelativistic problem (charges moving much slower than light), so there is no reason to insist on writing only manifestly Lorentz-invariant formulas.

Moreover, we'll see that the apparently leading term will not give rise to any radiation. Thus, dropping the subleading term just mentioned would fool us into thinking radiation is not possible at all!

Still-higher terms really may be dropped in far-field approximation. Thus,

$$\underline{A}^\mu(t, \vec{r}) = \frac{\mu_0}{4\pi r} \int d^3 r_* \underline{J}^\mu(t - r/c + \hat{r} \cdot \vec{r}_*/c, \vec{r}_*). \quad \text{far field} \quad (42.3)$$

Equation 42.3 is the desired generalization of Equation 23.8 (page 263) to situations where the net charge density is not everywhere zero.

42.3 HARMONIC TIME VARIATION

Let's suppose that the source charges and currents \underline{J} vary harmonically in time with some angular frequency ω . That is, assume²

$$\underline{J}^\mu(t, \vec{r}_*) = \frac{1}{2} e^{-i\omega t} \bar{\underline{J}}^\mu(\vec{r}_*) + \text{c.c.},$$

where $\bar{\underline{J}}^\mu$ are four complex functions of position \vec{r}_* only. Then

$$\underline{A}^\mu(t, \vec{r}) = \frac{1}{2} \frac{\mu_0}{4\pi r} e^{-i\omega(t-r/c)} \int d^3 r_* e^{-i\omega \hat{r} \cdot \vec{r}_*/c} \bar{\underline{J}}^\mu(\vec{r}_*) + \text{c.c.} \quad (42.4)$$

Everything inside the integral is independent of the observer's distance r . However, the observer's *direction* \hat{r} is still present inside the integral.

42.4 MULTIPOLE APPROXIMATION

Equation 42.4 is still a bit complicated, but fortunately another approximation is often justified: Often the quantity c/ω is much bigger than the source size³ a . That is, the dimensionless quantity

$$\epsilon_{\text{multi}} = \omega a / c \quad \text{multipole parameter} \quad (42.5)$$

is much smaller than 1.

In that case, we may replace the exponential inside the integral by its Taylor series: $1 - i\epsilon_{\text{multi}}(\hat{r} \cdot \vec{r}_*/a) + \dots$. Making this approximation, and truncating after a finite number of terms, is called **multipole approximation**. Keeping only the *first* term (that is, 1) is called **electric dipole approximation**, for reasons that will be clear soon.

That is, the far-field, multipole approximation is a *double power series expansion* in both a/r and ϵ_{multi} .

²If that's not the case, we can nevertheless decompose \underline{J} into Fourier components, use the analysis below on each one, and ultimately add all their contributions.

³If the charges are oscillating or doing circular motion, this condition says that their speed $\approx a\omega$ is much smaller than c . This is certainly true of electrons in a radiating atom or molecule.

42.5 LEADING ORDER: ELECTRIC DIPOLE RADIATION

42.5.1 A time-varying ED moment leads to $1/r$ potentials

Equation 42.4 has become

$$\underline{A}^\mu(t, \vec{r}) = \frac{1}{2} \frac{\mu_0}{4\pi r} e^{-i\omega(t-r/c)} \underbrace{\int d^3r_* \bar{J}^\mu(\vec{r}_*)}_{\text{retarded time}} + \text{c.c.} = \frac{\mu_0}{4\pi r} \int d^3r_* \underline{J}^\mu(t_r, \vec{r}_*). \quad (42.6)$$

In this expression, t_r is shorthand for the **retarded time** $t - r/c$, which does not depend on \vec{r}_* .

We can now get an even simpler formula for the spatial components of \underline{A} .⁴ First, the divergence theorem implies

$$\int d^3r_* \vec{\nabla}_i(\vec{r}_m \vec{j}_i)|_{\vec{r}_*} = 0$$

for each of $m = 1, 2, 3$. (Remember that $\vec{j} \rightarrow 0$ outside the finite region where the source is located.) So

$$\int d^3r_* \delta_{im} \vec{j}_i(\vec{r}_*) = - \int d^3r_* \vec{r}_{*m} \vec{\nabla} \cdot \vec{j}|_{\vec{r}_*} = + \int d^3r_* \vec{r}_{*m} \frac{\partial}{\partial t} \rho_q(\vec{r}_*) = \frac{d}{dt} \vec{\mathcal{D}}_{E,m}. \quad (42.7)$$

The final step made use of the definition of electric dipole moment $\vec{\mathcal{D}}_E$.

Equation 42.7 says that $\int d^3r_* \vec{j}_m = \frac{d}{dt} \vec{\mathcal{D}}_{E,m}$. So the three spatial components of Equation 42.6 reduce to

$$\vec{A}_m^{[ED]}(t, \vec{r}) = \frac{\mu_0}{4\pi r} \left. \frac{d\vec{\mathcal{D}}_{E,m}}{dt} \right|_{\text{ret}}. \quad \text{ED approximation, far field} \quad (42.8)$$

The notation “ret” indicates that the derivative is to be evaluated at time $t_r = t - r/c$, the retarded time.⁵

Your Turn 42A

Evaluate \underline{A}^0 using Equation 42.6 and check that it agrees with Your Turn 37Aa (page 404).

42.5.2 Pure dipole limit

Chapter 37 pulled a spherical wave solution out of a hat and then showed it was an exact solution.⁶ Here, we obtained it as an *approximate* solution to a real physical problem. We can recover the exact solution by considering the pure-dipole limit, in which $a \rightarrow 0$ holding fixed the amplitude $\vec{\mathcal{D}}_E$. In this limit, the ED approximation really does become exact.

⁴The following derivation should be familiar from Chapter 3.

⁵Chapter 40 used “ret” to mean evaluation at $t - R/c$, where $r = \|\vec{r} - \vec{r}_*\|$, but formulas like Equation 42.8 no longer involve R so there is no ambiguity in the change of notation.

⁶Or see P+S, §14.5.4.

42.6 RADIATION

42.6.1 Electric and magnetic fields

We now need a physical interpretation of our answer, Equation 42.8. One good step would be to find the physical fields \vec{E} and \vec{B} . This calculation, too, is greatly simplified in far-field approximation.⁷ The point is that, when taking derivatives, we never need to differentiate the $1/r$ factor, because that will give $1/r^2$, which we will see is not leading order. Recall that \vec{A}_i denotes the spatial components of \underline{A} .

$$\begin{aligned}\vec{B}_k &= \varepsilon_{kmi} \vec{\nabla}_m \vec{A}_i = \varepsilon_{kmi} \frac{\mu_0}{4\pi} \frac{\partial}{\partial \vec{r}_m} \left(\frac{1}{r} \frac{d\vec{\mathcal{D}}_{E,i}}{dt} \Big|_{t-r/c} \right) \\ &= \varepsilon_{kmi} \frac{\mu_0}{4\pi r} \frac{d^2 \vec{\mathcal{D}}_{E,i}}{dt^2} \Big|_{\text{ret}} (-\hat{r}_m/c) + \text{subleading}.\end{aligned}$$

(To get the second line, use the chain rule.) More compactly,

$$\vec{B}^{[ED]} = -\frac{\mu_0}{4\pi r c} \hat{r} \times \frac{d^2 \vec{\mathcal{D}}_E}{dt^2} \Big|_{\text{ret}} \quad \text{far-field} \quad (42.9)$$

We can see that

- Indeed, the only aspect of the source that matters in this approximation is its time-varying electric dipole moment, which explains our name “electric dipole approximation.”
- Specifically, the \vec{B} field is proportional to the *acceleration* of the charge.
- The far field wavecrests are spherical and move radially outward at speed c , because \vec{B} depends on observer’s distance and time only through the combination $r - ct$.
- The far field is everywhere transverse (\vec{B} points perpendicular to its direction of propagation \hat{r}).
- The far field falls off with distance like r^{-1} .

We could now obtain \vec{E} by returning to Equation 42.6, this time working out \underline{A}^0 , and using the formula for \vec{E} in terms of the vector and scalar potential. But there’s an easier way. Recall that Ampère’s law says $\vec{\dot{E}} = c^2 \vec{\nabla} \times \vec{B}$, and we just found \vec{B} . Again use the fact that derivatives of r^{-1} will be subleading and may be dropped in far-field approximation. Furthermore, derivatives of \hat{r} fall with distance like r^{-1} , and hence will also generate subleading terms. The *leading* contribution to \vec{E} comes once again from the retardation factor: $\vec{\nabla}(t - r/c) = -\hat{r}/c$. So

$$\frac{d\vec{E}}{dt} = c^2 \left(-\frac{\mu_0}{4\pi r c} \right) \left(-\frac{\hat{r}}{c} \right) \times \left(\hat{r} \times \frac{d^3 \vec{\mathcal{D}}_E}{dt^3} \Big|_{\text{ret}} \right).$$

Because everything is harmonic in time, we can just drop one time derivative from

⁷The following derivation is essentially a solution to Your Turn 37B.

both sides of this equation:

$$\vec{E}^{[ED]} = \frac{\mu_0}{4\pi r} \hat{r} \times \left(\hat{r} \times \frac{d^2}{dt^2} \vec{D}_E|_{\text{ret}} \right). \quad \text{far-field}$$

Like \vec{B} , the electric field is transverse to the line of sight \hat{r} , falls like r^{-1} , and involves acceleration of the charge. Moreover, \vec{E} is also perpendicular to \vec{B} , a property that we observed some time ago for plane waves. What's new is that now we know the quantitative relations between the charge's motion and the amplitude and polarization of the wave.

42.7 CONCRETE EXAMPLES

42.7.1 Electric dipole antenna

Usually when we introduce “wires,” we implicitly assume an approximation in which no charge builds up anywhere. That is, usually we ignore the *capacitance* of a system of “wires”; for example, if the wires do not form a closed circuit, we assume that no current flows.

However, if we attach an alternating potential source to two diverging wires, then some current will flow into and out of them, particularly at high frequency. That current alternately builds up charge along the wires, which in turn creates an oscillating electric dipole moment, which we now know can radiate.

The exact theory of such an “electric dipole antenna” is complicated and involves self-consistently solving for the fields, currents, and charges. Instead of doing this, we now *assume* a simple form for the currents and charges that is at least consistent with the continuity equation. Suppose that one wire segment stretches from the origin along the z axis to $z = a/2$. Another wire segment stretches the other direction to $z = -a/2$. Alternating current is fed into the top wire at the origin; we will suppose that its amplitude falls linearly to zero at the end of the wire. An equal and opposite current is fed into the lower wire at the origin, so that overall the antenna is always net neutral. Moreover, because the wires run in opposite directions, their respective currents are always *parallel*.

In a formula, the current in each wire is

$$I(t, z) = \bar{I} \cos(\omega t) (1 - |z|/(2a)) \quad \text{for } |z| < a/2.$$

Current is 1D charge flux, so the 1D continuity equation says

$$\frac{d\rho_q^{[1D]}}{dt} = -\frac{dI}{dz} = -(\bar{I} \cos \omega t)(\pm 2/a)$$

for the upper and lower wires respectively. Thus $\rho_q^{[1D]} = \pm \frac{2\bar{I}}{a\omega} \sin \omega t$.

We can now find the dipole moment:

$$\vec{D}_E = \hat{z} \sin \omega t \left[\int_{-a/2}^0 z dz \frac{-2\bar{I}}{a\omega} + \int_0^{a/2} z dz \frac{2\bar{I}}{a\omega} \right] = \hat{z} \frac{\bar{I}a}{8\omega} \sin \omega t.$$

Substituting into the general dipole radiation formulas then gives the radiation created by this antenna. A distant observer in the xy plane will see radiation linearly polarized along \hat{z} . A distant observer along the z axis will see nothing. A distant observer along any other direction will see radiation linearly polarized along the direction obtained by projecting \hat{z} to the plane perpendicular to the line of sight.

42.7.2 Greenhouse gases

Absorption and emission by single molecules should properly be treated quantum mechanically (Chapter 55); however, some qualitative features can be understood in our classical picture.

Earth's surface is kept considerably warmer than would otherwise be the case by its atmosphere. Our atmosphere is largely transparent to visible light from the Sun, yet it intercepts infrared radiation and impedes its escape to space. Different gas molecules have very different abilities to absorb and reemit infrared photons, however.

Optical absorption by a molecule involves its distribution of charge and current. Similarly to what we have seen in this chapter, the most important term is controlled by the “transition dipole,” which is the matrix element of the electric dipole moment operator between the ground and excited molecular states.

The molecules O_2 and N_2 , which constitute the bulk of Earth's atmosphere, are called **homonuclear**, because they contain two identical nuclei. A homonuclear diatomic molecule is symmetric under inversion, even when strained away from its normal chemical bond length, and hence can have no dipole moment. Thus, the transition dipole between the ground state and either a rotational or vibrational excited state must equal zero. Such excited states are typically separated from the ground state by an energy gap in the infrared region, but a homonuclear molecule cannot use photons to enter or leave those states, and hence is a poor absorber of infrared light.

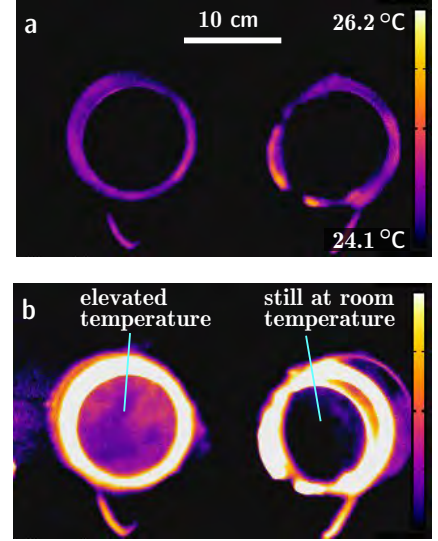
Non-homonuclear diatomic molecules, notably nitric oxide (NO), have nonzero dipole moment in their ground state, which changes when the molecule is set into rotational motion. Moreover, the vibrational modes of such a molecule change its dipole moment. The transition dipoles between the ground state and the rotational and vibrational excited states are therefore non zero, making NO a strong absorber in the infrared. It is therefore referred to as an infrared-active (“greenhouse”) gas.

A bent triatomic molecule, such as water (H_2O), also has a permanent dipole moment; water vapor is also a potent infrared-active gas. The carbon dioxide molecule has three nuclei in a linear arrangement, and hence zero dipole moment in its ground state. Thus, its transition dipoles between ground and rotationally excited states vanish. However, it develops an oscillating dipole moment in some of its vibrational states; transition dipoles therefore exist for these and also for mixed rotation-vibration states, making CO_2 another infrared-active gas (Figure 42.1).

42.8 ENERGY FLUX

In this section, we'll omit the suffix “ret” for brevity and continue to work in the far field, in electric dipole approximation.

Figure 42.1: [Infrared photographs.] **Energy absorption by an IR-active gas.** Two identical, cylindrical chambers with transparent ends, viewed in the wavelength band 7.5–14 μm . False color indicates radiance in this band (reds are higher than blues); the scale bar is labeled with approximate inferred temperature values in degrees Celsius. The chamber on the right contains dry air. The one on the left contains carbon dioxide. Both have axial length 23 cm and are viewed end-on. (a) Both chambers started at room temperature. False color when looking into each chamber matches the backdrop. (b) The chambers were briefly exposed to infrared light. After irradiation was stopped, the one containing IR-active gas was observed to be slightly warmer for about one minute. That is, more infrared light was observed coming out of this chamber than was the case for either the backdrop or the other chamber. [See also youtu.be/0eI9zzZoiP4 and Sieg et al., 2019.]



Now at last we can see how energy is transported: Its flux is

$$\vec{S}^{[ED]} = \frac{1}{\mu_0} \vec{E} \times \vec{B} = -\mu_0^{-1} \left(\frac{\mu_0}{4\pi r} \right)^2 \frac{1}{c} \underbrace{\left[\hat{r} \times \left(\hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) \right]}_{\hat{r} \cdot \left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) - \frac{d^2}{dt^2} \vec{\mathcal{D}}_E} \times \left[\hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right].$$

The factor in the brace is $\hat{r} \cdot \left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) - \frac{d^2}{dt^2} \vec{\mathcal{D}}_E$. Now use the triple cross product formula again:

$$\vec{S} = -\mu_0^{-1} \left(\frac{\mu_0}{4\pi r} \right)^2 \frac{1}{c} \left(\hat{r} \left[\left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) - \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right] \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) - \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \left[\left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right) - \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \cdot \hat{r} \right]$$

$$\vec{S}^{[ED]} = \hat{r} \frac{\mu_0}{(4\pi r)^2} \frac{1}{c} \left(\left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \Big|_{\text{ret}} \right\|^2 - \left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \Big|_{\text{ret}} \right)^2 \right). \quad \text{far-field} \quad (42.10)$$

Thus, the energy flux vector always points radially outward. It's not spherically symmetric, however, because its magnitude depends on the direction \hat{r} to the observer.

The *total* power output is the rate at which energy passes through a large spherical shell:⁸

$$\mathcal{P}^{[ED]} = \lim_{B \rightarrow \infty} \int_{r=B} d^2 \vec{\Sigma} \cdot \vec{S}^{[ED]} = \frac{\mu_0}{(4\pi)^2 c} \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \Big|_{\text{ret}} \cdot \left[\int d^2 \hat{r} (\vec{\mathbf{1}} - \hat{r} \hat{r}) \right] \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \Big|_{\text{ret}}$$

The first term inside the square brackets is the integral over all directions of a constant tensor, that is, $4\pi \vec{\mathbf{1}}$. The second term is 4π times the average over all directions of $\hat{r} \hat{r}$. It has no dependence on the observer's position. Thus it must be a rotationally-invariant, yet constant, 3-tensor of rank 2. There is only one possibility: This term

⁸Because we only want energy that makes it all the way out to infinity, the far-field approximation is automatically satisfied.

must be a constant times the unit tensor. Moreover, its trace must be $\int d^2\hat{r} = 4\pi$, which fixes the constant to be $1/3$. All together, then, the factor in square brackets is $4\pi(1 - \frac{1}{3})\hat{\mathbb{1}}$, and we have

$$\mathcal{P}^{[ED]} = \frac{\mu_0}{4\pi c} \frac{2}{3} \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_E|_{\text{ret}} \right\|^2 \quad \text{total power output, ED approximation} \quad (42.11)$$

[T2] Section 42.8' (page 446) justifies a claim made in this section.

42.9 LINEAR POLARIZATION

Consider the case in which $\vec{\mathcal{D}}_E$ is always directed along a single direction (linear polarization). We can choose coordinates to make that direction be the z -axis: $\vec{\mathcal{D}}_E = \mathcal{D}_E(t)\hat{z}$. First note a relation between the spherical directions:

$$\hat{z} = \hat{r} \cos \theta - \hat{\theta} \sin \theta.$$

Your Turn 42B

Show that

$$\begin{aligned} \vec{B}^{[ED]} &= \hat{\varphi} \frac{\mu_0}{4\pi r c} \frac{d^2}{dt^2} \mathcal{D}_E|_{\text{ret}} \sin \theta \\ \vec{E}^{[ED]} &= \hat{\theta} \frac{\mu_0}{4\pi r} \frac{d^2}{dt^2} \mathcal{D}_E|_{\text{ret}} \sin \theta. \end{aligned}$$

In any direction, we see a linearly polarized plane wave.

Turning now to the energy flux,

$$\begin{aligned} \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right\|^2 &= \left[\frac{d^2}{dt^2} \mathcal{D}_E \right]^2 \\ \left(\hat{r} \cdot \frac{d^2}{dt^2} \vec{\mathcal{D}}_E \right)^2 &= \left(\frac{d^2}{dt^2} \mathcal{D}_E \hat{r} \cdot \hat{z} \right)^2 = \left[\frac{d^2}{dt^2} \mathcal{D}_E \right]^2 \cos^2 \theta \\ \vec{S}^{[ED]} &= \mu_0^{-1} \vec{E} \times \vec{B} = \hat{r} \frac{\mu_0}{(4\pi r)^2} \frac{1}{c} \left[\frac{d^2}{dt^2} \mathcal{D}_E|_{\text{ret}} \right]^2 \sin^2 \theta. \end{aligned} \quad (42.12)$$

Equation 42.12 shows the angular dependence explicitly: Energy mostly comes out near the equatorial plane (here the xy plane). We can now get the total power output from Equation 42.11.

If the dipole varies harmonically in time, then we can write $\mathcal{D}_E(t)$ in terms of the amplitude (maximum value) $\bar{\mathcal{D}}_E$ as $\mathcal{D}_E(t) = \frac{1}{2} e^{-i\omega t} \bar{\mathcal{D}}_E + \text{c.c.}$ Then the time-averaged power output is

$$\langle \mathcal{P}^{[ED]} \rangle = \frac{\mu_0}{12\pi c} \omega^4 |\bar{\mathcal{D}}_E|^2, \quad \begin{array}{l} \text{total power output,} \\ \text{harmonic source} \end{array} \quad (42.13)$$

a famous result.

Your Turn 42C

Repeat the exercise, but with $\vec{\mathcal{D}}_{\text{E}}(t) = \mathcal{D}_{\text{E}0} \begin{pmatrix} \cos \omega t \\ \sin \omega t \\ 0 \end{pmatrix}$ and interpret the result.

FURTHER READING

Greenhouse gases:

Bohren & Clothiaux, 2006, Chapt. 2.

T₂ Track 2

42.8' Only one rank-2 tensor from Heaven

The main text claimed that taking the average of a polarizability tensor over all orientations must yield a constant times the unit tensor. Certainly that average must be a rotationally-invariant, symmetric, rank-2 tensor. Certainly the unit tensor is one possibility (Chapter 13). Can we be sure there are no others?

Consider the effect of *infinitesimal* rotations on a symmetric, rank-2, 3-tensor $\overset{\leftrightarrow}{T}$. Choose a cartesian coordinate system and let the matrix T represent the tensor. Rotational invariance is the statement that

$$(\mathbf{1} - \epsilon\Omega)^t \mathsf{T} (\mathbf{1} + \epsilon\Omega) = \mathsf{T} + \mathcal{O}(\epsilon^2)$$

for any antisymmetric “generator” matrix Ω .⁹ That is, the matrix T must commute with any such Ω .

For example, consider the generator $\overset{\leftrightarrow}{\Omega}_{ij} = \delta_{i2}\delta_{j3} - \delta_{i3}\delta_{j2}$, which generates rotations about the \hat{x} axis. Then $\overset{\leftrightarrow}{T}$ must obey

$$0 = \overset{\leftrightarrow}{T}_{m2}\delta_{j3} - \overset{\leftrightarrow}{T}_{m3}\delta_{j2} - \overset{\leftrightarrow}{T}_{3j}\delta_{m2} + \overset{\leftrightarrow}{T}_{2j}\delta_{m3}$$

for any m and j . Specialize further to $m = 1$; this gives $\overset{\leftrightarrow}{T}_{12} = 0$ and $\overset{\leftrightarrow}{T}_{13} = 0$. If we instead specialize to $m = 2$, we find $\overset{\leftrightarrow}{T}_{22} = \overset{\leftrightarrow}{T}_{33}$.

Continuing in this way we find that all off-diagonal entries of $\overset{\leftrightarrow}{T}$ are zero, whereas all diagonal entries are equal. That is, $\overset{\leftrightarrow}{T}$ is a constant times the identity tensor.

⁹See Problem 17.5 (page 215) and Problem 12.1 (page 148).

PROBLEMS

42.1 Beyond far-field approximation

Background: In class I derived expressions for the exact electric and magnetic fields outside an arbitrary charge/current distribution. Then I simplified the result by assuming (i) harmonic time dependence of the sources, (ii) observer is far away, so we may discard $\mathcal{O}(r^{-2})$ terms, (iii) long wavelength (low frequency, nonrelativistic source motion), so we may discard all but the leading term in an expansion in powers of a/λ , where a is the source size.

I did however *sketch* the complete fields for an oscillating electric dipole, both near- and far-field. Close to the source, at each instant of time the electric field looked like the field around a *static* dipole. Thus each field line starts on a net + charge and terminates on a net - charge. Farther from the source, however, the field lines detach from the source and move outward on closed, banana-shaped paths (they don't terminate anywhere).

How and where does this detachment occur? Let's see. To keep things interesting, however, we'll do a slightly different problem.

Problem: Consider a circular loop of wire of radius a in the xy plane, carrying a prescribed, harmonically oscillating current $I(t) = \frac{1}{2}[\bar{I}e^{-i\omega t} + \text{c.c.}]$. Thus \bar{I} is one half of the peak-to-peak current amplitude. You are to find and plot the magnetic field $\vec{B}(t, \vec{r})$ everywhere, at various times. This "simply" involves evaluating numerically a formula we obtained in class. Thus we keep assumption (i) above but drop (ii-iii).

Because of the azimuthal symmetry, it's enough to examine $\vec{B}(t, \vec{r})$ only for \vec{r} in the xz plane, and indeed to look only at $x > 0$.

Because this time we are examining \vec{B} (not \vec{E}), none of the field lines will terminate (all are closed loops). Nevertheless we'll find a distinction between those lines attached to the source and those that have detached.

Steps:

Measure all lengths in units of a . (Or equivalently, measure lengths in meters and take $a = 1$ m.)

- Write the general expression to be evaluated, specialized for the situation in the problem. This expression involves an integral, which you'll eventually have to do numerically, but not yet. Using this unevaluated expression, show that one of the three cartesian components of \vec{B} equals zero everywhere on the xz plane.

That's convenient: it means that every integral curve of \vec{B} that starts in the xz plane will remain completely in that plane. These curves are Faraday's magnetic "field lines."

- Set $\omega = 0$ and evaluate \vec{B} on a grid of points out to a distance of, say, $r = 5a$ from the origin. Display your answer as a tiny-arrows plot. It may take some futzing to make your plot look nice (i.e., physically informative).

I found it hard to visualize the answer because the arrows were of such differing lengths. So I found it better to display instead the direction $\hat{B} = \vec{B}/|\vec{B}|$. This normalized vector field has the same integral curves as \vec{B} . (Also, of course overall factors like $\mu_0\bar{I}/(4\pi r)$ drop out when we normalize.)

- Draw by hand on your plot some of the integral curves obtained by following the arrows. Some or all of your integral curves will have the property that they *link* (pass through) the current loop: We say they are “attached to the source.” Find which ones have this property and comment.
- Repeat taking $\omega = 2\pi c/(3a)$ and time $t = 0$. Thus we expect the far fields to be waves with wavelength $3a$. Comment on the behavior you observe both close to and far from the origin; on the z axis versus on the equatorial plane; etc. If some field lines are not attached to the source, find the locus separating the attached lines from the detached ones.
- Also compute the longitudinal part of \hat{B} , that is, $\hat{r} \cdot \vec{B}$ at time zero, and plot in the xz plane. If there’s something surprising about your answer, explain it; otherwise explain why it’s not so surprising. [*Remark:* To show a function of two variables, you may want to use a contour plot.]
- Try some other values of t , e.g. $t\omega = \pi/4$. Can you find a time where an attached field line is just about to detach?
- Show some initiative. Suppose these are figures in a paper you’re trying to publish—figure out some improvements in presentation, informative labels, etc. If you think that the range $0 < r < 5a$ doesn’t show the physics optimally, choose some better range. Play.

42.2 [Not ready yet.]

CHAPTER 43

Higher-Multipole Radiation

43.1 FRAMING

Chapter 42 found a solution to Maxwell's equations that, in the far-field region, becomes approximately a spherical wave potential with amplitude proportional to the time derivative of the electric dipole moment (compare Equations 37.1 and 42.8). Does that mean that a charge and current distribution with electric dipole moment equal to zero (or a constant) cannot radiate? No, we already found in Chapter 23 that a purely magnetic dipole also creates far fields that fall like r^{-1} , indeed as a different sort of spherical wave.

To see what's going on, recall a second approximation made in Section 42.4: The electric dipole approximation retained only the first term in the multipole approximation. If that term vanishes, then the leading behavior may nevertheless involve some higher term. In this chapter we'll pursue such terms, while still making the far-field approximation. When convenient, we'll again suppose that the current and charge distribution is harmonic in time with frequency ω . In short, we'll take a second look at

$$\underline{A}^\mu(t, \vec{r}) = \frac{\mu_0}{4\pi r} e^{-i\omega(t-r/c)} \int d^3r_* e^{-i\epsilon_{\text{multi}} \hat{r} \cdot \vec{r}_*/a} \frac{1}{2} \underline{J}^\mu(\vec{r}_*) + \text{c.c.} \quad [42.4, \text{page 438}]$$

Recall that in this formula, t and \vec{r} (and hence also \hat{r}) refer to the observation, whereas \vec{r}_* is a source point. ϵ_{multi} is the small quantity controlling the multipole expansion (Equation 42.5, page 438), and a is the overall source size (upper bound on $\|\vec{r}_*\|$).

43.2 NEXT-ORDER TERMS

43.2.1 Order-one terms in ϵ_{multi} can be divided into two tensor structures

Proceeding as before, we now expand the exponential factor inside the integral in Equation 42.4. Chapter 42 evaluated the zeroth-order term, which we'll now call $\vec{A}^{[0]}$; instead, now we focus on first order in ϵ_{multi} . We'll call the three spatial components of that term $\vec{A}^{[1]}$:

$$\vec{A}^{[1]}(t, \vec{r}) = \frac{\mu_0}{4\pi r} e^{-i\omega(t-r/c)} \int d^3r_* (-i\epsilon_{\text{multi}} \hat{r} \cdot \vec{r}_*/a) \frac{1}{2} \vec{j}(\vec{r}_*) + \text{c.c.}$$

We can write $-i\epsilon_{\text{multi}}/a$ as $c^{-1} \frac{d}{dt}$:

$$= \frac{\mu_0}{4\pi r c} \hat{r} \cdot \frac{d}{dt} \left[\underbrace{\int d^3r_* \vec{r}_* \vec{j}(t-r/c, \vec{r}_*)}_{\text{}} \right].$$

The expression in the brace is a rank-2 3-tensor that depends on the observer's position only via the retarded time $t_r = t - r/c$. We'll call it $\vec{\Gamma}(t_r)$; it is a kind of moment.

Like any second-rank tensor, $\vec{\Gamma}$ can be written as the sum of its symmetric and antisymmetric pieces, which we'll call

$$\vec{\Gamma} = \vec{\Gamma}^{[EQ]} + \vec{\Gamma}^{[MD]}$$

respectively.

43.2.2 Antisymmetric part of the moment

Like any antisymmetric second-rank 3-tensor, we may reexpress the three independent entries of $\vec{\Gamma}^{[MD]}$ in terms of a single pseudovector (Equation 16.5, page 188):

$$\vec{\Gamma}_{np}^{[MD]} = \epsilon_{npi} \vec{\mathcal{D}}_{M,i} \quad \text{where} \quad \vec{\mathcal{D}}_{M,i} = \frac{1}{2} \epsilon_{iks} \int d^3r_* \vec{r}_{*k} \vec{j}_s. \quad (43.1)$$

Your Turn 43A

- Establish Equation 16.5.
- Show that $\vec{\Gamma}^{[MD]}$ contributes

$$\vec{A}^{[MD]} = -\frac{\mu_0}{4\pi r c} \hat{r} \times \frac{d}{dt} \vec{\mathcal{D}}_M \Big|_{\text{ret}} \quad \text{far field}$$

to $\vec{A}^{[1]}$.

Your result implies that

- This part of the far field is also a spherical wave (because $\vec{A}^{[MD]}$ depends harmonically on time via $t - r/c$).
- It falls like r^{-1} , and hence can potentially transport energy to infinity.

Your Turn 43B

- Do a calculation similar to the one in Section 42.6.1 to show that

$$\vec{B}^{[MD]} = \frac{\mu_0}{4\pi r c^2} \hat{r} \times \left(\hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_M \Big|_{\text{ret}} \right). \quad \text{far field} \quad (43.2)$$

- Then use Ampère's law to find $\vec{E}^{[MD]}$.

Remarkably

The MD contribution to the magnetic field looks just like the ED contribution to the electric field. The MD contribution to the electric field looks just like the ED contribution to the magnetic field.

Consider a circular loop of wire in the xy plane, carrying current with amplitude \bar{I} and frequency ω , has no net charge anywhere, and hence vanishing electric dipole and quadrupole moments. But you found in Your Turn 16A (page 188) that the magnetic dipole moment is nonzero: $\vec{\mathcal{D}}_M = (\hat{z}\Sigma)(\bar{I} \cos \omega t)$.

Your Turn 43C

- Find the far electric and magnetic fields and compare to your earlier result obtained in Coulomb gauge (Your Turn 23E, page 264).
- Find the Poynting vector and compare with the result in Coulomb gauge (Your Turn 23E).
- Integrate the Poynting vector over all directions \hat{r} .
- Time-average your result from (c) to show that

$$\langle \mathcal{P}^{[MD]} \rangle = \left(\frac{\omega}{c}\right)^4 \frac{(\bar{I}\Sigma)^2}{12\pi\epsilon_0 c}. \quad (43.3)$$

43.2.3 Symmetric part of the moment

To simplify $\vec{\Gamma}^{[EQ]}$, we now use a trick remembered from magnetostatics (Section 16.1): The divergence theorem gives that

$$0 = \int d^3r_* \vec{\nabla}_{*i} (\vec{r}_{*k} \vec{r}_{*m} \vec{j}_i(\vec{r}_*)),$$

where $\vec{\nabla}_*$ denotes partial derivatives with respect to \vec{r}_* . Thus

$$\vec{\Gamma}_{mk}^{[EQ]} = \frac{1}{2} \int d^3r_* (\vec{r}_{*m} \vec{j}_k + \vec{r}_{*k} \vec{j}_m) = -\frac{1}{2} \int d^3r_* \vec{r}_{*m} \vec{r}_{*k} \vec{\nabla} \cdot \vec{j}.$$

For a static current distribution, this quantity would be zero by the continuity equation. More generally, however, we get

$$= \frac{1}{2} \frac{d}{dt} \int d^3r_* \vec{r}_{*m} \vec{r}_{*k} \rho_q|_{\text{ret}}.$$

That is, this term involves the second moment of electric charge. We can write that moment as its traceless part plus the rest, using Equation 3.2 (page 31):

$$\frac{1}{3} \vec{\mathcal{Q}}_{E,ml}|_{\text{ret}} + \frac{1}{3} \mathbf{1}_{ml} \int d^3r_* r_*^2 \rho_q|_{\text{ret}}.$$

So the contribution of $\vec{\Gamma}^{[EQ]}$ to the first-order term of the vector potential, $\vec{A}^{[1]}$, can be written as

$$\begin{aligned} \vec{A}^{[EQ]} &= \frac{1}{6} \frac{\mu_0}{4\pi r c} \frac{d}{dt} \hat{r} \cdot \left(\frac{d}{dt} \vec{\mathcal{Q}}_E|_{\text{ret}} + \mathbf{1} \int d^3r_* r_*^2 \frac{d}{dt} \rho_q(t-r/c, \vec{r}_*) \right) \\ &= \frac{\mu_0}{24\pi c} \left[r^{-1} \hat{r} \cdot \frac{d^2 \vec{\mathcal{Q}}_E}{dt^2} \Big|_{\text{ret}} + \hat{r} r^{-1} \int d^3r_* r_*^2 \frac{d}{dt} \rho_q(t-r/c, \vec{r}_*) \right]. \end{aligned}$$

The second term of this expression looks complicated, but it's purely a *gradient*, and hence cannot contribute to the magnetic field. Equivalently, it can be removed by an appropriate gauge transformation, leaving

$$\vec{A}^{[EQ]} = \frac{\mu_0}{24\pi c r} \hat{r} \cdot \frac{d^2 \vec{\mathcal{Q}}_E}{dt^2} \Big|_{\text{ret}} \quad \text{far field} \quad (43.4)$$

Once again, we have found an outgoing spherical wave (the potential depends harmonically on $t - r/c$), falling in the far field region like r^{-1} . Compared with electric dipole radiation, EQ radiation is suppressed by an extra factor of $\epsilon_{\text{multi}} = \omega a/c$, but it can be the leading term for a source with dipole moments everywhere equal to zero.

43.3 HIGHER ORDERS

Clearly we could carry out the expansion to next order in ϵ_{multi} to find $\vec{A}^{[2]}$, with contributions from magnetic quadrupole and other terms. In the electrostatic and magnetostatic multipole expansions, we found that each successive order gave fields falling off with distance faster than the previous one. In contrast, for time-varying sources

Every order of the multipole expansion gives a contribution whose leading far-field behavior is always $1/r$. Each order is suppressed relative to the previous one by an additional factor of frequency, not $1/r$.

Thus, all of the orders create outgoing spherical waves, so they can all transport energy to infinity.

In greater detail, we have in the far field approximation

$$\frac{d}{dt}\vec{E} = c^2\vec{\nabla} \times \vec{B} \approx -c\hat{r} \times \vec{B}$$

so

$$S = \frac{1}{\mu_0}\vec{E} \times \vec{B} = \frac{c}{\mu_0}\vec{B} \times (\hat{r} \times \vec{B}) = \frac{c}{\mu_0}\hat{r}\|\vec{B}\|^2,$$

and each nonzero term of

$$\|\vec{B}^{[0]} + \vec{B}^{[1]} + \vec{B}^{[2]} + \dots\|^2 \quad (43.5)$$

falls with distance as r^{-2} .

Let's consider the various contributions according to their order in the multipole expansion parameter. Equation 42.10 (page 443) gave the $\|\vec{B}^{[0]}\|^2$ term (electric dipole), and Equation 42.13 (page 444) gave its integral over all directions. If this term is nonzero, it's the most important one.

The cross term $2\vec{B}^{[0]} \cdot \vec{B}^{[1]}$ integrated over angles gives zero.¹ So the next most important terms involve $\|\vec{B}^{[1]}\|^2$ (magnetic dipole and electric quadrupole, Problem 43.7) and $2\vec{B}^{[0]} \cdot \vec{B}^{[2]}$ (the “anapole” term²).

43.4 PLUS ULTRA

A *spherically symmetric* charge distribution will not radiate, no matter how it depends on time. For example, its monopole moment is fixed by charge conservation, and hence has vanishing time dependence. We also saw above how the first orders of the expansion involve \vec{D}_E , \vec{D}_M , \vec{Q}_E , etc., all of which are zero for a spherically symmetric distribution.

¹See Problem 43.6.

²See Rovenchak & Krynytskyi, 2018.

The deeper point is that a spherically symmetric distribution, for example a shell of charge that grows and shrinks over time, carries no angular momentum, and yet photons, which have spin 1, must carry net angular momentum.

FURTHER READING

General: Zangwill, 2013, §20.8.

Anapole moment: Stefan Nanz, *Toroidal Multipole Moments in Classical Electrodynamics* (Springer 2016).

Anapole radiation, frequently neglected in textbooks: Rovenchak & Krynytskyi, 2018.

PROBLEMS

43.1 *MD antenna*

Chapter 23 discussed the radiation we see when standing far away from an oscillating magnetic dipole. Specifically, the dipole was oriented with its moment in the $\pm\hat{z}$ direction, we imagined measuring the fields at $\vec{r} = (L, 0, 0)$, and we only asked for the leading order term in powers of $1/L$. You found a formula for the vector potential (Your Turn 23E), but even with the far-field limit it was still involved a complicated integral. In this problem, you'll find a simplified expression in a special limiting case.

Consider a series of loops with smaller and smaller radii b . However, each loop also has a larger current than the previous one, in such a way that the magnetic dipole moment $\mathcal{D}_M(t) = \bar{\mathcal{D}}_M \cos(\omega t)$ is the same for all. In this limit (and also the far-field limit), find a simplified form for the vector potential, magnetic field, and electric field observed far from the source along the x axis. If the outgoing wave is polarized, describe its polarization. Also characterize how the energy density falls with distance.

43.2 *Double-loop antenna*

In class we considered an antenna consisting of a circular loop of wire driven by an oscillator. In this problem we consider an antenna consisting of *two* circular loops, each of radius a and parallel to the xy plane, and centered on the z axis at heights $z = \pm a$. Here $a \ll r$. The currents in these loops are $\pm \frac{1}{2}(\bar{I}e^{i\omega t} + \text{c.c.})$ respectively.

Find the lowest-order multipole radiation fields produced by this system.

43.3 *Pulsar I*

A pulsar is a compact star with a large magnetic dipole moment $\bar{\mathcal{D}}_M$ frozen into it. The pulsar rotates uniformly as a solid body (carrying the dipole moment along), with angular velocity ω . The dipole moment is located at the center and oriented at angle α relative to the rotation axis.

The moment $\bar{\mathcal{D}}_M$ is related to the strength of B_{pole} of the magnetic field at the magnetic pole by $|\bar{\mathcal{D}}_M| = \kappa B_{\text{pole}}$, where κ is a constant.

a. Find the rate at which the pulsar radiates electromagnetic energy, as a function of κ , B_{pole} , ω , and α . [*Hint*: The magnetic dipole radiation formula has the same general form as the electric dipole radiation formula.]

- b. If the source of the energy is the pulsar's rotational kinetic energy, $\mathcal{E} = \frac{1}{2}I\omega^2$ with I = pulsar's moment of inertia, find the characteristic slowdown time scale $T \equiv -\frac{\omega}{d\omega/dt}$ as a function of κ , B_{pole} , ω , and α , and I .
- c. Suppose that the pulsar has radius R , and get a formula for κ in terms of R .
 [Remark: Thus we get a prediction of the slowdown in terms of B_{pole} , R , ω , α , and I .]

43.4 Pulsar 2

A pulsar can be modeled as a rotating neutron star. Such a star is likely to have strong magnetic field B_0 at its surface, because it traps lines of force during its collapse, and it's also likely to be spinning rapidly, by conservation of angular momentum during the collapse. Suppose that the magnetic field is predominantly dipole. If the magnetic dipole axis does not align with the rotation axis, the star will have a time-dependent magnetic moment $\vec{\mathcal{D}}_M(t)$; as with an ordinary permanent magnet, $\vec{\mathcal{D}}_M$ is frozen with respect to a body-fixed coordinate frame.

It may seem hard to measure observationally the value of $|\vec{\mathcal{D}}_M|$, but it's related to the strength of B_0 of the magnetic field at the magnetic pole, and one can at least estimate B_0 from the Zeeman splitting of spectral lines (Problem 17.7). Thus the input parameters of the problem are B_0 and the mass M , star radius R , angle between axes α , and rotation angular velocity ω .

- a. Find an expression for the radiated power \mathcal{P} , in terms of ω , $|\vec{\mathcal{D}}_M|$, and α .
- b. Assume that rotational kinetic energy of the star is the ultimate source of the energy given off by the pulsar, and that the pulsar is a sphere of uniform mass density. Find a formula for the characteristic slowdown time scale $T \equiv -\frac{\omega}{d\omega/dt}|_{t=0}$, as a function of B_0 , R , α , M , and the initial value of ω .
- c. Use typical numbers $M = 1$ solar mass $= 2 \cdot 10^{30}$ kg, $R = 10$ km, $B_0 = 10^8$ T, and assume $\alpha = 90^\circ$. Evaluate \mathcal{P} and τ for $\omega(0) = 10^4 \text{ s}^{-1}$, a frequency thought to be typical of newly formed pulsars.

43.5 Exact MD wave

We have found the general solution to Maxwell's equations in Lorenz gauge with the outgoing wave boundary condition, but it's too awkward for many purposes. We did find some simple, exact solutions (plane waves), but they're not appropriate for outgoing radiation from a finite-size source. We also found some approximate solutions (multipole far fields), but they're... approximate.

I did give you one *exact* solution describing outgoing waves from a point (see the Chapter 37), but I hinted that it was just the first in a series of such exact solutions. Can we find any others, (almost) equally simple?

To explore this question, consider again the situation with a circular loop of wire, of radius b , in the xy plane, centered on the origin, carrying alternating current $I(t) = \frac{1}{2}\bar{I}e^{-i\omega t} + \text{c.c.}$ The corresponding fields are complicated. They become simpler, however in the "pure magnetic dipole" limit, in which the radius of the loop $b \rightarrow 0$ while the magnetic dipole moment $\vec{m} = \pi b^2 \bar{I} \hat{z}$ is held constant.³

Evaluate the Lorenz-gauge vector potential $\vec{A}(t, \vec{r})$ exactly in this limit. That is, don't make the far-field approximation, in which higher powers of $1/r$ are dropped.

³The observer's location \vec{r} , and the angular frequency ω , are also held constant in the limit.

You'll get an expression for $\vec{A}(t, \vec{r})$ that is almost as simple as the far field approximation that we worked out in class. Confirm directly that it does give an exact solution to Maxwell's equations in Lorenz gauge, everywhere away from the singularity at the origin.

43.6

Using Equations 42.9 (page 440), 43.2, and 43.4, show that the cross term $2\vec{B}^{[0]} \cdot \vec{B}^{[1]}$ integrated over angles gives zero. Thus there is no term in the total radiated power that is first order in the small parameter ϵ_{multi} . [*Hint:* You will encounter the angular average of $\hat{r}_i \hat{r}_j \hat{r}_k$. It must be a rotationally invariant, rank-3, 3-tensor. There's no such thing, so this average must equal zero.]

43.7 *Electric quadrupole radiation*

When we expand Equation 43.5, the term $\|\vec{B}^{[1]}\|^2$ includes the cross term $2\vec{B}^{[EQ]} \cdot \vec{B}^{[MD]}$. Show that in the far-field approximation, this term gives zero when integrated over outgoing directions \hat{r} , leaving only the contribution already found in Equation 43.3 (page 451), plus one other subterm that you are to find.

CHAPTER 45

The Microwave Polarizer

45.1 FRAMING

On the first day of class I brought in a microwave generator. We now know how it emits linearly polarized radiation. I also had a detector with a similar antenna, which was therefore sensitive to just one polarization. Finally, I introduced a polarizer (a planar array of long, thin, parallel copper wires). Interesting Electromagnetic Phenomena ensued.

45.2 A SIMPLIFIED CASE

The wire spacing was smaller than the wavelength, so let's model the microwave polarizer as a thin, planar conducting sheet at $z = 0$. It's highly anisotropic, conducting easily in one direction but not the other. Thus, the surface current density $\vec{j}^{(2D)}$ is related to the field \vec{E} by a 2D tensor, the surface conductivity:

$$\vec{j}^{(2D)} = \vec{\kappa}_s \cdot \vec{E}, \quad \text{where} \quad \vec{\kappa}_s = \kappa_s \hat{x}\hat{x}. \quad (45.1)$$

We approximate the incoming fields far from the source as a plane wave traveling along \hat{z} , and begin by supposing that it is linearly polarized along the conducting direction:

$$\vec{E} = \frac{1}{2} \bar{E} \hat{x} e^{-i(\omega t - kx)} + \text{c.c.}, \quad \text{where} \quad k = \omega/c.$$

We will also simplify by considering a poor conductor, that is, κ_s is small. Then each surface element will have little influence on the others; each just responds to the incoming plane wave \vec{E}_{in} via our ohmic hypothesis. Each surface element responds in phase with the others. Each in turn radiates according to the Green function solution. For example, at a point along the $+\hat{z}$ axis we have a total radiation field from all surface elements given by

$$\vec{A}_{\text{rad}} = \hat{x} \frac{\mu_0}{4\pi} \int d^2 r_* \frac{1}{R} \kappa_s \frac{1}{2} \bar{E} e^{-i\omega(t-R/c)} + \text{c.c.}$$

The integral runs over the whole plane $z = 0$. Let $k = \omega/c$.

This kind of integral comes up in many contexts, and it has a surprising feature, so let's pause to consider it carefully. We switch to plane polar coordinates; the integral over φ_* just gives 2π and we are left with $r_* dr_*$. The integrand, $R^{-1} e^{i\omega R/c}$, is a messy function of r_* , but there is an amazing trick. At an observation point along the $+\hat{z}$ axis ($z > 0$), we have $R^2 = r_*^2 + z^2$, so $RdR = r_* dr_*$. Thus, we can change variables in the integral to get

$$\int_0^\infty d^2 r_* R^{-1} e^{ikR} = 2\pi \int_z^\infty dR e^{ikR}.$$

That integral is easy! But it's confusing:

$$= \frac{c}{i\omega} \left[e^{i\infty} - e^{ikz} \right].$$

To understand that first term, suppose that our plane had a large, but finite, extent L . Then this term would give a contribution to the potential that oscillates as we consider larger L . But let's compute the magnetic field, a physical quantity:

$$\vec{\nabla} \times \vec{A} = \frac{\mu_0}{4\pi} \frac{2\pi\kappa_s \bar{E}}{2ik} e^{-i\omega t} (ik)(-\hat{x}) \times \hat{z} \left(-e^{ikz} + \frac{z}{\sqrt{L^2 + z^2}} e^{ik\sqrt{L^2 + z^2}} \right) + \text{c.c.}$$

Taking $L \rightarrow \infty$ at fixed z , we see the second term may be dropped:

$$\vec{B} = -\hat{y} \frac{\kappa_s \bar{E} \mu_0}{4} e^{-i(\omega t - kz)} + \text{c.c.}$$

Your Turn 45A

Compute the electric field as usual, obtaining

$$\vec{E}_{\text{rad}} = -\hat{x} \frac{\kappa_s \mu_0 \bar{E} c}{4} e^{-i\omega(t-z/c)} + \text{c.c.}$$

Remarkably, the forward scattered field is again a plane wave traveling along \hat{z} , but *180 degrees out of phase with the incoming wave*. The total forward wave is then

$$\vec{E}_{\text{tot}} = \hat{x} \frac{1}{2} \bar{E} \left(1 - \frac{1}{2} \kappa_s \mu_0 c \right) e^{-i\omega(t-z/c)} + \text{c.c.} \quad (45.2)$$

The transmitted wave has lost some of its amplitude.

Where did that energy go? Its flux decreased by the square of the factor in parentheses, or $\approx (1 - \kappa_s \mu_0 c)$ (remember that we work only to lowest order in κ_s). You should work out the radiated wave \vec{E}_{rad} in the backward (reflected) direction, along $-\hat{z}$, but clearly its energy flux will be proportional to $(\kappa_s)^2$, and so cannot fully account for the effect that we found. Instead, we must look for the culprit elsewhere.

A conductor with finite conductance *dissipates* energy as heat. The total loss is

$$\int d^2 r_* \vec{E} \cdot \vec{j}^{(2D)}, \quad (45.3)$$

where the surface current density $\vec{j}^{(2D)}$ is given by Equation 45.1. The loss per unit area is just the integrand of Equation 45.3.

Your Turn 45B

Add it to the energy flux from Equation 45.2 and compare to the incoming energy flux.

45.3 EFFECT ON ARBITRARILY POLARIZED WAVE

Equation 45.1 says that the conductivity tensor's principal directions are \hat{x} (eigenvalue κ_s) and \hat{y} (eigenvalue 0). We just saw that an incoming wave polarized along \hat{x} will excite no currents and hence will be attenuated. However, a wave polarized along \hat{y} will excite no currents and hence will be *unaffected*—as we saw on the first day of class.

45.4 REGENERATION

Finally, suppose that the incoming wave polarization is linear but tilted by 45° : $\vec{E}_{\text{in}} = \bar{E}(\hat{x} + \hat{y})/\sqrt{2}$. Now we find the forward wave to be

$$\vec{E}_{\text{tot}} = \frac{1}{2}\bar{E}\left(\frac{\hat{x} + \hat{y}}{\sqrt{2}} - \frac{\kappa_s\mu_0 c\hat{x}}{2\sqrt{2}}\right)e^{-i\omega(t-z/c)} + \text{c.c.}$$

We can reexpress the second term in the tilted basis by noting that $\hat{x} = (\hat{x} + \hat{y})/2 + (\hat{x} - \hat{y})/2$. The first of these terms destructively interferes with the incoming beam as before. The other one, however, generates a “transmitted” wave with a *polarization not present in the incoming wave*, another Electromagnetic Phenomenon we observed on the first day of class.

PROBLEMS

45.1 *Another integral*

Another situation of interest involves a plane wave that impinges on a dielectric (nonconducting but polarizable) sheet. We then need an integral of the form $\int_0^\infty (2\pi\rho d\rho)(1 - \cos^2\alpha)e^{ikr}/(4\pi r)$. Here $r = \sqrt{\rho^2 + (z_*)^2}$ and $\cos\alpha = \rho/r$. k , z_* are constants.

Following the discussion in the *Feynman Lectures*,¹ we can wave our hands a bit and argue that this integral is approximately equal to $\frac{i}{2k}e^{ikz_*}$. You may or may not find this argument convincing, but either way, it's good to check. Unfortunately this integral is probably not one you have met in calculus. Fortunately, however, we can simplify it to the point where a computer can help us. Notice that the problem contains two parameters, z_* and k . There is only one dimensionless combination of these parameters; call it $M = kz_*$.

- Change variables in the integral from ρ to r . Define dimensionless variable $u = kr$, and express the thing that is to be shown in terms of it. Express it in the form (a certain integral) ≈ 1 .
- Figure out how to get your favorite mathematical software to do this integral. Evaluate it for various values of M and check our expectation. [One visually appealing way could be to graph the real and imaginary parts of the quantity you found in (a) as functions of M .]

¹Volume 1, sections (30-7)–(31-2).

c. Are there some values of M for which our expectation is more, or less, accurate?

Hint: To get started, you'll need to understand how to do integrals numerically. I had to look in MATLAB's help under `quad`. In order to understand the help, I then needed to look up `function handles`. I still got stuck until I noticed the remark "the function should accept a vector argument x and return a vector result, the integrand evaluated at each element of x ." Thus, `quad(@(x) x^2,0,1)` is an error, whereas `quad(@(x) x.^2,0,1)` yields $1/3$. Try it yourself with some other favorite integral of yours before doing the problem.

CHAPTER 46

Scattering by Free and Bound Charges

See also P+S §15.5.

46.1 SHAKE IT

When an EM wave encounters a charged particle, we've seen that it shakes the particle. Chapter 18 considered the rather fanciful situation of a particle subject to “viscous friction.” A more realistic case is a *free* charged particle. For example, in a low-density plasma like the early Universe just prior to recombination, atoms are dissociated into nuclei and electrons, each of which feels an overall potential due to all the others but is not bound to any specific partner.¹

Let's investigate the simplest case, with a *single* free charge q , of mass m . We will assume that the charge's motion is always nonrelativistic (and later justify that assumption, in a limit that we will make precise). Write an incident plane wave as

$$\vec{E}(t, \vec{r}) = \frac{1}{2} \vec{E} e^{-i(\omega t - \vec{k} \cdot \vec{r})} + c.c.$$

(and the associated \vec{B} field). The charge sits at $\vec{r} = 0$, so it feels an electric force $\vec{f} = q\vec{E}$. The transverse magnetic force is negligible because $E = cB$ so $q\vec{v} \times \vec{B} \sim q(v/c)E$. Our assumption of nonrelativistic motion, $v/c \ll 1$, means that we can neglect this part of the force.²

Write the resulting motion as $\vec{r}'(t) = \frac{1}{2} \vec{r} e^{-i\omega t} + c.c.$ Then Newton's law gives the amplitude of the shaking motion as $\vec{r} = -(q\vec{E})/(m\omega^2)$, whose velocity will be $\ll c$ if

$$\|q\vec{E}\| \ll m\omega c. \quad \text{condition for nonrelativistic motion} \quad (46.1)$$

So our assumption is justified for weak enough fields. In practice, this condition is nearly always well satisfied.³

46.2 THOMSON CROSS SECTION

Our shaking charge gives rise to a time-dependent dipole moment $\vec{D}_E(t) = q\vec{r}(t)$, so it will radiate at the same frequency. The charge's motion remains confined to a region of size $\|\vec{r}\|$. The criterion for the ED approximation is met:

$$\|\vec{r}\| \omega/c = (q\vec{E})/(m\omega^2 c) \omega \ll 1,$$

¹A situation effectively like this one also holds for some of the electrons in a metal.

²Chapter 18 studied the *longitudinal* force, for which the magnetic part was the leading term and so could not be dropped.

³OK, not in the free electron laser.

by virtue of Equation 46.1. We can therefore use the ED radiation formulas to find the energy flux in any direction.

Chapter 42 gave the energy flux for a time-dependent, linear dipole as

$$\vec{S} = \hat{r} \frac{\mu_0}{(4\pi r)^2} \frac{1}{c} \frac{d^2}{dt^2} \mathcal{D}_E^2 \sin^2 \vartheta, \quad [42.12, \text{page } 444]$$

where ϑ is the angle between the dipole moment and the direction of observation. In our case, suppose that the incoming wave is polarized along \hat{x} ; then $\vec{\mathcal{D}}_E(t) = \hat{x} \frac{1}{2} \bar{\mathcal{D}}_E e^{-i\omega t} + \text{c.c.}$, with $\bar{\mathcal{D}}_E = -q^2 \bar{E}/(m\omega^2)$. The power output per solid angle is then

$$\left\langle \frac{d\mathcal{P}}{d\Omega} \right\rangle = \langle r^2 \hat{r} \cdot \vec{S} \rangle = \frac{1}{(4\pi)^2} \frac{1}{\epsilon_0 c^3} \frac{q^4}{m^2} \frac{1}{2} \|\vec{E}\|^2 \sin^2 \vartheta. \quad (46.2)$$

Remarkably, the frequency ω drops out of this formula. Note, too, that the incident wave's *direction* \hat{k} is irrelevant, other than that it defines the plane of allowed directions for $\vec{\mathcal{D}}_E$. Finally, note that a free proton is much less effective at scattering than a free electron, due to the $1/m^2$ factor.

Equation 46.2 tell us something about how good our charge is at scattering radiation, but it's not intrinsic to the charge—it also depends on the strength of the incoming field. To get something intrinsic, we need to *normalize* it by some measure of the strength of the incoming wave. How should we do that? The total power transported by a plane wave is *infinite*, because of its infinite extent in the transverse directions. But most of that extent is irrelevant—bits of the wave that never come near the charge just cruise by without scattering.

The key insight is that the energy *flux* (power per unit area) is finite. Think about holding a penny in the sunlight. The energy removed from the incoming beam (reflected, absorbed, whatever) equals the solar energy flux times the cross-sectional area of the penny, or

$$\text{cross section} = (\text{energy removed from beam})/(\text{energy flux incoming}).$$

Note how the units work out: energy and time cancel, leaving behind $1/(1/\mathbb{L}^2)$, or area. The infinite transverse extent of the incoming beam is irrelevant, as desired.

We can similarly characterize how good a single electron is at scattering light by forming the same quotient; the intensity of the incoming beam cancels from numerator and denominator, leaving behind a quantity with units of area, which we will again call “cross section” by analogy to the macroscopic situation. We just need a formula for the denominator:

$$\langle \|\vec{S}_{\text{in}}\| \rangle = \langle \mu_0^{-1} \|\vec{E} \times \vec{B}\| \rangle = \frac{1}{2} \epsilon_0 c \|\vec{E}\|^2.$$

The cross-section is traditionally denoted σ . Extending our original thought experiment, we can subdivide this scattering cross section into bits attributable to scattering into particular angular bins $d\Omega$, or:

$$\frac{d\sigma}{d\Omega} = \left\langle \frac{d\mathcal{P}}{d\Omega} \right\rangle / \left\langle \|\vec{S}_{\text{in}}\| \right\rangle.$$

This quantity is generically called the **differential scattering cross-section**.

For the case of scattering from a single electron, in classical electrodynamics, combining the preceding generic formula with Equation 46.2 gives

$$\frac{d\sigma}{d\Omega} = \left(\frac{1}{4\pi\epsilon_0 c^2} \frac{q^2}{m} \right)^2 \sin^2 \vartheta. \quad \text{Thomson scattering cross-section}$$

Your Turn 46A

Confirm that the constants in brackets really do combine into a quantity with dimensions of length, and evaluate it for q and m appropriate for an electron. This quantity is called the **classical electron radius**, or r_c .

Often we don't care about angular dependence; we only want to know how much energy the electron scatters out of the beam. For this, we can integrate the Thomson formula over all directions, using

$$\int d\varphi d(\cos \vartheta) \sin^2 \vartheta = 8\pi/3.$$

The total scattering cross-section obtained in this way is $\sigma = (8\pi/3)r_c^2$, a useful number you should evaluate for electrons.

46.3 AN ASTROPHYSICAL APPLICATION

The Sun's interior is hot. There's a lot of light in there. And yet, that light takes a long time to make its way to the surface of the Sun. One way to think about this is to imagine the light constantly scattering, changing direction. Although any one electron in this plasma isn't very effective at scattering light, there are quite a lot of electrons. So the light must take a zigzag path; even though it's traveling at c between collisions, nevertheless that path will be much longer than the Sun's diameter, so traversing it takes a lot of time.

The quantity that characterizes the tortuous light trajectories is a "mean free path." Dimensional analysis suggests that, to get dimensions of length, we need to form the quantity $1/(r_c^2 n)$, where n is the density of free electrons. The mean free path for light is this quantity times some geometrical constants of order one.

46.4 POLARIZED INCOMING LIGHT

Suppose that the incoming light travels along \hat{z} , with polarization along \hat{x} . Then $\vec{D}_E \parallel \hat{x}$. The electric far field points along $\hat{x} - \hat{r}(\hat{r} \cdot \hat{x})$; that is, it lies in the plane spanned by \hat{r} and \hat{x} and (as always) transverse to \hat{r} .

Linearly polarized light always scatters to some kind of linearly polarized light, regardless of the scattering direction (or to *nothing* if we observe along the direction of polarization, $\hat{r} \parallel \hat{x}$).

46.5 UNPOLARIZED INCOMING LIGHT

So far, we have been considering a monochromatic, pure incoming wave, and in particular polarized. We can treat unpolarized light as an incoherent superposition of many pure waves.⁴ Scattering can *create* polarization from such light. For example, when viewed at 90° to the original wave's direction, the scattered light will be 100% linearly polarized: One component of the incoming light shakes electrons longitudinally to that viewing direction, so there is no reradiation in that direction at all. At other scattering angles, the light interpolates between that extreme value and 0% for the forward and backward directions.

Please read the posted pages from Dodelson's book about how we can use these observations to learn about the early Universe from the faint polarization pattern in the cosmic microwave background radiation.⁵

46.6 BOUND CHARGES

46.6.1 Rayleigh scattering cross section

Next suppose that the charge is bound, for example, to a heavy atomic nucleus. The simplest classical model we can make of that situation is to suppose that the charge gets a linear restoring force with some spring constant k . As usual with harmonic oscillators, it is convenient to introduce $\omega_0 = \sqrt{k/m}$. Then Newton's law becomes

$$-m\omega^2\vec{r} = -\omega_0^2\vec{r} + q\vec{E}, \quad \text{so} \quad \vec{r} = \frac{q\vec{E}}{m(\omega_0^2 - \omega^2)}.$$

Substituting this expression into earlier results then gives the Thomson expression for differential and total cross-sections, each multiplied by $(1 - (\omega_0/\omega)^2)^2$. Either of these formulas is called the **Rayleigh cross-section** formula. In particular, the differential cross section has the same polarization behavior as what we already observed for free charges.

Two limiting cases are noteworthy: At high frequency $\omega \gg \omega_0$, our results reduce to the Thomson formulas. In this regime, the fact that the charge is bound is immaterial to its response. In the opposite limit, we get the Thomson formulas multiplied by $(\omega/\omega_0)^4$: The cross-section is now strongly frequency dependent.

46.6.2 Blue sky

Earth's upper atmosphere consists of polarizable objects (molecules) that are much smaller than the wavelength of visible light, at low enough density that we may neglect their mutually interactions and treat them as independently scattering sunlight to our eyes. They are also randomly placed in space, which eliminates any coherent effects from multiple scattering. In such a situation, the fact that there are many such molecules just amplifies the scattering without changing its character. Indeed, we know that

⁴See Chapter 22.

⁵Dodelson, 2003, pp310–319.

- The scattered light is polarized in a way that depends on the direction of the line of sight relative to the incoming beam.
- The scattered light is bluer than sunlight itself (higher frequencies scatter more strongly).
- At sunset, we observe sunlight through a thicker layer of air than at noon, and direct (unscattered) light is redder (more depleted of high frequencies) than at noon.

46.6.3 A demo

I sent a beam of white light from a projector into a dilute suspension of nonfat milk. Milk is mainly water of course, but that component is irrelevant. Milk is also a colloidal suspension of fat globules, but I used nonfat milk for the demo. Milk contains dissolved lactose, etc., but that just gives a solution that's homogeneous on the scale of wavelength of light (it alters the index of refraction), and so again is irrelevant for scattering. What's *important* is that nonfat milk is a colloidal suspension of protein micelles, which (a) are well separated compared to light wavelength; (b) are themselves much smaller than wavelength of light (nanometer scale); (c) move randomly and independently; and (d) have polarizability different from that of the surrounding water. Thus, we were in a situation similar to that of sunlight on the upper atmosphere. You saw what you saw.

T₂ Track 2

46.2'

At high frequencies, the quantum character of light starts to matter. Dimensional analysis gives us a clue: We can form another length scale, the **Compton wavelength** $2\pi\hbar/(mc)$, by using Planck's constant. If the incoming light's wavelength is shorter than this, then we start to get billiard-ball collisions of electrons and single photons, the Thomson formula is no longer valid, and weirder still, the outgoing photon won't have the same frequency as the incoming one (Compton effect).

PROBLEMS

46.1 *Estimates and approximations*

A red laser gives a 100 mW beam that is approximately a plane wave with cross-sectional area 1 mm^2 .

- Find the electric field strength in this beam.
- Estimate the fractional deformation of a hydrogen atom placed in this beam, due to the electric field. Is it likely that we could make the approximation of working to first order in this deformation when we study polarizability?
- Suppose that this beam encounters a single free electron. The electron responds by oscillating. Justify our use of the nonrelativistic approximation for that motion.

46.2 *Diffusion of light*

Idealize the Sun as a highly ionized plasma with average free electron density about 10^{24} cm^{-3} .

- Use the Thomson formula to find the mean free path for electromagnetic radiation in the Sun, as a function of wavelength.
- Over lengths longer than the MFP, radiation takes a random-walk path out of the Sun. Estimate the time required for EM radiation to diffuse from the core to the outside, a distance of $7 \cdot 10^8 \text{ m}$.

PART VII

Light in Materials

FIG. 167.



CHAPTER 49

Isotropic, Linear Media

49.1 FRAMING

We now return to the study of nonconducting, but polarizable, media, in greater detail than Chapter 5. Thus, charges are not free to travel throughout the material; however, the individual molecules can deform slightly.

We will consider an approximation in which

- We assume that the medium consists of polarizable objects (or permanently polarized, unoriented objects which can become oriented by an external field). We only consider the dipole fields created by those objects.
- We will neglect all forms of energy dissipation. Thus, we exclude ohmic materials (conductors). It's not hard to add this feature to our equations.
- We assume that external fields vary over length scales much longer than the spacing between the polarizable constituents. We also suppose the latter to be finely enough divided (compared to the length scales of the disturbances we're studying) that they can be treated as a continuous density of dipole moment.¹

All formulas in this chapter are understood to be subject to the limitations of these approximations, whose domain of validity we won't explore.

49.2 POLARIZABLE MEDIA

49.2.1 Electric

First we must review and extend the discussion of dielectric materials from Chapter 5.² Let \vec{P} denote the net density of electric dipole moment. Chapter 5 argued that an interface, for example between a medium and vacuum, will develop a layer of **bound surface charge** with surface density σ_b given by

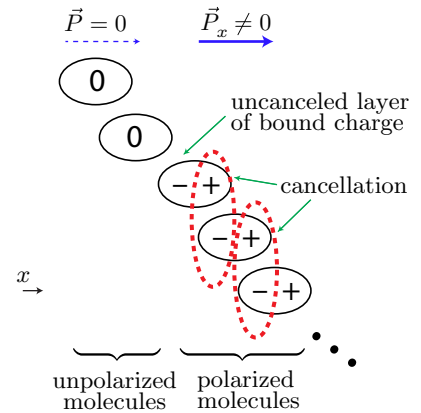
$$\sigma_b = \hat{n} \cdot \vec{P}, \quad [5.3, \text{page } 57]$$

where \hat{n} is the perpendicular unit vector directed outward. At an interface between two media, substitute the difference in \vec{P} values on either side.

¹Or equivalently, we average the effects of finite-size molecules over a length scale smaller than the one of interest, but much bigger than the molecular spacing. The quantities \vec{E} , \vec{P} , \vec{B} , \vec{M} below are all averages of this sort.

²Just don't confuse "dielectric material" with "dialectical materialism."

Figure 49.1: A collection of electrically polarizable “molecules” in a nonuniform electric field (magnitude increasing as we move to the right). Net bound charge appears that is minus the divergence of the polarization density, in this case $-\partial\vec{P}_x/\partial x$.



If \vec{P} is spatially nonuniform, it will also give rise to a **bound charge density** $\rho_{q,b}$ in the *interior* of a medium. Figure 49.1 shows a simple example of this effect. The general formula

$$\rho_{q,b} = -\vec{\nabla} \cdot \vec{P} \quad (49.1)$$

is rotationally invariant and agrees with the figure in the special case shown there. Mentally deleting the unpolarized “molecules” at the left of the figure also recovers our original discovery of Equation 5.3.

If the polarization is time-dependent, then the motion of bound charge will also give rise to a **bound charge flux** $\vec{j}_{b,P}$, via the continuity equation: $\partial\rho_{q,b}/\partial t = -\vec{\nabla} \cdot \vec{j}_{b,P}$. Substituting that result into Equation 49.1 gives

$$\vec{j}_{b,P} = \partial\vec{P}/\partial t \quad \text{electric contribution.} \quad (49.2)$$

To understand this result, suppose that \vec{P} is initially zero, then switches on to the form shown in Figure 49.1. Creation of the internal layer of negative bound charge requires net *flow* of charge to the right.

The **electric displacement** is defined by

$$\vec{D} = \epsilon_0\vec{E} + \vec{P}. \quad [5.5, \text{page 58}]$$

(We’ll just call it “the \vec{D} field.”) With these definitions, the electric Gauss law takes a simple form (Equation 49.6 below). The only source appearing explicitly in this formula is the free charge density.

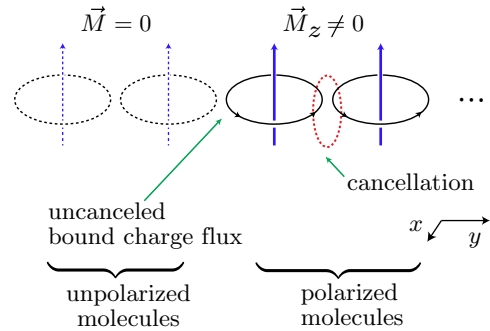
[T2] Section 49.2.1’ (page 509) introduces dissipation.

49.2.2 Magnetic

Let \vec{M} denote the net magnetic dipole moment density created by the motions of bound charges in individual polarizable objects. If \vec{M} is spatially nonuniform, it will give rise to a second contribution to the bound charge flux (in addition to Equation 49.2). Figure 49.2 shows a simple example of this effect. The general formula

$$\vec{j}_{b,M} = \vec{\nabla} \times \vec{M} \quad \text{magnetic contribution.} \quad (49.3)$$

Figure 49.2: A collection of magnetically polarizable “molecules” in a nonuniform magnetic field (magnitude increasing as we move to the right). *Black rings* indicate classical currents equivalent to those induced by the applied field. Net bound charge flux appears that is proportional to the curl of the magnetic moment density, in this case $\hat{x}(\partial\vec{M}_z/\partial y)$, which is directed out of the page as shown.



is rotationally invariant and agrees with the figure in the special case shown there. Mentally deleting the unpolarized “molecules” at the left of the figure also shows that at the boundary between medium and vacuum, we get a **bound surface charge flux**

$$\vec{j}_b^{(2D)} = \vec{M} \times \hat{n}, \quad (49.4)$$

where \hat{n} is the normal directed outward.

The **magnetic field intensity** is then defined by

$$\vec{H} = \mu_0^{-1}\vec{B} - \vec{M}. \quad (49.5)$$

(We’ll just call it “the \vec{H} field.”)

49.2.3 Maxwell

We wish to eliminate explicit mention of the bound charges and currents, a job that we began in Chapter 5. The remaining (non-bound) charges and currents are called “free”: $\rho_{q,f}$, \vec{j}_f . Excess static charges, which macroscopically violate charge neutrality, are considered free, for example, the charge delivered by the leads of a capacitor. Currents that transport net charge over macroscopic lengths are also considered free, for example, those in a coil of wire surrounding an inductor.

Your Turn 49A

Using Equations 49.1, 49.2, and 49.3, show that

$$\vec{\nabla} \cdot \vec{D} = \rho_{q,f} \quad \text{Gauss} \quad (49.6)$$

$$\vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = \vec{j}_f. \quad \text{Ampère} \quad (49.7)$$

Equation 49.6 extends the validity of Equation 5.4 to situations where the polarization is nonuniform. (The magnetic Gauss law and the Faraday law are unaffected because they do not involve charges or currents.)

49.2.4 Boundary conditions

We have already seen that the perpendicular component of the \vec{B} field must be continuous across a boundary between media:

$$\Delta B_{\perp} = 0. \quad \text{always} \quad [14.20, \text{page } 169]$$

We also saw that at a dielectric/vacuum interface, with no free surface charge nor current,

$$\hat{n} \cdot (\vec{E}^{\text{[vac]}} - \vec{E}^{\text{[1]}}) = \sigma_{\text{b}}/\epsilon_0, \quad [5.13, \text{page } 63]$$

$$\Delta \vec{E}_{\parallel} = 0, \quad \text{and} \quad [5.15, \text{page } 63]$$

$$\Delta \vec{B}_{\parallel} = \mu_0 \vec{j}_{\text{b}}^{\text{(2D)}} \times \hat{n}, \quad [14.21, \text{page } 170]$$

where \hat{n} points outward from medium 1 (toward the vacuum). At an interface between two magnetic media, or one such medium and vacuum, the contribution from bound currents can be incorporated into \vec{H} :

Your Turn 49B

Allow for free surface charge density and flux. Use Equations 5.13, 14.21, 5.3, and 49.4 to show that

$$\Delta D_{\perp} = \sigma_{\text{f}}; \quad \Delta \vec{H}_{\parallel} = \vec{j}_{\text{f}}^{\text{(2D)}} \times \hat{n}.$$

Here $\Delta D_{\perp} = (\vec{D}^{\text{[2]}} - \vec{D}^{\text{[1]}}) \cdot \hat{n}$, where \hat{n} is the unit normal vector pointing from medium 1 to medium 2; similarly for $\Delta \vec{H}_{\parallel}$.

These results are particularly useful when we have reason to believe that an interface has zero free surface charge density and zero free surface current. The other boundary conditions are the same as always:³

$$\Delta(B_{\perp}) = 0 \quad \text{and} \quad \Delta(\vec{E}_{\parallel}) = 0.$$

[\[T2\]](#) Section 49.2' (page 509) mentions more sophisticated ways to think about bound charge and current.

49.3 LINEAR REGIME

Our goal was to eliminate explicit mention of bound charges and currents from the Maxwell equations, but Equations 49.6–49.7 didn't yet succeed: Together with the remaining unmodified Maxwell equations, they have *doubled* the unknown fields, adding \vec{D} and \vec{H} to \vec{E} and \vec{B} . It is true that the new quantities are determined by the old ones, but in a way that does involve the charges and currents in the medium (Equations 5.5 and 49.5). We now introduce a further level of approximation that, when justified, finishes our job in a simple way.

³See Sections 5.10 and 14.7.

49.3.1 Electric

Many dielectric media are approximately linear:⁴ That is, \vec{P} is a linear function of \vec{E} , described by the **dielectric susceptibility** $\tilde{\chi}_e$ via the response function $\vec{P} = \epsilon_0 \tilde{\chi}_e \cdot \vec{E}$. The dielectric susceptibility describes how much induced electric dipole moment you get (deformation times charge per volume) per applied electric field (force per charge). That is, it is essentially a spring constant tensor, times density and charge squared. Like any spring constant tensor, it is symmetric.⁵

For simplicity, let's assume the medium is **isotropic** (χ_e is a scalar constant). A medium can be isotropic if its constituent polarizable objects are themselves spherical (like helium atoms), or if they are arranged with random orientations (like water molecules in liquid or vapor phase). Define the **permittivity** $\epsilon = \epsilon_0(1 + \chi_e)$. Then⁶

$$\vec{D} = \epsilon \vec{E}. \quad \text{constitutive relation for uniform, linear,} \quad [5.6, \text{page 58}]$$

isotropic, lossless, nonchiral dielectric

More general forms of the constitutive relation include dissipation (complex ϵ), anisotropy ($\vec{\epsilon}$ with tensor structure), and chirality.⁷

49.3.2 Magnetic

Many magnetic media are also approximately linear;⁸ that is, \vec{M} is a linear function of \vec{B} , described by the **magnetic susceptibility**⁹ $\tilde{\chi}_m$ via the response function $\vec{M} = \mu_0^{-1} \tilde{\chi}_m \vec{B}$. Define the **permeability** $\mu = \mu_0/(1 - \tilde{\chi}_m)$. Then

$$\vec{H} = \mu^{-1} \vec{B} \quad \text{constitutive relation for uniform, linear,} \quad (49.8)$$

isotropic, lossless nonchiral magnetic material

More general forms of the constitutive relation include dissipation (complex μ), anisotropy ($\vec{\mu}$ with tensor structure), and chirality.

⁴Exceptions include piezoelectric crystals under stress, or ferroelectrics (“electrets”), which have nonzero \vec{P} in zero applied field. Also, any medium will be linear only in some regime of weak enough applied fields. For example, the orientational ordering of water molecules must eventually saturate (100% alignment) at high applied fields. Much of optics deals with media in their linear regime, but there is also a big field of “nonlinear optics.”

⁵Unlike a quadrupole tensor, however, it is not traceless.

⁶[T2]Section 42.8' (page 446) showed that a rotationally-invariant rank-2 tensor must be a constant times the identity.

⁷Chapter 50 studies anisotropy. Section 49.6 below studies chirality.

⁸Exceptions include ferromagnets, which have nonzero \vec{M} at zero applied field. Also, again any medium is only linear for sufficiently weak applied fields.

⁹We follow a convention in Feynman et al., 2010b. Sadly, for historic reasons most people instead define a different quantity χ_m by $\vec{M} = \chi_m \vec{H}$. The two descriptions are equivalent: The relation between the susceptibilities is $\tilde{\chi}_m = \chi_m/(1 + \chi_m)$.

49.3.3 Maxwell

Equations 49.6–49.7 are general. For the special case of linear media, they can be combined with Equations 5.6 and 49.8, and the boundary conditions, to form a closed system that can be solved to give all fields in terms of free charges and currents.

That is, we can *forget about the medium* if it's linear; the Gauss law Equation 49.6 retains its vacuum form, but with a modified value of the permittivity. The Ampère law Equation 49.7 also retains its vacuum form, but with a modified value of the permeability. Only the *free* charge density and flux enter these equations. You also found in Your Turn 49B that the same is true for the boundary conditions.

In particular, in a bulk isotropic medium there will be the same wave solutions as in vacuum (two transverse polarizations), except that the velocity is $(\epsilon\mu)^{-1/2}$ instead of c . For example, dielectric polarizability ($\epsilon > \epsilon_0$) leads to a slowdown, that is, to a value of the refraction index that is larger than the vacuum value of 1.¹⁰

49.3.4 Macroscopic physical realizations

Consider a medium consisting of (or containing):

- A jumble of long, thin, straight strands of wire, oriented randomly. This medium is electrically polarizable and isotropic.
- A jumble of circular rings of conductor, oriented randomly. This medium is magnetically polarizable and isotropic.

49.3.5 Remarks and further examples

The preceding section imagined macroscopic polarizable objects, which could be relevant for radio or microwave propagation, but individual molecules are also polarizable. Although the details involve quantum mechanics that lies outside the scope of these notes, nevertheless for many purposes those details can be incorporated into phenomenological values of the susceptibilities.

Note that \vec{P} and \vec{M} may arise due to processes that are not instantaneous. Nevertheless, linearity and time-translation invariance of the Maxwell equations imply the existence of single-frequency solutions. But ϵ and μ will in general be frequency-dependent, leading to **dispersion**, that is, the dependence of wave velocity on frequency. Examples:¹¹

- $\epsilon \approx 81\epsilon_0$ for water at $\omega \rightarrow 0$; it's highly polarizable. But $\epsilon \approx (4/3)^2\epsilon_0$ for water at visible frequencies; the alignment of permanent dipoles is sluggish.
- For *split* rings, each ring can act as an RC circuit, and will resonate.¹²
- Dilute plasma: $\epsilon = \epsilon_0(1 - (\frac{\omega_P}{\omega})^2)$.

¹⁰Exotic “metamaterials” exist with ϵ that is not positive in certain frequency ranges, requiring special interpretation. *Anisotropic* polarizability, for example in a crystalline material, leads to birefringence (Chapter 50).

¹¹We also saw examples of dispersion in Thomson and Rayleigh scattering (Chapter 46).

¹²See Zangwill, 2013, §18.5.6.

49.4 FRESNEL EQUATIONS, “TOTAL” INTERNAL REFLECTION, AND THE EVANESCENT WAVE

Our discussion has justified the approach to optics used in Chapter 19, and extended it to magnetically responsive media.

[Not ready yet.]

49.5 CIRCULAR BIREFRINGENCE

Section 49.3.3 argued that waves propagate in an isotropic linear medium in much the same way as in vacuum: The medium slows light down, but cannot alter its polarization. Changing the value of ϵ , μ , or both just slows the waves down.

Real materials often consist of objects, such as water molecules, that are individually far from being isotropic. Nevertheless, in liquid water many molecules are jumbled together in random orientations. The same holds for a mixture, such as a solution, and even for an amorphous solid material such as glass. In each of these materials, the overall polarizability tensors are therefore averaged over all possible rotations, and hence are proportional to the identity tensor,¹³ effectively creating an isotropic medium. So we again predict no effect on polarization of light.


The prediction just made *fails* spectacularly, however, even for everyday material like a solution of sugar in water! For example, corn syrup (essentially a concentrated glucose solution) rotates the axis of linearly polarized light in a counterclockwise direction when viewed along \vec{k} . This electromagnetic phenomenon is called **circular birefringence**.¹⁴

What property could select this direction of rotation, a choice that breaks spatial inversion invariance? Because Maxwell’s equations are themselves invariant under inversions, the only source of optical rotation must be a property of the sugar molecules themselves—one not shared by, say, water molecules.

Indeed, glucose differs from H_2O by a property called **chirality**. An object that cannot be superimposed on its mirror image by any rotation or translation is called **chiral**.¹⁵ That is, a chiral object’s very presence breaks inversion symmetry.¹⁶

The hypothesis that molecular chirality is the source of optical rotatory power predicts that molecules that are mirror-images of each other should induce rotation in opposite directions. But we still face a paradox, because the argument given at the start of this section seems to apply to an isotropic solution of any kind of molecule, chiral or not.

We must be missing something crucial. Since we calculated that an effect is zero

¹³  See Section 42.8’ (page 446).

¹⁴ Some authors use the synonym “optical activity”; the medium is said to possess “optical rotatory power.” Circular birefringence is different from ordinary birefringence, which can happen even in a nonchiral crystal of nonchiral objects (Chapter 50).

¹⁵ Objects that are not chiral are called “nonchiral” or “achiral.” The two mirror images of a chiral object are called each other’s **enantiomers**.

¹⁶ In contrast, the oxygen, nitrogen, and argon making up most of our atmosphere are achiral, and hence the polarization of the blue sky is not washed out by different rotations from the many different distances that contribute to what an observer sees.

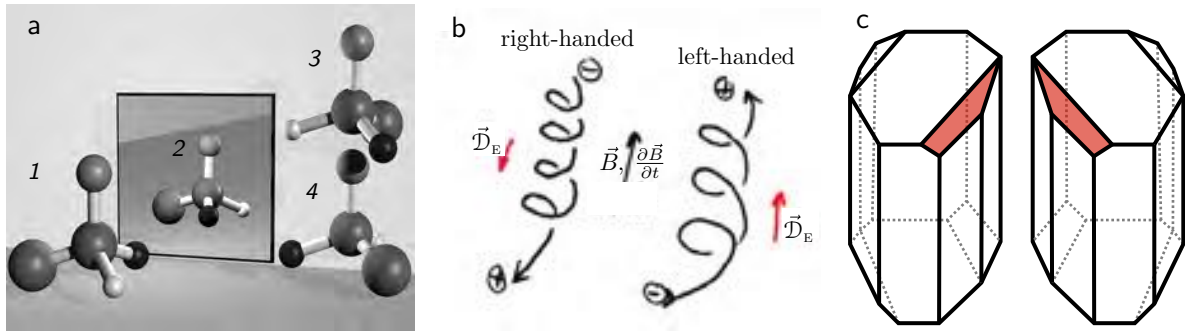


Figure 49.3: (a) A simple chiral molecule can be obtained by bonding four different atoms to a central carbon. The mirror image of 1 (shown as 2) cannot be brought into coincidence with any rotated version of 1; 3,4 show some failed attempts. (b) Helical wires as a model for chiral molecules. When $\vec{B}(t)$ is increasing in magnitude in the direction shown, the two loops of wire will experience opposite electric polarizations due to their chirality. The case of nonconstant applied $\vec{E}(t)$ involves a similar cartoon. (c) A macroscopic crystal that cannot be rotated into its mirror image.

and observed that it's not, maybe we made a bad approximation. A typical impulse is to wonder: Maybe we truncated a power series to an order at which the effect does not yet arise. But that's not the answer.

49.6 CROSS-SUSCEPTIBILITY

The resolution of our puzzle lies in another possibility that we've overlooked so far. The most general response function that is linear, homogeneous, isotropic, and lossless is:

$$\begin{pmatrix} \vec{P} \\ \vec{M} \end{pmatrix} = \begin{pmatrix} \epsilon_0 \chi_e \vec{\mathbb{1}} & ? \\ ? & \frac{1}{\mu_0} \tilde{\chi}_m \vec{\mathbb{1}} \end{pmatrix} \begin{pmatrix} \vec{E} \\ \vec{B} \end{pmatrix}.$$

That is, the constitutive relations (Equations 5.6 and 49.8) may in general have cross-terms. I'll call such terms **cross-susceptibilities**. As long as they, too, are proportional to $\vec{\mathbb{1}}$, they will still be rotationally invariant (isotropic).

49.6.1 Macroscopic physical realization

Are cross-terms of this sort really allowed? To see, let's invent another simple physical realization, along lines similar to Section 49.3.4. Consider a helix of wire open at each end (Figure 49.3b). This helix can be left- or right-handed. Its handedness has nothing to do with how it is oriented in space; for example, flipping it end-for-end does not change the handedness. In short, it is a chiral polarizable object, and that property will not be erased by rotational averaging.

Imagine a time-dependent \vec{E} field directed along the helical axis direction, with magnitude $\|\vec{E}\|$ increasing in time, so that $\partial\vec{E}/\partial t$ is parallel to \vec{E} . The applied field leads to an electric polarization \vec{D}_E as usual. Because it's time-dependent, we also get a *current* \vec{j}_z directed axially.

Your Turn 49C

Show that the helical structure also forces the current to have an azimuthal component, and hence generates a *magnetic* dipole moment: $\vec{\mathcal{D}}_M = (c\eta')(\partial\vec{E}/\partial t)$, where η' is a positive constant for the right-handed helix, or negative for the left-handed one.

Explicitly, if $\partial\vec{E}/\partial t$ points upward, positive charges flow up, regardless of the handedness of the helix. The shape of the helix then forces this charge to rotate about the axis as it moves. The direction of this azimuthal current, and hence the sign of η' , depends on the handedness of the helix.

Next, imagine a magnetic field directed along the helical axis direction with $\|\vec{B}\|$ increasing in time, so $\partial\vec{B}/\partial t$ is parallel to \vec{B} .

Your Turn 49D

Use the Faraday law to show that this field induces an EMF in the wire, creating a cylindrical current sheet that partially cancels the \vec{B} inside the coil. But the helical shape also imposes an *axial* motion of charge and hence an *electric* polarization: $\vec{\mathcal{D}}_E = -c\eta(\partial\vec{B}/\partial t)$, where η is a positive constant for the right-handed helix, or negative for the left-handed one.

That is, the constants η' and η always have the same sign. (The sign of the charge carriers is unimportant.)

Both of the arguments above are for \vec{E} and \vec{B} directed along the helical axis. But if the medium contains randomly oriented helices, then some fraction of them will have their axes along \vec{E} or \vec{B} .

49.6.2 General form

The discussion above suggests that in general a uniform, linear, isotropic, lossless, chiral medium will have¹⁷

$$\begin{bmatrix} \vec{P} \\ \check{M} \end{bmatrix} = \begin{bmatrix} \epsilon_0\chi_e & -\eta\frac{\partial}{\partial t} \\ \eta'\frac{\partial}{\partial t} & (\mu_0c^2)^{-1}\check{\chi}_m \end{bmatrix} \begin{bmatrix} \vec{E} \\ \check{B} \end{bmatrix}. \quad (49.9)$$

(Here $\check{M} = \vec{M}/c$ and $\check{B} = c\vec{B}$. These definitions simplify our formulas by giving all the entries in the matrix the same dimensions.)

Generally the constants χ_e, η, η' , and $\check{\chi}_m$ are tensors, but in isotropic medium such as aqueous solution they get replaced by their averages over orientation, that is, as 3-scalars times $\check{\mathbf{1}}$.¹⁸

As mentioned before, χ_e and $\check{\chi}_m$ may be frequency-dependent. Similarly, for disturbances at a specific frequency ω the cross-terms will be functions of frequency, both due to explicit frequency dependence of η and η' and because of the explicit time derivatives in the formula. By time-reversal invariance they must be *odd* functions, so these terms vanish at zero frequency.

¹⁷There is a slight change of notation here: Now η, η' include the density of the polarizable molecules.

¹⁸See Section 42.8' (page 446).

In the macroscopic physical realization of cross-polarization given above, you showed that η and η' have the same sign.¹⁹ Replacing the helices by their mirror images reverses the signs of both η and η' .

Thus cross-susceptibility is allowed in a medium that breaks spatial inversion invariance. That could occur either because the medium contains chiral molecules (such as most sugars, proteins, DNA, . . .), or nonchiral molecules arranged in a chiral crystal structure (such as in quartz).

49.7 THE ORIGIN OF CIRCULAR BIREFRINGENCE

Your Turn 49E

- Formulate a plane wave trial solution for the medium described by Equation 49.9. To keep things simple, you may (unrealistically) set $\chi_e = \tilde{\chi}_m = 0$, that is, neglect the ordinary susceptibilities and focus only on the cross-susceptibilities.
- Proceeding similarly to Section 50.1, show that the condition for a plane-wave solution simplifies if we expand the polarization vector in the *circular* polarization basis (helicity basis) $\vec{\zeta}_{(\pm)} = (\hat{x} \pm i\hat{y})/\sqrt{2}$.
- Show that *each circular polarization gets a different phase velocity*.

The two wave speeds you found can as usual be expressed as indices of refraction, c/n_{\pm} , explaining the term “”.

Similarly to Section 50.1, we can now ask what happens to an arbitrary linear combination of the two circularly-polarized eigenmodes of propagation. Specifically, if we feed in a *linearly* polarized plane wave, its frequency will not change, by time-translation invariance. Once the wave enters the medium, however, each circularly polarized component propagates with a different wavenumber $k_{(\pm)}$ (the two values you found in Your Turn 49E). After the wave reemerges into vacuum at the other end of a slab of medium, we can reassemble the two components and interpret the resultant.

Your Turn 49F

- Try this, and show that the result is again linearly polarized but in a direction rotated relative to the original. (This observation explains the term optical rotatory power.)
- Show that the angle of rotation is proportional both to $n_+ - n_-$ and to the thickness of the slab.
- In particular, show that $n_+ - n_-$ is proportional to the density of chiral polarizable objects (for example, concentration of a solution).

Your last result can be stated as saying that the total rotation depends on the “chiral optical depth,” that is, a constant characterizing the chiral molecule in question times the projected areal density of those molecules encountered by the light during its passage.

¹⁹ [T2] Indeed, Onsager reciprocity implies that $\eta' = \eta$. See Landau et al., 1984, Eq. 103.10. (Note that Landau uses spatial derivatives, but these can be converted to time derivatives by using Maxwell equations.)

In short:

- Cross-susceptibility is physically possible in a chiral medium, even if the medium is isotropic. The time derivatives in Equation 49.9 predict that it will be strongly dependent on wavelength. It gives rise to circular birefringence = optical activity = optical rotatory power.
- The macroscopic realization of these ideas in Section 49.6.1 involved a disordered sample of helical wires. Perhaps it seems plausible that the conclusions would extend to obviously helical molecules like DNA as well. Actually, however, the result is of far wider validity: *Any* chiral molecule, whether or not it looks helical, can give rise to optical activity. For example, we could take CH_4 and substitute the 3 hydrogen atoms with distinct things (maybe an OH group for one, a Cl atom for another, and a chain for the 3rd). Even if each group is itself nonchiral, the whole thing will break spatial inversion invariance (Figure 49.3a).
- However, air (O_2, N_2) or H_2O *won't* display this phenomenon—They are all disordered arrangements of nonchiral (inversion-invariant) objects.

49.8 DEMO

To follow up on those questions, we observed a beaker of corn syrup (concentrated sugar solution), and illuminated it with polarizers fixed above and below the dish. We noticed different colors based on the orientation of the polarizer. We did not observe this effect for H_2O or glass.

With a thicker layer of syrup, a greater rotation of the second polarizer relative to the first was required to obtain the same transmission of light.

Blue light rotates more than red. Had we diluted the solution by adding more H_2O to it, the total optical thickness would have gone up but the total projected density of sugar molecules/area would not; empirically, one indeed finds that the total polarization rotation doesn't change.

49.9 CODA

- Remarkably, Fresnel showed in 1825 (long before Maxwell) that a difference in refractive index for left- and right-circularly polarized light would lead to the polarization rotation observed in optically active liquids, and he predicted that merely letting a beam of unpolarized light enter at an angle into such a medium would separate it into circularly polarized beams.
- Louis Pasteur intuited the connection between chiral molecules and optical activity in 1849, also long before Maxwell's equations, just by thinking about symmetry. Pasteur crystallized synthetic tartaric acid and noticed that the tiny crystals came in two mirror image forms (Figure 49.3c). He painstakingly sorted a pile of these tiny crystals into two piles, in this way manually purifying the two enantiomers. Dissolving each one in water then yielded two solutions with opposite optical activity!

- Living organisms discriminate between the two enantiomers of each biomolecule, and only synthesize the one they need. In contrast, most artificial synthesis techniques make both enantiomers indiscriminately (they create a “racemic mixture”). Similarly, most purification techniques are also unable to separate enantiomers (apart from Pasteur’s heroic effort).²⁰ Thus, the presence of optical activity can in principle distinguish artificial from synthetic compounds, a circumstance that provided the crucial plot element in (at least one) novel from the classical era of British murder mysteries. But maybe I had better not tell you which one.
- A computational chemist might now proceed to formulate what quantum mechanical calculation to do to go from molecular structure to a prediction of the value of η . The calculation is long and hard, and in the end you have to discard most of your work by averaging over random orientations. And to a physicist it’s not so interesting—what’s interesting is how symmetry says there’s *just one* phenomenological parameter η characterizing the effect of chirality of an isotropic medium on light (to leading nontrivial order in frequency).
- The math predicts that optical rotatory power goes to zero at zero frequency, due to the time derivative in the constitutive relation (Equation 49.9). More generally, the entire spectrum of optical rotatory power is called the *optical rotatory dispersion*, and it forms a fingerprint of the constituent molecules, independent of the ordinary dispersion. ORD is a convenient thing to measure because the uninteresting water molecules in a solution don’t contribute to it.²¹
- There can also be chiral dissipation (“friction”) terms, leading to different absorption lengths for each helicity (each choice of $\hat{e}_{(\pm)}$). The entire spectrum of the differential absorption is called the material’s **circular dichroism** spectrum, yet another fingerprint of a molecule that can be observed in solution. An unexpected structural form of the DNA molecule called “Z-DNA” was first discovered via its nonstandard CD spectrum.
- A similar phenomenon can occur in an astrophysical plasma, if a uniform \vec{B} field is present. Although this is a very different kind of medium from sugar solution, nevertheless it breaks inversion symmetry in a way that is mathematically similar to what we have studied, again leading to different phase velocities for the two circular polarizations. Synchrotron radiation from an accretion disk is polarized, and so this rotation can be used to disclose strong magnetic fields.

49.10 PLUS ULTRA

Was it worth the effort? I’d like to suggest that it’s not much of an exaggeration to say that this story illustrates in miniature *how physicists think about nearly everything*. We saw the possibility of a surprising new coupling, we characterized it in terms

²⁰Even mass spectrometry cannot separate them, because they have the same charge/mass ratio.

²¹Although the polarization rotation angle is ambiguous by 180° , its differential rate of increase as depth increases is well defined. The (rotation angle)/(depth \times concentration) as a function of frequency is what characterizes the solute. Chemists sometimes use the baroque unit $\text{M}^{-1}\text{dm}^{-1}$ for this quantity; you should convince yourself that it has the same dimensions as area, and indeed is in some sense a scattering cross-section.

of symmetry, we looked for what sort of physical setup had the required (lack of) symmetry, we looked for how the math could transmit the key property from the physical setup to observable, quantitative predictions. Then we made the observation.

Again: The molecule shown in Figure 49.3a does not have any obviously helical character. After arguing that the structures in panel (a) would exhibit cross-susceptibility, we still might have no intuition about the molecule in panel (b). But from the symmetry viewpoint, they are the same: Both lack invariance under spatial reflections, even when averaged over orientations. And that invariance is the only thing that forbids cross-susceptibility, and its symptom optical activity. So we expect it with any chiral molecule—and there it is.

But... is it beautiful? Section 32.2.4 claimed that, to a physicist, “beauty” often means the combined effect of *inevitability* and *surprise*. So—I’d say yes. And then when you see the *colors*—that’s another level of beauty.

FURTHER READING

Historical: https://en.wikipedia.org/wiki/Louis_Pasteur#Molecular_asymmetry.

Physical realization of a cross-polarizable material: See Feynman et al., 2010a, §33-5; Hecht, 2002, §8.10.1; Nieves & Pal, 1994.

Experimental observation of double diffraction from optically active liquids: Ghosh et al., 2007.

Liquid crystals: de Vries & IUCr, 1951.

T₂ Quantum mechanical treatment of optical activity of molecules: Cantor & Schimmel, 1980, ch. 8; Craig & Thirunamachandran, 1998, ch. 8; Caldwell & Eyring, 1971.

T₂ Relativistic treatment of media: Landau et al., 1984, §76.

T₂ Track 2

49.2'

Our pictorial approach to bound charge and current summarizes the results of an analysis that is really only valid for a restricted class of materials, such as dilute gases, nonpolar liquids, and molecular solids with weak interactions between the molecules. If we want to predict bulk material parameters from microscopic details in such situations, we can make a multipole expansion of the fields from each constituent, spatially average over length scales relevant to the problem (but much longer than the size of the constituents), then find the effective continuous charge density and flux that *could have* given rise to the same fields.

For many materials, quantum-mechanical couplings between constituents invalidate this simple approach. A more general approach appears in Zangwill, 2013, ch. 6 and 13. However, this chapter's concerns were restricted to understanding general properties of linear response; the heuristic approach we gave motivated general formulas allowed by principles such as rotational, time inversion, and (when appropriate) spatial inversion invariance.

T₂ Track 2

49.2.1' Dissipation and frequency dependence

Suppose that an electric field varies harmonically in time: $\vec{E}(t) = \frac{1}{2}\vec{E}e^{-i\omega t} + \text{c.c.}$ In a medium that is itself time-translation invariant, we will then find that the displacement $\vec{D}(t) = \frac{1}{2}\vec{D}e^{-i\omega t} + \text{c.c.}$ If the medium is linear, then we will have

$$\vec{D} = \epsilon(\omega)\vec{E},$$

which defines the frequency-dependent permittivity function. We have tacitly assumed that ϵ is real, but this need not be the case if there is dissipation.

To understand complex permittivity, imagine a material consisting of polarizable “molecules” with density ρ_{mol} , consisting of a pair of charges $\pm q$ that can separate by Δx . Let $\chi_e(\omega) = (\epsilon(\omega)/\epsilon_0) - 1$ as usual. In response to the field, charge will separate by $\Delta x = \frac{1}{2}\overline{\Delta x}e^{-i\omega t} + \text{c.c.}$

The density of induced dipole moment is then $P = \rho_{\text{mol}}q\Delta x$. That result lets us find the velocity $v(t) = \frac{1}{2}\overline{v}e^{-i\omega t} + \text{c.c.}$, where

$$\overline{v} = \frac{-i\omega\epsilon_0\chi_e(\omega)\overline{E}}{q\rho_{\text{mol}}}.$$

The rate at which the field does work on the particle is qE times v , or

$$qEv = q\left(\frac{1}{2}\overline{E}e^{-i\omega t} + \text{c.c.}\right)\left(\frac{-i\omega\epsilon_0\chi_e\overline{E}}{2\rho_{\text{dip}}}e^{-i\omega t} + \text{c.c.}\right).$$

The time average of that power, per volume is thus

$$\frac{1}{4}(-i)\omega\epsilon_0\chi_e|\overline{E}|^2 + \frac{1}{4}(i)\omega\epsilon_0\chi_e^*|\overline{E}|^2 = \frac{1}{2}\omega\epsilon_0|\overline{E}|^2 \text{Im } \chi_e.$$

As claimed, if the frequency-dependent permittivity function is complex then the material dissipates energy (into heat). Similar remarks apply for the frequency-dependent magnetic permeability.

T₂ Track 2

49.6'a Just two enantiomers

Why are there just two of a chiral molecule? The point is that electromagnetism, including its quantum version, is invariant under the group $O(3)$ of orthogonal 3×3 matrix transformations of space. Any two molecules related by such a transformation will have the same energy, stability, excited states, etc. And this group is *twice as big* as the rotation group $SO(3)$: The coset space of $O(3)$ matrices modulo all rotations is just the group \mathbb{Z}_2 with two elements.

49.6'b Relativistic formulation

Equation 49.9 (page 504) involves a 6×6 matrix of susceptibilities, which is not obviously a 4-tensor. But in fact, we can define a response 4-tensor \underline{R} analogously to $\underline{F}^{\mu\nu}$, as

$$\underline{R}^{\mu\nu} = \begin{bmatrix} 0 & \vec{P}_x & \vec{P}_y & \vec{P}_z \\ -\vec{P}_x & 0 & -\vec{M}_z & \vec{M}_y \\ -\vec{P}_y & \vec{M}_z & 0 & -\vec{M}_x \\ -\vec{P}_z & -\vec{M}_y & \vec{M}_x & 0 \end{bmatrix}^{\mu\nu}. \quad (49.10)$$

where again $\vec{M}_i = c^{-1}\vec{M}_i$. This big formula can be summarized in the usual way by $\underline{R}^{0i} = -\underline{R}^{i0} = \vec{P}_i$ and $\underline{R}^{ij} = -\varepsilon_{ijk}\vec{M}_k/c$. Also, let \underline{J}_f denote the free charge flux 4-vector field.

In terms of these definitions, four of the Maxwell equations take the form

$$\partial_\mu \underline{H}^{\nu\mu} = c^{-1} \underline{J}_f^\nu, \quad (49.11)$$

where

$$\underline{H}^{\nu\mu} = c\varepsilon_0 \underline{F}^{\nu\mu} + \underline{R}^{\nu\mu}. \quad (49.12)$$

Thus $\underline{H}^{0m} = \vec{D}_m$ and $\underline{H}^{nm} = c^{-1}\varepsilon_{nm\ell}\vec{H}_\ell$, in parallel to the naming of elements of \underline{F} . We conclude that \underline{R} must be a tensor because the world is Lorentz invariant, and Equations 49.11–49.12 are only invariant if \underline{R} is a tensor.

The remaining four Maxwell equations are unchanged from the case of vacuum, because they have no source terms.

Linear response is the statement that \underline{R} is a linear function of \underline{F} :

$$\underline{R}^{\mu\nu} = \underline{K}^{\mu\nu}{}_{\lambda\sigma} \underline{F}^{\lambda\sigma}, \quad (49.13)$$

where the **susceptibility operator** \underline{K} is antisymmetric on its first two indices, and also on the last two.

Let's apply "Einstein thinking" to see what are the possibilities for the susceptibility 4-tensor. We know that \underline{R} and \underline{F} are 4-tensors, so Equation 49.13 implies that \underline{K} is a 4-tensor operator. Even an isotropic medium breaks Lorentz symmetry—unlike the vacuum, it can have states of motion. But isotropy and homogeneity do imply that the *only* quantity describing the state of the medium is its 4-velocity \underline{U} . Hence it must be possible to express \underline{K} as a combination of \underline{U} 's and invariant quantities describing the medium. \underline{K} must also be a symmetric operator in the sense that exchanging $\mu\nu$ with $\lambda\sigma$, and $\underline{\partial} \rightarrow -\underline{\partial}$, must leave it unchanged. Playing around shows that there are only three possible forms permitted by the

symmetries:²²

$$\begin{aligned} \underline{K}^{\mu\nu}{}_{\lambda\sigma} = & \frac{\alpha}{2} (\delta_{\lambda}^{\mu} \delta_{\sigma}^{\nu} - \delta_{\lambda}^{\nu} \delta_{\sigma}^{\mu}) + \frac{\tau}{2} (\underline{U}^{\mu} \underline{U}_{\sigma} \delta_{\lambda}^{\nu} - \underline{U}^{\nu} \underline{U}_{\sigma} \delta_{\lambda}^{\mu} - \underline{U}^{\mu} \underline{U}_{\lambda} \delta_{\sigma}^{\nu} + \underline{U}^{\nu} \underline{U}_{\lambda} \delta_{\sigma}^{\mu}) \\ & + \frac{\gamma}{2} (\varepsilon^{\mu\nu}{}_{\tau\lambda} \underline{U}^{\tau} \underline{U}_{\sigma} - \varepsilon^{\mu\nu}{}_{\tau\sigma} \underline{U}^{\tau} \underline{U}_{\lambda} - \varepsilon_{\lambda\sigma\tau}{}^{\mu} \underline{U}^{\tau} \underline{U}^{\nu} + \varepsilon_{\lambda\sigma\tau}{}^{\nu} \underline{U}^{\tau} \underline{U}^{\mu}) \underline{U}^{\rho} \partial_{\rho}. \end{aligned} \quad (49.14)$$

Here the 4-dimensional Levi-Civita pseudotensor has $\varepsilon_{0123} = +1$ etc.

Specialize this formula to an inertial coordinate system in which the medium is at rest, and show that the constants α , β , and γ can be chosen so that it reproduces Equation 49.9 (which also has three phenomenological parameters χ_e , $\tilde{\chi}_m$, and η).

Then substituting arbitrary 4-velocity at once tells us the appropriate form of the susceptibility tensor in a *moving* medium.²³

Every term in Equation 49.14 must be time-reversal invariant, because a static collection of molecules does not break time-reversal invariance.²⁴ (This is why the γ term needs a derivative.) Also, the α and β terms are invariant under spatial inversions—but not the γ term. Thus, γ must equal zero for an achiral medium, as we observed with liquid water in the demo.

²²More precisely, this is the most general structure to leading order in powers of derivatives. The logic is similar to what we've done before, e.g. in Section 34.5 (page 385). Some terms that may seem missing from our list are in fact redundant because of Maxwell's equations and the constraint that $\underline{U}_{\mu} \underline{U}^{\mu} = -c^2$.

²³You previously used similar logic in Problem 33.1 (page 380).

²⁴Ferromagnetism was not allowed.

PROBLEMS

49.1 *Electrorotation of cells*

[Not ready yet.]

49.2

Repeat Your Turn 49E, but this time without the unrealistic simplifying assumptions $\chi_e = \tilde{\chi}_m = 0$.

49.3 *Bulk conductor, II*

A stationary (time-independent) current distribution is established in a medium which is isotropic but not necessarily homogeneous. For example, that medium could be animal tissue.

Specifically, the charge flux \vec{j} is everywhere a scalar multiplier times $-\vec{\nabla}\psi$, but that multiplier (the conductivity κ) is not necessarily the same everywhere. However, you may assume that the dielectric constant ϵ/ϵ_0 is uniform and isotropic.

- Show that the medium will in general acquire a nonzero free electric charge density $\rho_{q,f}(\vec{r})$. Show that this charge density may be written as the dot product of $\vec{\nabla}\psi$ with a certain vector field, and find that vector field.
- Repeat for the case where ϵ is also nonuniform, though isotropic.

49.4 *Polarization of evanescent wave*

Preamble: Dr. Beausang discussed an experimental technique called “pol-TIRF,” for “polarized total internal reflection fluorescence microscopy.” The essential points were:

- . TIRF excitation improves-signal-to noise in fluorescence microscopy by only creating electric fields in a thin layer next to the floor of the experimental chamber.
- . These electric fields retain information about the polarization of the laser beam that gives rise to them, a fact that can be used to learn about the orientation of a single fluorescent molecule in the sample.

The first point is discussed in our text. We’d like to look closer into the second point.

A linearly polarized, monochromatic wave of frequency ω enters a sample chamber filled with water (index of refraction $n_2 \approx 1.33$) from a medium with larger index n_1 (typically quartz, ≈ 1.46). For this problem you may assume that the permeabilities are equal: $\mu_1 = \mu_2$.

The interface between media is the yz plane. The incoming wave (in the region $x < 0$) has wavevector \vec{k} lying in the xy plane; all fields are independent of z . The incoming \vec{k} makes angle θ with the normal to the interface, that is, $\vec{k} \cdot \hat{x} = \cos\theta$. We’ll eventually consider the case where the angle of incidence θ is large, but you should first work out the answers for arbitrary θ , then specialize to large θ .

It’s convenient to choose the following basis vectors for the incoming polarization:

- . “TE” polarization (also called “s-wave”): \vec{E} is parallel to \hat{z} .
- . “TM” polarization (also called “p-wave”): \vec{B} is parallel to \hat{z} .

Review P+S §13.2.1 for the definition of the critical angle θ_c , the transmitted wavevector \vec{k}' , and the reflected wavevector \vec{k}'' . Write the incident wave as

$$\vec{E}(t, x, y, z) = \frac{1}{2} \left[\vec{E} e^{i(\vec{k} \cdot \vec{r} - \omega t)} + \text{c.c.} \right] \quad x < 0.$$

Here \vec{E} is the incoming polarization vector. The transmitted and reflected waves are given by similar expressions with \vec{E}' , \vec{k}' , etc.; they all have the *same* value of ω .²⁵ Continue reading through §12.2.3 to see the method to find the transmitted and reflected waves.

The 3-vector \vec{E}' describes the amplitude, phase, and polarization of the transmitted wave. We want to know the polarization, particularly in the case where the transmitted wave is nonpropagating.

Problem:

- Consider a quartz-water interface and laser wavelength in vacuum 514 nm. Find the critical angle. Find the exponential amplitude falloff length scale, assuming $\theta = 70^\circ$.
- Find the amplitude and direction of the electric field²⁶ for $x > 0$, in the case of TE incident polarization. That is, suppose $\vec{E}' = \bar{E}\hat{z}$ where \bar{E} is a real constant. Then specialize to the case with $\theta > \theta_c$. Characterize in words the *type* of polarization you get for the evanescent electric field. Then substitute the numbers in (a) to get a quantitative characterization.
- Repeat for the TM polarization. Again characterize in words the *type* of polarization obtained, then substitute the numbers in (a) to get a quantitative characterization.

49.5 Relativistic formulation

Use Equations 49.11–49.12, 49.13, and 49.14 to derive the plane wave solutions for light in flowing water, relevant to the Fizeau experiment. You can also seek solutions corresponding to light propagating in an isotropic, chiral medium, such as sugar water, at rest.

²⁵One prime for transmitted, two primes for reflected.

²⁶We are not interested in any overall phase shift.

CHAPTER 50

Anisotropic Media

50.1 ORDINARY BIREFRINGENCE

The preceding chapter showed that circular birefringence was possible in an isotropic medium, due to electric/magnetic crossterms in the polarizability. We now consider a more common situation, a homogeneous medium in which magnetic effects are negligible but the polarizability is not isotropic. For example, a material may consist of molecules held in a crystal lattice, so that their polarizability is not averaged over rotations. Typical transparent, crystalline solids include quartz and many kinds of hard clear plastic. Even liquid crystals can have at least partial orientational order. The ensuing behavior of light is called **ordinary birefringence**, or more commonly just “birefringence.”



Figure 50.1: Zooplankton under parallel and crossed polarizers. Left to right: *Cyclosalpa floridana* (about 5 cm across); *Bolinopsis* sp. (about 5 cm long); *Salpa cylindrica* (about 2.5 cm long). “The photos were taken by Edith A. Widder, on a project that we were doing together on a ship.” – Sönke Johnsen

Thus we will consider a medium with $\vec{\chi}_e$ constant but not scalar and $\tilde{\chi}_m \approx 0$. Like any real symmetric matrix, $\vec{\chi}_e$ has a basis of three mutually perpendicular, real eigenvectors. Choose coordinates such that those eigenvectors are \hat{x} , \hat{y} , and \hat{z} and consider a trial solution that is a plane wave propagating along \hat{z} :

$$\vec{E}(t, \vec{r}) = \frac{1}{2} \vec{\zeta} e^{-i(\omega t - \vec{k} \cdot \vec{r})} + \text{c.c.} \quad \vec{B}(t, \vec{r}) = \frac{1}{2} \vec{\beta} e^{-i(\omega t - \vec{k} \cdot \vec{r})} + \text{c.c.}$$

Maxwell's equations¹ then say

$$i\vec{k} \cdot \vec{\epsilon} \cdot \vec{\zeta} = 0, \quad i\vec{k} \cdot \vec{\beta} = 0, \quad (50.1)$$

$$i\vec{k} \times \vec{\zeta} + (-i\omega)\vec{\beta} = 0, \quad (50.2)$$

$$i\vec{k} \times \vec{\beta} / \mu_0 - (-i\omega)\vec{\epsilon} \cdot \vec{\zeta} = 0. \quad (50.3)$$

Equation 50.1 tells us that $\vec{\zeta}$ and $\vec{\beta}$ must both be perpendicular to \vec{k} . Equation 50.2 tells us $\vec{\beta}$ in terms of $\vec{\zeta}$. Substituting into the last equation gives the dispersion relation, which is simple if $\vec{\zeta}$ is directed along either of the two transverse eigenvectors of the permittivity:

$$k = \omega \sqrt{\epsilon_{(\alpha)} \mu_0}, \quad \alpha = 1 \text{ or } 2.$$

Here $\epsilon_{(\alpha)}$ denotes one of the eigenvalues.

In other words, in this simple situation light propagating along a principal axis, and polarized along another principal axis, is transmitted without change. However, *those two linear polarizations propagate at different speeds* (phase velocities). We'll call those speeds $c/n_{(\alpha)}$ where $n_{(\alpha)} = \sqrt{\epsilon_{(\alpha)}/\epsilon_0}$.

50.1.1 Half-wave plate

We can now ask, what happens to a *mixture* of those two polarizations? In particular, consider a slab of this medium whose thickness z_{tot} is such that²

$$\omega(n_{(1)} - n_{(2)})z_{\text{tot}}/c = \pi. \quad \text{half-wave plate}$$

Now consider light that at time zero, and $z = 0$, is linearly polarized at some arbitrary angle to the x axis: $\vec{\zeta} = \hat{x} \cos \alpha + \hat{y} \sin \alpha$. How will it look when it emerges a distance z_{tot} from its entry point?

We may solve Maxwell's equations for each eigenvector component separately, then superpose the answers. Let $\bar{n} = (n_{(1)} + n_{(2)})/2$. From previous paragraphs, then,

$$\begin{aligned} \vec{E}(t, z) &= \frac{1}{2} e^{-i\omega(t - \bar{n}z_{\text{tot}}/c)} \left[\hat{x} e^{i\omega \Delta n z_{\text{tot}}/(2c)} \cos \alpha + \hat{y} e^{-i\omega \Delta n z_{\text{tot}}/(2c)} \sin \alpha \right] + \text{c.c.} \\ &= \sin(\omega(t - \bar{n}z_{\text{tot}}/c)) \left[\hat{x} \cos \alpha - \hat{y} \sin \alpha \right]. \end{aligned}$$

Once again, we find that if the wave enters polarized along \hat{x} or \hat{y} , it leaves in the same state. But in intermediate cases, its polarization gets rotated by an angle somewhere between zero and 90° . More precisely, it is *reflected* through a plane (Figure 50.2). We can arrange for whatever rotation angle we like by twisting the half-wave plate relative to the polarization of incoming light, making it a useful gadget for optical setups.

¹For example Equations 49.6–49.7 (page 498), with the constitutive relation Equation 5.6 (page 58).

²For simplicity, we suppose that the surfaces of the medium are perpendicular to \vec{k} , and hence to one of the eigenvectors of the material, so that there is no refraction.

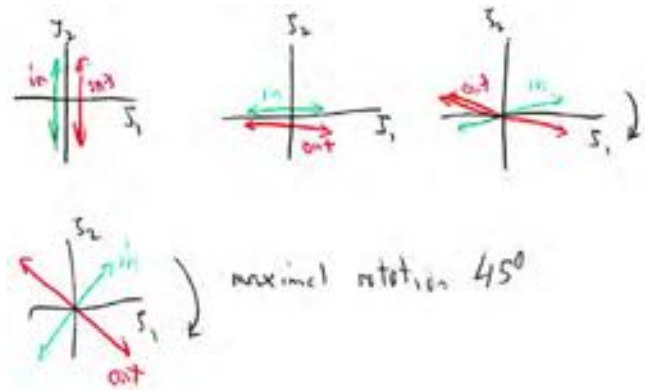


Figure 50.2: Change of polarization after passage through a half-wave plate, for various kinds of incident light. ζ_1 and ζ_2 are components of the polarization expanded in eigenvectors of the polarizability tensor; the third eigenvector points out of the page and coincides with the propagation direction.

50.1.2 Linear dichroism

A dense material in an electromagnetic wave can have dissipative losses that also depend on the polarization of the light. This means that different polarizations can be differentially absorbed.

Edwin Land experimented with polyvinyl alcohol chains aligned on plastic substrate. When the material is heated or stretched, the chains become electrically conducting, creating large polarizability in one direction. This comes along with dissipation (Chapter 45), and so Land's "polaroid filter" effectively blocked EM radiation with one linear polarization, much like the microwave polarizer shown in a class demo. Polaroid filters are not the only way to obtain polarized light, but they were much cheaper and more convenient than the alternatives available at that time.

50.2 OPTICAL TORQUE WRENCH

[Not ready yet.]

50.3 PLUS ULTRA

[Not ready yet.] Some animals have evolved wing scales that reflect sunlight preferentially in one circular polarization (Figure 50.3).

FURTHER READING

Optical torque wrench: La Porta & Wang, 2004.

Fig. 1. Photographs of the beetle *C. gloriosa*. (A) The bright green color, with silver stripes as seen in unpolarized light or with a left circular polarizer. (B) The green color is mostly lost when seen with a right circular polarizer.

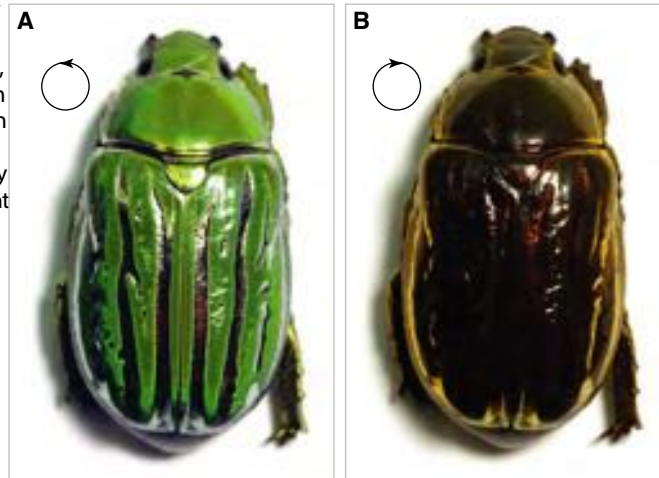


Figure 50.3: Sharma et al., 2009.

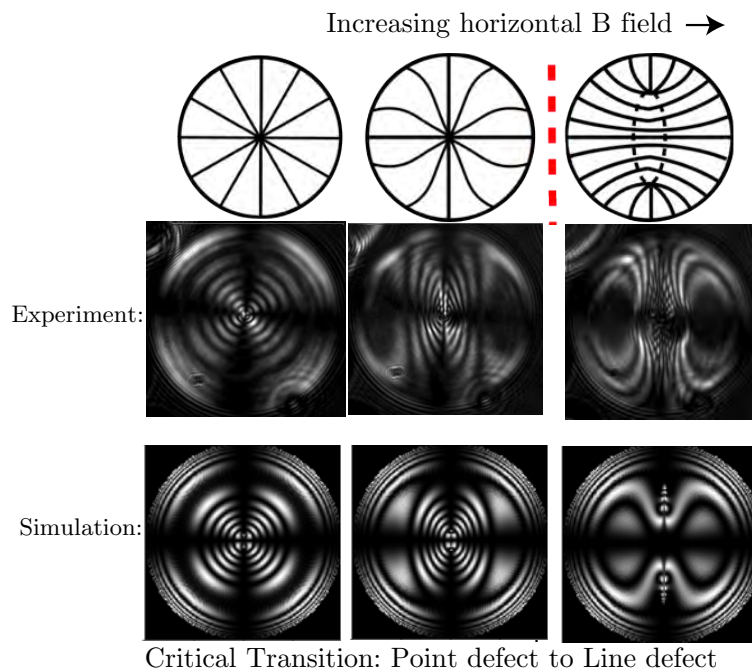


Figure 50.4: [Courtesy Sophie Ettinger, Dan Beller, and Arjun Yodh.]

T_2 Track 2

50.3' Magnetic anisotropy

“Nematic liquid crystal materials generally contain molecules that possess permanent dipole moments and that also exhibit an anisotropic electronic polarizability and an anisotropic diamagnetic susceptibility.” – Smith et al., 2007

PROBLEMS

50.1 *Quarterwave plate*

A quarterwave plate is a slab of birefringent dielectric material that has been cut so that its planar faces are perpendicular to one of its three principal axes. Light enters along that direction. The thickness of the slab is chosen such that, at a particular wavelength, the transit times for light linearly polarized along the other two principal axes differ by $1/4$ of a wave period.

Suppose that a linearly polarized plane wave enters this material, with polarization vector oriented *midway between* the two principal directions. What sort of wave emerges from the other side? Write a short formula to justify your answer.

50.2 *Circular polarizer*

[Not ready yet.]

50.3 *Optical torque wrench*

[Not ready yet.]

CHAPTER 51

Čerenkov radiation

51.1 FRAMING

When we think of the generation of radiation, we generally envision a charge that is shaking, braking, circulating, or otherwise accelerating. So it may come as a surprise to find that a charged particle in *uniform, straight-line motion* can generate radiation! How can that happen?

51.2 CHARGED PARTICLE IN VACUUM

To begin to answer, we will first revisit a problem we solved already in Section 32.3.2:¹ The fields created by a charge in uniform, straight-line motion in vacuum. Reassuringly, we'll see that we did it correctly the first time; there is no radiation. But the geometric approach we use here can be generalized to include a dielectric medium, such as water. Interesting and unexpected behavior will then appear in Problem 51.1.

Suppose that a point charge q moves along the z axis at speed βc . Thus, its trajectory can be written as $\vec{r}_q(t) = \beta ct \hat{z}$, and

$$\rho_q(t_*, \vec{r}_*) = q \delta^{(3)}(\vec{r}_* - \beta ct_* \hat{z}) \quad (51.1)$$

$$\vec{j}(t_*, \vec{r}_*) = q \beta c \hat{z} \delta^{(3)}(\vec{r}_* - \beta ct_* \hat{z}). \quad (51.2)$$

The Green function solution to Maxwell's equations in Lorenz gauge then gives²

$$\psi(t, \vec{r}) = \frac{1}{4\pi\epsilon_0} \int d^3r_* \|\vec{r} - \vec{r}_*\|^{-1} \rho_q(\underbrace{t - \|\vec{r} - \vec{r}_*\|/c}_{t_*}, \vec{r}_*),$$

and similarly for the vector potential. Substitute the expression in the brace for t_* in Equation 51.1:

$$\psi(t, \vec{r}) = \frac{1}{4\pi\epsilon_0} \int d^3r_* \|\vec{r} - \vec{r}_*\|^{-1} \delta^{(3)}(\vec{r}_* - \beta c \hat{z} (t - \|\vec{r} - \vec{r}_*\|/c)). \quad (51.3)$$

The three delta functions will eliminate the three integrals, but it's a bit tricky because \vec{r}_* appears in two places.

Let $\vec{w}(\vec{r}_*; \vec{r}, t)$ denote the argument of the delta functions. Then

$$\frac{\partial \vec{w}_i}{\partial \vec{r}_{*j}} = \delta_{ij} - \beta \delta_{i3} \frac{-1}{\|\vec{r} - \vec{r}_*\|} (\vec{r} - \vec{r}_*)_j (-1).$$

¹Also see P+S §15.6.1.

²See Chapter 40.

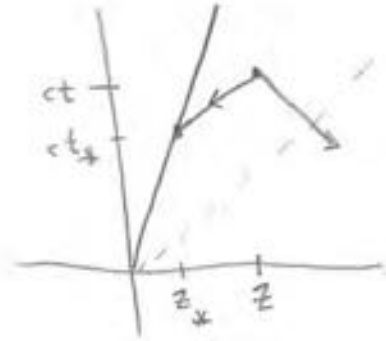


Figure 51.1: Graphical solutions of Equation 51.5. The special case in which the field point \vec{r} lies on the \hat{z} axis. Clearly, one of the two past-directed, light-speed lines from (ct, z) hits the particle trajectory exactly once, at (ct_*, z_*) ; the other such line misses the trajectory altogether.

Let $\vec{R} = \vec{r} - \vec{r}_*$ as usual. Then the rule for delta functions (Section 0.3.6) simplifies Equation 51.3 to

$$\psi(t, \vec{r}) = \frac{q}{4\pi\epsilon_0} \sum_{\ell} \left| \det \frac{\partial \vec{w}_i}{\partial \vec{r}_j} \right|^{-1} \|\vec{r} - \vec{r}_{*\ell}\|^{-1},$$

where the sum is over all the solutions $\vec{r}_{*\ell}$ to the equation $\vec{w}(\vec{r}_{*\ell}; r, t) = \vec{0}$.

Next note that the 3×3 determinant equals $1 - \beta \vec{R}_3/R$, so

$$\psi(t, \vec{r}) = \frac{q}{4\pi\epsilon_0} \sum_{\ell} |R - \beta \vec{R}_3|^{-1}. \quad (51.4)$$

To evaluate this expression, we need all the solutions to $\vec{w} = 0$. These are points \vec{r}_* with the properties that:

- The particle was there at some time t_* , and hence $\vec{r}_* = \beta c \hat{z} t_*$.
- The distance between that location and the observer at \vec{r} can exactly be covered by moving at speed c for time $t - t_*$.

That is, $R = c(t - t_*)$, or

$$\begin{aligned} \frac{R}{c} \beta c \hat{z} - M b c \hat{z} (t - t_*) &= 0 \\ R \beta \hat{z} - \beta c \hat{z} t + \vec{r}_* &= 0. \end{aligned} \quad (51.5)$$

In fact, the last equation always has exactly one solution, at a point we will call \mathbf{Q}_r . Figure 51.1 is a familiar spacetime diagram that establishes this claim in a special case, where the observer is sitting on the z axis. Even when that is not the case, we can use rotation invariance to choose coordinates for which $x = 0$ (though y need not be zero). Figure 51.2a then shows that for any point \mathbf{P} in the yz plane, exactly one of the circles drawn intersects \mathbf{P} .

It is worthwhile to give one more proof of the point just made. Again suppose that we have been given a choice of field point \mathbf{P} and observation time t . Figure 51.2b again shows a particular choice, along with the charged particle's position \mathbf{Q} at t . The given information determines the angle θ between the line $\overline{\mathbf{QP}}$ and the z axis. Extend this line beyond \mathbf{P} (dotted line on the figure).

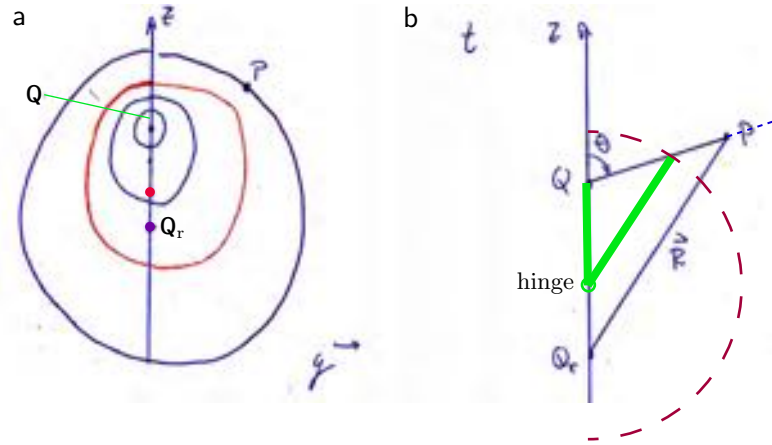


Figure 51.2: (a) Circles centered on $\beta c\hat{z}t_*$, of radii $c(t - t_*)$, for various choices of time t_* . A field point \mathbf{P} has been selected; exactly one of the nested circles shown passes through that point. (b) Two sticks (*green*) joined by a hinge. The configuration shown can be magnified, holding one endpoint at \mathbf{Q} , until its other endpoint passes through \mathbf{P} . The position of the hinge then determines \mathbf{Q}_r .

The figure also shows another point, \mathbf{Q}_r , which is the charge's position at an earlier time t_* . Thus, the distance $\overline{\mathbf{Q}\mathbf{Q}_r}$ equals $\beta c(t - t_*)$. The claim is that we may choose \mathbf{Q}_r such that also the distance $\overline{\mathbf{Q}_r\mathbf{P}}$ equals $c(t - t_*)$. To see this, imagine two sticks joined by a hinge. The ratio of the sticks' lengths is β . Place the free end of the shorter stick at \mathbf{Q} , and align it along the z axis. Hold the short stick in place and pivot the long stick about the hinge point. The long stick's end then sweeps out a circle (dashed in the figure), which intersects $\overline{\mathbf{Q}\mathbf{P}}$ in exactly one point. Now rescale both sticks by whatever overall factor is needed to make the long stick's endpoint pass exactly through \mathbf{P} while maintaining the length ratio β . There is always just one way to do this.

Your Turn 51A

Figures 51.1–51.2 were all drawn assuming that the observer is ahead of the charged particle at the time of observation, that is, $z > \beta ct$. Redraw them to make sure they still work in the contrary case.

We conclude that the sum in Equation 51.4 contains just one term. Now we must evaluate the expression $R - \beta\vec{R}_3$ that appears in that formula. Figure 51.3 shows a perpendicular dropped from \mathbf{Q} to the segment $\overline{\mathbf{Q}_r\mathbf{P}}$ in red. We know that $\overline{\mathbf{Q}_r\mathbf{Q}}$ has length $\beta c(t - t_r) = \beta\vec{R}_3$. However, we want

$$\begin{aligned} R - \beta\vec{R}_3 &= R - \overline{\mathbf{Q}_r\mathbf{M}} = \overline{\mathbf{M}\mathbf{P}} \\ &= \sqrt{\overline{\mathbf{Q}\mathbf{P}}^2 - \overline{\mathbf{M}\mathbf{Q}}^2} = \sqrt{\vec{r}_\perp^2 + (z - \beta ct)^2 - (\beta R \sin \psi)^2} \\ &= \sqrt{(1 - \beta^2)\vec{r}_\perp^2 + (z - \beta ct)^2}. \end{aligned}$$

The square root is always real, because $\beta < 1$. Finally, substitute this result into

Note that $\hat{u}_\perp u_\perp + \hat{z}\vec{u}_3$ is just \vec{u} , which is $\vec{r} - \beta ct\hat{z}$. Thus

$$\vec{E} = \frac{q}{4\pi\epsilon_0} \frac{\vec{r} - \beta ct\hat{z}}{(r_\perp^2 + \gamma^2(z - \beta ct)^2)^{3/2}}.$$

This is the same result we obtained by applying a Lorentz transformation to the electrostatic field surrounding a static point charge.

The solution that we have found corresponds to a lump of electric and magnetic field strength that moves at speed βc . The energy flux $\vec{E} \times \vec{B}$, is nonzero, but that just describes the translational motion of the lump of energy associated to those fields.

51.3 CHARGED PARTICLE IN A DIELECTRIC MEDIUM

We seem to have done a lot of work for nothing, but our real destination is to understand what happens when a charged particle passes through matter, for example, water. Section 5.5 argued that in this situation we may *forget* the medium and simply modify the Maxwell equations, replacing ϵ_0 by a larger permittivity ϵ . But now an interesting possibility arises: What if the particle moves faster than the speed of light *in medium*, that is, $\beta c > c/n$ where $n = \sqrt{\epsilon/\epsilon_0}$? It is true that the modified Maxwell equations have a Lorentz-like invariance, with $c_m = c/n$ playing the role of light speed, and we can use that invariance to find the fields if $\beta < c_m$. In the contrary case, however, *there's no Lorentz-type transformation* that can bring us to the rest frame of the particle, and this method breaks down.

Luckily, the proof that the radiation Green function solves the Maxwell equations is just as correct in the medium as it was in vacuum; we need only substitute $c \rightarrow c_m$ in the derivation of Section 51.2. However, the geometry is different when $v/c_m > 1$. In the language of Figure 51.2b, the stick held fixed on the z axis is *longer* than the pivoting stick. You'll explore the consequences of this difference in Problem 51.1, but the upshot is that [Not ready yet.].

51.4 INTERPRETATION

[Not ready yet.]The result just announced may seem paradoxical: How can a non-accelerating charge radiate? Remember, however, that the one charge we investigated is not the only one in the system. The medium that we added is polarizable because it contains many charges in the deformable molecules that constitute it. As the free charge of the particle flies past one such molecule, it gives that molecule a momentary jolt. The sum of the resulting fields from all of the molecules can and does include a radiation component, if $v > c_m$.

FURTHER READING

Smith, 1997; Ginzburg, 1989
Historical: Jelley, 1958.

“Cherenkov radiation, also known as Vavilov–Cherenkov radiation (VCR) (named after Sergey Vavilov and Pavel Cherenkov), is electromagnetic radiation emitted when a charged particle (such as an electron) passes through a dielectric medium at a speed greater than the phase velocity of light in that medium. The characteristic blue glow of an underwater nuclear reactor is due to Cherenkov radiation. It is named after Soviet scientist Pavel Cherenkov, the 1958 Nobel Prize winner who was the first to detect it experimentally. A theory of this effect was later developed within the framework of Einstein’s special relativity theory by Igor Tamm and Ilya Frank, who also shared the Nobel Prize. Cherenkov radiation had been theoretically predicted by the English polymath Oliver Heaviside in papers published in 1888–89.” – https://en.wikipedia.org/wiki/Cherenkov_radiation

PROBLEMS

51.1 *What a shock*

The main text worked out the fields created by a point charge in vacuum, in uniform, straight-line motion, by using the Green function solution.³ Not surprisingly, the fields were exactly the same as what we found by doing a Lorentz transformation on the fields of a charge at rest (Section 32.3.2).

In this problem, you’ll consider fields in a dielectric medium, perhaps water. There is an approximate regime (fields not too strong, time variation not too fast) in which, as I argued long ago, we may forget the medium and just replace ϵ_0 by some larger constant ϵ , the “permittivity” of the medium. We’ll neglect the analogous possibility for magnetic fields (it’s negligible for many dielectric media). Then we just get Maxwell’s equations, and in particular the wave equation, in their usual form apart from a reduced value of the speed of light $\tilde{c} = (\mu_0\epsilon)^{-1/2}$. For example, the Green function is the same apart from that one change.

We can now consider the problem of a charged particle that cruises through this medium at uniform speed $\beta\tilde{c}$. If $\beta < 1$, then everything is exactly the same as before, and we find that (in this approximate treatment of the medium) the charged particle just carries a blob of field energy along with it, and in particular there is no energy radiated out to infinity.

The interesting new electromagnetic phenomenon concerns the possibility that now β may *exceed 1*. No physical law forbids a particle from moving through water at, say $0.9c$, which is $\approx 1.2\tilde{c}$. Now, however, we are on new territory. The modified Maxwell equations have a Lorentz-type invariance, but no transformation of this form can bring a particle from rest to faster than \tilde{c} , so we may not obtain the fields in this easy way. Nevertheless, the proof that the Green function solves the equations is still valid, so we can still use that method.

The main text argued that, for $\beta < 1$, there was always exactly one source point in the past light-cone of any observation point.

- a. Show that, for $\beta > 1$, at any time t some observation points have *no* source point in their past light-cone. The fields at such points, at time t , must equal *zero*.

³See also P+S §15.6.1.

Characterize the set of all such points. [*Hint*: Start by finding the appropriate 1D spacetime diagram, that is, in the $(\tilde{t})(z)$ plane. Then generalize to two space dimensions, adapting the “two sticks” argument to show that some angles θ cannot be attained.]

- b. Show that all observation points inside the allowed region each have *two* source points in their past light cone.⁴ [*Hint*: Make the needed changes to the “third proof” in Section 51.2.]
- c. Then get expressions for the scalar and vector potentials.
- d. Compute appropriate derivatives to find what direction \vec{E} and \vec{B} , and hence the Poynting vector, point. Which way does energy flow? Will it just stay concentrated along the z axis, or flow outward?

⁴Right on the edge of the allowed region, those two points merge into one.

CHAPTER 55

Field Quantization, Polarization, and the Orientation of a Single Molecule

55.1 A SINGLE MOLECULE EMITS PHOTONS IN A DIPOLE DISTRIBUTION

Before we unleash a lot of formulas, let's first frame the issues with an experimental observation. A concrete example of what we'd like to understand is the pattern of light seen from a single immobilized fluorophore, for example in defocused orientation imaging (Figure 55.1). The distribution of photon arrivals resembles the dipole radiation pattern found in Chapter 42, but the emission of *single photons* by a *single molecule* is as far from being classical as one can get. Is the observed agreement in radiation patterns just a coincidence? This chapter will argue that in fact, a quantum-mechanical treatment recapitulates the classical distribution of energy flow as a probability density function for photon arrivals.

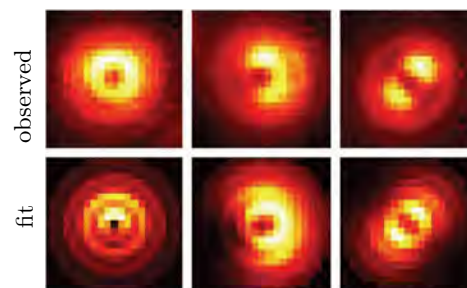
55.2 MAXWELL EQUATIONS AS DECOUPLED HARMONIC OSCILLATORS

Classical electrodynamics describes a system whose states are field configurations. But Nature is described by quantum probability amplitudes, not classical state variables. The goal of this section is therefore to recast Maxwell's eight equations for the electric and magnetic fields in a form that is suitable for quantization. Later, Section 55.3 will recover the photon concept as a consequence of field quantization.

As usual, we can represent electric and magnetic fields via a scalar potential field, $\psi(t, \vec{r})$, and a vector potential field, $\vec{A}(t, \vec{r})$:

$$\vec{E} = -\frac{\partial}{\partial t}\vec{A} - \vec{\nabla}\psi; \quad \vec{B} = \vec{\nabla} \times \vec{A}. \quad [17.22, \text{page 208}]$$

Figure 55.1: Defocused orientation imaging. [Experimental data and fits.] *Top:* Observed point spread functions for three single fluorophores. Lighter colors correspond to pixels with larger photon counts. *Bottom:* Corresponding theoretical predictions, after finding the best-fit value of the angle between the transition dipole and the centerline. From left to right, the fit values of this angle were 10 deg, 60 deg, and 90 deg. The in-plane orientation (azimuth) was also obtained by fitting. [From Toprak et al., 2006.]



Chapter 17 showed that in this representation, half of Maxwell's equations are identities (automatically true). We will choose to work in Coulomb gauge, that is, use only vector potentials that obey $\vec{\nabla} \cdot \vec{A} = 0$.¹ Section 17.8.4 showed that in a world with no charged particles we can always specialize further, supplementing Coulomb gauge with the extra condition that the scalar potential $\psi = 0$ everywhere. (Later sections will reinstate ψ when we consider coupling of the field to electrons.)

We wish to show that Maxwell's equations reduce to a set of simple, decoupled dynamical systems. It's convenient to imagine a finite world of some very large size L , which will ultimately be taken to be infinity, and specifically to take that world to be a cube with periodic boundary conditions. Then the vector potential can be expanded as

$$\vec{A}(t, \vec{r}) = \frac{1}{2} \sum_{\vec{k}}' (\vec{A}_{\vec{k}}(t) e^{i\vec{k} \cdot \vec{r}} + \text{c.c.}). \quad (55.1)$$

In this formula, each coefficient $\vec{A}_{\vec{k}}$ is a complex 3D vector depending on time. There are many such vectors, indexed by a discrete label \vec{k} with components of the form $2\pi\eta_i/L$; the η_i are integers, not all of which are zero. The primed summation means that for each such wavevector \vec{k} , we exclude the redundant $-\vec{k}$.

The Coulomb gauge condition implies that $\vec{k} \cdot \vec{A}_{\vec{k}} = 0$, or in other words that the component of each $\vec{A}_{\vec{k}}$ along its \vec{k} must equal zero. The other two components are unrestricted, so for each \vec{k} , we choose a basis of two real unit vectors perpendicular to it and to each other; we denote these **polarization basis vectors** by $\hat{\zeta}_{(\alpha, \vec{k})}$, where the index α runs from 1 to 2. Then Equation 55.1 becomes

$$\vec{A}(t, \vec{r}) = \frac{1}{2} \sum_{\vec{k}, \alpha}' (A_{\vec{k}, \alpha}(t) \hat{\zeta}_{(\alpha, \vec{k})} e^{i\vec{k} \cdot \vec{r}} + \text{c.c.}). \quad (55.2)$$

The polarization basis vectors are not dynamical variables. The dynamical variables, whose equations of motion we wish to find and quantize, are the mode expansion coefficients $A_{\vec{k}, \alpha}(t)$.

Your Turn 55A

Show that, with these definitions, Maxwell's equations in Coulomb gauge become simple:

$$\frac{d^2}{dt^2} A_{\vec{k}, \alpha} = -(ck)^2 A_{\vec{k}, \alpha}. \quad (55.3)$$

Here α runs over 1, 2, \vec{k} runs over the nonredundant set described earlier, and k denotes the length of the vector \vec{k} (that is, $\|\vec{k}\|$).

Equation 55.3 shows that every distinct combination of polarization α and wavevector \vec{k} corresponds to an independent dynamical system, decoupled from the others. To make the system more familiar, we now give separate names to the real and imaginary parts of $A_{\vec{k}, \alpha}$:²

$$A_{\vec{k}, \alpha} = (\epsilon_0 L^3 / 2)^{-1/2} (X_{\vec{k}, \alpha} + iY_{\vec{k}, \alpha}). \quad (55.4)$$

¹Section 17.8.3 (page 208).

²The overall rescaling chosen in the definitions of X and Y will simplify some later formulas.

The real scalar quantities $X_{\vec{k},\alpha}$ and $Y_{\vec{k},\alpha}$ separately obey Equation 55.3, so we see that

Maxwell's equations in vacuum are mathematically equivalent to a set of decoupled harmonic oscillators. (55.5)

The harmonic oscillator has a well known quantum-mechanical formulation, so Idea 55.5 achieves the first goal of this section.

To understand the meaning of these oscillators better, we now express the electromagnetic field energy \mathcal{E} and momentum \vec{P} in terms of the new variables X and Y . Let $\dot{\vec{A}}$ denote the time derivative $\partial\vec{A}/\partial t$. Then Your Turn 34Ca (page 386) gives

$$\begin{aligned}\mathcal{E} &= \frac{\epsilon_0}{2} \int d^3r (\vec{E}^2 + c^2 \vec{B}^2) = \frac{\epsilon_0}{2} \int d^3r ((-\dot{\vec{A}})^2 + c^2 (\vec{\nabla} \times \vec{A})^2) \\ &= \frac{\epsilon_0}{2} \sum'_{\vec{k}_1, \alpha} \sum'_{\vec{k}_2, \beta} \int d^3r \left[\frac{1}{2} (-\dot{A}_{\vec{k}_1, \alpha} \hat{\zeta}_{(\alpha, \vec{k}_1)} e^{i\vec{k}_1 \cdot \vec{r}} + \text{c.c.}) \cdot \frac{1}{2} (-\dot{A}_{\vec{k}_2, \beta} \hat{\zeta}_{(\beta, \vec{k}_2)} e^{i\vec{k}_2 \cdot \vec{r}} + \text{c.c.}) \right. \\ &\quad \left. + c^2 \frac{1}{2} (A_{\vec{k}_1, \alpha} i\vec{k}_1 \times \hat{\zeta}_{(\alpha, \vec{k}_1)} e^{i\vec{k}_1 \cdot \vec{r}} + \text{c.c.}) \cdot \frac{1}{2} (A_{\vec{k}_2, \beta} i\vec{k}_2 \times \hat{\zeta}_{(\beta, \vec{k}_2)} e^{i\vec{k}_2 \cdot \vec{r}} + \text{c.c.}) \right].\end{aligned}\quad (55.6)$$

The integrals are easy to do, because most of them vanish: Only those cross-terms with $\vec{k}_1 = \vec{k}_2$, and hence involving $e^{i\vec{k}_1 \cdot \vec{r}} e^{-i\vec{k}_1 \cdot \vec{r}} = 1$, survive. Moreover, we have $\hat{\zeta}_{(\alpha, \vec{k})} \cdot \hat{\zeta}_{(\beta, \vec{k})} = \delta_{\alpha\beta}$, leaving

$$\begin{aligned}\mathcal{E} &= \frac{\epsilon_0 L^3}{4} \sum'_{\vec{k}, \alpha} (|\dot{A}_{\vec{k}, \alpha}|^2 + (ck)^2 |A_{\vec{k}, \alpha}|^2) \\ &= \frac{1}{2} \sum'_{\vec{k}, \alpha} (\dot{X}_{\vec{k}, \alpha}^2 + (ck)^2 X_{\vec{k}, \alpha}^2 + \dot{Y}_{\vec{k}, \alpha}^2 + (ck)^2 Y_{\vec{k}, \alpha}^2).\end{aligned}\quad (55.7)$$

The field momentum is given by a similar calculation, starting with the Poynting vector (Your Turn 34C (page 386)b):

$$\begin{aligned}\vec{P} &= \epsilon_0 \int d^3r \vec{E} \times \vec{B} \\ &= \epsilon_0 \sum'_{\vec{k}_1, \alpha} \sum'_{\vec{k}_2, \beta} \int d^3r \frac{1}{2} (-\dot{A}_{\vec{k}_1, \alpha} \hat{\zeta}_{(\alpha, \vec{k}_1)} e^{i\vec{k}_1 \cdot \vec{r}} + \text{c.c.}) \times \left(\vec{\nabla} \times \frac{1}{2} (A_{\vec{k}_2, \beta} \hat{\zeta}_{(\beta, \vec{k}_2)} e^{i\vec{k}_2 \cdot \vec{r}} + \text{c.c.}) \right) \\ &= -\frac{\epsilon_0 L^3}{4} \sum'_{\vec{k}, \alpha} \sum_{\beta} (\dot{A}_{\vec{k}, \alpha} A_{\vec{k}, \beta}^* \hat{\zeta}_{(\alpha, \vec{k})} \times (-i\vec{k} \times \hat{\zeta}_{(\beta, \vec{k})}) + \text{c.c.}) \\ &= \frac{\epsilon_0 L^3}{4} \sum'_{\vec{k}, \alpha} (i\vec{k} \dot{A}_{\vec{k}, \alpha} A_{\vec{k}, \alpha}^* + \text{c.c.}) \\ &= \frac{1}{2} \sum'_{\vec{k}, \alpha} \vec{k} ((i\dot{X}_{\vec{k}, \alpha} - \dot{Y}_{\vec{k}, \alpha})(X_{\vec{k}, \alpha} - iY_{\vec{k}, \alpha}) + \text{c.c.}) \\ &= \sum'_{\vec{k}, \alpha} \vec{k} (\dot{X}_{\vec{k}, \alpha} Y_{\vec{k}, \alpha} - \dot{Y}_{\vec{k}, \alpha} X_{\vec{k}, \alpha}).\end{aligned}\quad (55.9)$$

We now have compact formulas for the energy and momentum of the electromagnetic field in terms of the harmonic-oscillator representation (Equation 55.7 and

55.9). The interpretation is that every mode of the field, labeled by \vec{k} and α , makes an independent contribution to \mathcal{E} , and also to each component of \vec{P} . Note, however, that the momentum gets mixed contributions from the X and Y oscillators. We will soon remove this remaining inconvenience.

55.3 QUANTIZATION REPLACES FIELD VARIABLES BY OPERATORS

Finding the quantum-mechanical version of a harmonic oscillator is a standard problem which will be easy after we make a rather involved change of variables. To motivate the required change, we will break it down into four steps. It is worthwhile to verify each of the steps, which are straightforward if a bit tedious; ultimately the goal is to replace the X and Y variables by a set of quantum operators called \mathbf{Q} and their Hermitian conjugates (Equation 55.21). Note that this chapter uses different typefaces to distinguish quantum operators from their corresponding classical dynamical variables.

Step 1: Quantize

For brevity, at first consider only one pair of modes X and Y , that is, only a particular \vec{k}, α . We introduce two Hermitian operators³ \mathbf{X} and \mathbf{U} , with the property that their commutator is $[\mathbf{X}, \mathbf{U}] = i\hbar$. In the energy function, Equation 55.7, we substitute $X \rightarrow \mathbf{X}$ and $\dot{X} \rightarrow \mathbf{U}$ to obtain the Hamiltonian operator for X :

$$\mathbf{H}_X = \frac{1}{2}(\mathbf{U}^2 + (ck)^2\mathbf{X}^2). \quad (55.10)$$

This operator both represents the energy of a quantum state and also determines its time evolution. For example, the time evolution of $|\Psi(t)\rangle$ is given by $\exp(-i\mathbf{H}_X t/\hbar)|\Psi\rangle$. It implies that

$$\begin{aligned} \frac{d^2}{dt^2}\langle\Psi_1|\mathbf{X}|\Psi_2\rangle &= \frac{d}{dt}\langle\Psi_1|\frac{i}{\hbar}[\mathbf{H}_X, \mathbf{X}]|\Psi_2\rangle = \frac{d}{dt}\langle\Psi_1|\mathbf{U}|\Psi_2\rangle = \langle\Psi_1|\frac{i}{\hbar}[\mathbf{H}_X, \mathbf{U}]|\Psi_2\rangle \\ &= -(ck)^2\langle\Psi_1|\mathbf{X}|\Psi_2\rangle, \end{aligned} \quad (55.11)$$

which implements the classical equation of motion for the harmonic oscillator in Equation 55.3.

We proceed in the same way with the other oscillator family, introducing operators \mathbf{Y} and \mathbf{V} analogous to \mathbf{X} and \mathbf{U} . Then the operator corresponding to $A_{\vec{k}, \alpha}$ in Equation 55.4 is

$$\mathbf{A} = (\epsilon_0 L^3/2)^{-1/2}(\mathbf{X} + i\mathbf{Y}). \quad (55.12)$$

Step 2: Diagonalize energy

We could now finish constructing the state space, for example, by writing and solving a set of decoupled Schrödinger equations for each pair of operators (\mathbf{X}, \mathbf{U}) and (\mathbf{Y}, \mathbf{V}) . However, the harmonic oscillator problem has an elegant reformulation that simplifies the math. Change variables once again by defining new operators

$$\mathbf{S} = (2\hbar ck)^{-1/2}(ck\mathbf{X} + i\mathbf{U}) \text{ and } \mathbf{R} = (2\hbar ck)^{-1/2}(ck\mathbf{Y} + i\mathbf{V}). \quad (55.13)$$

³In the analogy to a harmonic oscillator, these represent the position and momentum respectively, but in electrodynamics they have no direct connection to physical position \vec{r} or field momentum \vec{P} .

Then it is straightforward to verify that

$$[S, S^\dagger] = 1, \quad [R, R^\dagger] = 1, \quad [S, R] = [S, R^\dagger] = 0, \quad (55.14)$$

$$H = H_X + H_Y = \hbar ck(S^\dagger S + R^\dagger R + 1), \text{ and} \quad (55.15)$$

$$\vec{P} = i\hbar\vec{k}(S^\dagger R - \text{h.c.}). \quad (55.16)$$

In the last formula, “h.c.” denotes the Hermitian conjugate, that is, $R^\dagger S$.

Step 3: Diagonalize momentum

The Hamiltonian operator has the nice property that S and R make independent, additive contributions to it. The momentum operator still mixes S and R , but we can diagonalize it, without spoiling H , by a unitary transformation. Define two new **lowering operators** by

$$Q = (S + iR)/\sqrt{2}, \quad \tilde{Q} = (S - iR)/\sqrt{2}. \quad (55.17)$$

Your Turn 55B

Show that

$$[Q, Q^\dagger] = 1, \quad [\tilde{Q}, \tilde{Q}^\dagger] = 1, \quad [Q, \tilde{Q}] = [Q, \tilde{Q}^\dagger] = 0, \quad (55.18)$$

$$H = \hbar ck(Q^\dagger Q + \tilde{Q}^\dagger \tilde{Q} + 1), \text{ and} \quad (55.19)$$

$$\vec{P} = \hbar\vec{k}(Q^\dagger Q - \tilde{Q}^\dagger \tilde{Q}). \quad (55.20)$$

We now have new field operators Q and \tilde{Q} that, unlike S and R , enter independently into both the field energy and momentum.

Step 4: Relabel

We now reinstate the mode indices \vec{k} and α . Until now, all mode sums were over a half-space of discrete \vec{k} values, but now we can simplify the notation: Define operators for *all* nonzero \vec{k} by renaming $\tilde{Q}_{\vec{k},\alpha}$ as $Q_{-\vec{k},\alpha}$. Then

$$[Q_{\vec{k}_1,\alpha}, Q_{\vec{k}_2,\beta}^\dagger] = \delta_{\alpha\beta}\delta_{\vec{k}_1,\vec{k}_2}, \quad [Q_{\vec{k}_1,\alpha}, Q_{\vec{k}_2,\beta}] = 0, \quad \text{for all nonzero } \vec{k}_1 \text{ and } \vec{k}_2. \quad (55.21)$$

Our final formulas then become unrestricted sums:

$$H = \sum_{\vec{k},\alpha} \hbar ck(Q_{\vec{k},\alpha}^\dagger Q_{\vec{k},\alpha} + \frac{1}{2}), \text{ and} \quad (55.22)$$

$$\vec{P} = \sum_{\vec{k},\alpha} \hbar\vec{k}(Q_{\vec{k},\alpha}^\dagger Q_{\vec{k},\alpha}). \quad (55.23)$$

We now have a set of operators in terms of which the energy and momentum of light will have simple interpretations.

55.4 PHOTON STATES

55.4.1 Basis states can be formed by applying creation operators to the vacuum state

We have found a set of field-like *operators* that obey Maxwell-like equations, and recast them in terms of the Q and Q^\dagger operators. Besides giving an elegant approach to quantization, this formulation gives a basis of states that is readily interpretable.

Your Turn 55C

Show that

$$[H, Q_{\vec{k},\alpha}] = -\hbar ck Q_{\vec{k},\alpha} \quad \text{and} \quad [\vec{P}, Q_{\vec{k},\alpha}] = -\hbar \vec{k} Q_{\vec{k},\alpha}. \quad (55.24)$$

Equations 55.24 justify the term “lowering operator”:

Applying the lowering operator $Q_{\vec{k},\alpha}$ to a state lowers its energy by $\hbar ck$, and changes its momentum by $-\hbar \vec{k}$. Conversely, applying the raising operator $Q_{\vec{k},\alpha}^\dagger$ has the opposite effects. (55.25)

Next, note that both of the terms in the classical electromagnetic energy function (Equation 55.6) are nonnegative. So it must not be possible to lower that energy indefinitely; there must be a state for which any lowering operator yields *zero*. We’ll denote that **photon ground state** by the symbol $|0\rangle$. Any other state is obtained from this one by the actions of the various raising operators, each of which may be applied any number of times, always raising the energy by $\hbar ck$ and changing the momentum by $\hbar \vec{k}$. The spectrum of allowed energy and momentum values suggests a description: It is exactly the same as that of a *gas of noninteracting particles*, each carrying energy $\hbar ck$ and momentum $\hbar \vec{k}$.

Your Turn 55D

Show that when a raising operator acts n times, we can obtain a normalized state as follows:

$$|n_{\vec{k},\alpha}\rangle = \sqrt{\frac{1}{n!}} (Q_{\vec{k},\alpha}^\dagger)^n |0\rangle. \quad (55.26)$$

More generally, we can define $|n_{\vec{k}_1,\alpha_1}; n_{\vec{k}_2,\alpha_2}, \dots\rangle$ as a state obtained by applying several different raising operators to the ground state, each multiple times, and then normalizing. States of this form with different sets of **occupation numbers** are all linearly independent and orthogonal. In fact,

The quantum states of light form a linear space spanned by basis vectors of this form, which act like states of noninteracting particles (“photons”). (55.27)

That is, each one-photon basis state is labeled by a wavevector and a polarization, and carries energy and momentum related by Equation 55.24:

$$\mathcal{E}_{\vec{k},\alpha} = \hbar ck; \quad \vec{p}_{\vec{k},\alpha} = \hbar \vec{k}; \quad \text{so} \quad \mathcal{E}_{\vec{k},\alpha} = c \|\vec{p}_{\vec{k},\alpha}\|, \quad (55.28)$$

implying that photons are massless (Equation 30.15 (page 329)). For multiphoton states, we add the corresponding quantities, just as we would do with any noninteracting particles.

The interpretation of the quantum basis states as containing particles motivates another commonly used set of terms for the raising and lowering operators: Because they can be interpreted as raising and lowering the *number of photons* in a state, they are also called **creation and destruction operators**; $|0\rangle$ is also called the **vacuum state**. We may guess that these concepts will be key to understanding how a fluorescent molecule in its excited state can create photons from “nothing” (and how other processes can make photons disappear).

The fact that the collection of occupation numbers, $\{n_{\vec{k},\alpha}\}$, fully determines a basis state is the key insight that leads to the famous spectrum of thermal (“black body”) radiation. This aspect of light can alternatively be expressed by saying that the particles of light with given \vec{k}, α are *indistinguishable*: they have no further attributes, so all we need to state is how many are present. For example, it doesn’t matter in what order we build a photon state by applying raising operators, because those operators all commute with one another.⁴

55.4.2 Coherent states mimic classical states in the limit of large occupation numbers

The states we have called “one-photon” are far from being classical. Indeed, no state with a definite number of photons can be an eigenvector of the field operators corresponding to the classical electric and magnetic field, because $\vec{A}(\vec{r})$ involves both raising and lowering operators:

Your Turn 55E

Use Equations 55.2, 55.12, 55.13, and 55.17 to show that

$$\vec{A}(\vec{r}) = \sum_{\vec{k},\alpha} \sqrt{\frac{\hbar}{2L^3\epsilon_0 c k}} \hat{\zeta}_{(\alpha,\vec{k})} (\mathcal{Q}_{\vec{k},\alpha} e^{i\vec{k}\cdot\vec{r}} + \text{h.c.}). \quad (55.29)$$

However, we can find eigenvectors of $\mathcal{Q}_{\vec{k},\alpha}$, called **coherent states**: For any complex number u , define

$$|u, \vec{k}, \alpha\rangle = \exp(-\frac{1}{2}|u|^2) \sum_{n=0}^{\infty} (n!)^{-1/2} (u)^n |n_{\vec{k},\alpha}\rangle. \quad (55.30)$$

⁴More precisely, a class of particles that are indistinguishable in this way is called “bosonic.” Another possibility, called “fermionic” particles, has raising operators that mutually *anticommute*.

Your Turn 55F

- Show that the states $|u, \vec{k}, \alpha\rangle$ just defined are all properly normalized for any complex number u .
- Show that $Q_{\vec{k}, \alpha}|u, \vec{k}, \alpha\rangle = u|u, \vec{k}, \alpha\rangle$, and hence also $\langle u, \vec{k}, \alpha|Q_{\vec{k}, \alpha}^\dagger = u^*\langle u, \vec{k}, \alpha|$.
- Then show that Equation 55.29 implies

$$\langle u, \vec{k}, \alpha|\vec{A}(\vec{r})|u, \vec{k}, \alpha\rangle = (2L^3\epsilon_0ck/\hbar)^{-1/2}\hat{\zeta}_{(\alpha, \vec{k})}ue^{i\vec{k}\cdot\vec{r}} + \text{c.c.}$$

Your results show that the coherent state based on a particular wavevector and polarization is the quantum analog of a classical single-mode state (Equation 55.2, page 536). Moreover, as the amplitude $|u|$ becomes large (and hence also the expectation of the photon number), the relative standard deviation of the electric field in this state goes to zero, leading to classical behavior. In this limit, the coherent states correspond to classical states of the electromagnetic field, for example the radiation emitted by a radio broadcast antenna.⁵

Your Turn 55G

The coherent states are superpositions of states with different numbers of photons. Find the length-squared of the individual terms of Equation 55.30 to get the probabilities of getting exactly ℓ photons in a measurement on that state. Is this a distribution you have seen previously?

This section has established contact between the field quantization procedure in this chapter, the particle picture from earlier chapters, and Maxwell's original classical fields.

55.5 INTERACTION WITH ELECTRONS**55.5.1 Classical interactions involve adding source terms to the field equations**

If we wish to study the creation of light by a molecule, then we must acknowledge that the light field interacts with that molecule's electrons. In the presence of charged matter, we can no longer find a gauge transformation that eliminates the scalar potential ψ , though we can still impose $\vec{\nabla} \cdot \vec{A} = 0$. The electric Gauss law then says

$$\vec{\nabla} \cdot \vec{E} = -\nabla^2\psi = \rho_q/\epsilon_0, \quad [2.3, \text{page } 22]$$

where ρ_q is the charge density. This formula looks just like the corresponding equation in electrostatics, and it leads to the usual potential that binds the molecule's electrons to its nuclei.

Ampère's law also involves charges, via the electric current density $\vec{j}(t, \vec{r})$:

$$\vec{\nabla} \times \vec{B} = \mu_0\vec{j} + \mu_0\epsilon_0\frac{\partial}{\partial t}\vec{E}. \quad [17.15, \text{page } 202]$$

⁵Books on quantum optics show that the light created by a single-mode laser, operated well above threshold, is also a coherent state (Loudon, 2000, chapt. 7).

Casting everything into plane wave mode expansions as before gives the full Maxwell equations as

$$k^2 \psi_{\vec{k}} = \frac{1}{\epsilon_0} \rho_{\mathbf{q}, \vec{k}} \quad \text{and} \quad (55.31)$$

$$\frac{d^2}{dt^2} \vec{A}_{\vec{k}} + (ck)^2 \vec{A}_{\vec{k}} = -i\vec{k} \frac{d\psi_{\vec{k}}}{dt} + \frac{1}{\epsilon_0} \vec{j}_{\vec{k}}, \quad (55.32)$$

where $c = (\mu_0 \epsilon_0)^{-1/2}$ and $\psi_{\vec{k}}$, $\rho_{\mathbf{q}, \vec{k}}$, and $\vec{j}_{\vec{k}}$ are the plane-wave components of ψ , $\rho_{\mathbf{q}}$, and \vec{j} , respectively. We now take the dot product of both sides of Equation 55.32 with the two transverse basis vectors $\hat{\zeta}_{(\alpha, \vec{k})}$ to find the desired generalization of Equation 55.3:

$$\frac{d^2}{dt^2} A_{\vec{k}, \alpha} = -(ck)^2 A_{\vec{k}, \alpha} + \frac{1}{\epsilon_0} \vec{j}_{\vec{k}} \cdot \hat{\zeta}_{(\alpha, \vec{k})} \quad \text{for each } \vec{k}, \alpha. \quad (55.33)$$

The scalar potential ψ has dropped out of this equation of motion.

55.5.2 Electromagnetic interactions can be treated perturbatively

There is no need to quantize the scalar potential ψ , because Equation 2.3 shows that in Coulomb gauge, it is not an independent dynamical variable: It just tracks whatever the charge density is doing.

The last term of Equation 55.33 describes the interaction of the vector potential with charge flux. To discuss the radiation of a molecule, we treat this term as a perturbation. That is, we set up an “unperturbed” Hamiltonian operator describing the quantum mechanics of the electrons making up the molecule, with their Coulomb attraction to the nuclei mediated by the scalar potential ψ as usual. There is another term describing the free electromagnetic field (Equation 55.22). To these terms we then add the perturbation

$$- \int d^3r \vec{j}(\vec{r}) \cdot \vec{A}(\vec{r}), \quad (55.34)$$

where $\vec{j}(\vec{r})$ is the operator version of the current density and $\vec{A}(\vec{r})$ is given by Equation 55.29. This term modifies the quantum equations of motion, introducing the last part of Equation 55.33.

Each electron in the atom or molecule of interest contributes a delta function to \vec{j} that is localized at the electron’s position \vec{r}_e , with strength equal to its charge, $-e$, times its velocity, \vec{p}_e/m_e . Thus, each electron makes a contribution to the integral in Equation 55.34 equal to

$$- \sum_{\vec{k}, \alpha} \sqrt{\frac{\hbar}{2L^3 \epsilon_0 ck}} \hat{\zeta}_{(\alpha, \vec{k})} \cdot (-e)(\vec{p}_e/m_e) (\mathbf{Q}_{\vec{k}, \alpha} e^{i\vec{k} \cdot \vec{r}_e} + \text{h.c.}). \quad (55.35)$$

The effect of this perturbation is to allow transitions between eigenstates of the unperturbed Hamiltonian operator, that is, between states that would be stationary were it not for the perturbation term. For example, the transitions that interest us are those from a molecule with initially excited electron state and no photons present, to a deexcited electron state and one photon present. To find the probability per

unit time that this transition will occur, we need to compute the modulus squared of Equation 55.35 sandwiched between the initial and final states.⁶ The Hermitian conjugate term, involving $Q_{\vec{k},\alpha}^\dagger$, can create the photon, so we want the matrix element of the remaining factors of this term sandwiched between the molecular states.

To make progress, notice that for transitions in the visible spectrum, $k \approx 10^{-2} \text{ nm}^{-1}$. But r_e cannot exceed the size of the atom or molecule, typically $\approx 1 \text{ nm}$, so $\vec{k} \cdot \vec{r}_e$ is a small dimensionless number. Accordingly, we will approximate $\exp(i\vec{k} \cdot \vec{r}_e)$ by its leading-order Taylor series term, which is 1—the electric dipole approximation.⁷

55.5.3 The dipole emission pattern

We now ask for the probability that the emitted photon will be observed to be traveling in a particular direction with a particular energy and polarization. The preceding section argued that dropping overall constant factors, the answer is proportional to

$$\begin{aligned} & \left| \langle \text{ground}; \vec{k}, \alpha | \sum_{\vec{k}', \beta} Q_{\vec{k}', \beta}^\dagger \hat{\zeta}_{(\beta, \vec{k}')} \cdot \vec{p}_e | \text{excited} \rangle \right|^2 \\ &= \left| \langle \text{ground} | \vec{p}_e | \text{excited} \rangle \cdot \hat{\zeta}_{(\alpha, \vec{k})} \right|^2. \end{aligned} \quad (55.36)$$

One further transformation helps to clarify the meaning of this quantity. The electron momentum operator, whose matrix element we need, can be rephrased in terms of the electron *position* operator, as the commutator

$$[H_e, \vec{r}_e] = \frac{-i\hbar}{m} \vec{p}_e.$$

Sandwich this relation between the ground and excited states to find

$$\langle \text{ground} | (E_0 \vec{r}_e - \vec{r}_e E_*) | \text{excited} \rangle = \frac{-i\hbar}{m} \langle \text{ground} | \vec{p}_e | \text{excited} \rangle.$$

The right-hand side of this formula is a constant times the quantity needed in Equation 55.36. The left-hand side can be written in terms of the electric dipole moment operator, $\vec{D}_E = -e\vec{r}_e$, so we find that the probability of photon emission involves the matrix element of the dipole moment, a vector called the molecule's **transition dipole**. This is encouraging news: In classical electrodynamics the rate of energy radiation is also proportional to the amplitude squared of the electric dipole moment.

If the molecular states are such that the transition dipole is nonzero, then we can choose a coordinate system in which it points along the z axis:

$$\langle \text{ground} | \vec{D}_E | \text{excited} \rangle = \mathcal{D}_E \hat{z}. \quad (55.37)$$

Suppose that, as is the case in many experiments, we record every photon received regardless of its polarization. The sum of Equation 55.36 over α includes the factor⁸

$$\sum_{\alpha} \hat{z} \cdot \hat{\zeta}_{(\alpha, \vec{k})} \hat{\zeta}_{(\alpha, \vec{k})} \cdot \hat{z}. \quad (55.38)$$

⁶Quantum mechanics textbooks call this scheme the “Golden Rule” of time-dependent perturbation theory.

⁷See Chapter 42.

⁸We chose $\hat{\zeta}_{(\alpha, \vec{k})}$ to be real vectors in Section 55.2.

We can simplify this expression by realizing that it involves the *projection* of \hat{z} onto the plane perpendicular to \vec{k} . Another expression for that projection operator is $1 - \hat{k}\hat{k}$, so we get

$$\hat{z} \cdot (1 - \hat{k}\hat{k}) \cdot \hat{z} = \hat{z} \cdot \hat{z} - (\hat{z} \cdot \hat{k})^2 = 1 - \cos^2 \theta = \sin^2 \theta, \quad (55.39)$$

where θ is the polar angle between the direction of observation, \hat{k} , and the transition dipole.

Equations 55.39 and 55.36 show that the probability density function for the angles at which photons are emitted has a “dipole doughnut” pattern: No photons are emitted along $\pm\hat{z}$; instead, they are preferentially emitted in the equatorial belt $\theta \approx \pi/2$.⁹ A similar argument shows that the probability to *absorb* light also follows a dipole pattern.

The mean rate at which photons are emitted is determined by the transition dipole \mathcal{D}_E defined by Equation 55.37, which itself is essentially the matrix element of the molecule’s electric dipole moment operator.

If the matrix element of the dipole moment operator is nonzero, then the dominant mechanism of energy loss by a molecule is the one just described, with its characteristic angular distribution $\wp(\theta, \varphi) \propto \sin^2 \theta$. (55.40)

This section has resolved the puzzle posed at the start of this chapter: The pattern of photon emission observed in defocused orientation imaging (Figure 55.1) agrees with the dipole radiation pattern in classical electrodynamics because the same angular factors enter each calculation.

55.6 VISTAS

55.6.1 Some transitions are far more probable than others

Section 55.5 focused on the relative mean rates to emit photons in different directions. To find the absolute rates, we need various other factors provided by the “Golden Rule” of time-dependent perturbation theory. The derivation of the rule also shows why energy must be conserved in photon emission and absorption, or more precisely, it must be conserved to within a tolerance set by the uncertainty relation.

For simplicity, Section 55.2 chose to expand the vector potential \vec{A} in a basis of linearly polarized, plane wave states. Other bases may be better adapted to the problem at hand, for example, a basis of circularly polarized plane waves. Also, a basis of outgoing *spherical* waves, centered on the emitting object, is better suited to study light emitted by a very small object and traveling out to infinity. That basis can be chosen such that each element carries definite angular momentum away from the emitter. When we do this, we find that certain kinds of photons cannot be emitted at all by certain kinds of transitions, because doing so would violate the conservation of angular momentum. Other transitions appear impossible when we make the approximation $\exp(i\vec{k} \cdot \vec{r}_e) \approx 1$, as was done in Section 55.5.2, but not when we retain higher terms in the Taylor series. Such transitions are called “forbidden,” but more precisely their rates are just suppressed by powers of the small factor $(kr_e)^2$.

⁹See Section 37.2.2.

The statement that some transitions are “forbidden” is an example of a **selection rule**. Another class of selection rules arises from considerations of electron spin in multi-electron atoms or molecules. It is possible for a molecule to get trapped in an excited state, from which transitions to the ground state are suppressed by a spin selection rule. Such an excited state can eventually make its transition, but with mean rate far slower than most fluorescence transitions, leading to the phenomenon of **phosphorescence** (ultra-slow fluorescence). Spin selection rules also ensure very slow exit from the dark states of some fluorophores, which is useful for localization microscopy.

55.6.2 Lasers exploit a preference for emission into an already occupied state

Sections 55.5.2–55.5.3 restricted attention to the case in which a photon is emitted into a world originally containing *no* photons. Although photons do not interact in the usual sense of colliding, nevertheless a very important new phenomenon arises when we consider *adding* a photon to a state that is already occupied. If a mode initially contains n photons, Equation 55.26 (page 540) implies

$$\langle n+1 | Q^\dagger | n \rangle = \langle 0 | \frac{1}{\sqrt{(n+1)!}} Q^{n+1} (Q^\dagger)^{n+1} \frac{1}{\sqrt{n!}} | 0 \rangle = \langle 0 | \sqrt{\frac{(n+1)!}{n!}} | 0 \rangle = \sqrt{n+1}.$$

This factor gets squared when it enters into the rate for photon emission into this mode. Because this matrix element depends on n , we conclude that

When an atom or molecule emits a photon, it preferentially chooses a mode that is already occupied. (55.41)

If we have a population of many excited atoms or molecules, then this result implies that there can be an avalanche-type effect, in which one particular mode gets the vast majority of all emitted photons. This mechanism for obtaining nearly single-mode light is called **light amplification by stimulated emission of radiation**—the **laser**.

FURTHER READING

Semipopular:

Walmsley, 2015.

Intermediate:

Quantum mechanics and the radiation field: Feynman et al., 2010c, chapt. 9.

Specifically on the quantum theory of light: Lipson et al., 2011; Leonhardt, 2010; Loudon, 2000; Nelson, 2017.

Radiation; forbidden transitions: van der Straten & Metcalf, 2016.

Technical:

General: Berman & Malinovsky, 2011; Mandel & Wolf, 1995.

Defocused orientation imaging: Toprak et al., 2006; Böhmer & Enderlein, 2003; Bartko & Dickson, 1999a; Bartko & Dickson, 1999b.

Epilogue

Above stands the marble smile of implacable Nature, which has endowed us more with longing than with intellectual capacity.

— *Einstein*

The course started slowly but in the end we covered a lot. What was it about? As I said at the start, “organize/systematize/consolidate/integrate.” But there was more than that.

Did you walk into this room in January thinking that classical electrodynamics ended with Maxwell? My attitude has been that, besides enabling a lot of technology (2d industrial revolution...), it is also the springboard to all of modern physics theories, largely in ways you weren't taught in undergrad. Let's think where we've been.

We were always interested in, and guided by, real-world Electromagnetic Phenomena, such as:

- The multipole expansion gave us a rough guide to the physical properties of various substances (for example, boiling points) just based on the geometry of their molecular constituents. Besides cleaning up (unifying) a lot of old ideas in statics (and in mechanics), the multipole expansion came back to organize our study of magnetostatics and then radiation.
- Not only multipoles, but also the geometrical description of the shapes of fluid interfaces benefited from formulation in terms of tensors. Besides cleaning up (unifying) a lot of old ideas, 3-tensors and their 4D big brothers helped us to see new things about relativity.
- Electrostatics makes nontrivial testable predictions even in world of biomacromolecules.
- Magnetic tweezers exploit multipole forces to let us manipulate microscopic objects.
- Multipole expansion gave the distance dependence of FRET, establishing a “spectroscopic ruler.”
- Electrostatics and magnetostatics predict the velocity of light.
- Fields can transport momentum and angular momentum, not just energy: Jumping ring, radiation pressure, optical tweezers, comet tails, orbital angular momentum “tractor beams,” ...
- Mirages, rainbows, and other optical phenomena that are exotic enough to be interesting, yet common enough to see without fancy instruments.
- Waves interfere, but with two independent polarizations.

- A number of non-null experiments quantitatively distinguish Lorentz from Galilean: Doppler, Fizeau, aberration, muon lifetime, breakdown of the cyclotron principle, $E = mc^2$, . . .
- Microwaves behave in many ways similarly to light.
- Cosmic microwave background radiation: Its dipole anisotropy, polarization patterns, . . .
- IR absorption of greenhouse gases.
- Photonic bandgap materials, IR laser scalpel, . . .
- Optical rotatory power, even in isotropic media. Birefringence, dichroism, and related optical effects in media.
- Synchrotron radiation, bremsstrahlung, transition radiation, with all their technological impact.
- Besides cleaning up (unifying) a lot of old ideas, the energy–momentum flux tensor came back as fundamental in finding the appropriate quantum generalization of electrodynamics. That gave us insight into the dipolar probability distribution of individual photons from a single fluorescent molecule.

The basic epistemological miracle of science is that sometimes—more often than one might have a right to expect—a good idea has a far greater range of validity than what was envisioned when it was proposed (then we call it a “principle”). Physics looks at those principles whose consequences are not visible to the naked eye, but become visible when we bring the lens of mathematical analysis to bear on them. Then we get a bigger miracle: Sitting in our armchair, a long chain of mathematical reasoning can actually lead to *new knowledge*. It’s pointless for philosophers to quibble that that new knowledge was really latent in the starting facts—it’s something not previously known to any human being. (Of course it’s provisional until tested by experiment.)

This procedure doesn’t always work, but in physics it has sometimes worked in the past. In the other sciences you essentially never get new knowledge in this way. It’s a nontrivial synthetic step. It has happened a lot in electrodynamics:

- Maxwell... need the extra term in Ampère law... waves... radio technology...
- Einstein... no aether... no Galilean, but yet Lorentz invariance... sounded abstract, but it led to specific experimental predictions.
- More generally, the systematic exploitation of overarching symmetry principles (implemented with tensor methods), gave us “Einstein thinking,” which then led us to many payoffs.
- A lot of vistas now open up. What if mu and epsilon are negative... the perfect lens, maybe cloaking etc... What if we set up an evanescent wave; it could still excite fluorescence etc... Polarization in CMBR tells us about mass distribution... what if we could engineer a material that neither absorbs IR light nor allows it to propagate...

Genius is helpful but not required; the tools we developed represent the *distilled genius* of others. Maybe this is the first physics course where you didn’t learn how to solve some new differential equation analytically. Instead we focused on multipole expansion; group theory; tensor analysis. Other uses abound: SU(5) symmetry brings

its own flavor of tensor analysis as does liquid crystal physics; general relativity takes our tensor analysis as point of departure.

Of course you can make errors. Experiment is still the decisive step. Experience and practice help you to make fewer errors, so you can do more relevant experiments that test more promising ideas.

Finally, we have also been interested in how you get answers when analytic methods are too difficult. For this we have sometimes turned to numerical methods. Even if the calculations are not too difficult, the computer can give us a lot of insight via fast accurate visualization of our formulas.

This is a solemn, poignant moment in your intellectual journey: The last moment when you'll all be in the same room, thinking about the same physics. But it needn't be so. Yes, you're about to scatter to your respective research specialties. But you can still keep talking to each other about each others' projects. That has made me happy, or at least happier than I would otherwise have been. Good luck with your own search.

APPENDIX A

Units and Dimensional Analysis

x.

— y

Some physical quantities are naturally integers, like the number of discrete clicks made by a Geiger counter. But others are continuous, and most continuous quantities must be expressed in terms of conventional units. This book uses the *Système International*, or **SI units**, but you'll need to be able to convert units when reading other works. Units and their conversions in turn form part of a larger framework called **dimensional analysis**.

Dimensional analysis gives a powerful method for catching algebraic errors, as well as a way to organize and classify numbers and situations, and even to guess new physical laws, as we'll see in Section A.4.

To handle units systematically, remember that

*A “unit” acts like a symbol representing an unknown quantity. Most continuous physical quantities should be regarded as the **product** of a pure number times one or more units.*

(A few physical quantities, for example, those that are intrinsically integers, have no units and are called **dimensionless**.) We carry the unit symbols along throughout our calculations. They behave just like any other multiplicative factor; for example, a unit can cancel if it appears in the numerator and denominator of an expression.¹ We know relations among certain units; for example, we know that 1 inch \approx 2.54 cm. Dividing both sides of this formula by the numeric part, we find 0.39 inch \approx 1 cm, and so on.

A.1 BASE UNITS

The SI chooses “base” units for length, time, mass, and electric charge: Lengths are measured in meters (abbreviated m), masses in kilograms (kg), time in seconds (s), and electric charge in coulombs (which this book abbreviates as coul).² The system also creates related units via the prefixes giga ($=10^9$), mega ($=10^6$), kilo ($=10^3$), deci ($=10^{-1}$), centi ($=10^{-2}$), milli ($=10^{-3}$), micro ($=10^{-6}$), nano ($=10^{-9}$), pico ($=10^{-12}$), or femto ($=10^{-15}$), abbreviated as G, M, k, d, c, m, μ , n, p, and f respectively. Thus, 1 nm is a nanometer (or 10^{-9} m), 1 μ g is a microgram, and so on.

A symbol like μm^2 means $(\mu\text{m})^2 = 10^{-12} \text{m}^2$, not “ $\mu(\text{m}^2)$.”

¹One exception involves temperatures expressed using the Celsius and Fahrenheit scales, each of which differ from the absolute (Kelvin) scale by an offset.

²The standard abbreviation is C, but this risks confusion with the speed of light, a concentration or capacitance variable, or a generic constant.

A.2 DIMENSIONS VERSUS UNITS

Other quantities, such as electric current, derive their standard units from the base units. But it is useful to think about current in a way that is less strictly tied to a particular unit system. Thus, we define abstract **dimensions**, which tell us *what kind of quantity* a variable represents. For example,

- The symbol \mathbb{L} denotes the *dimension* of length. The SI assigns it a base *unit* called “meters,” but other units exist with the same dimension (for example, miles or centimeters). Once we have chosen a unit of length, we then also get derived units for area (m^2) and volume (m^3), which have dimensions \mathbb{L}^2 and \mathbb{L}^3 , respectively.
- The symbol \mathbb{M} denotes the dimension of mass. Its SI base unit is the kilogram.
- The symbol \mathbb{T} denotes the dimension of time. Its SI base unit is the second.
- The symbol \mathbb{Q} denotes the dimension of electric charge.³ Its SI base unit is the coulomb.
- Electric current has dimensions $\mathbb{Q}\mathbb{T}^{-1}$. The SI assigns it a standard unit coul/s, also called “ampere” and abbreviated A.
- Energy has dimensions $\mathbb{M}\mathbb{L}^2\mathbb{T}^{-2}$. The SI assigns it a standard unit $\text{kg m}^2/\text{s}^2$, also called “joule” and abbreviated J.
- Power (energy per unit time) has dimensions $\mathbb{M}\mathbb{L}^2\mathbb{T}^{-3}$. The SI assigns it a standard unit $\text{kg m}^2/\text{s}^3$, also called “watt” and abbreviated W.

Suppose that you are asked on an exam to compute an electric current. You work hard and write down a formula made out of various given quantities. To check your work, write down the dimensions of each of the quantities in your answer, cancel whatever cancels, and make sure the result is $\mathbb{Q}\mathbb{T}^{-1}$. If it’s not, you may have forgotten to copy something from one step to the next. It’s easy, and it’s amazing how quickly you can spot and fix errors in this way.

When you multiply or divide two quantities, the dimensions combine like numerical factors: Photon flux irradiance ($\mathbb{T}^{-1}\mathbb{L}^{-2}$) times area (\mathbb{L}^2) has dimensions appropriate for a rate (\mathbb{T}^{-1}). On the other hand, you cannot add or subtract terms with different dimensions in a valid equation, any more than you can add rupees to centimeters. Equivalently, an equation of the form $X = Y$ cannot be valid if X and Y have different dimensions. (If either X or Y equals zero, however, then we may omit its units without ambiguity.)

You *can* add dollars to yuan, with the appropriate conversion factor, and similarly cubic centimeters to fluid ounces. Cubic centimeters and fluid ounces are different units that both have the same dimensions (\mathbb{L}^3). We can automate unit conversions, and reduce errors, if we restate the conversion $1 \text{ US fluid ounce} \approx 29.6 \text{ cm}^3$ in the form

$$1 \approx \frac{\text{US fluid ounce}}{29.6 \text{ cm}^3}.$$

Because we can freely insert a factor of 1 into any formula, we may introduce as many factors of the above expression as we need to cancel all the ounce units in that

³Some authors use $\mathbb{I} = \mathbb{Q}/\mathbb{T}$, a “current” dimension, instead of \mathbb{Q} .

expression. This simple prescription (“multiply or divide by 1 as needed to cancel unwanted units”) eliminates confusion about whether to place the numeric factor 29.6 in the numerator or denominator.

Functions applied to dimensional quantities

If $x = 1 \text{ m}$, then we understand expressions like $2\pi x$ (with dimensions \mathbb{L}), and even x^3 (with dimensions \mathbb{L}^3). But what about $\sin(x)$ or $\log_{10} x$? These expressions are meaningless;⁴ more precisely, they don’t transform in any simple multiplicative way when we change units, unlike say $x/26$ or x^2 .

Additional SI units

frequency: One hertz (Hz) equals one complete cycle per second, or $2\pi \text{ rad/s}$.

temperature: One kelvin (K) can be defined by saying that the atoms of an ideal monoatomic gas have mean kinetic energy $(3/2)k_{\text{B}}T$, where $k_{\text{B}} = 1.38 \cdot 10^{-23} \text{ J K}^{-1}$.

resistance and conductance: One ohm (Ω) equals one volt per ampere. One siemens is an inverse ohm: $1 \text{ S} = 1 \Omega^{-1}$.

electric potential: One volt (volt) equals 1 J/coul .

Traditional but non-SI units

mass: One dalton (also called “unified atomic mass unit,” and abbreviated u) is $1 \text{ Da} = 931.5 \text{ MeV}/c^2$.

time: One minute is 60 s , and so on.

length: One Ångstrom unit (Å) equals 0.1 nm .

volume: One liter (L) equals 10^{-3} m^3 . Thus, $1 \text{ mL} = 1 \text{ cm}^3$.

number density: A 1 M solution has a number density of $1 \text{ mole/L} = 1000 \text{ mole m}^{-3}$, where “mole” represents the number $\approx 6.02 \cdot 10^{23}$.

energy: An electron volt (eV) equals $e \times (1 \text{ volt}) = 1.60 \cdot 10^{-19} \text{ J} = 96 \text{ kJ/mole}$.

Here e is the electric charge on a proton. An erg (erg) equals 10^{-7} J . Thus, $1 \text{ kcal mole}^{-1} = 0.043 \text{ eV} = 6.9 \cdot 10^{-21} \text{ J} = 6.9 \cdot 10^{-14} \text{ erg} = 4.2 \text{ kJ mole}^{-1}$.

A.3 ABOUT GRAPHS

When you make a graph involving a continuous quantity, state the units of that quantity in the axis label. For example, if the axis label says **waiting time [s]**, then we understand that a point aligned with the tick mark labeled 2 represents a measured waiting time that, when divided by 1 s , yields the pure number 2.

The same interpretation applies to logarithmic axes. If the axis label says **flash photon density [photons/ μm^2]**, and the tick marks are unequal, then we understand that a point aligned with the first minor tick after the one labeled 10 represents a quantity that, when divided by the stated unit, yields the pure number 20 (in this case,

⁴One way to see why such expressions are meaningless is to use the Taylor series expansion of $\sin(x)$, and notice that it involves adding terms with incompatible units.

20 photons / μm^2). Alternatively, we can make an ordinary graph of the logarithm of a quantity x , indicating this in the axis label, which says $\log_{10} x$ or $\ln x$ instead of x . The disadvantage of the second system is that, if x carries units, then strictly speaking we must instead write something like $\log_{10}(x/(1 \text{ m}^2))$ or $\log_{10}(x \text{ [a.u.]})$, because the logarithm of a quantity with dimensions has no meaning.

A.3.1 Arbitrary units

Sometimes a quantity is given in some unknown or unstated unit. It may not be necessary to be more specific, but you should alert your reader by saying something like `emission spectrum [arbitrary units]`. Many authors abbreviate this as “[a.u.]”

A.3.2 Angles

Angles are dimensionless: We get the angle between two intersecting rays, in the dimensionless unit radians (abbrevated `rad`), by drawing a circular arc of any radius r between them and centered on the intersection, then dividing the length of that arc (with dimensions \mathbb{L}) by r (with dimensions \mathbb{L}). Another clue is that if θ carried dimensions, then trigonometric functions like sine and cosine wouldn't be defined (see Section A.2). The angle corresponding to a complete circle is $2\pi \text{ rad}$. An alternative expression for this quantity is `360 deg`.

Angular area (also called solid angle) is also dimensionless. Given a patch on the surface of a sphere, we get its angular area, in the dimensionless unit steradians (abbrevated `sr`), by finding the area of that patch and dividing by the sphere's radius squared.

A.4 PAYOFF

APPENDIX B

Global List of Symbols

Good notation should serve you—not the other way round.
—Howard Georgi

Throughout these notes the word “vector” is used specifically to mean a set of three numbers that *points* in space (or four numbers that point in spacetime). More abstract notions of vector, like the state vector of quantum mechanics, exist but don’t follow the particular transformation rules we use here.¹

B.1 MATHEMATICAL NOTATION

We need a notational system that is precise enough to express intricate ideas unambiguously, yet flexible enough to not be a burden when we know what we’re doing. If possible, we also want a system in which it’s *harder to write down wrong formulas than it is to write correct formulas*.

Abbreviated words

c.c. Complex conjugate of the preceding term(s).

$|_{\text{ret}}$ Evaluated at “retarded time” (observation time minus R/c); see Section 23.4.1.

Operations

$\|\vec{b}\|$ Length of a real 3-vector, $=\sqrt{\vec{b}\cdot\vec{b}}$. For a complex vector it means $\sqrt{\vec{b}^*\cdot\vec{b}}$.

z^* Complex conjugate of a complex number z .

$|z|$ Absolute value of a complex number, $=\sqrt{z^*z}$.

$\|\underline{X}\|^2$ Invariant norm-squared of a 4-vector.

∇^2 Laplace operator.

\square D’Alembert operator.

\star Hodge dual operation.

$\vec{a}\vec{b}$ dyad (tensor) product of two vectors (itself a rank-2 tensor). (Others may call it “tensor product” or “outer product,” and denote it by the symbol \otimes .)

$\vec{T}^{[S]}$ and $\vec{T}^{[A]}$ symmetric and antisymmetric parts of a rank-2 tensor (Equation 31.8, page 338).

¹A mathematician might therefore say “rank-1 tensor” wherever I say “vector” below.

Other modifiers

An overbar on a symbol can denote peak value (amplitude) of a sinusoidally varying quantity with the same letter name, for example, $f(t) = \bar{f} \cos(\omega t)$. More generally, such quantities may be complex; then $f(t) = \frac{1}{2} \bar{f} e^{-i\omega t} + \text{c.c.}$

Sometimes an overbar can instead be used to indicate the nondimensionalized version of some quantity.

A dot over a function name can mean a derivative with respect to time. A prime following a function name can mean a derivative with respect to a spatial coordinate. Primes have other uses, however; see below.

3-vectors and -tensors

$\vec{\nabla}$ Spatial gradient operator.

$\overleftrightarrow{\mathbb{1}}$ Unit operator regarded as a 3-tensor. Its components $\overleftrightarrow{\mathbb{1}}_{ij}$ are usually written as the “Kronecker delta” symbol: $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Most books use boldface type to denote 3-vectors and 3-tensors. I can’t draw that on the blackboard, so I use an arrow above the variable’s name to denote a 3-vector and a double arrow to denote a 3-tensor of rank two. Tensors of higher rank will always appear with explicit indices indicating their components and no arrow, for example, ε_{ijk} .

When a letter that is normally used for a vector appears without an overarrow or index, that notation usually refers to the *length* of the corresponding vector; for example, r indicates the length of \vec{r} . However, d^3r denotes $dx dy dz$ (which is not a vector).

A differential element of surface has area denoted $d^2\Sigma$. When multiplied by an outward-pointing perpendicular unit vector, this becomes the vector $d^2\vec{\Sigma}$.

If a 3-vector is normalized to unit length, it gets a hat (circumflex) instead of an arrow, for example, the coordinate basis vectors \hat{x} , \hat{y} , and \hat{z} .² These are constant unit vectors, but the radial unit vector $\hat{r} = \vec{r}/r$ is a vector *field*.

Most books drop the boldface or arrow when referring to the *components* of a vector or tensor, but I will retain it, to emphasize that these numbers have a particular transformation rule under change of coordinate system, for example, \vec{V}_i . Latin letters are used for 3D indices denoting components, especially i, j, \dots . 3D indices are always written as subscripts.³

When we have a collection of related vectors, for example, the positions of many particles, they may be distinguished by a subscript in parentheses, to avoid confusion with a vector component index. Thus $\vec{r}_{(\ell)}$ is the position of particle ℓ ; its x component is then $\vec{r}_{(\ell)1}$ and so on.

A few “alternate” versions of vector quantities will even get an upside-down hat (háček) instead of an arrow.

When a letter that is normally used for a rank-2 tensor appears without an overarrow, that may indicate that in this instance, the tensor is assumed to be an

²Some authors use the symbols $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$, or simply $\mathbf{i}, \mathbf{j}, \mathbf{k}$, to represent the unit vectors that these notes call $\hat{x}, \hat{y}, \hat{z}$.

³We would need to be more careful expressing tensors in curvilinear coordinates or on curved space, but these notes don’t do that.

overall scalar times the identity tensor. For example, an isotropic polarizability may be written as α , shorthand for $\alpha \vec{\mathbf{1}}$.

Tilde versus prime

Sometimes each member of a collection of vectors will be related to a corresponding member of another collection by a common operation, for example, a physical, or “active,” rotation. We may use the same symbol for each set to emphasize the correspondence, but distinguish the modified ones with a tilde: \tilde{V} in place of \vec{V} , or even $\tilde{r}_{(\ell)}$ in place of $\vec{r}_{(i)}$.

Primes will usually indicate a completely different concept. Sometimes we will express a *single* vector in terms of more than one coordinate system. Then the *components* (ordinary numbers) used to represent that vector will have two different forms, which we will write as \vec{V}_i , $i = 1, 2, 3$ and \vec{V}'_a , $a = 1, 2, 3$ respectively. In each case, we are referring to the *same* vector \vec{V} . What’s being rotated is the coordinate system, not \vec{V} , but this introduces a “passive” transformation on the components.

Occasionally, prime will instead be used to mean a derivative with respect to a spatial coordinate.

Similar remarks apply to higher-rank 3-tensors.

4-vectors and -tensors

Many books use no typographical signal to indicate 4-vectors and 4-tensors; I use an underbar, regardless of rank. As with 3-quantities, I’ll retain the bar even when referring to specific components, to emphasize that they have particular transformation rules under change of coordinate system, for example \underline{p}_μ . Greek letters are used for 4D indices denoting components, especially μ, ν, \dots . Subscript indices are distinct from superscript indices, as explained in Chapters 31–32.

Often, a 4D quantity has a name similar to that of the 3D quantity related to its spatial components.

When the same letter of the alphabet is used for both a 3-vector and a 4-vector, it is understood that the spatial part of the 4-vector is the same as the corresponding 3-vector in some coordinate system.

The usage of tilde (active) and prime (passive) is the same as for three-dimensional objects.

$\underline{\partial}$ Spacetime gradient operator.

$\binom{p}{q}$ Denotes the rank of a tensor with p upper and q lower indices.

Matrices

Matrices are set in sans-serif type, M. They are arrays of numbers that do not necessarily transform in the specific manner of tensors upon coordinate change.

$\mathbf{1}$ Unit matrix.

S 3D rotation matrix.

Λ 4D Lorentz transformation matrix.

Relations

- \sim Has the same dimensions as.
- \approx Is approximately equal to.

Miscellaneous

The usual square root of minus one is indicated in roman type (i) to distinguish it from say, an index. Some engineering texts instead use the letter j to represent this quantity. Some computer math systems instead refer to this quantity as I or as j. The other square root of minus one is then $-i$.

The base of natural logarithms is indicated in roman type (e) to distinguish it from the charge on a proton (e), a constant of nature.

The differential symbol is indicated in roman type (d) to distinguish it from any variable called d , which might denote a distance.

B.2 NAMED QUANTITIES

We have a lot of quantities, and only a limited number of letters of the alphabet, so inevitably some symbols will be overloaded with more than one meaning. Sometimes the meanings will be disambiguated by upper/lower case, or by tensor rank. In other cases, you just have to determine the desired meaning by context.

Latin alphabet

- a Size of a finite distribution of charge and/or current; thickness of a layer.
- \vec{A} Three-dimensional magnetic vector potential.
- \underline{A} Four-vector potential.
- b Generic name for a constant. \bar{b} , generic name for the amplitude of a sinusoidally-varying quantity.
- B_{ij} Shape operator for a 2D surface in 3-space.
- \vec{B} Magnetic induction (often called “magnetic field”) (a pseudovector); \check{B} , modified form, $= c\vec{B}$ (same units as electric field).
- c Speed of light in vacuum. c_s , speed of vibrations in a medium, e.g. a spring.
- c_e, c_{ion} , etc. Number density of electrons, ions, etc. [dimensions \mathbb{L}^{-3}].
- C Capacitance.
- \mathcal{C} Areal density of capacitance.
- D_{ion} Diffusion constant for some species of ions in solution.
- D_{r} Retarded green function for the D’Alembert operator.
- \vec{D} Electric displacement (analog of $\epsilon_0\vec{E}$ in a medium).
- $\vec{\mathcal{D}}_{\text{E}}$ Electric dipole moment. $\vec{\mathcal{D}}_{\text{E}}$, its quantum version.
- $\vec{\mathcal{D}}_{\text{M}}$ Magnetic dipole moment (a pseudovector); $\check{\mathcal{D}}_{\text{M}} = \vec{\mathcal{D}}_{\text{M}}/c$, modified form (with same units as electric dipole moment).
- $\hat{e}_{(i)}$ Basis of three mutually perpendicular, unit 3-vectors defined by a cartesian coordinate system.

- e Charge on a proton.
- \vec{E} Electric field.
- \mathcal{E} Energy. $\check{\mathcal{E}}$, specifically the relativistic energy when it is necessary to distinguish it from the newtonian quantity.
- F Linear tension, for example in a spring or along a 1D interface [dimensions of force] (Chapters 6,26).
- \mathcal{F} Helmholtz free energy.
- \underline{F} Faraday 4-tensor.
- g Conductance per area.
- G Conductance.
- G Gauss curvature of a surface in space.
- G_N Newton gravitation constant.
- \underline{g} Metric 4-tensor. In special relativity, this is a rank- $\binom{0}{2}$ tensor whose 16 components in any E-inertial coordinate system, $\underline{g}_{\mu\nu}$, are always the same numerical constants. The same letter \underline{g} can also be used to refer to the dual metric tensor, a rank- $\binom{2}{0}$ tensor whose 16 components in any E-inertial coordinate system, $\underline{g}^{\mu\nu}$, are the *same* numerical constants as those of $\underline{g}_{\mu\nu}$. The notation is unambiguous because applying the index-raising operation to the first version does yield the second one.
- \vec{h} Generic symbol for a distance, or specifically displacement (position) of an object relative to the origin of coordinates or other reference point.
- H Mean curvature of a surface in space.
- \vec{H} Magnetic intensity (analogous to \vec{B}/μ_0 but includes a medium) (a pseudovector).
- \mathbf{H} Hamiltonian operator (Chapter 55) [dimensions ML^2T^{-2}].
- I Electric current.
- $\vec{\mathcal{J}}$ Moment of inertia tensor of a rigid body.
- \vec{j} Electric charge flux (sometimes called “current density”); $j^{(1D)}$, one-dimensional version. \vec{j} , its quantum version.
- $\vec{j}_{\mathcal{E}}$ Flux of energy.
- j_{ion} Number flux of ions of some species [dimensions $\text{L}^{-2}\text{T}^{-1}$].
- \underline{J} Electric charge 4-flux (sometimes called “4-current”); \underline{j} , scalar analog sometimes used in simplified formulas.
- k Generic name for a Hooke-law spring constant.
- k_B Boltzmann constant; $k_B T$, thermal energy; $k_B T_r$, at room temperature.
- K Temporary name for relativistic energy/ c , later named \underline{p}^0 .
- $\vec{j}^{(2D)}$ Surface charge flux (sometimes called “surface current density”); $\vec{j}_{\mathbf{f}}^{(2D)}$, free surface charge flux.
- \vec{K} Hooke-law spring constant tensor.
- $\underline{K}^{\mu\nu}_{\lambda\sigma}$ Susceptibility operator (Section 49.6'b (page 510)).
- ℓ Generic index for enumeration, for example, a set of particles or elements of a continuous source. Can also indicate which of several ion species is under consideration.

- ℓ_B Bjerrum length.
- $\vec{\ell}$ Parametric representation of curve in space; $d\vec{\ell}$, small element.
- L Inductance.
- \vec{L} Angular momentum (a pseudovector).
- m Mass.
- m Generic 3-space index.
- \vec{M} Volume density of magnetic dipole moment (a pseudovector); $\check{M} = \vec{M}/c$, modified form (same units as \vec{P}).
- n Index of refraction.
- p Order of a multipole (called a “ 2^p -pole”), equal to the rank of the 3-tensor that specifies its moment. Rank of a generic 3-tensor.
- p Pressure.
- \vec{p} A particle’s 3-momentum. \vec{p} , specifically the relativistic momentum when it is necessary to distinguish it from the newtonian quantity.
- \underline{p} A particle’s 4-momentum.
- \vec{P} volume density of electric dipole moment (“polarization density”).
- \vec{P} momentum of electromagnetic field (Equation 55.8, page 537) [dimensions MLT^{-1}]. \hat{P} , corresponding quantum operator.
- \mathcal{P} Power.
- Prob Probability (a real, dimensionless quantity between 0 and 1). \wp , probability density (a nonnegative real function).
- \vec{p}_e electron momentum operator (Section 55.5.2, page 543) [dimensions MLT^{-1}].
- q Electric charge.
- Q, Q^\dagger lowering (destruction) and raising (creation) operators, respectively, for electromagnetic field (Equation 55.17, page 539) [dimensionless].
- \vec{Q}_E Electric quadrupole 3-tensor.
- r_c classical electron radius.
- \vec{r} Three-dimensional position vector, with cartesian components $\vec{r}_i = (x, y, z)^t$.
- \vec{r}_e Electron position [dimensions L]; \hat{r}_e , corresponding quantum operator (Equation 55.35, page 543).
- s Arclength parameter along a curve in 3-space.
- S A 3D rotation, or the 3×3 matrix representing it; S_{ij} , its explicit components.
- t Time, as measured in an inertial coordinate system (either G-inertial in newtonian physics or E-inertial in relativistic physics).
- T Interfacial surface tension (Chapter 6).
- \vec{T} Stress tensor.
- \vec{T} Generic name for a 3-tensor.
- \underline{T} Energy-momentum flux tensor (sometimes called “stress-energy tensor”).
- u, v Light-cone coordinates.
- u Energy density of electromagnetic field.
- u Displacement of a continuous spring.

- U Potential energy of a particle.
- \underline{U} Four-velocity. Its three spatial components are not equal to the components of ordinary velocity \vec{v} .
- v Velocity; that is, the time derivative of the position of a particle in an inertial coordinate system (either G-inertial in newtonian physics or E-inertial in relativistic physics). v_* , velocity of a Galilean or Lorentz boost. v_m , velocity of a material medium that supports waves (spring, water, æther, ...).
- V A region in 3-space, or its volume; ∂V , the boundary of V , that is a closed surface. An area element $d^2\vec{\Sigma}$ of that surface is conventionally taken to point outward. $\check{B} = c\vec{B}$ modified magnetic induction (same units as \vec{E}).
- x, y, z Right-handed cartesian coordinates of 3-space, or spatial components of a right-handed E-inertial coordinate system on spacetime.

Greek alphabet

- α Electric polarizability of a molecule or other small object; α_m , magnetic polarizability. $\vec{\alpha}$, polarizability tensor of an anisotropic object.
- β Cross-polarizability of a single chiral molecule.
- $\vec{\beta}$ Velocity of a particle divided by c .
- γ Abbreviation for $1/\sqrt{1-\beta^2}$.
- $\underline{\Gamma}(\xi)$ Parametric representation of a trajectory (curve in spacetime).
- $\vec{\Gamma}$ Alternate representation of the magnetic dipole moment as an antisymmetric rank-2 3-tensor.
- ϵ Dielectric permittivity of a medium; ϵ_0 , permittivity of vacuum. The dimensionless ratio ϵ/ϵ_0 is called the “dielectric constant,” but we don’t assign any symbol to it.
- ϵ_{ijk} Components of the 3D Levi-Civita tensor in a particular cartesian coordinate system (a pseudotensor).
- ϵ_{multi} Multipole parameter (Equation 42.5) [dimensionless].
- $\vec{\zeta}$ Polarization 3-vector for a plane EM wave; $\hat{\zeta}_{(1)}$, $\hat{\zeta}_{(2)}$, linear polarization basis (real); $\hat{\zeta}_{(+)}$, $\hat{\zeta}_{(-)}$, circular polarization basis (complex). $\hat{\zeta}_{(\alpha, \vec{k})}$ basis of unit polarization vectors ($\alpha = 1, 2$) for plane waves traveling along \vec{k} (Equation 55.2, page 536) [dimensionless].
- $\underline{\zeta}$ Polarization 4-vector.
- η Viscous drag coefficient for a particle in fluid.
- η Bulk cross-polarizability of a chiral material.
- η_i integers specifying a mode in a cavity (Section 55.2) [dimensionless].
- θ Polar angle in spherical polar coordinates [dimensionless].
- ϑ Angle between an incoming wave’s linear polarization and the line of sight to an observer.
- ϑ Velocity of neural action potential.
- κ Electric conductivity of a medium.
- κ Elastic stretch modulus of a continuous spring.

- λ Wavelength of a plane or spherical wave.
- λ_D Debye length.
- λ_{cable} Space constant of a nerve axon or other cable.
- Λ A Lorentz transformation linking two E-inertial coordinate systems, or the 4×4 matrix representing it; Λ^μ_ν , its explicit components.
- μ Magnetic permeability of a medium; μ_0 , permeability of vacuum.
- ν Frequency of a sinusoidally varying quantity (cycles per unit time) [dimensions \mathbb{T}^{-1}].
- ξ Generic parameter for a curve in space (not necessarily arclength) or spacetime (not necessarily proper time). $\vec{\xi}$, constant 3-vector used when constructing a dipole spherical wave.
- Ξ Gauge-transformation parameter.
- ρ Radial coordinate in cylindrical coordinates.
- ρ Generic symbol for volume density of a continuous scalar quantity; ρ_q , electric charge density [dimensions $\mathbb{Q}\mathbb{L}^{-3}$]; $\rho_\mathcal{E}$, energy density; ρ_m , mass density. $\rho_q^{(1D)}$, linear electric charge density (coul/m); $\rho_\mathcal{E}^{(1D)}$, linear energy density; $\rho_m^{(1D)}$, linear mass density (kg/m).
- σ Generic symbol for surface density of a scalar quantity; σ_q , surface charge density; σ_f , free surface charge density; σ_b , bound surface charge density.
- σ Scattering cross section.
- Σ A 2D surface, or its area; $d\vec{\Sigma}$, infinitesimal surface element, including a choice of perpendicular vector, that is, differential of area times the chosen unit vector. $\partial\Sigma$, boundary of a surface Σ , that is, a closed curve with a direction chosen by applying the right-hand rule to the chosen perpendicular.
- τ A particle's proper time; equivalently, proper time parameter along a trajectory in spacetime; equivalently, the time recorded by an imagined clock carried along with the particle. If the particle's trajectory is accelerating, then proper time will not agree with time t in any fixed E-inertial coordinate system.
- τ_{cable} Time constant of a nerve axon or other cable.
- Υ Rapidity parameter of a Lorentz boost.
- φ Azimuthal angle in either cylindrical or spherical polar coordinates.
- ϕ Phase shift of one sine function relative to another.
- ϕ_N Newtonian gravitational potential.
- $\Phi_{\vec{k},\omega}$ The complex function $e^{i(\vec{k}\cdot\vec{r}-\omega t)}$ (dimensionless).
- Φ_B Magnetic flux. $\check{\Phi}_B = \Phi_B/c$, modified version.
- χ_e Dielectric susceptibility (polarizability of an isotropic medium); χ_m , magnetic susceptibility (polarizability of an isotropic medium); $\check{\chi}_m$, modified form. For anisotropic media, these are replaced by tensors.
- ψ Scalar potential field, also called electric potential. In electrostatics, also called the electrostatic potential. $\bar{\psi}$, its dimensionless form (in static or quasistatic situations), $\tilde{\psi}$, amplitude of a potential varying sinusoidally in time. $\psi^{[p]}$, standard 2^p -pole potentials. ψ_{in} , potential inside a neuron; ψ_{out} , potential outside. ψ^{Nernst} ,

Nernst potential; ψ^0 , quasisteady resting potential; v , membrane potential relative to ψ^0 .

ω Angular frequency (radians per unit time).

ω_P Plasma frequency.

$\vec{\omega}$ Angular frequency of rigid body rotation, with direction corresponding to its axis of rotation via the right-hand rule (a pseudovector).

$\vec{\omega}$ Alternate representation of \vec{B} as an antisymmetric, rank-2 3-tensor.

Ω Solid angle (sometimes called angular area).

APPENDIX C

Numerical Values

If the model explains all the facts, then there's something wrong—because always some of the facts are wrong.

— Aharon Katchalsky

For salt solution at concentration 100 mM, $\kappa \approx 0.1 \Omega^{-1} \text{m}^{-1}$.
Liquid water: $\epsilon \approx 80\epsilon_0$ at low frequency.

C.1 FUNDAMENTAL CONSTANTS

Newtonian gravitation constant, $G_N \approx 6.7 \cdot 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$.

Planck constant (reduced), $\hbar \approx 1.05 \cdot 10^{-34} \text{ J s}$.

Proton charge, $e \approx 1.6 \cdot 10^{-19} \text{ coul}$. Electron charge is $-e$.

Electron mass, $m_e \approx 9.1 \cdot 10^{-31} \text{ kg}$.

Speed of light, $c \approx 3.0 \cdot 10^8 \text{ m/s}$.

Avogadro number, $N_{\text{mole}} \approx 6.02 \cdot 10^{23}$.

Boltzmann constant, $k_B \approx 1.38 \cdot 10^{-23} \text{ J K}^{-1}$. Typical thermal energy at room temperature $k_B T_r \approx 4.1 \text{ pN nm} \approx 4.1 \cdot 10^{-21} \text{ J} \approx 2.5 \text{ kJ mole}^{-1} \approx 0.59 \text{ kcal mole}^{-1} \approx 0.025 \text{ eV}$.

Permittivity of vacuum, $\epsilon_0 \approx 8.85 \cdot 10^{-12} \text{ coul}^2 \text{ N}^{-1} \text{ m}^{-2}$. Permeability of vacuum, $\mu_0 \approx 4\pi \cdot 10^{-7} \text{ m kg coul}^{-2}$.

C.2 OPTICS

C.2.1 Index of refraction for visible light

These approximate values neglect dispersion (dependence on wavelength).

Air at standard temperature and pressure: $n_{\text{air}} \approx 1.0003$. This book uses the approximate value 1, except when studying the mirage phenomenon; there, we use more precise values for light of wavelength 633 nm. At 30°C : $n_{\text{air}} \approx 1.00026$; at 50°C : $n_{\text{air}} \approx 1.00024$.

Water: $n_w \approx 1.33$.

Glass: 1.5–1.7. This book uses the illustrative value 1.52.

C.2.2 Miscellaneous

Earth's magnetic field strength at surface, approx $5 \cdot 10^{-5} \text{ T}$.

Earth radius $6.4 \cdot 10^6$ m.

Maximum energy of solar radiation per area at Earth surface: 1.4 kW/m^2 .

Mass of Sun $2.0 \cdot 10^{30}$ kg.

APPENDIX D

Formulas

Here are the bare bones of our topic. When you can derive all the formulas in this document, know the meanings of the symbols they contain, and know when they are applicable, then you'll know a lot of electrodynamics.

D.0 PROLOGUE

D.0.1 Vector calculus

See Section 0.2.

D.0.2 The Maxwell equations

$$\vec{\nabla} \cdot \vec{E} = \rho_q / \epsilon_0 \quad \text{Gauss} \quad (\text{D.1})$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{Gauss} \quad (\text{D.2})$$

$$\vec{\nabla} \times \vec{E} + \dot{\vec{B}} = 0 \quad \text{Faraday} \quad (\text{D.3})$$

$$\vec{\nabla} \times \vec{B} - \mu_0 \epsilon_0 \dot{\vec{E}} = \mu_0 \vec{j} \quad \text{Ampère.} \quad (\text{D.4})$$

The dots represent $\partial/\partial t$. The constants have numerical values $\mu_0 \approx 4\pi \cdot 10^{-7} \text{ m kg coul}^{-2}$ (the magnetic permeability of vacuum), and $\epsilon_0 \approx 8.85 \cdot 10^{-12} \text{ coul}^2 \text{ N}^{-1} \text{ m}^{-2}$ (the electric permittivity of vacuum). (For info about the units, see Chapter 15.)

See Section D.14.2 for definitions of the charge density ρ_q and charge flux \vec{j} .¹

The official name for \vec{E} is “electric field intensity”; \vec{B} is the “magnetic induction.” We'll just call them the electric and magnetic fields. Some formulas are neater when expressed in terms of a quantity I'll call $\vec{B} \equiv c\vec{B}$,² because this quantity has the same dimensions as \vec{E} .

D.0.3 Lorentz force law

The Lorentz force law is the equation of motion of a point charge:³

$$\frac{d}{dt} \vec{p} = q \left(\vec{E} + \vec{v} \times \vec{B} \right) + \vec{f}_{\text{other}}. \quad (\text{D.5})$$

¹Some authors call this quantity “current density.”

²We won't give this quantity any particular identifying phrase. (Confusingly, gaussian people call it “the magnetic induction,” even though it's defined as the magnetic induction times c .)

³A “point charge” is an idealization, having no (or negligible) multipole moments other than its total charge. Thus for some purposes even an electron cannot be regarded as a point charge, because it has a magnetic dipole moment! In classical electrodynamics, we assume that any charged macroscopic body can be regarded as a collection of point charges.

This time, the dot represents d/dt . The fields \vec{E} , \vec{B} are to be evaluated at some time t and at the position $\vec{r}(t)$ of the particle at that time; $\vec{v} = d\vec{r}/dt$ at that time. q and m are constants completely characterizing the point charge. \vec{f}_{other} represents any non-electromagnetic force acting on the charged bodies in the system.⁴ And the momentum $\vec{p} = m\vec{v}$, for velocities much smaller than 10^8 m/s. (See Equation D.43 later.)

A test body refers to a limiting case of a point object with charge and mass infinitesimally small, but q/m a finite constant. In practice, a test body is a point charge so small that it measures, but does not significantly perturb, surrounding fields set up by other charges.

D.1 NEWTONIAN GRAVITY

See Chapter 1.

D.2 ELECTROSTATICS

If all charges are at rest, then we may set time derivatives equal to zero and also the charge flux $\vec{j} = 0$. Then the magnetic field is zero and \vec{E} can be expressed as the gradient of an electric potential: $\vec{E} = -\vec{\nabla}\psi$. Warning: this useful trick doesn't work in the nonstatic case, where $\vec{\nabla} \times \vec{E} \neq 0$. But in the static case it's very helpful: Instead of solving three PDEs for the electric field, we only need to solve one PDE for the potential.

Confusing standard terminology: The electrostatic potential is *not* the potential energy of a test body. Instead, ψ determines each particle's potential energy by $U(\vec{r}) = q\psi(\vec{r})$. That's consistent with the Lorentz force law, which in electrostatics says $\vec{f} = -\vec{\nabla}(q\psi)$.⁵ Thus $\nabla^2\psi = -\rho_q/\epsilon_0$ in vacuum (the Poisson equation).⁶ Boundary conditions: $\vec{E}_{\parallel} = 0$ just outside a perfect conductor. $E_{\perp} = \hat{n} \cdot \vec{E}$ jumps from zero inside a perfect conductor to a value just outside that is related to the surface charge density σ .

D.3 ELECTROSTATIC MULTIPOLE EXPANSION

The field outside an arbitrary static charge distribution is complicated, but frequently we are in a special case where the answer can be simplified. Consider a charge distribution localized to a region of size a , viewed at a distance $r \gg a$ from a reference point located inside the distribution. We'd like an expression for the field that is organized as a power series expansion in a/r . Moreover, we'd like to display each term as the product of a factor characterizing the charge distribution (but independent of the

⁴Sometimes it's appropriate to instead introduce a constraint. For example, we may say "suppose that charge is fixed onto on a spinning disk."

⁵Warning: Later on, when we consider time-dependent fields we'll again introduce a scalar potential, but it will no longer have a simple interpretation as potential energy per charge.

⁶The special case $\rho_q = 0$ is sometimes called the Laplace equation.

observer's location), times a generic spatial dependence (i.e. one that is independent of the details of the charge distribution). The former factors are called the multipole moments of the distribution; the latter ones are the multipole potentials.

Place the origin of the coordinates at the chosen reference point. The field point (observer location) is \vec{r} , and we are assuming that the charge distribution $\rho_q(\vec{r}_*)$ is nonzero only for $(\vec{r}_*)^2 < a^2$, for some $a \ll r$. Then the leading terms of the expansion of $\psi(\vec{r})$ in powers of a/r can be written as

$$\psi(\vec{r}) = q_{\text{tot}}\psi^{[0]}(\vec{r}) + \vec{D}_E \cdot \vec{\psi}^{[1]}(\vec{r}) + \text{Tr} \left[\vec{Q}_E \cdot \vec{\psi}^{[2]}(\vec{r}) \right] + \mathcal{O}(a^3/r^4). \quad (\text{D.6})$$

In this formula,⁷

$$\psi^{[0]}(\vec{r}) = \frac{1}{4\pi\epsilon_0 r}; \quad \vec{\psi}^{[1]}(\vec{r}) = \frac{1}{4\pi\epsilon_0 r^2} \hat{r}; \quad \vec{\psi}^{[2]}(\vec{r}) = \frac{1}{8\pi\epsilon_0 r^3} (\hat{r}\hat{r} - \frac{1}{3}\vec{\mathbb{1}}). \quad (\text{D.7})$$

These formulas define a single monopole field, a set of three dipole fields, and a set of five independent quadrupole fields. (\vec{Q}_E and $\vec{\psi}^{[2]}$ are both symmetric and traceless, and hence each only has five independent entries.)

The corresponding moments are sets of numbers, not functions of \vec{r} :⁸

$$q_{\text{tot}} = \int d^3r_* \rho_q(\vec{r}_*), \quad \vec{D}_E = \int d^3r_* \rho_q(\vec{r}_*) \vec{r}_*, \quad \vec{Q}_E = 3 \int d^3r_* \rho_q(\vec{r}_*) (\vec{r}_* \vec{r}_* - \frac{1}{3}(r_*)^2 \vec{\mathbb{1}}). \quad (\text{D.8})$$

Corresponding to the multipole potentials, these formulas define a single monopole moment q_{tot} , three components of the dipole moment \vec{D}_E , and five independent components of the quadrupole moment \vec{Q}_E .

If the reference point is not at the origin of coordinates, but instead is at \vec{h} , then substitute $\vec{r} - \vec{h}$ for \vec{r} in Equation D.6. Notice that if we change our choice of the reference point inside the object, then in general all the multipole moments except q_{tot} will change, even though we did nothing to the charge distribution.⁹

Sometimes we know in advance that some moments must equal zero. For example, a spherically-symmetric charge distribution has all its multipole moments (other than q_{tot}) equal to zero (Birkhoff's theorem). You should prove this for the special case of the dipole moment and quadrupole tensor (see Section D.12.1).

Consider the operation that reflects the location of every charge through the reference point. If this operation leaves our charge distribution unchanged, then all the 2^p moments with p odd (dipole, octupole, ...) equal zero. If this operation, followed by also reversing the sign of every charge, leaves our charge distribution unchanged, then all the *even* moments (monopole, quadrupole, ...) equal zero.

⁷See Section D.12 for more about the tensor notation in the last formula. The $\vec{\mathbb{1}}$ terms in Equations D.7 and D.8 are redundant. You can omit either (but not both) without changing ψ . I included both to emphasize that: (a) The potential at order r^{-3} has a traceless character, no matter what the charge distribution (Equation D.6); and (b) the trace of the second moment of charge cannot contribute at all to the parts of the field that are of order r^{-3} or less (Equation D.8).

⁸Some authors move the factor $\frac{1}{2}$ into the definition of \vec{Q}_E ; others instead use the convention given here.

⁹More precisely, suppose that the first few moments are all zero. Then they, and the first nonzero moment, don't change when we change our choice of reference point. Higher moments do change, but in simple ways.

Equation D.6 gives the dipole part of the field strength as the gradient:

$$\vec{E}^{\text{dipole}} = -\vec{\nabla}(\vec{\psi}^{[1]} \cdot \vec{\mathcal{D}}_{\text{E}}) = \frac{1}{4\pi\epsilon_0 r^3} \left(3(\hat{r} \cdot \vec{\mathcal{D}}_{\text{E}})\hat{r} - \vec{\mathcal{D}}_{\text{E}} \right).$$

Notice that it falls off with distance like r^{-3} , faster than the Coulomb field from a single point charge. You should do a similar derivation to find the quadrupole part of the electric field outside a static distribution of charge.

Sometimes we will use a mathematical limit called “pure electric dipole,” which is two charges $\pm q$, separated by a , in the limit $a \rightarrow 0$, $q \rightarrow \infty$ holding $\mathcal{D}_{\text{E}} = qa$ fixed. In this limit, the distribution’s higher moments all approach zero, leaving only the dipole moment.

D.3.1 Electrostatic force and torque on a charge distribution

Consider a charge distribution, free to rotate or translate, but otherwise rigid. This charge distribution is placed, with its reference point at \vec{h} , in an externally generated potential ψ_{ext} . It is confined to a spatial region of size much smaller than the characteristic length scale of variation of ψ_{ext} . Then its potential energy due to the external field is:

$$U(\vec{h}) = \text{const.} + q_{\text{tot}}\psi_{\text{ext}}(\vec{h}) - \vec{\mathcal{D}}_{\text{E}} \cdot \vec{E}_{\text{ext}}(\vec{h}) + \dots$$

where $\vec{\mathcal{D}}_{\text{E}}$ is the distribution’s dipole moment. The constant depends on the details of the distribution, but not on the orientation nor the position \vec{h} . We can find the force and torque on the dipole by differentiating this expression with respect to rigid translations and rotations, respectively.

You should work out the potential energy of one rigid dipole in the presence of another rigid dipole (to leading order in $1/r$).

A charge distribution need not be rigid. For example, if the dipole moment of a charge distribution equals zero, nevertheless it may be **polarizable** by an external field. In this case our object will again feel a force when placed in a nonuniform electrostatic field, proportional to $\vec{\nabla}(\vec{E}^2)$ and also to the object’s polarizability.

D.4 CURVILINEAR COORDINATES

D.4.1 Laplace operator

Useful formulas for the Laplace operator: In plane polar coordinates,

$$\nabla^2\psi = r^{-1} \frac{\partial}{\partial r} \left(r \frac{\partial\psi}{\partial r} \right) + r^{-2} \frac{\partial^2\psi}{\partial\varphi^2}.$$

For cylindrical coordinates, we add $\frac{\partial^2\psi}{\partial z^2}$ and rename r as ρ . In spherical coordinates, you should work out

$$\nabla^2\psi = r^{-2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial\psi}{\partial r} \right) + \frac{1}{r^2 \sin\theta} \frac{\partial}{\partial\theta} \left(\sin\theta \frac{\partial\psi}{\partial\theta} \right) + \frac{1}{r^2 \sin^2\theta} \frac{\partial^2\psi}{\partial\varphi^2}.$$

To derive these and other useful formulas, we introduced a general trick. Suppose f, g are two functions on the plane. The gradient of f in cartesian components can be written as

$$\vec{\nabla} f = \mathbf{J} \begin{pmatrix} \partial f / \partial r \\ \partial f / \partial \varphi \end{pmatrix} \quad \text{where } \mathbf{J} = \begin{pmatrix} \partial r / \partial x & \partial \varphi / \partial x \\ \partial r / \partial y & \partial \varphi / \partial y \end{pmatrix}.$$

Then

$$\int dx dy \vec{\nabla} f \cdot \vec{\nabla} g = \int (r dr d\varphi) (\partial f / \partial r, \partial f / \partial \varphi) \mathbf{J}^t \mathbf{J} \begin{pmatrix} \partial g / \partial r \\ \partial g / \partial \varphi \end{pmatrix}.$$

Integrating both sides by parts gave us the formula for ∇^2 in plane polar coordinates, and a similar method works for other separable coordinate systems too.

D.5 CAPACITOR

The **capacitance** is $C = q / \Delta\psi$. The energy stored in a capacitor is then $Q^2 / (2C)$ or $(\Delta\psi)^2 C / 2$. Think about why C appears in the numerator of one of these expressions, but in the denominator of the other one.

Consider two parallel conducting plates separated by vacuum. Neglecting edge effects, in the region between the plates $\vec{E} = \hat{n} \rho_q / \epsilon_0$, where \hat{n} is perpendicular to the plates and ρ_q is the charge per area. Thus $C = \Sigma \epsilon_0 / a$ where Σ is the plate area and a is the separation.

Many insulating media can be modeled as **linear dielectrics**: we can forget about the medium if we just replace ϵ_0 by some other constant ϵ in the Gauss law (the **electric permittivity** of the medium). The ratio ϵ / ϵ_0 is usually > 1 and is called the *dielectric constant* of the material. Thus adding a dielectric to a capacitor typically increases its capacitance.

Consider a spherical shell of charge, for example the distribution of excess charge on a conducting sphere. You can find its potential energy by starting with zero charge, then incrementally adding little bits dq until you arrive at the desired total. Please work out this formula, even if you already know it.

D.5.1

A solid sphere of dielectric, of volume V , uniformly polarized, creates a pure dipole electrostatic field outside it of strength $\vec{p} = \vec{P} V$. Inside the sphere the field is uniform, $\vec{E}_{\text{in}} = -\vec{P} / (3\epsilon_0)$. Note that the latter formula is independent of the size of the sphere.

A spherical *cavity* inside an infinite, uniformly polarized medium again has uniform $\vec{E}_{\text{in}} = \vec{E}_{\infty} + \vec{P} / (3\epsilon_0)$.

Consider a medium consisting of “molecules” with density N and polarizability α . Some external free charges polarize the medium uniformly, creating an average field E inside it. We suppose that each “molecule” sits at the center of a roughly spherical cavity (true on average for a fluid), and responds to a local mean field E_{in} created by the external free charge plus all the other “molecules,” regarded as a continuous dielectric medium. Then that medium’s polarization is $\vec{P} = \frac{N\alpha}{1 - N\alpha / (3\epsilon_0)} \vec{E}$ (the **Clausius–Mossotti formula**).

D.6 ELECTROHYDROSTATICS

To first order in perpendicular displacement f , the total area change is the integral over the surface of $-2Hf$, where H is the mean curvature, a scalar function on the surface.

The volume occupied by one side grows, and the other side shrinks, by an amount proportional to the integral of f without any factor of curvature.

D.7

D.7.1 Continuity equation

We assume that the quantities q_ℓ are constants characterizing each particle, and that they are locally conserved even if a particle splits into two, merges with another, etc. Then Equations D.16–D.17) imply the **continuity equation**:

$$\vec{\nabla} \cdot \vec{j} + \frac{\partial}{\partial t} \rho_q = 0. \quad (\text{D.9})$$

This equation is kinematic, not dynamic: It reflects only the definitions of ρ_q and \vec{j} , and the assumption of local charge conservation. It's true regardless of whether the charged particle trajectories obey any equation of motion.

The Maxwell equations may appear to be overdetermined (eight differential equations in six unknown functions), but when we compute the divergence of Equations 0.3 and 0.4 we find that two of these six equations are redundant with Equations 0.1, 0.2, and D.9.

D.7.2 Quasi-static

In a resistive conducting material, like salt water, the charge flux can sometimes be taken to be **ohmic**: $\vec{j} = \kappa \vec{E}$.¹⁰ The constant κ is called the **conductivity**; its units are $\Omega^{-1} \text{m}^{-1}$.

The power dissipation density (watts per cubic meter) is j^2/κ or $E^2\kappa$. (Compare the freshman-physics formulas I^2R and $(\Delta\psi)^2/R$.)

The characteristic time scale for net charge density to dissipate (the **relaxation time**) is ϵ/κ where ϵ is the permittivity of the material.¹¹ For fields varying on much longer time scales than the relaxation time (the **quasistatic** case), $\rho_q \approx 0$ and so we have $\vec{\nabla} \cdot \vec{j} = 0$. Generally on such time scales the magnetic fields created by the currents are weak and may be neglected in Faraday's law; then we may usefully introduce the potential function, and $\nabla^2\psi = 0$, even though strictly speaking we're not static.

D.8 CELL MEMBRANE CAPACITANCE

[Not covered this year.]

¹⁰Here we assume that the conductivity is isotropic and spatially uniform in the medium.

¹¹For water at standard conditions and at zero frequency, the dielectric constant ϵ/ϵ_0 is about 80.

D.9 AQUEOUS SOLUTIONS

[Not covered this year.]

D.10 CABLE EQUATION

[Not covered this year.]

D.11 NERVE IMPULSES

[Not covered this year.]

D.12 3-TENSORS

The **dyad product** $\vec{r}\vec{r}$ is a 3×3 matrix of entries;¹² the entry in row i and column j is the product $\vec{r}_i\vec{r}_j$.¹³ Contrast the scalar product $\vec{r} \cdot \vec{r}$, which is a single number; omitting the dot means we instead intend the dyad product. Note that $\vec{a}\vec{b} \neq \vec{b}\vec{a}$ (each of these is the other's transpose).

More generally, a “three-tensor of rank two” can be defined by stating its components in a cartesian coordinate system, that is, via an indexed set of nine quantities $\overset{\leftrightarrow}{T}_{ij}$.¹⁴

The dot product $\overset{\leftrightarrow}{T} \cdot \vec{b}$ is the vector with i th component equal to $\overset{\leftrightarrow}{T}_{ij}\vec{b}_j$ etc. The **trace** of a tensor is defined in a cartesian coordinate system by $\text{Tr } \overset{\leftrightarrow}{T} = \overset{\leftrightarrow}{T}_{ii}$. If it's zero, we say the tensor is **traceless**. A second rank tensor may also be **symmetric**, **antisymmetric**, or neither. For example, $\vec{r}\vec{r}$ is symmetric, but $\vec{a}\vec{b} - \vec{b}\vec{a}$ is antisymmetric. These are invariant properties: If the components satisfy any of them in one cartesian coordinate system, then the same is also true in any other such system.

The **unit tensor** (or “identity tensor”) $\overset{\leftrightarrow}{\mathbf{1}}$ is defined in any cartesian coordinate system to have components $\overset{\leftrightarrow}{\mathbf{1}}_{ij} = \delta_{ij}$. Thus $\overset{\leftrightarrow}{\mathbf{1}} \cdot \vec{v} = \vec{v}$ for any vector \vec{v} .

D.12.1 Tensors and spherical symmetry

The moment of inertia involves the “second moment of mass” tensor:

$$\overset{\leftrightarrow}{J}_{ij} = \delta_{ij} \overset{\leftrightarrow}{\mu}_{kk}^{[2,m]} - \overset{\leftrightarrow}{\mu}_{ij}^{[2,m]} \quad \text{where} \quad \overset{\leftrightarrow}{\mu}_{ij}^{[2,m]} = \int d^3r_* \rho_m(\vec{r}_*) \vec{r}_{*i} \vec{r}_{*j}. \quad (\text{D.10})$$

[Note that the electric quadrupole tensor can be expressed as

$$\overset{\leftrightarrow}{Q}_E = 3\overset{\leftrightarrow}{\mu}^{[2,q]} - \overset{\leftrightarrow}{\mathbf{1}}(\text{Tr } \overset{\leftrightarrow}{\mu}^{[2,q]}), \quad (\text{D.11})$$

¹²Some authors write $\vec{r} \otimes \vec{r}$ for the dyad product and call it the “tensor product” or “outer product.”

¹³Most authors write the components of a vector \vec{r} as r_i (i.e., without the arrow). In these notes we are being ultra-explicit, so we keep the arrow to emphasize the vector status of \vec{r} .

¹⁴Later we'll also require that these quantities have the same transformations under rotation or inversion of the spatial axes as $\vec{r}\vec{r}$.

where $\vec{\mu}^{[2,q]}$ is the second moment of charge, analogous to Equation D.10.]

Suppose that the mass density ρ_m depends only on r . Carry out the angular integrals to show that the moment of inertia tensor for a spherically symmetric body equals a constant times the unit tensor; in this case, the angular momentum \vec{L} really is always parallel to the angular velocity $\vec{\omega}$.

D.13 3-TENSORS FROM HEAVEN

Higher-order tensors have more than two indices. No over-arrow convention is used to denote them. A particularly useful one is the 3D **Levi-Civita tensor**. In any right-handed cartesian coordinate system, we define the tensor whose components ε_{ijk} are the 27 quantities defined in Section 0.3 (page 7). Note the useful identities

$$\varepsilon_{ijk}\varepsilon_{ijk} = 6, \quad \varepsilon_{ijk}\varepsilon_{ijl} = 2\delta_{kl}, \quad \varepsilon_{ijk}\varepsilon_{ilm} = \delta_{jl}\delta_{km} - \delta_{jm}\delta_{kl}. \quad (\text{D.12})$$

Please show that these are consistent with each other, and that the last one implies the triple cross product formula. Then derive the formula Equation D.10 for the moment of inertia tensor, starting from the relation between the angular frequency vector and the velocity of an element of a rigid body.

Any second-rank 3-tensor \vec{T} can be separated into its symmetric and antisymmetric parts. There is a useful decomposition:

$$\vec{T}_{ij} = \vec{S}_{ij} + \vec{\Omega}_{ij}, \quad \text{where } \vec{S}_{ij} = \frac{1}{2}(\vec{T}_{ij} + \vec{T}_{ji}) \text{ and } \vec{\Omega}_{ij} = \frac{1}{2}(\vec{T}_{ij} - \vec{T}_{ji}) = \frac{1}{2}\varepsilon_{ijk}(\varepsilon_{klm}\vec{T}_{lm}).$$

Suppose that \vec{T} is symmetric; that is, $\vec{T}_{ij} = \vec{T}_{ji}$. Then the last expression above is zero, because the contraction of \vec{T} on both its indices with any *antisymmetric* tensor must equal zero.

D.14 MAGNETOSTATICS

In this case we allow nonzero but steady currents. That is, we don't require *static* sources (invariant under both time translation and time reflection), but we do require that they be *stationary* (invariant under time translation only). In this case \vec{B} is divergenceless, but does not have vanishing curl in regions where there is current. Nevertheless we can usefully introduce a potential: The Poincaré lemma says that if a three-dimensional vector field is divergence free, then locally it can be written as a curl: $\vec{B} = \vec{\nabla} \times \vec{A}$.

D.14.1 Poincaré lemma

If $\vec{\omega}_{jk}$ is an antisymmetric tensor field, on a contractible region of space, satisfying

$$\vec{\nabla}_i \vec{\omega}_{jk} \pm (\text{all permutations of } i, j, k) = 0, \quad (\text{D.13})$$

then there's a vector field \vec{A}_i such that

$$\vec{\omega}_{ij} = \vec{\nabla}_i \vec{A}_j \pm (\text{all permutations of } i, j). \quad (\text{D.14})$$

Formulated in this way, the Poincaré Lemma is valid in any number D of dimensions, and analogous statements also hold for antisymmetric tensors of any rank p :

- a. In electrostatics, we choose $D = 3$, $p = 1$. In this case there's only one index in Equation D.14, so the lemma says that any curl-free \vec{E} is the gradient of some scalar function (which we have named $-\psi$).
- b. In magnetostatics, we choose $D = 3$, $p = 2$: The tensor field

$$\vec{\omega}_{ij} = \frac{1}{2}\varepsilon_{ijk}\vec{B}_k \quad \text{magnetic field tensor} \quad (\text{D.15})$$

satisfies Equation D.13, because the left side equals $\vec{\nabla}_i\vec{\omega}_{jk} + \vec{\nabla}_j\vec{\omega}_{ki} + \vec{\nabla}_k\vec{\omega}_{ij}$, and this is essentially $\vec{\nabla} \cdot \vec{B}$. So the lemma says that $\vec{\omega}$ may be written as the antisymmetrized derivative of some vector field (which we call \vec{A}). This in turn implies that \vec{B} is the curl of \vec{A} .

- c. Coming up: In time-dependent situations \vec{E} is not curl-free, but nevertheless we'll find we can apply the Poincaré lemma with $D = 4$, $p = 2$ to construct a *4-vector potential*.

D.14.1.1 For theory enthusiasts only a. Try rephrasing the Lorentz force law using $\vec{\omega}$, not \vec{B} . What Good Thing happens?

- b. The Poincaré lemma is valid even in curved space, so it will remain a useful tool when you study General Relativity.
- c. It's also valid for totally antisymmetric tensors of any rank (not just 1 or 2). Mathematicians refer to totally antisymmetric tensors as **differential forms** or **p-forms** where p is the rank. The antisymmetrized derivative is called the **exterior derivative**; if it equals zero, the form is called **closed**. If the form can itself be written as the exterior derivative of something, it's called **exact**. The Poincaré lemma states that

If a form is closed over some contractible region of space, then it is exact on that region.

In D dimensions there is always a special Levi-Civita tensor of rank D . Using it, we can convert back and forth between a p -form and a corresponding $(D - p)$ -form. The two forms are called each others' **Hodge duals**, and the transformation is the **Hodge dual** (or "Hodge star") operation.¹⁵ For example, the 1-form \vec{B} is dual to the 2-form $\vec{\omega}$ via Equation D.15.

Reexpressing a set of physics formulas in terms of a Hodge dual quantity sometimes clarifies its symmetry. For example, the parity invariance of the Maxwell/Lorentz system is obscured when we express it in terms of \vec{B} , but manifest when we use $\vec{\omega}$ instead.

- d. Quantum mechanics attributes a more fundamental role to the vector potential than does classical physics. Classical electrodynamics regards the vector potential as merely a convenient way to solve certain problems involving the "real" electric and magnetic fields, and gauge invariance as merely a redundancy of this description. But in quantum mechanics, the interaction of a charged particle with EM fields can *only* be formulated using the vector potential. Experimental signatures such as the Bohm–Aharonov effect, quantization of magnetic flux, etc. attest to the reality of the vector potential.

¹⁵The Levi-Civita tensor, and the Hodge dual operation, are rotation-invariant but not invariant under spatial inversions ("parity" transformations). Thus strictly speaking they are not defined until we specify an "handedness" convention on space (or spacetime).

D.14.2 Charge and flux

If the field sources consist of point particles with charges $\{q_\ell\}$, moving along known trajectories $\vec{r}_\ell(t)$, then the **charge density** is the function

$$\rho_q(t, \vec{r}) = \sum_{\ell} q_\ell \delta^{(3)}(\vec{r} - \vec{r}_\ell(t)) \quad (\text{D.16})$$

and the **charge flux** is the vector field¹⁶

$$\vec{j}(t, \vec{r}) = \sum_{\ell} q_\ell \dot{\vec{r}}_\ell(t) \delta^{(3)}(\vec{r} - \vec{r}_\ell(t)). \quad (\text{D.17})$$

The delta function $\delta(x)$ has the property that it equals one when integrated over any region of x that contains $x = 0$. The symbol $\delta^{(3)}(\vec{r})$ means the product of three delta functions, one for each component of \vec{r} . So the integral of ρ_q over a region of space, at a fixed time, is the total charge in that region at that time. And the integral of $\int d\vec{\Sigma} \cdot \vec{j}$ over a surface in space, at fixed time, is the net electric current passing through that surface at that time in the direction specified by the perpendicular direction we chose for $d\vec{\Sigma}$.

The Lorentz force law has a differential form for the force acting on a small volume d^3r of a continuous charge distribution:

$$d\vec{f} = (\rho_q \vec{E} + \vec{j} \times \vec{B}) d^3r.$$

Somewhere between the realistic case of a spread-out distribution and the idealization of a point charge there lies an intermediate idealization called the “thin wire.” Here we imagine a total current I (units A) to be trapped on a mathematical curve in space, $\vec{\ell}(s)$. The steady-current condition implies that I is uniform along that curve. Then the charge flux is

$$\vec{j}(\vec{r}) = I \oint d\vec{\ell} \delta^{(3)}(\vec{r} - \vec{\ell}). \quad (\text{D.18})$$

You should check how the units work in this formula.

D.14.3 Gauge freedom; fundamental solutions

There is still an ambiguity (**gauge invariance**) in this description of the fields: The magnetic field is unchanged after a **gauge transformation** $\vec{A} = \vec{A} + \vec{\nabla}\Xi$, where $\Xi(\vec{r})$ is any function. The freedom to transform \vec{A} in this way allows us to impose a subsidiary condition called **Coulomb gauge**:¹⁷ $\vec{\nabla} \cdot \vec{A} = 0$. In Coulomb gauge, for a stationary current distribution, Ampère’s law becomes three independent copies of the Poisson equation, which we’ve already solved:

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \int \frac{\vec{j}(\vec{r}_*) d^3r_*}{\|\vec{r} - \vec{r}_*\|}. \quad (\text{D.19})$$

¹⁶Some authors call this quantity “current density,” but that name might lead us to think that it’s a quantity with units like A/m³. Instead Equation D.16 shows that its units are A/m². Its relation to ordinary current I is that you get the current passing through a surface by integrating \vec{j} over that surface.

¹⁷Other “gauge choices” are possible, but this one is the standard choice for magnetostatic problems.

Suppose that a current I is confined to a thin wire following a closed curve $\vec{\ell}(s)$, where s is a parameter. Then Equations D.19 and D.18 yield the **Biot-Savart formula**:

$$\vec{B}(\vec{r}) = \frac{\mu_0 I}{4\pi} \oint d\vec{\ell} \times \frac{\vec{r} - \vec{\ell}}{\|\vec{r} - \vec{\ell}\|^3}.$$

In this formula, $\oint d\vec{\ell}$ is short for $\int_0^{s_{\text{tot}}} ds (d\vec{\ell}/ds)$ around the loop; $\vec{\ell}$ is short for $\vec{\ell}(s)$. If we parameterize the curve by arclength, then $d\vec{\ell}$ can also be regarded as $\hat{t}(s)ds$, where $\hat{t}(s)$ is the unit tangent vector to the curve at s .

D.15 UNITS

See Chapter 15.

D.16 MAGNETIC MULTIPOLE EXPANSION

Consider a current distribution localized to a region of size d , viewed at a distance $r \gg d$ from a reference point inside the distribution. We take the reference point to be at the origin of coordinates; then the 3-vector potential created by the current distribution is

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi r^3} \vec{r} \cdot \vec{\Gamma} + \mathcal{O}(r^{-3}), \quad (\text{D.20})$$

where the **magnetic dipole moment tensor** is defined by

$$\vec{\Gamma} = \int d^3r_* \vec{r}_* \vec{j}(\vec{r}_*). \quad (\text{D.21})$$

Section 16.1 proved that $\vec{\Gamma}$ is antisymmetric, so it contains the same information as the more traditional **magnetic dipole moment vector**, defined as

$$\vec{\mathcal{D}}_M = \frac{1}{2} \int d^3r_* (\vec{r}_* \times \vec{j}(\vec{r}_*)). \quad (\text{D.22})$$

Similarly to Section D.3, the magnetic dipole moment doesn't depend on our choice of a reference point.¹⁸ In terms of this, Equation D.20 becomes

$$\vec{A}(\vec{r}) = \frac{\mu_0}{4\pi} \frac{\vec{\mathcal{D}}_M \times \hat{r}}{\|\vec{r}\|^2} + \mathcal{O}(r^{-3}). \quad (\text{D.23})$$

Either formulation (Equations D.20 or D.23)) accomplishes the same sort of goal as the electrostatic multipole expansion: The far potential is expressed as a linear combination of three universal dipole vector potentials (functions of observer position \vec{r} , independent of the character of the source). The coefficients of the linear combination (the components of $\vec{\mathcal{D}}_M$ or $\vec{\Gamma}$) summarize the source; they don't depend on the observer.

Had we kept the higher terms in the Taylor expansion, we'd have found that they involve other moments of the current distribution (magnetic quadrupole and so on).

¹⁸Also similarly to the electrostatic case, higher moments may depend on this choice; more precisely, only the first nonzero moment is unambiguously defined.

D.16.1 More about magnetic dipoles

For example, for a circular current loop spanning area Σ we find $\vec{\mathcal{D}}_M = I\hat{n}\Sigma$. Here \hat{n} is the unit vector perpendicular to the surface spanning the loop, with its sign chosen by applying the right-hand rule to the direction of the current.

Sometimes we will use a mathematical limit called “pure magnetic dipole,” which imagines a circular loop of area Σ , carrying current I , in the limit $\Sigma \rightarrow 0$, $I \rightarrow \infty$ holding $\vec{\mathcal{D}}_M = I\vec{\Sigma}$ fixed. In this limit, Equation D.23 is exact.

Some formulas will look nicer when expressed in terms of the quantity $\check{\mathcal{D}}_M \equiv \vec{\mathcal{D}}_M/c$, because this quantity has the same units as the electric dipole moment.

Consider a current distribution, free to rotate or translate, but otherwise rigid.¹⁹ This current distribution is placed in an externally generated field \vec{B}^{ext} . It is much smaller than the characteristic length scale of variation of \vec{B}^{ext} . Then the net force on the distribution is $\vec{\nabla}(\vec{\mathcal{D}}_M \cdot \vec{B}^{\text{ext}})$ (plus subleading terms). The net torque, taken about the chosen reference point, is $\vec{\tau} = \vec{\mathcal{D}}_M \times \vec{B}^{\text{ext}}$ (plus subleading terms). These quantities can be expressed as derivatives of an “effective potential energy”²⁰

$$U_{\text{eff}} = \text{const.} - \vec{\mathcal{D}}_M \cdot \vec{B}^{\text{ext}} + \dots \quad (\text{D.24})$$

[Theory enthusiasts should find a formula equivalent to Equation D.24 but expressed in terms of $\vec{\omega}^{\text{ext}}$ (Equation D.15) and $\vec{\Gamma}$ (Equation D.21) instead of \vec{B}^{ext} and $\vec{\mathcal{D}}_M$. No Levi-Civita tensors will appear.]

D.16.2 Magnetic polarizability

If an object does not have a permanent magnetic dipole moment, it may nevertheless be magnetically polarizable. That is, it develops an induced moment $\vec{\mathcal{D}}_M \propto \vec{B}$, so the translational force is a constant times $\vec{\nabla}\|\vec{B}\|^2$. A bulk material consisting of such objects develops a magnetic dipole moment density $\vec{M}(\vec{r})$. If it's directed parallel to \vec{B} we call the material **paramagnetic**. If \vec{M} points opposite to \vec{B} the material is called **diamagnetic**. (Anisotropic response is also possible, in which \vec{M} is neither parallel nor antiparallel to \vec{B} .) E.g. a paramagnetic bead will be pulled toward a region of strong magnetic field.

A **ferromagnetic** material has nonzero net magnetic polarization even in the absence of external field.

D.17 NONSTATIC FIELDS

We now consider fields and sources that are time-dependent. One warmup problem concerned a helical coil of wire (solenoid). Faraday's law implies an induced electric field that opposes changes in current through the coil. To raise the current from zero to I requires that we do work to overcome this induced field, storing energy

¹⁹In particular, some fixed-current source maintains the magnitude of the current constant, regardless how we move the distribution.

²⁰But the sense in which U_{eff} is “really” a potential energy is subtle; see P+S p. 290 or Feynman II §§15-1–15-2.

$\mathcal{E}/\text{volume} = \frac{1}{2\mu_0} \|\vec{B}\|^2$. That result seems to attribute energy to the *empty space* inside the coil, because it's proportional to the volume of that space. It's reminiscent of the energy of a charged capacitor, which was also proportional to volume and to field strength squared.

D.17.1 Potentials, nonstatic case

Again we may always write the magnetic field as $\vec{B} = \vec{\nabla} \times \vec{A}$. But now $\vec{E} = -\dot{\vec{A}} - \vec{\nabla}\psi$. We call ψ the **scalar potential**, no longer the “electrostatic potential”; it no longer has a direct relationship to any potential energy of a test particle. For example, in a changing magnetic field there can be a net “electromotive force” around a loop; if such a force is present, it is clearly not conservative.

There is again a redundancy in the description of the fields by potentials: The electric and magnetic fields are unchanged after a **gauge transformation**

$$\vec{A} \rightarrow \vec{A} + \vec{\nabla}\Xi, \quad \psi \rightarrow \psi - \dot{\Xi},$$

where now $\Xi(t, \vec{r})$ is any function of space and time. The freedom to transform \vec{A}, ψ in this way allows us again to insist on **Coulomb gauge**, $\vec{\nabla} \cdot \vec{A} \equiv 0$.

In a region with no charges, we can go further, supplementing Coulomb gauge by the extra condition $\psi \equiv 0$. With these conditions, the Maxwell equations in vacuum reduce to a single vector equation: As always working in cartesian coordinates, it's just three independent copies of the **wave equation**:

$$\left(\nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) X = 0, \quad (\text{D.25})$$

where X represents \vec{A}_1, \vec{A}_2 , or \vec{A}_3 .

D.17.2 Plane wave solutions

We studied Maxwell in vacuum. We considered the Coulomb-gauge potentials

$$\psi \equiv 0; \quad \vec{A}(t, \vec{r}) = \frac{1}{2} [\vec{\zeta} e^{-i(\omega t - \vec{k} \cdot \vec{r})} + \text{c.c.}] \quad (\text{D.26})$$

where \vec{k} (the **wavevector**), ω (the **angular frequency**) and $\vec{\zeta}$ (the **polarization vector**) are constants.²¹ In particular, $\vec{\zeta}$ is in general a vector with *complex* entries. If it satisfies $\vec{\zeta} \cdot \vec{k} = 0$, and if $\omega = c\|\vec{k}\|$, then these potentials are in Coulomb gauge and satisfy Maxwell's equations in vacuum. Physically, this represents a wave traveling along \vec{k} , with \vec{E} and \vec{B} both perpendicular to \vec{k} , and to each other. More precisely, $\vec{E} \times \vec{B}$ points along \vec{k} , the direction of wave propagation.

D.17.3 Polarization of waves

For each \vec{k} , Equation D.26 gives two independent modes (**polarizations**), which can be expressed in various bases, e.g. via linearly or circularly polarized mode expansions.

²¹So the **wavelength** is $\lambda = 2\pi/\|\vec{k}\|$; the **frequency** is $\nu = \omega/2\pi$; the **phase velocity** is $\omega/\|\vec{k}\|$, which equals c in vacuum.

The wave is linearly polarized if $\vec{\zeta}$ is real (or an overall complex constant times a real vector). Otherwise the wave is circularly or elliptically polarized.

Starting from a particular \vec{k} , choose a pair of real unit vectors $\hat{\zeta}_{(1)}, \hat{\zeta}_{(2)}$ perpendicular to \vec{k} and forming a right-handed triad with it. That is, $\hat{\zeta}_{(1)} \times \hat{\zeta}_{(2)} = \hat{k}$. Then let $\hat{\zeta}_{(\pm)} = (\hat{\zeta}_{(1)} \pm i\hat{\zeta}_{(2)})/\sqrt{2}$ (the **helicity basis**). Any polarization vector $\vec{\zeta}$ can be written as a (complex) linear combination of $\hat{\zeta}_{(1,2)}$, or of $\hat{\zeta}_{(\pm)}$. If the polarization vector is purely along $\hat{\zeta}_{(+)}$, then the wave is said to be **circularly polarized with positive helicity**, and similarly for a pure $\hat{\zeta}_{(-)}$ wave (which is **negative helicity**).²²

If we choose a different pair of unit vectors $\hat{\zeta}'_{(1)}, \hat{\zeta}'_{(2)}$, then we get essentially the same helicity basis. That is, $\hat{\zeta}'_{(+)}$ is a complex constant times $\hat{\zeta}_{(+)}$ and similarly for $\hat{\zeta}'_{(-)}$. Note that $\hat{k} \times \hat{\zeta}_{(\pm)} = \mp i\hat{\zeta}_{(\pm)}$; the helicity basis vectors are eigenvectors of this operation, which represents infinitesimal rotation about \hat{k} . Also $\hat{\zeta}_{(\pm)} \cdot \hat{\zeta}_{(\pm)}^* = 1$ and $\hat{\zeta}_{(\pm)} \cdot \hat{\zeta}_{(\mp)}^* = 0$.

D.18 ENERGY AND MOMENTUM TRANSPORT BY WAVES

Electromagnetic waves can carry energy and momentum across empty space. We are not ready to discuss this in full generality yet, but I at least showed that a plane wave carries energy and momentum directed along its propagation vector, and that the corresponding fluxes are proportional to the field amplitudes squared. (Compare water or sound waves, which transport energy but not momentum.) We didn't find the constant of proportionality yet, because all I did was to study the absorption of energy and momentum by a single test particle. However, I did show a relation between the fluxes: (energy flux) = (momentum flux) $\times c$.

Quantum mechanics reinterprets light as a stream of packets called photons. In order to have Maxwell electrodynamics as its classical limit, each of those photons must have the same relation: (energy of a packet) = (momentum of a packet) $\times c$. Interestingly, this formula disagrees with what newtonian physics would have said by a factor of two: the newtonian answer is $\mathcal{E} = \frac{1}{2}mv^2 = (\frac{1}{2})(mv)(v)$.

D.19 RAY OPTICS

[Not covered this year.]

D.20 DIFFRACTION

[Not covered this year.]

²²Sadly there is little agreement among authors about which helicity is which. You must read closely to find what convention is being used (and write clearly which one you are using).

D.21 RAINBOWS

[Not covered this year.]

D.22 PARTIAL POLARIZATION

[Not covered this year.]

D.23 GENERATION OF RADIATION

When currents are present, but the net charge density is zero, then each component of the vector potential in Coulomb gauge obeys a modified form of Equation D.25 (the **inhomogeneous wave equation**), in which the right hand side is replaced by $-\mu_0 \vec{j}$. We found that, for any “source function” $\mathcal{J}(t, \vec{r})$, the expression

$$\phi(t, \vec{r}) = \int d^3 r_* \frac{1}{\|\vec{r} - \vec{r}_*\|} \mathcal{J}(t - \|\vec{r} - \vec{r}_*\|/c, \vec{r}_*) \quad (\text{D.27})$$

is a solution to a simplified form of the wave equation.

D.24 GALILEAN RELATIVITY

D.24.1 Rotations

A trajectory is a chain of events in spacetime. A rotation leaves time unchanged but moves to a new point in space (**active viewpoint**), or alternatively leaves the time axis unchanged but changes the coordinate system by altering the space axes (**passive viewpoint**).

D.24.2 Active viewpoint

A point P with coordinates \vec{r} moves to a new point Q with coordinates \vec{r}' related to the others via $\vec{r}'_i = S_{ij}^{-1} \vec{r}_j$.

Given a trajectory $\vec{r}(t)$ we construct a new, different trajectory by subjecting \vec{r} at every instant to a rotation. Suppose that S_{ij} is an **orthogonal matrix** (that is, $S^t S = \vec{\mathbf{1}}$). Then the new trajectory $\vec{r}'_i(t)$ solves Newton’s laws iff $\vec{r}_j(t)$ does. We say newtonian physics has a **symmetry** under this **active transformation**.²³ The particular property of symmetry under rotations is called **isotropy**.

D.24.3 Passive viewpoint

Equivalently, we can take a trajectory $\vec{r}(t)$ and reexpress the *same* trajectory in a new coordinate system on space. In the new system the coordinates of a point P are related to the coordinates of the *same point* via $\vec{r}'(P)_a = S_{aj} \vec{r}_j(P)$. Applying this

²³If there are many particles, we must apply the same rotation to each one’s trajectory.

coordinate transformation point by point to a trajectory, specified by three functions of time $\vec{r}_i(t)$, yields three new functions $\vec{r}'_a(t)$.

We can also reexpress the equations of motion obeyed by the original three functions in terms of the new coordinates. Suppose that \mathbf{S} is an orthogonal matrix. Then we find that the new equations have exactly the same form as the old ones, including the numerical values of any constants of Nature (e.g. Newton's gravitational constant). We say that newtonian physics has an **invariance** (or “is form-invariant”) under this **passive transformation**.

Symmetry and invariance are two viewpoints for saying the same thing.

The orthogonal 3×3 matrices form a **group** named $O(3)$. They include rotations and reflections. They *don't* include squashing (rescaling just one coordinate), nor even dilatation (rescaling all coordinates by the same amount).

D.24.4 Boosts

In fact, newtonian physics has symmetries (or equivalently invariances) corresponding to space rotation, space translation, time translation, and **galilean boosts** (plus some reflections); its full symmetry group is the **galilean group**, which has 10 continuous parameters. In the passive language, a galilean boost is a relabeling of the points of spacetime (called **events**) by four new numbers related to (t, \vec{r}) by

$$t' = t, \quad \vec{r}' = \vec{r} - \vec{v}_* t,$$

where \vec{v}_* is a constant vector. More general galilean transformations take the form

$$t' = t + B, \quad \vec{r}'_a = S_{aj} \vec{r}_j + \vec{A}_a - \vec{v}_{*a} t.$$

Here B , \vec{A} , and \vec{v}_* are constants, and \mathbf{S} is an orthogonal matrix.

When we perform one galilean transformation $(B_1, \mathbf{S}_1, \vec{A}_1, \vec{v}_{*1})$, then follow it by another $(B_2, \mathbf{S}_2, \vec{A}_2, \vec{v}_{*2})$, the net result is again of the same form, with:

$$B_{\text{tot}} = B_1 + B_2, \quad \mathbf{S}_{\text{tot}} = \mathbf{S}_2 \mathbf{S}_1, \quad \vec{A}_{\text{tot}} = \vec{A}_2 - \vec{v}_{*1} B_1 + \mathbf{S}_2 \vec{A}_1, \quad \vec{v}_{*,\text{tot}} = \vec{v}_{*2} + \mathbf{S}_2 \vec{v}_{*1}.$$

In the special case of pure boosts, this reduces to $\vec{v}_{*,\text{tot}} = \vec{v}_{*2} + \vec{v}_{*1}$, an unsurprising result (“galilean velocity addition formula”).

D.25 STRINGS/SPRINGS

Consider the dynamics of a 1-dimensional continuum elastic object (a coiled spring or “slinky”), allowed to move in one dimension. It's characterized by two material parameters: the linear mass density $\rho_m^{(1D)} \sim \text{kg/m}$, and an elastic constant $\kappa \sim \text{N}$. Thus a segment of length L has mass $\rho_m^{(1D)} L$ and spring constant κ/L . We let $u(t, x)$ denote the longitudinal displacement at time t of an element that in equilibrium would have been at position x . Then the equation of motion is

$$\frac{\partial^2 u}{\partial t^2} - (c_s)^2 \frac{\partial^2 u}{\partial x^2} = 0. \quad (\text{D.28})$$

In this formula the constant $c_s = \sqrt{\kappa/\rho_m^{(1D)}}$ is the speed of waves in the spring. Essentially the same equation also describes transverse waves of small amplitude along a tense string. Generalizing to 2 or 3 dimensions gives waves in a network of springs.

One can find local expressions for the kinetic and potential energy densities, and their sum $\rho_{\mathcal{E}}^{(1D)}(t, x)$. There is also a 1D flux of energy $j_{\mathcal{E}}^{(1D)}(t, x)$, which is the net rate at which the region to the left of x is doing work on the region to the right. The local quantities $\rho_{\mathcal{E}}^{(1D)}(t, x)$ and $j_{\mathcal{E}}^{(1D)}(t, x)$ obey a continuity equation, so we say that energy is locally conserved.

The slinky wave equation doesn't appear to be galilean invariant, but that's because we implicitly assumed that the undisturbed slinky is at rest with respect to the observer. Really, the slinky system has another dynamical variable, the velocity of the medium w.r.t. the observer. When we generalize to let this velocity be nonzero, and then transform *all* the dynamical variables (including the new one), then we find that the wave equation really is galilean-invariant. But for light, which has no material medium, we cannot rescue galilean invariance in this (or any other) way.

D.26

D.27 PROVISIONAL LORENTZ TRANSFORMATIONS

The incompletely-specified transformations discussed here got superseded by the full Lorentz transformations (below).

D.27.1 Addition of velocities

$$\beta_{\text{tot}} = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2}.$$

This result differs from the galilean velocity addition formula. That fact let us distinguish the alternatives by analyzing Fizeau's experiment on the speed of light in flowing water.

D.28 FULL LORENTZ TRANSFORMATIONS

Lorentz transformations

I observed that the wave equation, if not extended to include any medium velocity, does not have galilean invariance. But it does have a *different* invariance, which also relates coordinate systems in uniform motion w.r.t. each other:

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \gamma \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix} \quad \text{where} \quad \gamma^2 = (1 - \beta^2)^{-1}. \quad (\text{D.29})$$

We can see that two coordinate systems related by a transformation of the above form will disagree on whether or not two events 1 and 2 are simultaneous. For example, if $t_{(1)} = t_{(2)}$, we generally won't have $t'_{(1)} = t'_{(2)}$ (unless $x_{(1)} = x_{(2)}$ as well).

In contrast, two coordinate systems related by the transformation in Equation D.29 *will* agree on whether a given trajectory is moving at speed c or not, despite the fact that the two systems are moving relative to each other! For example, if one trajectory obeys $x = ct$, then you can easily show that the other one obeys $x' = ct'$ with the same numerical value of c .

Invariant interval

$$(\Delta\tau)^2 = c^{-2}((c\Delta t)^2 - (\Delta x)^2). \quad (\text{D.30})$$

Another form of Lorentz transformations

Another way to characterize the Lorentz transformations is as the linear transformations that preserve the form of the invariant interval (Equation D.30). That viewpoint inspired the guess that, almost exactly like rotations,

$$\begin{bmatrix} ct' \\ x' \end{bmatrix} = \begin{bmatrix} \cosh \Upsilon & \sinh \Upsilon \\ \sinh \Upsilon & \cosh \Upsilon \end{bmatrix} \begin{bmatrix} ct \\ x \end{bmatrix} \quad (\text{D.31})$$

does the job. This really is the same as Equation D.29, with $\gamma = \cosh \Upsilon$, $\gamma\beta = \sinh \Upsilon$, two expressions that obey Equation D.29 for any value of the **rapidity** Υ . The composition (matrix product) of two of these boosts has the same overall form as Equation D.31 with parameter $\Upsilon_{\text{tot}} = \Upsilon_{(1)} + \Upsilon_{(2)}$. Translating back to velocities recovers the formula for β_{tot} in terms of β_1, β_2 .

D.29 RELATIVISTIC PARTICLE ENERGY AND MOMENTUM

We regard the trajectory of a moving particle as a curve in spacetime parameterized by proper time τ . Then we define relativistic momentum as $\vec{p} = m(d\vec{r})/d\tau = m\gamma\vec{v}$ and relativistic energy as $\mathcal{E} = mc(d(ct)/d\tau) = mc^2\gamma$. Here m is a constant, invariant property of the particle with dimensions of mass; it's called “the mass” of the particle. Our proposed form of the conservation laws is then simply that the sum of any of these four quantities over all particles is the same before and after a collision.

The virtue of these definitions is that then the **four-momentum** $\begin{bmatrix} \mathcal{E}/c \\ \vec{p} \end{bmatrix}$ transforms in the same way under Lorentz boosts as $\begin{bmatrix} ct \\ \vec{r} \end{bmatrix}$, namely, it undergoes a linear transformation. That makes it easy to show that the proposed conservation of four-momentum is Lorentz invariant.

For particles moving at close to the speed of light, $\mathcal{E} \approx pc$.

D.29.1 Some experimentally testable consequences

1. If a radioactive nucleus at rest decays into products with total mass smaller than the original, then the missing mass goes into kinetic energy of the products. Or when two nuclei react and rearrange to make two (or more) products, same prediction.
2. When a photon in the x ray part of the spectrum smacks an electron, it scatters into a new direction with a new wavelength. The observed change in wavelength

(Compton scattering formula) agrees with that predicted from the conservation of relativistic, not newtonian, momentum.

D.30 FOUR-VECTORS AND -TENSORS

D.30.1 More about 3-vectors and 3-tensors

We continue to restrict attention to non-accelerating, cartesian coordinate systems. Different cartesian coordinate systems, at rest with respect to one another, are related by the transformations

$$\vec{r}'_a = S_{aj}\vec{r}_j, \text{ and } t' = t, \text{ for } a, j = 1, 2, 3 \tag{D.32}$$

where the 3×3 matrix $[S]$ satisfies

$$[S]^t[S] = \mathbf{1}. \quad \text{Defining condition for a rotation or reflection} \tag{D.33}$$

That is, it's an **orthogonal matrix**. This condition ensures that the pythagorean theorem takes the same form in both the old and new coordinate systems, that is, the length-squared of a vector is

$$(\vec{r}'_1)^2 + (\vec{r}'_2)^2 + (\vec{r}'_3)^2 = (\vec{r}_1)^2 + (\vec{r}_2)^2 + (\vec{r}_3)^2.$$

About primes: We will usually reserve the prime modifier specifically for representations of the same geometrical object in two different coordinate systems.²⁴ As an additional visual cue, we'll use indices i, j, k, \dots for one set of coordinates and a, b, \dots for a new set. For example, Equation D.32 says that $\{\vec{r}'_a\}$ are the components of the vector “in the primed coordinate system.”

When two quantities are closely related, in some *other* way (not by a coordinate transformation), we'll sometimes use a tilde or check (háček) instead of a prime to distinguish them.²⁵

A **3-scalar** is a representation of a physical quantity by a single number that is the same when expressed in any of the coordinate systems in Equation D.32. Examples include the mass or charge of a point particle, etc. Ordinary numbers, with or without units, are also 3-scalars, for example the constants $\pi, \epsilon_0, \mu_0, \hbar$, etc. A **3-vector** is a representation of a physical quantity by three numbers (**components**) that depend on our choice of coordinate system in the same way as the components of \vec{r} . Thus, if the components of a 3-vector are given in one cartesian coordinate system, then we find them in any other one by using Equation D.32.

We define tensors of higher rank similarly. For example, a **3-tensor of rank two** has entries that transform the same way as those of the dyad product $\vec{r}\vec{r}$:

$$\vec{b}'_a = S_{ai}\vec{b}_i, \quad \vec{Q}'_{ab} = S_{ai}S_{bj}\vec{Q}_{ij}. \tag{D.34}$$

²⁴One exception: Sometimes we also use the traditional prime notation for a derivative of a function with respect to a spatial variable.

²⁵When we have an indexed collection of related objects, but the index is not specifically a coordinate index, we sometimes emphasize that by putting parentheses around the index (e.g. see Section D.17.3). Later we'll introduce yet another way to distinguish related quantities (index raising/lowering).

Again: The *same* vector will have *numerically different* components \vec{b}_i or \vec{b}'_a when viewed in a different coordinate system, and similarly for higher tensors.²⁶

The sum of two tensors of the same rank, for example $\vec{T} + \vec{S}$, is an indexed collection of numbers, each of whose entries is the sum of the corresponding entries of \vec{T} and \vec{S} :

$$(\vec{T} + \vec{S})_{ij} = \vec{T}_{ij} + \vec{S}_{ij} \quad \text{for each } i \text{ and } j.$$

Because coordinate transformations are linear (e.g. Equation D.34), this sum is again a tensor of the same rank as its constituents.

If the components of a tensor are specified *a priori* in more than one of these coordinate systems, then those definitions must be related by Equation D.34. For example, although the components of $\vec{\mathbb{I}}$ are fixed constants in any coordinate system (namely δ_{ij} , each of which equals 1 or 0), nevertheless the unit tensor really is a tensor: Its components do satisfy the second of Equation D.34, i.e., they transform the same way as those of $\vec{r}\vec{r}$ under rotations.

D.30.1.1 For theory enthusiasts only A mathematician would call a 3-tensor a “linear representation of the rotation group $\text{SO}(3)$.”

All our notations \vec{r}_i , \vec{r} , \vec{r}'_a , \vec{r}' , ... refer to *coordinate representations* of a vector. We have introduced no symbol to represent the “actual” geometrical vector itself.²⁷ We rely on consistent use of the same letter of the alphabet (here “*r*”) to remind us that these are all representations of the same thing.

Although the components of the Levi-Civita tensor are fixed constants, nevertheless they transform the same way as those of $\vec{r}_i\vec{r}_j\vec{r}_k$ under rotations. But under reflections, we find that $\varepsilon'_{123} = -1$ etc. So the L-C tensor is not quite invariant under $\text{O}(3)$.

Equivalently, if we define a tensor η in the new (left-handed) coordinates by using the usual rule, $\eta'_{123} = +1$, etc., then η is *not the same tensor* as ε . And if we were then to define the magnetic field by throwing test particles and interpreting their motions by a Lorentz force law defined using η , then we’d infer a magnetic field that differs from our \vec{B} by an overall minus sign in each component. This is what some authors mean when they call \vec{B} a **pseudovector** (and similarly for angular momentum): Its definition contains an odd number of ε ’s, and so depends on which coordinate choice we make when defining ε . (Similarly for **pseudoscalars**, such as the helicity of a circularly polarized plane wave.) We’ll just say that ε is always defined as in Section 0.2.2, starting from a right-handed coordinate system.²⁸

Finally, in quantum theory we must make a small extension to the Tensor Principle: If a theory is invariant under rotations, then actually we also must allow some physical quantities that transform according to representations of the “covering group” of $\text{SO}(3)$.²⁹ Some of these cannot be written as tensors. Such “spinor representations” correspond to half-odd spin particles like electrons in nonrelativistic quantum theory.

²⁶ S_{aj} is not itself a tensor because it carries mixed indices (j is in the old system; a is in the new system).

²⁷When we draw an arrow on a diagram, *that* refers to the actual geometric object.

²⁸We can physically select RH systems by weak interaction physics, which (unlike classical electrodynamics) makes an absolute distinction between left- and right-handed systems.

²⁹The covering group of $\text{SO}(3)$ turns out to be $\text{SU}(2)$.

D.30.2 3-scalar, 3-vector, and 3-tensor fields

A **3-scalar field** is a function of position, for example, energy density or electrostatic potential. Its value at a point P is the same regardless of coordinate system choice. However, when regarded as a function of three variables \vec{r}_i , then that function does depend on what coordinates we choose to represent points in space:

$$\psi'(\vec{r}'_1, \vec{r}'_2, \vec{r}'_3) = \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3) \quad \text{where } \vec{r}_i = (\mathbf{S}^{-1})_{ia} \vec{r}'_a. \quad (\text{D.35})$$

This formula defines its left side (a function of the three variables \vec{r}'_p) by setting it equal to the right side (also a function of \vec{r}'_a , once we use the definitions of \vec{r}_i as functions of \vec{r}'_a).

A **3-vector field** is a vector depending on position, for example the electric charge flux, $\vec{j}(\vec{r})$. Its value at a point P can be represented in the usual way relative to some coordinate system by a triple of numbers, each of which depends on the three coordinates of P . That is, we represent the field by three functions of three variables. In a different system we represent the *same* vector field \vec{b} by the new functions

$$\vec{b}'_a(\vec{r}'_1, \vec{r}'_2, \vec{r}'_3) = S_{ai} \vec{b}_i(\vec{r}_1, \vec{r}_2, \vec{r}_3) \quad \text{where } \vec{r}_i = (\mathbf{S}^{-1})_{ia} \vec{r}'_a. \quad (\text{D.36})$$

The above formula defines its left side (three functions of the three variables \vec{r}'_a) by setting it equal to its right side (also three functions of the three variables \vec{r}'_a). It is often written in the terse abbreviation $\vec{b}'_a(\vec{r}') = S_{ai} \vec{b}_i(\vec{r})$, with the understanding that the three variables $\{\vec{r}_i\}$ are to be regarded as functions of the $\{\vec{r}'_a\}$.

The transformation rules for the components of higher **3-tensor fields** are similar to Equation D.36.³⁰

D.30.3 4-vector notation

I'll use underscore to denote 4-vectors and other 4-tensor objects.³¹

Einstein agrees with Newton that: Trajectories are chains of “events,” and it takes four numbers to specify an event. But those giants disagree on what exactly are the transformations that are invariances of Nature.

In this course we are exploring the hypothesis that there's at least one “good” (or “E-inertial”) coordinate system (one in which particles and fields obey the Maxwell equations and Lorentz force law in their usual form). Choosing such a system, we let³²

$$\underline{X}^\mu = \begin{bmatrix} ct \\ \vec{r} \end{bmatrix}^\mu = \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}^\mu. \quad (\text{D.37})$$

³⁰An ordinary tensor may be regarded as a tensor field each of whose components is a constant. Equation D.36 shows that if the entries are constant in one cartesian coordinate system, they'll also be constant in any rotated system.

³¹Sometimes the symbol's *name* will also be a clue: $\vec{\nabla}_i$ are components of a three-vector operator, whereas we'll soon define $\underline{\partial}_\mu$ to be components of a four-covector operator. Sometimes a 4-vector quantity will be given a name with a capital letter to distinguish it from the corresponding 3-vector.

³²Comparing Equation D.37 to Equation 0.6 gives the bizarre-looking, but correct, formula $\underline{X}^i = \vec{r}_i$ for $i = 1, 2, 3$. This clash of 3-space and 4-space conventions is one reason why we keep the arrows and bars on all quantities in this document, to make clear which is which. Luckily, in E-inertial coordinates we have $\underline{g}_{ij} = \delta_{ij}$, so if we stick to such systems then we also have the saner-looking formula $\underline{X}_i = \vec{r}_i$.

Greek indices run through 0,1,2,3. Thus $\underline{X}^0 = ct$ etc. $[\underline{X}]$ is an abbreviation for \underline{X}^μ in which we abbreviate by omitting the index and using the rules of matrix multiplication to imply summations. $[\underline{X}]$ is a column vector.

Write a general linear transformation of coordinates as

$$\underline{X}'^\alpha = \Lambda^\alpha_\nu \underline{X}^\nu. \quad \text{transformation of a 4-vector} \quad (\text{D.38})$$

As in 3-space, we distinguish old and new coordinate systems by using one part of the alphabet μ, ν, \dots for the old and another part α, β, \dots for the new.

We can usefully think of the Lorentz transformations as those linear transformations on \underline{X}^μ that preserve the form of the following quantity:

$$\|\underline{X}\|^2 = -(ct)^2 + (\vec{r})^2 = \underline{X}^\mu \underline{g}_{\mu\nu} \underline{X}^\nu = [\underline{X}]^t [\underline{g}] [\underline{X}].$$

Here the **metric tensor**³³ $\underline{g}_{\mu\nu}$ refers to a matrix of constants:

$$\underline{g}_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_{\mu\nu}. \quad (\text{D.39})$$

That is, $\underline{g}_{00} = -1$, and so on. The condition for form-invariance of $\|\underline{X}\|^2$ is

$$\Lambda^\alpha_\mu \Lambda^\beta_\nu \underline{g}_{\alpha\beta} = \underline{g}_{\mu\nu}, \quad \text{Defining condition for a Lorentz transformation} \quad (\text{D.40})$$

where $\underline{g}_{\alpha\beta}$ is the *same* set of 16 constants as in Equation D.39.

For example, Equation D.29 amounts to

$$\Lambda^\alpha_\nu = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}_\nu^\alpha \quad \text{for a boost along the } +\hat{x} \text{ axis}$$

(where γ is defined by Equation D.29).

$[\Lambda]$ is an abbreviation for Λ^α_ν regarded as a 4×4 matrix. Thus we can abbreviate Equation D.38 as $[\underline{X}'] = [\Lambda][\underline{X}]$. (As soon as a formula starts to have more than a couple of indices, however, this level of abbreviation may become too confusing.) And Equation D.40 says $[\Lambda^t][\underline{g}][\Lambda] = [\underline{g}]$. *Notice* that this does *not* say that $[\Lambda]$ is an orthogonal matrix (unlike the case in 3D). That is, the group of all matrices satisfying this condition is not $O(4)$; it's a new group, the **Lorentz group** also called $O(3,1)$.

Note that despite the fact that we write it as a square, the quantity $\|\underline{V}\|^2$ need not be a positive number. For example, we can have $\|\underline{V}\|^2 = 0$ even if \underline{V} itself is not zero. Any 4-vector with this property is called **lightlike**, or **null**. If $\Delta\underline{X}$ is the spacetime separation between two events, then we call the three cases $\|\Delta\underline{X}\|^2 < 0$ and > 0 by the names **timelike** and **spacelike separation** respectively. A material particle always moves slower than c , so it will always move to a new spacetime point that is separated by a timelike displacement 4-vector from its original point.

We now ask, if the wave equation takes its usual form in one coordinate system, are there any *other* systems in which it also takes its usual form? To answer, we first

³³Warning: Some authors use a convention where they take the metric tensor to be minus our metric tensor.

consider the wave equation for a scalar field³⁴: $\left(-\frac{\partial^2}{\partial(ct)^2} + \vec{\nabla}^2\right)\Phi = 0$. This can be elegantly written as³⁵

$$\underline{g}^{\mu\nu} \partial_\mu \partial_\nu \Phi = 0. \quad (\text{D.41})$$

Here $\underline{g}^{\mu\nu}$ is numerically the *same* set of constants as the ones appearing in Equation D.39. (We will soon explain why we introduce two slightly different names for the same set of constants!)

Lorentz transformations leave the form of Equation D.41 unchanged.³⁶ The proof is not as simple as in the 3D case, but it follows from Equation D.40.

Equation D.41 has plane-wave solutions, which can be compactly expressed as

$$\exp(i\underline{k}^\mu \underline{g}_{\mu\nu} X^\nu) \quad (\text{D.42})$$

Here \underline{k} is a lightlike 4-vector consisting of \underline{k}^0 (previously called ω/c) and \underline{k}^i (previously called \vec{k}_i).

D.30.4 Relativistic particles

D.30.4.1 Massive We describe a trajectory $\underline{\Gamma}$ as a sequence of events depending on some parameter, i.e. $\underline{\Gamma}(\xi)$. Because any massive particle must travel slower than c , we must have $\|\underline{d}\underline{\Gamma}/d\xi\|^2 < 0$. We can make a unique choice of parameterization by requiring that $\|\underline{d}\underline{\Gamma}/d\xi\|^2 = -c^2$; with this choice, we call ξ the particle's **proper time** and rename it τ .³⁷ Then we define the **4-velocity** as³⁸

$$\underline{U} = \frac{d\underline{\Gamma}}{d\tau}.$$

Because τ is defined by an invariant equation, it's a scalar. The derivative of a 4-vector with respect to a scalar is a 4-vector, so \underline{U}^μ deserves its name. (Note that all four of its components have the units of velocity.)

Equivalently, for *any* parameterization we can define

$$\underline{U} = \frac{c}{\sqrt{-\|\underline{d}\underline{\Gamma}/d\xi\|^2}} \frac{d\underline{\Gamma}}{d\xi}$$

which again shows that \underline{U} really is a 4-vector. (The choice of parameterization ξ drops out of this formula.)

Note that

$$\underline{U}^\mu = \gamma \begin{bmatrix} c \\ v_x \\ v_y \\ v_z \end{bmatrix}^\mu.$$

³⁴This is a simplification of the electromagnetic wave equation, which actually involves the vector potential. But physics does make use of the scalar wave equation, e.g. for describing spinless particles like the pion. (To be ultraprecise, the pion field is a “pseudoscalar.”)

³⁵The Laplace-type operator appearing in Equation D.41 is called the **wave operator** or D'Alembert operator (sometimes written \square).

³⁶Actually, we get an invariance even if $\Lambda^\alpha_\mu \Lambda^\beta_\nu \underline{g}^{\mu\nu}$ is a constant times $\underline{g}^{\alpha\beta}$. As before, we are eliminating the spurious dilatation transformations because we know they can't be invariances of all of Nature.

³⁷Thus τ is analogous to contour length (also called arclength), along a curve in ordinary euclidean geometry.

³⁸P+S use the letter $\underline{\eta}$ instead of \underline{U} for this set of quantities.

In particular, lab time along the trajectory is related to proper time by $dt/d\tau = \gamma$.

We define the 4-momentum of a particle with this trajectory by the formula

$$\underline{p} = m\underline{U}, \quad (\text{D.43})$$

where m is a scalar constant characterizing the particle. We also give special names to its components, namely $\underline{p}^0 = E/c$ and $\underline{p}^i = p_i$. To justify these formulas, which differ from Newton's, we first note that the conservation law

$$\underline{p}_{\text{initial}}^\mu - \underline{p}_{\text{final}}^\mu = 0$$

is Lorentz invariant (and hence a candidate for a law of Nature). In contrast, Newton's versions of energy and momentum conservation rules are not Lorentz-invariant.

From its definition we see that $\|\underline{p}\|^2 = -(mc)^2$. Thus the relativistic energy and momentum obey the relation

$$\mathcal{E}^2 - (c\vec{p})^2 = m^2c^4. \quad (\text{D.44})$$

In an E-inertial coordinate system where $|\vec{v}| \ll c$, the spatial components of momentum reduce to $\vec{p} \approx m\vec{v} + \dots$. Also, $\mathcal{E} \approx mc^2 + \frac{1}{2}m|\vec{v}|^2 + \dots$; for collisions in which particle masses don't change, we can ignore the first term and recover the newtonian formula for kinetic energy conservation.

D.30.4.2 Massless To see what a "massless particle" could possibly mean, notice that setting $m \rightarrow 0$ in Equation D.44 yields $\|\vec{p}\|^2 = (\mathcal{E}/c)^2$ or

$$\underline{p}^\mu = \begin{bmatrix} p \\ \vec{p} \end{bmatrix}^\mu \quad \text{where } p = \|\vec{p}\|. \quad (\text{D.45})$$

The corresponding trajectories are null lines in spacetime. Although such a trajectory has no proper-time parameter, and hence \underline{U} is undefined, nevertheless Equation D.45 defines a legitimate momentum 4-vector; the corresponding particle does carry nonzero momentum and energy. Moreover, these quantities are related by $\mathcal{E} = pc$, just like the energy and momentum fluxes of light (sect. D.18).

In quantum theory, $\underline{p} = \hbar\underline{k}$ where \underline{k} is the wave vector from Equation D.42. This relation encompasses both deBroglie's $\vec{p} = \hbar\vec{k}$ and Einstein's $\mathcal{E} = \hbar\omega$.

D.31 THE FARADAY TENSOR

D.31.1 Manifestly invariant form of Lorentz force law

The **Faraday tensor** unifies electric and magnetic fields:

$$\underline{F}^{\mu\nu} = c^{-1} \begin{bmatrix} 0 & \vec{E}_1 & \vec{E}_2 & \vec{E}_3 \\ -\vec{E}_1 & 0 & \check{B}_3 & -\check{B}_2 \\ -\vec{E}_2 & -\check{B}_3 & 0 & \check{B}_1 \\ -\vec{E}_3 & \check{B}_2 & -\check{B}_1 & 0 \end{bmatrix}^{\mu\nu} \quad (\text{D.46})$$

where $\check{B}_i = c\vec{B}_i$. This big formula can be summarized compactly by $\underline{F}^{0i} = -\underline{F}^{i0} = \vec{E}_i/c$ and $\underline{F}^{ij} = \varepsilon_{ijk}\vec{B}_k$.³⁹

The Lorentz force law says that a particle’s trajectory $\underline{\Gamma}(\tau)$ obeys the four ordinary differential equations

$$\frac{d}{d\tau} P^\mu = q \underline{F}^{\mu\nu}(\underline{\Gamma}(\tau)) \underline{U}_\nu(\tau). \quad \text{Lorentz force law} \quad (\text{D.47})$$

Suppose that we have a region of uniform \vec{B} . A charged particle launched perpendicular to \vec{B} will move in a circular orbit. The period of the orbit is independent of the initial particle speed, as long as that speed is much smaller than c . At relativistic speed, however, the period begins to change in a readily measurable way that agrees with the prediction from Equation D.47.

D.32 MANIFESTLY INVARIANT FORM OF MAXWELL

D.32.1 4-covectors et al.

Any set of four physical quantities whose values change in the same way as \underline{X}^μ when remeasured in another E-inertial coordinate system (see Equation D.38) is called a four-vector, or **4-tensor of rank** $\binom{1}{0}$.⁴⁰ For example, the energy/ c and momentum of a point particle together form a 4-vector \underline{p}^μ .

A **4-tensor of rank** $\binom{2}{0}$ is a set of 16 quantities that transform under change of E-inertial coordinate system like the products $\underline{X}^\mu \underline{X}^\nu$.

If we try to contract a 4-vector with another 4-vector, we know that the result $\sum_{\mu=0}^3 A^\mu B^\mu$ is not invariant, because Lorentz transformations are not orthogonal matrices. Asking what sort of quantity *will* be invariant when contracted with a 4-vector leads us to define a **4-covector**, or **4-tensor of rank** $\binom{0}{1}$, as a set of four measurable quantities \underline{b}_μ that transform under change of E-inertial coordinate system according to

$$\underline{b}'_\alpha = \underline{b}_\mu (\Lambda^{-1})^\mu_\alpha. \quad \text{Lorentz transformation of a 4-covector} \quad (\text{D.48})$$

In matrix notation, we can say $[\underline{b}'] = [\Lambda]^{-1t}[\underline{b}]$. Since $[\Lambda]$ is not orthogonal, this transformation is not the same as Equation D.38. So we must distinguish 4-vectors from 4-covectors; we do this via index placement.

A **4-tensor of rank** $\binom{p}{q}$ has components of the form $\underline{A}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$, which undergo a mixed transformation, with p copies of $[\Lambda]$ and q copies of $[\Lambda]^{-1t}$.

Another way of expressing the form-invariance of the invariant interval is to say that $\underline{g}_{\mu\nu}$ is an **invariant constant tensor** of rank $\binom{0}{2}$ (a “4-tensor from Heaven,” just as δ_{ij} is a rotationally invariant constant tensor of rank 2 in 3D). Similarly $\underline{g}^{\mu\nu}$ is

³⁹Section D.14 called this 3×3 block $2\vec{\omega}_{ij}$.

⁴⁰P+S use the term **contravariant vector**, but I can never remember which is co- and which is contra-variant. The notation $\binom{1}{0}$ says just what it means, namely “one index up, none down.”

also an invariant constant tensor of rank $\binom{2}{0}$. The Kronecker symbol δ_{μ}^{ν} , whose entries are 1 if $\mu = \nu$ or 0 otherwise, is an invariant constant tensor of rank $\binom{1}{1}$.

When we multiply the components of a $\binom{p}{q}$ tensor by those of a $\binom{p'}{q'}$ tensor, we obtain the components of a $\binom{p+p'}{q+q'}$ tensor. When we add the components of a $\binom{p}{q}$ tensor to the corresponding components of another tensor of the same rank, we again get a $\binom{p}{q}$ tensor.

It may seem burdensome to keep track of two different transformation laws, but in practice it's easy because we can easily convert a covector to a vector and vice versa: Given a 4-covector \underline{k}_{μ} , the four quantities $\underline{g}^{\nu\mu}\underline{k}_{\mu}$ form a 4-vector. We name these quantities \underline{k}^{ν} , that is, we use the same letter of the alphabet but different index placement, to emphasize that they are closely related to \underline{k}_{μ} , and we call the conversion **index raising**.

Similarly, given a 4-vector \underline{A}^{μ} , the four quantities $\underline{g}_{\nu\mu}\underline{A}^{\mu}$ form a 4-covector. We name these quantities \underline{A}_{ν} , that is, we use the same letter of the alphabet but different index placement, to emphasize that they are closely related to \underline{A}^{μ} , and we call the conversion **index lowering**.

Index raising and lowering are invariantly defined operations, because $\underline{g}^{\mu\nu}$ and $\underline{g}_{\mu\nu}$ are invariant constant tensors and contraction is invariant. These operations invert each other, because $\underline{g}^{\mu\nu}\underline{g}_{\nu\lambda} = \underline{\delta}_{\lambda}^{\mu}$.

So really, we only need to remember one transformation law. If we want to find the transformation of \underline{k}_{ν} , we can raise its index, apply Equation D.38 to \underline{k}^{μ} , then lower the index.

Let $\underline{A}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$ be a 4-tensor of rank $\binom{p}{q}$. Then:

- . If \underline{A} is antisymmetric on some or all of its upper indices in one coordinate system, it will have that same property in any other E-inertial system (and similarly for lower indices, and similarly for symmetry). That is, the statement that a tensor is (anti)symmetric is a Lorentz-invariant property.
- . Also the operation of antisymmetrizing (or symmetrizing) a tensor on some or all of its upper (or lower) indices is invariantly defined.
- . But warning: There is no invariant sense to (anti)symmetry between an upper and a lower index. We must lower one index, or raise the other, before we can speak invariantly of (anti)symmetry.

D.32.2 4-tensor fields

A **4-vector field** is a 4-vector defined throughout a region of spacetime, with transformation law analogous to the one for a 3-tensor field:

$$\underline{A}'^{\alpha}(\underline{X}') = \Lambda^{\alpha}_{\mu} \underline{A}^{\mu}(\underline{X}), \quad \text{where } [\underline{X}] = [\Lambda^{-1} \underline{X}']. \quad (\text{D.49})$$

Similarly we define other 4-tensor fields. (A **4-scalar field** is just a fancy name for an ordinary function of \underline{X}^{μ} .)

We can take derivatives by using the 4-gradient operator $\underline{\partial}_{\mu} = \partial/\partial \underline{X}^{\mu}$. Notice the notation: An upper index in the denominator counts as a lower index. You should confirm that indeed the derivative of a 4-scalar field $\underline{\partial}_{\mu}\Phi$ is a covector field (Equation D.48), as implied by the lower index. Thus also the directional derivative of Φ along a 4-vector field \underline{V} is a scalar field (namely $\underline{V}^{\mu}\underline{\partial}_{\mu}\Phi$).

We fuse the charge density and charge flux to make a single 4-vector field⁴¹

$$\underline{J}^\mu(\underline{X}) = \left[\begin{array}{c} c\rho_q(t, \vec{r}) \\ \vec{j}(t, \vec{r}) \end{array} \right]^\mu. \tag{D.50}$$

One way to prove that these four quantities indeed form a 4-vector is to note that the continuity equation, which is valid in any coordinate system (invariant), can be written as $\partial \underline{J}^\mu / \partial \underline{X}^\mu = 0$. Making the useful abbreviation

$$\underline{\partial}_\mu = \partial / \partial \underline{X}^\mu$$

lets us rephrase compactly as $\underline{\partial}_\mu \underline{J}^\mu = 0$. Since we know the transformation rule of $\underline{\partial}_\mu$, we infer that of \underline{J}^μ .⁴²

An ordinary constant may be regarded as a scalar field that happens to be independent of \underline{X} . Similarly, an ordinary 4-vector can be regarded as a 4-vector field with constant components, as long as we stick to E-inertial (cartesian) coordinates.

D.32.3 Invariants and invariant equations

The metric tensor can be used to create a scalar given two 4-vectors, via $\underline{A}^\mu \underline{B}^\nu g_{\mu\nu}$, or equivalently its abbreviation $\underline{A}^\mu \underline{B}_\mu$. The quantity $\|\underline{X}\|^2$ is just the special case where both \underline{A} and \underline{B} are taken to be the separation $\Delta \underline{X}$:

$$c^{-2} \|\Delta \underline{X}\|^2 = c^{-2} (\Delta \underline{X}^\mu) (\Delta \underline{X}_\mu).$$

More generally, we can **contract** any upper index with any lower index to convert a $\binom{p}{q}$ tensor to a $\binom{p-1}{q-1}$ tensor. If we reduce the rank all the way to zero, we've got a scalar (invariant quantity).

Here are a few simple Rules analogous to the ones in 3D:

1. A formula will be invariant if it takes the form (4-tensor) = 0.
2. The left side will be guaranteed to be a tensor if all dynamical quantities appearing in it are themselves 4-tensors, and...
3. ... all constants are either 4-scalars (for example, mass m , charge q , c , ϵ_0 , μ_0), or the special 4-tensors $\underline{g}^{\mu\nu}$, $\underline{g}_{\mu\nu}$, δ_ν^μ .⁴³
4. Indices may be contracted in up/down pairs.
5. You can get an ordinary tensor starting from a tensor field by integrating over all spacetime using d^4X .⁴⁴ In particular, you can integrate a scalar field to get a scalar. But note that integrating over all *space*, at a fixed value of time, does not necessarily give a scalar.

When an equation is Lorentz-invariant because it obeys these rules, some authors call it **covariant** or **manifestly invariant**.

⁴¹Section D.32.6 gives another formulation for \underline{J} .

⁴²Section D.32.6 gives another form of the definition of \underline{J} .

⁴³There is also a 4-dimensional Levi-Civita tensor (invariant under those Lorentz transformations that are not reflections), but we won't need it.

⁴⁴As in 3D, we must introduce a jacobian factor if we wish to use curvilinear coordinates.

D.32.3.1 For theory enthusiasts only Again there is an extension to the Tensor Principle in relativistic quantum theory, analogous to the one in nonrelativistic theory: To write a classical field theory that when quantized will yield fermions, we must generalize to allow representations of the covering group Spin(3,1).

D.32.4

Equation D.50 combined the electric charge density and electric charge flux into a 4-component object. Then the following equations are manifestly Lorentz-invariant:

$$\boxed{\partial_\nu \underline{F}^{\mu\nu} = \mu_0 \underline{J}^\mu \quad \text{and} \quad \partial_\mu \underline{F}_{\nu\lambda} + \partial_\nu \underline{F}_{\lambda\mu} + \partial_\lambda \underline{F}_{\mu\nu} = 0. \quad \text{Maxwell equations}} \quad (\text{D.51})$$

Notice that no Levi-Civita tensor appears; thus, these equations are also manifestly invariant under inversions.

The first Maxwell equation, involving \underline{J} , is really four equations, because each side has one loose index. The second one appears to be $4^3 = 64$ equations, because it has three loose indices. Really, however, most of these equations are redundant, because the left side is totally antisymmetric on its three indices. A totally antisymmetric 4-tensor of rank $\binom{0}{3}$ has only four independent components.

So in all there are eight distinct equations, and indeed, substituting Equations D.46 and D.50 shows that Equation D.51 are precisely the Maxwell equations.

It may seem that eight equations in six unknown functions is too many, but two combinations of the equations are tautologies, vacuously satisfied regardless of what the fields and particles are doing. To see this,

- Take the divergence of the first set of equations and recall that $\partial_\mu \underline{J}^\mu = 0$ identically. So one combination of these four equations does not constrain $\underline{F}^{\mu\nu}$.
- Apply $\underline{\varepsilon}^{\mu\nu\lambda\kappa} \partial_\kappa$ to the second set of equations and recall that partial derivatives commute. Here $\underline{\varepsilon}$ is the 4D analog of the Levi-Civita tensor, with $\underline{\varepsilon}_{0123} = +1$. Again, you find that one combination of these four equations is always automatically satisfied.

(Compare the discussion in Section D.7.1.)

D.32.4.1 For theory enthusiasts only We can rewrite Maxwell's equations even more compactly as

$$\underline{d} \underline{F} = 0 \quad \text{and} \quad \underline{d}(*\underline{F}) = \mu_0(*\underline{J}). \quad (\text{D.52})$$

Here \underline{d} is the exterior derivative operator (Section D.14.1.1) and the **dual Faraday tensor** is:

$$*\underline{F}^{\mu\nu} = \frac{1}{2} \underline{\varepsilon}^{\mu\nu\alpha\beta} \underline{F}_{\alpha\beta} = c^{-1} \begin{bmatrix} 0 & -\underline{B}_x & -\underline{B}_y & -\underline{B}_z \\ \underline{B}_x & 0 & E_z & -E_y \\ \underline{B}_y & -E_z & 0 & E_x \\ \underline{B}_z & E_y & -E_x & 0 \end{bmatrix}^{\mu\nu}. \quad (\text{D.53})$$

P+S call this tensor $-\underline{G}^{\mu\nu}$ (their Levi-Civita tensor differs from mine by an overall minus sign). The other star appearing in Equation D.52 is defined by a formula similar to Equation D.53: $(*\underline{J})_{\mu\nu\lambda} = \underline{J}^\sigma \underline{\varepsilon}_{\sigma\mu\nu\lambda}$.

In fact both the exterior derivative and the **Hodge dual** operation defined above can be defined in any coordinate system (not just E-inertial coordinates), using the metric tensor.

So Equation D.52 shows that Maxwell's equations can be written in a completely coordinate invariant way. And that is valuable when it comes time to introduce gravitation, for example to find the gravitational bending and redshifting of light as it passes by a massive object: All we need to do is to replace the flat spacetime metric we've been using so far by the curved metric appropriate to this situation.

D.32.5 Four-vector potential

The Poincaré Lemma (Section D.14), together with the second Maxwell equation in Equation D.51, says that on a contractible region we can always write the Faraday tensor in terms of a **four-vector potential**:⁴⁵

$$F^{\mu\nu} = \partial^\mu \underline{A}^\nu - \partial^\nu \underline{A}^\mu, \quad \text{where } \underline{A}^\mu = \left[\frac{\psi/c}{\underline{A}} \right]^\mu. \quad (\text{D.54})$$

Thus typical units of \underline{A} are $\text{kg m coul}^{-1} \text{s}^{-1}$. The known transformation properties of ∂ and F imply that \underline{A} is a four-vector field.

When we write Maxwell's equations in terms of \underline{A} , one set is automatically solved; the other becomes the 4-vector wave equation

$$-\partial_\mu \partial^\mu \underline{A}^\nu + \partial_\mu \partial^\nu \underline{A}^\mu = \mu_0 \underline{J}^\nu. \quad (\text{D.55})$$

This is four equations in four unknowns. However, one combination of the potentials is a fake: **Gauge invariance** is the observation that the Faraday tensor doesn't change when we replace \underline{A}^μ by $\tilde{\underline{A}}^\mu = \underline{A}^\mu + \partial^\mu \Xi$. Luckily, one combination of Equations D.55 is still vacuously satisfied, as we see by taking the 4-divergence of both sides. So we have effectively reduced to 3 equations in 3 unknowns.

D.32.6 More about the charge–flux 4-vector field

We can reexpress \underline{J} in a way that makes its 4-vector character more explicit.⁴⁶ Suppose that we have a lot of charged particles labeled by ℓ . The charges are q_ℓ and the trajectories are $\underline{\Gamma}_{(\ell)}(\tau)$ where τ is proper time. Then

$$\underline{J}(\underline{X}) = c \sum_\ell q_\ell \int d\tau \underline{U}_\ell(\tau) \delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau)). \quad (\text{D.56})$$

As before, \underline{X} is the observation point, whereas $\underline{\Gamma}_{(\ell)}(\tau)$ is the location of particle ℓ at its proper time τ . In this formula, the separate ingredients q_ℓ , $\int d\tau$, and $\delta^{(4)}(\underline{X} - \underline{\Gamma}_{(\ell)}(\tau))$ are all Lorentz scalars. And \underline{U}_ℓ is a 4-vector, so the whole thing is a 4-vector.

D.32.6.1 For theory enthusiasts only In Minkowski spacetime, the delta function appearing in Equation D.56 is just the product of four ordinary delta functions. In curved spacetime, it needs an additional factor related to the metric.

Actually, it is more geometrically natural to think of charge flux in terms of the Hodge dual to \underline{J} (Section D.32.4.1). This 3-form $*\underline{J}$ takes a signed 3-volume element and spits out the net charge of all particle trajectories that pierce it; it is defined without any recourse to the metric. And then the continuity equation takes the elegant, and fully coordinate-invariant, form $\underline{d} * \underline{J} = 0$.

⁴⁵So we've more elegantly rederived a result in Section D.14.1.

⁴⁶Compare Equations D.16–D.17.

D.33 ENERGY AND MOMENTUM OF FIELDS, GENERAL

D.33.1

We define a set of four 4-vector fields analogous to \underline{J} in Equation D.56, but replacing each particle's charge q_ℓ by the components of its momentum $\underline{p}_{(\ell)}^\mu(\tau)$. We assemble these four 4-vectors into a single tensor $\underline{T}_{\text{part}}^{\mu\nu}$, the **energy-momentum flux tensor**⁴⁷ of the particles. You should make sure you understand these identifications of the blocks of \underline{T} :

$$\begin{aligned}\underline{T}^{00} &= c \cdot (\text{density of energy}/c) \\ \underline{T}^{i0} &= c \cdot (\text{density of } p_i) \\ \underline{T}^{0j} &= (\text{flux of energy}/c \text{ along } j) \\ \underline{T}^{ij} &= (\text{flux of } p_i \text{ along } j).\end{aligned}$$

Thus typical units of \underline{T} are $\text{kg m}^{-1} \text{s}^{-2}$.

Unlike charge, momentum can change along each particle's trajectory. So $\partial_\mu \underline{T}_{\text{part}}^{\mu\nu} \neq 0$. But the energy-momentum flux tensor for the electromagnetic fields:

$$\underline{T}_{\text{field}}^{\mu\nu} = -\mu_0^{-1} \left(\underline{F}^{\mu\lambda} \underline{F}_{\lambda}{}^{\nu} + \frac{1}{4} g^{\mu\nu} (\underline{F}^{\lambda\sigma} \underline{F}_{\lambda\sigma}) \right) \quad (\text{D.57})$$

has the property that

$$\partial_\mu (\underline{T}_{\text{part}}^{\mu\nu} + \underline{T}_{\text{field}}^{\mu\nu}) = 0.$$

This formula is a continuity equation expressing local conservation of total energy and momentum.⁴⁸ For example, the density of field energy is $u = \underline{T}_{\text{field}}^{00} = (2\mu_0)^{-1} (\vec{E}^2/c^2 + \vec{B}^2)$. The flux of field energy (**Poynting vector**) is $\vec{S}_i = c \underline{T}^{i0} = c \underline{T}^{0i} = \mu_0^{-1} (\vec{E} \times \vec{B})_i$. And $\frac{1}{3} \underline{T}_{\text{field}}^{ii}$ is the pressure exerted by EM fields (averaged over directions). You should work it out in terms of the components of \vec{E} and \vec{B} .

D.33.1.1 For theory enthusiasts only: Angular momentum of fields Define

$$\underline{M}^{\lambda\mu\nu} = \underline{X}^\mu \underline{T}^{\nu\lambda} - \underline{X}^\nu \underline{T}^{\mu\lambda}.$$

Then $\partial_\lambda \underline{M}^{\lambda\mu\nu} = 0$, so the six quantities

$$\underline{J}^{\mu\nu} = \int d^3r \underline{M}^{0\mu\nu},$$

where the integral is over any fixed time, are conserved. The spatial components \underline{J}^{ij} are total angular momenta of fields and particles.

D.34

D.35 PLANE WAVES IN 4D LANGUAGE

⁴⁷Usually abbreviated "energy-momentum tensor." Some authors instead say "stress-energy tensor."

⁴⁸So it's a generalization of **Poynting's theorem**.

D.35.1 Scalar waves

The scalar wave equation has plane-wave solutions of the form $\Phi(\underline{X}) = \frac{1}{2}(\exp(i\underline{k}_\mu \underline{X}^\mu) + \text{c.c.})$, characterized by a 4-vector $\underline{k}^\mu = \left[\frac{\omega/c}{\underline{k}}\right]^\mu$ (the **4-wavevector**). Such a solution solves the scalar wave equation if $\|\underline{k}\|^2 = 0$ (“ \underline{k} is a **null 4-vector**”).

We could use gauge freedom to insist on Coulomb gauge as before. But it’s nicer to insist on a Lorentz-invariant condition,

$$\partial_\mu \underline{A}^\mu = 0. \quad \text{Lorenz gauge condition} \quad (\text{D.58})$$

This condition simplifies Equation D.55, which becomes four uncoupled copies of the scalar wave equation:

$$-\square \underline{A} = \mu_0 \underline{J}. \quad (\text{D.59})$$

In particular, this equation is still manifestly Lorentz-invariant, because we imposed a manifestly Lorentz-invariant condition on \underline{A} . These equations are valid regardless whether the charge density is zero or not.

D.35.1.1 For theory enthusiasts only Although \vec{E} and \vec{B} have messy transformation laws, remarkably two quadratic combinations of their entries are Lorentz-invariant. This becomes obvious, not remarkable, when we apply “Einstein thinking”: note that

$$\underline{F}_{\mu\nu} \underline{F}^{\mu\nu} = 2(\vec{B}^2 - \vec{E}^2/c^2) \quad (\text{D.60})$$

and (less obviously)

$$\underline{F}^{01} \underline{F}^{23} \pm (\text{all permutations of } 0123) = \frac{8}{c} \vec{E} \cdot \vec{B}. \quad (\text{D.61})$$

Suppose that we are asked to find the orbit of a particle moving in imposed \vec{E} and \vec{B} fields. The first of these invariants can be used to figure out whether we can simplify the problem by passing to an E-inertial coordinate system in which $\vec{E} = 0$ or $\vec{B} = 0$.

Equation D.60 is also suitable as a Lagrangian density functional for a variational formulation of the Maxwell equations.

Equation D.61 isn’t, because it’s a total derivative. To see this, and also to see that this term really invariant under rotations and boosts (but not under inversions), it can be expressed as $\underline{F}^{\mu\nu} \underline{F}^{\lambda\sigma} \varepsilon_{\mu\nu\lambda\sigma}$.

D.35.2

Vacuum solutions include the plane waves. Similarly to the scalar wave equation (Section D.35.1), each is characterized by a null wavevector \underline{k} . Unlike the scalar field case, each wave is also characterized by a polarization 4-vector $\underline{\zeta}$:

$$\underline{A}^\mu(\underline{X}) = \frac{1}{2}(\underline{\zeta}^\mu \exp(i\underline{k}_\nu \underline{X}^\nu) + \text{c.c.}).$$

This 4-vector field will be in Lorenz gauge if $\underline{k}_\mu \underline{\zeta}^\mu = 0$. It also satisfies the Maxwell equations if in addition $\underline{k}_\mu \underline{k}^\mu = 0$.

Additionally, gauge invariance implies that we may add any multiple of \underline{k} to $\underline{\zeta}$ without changing the field strengths; note that this gauge transformation does not spoil Equation D.58. We can use this freedom to insist that also $\underline{\zeta}^0 = 0$. With that choice, $\vec{\zeta}$ is parallel to the electric field and perpendicular to \vec{k} .

D.35.2.1 For theory enthusiasts only Maxwell's equations may be valid only on a region of spacetime that's *not* contractible. For example, Grand Unified theories predict the existence of small regions where the electroweak and strong-interaction fields have nonzero values; Maxwell's equations are not valid inside such a region. In such a situation, even in the exterior region it may not be possible to find a four-vector potential that's everywhere smoothly defined, because the Poincaré lemma only guarantees this in a contractible region.

This loophole is what makes it possible to have magnetic monopole solutions in Grand Unified field theories, despite the absence of any magnetic-monopole source terms in Maxwell.

D.36 SPHERICAL WAVES

Lorenz gauge makes it easy to guess a **spherical wave** solution:

$$\vec{A}(t, \vec{r}) = \frac{1}{2} [\vec{\xi} r^{-1} e^{-i\omega t \pm ikr} + \text{c.c.}]. \quad (\text{D.62})$$

You should find the scalar potential corresponding to this solution.

D.37 BEAMS

[Not covered this year.]

D.38 VARIATIONAL FORMULATION

[Not covered this year.]

D.39 RADIATION GREEN FUNCTION

D.39.1

The **retarded (causal) Green function** for the wave equation (Equation D.55) is

$$D_r(\Delta \underline{X}) = \frac{1}{2\pi} \Theta(\Delta t) \delta(\|\Delta \underline{X}\|^2). \quad (\text{D.63})$$

$$= \frac{1}{4\pi c \|\Delta \vec{r}\|} \delta(\Delta t - c^{-1} \|\Delta \vec{r}\|). \quad (\text{D.64})$$

In the first formulation, Θ is the step function and $\|\Delta \underline{X}\|^2$ is the invariant length-squared. This formulation makes it clear that D_r is Lorentz invariant. To see this, note that $\delta(\|\underline{X}\|^2)$ is a function of an invariant expression, and hence invariant. The step function by itself would not be invariant, because the 0-component of a 4-vector can change sign under a Lorentz transformation (relativity of simultaneity). But this ambiguity does not arise if \underline{X} is a null vector (as it must be for the delta function to be nonzero).

The second formulation is often more convenient, as its delta-function is simpler. Here $\|\vec{r}\|^2$ is the ordinary length-squared.

What the Green function does for you: Introducing abbreviations $\vec{R} = \Delta\vec{r} = \vec{r} - \vec{r}_*$, $R = \|\vec{R}\|$, $\hat{R} = \vec{R}/R$, and $\rho_{\text{q}}|_{\text{ret}} = \rho_{\text{q}}(t - R/c, \vec{r}_*)$, we showed that for any “source function” $\mathcal{J}(\underline{X})$, the expression

$$\phi(\underline{X}) = \int d^4 X_* D_{\text{r}}(\underline{X} - \underline{X}_*) \mathcal{J}(\underline{X}_*)$$

reduces to the expression Equation D.27 and hence is a solution to $\partial_{\mu} \partial^{\mu} \phi = -\mathcal{J}$. Check that these formulas imply that D_{r} must have the dimensions \mathbb{L}^{-2} ; show that it does have those dimensions by using Equations D.63–D.64.

Taking \mathcal{J} to be μ_0 times any component of \underline{J}^{μ} , we get the corresponding component of \underline{A}^{μ} (the **retarded potentials**), in Lorenz gauge.

D.40

[Not covered this year.]

D.41 ELECTRIC DIPOLE RADIATION

D.41.1 Far fields of an oscillating electric dipole

Chapter 42 develops a double power-series expansion for the fields from a specified distribution of charges and currents. The first expansion (**far field**) is in powers of a/r , where a is the size of a region outside of which there are no charges or currents. The second expansion (**multipole**) is in powers of $a\omega/c$, where we assume a harmonic source with angular frequency ω .

Far-field expansion is always valid if we stand far enough away, for example if we only want to know how much radiation escapes to infinity.⁴⁹ Multipole expansion is useful for radiation by single atoms and molecules⁵⁰ and for many kinds of radio antennas etc.

The lowest term of the multipole expansion is called **electric dipole** approximation. In ED approximation we got an extremely simple result: A time-varying electric dipole moment $\vec{\mathcal{D}}_{\text{E}}(t)$ yields far fields

$$\vec{A}^{[ED]}(t, \vec{r}) = \frac{\mu_0}{4\pi r} \frac{d}{dt} \vec{\mathcal{D}}_{\text{E}}|_{\text{ret}}, \quad (\text{D.65})$$

$$\vec{B}^{[ED]}(t, \vec{r}) = -\frac{\mu_0}{4\pi r c} \hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_{\text{E}}|_{\text{ret}}, \quad \vec{E}^{[ED]}(t, \vec{r}) = -c\hat{r} \times \vec{B}(t, \vec{r}).$$

In the special case of linear polarization, $\vec{\mathcal{D}}_{\text{E}} = \hat{n} \mathcal{D}_{\text{E}}$. Then the Poynting vector is

$$\vec{S}^{[ED]} = \hat{r} \frac{\mu_0}{16(\pi r)^2 c} (\sin^2 \vartheta) \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_{\text{E}}|_{\text{ret}} \right\|^2, \quad (\text{D.66})$$

⁴⁹However, to analyze “near-field scanning optical microscopy” we need the exact fields.

⁵⁰Actually we need its quantum-mechanical generalization for that purpose.

where ϑ is the angle between the observer's position \hat{r} and the direction \hat{n} of the dipole moment. This is the **dipole doughnut** pattern.⁵¹

The energy flux per solid angle is then $d\mathcal{P}/d\Omega = r^2 \hat{r} \cdot \vec{S}$, and the total power sent out to infinity is

$$\mathcal{P}^{[ED]} = \frac{\mu_0}{6\pi c} \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_E|_{\text{ret}} \right\|^2. \quad (\text{D.67})$$

It's often appropriate to assume the source is harmonic at a single frequency; for example, if it points along a constant direction \hat{n} then $\vec{\mathcal{D}}_E(t) = \frac{1}{2}[\hat{n}\bar{\mathcal{D}}_E e^{-i\omega t} + \text{c.c.}]$. Here $|\bar{\mathcal{D}}_E|$ is the peak value of the magnitude of the electric dipole moment vector. In this situation, Equation D.65 agrees with the long-distance limit of the exact spherical wave solution, but now we can connect the outgoing wave to its cause (the oscillating dipole):

$$\vec{\xi} = \frac{\mu_0}{4\pi} (-i\omega) \bar{\mathcal{D}}_E \hat{n}.$$

We get another famous formula when we specialize the power-output formula (Equation D.67) to the harmonic case: Then the time average is $\langle \|\vec{\mathcal{D}}_E\|^2 \rangle = |\bar{\mathcal{D}}_E|^2/2$, and

$$\langle \mathcal{P}^{[ED]} \rangle = \frac{\mu_0}{12\pi c} \omega^4 |\bar{\mathcal{D}}_E|^2.$$

More complicated sources can be obtained with a more complicated constant vector $\vec{\mathcal{D}}_E$. For example, a rotating dipole has a *complex* $\vec{\mathcal{D}}_E$.

D.42 HIGHER MULTIPOLE RADIATION

If a time-varying source has electric dipole moment always equal to zero (or constant in time), then its radiation in the ED approximation will be zero. Nevertheless, it may still radiate. The next terms in the multipole expansion are called **magnetic dipole** (MD) and **electric quadrupole** (EQ), because they involve time derivatives of those quantities (Equations D.8, D.22). The MD term is of particular interest:

$$\vec{B}^{[MD]}(t, \vec{r}) = \frac{\mu_0}{4\pi c^2 r} \hat{r} \times \left(\hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_M|_{\text{ret}} \right) \quad \vec{E}^{[MD]}(t, \vec{r}) = \frac{\mu_0}{4\pi c r} \hat{r} \times \frac{d^2}{dt^2} \vec{\mathcal{D}}_M|_{\text{ret}}.$$

These formulas look like the ED formulas with \vec{E} and $c\vec{B}$ exchanged, so we can recycle our effort and just copy down the formula for energy flux per solid angle (Equations D.66 and D.67): for linear polarization,

$$d\mathcal{P}^{[MD]}/d\Omega = \frac{\mu_0}{16\pi^2 c^3} \sin^2 \vartheta \left\langle \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_M|_{\text{ret}} \right\|^2 \right\rangle \quad \text{or total power} \quad \mathcal{P}^{[MD]} = \frac{\mu_0}{6\pi c^3} \left\langle \left\| \frac{d^2}{dt^2} \vec{\mathcal{D}}_M|_{\text{ret}} \right\|^2 \right\rangle.$$

D.43

[Not covered this year.]

⁵¹So called because a polar plot of the surface $r = \sin^2 \theta$ looks like a doughnut. (Maybe it looks more like a bialy, or a red blood cell.)

D.44

D.45

[Not covered this year.]

D.46

[Not covered this year.]

D.47

D.47.1 Isotropic, Linear Media

We discussed an approximation in which

- We assume that the medium consists of polarizable objects (or permanently polarized, unoriented objects which can become oriented by an external field). We only consider the dipole fields created by those objects.
- We assume that external fields vary over length scales much longer than the spacing between the dipoles. That is, we suppose them to be finely enough divided (compared to the length scales of the disturbances we're studying) that they can be regarded as a continuously varying distribution of dipole moment density.⁵²

Let \vec{P} denote the dipole moment density. If \vec{P} is nonuniform, it gives rise to a **bound charge density** $\rho_{q,b} = -\vec{\nabla} \cdot \vec{P}$. At a boundary between dielectric and vacuum, we get a **bound surface charge density** $\sigma_b = \hat{n} \cdot \vec{P}$, where \hat{n} is the normal directed outward (toward the vacuum).

If the polarization is time-dependent, then the bound charge density will also give rise to a bound *current*, via the continuity equation; see Section D.47.2.

The “electric displacement” is defined by $\vec{D} = \epsilon_0 \vec{E} + \vec{P}$. (We'll just call it **the \vec{D} field**.) With these definitions, the electric Gauss law takes a simple form:

$$\vec{\nabla} \cdot \vec{D} = \rho_{q,f}. \quad (\text{D.68})$$

The only source appearing explicitly in this formula is the free charge density.

Many media are approximately linear:⁵³ That is, \vec{P} is a linear function of \vec{E} , described by the **dielectric susceptibility** $\vec{\chi}_e$ via $\vec{P} = \epsilon_0 \vec{\chi}_e \cdot \vec{E}$. For simplicity,

⁵²Or equivalently, we average the effects of finite-size molecules over a length scale smaller than the one of interest, but much bigger than the molecular spacing. The quantities \vec{E} , \vec{P} , \vec{B} , \vec{M} below are all averages of this sort.

⁵³Exceptions include piezoelectric crystals under stress, or ferroelectrics (“electrets”), which have nonzero \vec{P} in zero applied field. Also, any medium will be linear only in some regime of weak enough applied fields. For example, the orientational ordering of water molecules must eventually saturate (100% alignment) at high applied fields.

assume the medium is isotropic (χ_e is a scalar constant). Then we can define the **permittivity** $\epsilon = \epsilon_0(1 + \chi_e)$, so:

$$\vec{D} = \epsilon \vec{E} \quad \text{constitutive relation for homogeneous, linear, isotropic, lossless, nonchiral dielectric} \quad (\text{D.69})$$

Note that in general ϵ is a function of frequency: $\epsilon(\omega)$.

Thus, *we can forget about the medium* if it's linear; the Gauss law Equation D.68 retains its vacuum form, but with a modified value of the permittivity. All we need to keep track of is the *free* charge density. And the Faraday law, which has no sources anyway, remains completely unchanged.

More general forms of the constitutive relation include dissipation (complex ϵ), anisotropy ($\vec{\epsilon}$ with tensor structure), and chirality (see later).

D.47.2 Magnetic materials

In an approximation scheme similar to Section D.47.1, let \vec{M} denote the magnetic dipole density created by the motions of bound charges on individual polarizable objects. The “magnetic field intensity” is then defined by $\vec{H} = \mu_0^{-1}\vec{B} - \vec{M}$. (We'll just call it **the \vec{H} field**.)

If \vec{M} is nonuniform, it gives rise to a contribution to the **bound charge flux** (in addition to $\partial\vec{P}/\partial t$) equal to $\vec{\nabla} \times \vec{M}$. A special case is the boundary between medium and vacuum; here we get a **bound surface current density** $\vec{K}_b = \vec{M} \times \hat{n}$, where \hat{n} is the normal directed outward (toward the vacuum).

Many media are linear⁵⁴; that is \vec{M} is a linear function of \vec{B} , described by the **magnetic susceptibility** χ_m via $\vec{M} = \mu_0^{-1}\tilde{\chi}_m\vec{B}$. Then we can define the **permeability** $\mu = \mu_0/(1 - \tilde{\chi}_m)$, so

$$\vec{H} = \mu^{-1}\vec{B} \quad \text{constitutive relation for homogeneous, linear, isotropic, lossless nonchiral magnetic material} \quad (\text{D.70})$$

Note that in general μ is a function of frequency: $\mu(\omega)$.

More general forms of the constitutive relation include dissipation (complex μ), anisotropy ($\vec{\mu}$ with tensor structure), and chirality (see later).

D.47.3 Maxwell equations in media

We wish to eliminate explicit mention of the bound charges and currents. The remaining charges and currents are called “free”: $\rho_{q,f}$, \vec{j}_f . For example, excess (static) charges, which macroscopically violate charge neutrality, are considered free, as are currents that transport net charge over macroscopic lengths. Using the above formulas

⁵⁴Exceptions include ferromagnets, which have nonzero \vec{M} at zero applied field. Also, again any medium is only linear for sufficiently weak applied fields. Much of optics deals with media in their linear regime, but there is also a big field of “nonlinear optics.”

for bound charge and charge flux, we get

$$\vec{\nabla} \cdot \vec{D} = \rho_{q,f} \quad \text{Gauss} \quad (\text{D.71})$$

$$\vec{\nabla} \cdot \vec{B} = 0 \quad \text{Gauss} \quad (\text{D.72})$$

$$\vec{\nabla} \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \quad \text{Faraday} \quad (\text{D.73})$$

$$\vec{\nabla} \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = \vec{j}_f. \quad \text{Ampère} \quad (\text{D.74})$$

These equations are general. For the special case of linear media, they can be combined with Equations D.69–D.70, to form a closed system that can be solved to give all fields in terms of free charges and currents. Then just as we saw for the first Gauss law, *we can forget about* the medium, incorporating all bound currents and charges into the material parameters ϵ and μ ; for example, all we see in Equation D.74 is the *free* charge flux.

The boundary conditions at the interface between media can also be expressed entirely in terms of free surface charge density and current:

$$\Delta(D_{\perp}) = \sigma_f; \quad \Delta(B_{\perp}) = 0; \quad \Delta(\vec{E}_{\parallel}) = 0; \quad \Delta(\vec{H}_{\parallel}) = \vec{j}_f^{(2D)} \times \hat{n}.$$

Here $\Delta D_{\perp} = (\vec{D}^{[2]} - \vec{D}^{[1]}) \cdot \hat{n}$, where \hat{n} is the unit normal vector pointing from medium 1 to medium 2; similarly for $\Delta \vec{H}_{\parallel}$.

D.47.4 Cross-susceptibility

The constitutive relations Equations D.69 and D.70 are not the most general possible, even if we restrict to homogeneous, linear, isotropic, lossless materials. We may in addition have **cross-susceptibility**:

$$\begin{bmatrix} \vec{P} \\ \vec{M} \end{bmatrix} = \begin{bmatrix} \epsilon_0 \chi_e & -\eta \frac{\partial}{\partial t} \\ \eta' \frac{\partial}{\partial t} & (\mu_0 c^2)^{-1} \tilde{\chi}_m \end{bmatrix} \begin{bmatrix} \vec{E} \\ \vec{B} \end{bmatrix}. \quad (\text{D.75})$$

Recall $\check{M} = \vec{M}/c$ and $\check{B} = c\vec{B}$. These definitions simplify our formulas by giving all the entries in the matrix the same dimensions.

Generally the constants χ_e , η , η' , and $\tilde{\chi}_m$ are tensors, but in isotropic solution they get replaced by their averages over orientation, which are all 3-scalars. I argued from a macroscopic model (a helical wire) that η and η' have the same sign, which reflects the geometry (handedness) of the polarizable objects constituting the medium.⁵⁵ Replacing the helices by their mirror images changes the signs of both η and η' .

As mentioned before, χ_e and $\tilde{\chi}_m$ may be frequency-dependent. Similarly, for disturbances at a specific frequency ω the cross-terms may be functions of frequency. (By time-reversal invariance they must be *odd* functions, so these terms vanish at zero frequency.)

D.47.4.1 For theory enthusiasts only Equation D.75 has not been written in a manifestly Lorentz-invariant way: It involves a 6×6 matrix of susceptibilities, which is not obviously a

⁵⁵In fact, $\eta = \eta'$ by Onsager reciprocity.

Lorentz tensor. But in fact, we can define a “response” tensor as

$$\underline{R}^{\mu\nu} = \begin{bmatrix} 0 & \vec{P}_x & \vec{P}_y & \vec{P}_z \\ -\vec{P}_x & 0 & -\vec{M}_z & \vec{M}_y \\ -\vec{P}_y & \vec{M}_z & 0 & -\vec{M}_x \\ -\vec{P}_z & -\vec{M}_y & \vec{M}_x & 0 \end{bmatrix}^{\mu\nu} \quad (\text{D.76})$$

where $\vec{M}_i = c^{-1}\vec{\dot{M}}_i$. This big formula can be summarized compactly by $\underline{R}^{0i} = -\underline{R}^{i0} = \vec{P}_i$ and $\underline{R}^{ij} = -\varepsilon_{ijk}\vec{M}_k/c$.

The above definitions let us formulate four of the Maxwell’s equations as

$$\partial_\mu \underline{H}^{\nu\mu} = c^{-1} \underline{J}_\nu^\nu \quad \text{where} \quad \underline{H}^{\nu\mu} = c\varepsilon_0 \underline{F}^{\nu\mu} + \underline{R}^{\nu\mu}.$$

The four homogeneous equations are unchanged from Section D.32.4.

Linear response means that \underline{R} is a linear function of \underline{F} :

$$\underline{R}^{\mu\nu} = \underline{K}^{\mu\nu}{}_{\lambda\sigma} \underline{F}^{\lambda\sigma},$$

where the **susceptibility operator** \underline{K} is antisymmetric on its first two indices, and also on the last two.

Even an isotropic medium breaks Lorentz symmetry—unlike the vacuum, it can have states of rigid motion. But isotropy does imply that \underline{K} may only be constructed from one extrinsic quantity, the 4-velocity of the medium \underline{U} . Apart from that, it can only involve invariant tensors and scalar constants. Playing around shows that the only possible forms are dictated by the symmetries:

$$\begin{aligned} \underline{K}^{\mu\nu}{}_{\lambda\sigma} = & \frac{\alpha}{2} (\delta_\lambda^\mu \delta_\sigma^\nu - \delta_\lambda^\nu \delta_\sigma^\mu) + \frac{\tau}{2} (\underline{U}^\mu \underline{U}_\sigma \delta_\lambda^\nu - \underline{U}^\nu \underline{U}_\sigma \delta_\lambda^\mu - \underline{U}^\mu \underline{U}_\lambda \delta_\sigma^\nu + \underline{U}^\nu \underline{U}_\lambda \delta_\sigma^\mu) \\ & + \frac{\gamma}{2} (\varepsilon^{\mu\nu}{}_{\tau\lambda} \underline{U}^\tau \underline{U}_\sigma - \varepsilon^{\mu\nu}{}_{\tau\sigma} \underline{U}^\tau \underline{U}_\lambda - \varepsilon_{\lambda\sigma\tau}{}^\mu \underline{U}^\tau \underline{U}^\nu + \varepsilon_{\lambda\sigma\tau}{}^\nu \underline{U}^\tau \underline{U}^\mu) \underline{U}^\rho \partial_\rho. \end{aligned} \quad (\text{D.77})$$

Here the 4-dimensional Levi-Civita tensor has $\varepsilon_{0123} = +1$ etc. You can specialize this formula to a coordinate system in which the medium is at rest, and find that the constants α , β , and γ can be chosen so that it reproduces Equation D.75. Then substituting arbitrary 4-velocity at once tells us the appropriate form of the susceptibility tensor in a *moving* medium.

Every term in Equation D.77 must be time-reversal invariant, because a static collection of molecules does not break time-reversal invariance.⁵⁶ (This is why the γ term needed to have a derivative.) The α and β terms also are invariant to spatial inversions—but not the γ term. \underline{K} must also be a symmetric operator in the sense that exchanging $\mu\nu$ with $\lambda\sigma$, and $\underline{\partial} \rightarrow -\underline{\partial}$, must leave it unchanged. This fixes the above form as the complete list of allowed terms, to leading order in powers of derivatives.

D.47.5 Waves in media

In a simple, nonchiral medium the Maxwell equations with no free charges or currents have plane wave solutions with dispersion relation $|\vec{k}|^2 = \omega^2 \mu \varepsilon$. Note that although the right hand side of this formula is positive, still one component of \vec{k} could be imaginary (as long as the other two components are real and large enough). This could occur, for example, at the interface between media, giving rise to total internal reflection.

Another interesting situation is if the free charge flux is not zero, but instead satisfies Ohm’s law. Then we get wave propagation with attenuation.

⁵⁶Ferromagnetism was not allowed.

D.47.6 Optical activity

We found that when linearly polarized light passes through an isotropic but *chiral* medium, its polarization vector in general rotates by an amount proportional to the path length in the medium.⁵⁷

When *circularly* polarized light passes through such a medium, it emerges with the same circular polarization. However, the two helicity states propagate at different phase velocities. That is, they have different dispersion relations:

$$\omega^2 = \frac{k^2}{\mu(\epsilon \pm 2\eta ck)}$$

Equivalently, dividing the phase velocity by c yields two different refractive indices. This phenomenon is called **optical activity** (or “optical rotatory power” or “circular birefringence”).

A distinct but related phenomenon arises when we allow for the dissipation (energy conversion to heat) present in real materials: An isotropic chiral material can display **circular dichroism**, that is, it can have different energy absorption coefficients for the two circular polarizations of light.

D.48 ANISOTROPIC MEDIA

D.48.1 Ordinary birefringence

A homogeneous, lossless, *achiral*, but *anisotropic* medium, in its linear regime, is characterized by susceptibility *tensors*. Let’s suppose it’s nonmagnetic, $\mu = \mu_0$. The electric susceptibility tensor is a symmetric, rank-2 3-tensor, so (like the moment of inertia) it has three real, orthogonal, eigenvectors (whose directions are called **principal directions**). Such a medium has simple plane wave solutions propagating along one of the three principal directions and polarized along another one. But the two possible polarizations of such a wave travel with different phase velocities, a phenomenon called (ordinary) **birefringence**.⁵⁸ Similarly, a lossy medium may have different absorption coefficients in each of two principal directions, leading to (ordinary) **dichroism**.⁵⁹

⁵⁷Sorry, but “linear” is used in two different senses: “linear polarization” means that \vec{E} oscillates in a straight line, as opposed to circular or elliptical polarization. “Linear medium” means that the response of the medium is approximately linear in the strength of the applied fields.

⁵⁸As opposed to circular birefringence. Most people drop the qualifier “ordinary.”

⁵⁹As opposed to circular dichroism. Most people drop the qualifier “ordinary.”

Bibliography

Looking through this volume... was like roaming through an exquisite palace while its inhabitants slept.

— Orhan Pamuz

Many of the articles listed below are published in high-impact scientific journals. It is important to know that frequently such an article is only the tip of an iceberg: Many of the technical details (generally including specification of any physical model used) are relegated to a separate document called Supplementary Information, or something similar. The online version of the article will generally contain a link to that supplement.

- APPLEYARD, D C, VANDERMEULEN, K Y, LEE, H, & LANG, M J. 2007. Optical trapping for undergraduates. *Am. J. Phys.*, **75**(1), 5–14.
- BACKLUND, MIKAEL P, ARBABI, AMIR, PETROV, PETAR N, ARBABI, EHSAN, SAURABH, SAUMYA, FARAON, ANDREI, & MOERNER, W E. 2016. Removing orientation-induced localization biases in single-molecule microscopy using a broadband metasurface mask. *Nature Photon*, **10**(7), 459–+.
- BAKER, B B, & COPSON, E T. 1950. *The mathematical theory of Huygens' principle*. Oxford UK: Oxford Univ. Press.
- BARGMANN, V. 1954. On unitary ray representations of continuous groups. *The Annals of Mathematics*, **59**(1), 1.
- BARTKO, A, & DICKSON, R. 1999a. Imaging three-dimensional single molecule orientations. *J. Phys. Chem. B*, **103**, 11237–11241.
- BARTKO, A, & DICKSON, R. 1999b. Three-dimensional orientations of polymer-bound single molecules. *J. Phys. Chem. B*, **103**, 3053–3056.
- BECHHOEFER, J, & WILSON, S. 2002. Faster, cheaper, safer optical tweezers for the undergraduate laboratory. *Am. J. Phys.*, **70**(4), 393–400.
- BENEDEK, G B, & VILLARS, F M H. 2000. *Physics with illustrative examples from medicine and biology*. 2nd ed. Vol. 3. New York: AIP Press.
- BERESTETSKIĬ, V B, LIFSHITZ, E M, & PITAEVSKIĬ, L P. 1971. *Relativistic quantum theory*. New York: Pergamon Press.
- BERMAN, P R, & MALINOVSKY, V S. 2011. *Principles of laser spectroscopy and quantum optics*. Princeton NJ: Princeton Univ. Press.
- BERRY, M V. 1981. Singularities in waves and rays. *Pages 453–543 of: BALIAN, R D, KLEMAN, M, & POIRIER, J-P (Eds.), Physics of defects*. Amsterdam: North-Holland.
- BERRY, M V. 2015. Nature's optics and our understanding of light. *Contemporary Physics*, **56**(1), 2–16.
- BERRY, M V. 2017. *A half-century of physical asymptotics and other diversions: Selected works by Michael Berry*. World Scientific.
- BERRY, M V, & GEIM, A K. 1997. Of flying frogs and levitrons. *Eur. J. Phys.*, **18**(4), 307–313.
- BETH, RICHARD A. 1935. Direct Detection of the Angular Momentum of Light. *Phys. Rev.*, **48**(Sept.), 471–471.
- BETH, RICHARD A. 1936. Mechanical Detection and Measurement of the Angular Momentum of Light. *Phys. Rev.*, **50**(July), 115–125.

- BODANIS, D. 2005. *Electric universe: How electricity switched on the modern world*. New York: Three Rivers Press.
- BÖHMER, M, & ENDERLEIN, J. 2003. Orientation imaging of single molecules by wide-field epifluorescence microscopy. *J. Opt. Soc. Am. B*, **20**, 554–559.
- BOHREN, & HUFFMAN. 1993. *Absorption and scattering of light by small particles*. New York: Wiley.
- BOHREN, C F, & CLOTHIAUX, E E. 2006. *Fundamentals of atmospheric radiation*. Weinheim: Wiley-VCH.
- BORK, ALFRED M. 2005. Maxwell, Displacement Current, and Symmetry. *Am. J. Phys.*, **31**(11), 854–859.
- BORN, M, & WOLF, E. 1999. *Principles of optics*. 7th ed. Cambridge UK: Cambridge Univ. Press.
- BRESSLOFF, P C. 2014. *Waves in neural media*. New York: Springer.
- BURKE, W L. 1985. *Applied differential geometry*. Cambridge UK: Cambridge Univ. Press.
- BUSTAMANTE, C, SMITH, S B, LIPHARDT, J, & SMITH, D. 2000. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.*, **10**, 279–285.
- BUTT, H-J, & KAPPL, M. 2018. *Surface and interfacial forces*. 2d ed. Wiley-VCH.
- CAGNET, M, FRANCON, M, & THRIERR, J C. 1962. *Atlas optischer Erscheinungen. Atlas de phénomènes d'optique. Atlas of optical phenomena*. Berlin: Springer.
- CALDWELL, DENNIS J, & EYRING, H. 1971. *The theory of optical activity*. New York NY: Wiley Interscience.
- CANTOR, C R, & SCHIMMEL, P R. 1980. *Biophysical chemistry part II: Techniques for the study of biological structure and function*. San Francisco: W. H. Freeman and Co.
- CHEN, B-C, LEGANT, W R, WANG, K, SHAO, L, MILKIE, D E, DAVIDSON, M W, JANETOPOULOS, C, WU, X S, HAMMER, J A, LIU, Z, ENGLISH, B P, MIMORI-KIYOSUE, Y, ROMERO, D P, RITTER, A T, LIPPINCOTT-SCHWARTZ, J, FRITZ-LAYLIN, L, MULLINS, R D, MITCHELL, D M, BEMBENEK, J N, REYMANN, A-C, BOHME, R, GRILL, S W, WANG, J T, SEYDOUX, G, TULU, U S, KIEHART, D P, & BETZIG, E. 2014. Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, **346**(6208), 1257998–1257998.
- CHEN, B G-G, DERBES, D, GRIFFITHS, D, HILL, B, SOHN, R, & TING, Y-S (Eds.). 2019. *Lectures of Sidney Coleman on quantum field theory*. Singapore: World Scientific.
- CORLISS, W R. 1984. *Rare halos, mirages, anomalous rainbows and related electromagnetic phenomena*. Glen Arm MD: Sourcebook project.
- COWLEY, A. C., FULLER, N. L., RAND, R. P., & PARSEGAN, V. A. 1978. Measurements of repulsive forces between charged phospholipid bilayers. *Biochemistry*, **17**, 3163–3168.
- CRAIG, D P, & THIRUNAMACHANDRAN, T. 1998. *Molecular quantum electrodynamics: An introduction to radiation-molecule interactions*. New York: Dover Publications.
- CURTIS, JENNIFER, & GRIER, DAVID. 2003. Structure of Optical Vortices. *Phys. Rev. Lett.*, **90**(13), 133901.
- DAVIS, RICHARD S. 2017. Determining the value of the fine-structure constant from a current balance: Getting acquainted with some upcoming changes to the SI. *Am. J. Phys.*, **85**(5), 364.
- DE VRIES, H, & IUCR. 1951. Rotatory power and other optical properties of certain liquid crystals. *Acta Crystallographica*, **4**(3), 219–226.
- DODELSON, S. 2003. *Modern cosmology*. San Diego CA: Academic Press.
- DUBROVIN, B A, FOMENKO, A T, & NOVIKOV, S P. 1992. *Modern geometry—methods and applications*. 2nd ed. Vol. 1. New York: Springer-Verlag.
- DURNIN, J. 1987. EXACT-SOLUTIONS FOR NONDIFFRACTING BEAMS .1. THE SCALAR THEORY. *J. Opt. Soc. Am. A*, **4**(4), 651–654.
- DURNIN, J, MICELI, JJ, & EBERLY, JH. DIFFRACTION-FREE BEAMS. *J. Opt. Soc. Am. A*, **3**(13), P128.
- DURNIN, J, EBERLY, J, & MICELI, J. 1988. Comparison of Bessel and Gaussian beams. *Opt. Lett.*

- EID, JOHN, FEHR, ADRIAN, GRAY, JEREMY, LUONG, KHAI, LYLE, JOHN, OTTO, GEOFF, PELUSO, PAUL, RANK, DAVID, BAYBAYAN, PRIMO, BETTMAN, BRAD, BIBILLO, ARKADIUSZ, BJORNSON, KEITH, CHAUDHURI, BIDHAN, CHRISTIANS, FREDERICK, CICERO, RONALD, CLARK, SONYA, DALAL, RAVINDRA, DEWINTER, ALEX, DIXON, JOHN, FOQUET, MATHIEU, GAERTNER, ALFRED, HARDENBOL, PAUL, HEINER, CHERYL, HESTER, KEVIN, HOLDEN, DAVID, KEARNS, GREGORY, KONG, XIANGXU, KUSE, RONALD, LACROIX, YVES, LIN, STEVEN, LUNDQUIST, PAUL, MA, CONGCONG, MARKS, PATRICK, MAXHAM, MARK, MURPHY, DEVON, PARK, INSIL, PHAM, THANG, PHILLIPS, MICHAEL, ROY, JOY, SEBRA, ROBERT, SHEN, GENE, SORENSON, JON, TOMANEY, AUSTIN, TRAVERS, KEVIN, TRULSON, MARK, VIECELI, JOHN, WEGENER, JEFFREY, WU, DAWN, YANG, ALICIA, ZACCARIN, DENIS, ZHAO, PETER, ZHONG, FRANK, KORLACH, JONAS, & TURNER, STEPHEN. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, **323**(5910), 133–138.
- ELMORE, W C, & HEALD, M A. 1969. *Physics of waves*. New York: Dover Publications.
- FEYNMAN, R P, LEIGHTON, R, & SANDS, M. 2010a. *The Feynman lectures on physics*. New millennium ed. Vol. 1. New York: Basic Books. Free online: <http://www.feynmanlectures.caltech.edu/>.
- FEYNMAN, R P, LEIGHTON, R, & SANDS, M. 2010b. *The Feynman lectures on physics*. New millennium ed. Vol. 2. New York: Basic Books. Free online: <http://www.feynmanlectures.caltech.edu/>.
- FEYNMAN, R P, HIBBS, A R, & STYER, D F. 2010c. *Quantum mechanics and path integrals*. Emended ed. Mineola NY: Dover.
- FIZEAU, M. 1859. Sur les hypothèses relatives à l'éther lumineux, et sur une expérience qui paraît démontrer que le mouvement des corps change la vitesse avec laquelle la lumière se propage dans leur intérieur. *Ann. de Chim. et de Phys.*, **3e série t. LVII**, 385–404.
- FLEISCH, D. 2008. *A student's guide to Maxwell's equations*. Cambridge Univ. Press.
- FLEISCH, D. 2012. *A student's guide to vectors and tensors*. Cambridge Univ. Press.
- FOSTER, J, & NIGHTINGALE, J D. 2006. *A short course in general relativity*. 3rd ed. Springer.
- FRASER, ALISTAIR B. 1983a. Chasing rainbows. *Weatherwise*, **36**(6), 280–289.
- FRASER, ALISTAIR B. 1983b. Why can the supernumerary bows be seen in a rain shower? *J. Opt. Soc. Am.*, **73**(12), 1626–1628.
- FREEMAN, R, KING, J, & LAFYATIS, G. 2019. *Electromagnetic radiation*. Oxford UK: Oxford Univ. Press.
- GALISON, P. 2003. *Einstein's clocks, Poincaré's maps: Empires of time*. W.W. Norton.
- GARRITY, T F. 2002. *All the mathematics you missed - but need to know for graduate school*.
- GHOSH, AMBARISH, FAZAL, FURQAN M, & FISCHER, PEER. 2007. Circular differential double diffraction in chiral media. *Opt. Lett.*, **32**(13), 1836–1838.
- GILLESPIE, DANIEL T. 1998. The mathematics of Brownian motion and Johnson noise. *Am. J. Phys.*, **64**(3), 225–240.
- GINZBURG, V. L. 1989. *Applications of electrodynamics in theoretical physics and astrophysics*. Gordon and Breach.
- GRATIY, SERGEY L, HALNES, GEIR, DENMAN, DANIEL, HAWRYLYCZ, MICHAEL J, KOCH, CHRISTOF, EINEVOLL, GAUTE T, & ANASTASSIOU, COSTAS A. 2017. From Maxwell's equations to the theory of current-source density analysis. *Eur. J. Neurosci.*, **45**(8), 1013–1023.
- GRIER, DAVID G. 2003. A revolution in optical manipulation. *Nature*, **424**(6950), 810–816.
- GRODZINSKY, ALAN J. 2011. *Fields, forces, and flows in biological systems*. London UK: Garland Science.
- HASEGAWA, A, & KODAMA, Y. 1995. *Solitons in optical communications*. Oxford UK: Oxford Univ. Press.
- HECHT, E. 2002. *Optics*. 4th ed. Reading, MA: Addison-Wesley.
- HIRSHFELD, A. 2006. *The electric life of Michael Faraday*. New York: Walker and Co.
- HOBBIE, R K, & ROTH, B J. 2015. *Intermediate physics for medicine and biology*. 5th ed. New York: Springer.
- HODGKIN, A L, & KATZ, B. 1949. Effect of sodium ions on the electrical activity of the giant axon of the squid. *J. Physiol. Lond.*, **108**, 37–77.

- HULST, H C VAN DE. 1957. *Light scattering by small particles*. New York: Wiley.
- HUNT, B J. 2012. Oliver Heaviside: A first-rate oddity. *Phys. Today*, **65**(11), 48.
- ISRAELACHVILI, J N. 1992. *Intermolecular and surface forces*. 2nd ed. London: Academic Press.
- IVANOV, D T, & NIKOLOV, S N. 2016. A New Way to Demonstrate the Rainbow. *Physics Teach.*, **54**(8), 460–463.
- JAGGER, W S, & SANDS, P J. 1999. A wide-angle gradient index optical model of the crystalline lens and eye of the octopus. *Vision Res.*, **39**(17), 2841–2852.
- JELLEY, J V. 1958. *Cerenkov radiation and its applications*.
- JOANNOPOULOS, J D, MEADE, R D, & WINN, J N. 2008. *Photonic crystals: molding the flow of light*. 2d ed. Princeton NJ: Princeton Univ. Press.
- JONES, P H, MARAGÒ, O M, & VOLPE, G. 2015. *Optical tweezers: Principles and applications*. Cambridge UK: Cambridge Univ. Press.
- KEENER, J, & SNEYD, J. 2009. *Mathematical physiology I: Cellular physiology*. 2d ed. New York: Springer.
- KINDER, J M, & NELSON, P. 2018. *A student's guide to Python for physical modeling*. Updated ed. Princeton NJ: Princeton Univ. Press.
- LA PORTA, ARTHUR, & WANG, MICHELLE D. 2004. Optical torque wrench: angular trapping, rotation, and torque detection of quartz microparticles. *Phys. Rev. Lett.*, **92**(19), 190801.
- LAKSHMINARAYANAN, VASUDEVAN, & ENOCH, J M. 2011. Biological waveguides. *Chap. 9 of: BASS, M, DECUSATIS, C M, ENOCH, J, LAKSHMINARAYANAN, V, LI, G, MACDONALD, C, MAHAJAN, V N, & VAN STRYLAND, E (Eds.), Handbook of optics*, Vol. 3. OSA/McGrawHill.
- LANDAU, L D, & LIFSHITZ, E M. 1977. *Quantum mechanics: Non-relativistic theory*. 3rd ed. New York: Pergamon Press.
- LANDAU, L D, & LIFSHITZ, E M. 1979. *The classical theory of fields*. 4th rev. english ed. with corrections ed. Oxford UK: Pergamon Press.
- LANDAU, L D, LIFSHITZ, E M, & PITAEVSKII, L P. 1984. *Electrodynamics of continuous media*. 2d ed. Oxford UK: Pergamon Press.
- LEE, SH, ROICHMAN, Y, & GRIER, DG. 2010. Optical solenoid beams. *Opt. Express*, **18**(7), 6988–6993.
- LEE, JR., R L, & FRASER, A B. 2001. *The rainbow bridge: Rainbows in art, myth, and science*. University Park PA and Bellingham WA: Pennsylvania State University Press and SPIE Press.
- LEONHARDT, U. 2010. *Essential quantum optics*. Cambridge UK: Cambridge Univ. Press.
- LERCHE, I. 1977. The Fizeau effect: Theory, experiment, and Zeeman's measurements. *Am. J. Phys.*, **45**, 1154–1163.
- LEVENE, M J, KORLACH, J, TURNER, S W, FOQUET, M, CRAIGHEAD, H G, & WEBB, W W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. **299**(5607), 682–686.
- LEW, MATTHEW D, & MOERNER, W E. 2014. Azimuthal Polarization Filtering for Accurate, Precise, and Robust Single-Molecule Localization Microscopy. *Nano Lett.*, **14**(11), 6407–6413.
- LIPSON, A, LIPSON, S G, & LIPSON, H. 2011. *Optical physics*. 4th ed. Cambridge UK: Cambridge Univ. Press.
- LORRAIN, P, CORSON, D R, & LORRAIN, F. 1988. *Electromagnetic fields and waves - including electric circuits*. 3rd ed. W H Freeman and Co.
- LOUDON, R. 2000. *The quantum theory of light*. 3d ed. Oxford UK: Oxford Univ. Press.
- LYNCH, D K, & LIVINGSTON, W. 2001. *Color and light in nature*. 2nd ed. Cambridge Univ. Press.
- LYNCH, DAVID K, & SCHWARTZ, PTOLEMY. 1991. Rainbows and fogbows. *Appl. Opt.*, **30**(24), 3415–3420.
- MAHON, B. 2003. *The man who changed everything: The life of James Clerk Maxwell*. Chichester UK: Wiley.
- MAHON, BASIL. 2017. *The forgotten genius of Oliver Heaviside: A maverick of electrical science*. Prometheus Books.
- MALMIVUO, J, & PLONSEY, R. 1995. *Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetic Fields*. New York: Oxford University Press. <http://www.bem.fi/book/>.

- MANDEL, L, & WOLF, E. 1995. *Optical coherence and quantum optics*. Cambridge UK: Cambridge Univ. Press.
- MAXWELL, J, & JENKIN, F. 1865. LXI. On the elementary relations between electrical measurements, part I: Introductory. *Philosophical Magazine Series 4*, **29**(198), 436–460.
- MEHTA, A D, RIEF, M, SPUDICH, J A, SMITH, D A, & SIMMONS, R M. 1999. Single-molecule biomechanics with optical methods. *Science*, **283**, 1689–1695.
- MICHELSON, A, & MORLEY, E. 1886. Influence of motion of the medium on the velocity of light. *American Journal of Science*, **31**, 377–386.
- MICHILLI, D, SEYMOUR, A, HESSELS, J W T, SPITLER, L G, GAJJAR, V, ARCHIBALD, A M, BOWER, G C, CHATTERJEE, S, CORDES, J M, GOURDJI, K, HEALD, G H, KASPI, V M, LAW, C J, SOBEY, C, ADAMS, E A K, BASSA, C G, BOGDANOV, S, BRINKMAN, C, DEMOREST, P, FERNANDEZ, F, HELLBOURG, G, LAZIO, T J W, LYNCH, R S, MADDOX, N, MARCOTE, B, McLAUGHLIN, M A, PARAGI, Z, RANSOM, S M, SCHOLZ, P, SIEMION, A P V, TENDULKAR, S P, VAN ROOY, P, WHARTON, R S, & WHITLOW, D. 2018. An extreme magneto-ionic environment associated with the fast radio burst source FRB 121102. *Nature*, **553**(7687), 182–185.
- MINNAERT, M G J. 1993. *Light and color in the outdoors*. New York: Springer.
- MÖLLER, K. D. 2007. *Optics: Learning by computing, with model examples using Maple, MathCad, Mathematica, and MATLAB*. 2d ed. Springer.
- NAHIN, P J. 1988. *Oliver Heaviside, sage in solitude: The life, work, and times of an electrical genius of the Victorian age*. New York: IEEE Press.
- NELSON, P. 2014. *Biological physics: Energy, information, life—With new art by David Goodsell*. New York: W. H. Freeman and Co.
- NELSON, P. 2017. *From photon to neuron: Light, imaging, vision*. Princeton NJ: Princeton Univ. Press.
- NELSON, P, & DODSON, T. 2015. *Student's guide to MATLAB for physical modeling*.
<https://github.com/NelsonUpenn/PMLS-MATLAB-Guide>.
- NIEVES, JOSÉ F, & PAL, PALASH B. 1994. Third electromagnetic constant of an isotropic medium. *Am. J. Phys.*, **62**(3), 207–216.
- NOVOTNY, L, & HECHT, B. 2012. *Principles of Nano-Optics*. 2nd ed. Cambridge UK: Cambridge Univ. Press.
- NUSSENZVEIG, H M. 1992. *Diffraction effects in semiclassical scattering*. Cambridge UK: Cambridge Univ. Press.
- NYE, J F. 1999. *Natural focusing and fine structure of light: Caustics and wave dislocations*. Bristol UK: Institute of Physics Pub.
- OHANIAN, H. 2001. *Special relativity: A modern introduction*. Lakeville MN: Physics Curriculum and Instruction.
- OHANIAN, H C. 2007. *Classical electrodynamics*. 2d ed. Hingham MA: Infinity Science Press.
- OHANIAN, H C. 2008a. 0805.1400. *arxiv.org*, Aug., 1–10.
- OHANIAN, H C. 2008b. *Einstein's mistakes - the human failings of genius*. New York: Norton.
- OHANIAN, H C, & RUFFINI, R. 1994. *Gravitation and spacetime*. 2d ed. New York: Norton.
- PANTANO, C, GANAN-CALVO, A M, & BARRERO, A. 1994. Zeroth-order, electrohydrostatic solution for electro-spraying in cone-jet mode. *Journal of Aerosol Science*, **25**(6), 1065–1077.
- PARKER, ANDREW R, & TOWNLEY, HELEN E. 2007. Biomimetics of photonic nanostructures. *Nat. Nanotech.*, **2**(6), 347–353.
- PERKINS, THOMAS T. 2014. Ångström-precision optical traps and applications. *Annu. Rev. Biophys.*, **43**(1), 279–302.
- PESKIN, M E, & SCHROEDER, D V. 1995. *An introduction to quantum field theory*. Addison–Wesley.
- PHILLIPS, R, KONDEV, J, THERIOT, J, & GARCIA, H. 2012. *Physical biology of the cell*. 2nd ed. New York: Garland Science.
- PLANCHON, THOMAS A, GAO, L, MILKIE, D E, DAVIDSON, M W, GALBRAITH, JAMES A, GALBRAITH, CATHERINE G, & BETZIG, E. 2011. Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination. *Nat. Methods*, **8**(5), 417–423.

- POLLACK, G L, & STUMP, D R. 2002. *Electromagnetism*. San Francisco CA: Addison-Wesley.
- PURCELL, E M, & MORIN, D J. 2012. *Electricity and Magnetism*. 3d ed. Cambridge MA: Cambridge Univ. Press.
- RICHEY, LAUREN, STEWART, BAILEY, & PEATROSS, JUSTIN. 2006. Creating and Analyzing a Mirage. *Physics Teach.*, **44**(7), 460–464.
- RIOUX, MARC, & BÉLANGER, PIERRE-ANDRÉ. 1978. Ring pattern of a lens–axicon doublet illuminated by a Gaussian beam. *Appl. Opt.*, **17**(7), 1080–1088.
- ROHRlich, FRITZ. 2001a. The correct equation of motion of a classical point charge. *Physics Letters A*, **283**(5-6), 276–278.
- ROHRlich, FRITZ. 2001b. Why the principles of inertia and of equivalence hold despite self-interaction. *Physical Review D*, **63**(12), 621.
- ROICHMAN, YOHAJ, SUN, BO, ROICHMAN, YAEL, AMATO-GRILL, JESSE, & GRIER, DAVID. 2008. Optical Forces Arising from Phase Gradients. *Phys. Rev. Lett.*, **100**(1), 013602.
- ROTHMAN, T. 2003. *Everything's relative and other fables from science and technology*. New York: Wiley.
- ROVENCHAK, ANDRIJ, & KRYNYTSKYI, YURI. 2018. Radiation of the electromagnetic field beyond the dipole approximation. *Am. J. Phys.*, **86**(10), 727–732.
- RUFFNER, DAVID B, & GRIER, DAVID G. 2012. Optical forces and torques in nonuniform beams of light. *Phys. Rev. Lett.*, **108**(17), 173602.
- SAFRAN, S A. 2003. *Statistical thermodynamics of surfaces, interfaces, and membranes*. Boulder CO: Westview Press.
- SCHUTZ, B. 1990. *A first course in general relativity*. Cambridge MA: Cambridge Univ. Press.
- SHANKAR, R. 1995. *Basic training in mathematics: A fitness program for science students*. New York: Plenum.
- SHANKLAND, R S. 1964. Michelson-Morley experiment. *Am. J. Phys.*, **32**, 16–35.
- SHARMA, VIVEK, CRNE, MATIJA, PARK, JUNG OK, & SRINIVASARAO, MOHAN. 2009. Structural origin of circularly polarized iridescence in jeweled beetles. *Science*, **325**(5939), 449–451.
- SHEETZ, M P (Ed.). 1998. *Laser tweezers in cell biology*. Academic Press.
- SIEG, P G, BERNER, W, HARNISH, P K, & NELSON, P C. 2019. A demonstration of the infrared activity of carbon dioxide. *Physics Teach.*, **57**(4), 246–249.
- SIMON, D S. 2016. *A guided tour of light beams: From lasers to optical knots*. San Rafael CA: Morgan and Claypool.
- SMITH, A A T, BROWN, C V, & MOTTRAM, N J. 2007. Theoretical analysis of the magnetic Fréedericksz transition in the presence of flexoelectricity and ionic contamination. *Phys. Rev. E*, **75**(4), 315–8.
- SMITH, G S. 1997. *An introduction to classical electromagnetic radiation*. Cambridge UK: Cambridge Univ. Press.
- SMITH, STEPHEN P. 1999. Inexpensive optical tweezers for undergraduate laboratories. *Am. J. Phys.*, **67**(1), 26.
- SOHN, L L, SALEH, O A, FACER, G R, BEAVIS, A J, ALLAN, R S, & NOTTERMAN, D A. 2000. Capacitance cytometry: measuring biological cells one by one. *Proceedings of the National Academy of Sciences*, **97**(20), 10687–10690.
- SOMMERFELD, A. 1964a. *Electrodynamics*. New York: Academic Press.
- SOMMERFELD, A. 1964b. *Optics*. New York: Academic Press.
- SPIVAK, M. 1979. *A comprehensive introduction to differential geometry*. Vol. 2. Berkeley CA: Publish or Perish.
- STERNBERG, S. 1994. *Group theory and physics*. Cambridge UK: Cambridge Univ. Press.
- STRAUMANN, N. 2013. *General relativity*. 2d ed. New York: Springer.
- STREATER, R F, & WIGHTMAN, A S. 1964. *PCT, spin and statistics, and all that*. New York: W. A. Benjamin.
- STRUTT, J W (LORD RAYLEIGH). 1887. XVII. On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure. *Phil. Mag. Ser. 5*, **24**(147), 145–159.

- STRUTT, J W (LORD RAYLEIGH). 1888. XXVI. On the remarkable phenomenon of crystalline reflexion described by Prof. Stokes. *Phil. Mag. Ser. 5*, **26**(160), 256–265.
- STRUTT, J W (LORD RAYLEIGH). 1890. Measurements of the Amount of Oil Necessary in Order to Check the Motions of Camphor upon Water. *Proc. Roy. Soc Lond.*, **47**, 1–5.
- SVOBODA, K, & BLOCK, S M. 1994. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 247–285.
- TAYLOR, GEOFFREY. 1964. Disintegration of water drops in an electric field. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **280**(1382), 383–397.
- TOPRAK, E, ENDERLEIN, J, SYED, S, MCKINNEY, S A, PETSCHKEK, R G, HA, T, GOLDMAN, Y E, & SELVIN, P R. 2006. Defocused orientation and position imaging (DOPI) of myosin V. *Proc. Natl. Acad. Sci. USA*, **103**(17), 6495–6499.
- VAN DER STRATEN, P, & METCALF, H. 2016. *Atoms and molecules interacting with light: Atomic physics for the laser era*. Cambridge UK: Cambridge Univ. Press.
- VAN MAMEREN, J, WUITE, G J L, & HELLER, I. 2011. Introduction to optical tweezers: Background, system designs, and commercial solutions. *Meth. Mol. Biol.*, **783**, 1–20.
- VANDERLINDE, J. 2004. *Classical electromagnetic theory*. 2nd ed. Dordrecht NL: Kluwer Academic Publishers.
- WALMSLEY, I. 2015. *Light: A very short introduction*. Oxford UK: Oxford Univ. Press.
- WEINBERG, S. 1972. *Gravitation and cosmology*. New York: Wiley.
- WESS, J, & BAGGER, J. 1992. *Supersymmetry and supergravity*. 2nd ed. Princeton NJ: Princeton University Press.
- WHITTAKER, E T. 1951. *A history of the theories of aether and electricity*. London: T. Nelson.
- YABLONOVITCH, E. 2001. Photonic crystals: semiconductors of light. *Sci. Am.*, **285**(6), 47–51–54–5.
- ZANGWILL, A. 2013. *Modern electrodynamics*. Cambridge UK: Cambridge Univ. Press.
- ZHANG, Y-C. 1997. *Special relativity and its experimental foundations*. Singapore: World Scientific.

Index

Bold references indicate the main or defining instance of a key term. Symbol names and mathematical notation are defined in Appendix B.

- aberration
 - spherical, 229, 236, 240
- absolute temperature scale, 554
- absolute value of, **245**
- achiral object, 502
- action, **414**
- action functional, **414**
- action potential, 124, 135
 - mechanical analog, 132
 - speed, 135
- æther, **288**
- ampere (unit), **555**
- ampere (unit), **179**
- Ampère's law, 542
- angle, 557
- Ångstrom (unit), 556
- angular area, **8**
- angular frequency, **203, 244**
- angular velocity, **141**
- anion, **101**
- antenna
 - dipole, electric, 441
- arbitrary units, 557
- arclength, 240, 241
- area, angular, 557
- Avogadro number, 567
- axon, **85, 123**
 - giant, 136
- axoplasm, **65**

- back-EMF, **199**
- beauty, 359
- biaxial symmetry, **39, 78**
- bilayer membrane, **62**
- birefringence, **487, 501**
 - circular, *see* circular birefringence
 - ordinary, **514**
- Birkhoff's theorem, 34
- Bjerrum length, **105**
- Boltzmann constant, 567
- boost
 - galilean, **276**
- Born self-energy, **62, 100**
- bremstrahlung, 435

- cable equation
 - linear, 123
 - nonlinear, **134, 135**
- capacitance, **55, 107**
- capacitor, 107
- cation, **101**
- CD, *see* circular dichroism
- Celsius temperature scale, 554
- charge
 - density, **2**
 - bound, **497, 497, 603**
 - surface, 103
 - surface, bound, **57, 603**
 - surface, free, **57**
 - density, surface
 - bound, **496**
 - flux
 - 4-vector, **369**
 - bound, **497, 497, 498**
- charge flux, **2**
- charge renormalization, **113**
- chiral medium, *see* medium
- chirality, **502**
- circular birefringence, **502, 505–506**
- circular dichroism, **507**
- classical electron radius, **472**
- Clausius–Mossotti formula, 61, **573**
- closed differential form, **577**
- clumping catastrophe, 102
- CMBR, *see* cosmic microwave background radiation
- coherent states, **541**
- coion, 104, 112
- colloidal suspension, 102
- complex amplitude, **206**
- complex conjugate, **245**
- complex numbers, **245**
- compliance, **142**
- components of a tensor, **149**
- Compton scattering, **329**
- Compton wavelength, **475**
- conductance, **84**
 - resting, 133
- conductivity, **83**
 - electrical, 100, 121
- constitutive relation, **58**
 - electric, **500**
 - magnetic, **500**
- continuity equation, 82, **83, 166, 168, 201, 260, 286, 369, 371, 382–385, 393, 426, 451, 497, 585, 595, 597, 598, 603, 608**
 - charge, 417
 - energy of a spring, 286
 - momentum of a spring, 286
- coordinate system
 - cartesian, **150**
 - curvilinear, **47**
 - separable, *see* separable coordinate system
- cosmic microwave background radiation,
 - 280, 402, 473
- dipole anisotropy, 315
- Coulomb
 - attraction, 543
 - gauge, 536, 543
- coulomb (unit), **178, 554, 555**

- Coulomb gauge, 406, **578**
- Coulomb gauge condition, *see* gauge
- counterion, 101–111, 116, 117
 - cloud, 104
 - release, 111
- counterion release, 111
- covector, 4D, **362**
- creation and destruction operators, **541, 563**
- cross-section
 - scattering
 - differential, **471**
 - Thomson, **472**
- cross-susceptibilities, **503**
- curl, **7**
 - free vector field, **8**
- current
 - density, 542, 543
 - surface, *see* flux, charge, 2D
 - density (term not used in this book), *see* flux
 - electric, 555
 - through ion channel, 125
- curvature, **233**
 - curve in plane, **68**
 - surface
 - Gauss, **68, 72, 73, 77**
 - mean, 68, **72, 72–74, 76, 79, 80**
 - principal, **72**
- cutoff, **214**

- d'Alembert operator, *see* wave operator
- d'Alembertian, *see* wave operator
- debye (unit), **180**
- Debye screening length, **112, 113**
- defocused orientation imaging, 535, 545
- degree (temperature unit), 554
- delta function, **9, 14, see** Dirac delta function
- dendrites, **85**
- depletion layer, **25**
- depolarization, **86**
- deRham cohomology, **172**
- dialectical materialism, 496
- diamagnetism, **191**
- dielectric, **56**
- dielectric constant, **58**
- dielectric susceptibility, *see* susceptibility
- differential forms, **172, 577**
- diffraction pattern, **406**
- diffuse charge layer, **101, 108**
- diffusion equation, 134
- dimensional analysis, **554, 555–557**
- dimensionless quantities, **554, 557**
- dimensions, **555**
- dipole

- anisotropy of CMBR, **315**
approximation
 electric, 544
doughnut, **405**, 545
electric
 moment, 544, 545
 pure, 572
radiation pattern, 535, 545, *see* dipole
 doughnut
 transition, **544**, 545
Dirac equation, **379**
dispersion, **501**, 567
dispersion interaction, **38**
dispersion relation, **207**, **210**, **214**
displacement, **58**
displacement, electric, **497**
dissociation, 117
divergence operator, **4**
divergence theorem, **7**, 16, 48, 416
DNA
 dissociation, 101
 Z form, 507
Doppler shift
 longitudinal, **312**
 transverse, **312**
dot product, **4**, **153**
drag force, **142**
dyad product, **141**, 141, 558, **575**
- eikonal
equation
 in medium, **233**
 vacuum, **232**
function, **231**
eikonal trial solution, **231**
Einstein ring, **242**
Einstein thinking, **325**, 328, 346, 352, 374,
 379, 382, 385, 421, 423, 424, 510,
 599
Einstein, Albert, 17
electret, 500
electric
 charge, 555
 field, 535, 542
electric double layer, **25**, **101**
electrocardiogram, **88**
electrodynamics
 classical, 545
 quantum, 538–546
electroencephalography, **87**
electrolyte, **108**, 113
electromotive force, **198**
electron, 542, 543
 charge, 567
 mass, 567
 spin, 546
 volt (unit), 556
electrophoretic flux, **97**
electrostatic potential, 104
electrostatics, 542
EMF, **198**
enantiomers, **502**, 506, 507, 510
endoscopy, **230**
energy
 alternative units, 556
 dimensions, 555
 particle
 newtonian, **325**
 relativistic, **327**
 photon, 281, 540
 potential
 of capacitor, 107
 electrostatic, 100, 114
 solar, 235
 thermal, room temperature, 567
energy-momentum flux tensor, **383**
entropic force, 114
entropy
 counterions, 101, 104, 107, 110, 111
erg (unit), 556
esu (unit), **182**
event, **273**
exact differential form, **577**
exterior derivative, **577**
eye
 fish, 240
Fahrenheit temperature scale, 554
far fields, **263**
far-field approximation, **437**
farad (unit), **180**
Faraday tensor, *see* tensor, 4D
 plane wave, 398
Faraday, Michael, 198
feedback, 131, 133
ferroelectric, **57**
ferromagnetism, **191**
Fick's law, 97
field
 electric, **2**
 intensity, 2
 magnetic, **2**
 field lines, **393**
 field point, **25**
 field quantization, 535–546
 first fundamental form, *see* tensor, 2D
 metric
 fixed point, 134
 flagellum
 bacterial, 142
 flux, **83**
 1D, **81**
 charge, **81**, **82**
 2D, 170
 4D, free, 510
 surface (bound), **498**
 electrophoretic, **97**
 focus, **229**, 229, 236
 forbidden transition, 545
 form, differential, **377**
 forms, *p*, *see* differential forms
 Foucault pendulum, 320
 frame of reference, **18**, *see* coordinate
 system on spacetime
 Franklin, Benjamin, 74, 281
 free energy
 counterions, 107, 108, 111, 114–115
 frequency, **203**
 friction
 coefficient, viscous, 97
galilean boost, *see* boost
galilean invariance, **273**
gauge choice
 Coulomb, **165**, 166, 167, 208–210, 260,
 261, 397, 425, 536, 578, 581, 583,
 599
 gauge condition
 Lorenz, *see* Lorenz gauge
 gauge fixing, **165**
 gauge invariance, **164**, **370**, **578**
 gauge transformation, **164**, **578**
 gauss (unit), **181**
 Gauss Law
 at a surface, 103
 Gauss law, 103, 105, 106
 electric, 57, 103, 542
 in bulk, 103
 gaussian units, 180–182
 general relativity, xxxvi, 49, 76, 292, 335,
 342, 357, 359, 362, 381, 551
 generator
 rotation, 446
 generator of rotation, **37**
 Golden Rule, 544, 545
 Gouy–Chapman layer, **107**, 108
 gradient-index (GRIN) lens, 240
 gravitation, 22
 force due to, 97
 potential, *see* potential
 Green function, **17**
 causal, **425**
 Laplace operator, **25**
 laplace operator, 261
 retarded, **425**
 group, **337**, **342**
 galilean, **276**, **277**
Hamiltonian operator, 538, 539, 543, 562
harmonic oscillator, 537–539
Heaviside, Oliver, xxxi, 199, 200, 202, 385,
 524, 620
helicity
 negative, **211**
 positive, **211**
helicity basis, **211**, 505
henry (unit), **180**
hertz (unit), 556
Hodge dual, **577**
Hodgkin, Alan, 131
homonuclear molecule, **442**
Huxley, Andrew, 131
- index
 dummy, **7**
 loose, **7**
index lowering, **362**
index of refraction, **227**, 567
 effective gravitational, 237, 241
 graded, 236, 240
index raising, **362**
indistinguishable particles, 541
induced dipole, **38**
induction
 magnetic, 2
 gaussian units, 569
inductor, **213**
inertial coordinate system
 Einstein (E-), **308**
 galilean (G-), **275**
inertial frame
 Einstein (E-), 308, 310, 312, 314, 315,
 318, 322, 323, 325, 329, 330, 336,
 341, 343, 349, 363, 367, 368, 421,
 457, 562–565, 589, 592–596, 599
 galilean (G-), 275, 308, 421, 563, 564
integrability lemma, **23**
interference, **249**
invariance, **274**
invariant inner product, **344**
invariant interval, **308**, 317, 326, **343**
 (1+1)D, **308**
 (3+1)D, **310**
ion, 97
ion channel, **63**
 voltage-gated, 132

- ion channels, **86, 125**
- isotropic medium, *see* medium
- joule (unit), **555**
- Joule heating, *see* ohmic heating
- Keesom interaction, **38**
- kelvin (unit), 556
- Kelvin temperature scale, 554
- kink, 132
- Kirchhoff laws, 121
- Kronecker symbol, **7**
- lagrangian density, **414**, 414–416, 418, 419
- Laplace equation, **22**, 570
- Laplace operator, *see* laplacian
- laplacian, **4**, 25, 42, 47–49, 51–53, 89, 90, 166, 261, 308, 411, 558, 572
- laser, 542, **546**
- law of refraction
 - generalized, 234
- Legendre
 - equation, **50**, 75
 - function, 80
 - function, fractional order, 75
 - polynomial, 75
- Levi-Civita symbol
 - 3D, **6**
- light
 - speed of, 281, 567
 - unpolarized, **258**
- light-cone coordinates, **317**
- lightlike separation, **344**, 344, 422–424, **590**, 591
- linear cable equation, **122**
- liter (unit), 556
- localization microscopy, 546
- London force, **38**
- Lorentz boost, **307**
- Lorentz force law, **3**, 569
- Lorentz group, **300, 311**
- Lorentz transformation
 - provisional, 297
- Lorenz gauge, **397**, 397, 398, 401, 402, 404, 406, 407, 410, 412, 423, 425, 426, 429, 437, 454, 455, 599–601
- lowering operator, **539**, 540
- macroion, 101–102
- magnetar, **216**
- magnetic field, 535
 - intensity (H), **498**
- magnetic flux, **197**
- magnetic susceptibility, *see* susceptibility
- mass
 - relativistic (deprecated term), 326
 - rest (deprecated term), 326
- mass defect, **328**
- matrix
 - orthogonal, **150, 337**
 - rotation, 72
- Maxwell's equations, 536, 537, 543
- mean
 - rate, 545
- mean-field approximation, **104**, 105
- medium
 - chiral, **502**
 - isotropic, **500**
- metric, **342**
- mho (unit), **180**
- mirage, 235
- molar (unit), 556
- mole (unit), 556
- moment
 - dipole, 571
 - electric
 - dipole, induced, **143**
 - magnetic dipole
 - tensor form, *see* tensor
 - vector form, **188, 579**
 - monopole, 571
 - multipole, 571
 - quadrupole, 571
- moment of inertia tensor, **141**
- moments
 - seemultipole moments, 623
 - zeroth and first, **32**
- momentum
 - 4D, **345**
 - electromagnetic, 537, 539, 563
 - particle
 - newtonian, **324**
 - relativistic, **326**
- monochromatic light, **256**
- monovalent ion, **104**
- Mössbauer effect, 332
- multipole approximation
 - radiation, **438**
 - ED (electric dipole), **438**, 440, 441, 443, 444, 452
 - EQ (electric quadrupole), 451
 - MD (magnetic dipole), 450, 451
- multipole moments, 571
 - electric dipole, **31**
 - electric monopole, **31**
 - electric quadrupole, **31**
- multipole potentials, 571
- Nernst–Planck formula, **98**, 99
- Nernst potential, **125**
- Nernst relation, 98, **99**
- neuron, **85**
- Newton
 - constant, 567
- Newton, Isaac, 281
- newtonian potential, **14**
- Noether theorem, 414, 417
- nonchiral object, *see* achiral
- nonlinear optics, 500
- normal coordinates, **72**
- normal vector, **7**
- normalization
 - of a vector, **4**
- null
 - experiment, 300, 302, 305
 - separation, **344, 590**, 592, 600
 - wavevector, 398, 599
- O(3) (group), 337, 340, 510, 584, 588
- O(3,1) (Lorentz group), 364, 378, 590
- observer, **18**, *see* coordinate system on spacetime
- occupation numbers, **540**
- oersted (unit), **181**
- ohm (unit), 84, **180**, 556
- Ohm's law, *see* ohmic material
- ohmic
 - heating, **84**
 - hypothesis (neuron), **125, 129, 131**
 - material, **83**, 83, 112, 143, 466, 606
- one-shot action potential model, 132–136
- optical activity, *see* circular birefringence
- optical rotatory power, 505, *see* circular birefringence
- optical tweezers, **38, 228**
- osmotic pressure, 108
- outer product, *see* dyad product
- p-forms, **577**
- paramagnetism, **191**
- passive-spread solution, **123**
- patch-clamp, **91**
- period, **204**
- permeability
 - magnetic, **500**
 - vacuum, **2, 178**
- permeability of vacuum, 567
- permittivity, **58, 500**
 - vacuum, **2, 178**
- permittivity of vacuum, 567
- phosphorescence, **546**
- photon
 - ground state, **540**
 - indistinguishability of, 541
 - momentum, 540
- photonic bandgap materials, **527**
- piezoelectricity, **57**, 500
- Planck
 - constant, 567
- plane wave, **203**
- plasma membrane, **91**
- Poincaré lemma, **164**
- Poincaré sphere, **257**
- point dipole, **35**
- Poisson equation, **22**, 105, 570
- Poisson–Boltzmann equation, **105**, 107, 109, 112–114, 117
 - boundary condition, 106, 107
- polarizability
 - anisotropic, 501
 - bulk, **57**
 - molecular, **38**
- polarizable medium, **572**
- polarization
 - circular, **211**
 - elliptical, **211**
 - of light, 536, 540, 542, 544, 545
 - basis vectors, **536**, 564
 - linear, **210**
 - magnetic, 191
 - partial, degree, **258**
- polarization vector, **210**
- polaroid filter, 516
- positron emission tomography, 24
- potential
 - 4D, **370**
 - electric, 535, 542, 543, 556, 570
 - electrostatic, **22**
 - multipole, 571
 - newtonian (gravitational), **14**
- power, 555
- Poynting
 - theorem, 385, 598
 - vector, 387, 451, 461, 525, 532, 598, 601, 608
- Poynting vector, 537
- pressure, **143**
 - radiation, 220, 224, 387
 - isotropic, 399
- Principle of Relativity:, **272**
- probability
 - density function, 535, 545
- probability distribution
 - Boltzmann
 - and Nernst relation, 99
 - and Poisson equation, 104, 105, 112

- product
 dyad, **141**, *see* dyad product
 tensor, *see* tensor product
 propagator, *see* Green function
 proper time, **343**
 pseudotensor, 152
 pure dipole, **35**
- quadrupole
 magnetic, **189**
- racemic mixture, 507
 radian (unit), 557
 radiation
 anapole, **452**
 rainbow, 254
 raising operator, **540**
 rank
 4-tensor, 353
 rapidity parameter, **309**
 ray
 equation, **234**, 240, 241
 of light, 229, 255
 ray of light, **231**
 ray optics, **232**
 Rayleigh cross-section, **473**
 real part, **245**
 refraction
 law of, 229, 240, 241
 Relativity Strategy, **278**, 298, 300, 311, 314
 residue, 102
 resistance, **84**
 electrical, 100
 resistivity, **84**
 response function, **57**, 500
 rest frame, **314**
 retarded time, **262**
 rotation matrix, **37**
- scalar
 3D, **337**
 4D (4-scalar), **343**, 351
 scalar product
 3D, **4**
 second fundamental form, *see* tensor, 2D
 curvature
 selection rule, **546**
 self-inductance, **197**
 separable coordinate system, **47**
 short circuit, 84
 siemens (unit), 84, **121**, **180**
 simultaneity, relativity of, **422**
 $SL(2, \mathbb{C})$ (group), 378, 379
 $SO(3)$ (group), 337, 377–379, 488, 510, 588
 solenoid, **196**
 solid angle, **8**
 soliton, **138**
 source point, **24**
 space constant, of axon, **122**
 spacelike separation, **344**, 422, 424, **590**
 spacetime, **273**
 spherical aberration, **229**
 spherical harmonics, 39
 spherical wave, **404**
 $Spin(3)$ (group), 378
 $Spin(3,1)$ (group), 378, 596
 spinor, 141
 spring constant tensor, **142**
- standard deviation, also called
 root-mean-square deviation
 (RMSD)
 relative (RSD), 542
 static system, **165**
 stationarity, **165**
 stationary-phase
 path, 240, 241
 steady state, 133
 steradian (unit), 557
 stereospecific binding, **102**
 Stokes parameters, **257**
 Stokes Theorem, **7**
 streamline, **230**
 stress tensor, **143**
 stress-energy tensor, 383
 $SU(2)$ (group), 359, 378, 588
 $SU(3)$ (group), 359
 $SU(5)$ (group), 550
 subgroup, **306**
 summation convention, **7**
 surface attraction, electrostatic, 111
 surface repulsion, electrostatic, 108
 susceptibility
 4-tensor operator, **510**
 dielectric, **57**
 susceptibility, dielectric, **500**
 susceptibility, magnetic, **500**
 symmetry, **273**
- Taylor cone, 75
 Taylor's theorem, 556
 tension
 interfacial or surface, 67, 68, **70**, 70, 71,
 73, 75, 80
 line, **70**, 70, 71
- tensor
 antisymmetric, 575
 4D, **330**, **351**, **353**, **363**
 Faraday, **353**
 product, 141, 378, 379, 575
 Riemann, 77
 symmetric, **142**, 145, 575
 3D, **140**, **336**
 compliance, **142**
 Levi-Civita, **145**
 magnetic dipole moment, **187**, **579**
 metric, **143**
 mobility, **143**, *see* mobility tensor
 moment of inertia, **141**, *see* moment
 of inertia tensor
 polarizability, **143**, *see* polarizability
 tensor
 rank 2, **141**, **338**
 rank 3, **145**
 spring constant, **142**, *see* spring
 constant tensor
 stress, *see* stress tensor
 totally antisymmetric, 145
 viscous drag, **142**, *see* viscous drag
 tensor
 2D, **140**
 curvature, 76, 78
 metric, 78
 Tensor Principle, 359
 3D, **340**
 4D, **351**
 tesla (unit), **180**
- test body, **3**, 570
 threshold stimulus, 134, 136
 time
 proper, **326**
 retarded, **439**
 time constant, of axon, **122**
 timelike separation, **344**, 422, **590**
 TIR, *see* total internal reflection
 TIRF microscopy, **230**
 total internal reflection, 227, **229**
 trace, **575**
 trajectory, **273**
 transfer matrix, **529**
 transformation
 active, **273**
 galilean, **276**
 Lorentz, *see* Lorentz transformation
 passive, **274**
 traveling wave, 123, 136
 twinlead cable, 56
- uncertainty relation, 545
 uniaxial symmetry, **39**, 78
 uniform transparent medium, 240
 units, 554–557
 base, **176**, 554
 Système Internationale (SI), **554**
- vacuum state, **541**
 van der Waals, **38**
- vector
 3D, **4**, **336**
 4D, **327**
 null, *see* null
 4D (4-vector), **342**, 351
 vector potential, 535, 536, 545
 velocity
 4D, **344**
 velocity addition
 galilean, **276**
- vesicle, 108
 viscous drag tensor, **182**
 viscous stresses, **143**
 volt (unit), **180**
 voltage-gating hypothesis, 132, 134, 136
 simplified, 131
- watt (unit), **555**
 wave equation, **283**
 inhomogeneous, **260**
 wave operator, **261**, 261, 317, **363**, 427,
 558, 561, **591**
 wavelength, **204**
 wavenumber, **203**, **244**
 spectroscopic (alternate definition not
 used in this book), 203
 wavevector, 536
 4D, **345**, **398**
 weak interaction, 318
 weber (unit), **180**
 Weyl equation, **379**
- Young–Laplace formula, 73, 76, 80
 Yukawa equation, **416**
- Zeeman
 effect, 215, 454
 experiment on light velocity, 300