

Key Notes in Plasma Physics

Hongyang Zhou

2026-01-08

Table of contents

Welcome	3
1 About	4
2 Introduction	5
2.1 Occurrence of Plasma in Nature	7
2.2 Definition of Plasma	9
2.3 Concept of Temperature	10
2.4 Debye Shielding	12
2.5 Criteria for Plasmas	16
2.6 Plasma Frequency	17
2.7 Spatial Scales	18
2.8 How to Study Plasma Physics	20
2.9 Common Acronyms	20
3 Math	21
3.1 Vector Identities	21
3.2 Differential of Line Segments	22
3.3 Lagrangian vs Eulerian Descriptions	23
3.4 Complex Analysis	23
3.5 Electromagnetic Four-Potential	27
3.5.1 Multipolar gauge	27
3.6 Helmholtz's Theorem	28
3.7 Toroidal and Poloidal Decomposition	28
3.8 Magnetic Dipole Field	30
3.9 Green's Function	33
3.10 Linearization	34
3.11 Wave Equations	34
3.11.1 Airy Function	37
3.11.2 Uniformly Valid WKBJ Solution Across the Turning Point	44
3.11.3 Stokes Phenomenon	47
3.11.4 Application of Stokes Lines	51
3.11.5 "Exact" WKBJ Solution	60
3.12 Normalization	61
3.13 Wavelet Analysis	61

4 Theoretical Mechanics	63
4.1 Generalized Coordinates	63
4.2 D'Alembert's Principle	63
4.3 Lagrangian Mechanics	64
4.4 Hamiltonian Mechanics	64
4.5 Hamilton-Lagrange Formalism of Lorentz Equation	65
5 Relativity	70
5.1 Velocity	70
5.2 Theory of Relativity	70
5.2.1 Four-velocity	71
6 Unit Conversion	73
6.1 Formula Conversion	73
6.1.1 Maxwell's Equations	74
6.2 Unit System Conversion	76
6.2.1 Example: $\rho, \rho\mathbf{u}, \mathcal{E}, \mathbf{B}$ to n, u, T	78
7 Single-Particle Motions	79
7.1 Uniform E and B Fields	80
7.1.1 $E=0$	80
7.1.2 Finite E	82
7.1.3 Gravitational Field	83
7.2 Nonuniform B Field	84
7.2.1 B B: Grad-B Drift	85
7.2.2 Curved B: Curvature Drift	86
7.2.3 B B: Magnetic Mirrors	88
7.3 Nonuniform E Field	91
7.4 Time-Varying E Field	93
7.5 Time-Varying B Field	95
7.6 Drifts in slowly changing arbitrary fields	96
7.7 Summary of Guiding Center Drifts	101
7.8 Adiabatic Invariants	102
7.8.1 Adiabatic Invariant of a Pendulum	103
7.8.2 Extension of WKB method to general adiabatic invariant	106
7.8.3 Proof for the general adiabatic invariant	108
7.8.4 The First Adiabatic Invariant	111
7.8.5 The Second Adiabatic Invariant	113
7.8.6 The Third Adiabatic Invariant	115
7.9 Ring Current	116
7.10 Particle Acceleration	119
7.11 Test Particle Model	123

8 Plasmas as Fluid	124
8.1 Definitions	124
8.2 Relation of Plasma to Ordinary Electromagnetics	126
8.2.1 Maxwell's Equations	126
8.2.2 Classical Treatment of Magnetic Materials	126
8.2.3 Classical Treatment of Dielectrics	128
8.2.4 The Dielectric Constant of a Plasma	128
8.3 Fluid Equations	129
8.3.1 Equation of Continuity	130
8.3.2 Momentum Equation	130
8.3.3 Equation of State	134
8.4 Two-Fluid Model	135
8.4.1 Five-moment	135
8.4.2 Six-moment	136
8.4.3 Five/Ten-moment	137
8.4.4 Characteristic wave speeds	139
8.5 Generalized Ohm's Law	139
8.5.1 Basic definitions and assumptions	140
8.5.2 Equation of motion for each species	141
8.5.3 Preconditions	143
8.5.4 Derivation	143
8.5.5 Simplification of Generalized Ohm's Law	145
8.5.6 Discussions	146
8.5.7 Ohm's Law in a Multi-Ion System	147
8.6 Magnetohydrodynamics	147
8.6.1 MHD Ohm's law	150
8.6.2 Ideal MHD	150
8.6.3 Hall MHD	159
8.7 Fluid Drifts Perpendicular to B	160
8.8 Fluid Drifts Parallel to B	164
8.9 The Plasma Approximation	166
9 Diffusion and Resistivity	168
9.1 Diffusion and Mobility in Weakly Ionized Gases	168
9.2 Collision Parameters	168
9.3 Diffusion Parameters	169
9.4 Decay of a Plasma by Diffusion	170
9.4.1 Ambipolar Diffusion	170
9.4.2 Diffusion in a Slab	172
9.4.3 Diffusion in a Cylinder	174
9.5 Steady State Solutions	174
9.5.1 Constant Ionization Function	175
9.5.2 Plane Source	175

9.5.3	Line Source	175
9.6	Recombination	176
9.7	Diffusion Across a Magnetic Field	176
9.7.1	Ambipolar Diffusion Across B	177
9.8	Collisions in Fully Ionized Plasmas	178
9.8.1	Plasma Resistivity	179
9.8.2	Mechanics of Coulomb Collisions	180
9.8.3	Physical Meaning of	181
9.8.4	Numerical Values of	183
9.8.5	Pulsed Currents	184
9.8.6	Collisions Between Species	184
9.8.7	Conductivity Tensor	184
9.9	Diffusion of Fully Ionized Plasmas	187
9.9.1	Small R_m condition	188
9.9.2	Large R_m condition	192
9.10	Pitch Angle Scattering	194
9.10.1	Definition of Pitch Angle Scattering	195
9.11	Cosmic Ray Diffusion	197
9.11.1	Perpendicular Superdiffusion of Mirroring CRs	198
9.11.2	Parallel Mirror Diffusion	198
9.11.3	Transition between Mirror Diffusion and Scattering Diffusion	198
9.12	Solutions of the Diffusion Equation	199
9.12.1	Time Dependence	199
9.12.2	Time-Independent Solutions	199
9.13	Bohm Diffusion and Neoclassical Diffusion	199
10	Waves	200
10.1	Basic Properties	200
10.2	Polarization	200
10.2.1	Dispersion Relation	200
10.2.2	Damping/Growth Rate	201
10.2.3	Dielectric Function	201
10.3	Plasma Oscillations	202
10.4	Classification of EM Waves in Uniform Plasma	205
10.5	ES vs. EM Waves	205
10.6	Cold Uniform Plasma	207
10.6.1	Wave Modes	212
10.6.2	Resonances	212
10.6.3	Cutoffs	213
10.6.4	Polarizations	214
10.6.5	Low Frequency Limit	214
10.6.6	Faraday Rotation	216
10.6.7	Perpendicular Wave Propagation	217

10.7	MHD Waves	219
10.7.1	Cold MHD	219
10.7.2	Hot MHD	220
10.7.3	Alfvén Wave	221
10.7.4	Fast and Slow Wave	227
10.7.5	MHD Waves in CPDR	229
10.7.6	Limitations of the MHD model	230
10.8	Two-fluid model of Alfvén modes	231
10.8.1	Two-fluid shear modes	236
10.8.2	Two-fluid compressional modes	240
10.8.3	Differences between the two-fluid and MHD descriptions	241
10.8.4	KAW properties	241
10.9	Kinetic Slow Modes	243
10.10	Particle Motions	244
10.11	Cold Nonuniform Plasma	244
10.11.1	Simple EM Wave	244
10.11.2	Electron Cyclotron Resonance Heating	248
10.11.3	O-mode	249
10.12	Warm Uniform Plasma	249
10.12.1	Two-fluid approach	250
10.12.2	Vlasov approach	251
10.12.3	Ion-acoustic wave	253
10.13	Electrostatic Wave in a Magnetized Plasma	254
10.14	CMA Diagram	257
10.15	Wave-Particle Interactions	257
10.15.1	Pitch Angle Scattering	259
10.16	Whistler Wave	259
10.16.1	Ducting of Whistler Waves	262
10.16.2	Whistlers in EMHD	262
10.16.3	Importance of Whistler	263
10.17	Wave Identification	263
10.18	Animation of Waves	265
11	EMHD	266
11.1	Key Concepts in EMHD	266
11.2	Equations of EMHD	266
12	Kinetic Theory	268
12.1	Phase Space	268
12.2	Distribution Function	269
12.2.1	Moments of the distribution function	269
12.2.2	Maxwellian Distribution	271
12.2.3	Kappa Distribution	275

12.2.4	Entropy of a distribution	275
12.2.5	Effect of collisions on entropy	278
12.3	Equations of Kinetic Theory	281
12.3.1	Vlasov equation	281
12.3.2	Boltzmann equation	283
12.4	Derivation of the Fluid Equations	286
12.5	Plasma Oscillations and Landau Damping	292
12.6	The Meaning of Landau Damping	296
12.6.1	The Kinetic Energy of a Beam of Electrons	298
12.6.2	The Effect of Initial Conditions	300
12.7	A Physical Derivation of Landau Damping	300
12.7.1	The Resonant Particles	305
12.7.2	Two Paradoxes Resolved	305
12.8	BGK and Van Kampen Modes	306
12.9	Ion Landau Damping	306
12.9.1	The Plasma Dispersion Function	307
12.9.2	Ion Waves and Their Damping	309
12.10	Kinetic Effects in a Magnetic Field	311
12.10.1	The Hot Plasma Dielectric Tensor	312
12.10.2	Bernstein Waves	313
13	Stability	316
13.1	Analogy of the Energy Principle	318
13.2	Implication of Single Particle Motion on Instabilities	318
13.2.1	Equilibrium stability of a plasma from drifts	318
13.2.2	Stability of magnetic mirror in the scope of single particle motion	319
13.2.3	Stability in the Tokamak	322
13.3	Two-Stream Instability	322
13.4	Rayleigh-Taylor Instability	328
13.4.1	2-Fluid Diffuse Boundary Model	328
13.4.2	Single fluid MHD method	332
13.4.3	2-fluid sharp boundary model	334
13.5	Kelvin-Helmholtz instability	337
13.5.1	Diocotron instability on electron sheet	340
13.6	MHD Stability	343
13.6.1	Harris Current Sheet	343
13.6.2	-Pinch and Z-Pinch	345
13.6.3	Force-Free Field	345
13.7	MHD Modes	347
13.7.1	Kink Mode	347
13.7.2	Sausage Mode	357
13.7.3	Kink Mode	358
13.7.4	Ballooning Mode	359

13.7.5 Tearing Mode	359
13.8 Magnetic Reconnection	363
13.8.1 Reconnection Rate	364
13.8.2 Hall Field	365
13.8.3 Kinetic Signatures of Diffusion Region	368
13.8.4 D-shape Distribution	370
13.8.5 Classification of Discontinuity Types	371
13.8.6 Reconnection Efficiency	371
13.8.7 Dissipation Mechanisms	372
13.8.8 Sweet-Parker Solution	372
13.8.9 Petschek Solution	373
13.8.10 3D Nature of Reconnection	374
13.9 Lower Hybrid Drift Instability	375
13.10 Kinetic Mode	376
13.11 Pressure Anisotropy Instabilities	376
13.11.1 Firehose Instability	376
13.11.2 Mirror Instability	377
13.11.3 Weibel Instability	379
13.12 Cyclotron Instability	379
13.12.1 Ion Cyclotron Instability	379
14 Kinetic MHD	380
14.1 Firehose Instability: Linear Theory	387
14.2 Mirror Instability: Linear Theory	389
14.2.1 Comparison With Slow Mode	394
14.3 Origin of Pressure Anisotropy	396
14.4 Remarks	397
14.4.1 Remark I	397
14.4.2 Remark II	398
15 Gyrokinetics	400
15.1 A Brief Historical Review	400
15.2 Coordinate Transformation	400
15.3 The Gyrokinetic Ordering	403
15.3.1 From Kinetics to Gyrokinetics	403
15.4 A Lagrangian Approach	405
15.4.1 Linear Gyrokinetics	406
15.4.2 Nonlinear Gyrokinetics	411
15.5 Applications	415
15.6 Limitations	416
16 Field Line Resonance	417
16.1 Historial Review	417

16.2	Theory	418
16.2.1	Axisymmetric Spherical Coordinates	419
16.2.2	Cylindrical Coordinates	421
16.2.3	Cartesian Box Model	423
16.3	Energy Consideration	441
16.4	Non-MHD Effects	442
16.5	Nonlinear Effects	444
16.6	3D FLR	444
17	Nonlinear Effects	445
17.1	The Necessity for Sheaths	446
17.1.1	The Planar Sheath Equation	447
17.1.2	The Bohm Sheath Criterion	448
17.1.3	The Child-Langmuir Law	449
17.1.4	Electrostatic Probes	450
17.2	Ion Acoustic Shock Waves	451
17.2.1	The Sagdeev Potential	452
17.2.2	The Critical Mach Numbers	453
17.2.3	Wave Steepening	454
17.2.4	Double Layers	455
17.3	The Ponderomotive Force	455
17.4	Parametric Instabilities	458
17.4.1	Coupled Oscillators	458
17.4.2	Frequency Matching	459
17.4.3	Instability Threshold	461
17.4.4	Physical Mechanism	462
17.4.5	The Oscillating Two-Stream Instability	463
17.4.6	The Parametric Decay Instability	463
17.5	Plasma Echoes	463
17.6	Nonlinear Landau Damping	463
17.7	Equations of Nonlinear Plasma Physics	463
17.7.1	The Korteweg-de Vries Equation	463
17.7.2	The Nonlinear Schrödinger Equation	464
18	Turbulence	465
18.1	Weak and Strong Turbulence	466
18.2	MHD	467
18.2.1	Comparison with Hydrodynamics	468
18.2.2	Incompressible MHD equations	468
18.2.3	Coupling between charged fluid and magnetic field	469
18.2.4	Importance of Alfvén Waves	471
18.2.5	Wave Interaction Phenomenology	471
18.2.6	Nondimensional parameters	472

18.2.7	Critical balance	472
18.2.8	Residual energy	473
18.2.9	Normalized cross helicity	474
18.3	Solar Wind Turbulence	476
18.3.1	Taylor's hypothesis	476
18.3.2	Solar Wind Power Spectrum	476
18.4	Turbulence and Reconnection	480
19	Geometry	482
19.1	Local Coordinate System	482
19.1.1	Dimensionality	483
19.2	Reconnection Identification	484
19.2.1	2D	484
19.2.2	3D	485
19.2.3	Critical Points	485
19.2.4	Separatrix	487
19.3	Helicity	489
19.4	Magnetic Helicity	490
19.5	Cross Helicity	492
19.6	Flux Rope Identification	493
19.7	Magnetopause Identification	493
20	Solar Wind	494
20.1	Parker Solar Wind Model	496
20.2	Parker Spiral	499
20.3	Anisotropy in the Solar Wind	499
20.4	Switchbacks	501
20.5	MHD Description	502
20.6	Wave Modes in the Solar Wind	502
21	Shock	503
21.1	MHD Theory	505
21.1.1	Evolutionarity	510
21.1.2	Coplanarity	510
21.1.3	Criticality	511
21.1.4	Parallel Shock	515
21.1.5	Perpendicular Shock	518
21.1.6	Oblique Shock	519
21.1.7	Switch-On and Switch-Off Shocks	523
21.2	Double Adiabatic MHD	524
21.2.1	Perpendicular Shock	526
21.3	Subcritical Shocks	527

21.4	Supercritical Perpendicular Shock	528
21.4.1	Particle Dynamics	528
21.4.2	Foot Formation and Acceleration	532
21.4.3	Shock Potential Drop	534
21.4.4	Shock Structure	537
21.4.5	Shock Reformation	543
21.4.6	Ion Dynamics	548
21.4.7	Electron Dynamics	554
21.4.8	Stationarity	554
21.5	Supercritical Parallel Shock	556
21.5.1	Basic Properties	557
21.5.2	Foreshock	558
21.5.3	Low-Frequency Upstream Waves	571
21.5.4	Electron Foreshock	583
21.5.5	Quasi-Parallel Shock Reformation	590
21.5.6	Parallel Shock Particle Reflection	607
21.6	Instabilities and Waves	608
21.6.1	Ion Instabilities $\omega \leq \omega_{ci}$	609
21.6.2	Electron Instabilities and Radiation $\omega \sim \omega_{pe}$	612
21.6.3	Whistlers	612
21.7	Shock Particle Reflection	613
21.7.1	Specular Reflection	615
21.7.2	Consequences of Shock Reflection	616
21.8	Shock Particle Acceleration	617
21.8.1	Electron Acceleration and Heating at Shocks	621
21.9	Foreshock Transients	621
21.9.1	Hot Flow Anomaly	621
21.9.2	Spontaneous Hot Flow Anomaly	623
21.9.3	Foreshock Bubble	623
21.9.4	Foreshock Cavity	623
21.9.5	Foreshock Caviton	624
21.9.6	Foreshock Compressional Boundary	624
21.9.7	SLAMS and Shocklets	624
21.9.8	Density Hole	625
21.10	Location of Shocks	625
21.11	Earth Bow Shock	625
22	Magnetosphere	626
22.1	Earth's Magnetic Environment	626
22.1.1	Basic Structures	626
22.1.2	The Dipole Field	627
22.1.3	Current Systems	629
22.1.4	Geomagnetic Activity Indices	637

22.2	Particles	639
22.2.1	Outer Magnetosphere	639
22.2.2	Inner Magnetosphere	640
22.2.3	Cosmic Rays	643
22.3	Magnetospheric Dynamics	644
22.3.1	Magnetospheric Convection	645
22.3.2	Geomagnetic Storms	647
22.3.3	Substorms	650
22.4	ULF Waves	652
22.4.1	Pc1 & Pc2	653
22.4.2	Pc3 & Pc4	657
22.4.3	Pc5	658
22.4.4	Sinks	659
22.4.5	Interaction with Energetic Particles	659
22.4.6	EMIC Wave Events	662
22.5	Bow Shock	664
22.5.1	Foreshock	665
22.5.2	Transmission of Waves	669
22.6	Magnetosheath	669
22.6.1	Jets	669
22.7	Magnetosphere	670
22.7.1	Magnetopause	670
22.7.2	Transport of mass, momentum, and energy	672
22.7.3	Dungey Cycle	673
22.7.4	Vasyliunas Cycle	673
22.7.5	Low-Latitude Boundary Layer	673
23	Ionosphere	674
23.1	Current Systems	674
23.2	M-I Coupling	675
23.2.1	Caveats	677
23.3	Ionosphere Modeling	679
23.3.1	Chemistry	680
23.3.2	Ion Advection	680
23.4	Ionosphere Waveguide	681
23.5	Incoherent Scatter Radar	681
24	Radiation Belt	682
24.1	Waves in the Radiation Belt	683
24.1.1	Sources of Waves	684
24.2	Modeling	684
24.2.1	Coupling physical processes at different time scales	687
24.3	Loss	687

24.4 Plasmasphere	687
25 Aurora	688
25.1 Auroral Light Emission	688
25.2 Diffuse Aurora	689
25.3 Auroral Waves and Ion Heating	691
25.4 Substorms and Discrete Aurora	692
25.4.1 Knight's relation	694
25.4.2 Current circuit	695
25.4.3 Omega Band Auroras	697
25.4.4 Theta Auroras	698
25.4.5 Throat Auroras	698
25.4.6 Pulsation Auroras	698
25.5 Region I and Region II Currents	698
25.6 Aurora Oval	698
25.7 Summary	699
26 Guiding Center	700
26.1 Summarized Version	700
26.1.1 Basic Concepts and Variables	700
26.1.2 The Effective Fields	701
26.1.3 Hamiltonian	701
26.1.4 Equations of Motion	702
26.1.5 Numerical Implementation Guide	702
26.1.6 Conversion from Full Phase Space to Guiding Center	703
26.1.7 Conversion from Guiding Center to Full Phase Space	705
26.2 Classical Guiding Center Theory	706
26.2.1 Lorentz equation in tensor form	706
26.2.2 Fundamental assumptions	707
26.2.3 Derivations of GC equation	707
26.2.4 First order solution for the GCA equations of motion	711
26.2.5 Recursive solution	712
26.2.6 Energy conservation	712
26.3 Validity and Higher-Order Corrections	713
26.3.1 Numerical implementation	714
26.3.2 Nonrelativistic case	714
26.4 Hamiltonian Theory of Guiding Center Motion	714
26.4.1 Noncanonical Hamiltonian guiding center theory	715
27 Cosmic Ray	719
27.1 Basic Definitions	719
27.2 CR Continuity Equations	721
27.3 Parker Transport Equation	723

27.4	Quasilinear Theory	724
27.4.1	Derivation of the Fokker-Planck equation	725
27.4.2	The diffusion approximation	729
27.4.3	Computation of transport coefficients	731
27.4.4	Turbulence geometries and spectra	731
27.4.5	Field-line random walk	735
27.4.6	Short-comings of QLT	736
27.5	Test Particle Simulation	737
28	Boundary Conditions	739
28.1	Conducting Boundary	739
28.2	Perfectly Matched Layer	740
28.2.1	Derivation	741
29	Particle-in-Cell	742
29.1	Phase Space Sampling	742
29.2	Artificial Scalings	743
29.3	Numerical Heating and Thermalization	744
29.4	Computational Cost	745
29.5	f Method	745
30	Hybrid Methods	747
30.1	Classical Hybrid Model	747
30.1.1	Pros and Cons	750
30.1.2	Role of Electrons	750
30.2	Low Density Treatment	751
30.3	Finite Electron Inertia	752
30.4	Comparison with Hall MHD	753
30.5	Normalization	753
30.6	Numerical Stability	755
30.6.1	Finite Grid Instability	756
30.7	Boundary Conditions	756
30.7.1	Open Boundary Conditions	756
30.7.2	Magnetosphere Inner Boundary Conditions	757
30.8	Tests	758
30.8.1	Proton Cyclotron Anisotropy Instability	758
30.8.2	Landau Damped Ion Acoustic Wave	758
30.8.3	Magnetic Reconnection Island Coalescence	759
30.8.4	Collisionless Shock	759
31	Ray Tracing	760
31.1	Backgrounds	760
31.2	How it Works	761

32 Tests	762
32.1 Errors	762
32.2 Wave Tests	762
32.2.1 Linear wave	762
32.2.2 3D Adiabatic MHD	768
32.2.3 Anisotropic MHD	768
32.2.4 Circularly polarized Alfvén wave	769
32.2.5 Whistler wave	770
32.2.6 Light wave	772
32.2.7 Firehose instability	772
32.2.8 Mirror mode instability	772
32.2.9 Langmuir wave	772
32.3 Shock Tests	773
32.3.1 Brio-Wu shock tube	773
32.3.2 Ryu and Jones Test 2A	773
32.3.3 Spherical blast waves	774
32.4 Instability Tests	775
32.4.1 Kelvin-Helmholtz instability	775
32.4.2 Rayleigh-Taylor instability	775
32.5 Turbulence Tests	776
32.5.1 Orszag-Tang vortex	776
32.6 Reconnection Tests	777
32.6.1 GEM challenge	777
32.6.2 Current sheet	779
32.7 Divergence-free Field test	779
32.7.1 Magnetic field loop	779
33 Inductance	780
33.1 Unit Definition	780
33.2 Phenomena	781
33.2.1 Direct current	781
33.2.2 Alternating current	781
33.3 Inductors	782
33.3.1 Inductance of Wires and Coils	782
33.3.2 Applications	783
34 Capacitance	784
34.1 What Is a Capacitor?	784
Appendix	785
References	786

Welcome

This website is generated by [Quarto](#). The content is also available as [PDF](#).

1 About

These are my notes about electromagnetism and plasma physics from multiple courses and textbooks, built via [Quarto](#) and is made possible by the Julia programming language (Bezanson et al. 2017) and [pandoc](#).

Thanks to Prof. F. F. Chen, Y. Y. Lau, Yi Li, Yuming Wang, Richard Fitzpatrick, Paul Bellan, and Gábor Tóth through my journey of learning about plasmas. The backbone of the notes is (F. F. Chen 2016). I also learned a lot from Richard Fitzpatrick's online [Plasma Physics Course Notes](#). Y. Y. Lau's graduate course notes also have many references to (Bellan 2008). (Kilpua and Koskinen 2017) is a fairly short and concise introduction that contains many cute practical notes.

I wish to have separate notes on numerical simulations, but right now these are intertwined with the physics here. It is critical to realize that numerical simulations per se do not allow to make fundamental discoveries which go beyond the amount of information that is already contained in the equations one is going to solve on the computer. Therefore, in the first stage, more emphasis should be put on the governing equations as they are the keys to understand the behaviors of plasmas.

The future is about nuclear fusion: what do you think?

2 Introduction

- What is plasma?

Quasi-neutral ionized gases. The fourth state of matter: solid, liquid, gas, **plasma**.

- Where is plasma?
 - Astrophysics & geophysics
 - * Dynamo
 - * Shock
 - * Reconnection
 - * Particle acceleration
 - Engineering & application
 - * Controlled fusion
 - * Power conversion
 - * Reentry of intercontinental ballistic missiles and spacecrafts
 - * Plasma jet as a new propulsion for space vehicles
 - * Global warming
- Methodologies applied in plasma physics
 - Kinetic theory
 - * Microscopic point of view
 - * Positive, negative & neutral particles
 - * EM force, collisions
 - * Ordinary gas .vs. Plasma
 - binary collision
 - free between collisions

$$\begin{aligned} & nf(\mathbf{x} + d\mathbf{x}, \mathbf{p} + d\mathbf{p}, t + dt) d\mathbf{x} d\mathbf{p} - nf(\mathbf{x}, \mathbf{p}, t) d\mathbf{x} d\mathbf{p} \\ &= \frac{\partial(nf(\mathbf{x}, \mathbf{p}, t))}{\partial t} \Big|_{\text{collision}} d\mathbf{x} d\mathbf{p} dt \\ &= \frac{\partial(nf)}{\partial t} + \frac{\partial nf}{\partial \mathbf{x}} \cdot \frac{\mathbf{p}}{m} + \frac{\partial f}{\partial \mathbf{p}} \cdot \mathbf{F} \end{aligned}$$

This is also called the *Fokker-Planck* term.

- Theory of continuum
 - Macroscopic point of view, bulk motion
 - Fundamental equations from conservation laws
 - Electro-Magneto-Hydro-Dynamics (EMHD)

Definition: Plasma = *ionized gas*

In plasma physics we usually use energy units “eV” instead of temperature. Electron volt is a small unit, mostly used for particles, since 1 Joule in our daily life scale is too large for electrons. Given Boltzmann constant $k_B = 1.38 \times 10^{-23} J/K$,

$$1\text{eV} = (11600\text{K}) \times k_B = 1.602 \times 10^{-19} \text{J} ([V] = [J]/[C])$$

There are generally two kinds of plasma:

1. High temperature(> 1keV): fusion plasma
2. Low temperature(~ 1eV – 100eV): plasma processing, space propulsion

At the beginning, it is important to get the idea of size. Starting from two basic length scales:

- $a_0 \equiv$ atomic scale ($\sim 10^{-10}\text{m}$)
- $n =$ number density ($\sim 1/\text{cm}^3$ to $10^{20}/\text{cm}^3$)

Physically, n represents the number of particles in a unit volume, from which we know $n^{-1/3}$ has unit of length. It means a typical length for a single particle. Ionized state implies that $k_B T \gg E_i$ = ionization potential energy $\sim \frac{e^2}{4\pi\epsilon_0 a_0} \approx 14\text{eV}$.

Gaseous state implies that $n^{-1/3} \gg a_0$.

$$\begin{aligned} n^{-1/3} \gg a_0 &\sim \frac{e^2}{4\pi\epsilon_0 E_i} \gg \frac{e^2}{4\pi\epsilon_0 (k_B T)} \\ 4\pi n^{-1/3} n \left(\frac{\epsilon_0 k_B T}{e^2 n} \right) &\gg 1 \end{aligned}$$

$n^{2/3}$ has unit of length $^{-2}$, indicating that the other term should be in the unit of length 2 . This is derived using dimensional analysis, which gives

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T}{e^2 n}} \equiv \text{Debye length} = 740(\text{cm}) \sqrt{\frac{T(\text{eV})}{n(\text{cm}^{-3})}}$$

$$\begin{aligned} (4\pi)^{3/2} n \lambda_D^3 &\gg 1 \\ n \lambda_D^3 &\gg 1 \Rightarrow \text{Large number of particles in a sphere} \end{aligned}$$

This is necessarily satisfied for plasma under previous definitions.

Indication from the definition of plasma:

1. There are large number of particles in a Debye sphere.

Ex.1 laboratory plasma $10^8 \text{ cm}^{-3} \rightarrow 10^{20} \text{ cm}^{-3}$. Tokamak $n = 10^{14} \text{ cm}^{-3}$, $k_B T = 10 \text{ keV} = 10^4 \text{ eV}$, $n^{-1/3} = 2.15 \times 10^{-5} \text{ cm}$, $\lambda_D = 7.4 \times 10^{-3} \text{ cm} \gg n^{-1/3}$, $n\lambda_D^3 \gg 1$.

$L \gg \lambda_D \gg n^{-1/3}$, where L is the scale of the physical system ($\sim 1 \text{ mm}$ to 10 m).

Put it in another way, a criterion for an ionized gas to be a plasma is that it be dense enough that λ_D is much smaller than L .

Ex.2 Interstellar plasma $n = 1 \text{ cm}^{-3}$, $T = 1 \text{ eV}$, $\lambda_D = 740 \text{ cm} \sqrt{\frac{1}{1}} = 740 \text{ cm} \gg n^{-1/3}$, $n\lambda_D^3 \gg 1$.

2. Particle kinetic energy \gg average Coulombic interaction energy.

$$KE \sim k_B T$$

$$PE \sim \frac{e^2}{4\pi\epsilon_0(n^{-1/3})}$$

From the definition we can easily know that

$$k_B T \gg \frac{e^2}{4\pi\epsilon_0 n^{-1/3}} \Rightarrow KE \gg PE$$

This implies that particles are hard to be deflected by their neighbours.

3. Particles are scattered mostly by accumulative, multiple small angle collisions rather than by large angle collisions. This naturally follows the above energy discussion, and will be explained in further detail later.
4. Collective interactions (the basics of all instabilities) dominate over Coulombic interactions between individual particles. This has an analogy to economy: one has nearly no effect on the whole, but the whole has huge effect on individuals. EM fields stand as a bridge between particles.

2.1 Occurrence of Plasma in Nature

It is now believed that the universe is made of 69% dark energy, 27% dark matter, and 1% normal matter. All that we can see in the sky is the part of normal matter that is in the plasma state, emitting radiation. Plasma in physics, not to be confused with blood plasma, is an “ionized” gas in which at least one of the electrons in an atom has been stripped free, leaving a positively charged nucleus, called an ion. Sometimes plasma is called the “fourth state of matter”. When a solid is heated, it becomes a liquid. Heating a liquid turns it into a gas. Upon further heating, the gas is ionized into a plasma. Since a plasma is made of ions

and electrons, which are charged, electric fields are rampant everywhere, and particles “collide” not just when they bump into one another, but even at a distance where they can feel their electric fields. Hydrodynamics, which describes the flow of water through pipes, say, or the flow around boats in yacht races, or the behavior of airplane wings, is already a complicated subject. Adding the electric fields of a plasma greatly expands the range of possible motions, especially in the presence of magnetic fields.

Plasma usually exists only in a vacuum. Otherwise, air will cool the plasma so that the ions and electrons will recombine into normal neutral atoms. In the laboratory, we need to pump the air out of a vacuum chamber. In the vacuum of space, however, much of the gas is in the plasma state, and we can see it. Stellar interiors and atmospheres, gaseous nebulas, and entire galaxies can be seen because they are in the plasma state. On earth, however, our atmosphere limits our experience with plasmas to a few examples: the flash of a lightning bolt, the soft glow of the Aurora Borealis, the light of a fluorescent tube, or the pixels of a plasma TV. We live in a small part of the universe where plasmas do not occur naturally; otherwise, we would not be alive.

The reason for this can be seen from the Saha equation, which tells us the amount of ionization to be expected in a gas in thermal equilibrium:

$$\frac{n_i}{n_n} \approx 2.4 \times 10^{21} \frac{T^{3/2}}{n_i} \exp\left(-\frac{U_i}{k_B T}\right) \quad (2.1)$$

Here n_i and n_n are, respectively, the density (number per m^3) of ionized atoms and of neutral atoms, T is the gas temperature in Kelvin, k_B is Boltzmann’s constant, and U_i is the ionization energy of the gas — that is, the number of joules required to remove the outermost electron from an atom. For ordinary air at room temperature, we may take $n_n \approx 3 \times 10^{25} \text{ m}^{-3}$, $T \approx 300 \text{ K}$, and $U_i = 14.5 \text{ eV}$ (for nitrogen), where $1 \text{ eV} = 1.6 \times 10^{19} \text{ J}$. The fractional ionization $n_i/(n_n + n_i) \approx n_i/n_n$ is ridiculously low:

$$\frac{n_i}{n_n} \approx 10^{-122}$$

As the temperature is raised, the degree of ionization remains low until U_i is only a few times $k_B T$. Then n_i/n_e rises abruptly, and the gas is in a plasma state. Further increase in temperature makes n_n less than n_i , and the plasma eventually becomes fully ionized. This is the reason plasmas exist in astronomical bodies with temperatures of millions of degrees, but not on the earth. Life could not easily coexist with a plasma — at least, plasma of the type we are talking about. The natural occurrence of plasmas at high temperatures is the reason for the designation “the fourth state of matter”.

Although we do not intend to emphasize the Saha equation, we should point out its physical meaning. Atoms in a gas have a spread of thermal energies, and an atom is ionized when, by chance, it suffers a collision of high enough energy to knock out an electron. In a cold gas, such energetic collisions occur infrequently, since an atom must be accelerated to much higher than

the average energy by a series of “favorable” collisions. The exponential factor in Equation 2.1 expresses the fact that the number of fast atoms falls exponentially with $U_i/k_B T$. Once an atom is ionized, it remains charged until it meets an electron; it then very likely recombines with the electron to become neutral again. The recombination rate clearly depends on the density of electrons, which we can take as equal to n_i . The equilibrium ion fraction, therefore, should decrease with n_i ; and this is the reason for the factor n_i^{-1} on the right-hand side of Equation 2.1. The plasma in the interstellar medium owes its existence to the low value of n_i (about 1 per cm³), and hence the low recombination rate.

2.2 Definition of Plasma

Any ionized gas cannot be called a plasma, of course; there is always some small degree of ionization in any gas. A useful definition is as follows:

A plasma is a quasineutral gas of charged and neutral particles which exhibits collective behavior.

We must now define “quasineutral” and “collective behavior”. The meaning of quasineutrality will be made clear in Sect. 1.4 LINK?. What is meant by “collective behavior” is as follows.

Consider the forces acting on a molecule of, say, ordinary air. Since the molecule is neutral, there is no net electromagnetic force on it, and the force of gravity is negligible. The molecule moves undisturbed until it makes a collision with another molecule, and these collisions control the particle’s motion. A macroscopic force applied to a neutral gas, such as from a loudspeaker generating sound waves, is transmitted to the individual atoms by collisions. The situation is totally different in a plasma, which has charged particles. As these charges move around, they can generate local concentrations of positive or negative charge, which give rise to electric fields. Motion of charges also generates currents, and hence magnetic fields. These fields affect the motion of other charged particles far away.

Let us consider the effect on each other of two slightly charged regions of plasma separated by a distance r . ADD FIGURE! The Coulomb force between A and B diminishes as $1/r^2$. However, for a given solid angle (that is, $\Delta r/r = \text{constant}$), the volume of plasma in B that can affect A increases as r^3 . Therefore, elements of plasma exert a force on one another even at large distances. It is this long-ranged Coulomb force that gives the plasma a large repertoire of possible motions and enriches the field of study known as plasma physics. In fact, the most interesting results concern so-called “collisionless” plasmas, in which the long-range electromagnetic forces are so much larger than the forces due to ordinary local collisions that the latter can be neglected altogether. By “collective behavior” we mean motions that depend not only on local conditions but on the state of the plasma in remote regions as well.

The word “plasma” seems to be a misnomer. It comes from the Greek πλαστός, πλάστης, which means something molded or fabricated. Because of collective behavior, a plasma does not tend to conform to external influences; rather, it often behaves as if it had a mind of its own.

2.3 Concept of Temperature

Before proceeding further, it is important to review and extend our physical notions of “temperature”. A gas in thermal equilibrium has particles of all velocities, and the most probable distribution of these velocities is known as the Maxwellian distribution. For simplicity, consider a gas in which the particles can move only in one dimension. (This is not entirely frivolous; a strong magnetic field, for instance, can constrain electrons to move only along the field lines.) The one-dimensional Maxwellian distribution is given by

$$f(v) = A \exp\left(-\frac{1}{2}mv^2/k_B T\right) \quad (2.2)$$

where fdu is the number of particles per m^3 with velocity between u and $u + du$, $\frac{1}{2}mu^2$ is the kinetic energy, and k_B is Boltzmann’s constant. The density n , or number of particles per m^3 , is given by

$$n = \int_{-\infty}^{\infty} f(v) dv$$

The constant A is related to the density n by

$$A = n \left(\frac{m}{2\pi k_B T} \right)^{1/2}$$

The width of the distribution is characterized by the constant T , which we call the temperature. To see the exact meaning of T , we can compute the average kinetic energy of particles in this distribution:

$$E_{av} = \frac{\int_{-\infty}^{\infty} \frac{1}{2}mv^2 f(v) dv}{\int_{-\infty}^{\infty} f(v) dv} \quad (2.3)$$

Defining $y = v/v_t$ and

$$v_t = (2k_B T/m)^{1/2} \quad (2.4)$$

we can write Equation 2.2 as

$$f(v) = A \exp(-v^2/v_t^2)$$

and Equation 2.3 as

$$E_{av} = \frac{\frac{1}{2}mAv_t^3 \int_{-\infty}^{\infty} [\exp(-y^2)]y^2 dy}{Av_t \int_{-\infty}^{\infty} [\exp(-y^2)] dy}$$

The integral in the numerator is integrable by parts:

$$\begin{aligned} \int_{-\infty}^{\infty} y \cdot [\exp(-y^2)]y dy &= \left[-\frac{1}{2}[\exp(-y^2)]y \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -\frac{1}{2}[\exp(-y^2)]dy \\ &= \frac{1}{2} \int_{-\infty}^{\infty} [\exp(-y^2)] dy \end{aligned}$$

Canceling the integrals, we have

$$E_{av} = \frac{\frac{1}{2}mAv_t^3 \frac{1}{2}}{Av_t} = \frac{1}{4}mv_t^2 = \frac{1}{2}k_B T$$

Thus the average kinetic energy is $\frac{1}{2}k_B T$.

It is easy to extend this result to three dimensions. Maxwell's distribution is then

$$f(v_x, v_y, v_z) = A_3 \exp \left[-\frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)/k_B T \right]$$

where

$$A_3 = n \left(\frac{m}{2\pi k_B T} \right)^{3/2}$$

Because a Maxwellian distribution is isotropic (i.e. the form is symmetric in v_x , v_y , and v_z), we can separate each dimension. The average kinetic energy is then 3 times the single dimension result

$$E_{av} = \frac{3}{2}k_B T$$

The general result is that E_{av} equals $\frac{1}{2}k_B T$ per degree of freedom.

Since T and E_{av} are so closely related, it is customary in plasma physics to give temperatures in units of energy. To avoid confusion on the number of dimensions involved, it is not E_{av} but the energy corresponding to $k_B T$ that is used to denote the temperature. For $k_B T = 1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$, we have

$$T = \frac{q}{k_B} = 11600$$

Thus the conversion factor is

$$1 \text{ eV} = 11600 \text{ K}$$

By a 2-eV plasma we mean that $k_B T = 2 \text{ eV}$, or $E_{av} = 3 \text{ eV}$ in three dimensions.

It is interesting that a plasma can have several temperatures at the same time. It often happens that the ions and the electrons have separate Maxwellian distributions with different temperatures T_i and T_e . This can come about because the collision rate among ions or among electrons themselves is larger than the rate of collisions between an ion and an electron. Then each species can be in its own thermal equilibrium, but the plasma may not last long enough for the two temperatures to equalize. When there is a magnetic field \mathbf{B} , even a single species, say ions, can have two temperatures. This is because the forces acting on an ion along \mathbf{B} are different from those acting perpendicular to \mathbf{B} (due to the Lorentz force). The components of velocity perpendicular to \mathbf{B} and parallel to \mathbf{B} may then belong to different Maxwellian distributions with temperatures T_\perp and T_\parallel .

Before leaving our review of the notion of temperature, we should dispel the popular misconception that high temperature necessarily means a lot of heat. People are usually amazed to learn that the electron temperature inside a fluorescent light bulb is about 20000 K. “My, it doesn’t feel that hot!” Of course, the heat capacity must also be taken into account. The density of electrons inside a fluorescent tube is much less than that of a gas at atmospheric pressure, and the total amount of heat transferred to the wall by electrons striking it at their thermal velocities is not that great. For example, the temperature of a cigarette ash is high enough to cause a burn, but the total amount of heat involved is not. Many laboratory plasmas have temperatures of the order of 1,000,000 K (100 eV), but at densities of only $10^{18} - 10^{19}$ per m³, the heating of the walls is not a serious consideration.

2.4 Debye Shielding

(Problem 4 on P27 (Bellan 2008))

A fundamental characteristic of the behavior of plasma is its ability to shield out electric potentials that are applied to it. Suppose we tried to put an electric field inside a plasma by inserting two charged balls connected to a battery (Fig. 1.3 ADD IT!). The balls would attract particles of the opposite charge, and almost immediately a cloud of ions would surround the negative ball and a cloud of electrons would surround the positive ball. (We assume that a layer of dielectric keeps the plasma from actually recombining on the surface, or that the battery is large enough to maintain the potential in spite of this.) If the plasma were cold and there were no thermal motions, there would be just as many charges in the cloud as in the ball, the shielding would be perfect, and no electric field would be present in the body of the plasma outside of the clouds. On the other hand, if the temperature is finite, those particles that are at the edge of the cloud, where the electric field is weak, have enough thermal energy to escape from the electrostatic potential well. The “edge” of the cloud then occurs at the radius where the potential energy is approximately equal to the thermal energy $k_B T$ of the particles, and the shielding is not complete. Potentials of the order of $k_B T/e$ can leak into the plasma and cause finite electric fields to exist there.

Let us compute the approximate thickness of such a charge cloud. Let us put a point charge q_T into a quasi-neutral gas of charged particles. The gas is in *local thermodynamic equilibrium* (LTE) with ion and electron temperatures T_e and T_i (They are not necessarily the same, and in fact, quite different in nature.). (ADD FIGURE!) We wish to compute $\phi(x)$. For simplicity, we assume that the ion-electron mass ratio M/m is infinite, so that the ions do not move but form a uniform background of positive charge. To be more precise, we can say that M/m is large enough that the inertia of the ions prevents them from moving significantly on the time scale of the experiment.

Maxwell distribution:

$$f(\mathbf{v}) = n_0 \left(\frac{m}{2\pi kT} \right)^{3/2} \exp \left(- \frac{mv^2}{2kT} \right)$$

In the presence of a potential energy $q\phi$ the distribution function follows the Boltzmann distribution:

$$f(\mathbf{x}, \mathbf{v}) = n_0 \left(\frac{m}{2\pi kT} \right)^{3/2} \exp \left(-\frac{\frac{1}{2}mv^2 + q\phi(\mathbf{x})}{kT} \right)$$

It would not be worthwhile to prove this here. What this equation says is intuitively obvious: There are fewer particles at places where the potential energy is large, since not all particles have enough energy to get there.

Note that the concept of temperature is valid in LTE. From the Boltzmann equation,

$$\begin{aligned} \frac{df}{dt} &= \frac{\partial f}{\partial t} + \nabla_x \cdot (f\mathbf{v}) + \nabla_v \cdot (f\mathbf{a}) = 0 \\ &\Rightarrow \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_x f + \frac{\mathbf{F}}{m} \cdot \nabla_v f = 0 \end{aligned}$$

If we neglect the time derivatives, we obtain

$$(\mathbf{v}_j \cdot \nabla) f(\mathbf{r}, \mathbf{v}_j) + \left[-\frac{1}{m_j} \nabla(q_j \phi) \cdot \nabla_v \right] f(\mathbf{r}, \mathbf{v}_j) = 0 \quad (2.5)$$

The LTE distribution is in a Maxwellian shape,

$$f(\mathbf{r}, \mathbf{v}_j) = n_j(\mathbf{r}) \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{mv^2}{2k_B T} \right)$$

Substitute into Equation 2.5, we have

$$\begin{aligned} (\mathbf{v}_j \cdot \nabla) n_j + \frac{\mathbf{v}_j \cdot \nabla(q_j \phi)}{k_B T_j} n_j &= 0 \\ \Rightarrow n_j(\mathbf{r}) &= n_{j0} \exp \left[-\frac{q_j \phi_j(\mathbf{r})}{k_B T_j} \right] \end{aligned}$$

Then we get the Boltzmann distribution of each species with a separate temperature for each.

Consider a uniform plasma with $n_i = n_{i0} e^{-e\phi/k_B T_i}$, $n_e = n_{e0} e^{e\phi/k_B T_e}$, and with the quasi-neutral assumption, $n_{e0} = n_{i0} = n_0$. Introduce a point charge q_T in the initially neutralized plasma. Poisson equation gives

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0} = -\frac{1}{\epsilon_0} \left[e(n_e - n_i) + q_T \delta(\mathbf{r}) \right]$$

Due to symmetry, $\phi = \phi(r)$. When the temperature is high, $q\phi(r) \ll kT$, the exponent is small and we can expand in a Taylor series and leave only the first order term. In spherical coordinates,

$$\begin{aligned} \nabla^2 \phi &= \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) \\ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) + \frac{e}{\epsilon_0} \left(\underbrace{n_{e0} e^{e\phi/k_B T_e} - n_{i0} e^{-e\phi/k_B T_i}}_{\approx n_0 [e\phi/k_B T_e + e\phi/k_B T_i]} \right) &= -\frac{q_T}{\epsilon_0} \delta(\mathbf{r}) \\ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) - \underbrace{\frac{e^2 n_0}{\epsilon_0} \left(\frac{1}{k_B T_e} + \frac{1}{k_B T_i} \right)}_{\frac{1}{\lambda_D^2} \equiv \frac{1}{\lambda_{D_i}^2} + \frac{1}{\lambda_{D_e}^2}} \phi &= -\frac{q_T}{\epsilon_0} \delta(\mathbf{r}) \\ \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) - \frac{1}{\lambda_D^2} \phi &= -\frac{q_T}{\epsilon_0} \delta(\mathbf{r}) \end{aligned}$$

where

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T^*}{n_0 e^2}} \quad (2.6)$$

is the Debye length, $T^* = \left(\frac{1}{T_e} + \frac{1}{T_i} \right)^{-1}$. The Debye length, is a measure of the shielding distance or thickness of the sheath. Note that larger density indicates better shielding effects, as one would expect, since each layer of plasma contains more electrons. Furthermore, λ_D increases with increasing $k_B T^*$. Without thermal agitation, the charge cloud would collapse to an infinitely thin layer. Finally, most commonly it is the *electron* temperature ($T^* \approx T_e$) which is used in the definition of λ_D because the electrons, being more mobile than the ions, generally do the shielding by moving so as to create a surplus or deficit of negative charge. Only in special situations is this not true.

For $r > 0$ (away from the charge),

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \phi}{\partial r} \right) - \frac{1}{\lambda_D^2} \phi = 0$$

Let $\phi(r) = \frac{F(r)}{r}$, we have

$$\begin{aligned} \frac{d^2 F}{dr^2} - \frac{F}{\lambda_D^2} &= 0 \\ \Rightarrow F &= A e^{-r/\lambda_D} + B e^{r/\lambda_D} \quad (r > 0) \\ &= A e^{-r/\lambda_D} \end{aligned}$$

Anticipate when $r \rightarrow 0$, $\phi(r) \rightarrow \frac{q_T}{4\pi\epsilon_0 r} \approx \frac{A}{r}$. So

$$\phi(r) = \frac{q_t}{4\pi\epsilon_0 r} e^{-r/\lambda_D}$$

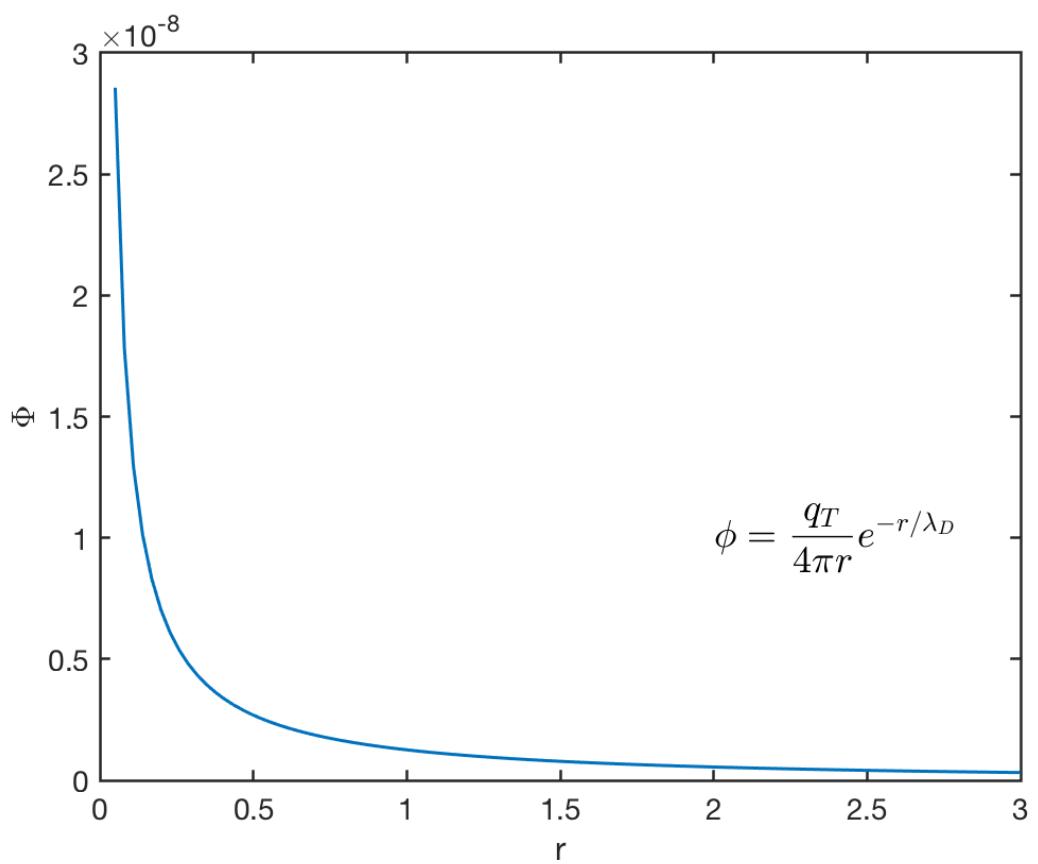


Figure 2.1: Electric potential.

The electric potential drops very quickly (Figure 2.1). In other words, if you insert a charge into a neutral plasma, its potential will only effect within a small range. Recall the Gauss's law,

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{Q_T}{\epsilon_0} \rightarrow \text{as } R \rightarrow \infty, \text{ E goes to zero.}$$

The following are useful forms of Equation 2.6:

$$\begin{aligned}\lambda_D &= 69(T_e/n)^{1/2} \text{ m, } T_e \text{ in K} \\ \lambda_D &= 7430(k_B T_e/n)^{1/2} \text{ m, } k_B T_e \text{ in eV}\end{aligned}$$

We are now in a position to define “quasineutrality.” If the dimensions L of a system are much larger than λ_D , then whenever local concentrations of charge arise or external potentials are introduced into the system, these are shielded out in a distance short compared with L , leaving the bulk of the plasma free of large electric potentials or fields. Outside of the sheath on the wall or on an obstacle, $\nabla^2\phi$ is very small, and n_i is equal to n_e . It takes only a small charge imbalance to give rise to potentials of the order of $k_B T/e$. The plasma is “quasineutral”; that is, neutral enough so that one can take $n_i \simeq n_e \simeq n$, where n is a common density called the *plasma density*, but not so neutral that all the interesting electromagnetic forces vanish. Plasma may be considered neutral over a length scale larger than Debye length.

Additional comments:

1. Plasma acts like dielectron.
2. There is a plasma sheath near to where materials contact, inside which charge neutrality is violated. This will be discussed in Section 17.1.

2.5 Criteria for Plasmas

We have given two conditions that an ionized gas must satisfy to be called a plasma. A third condition has to do with collisions. The weakly ionized gas in an airplane’s jet exhaust, for example, does not qualify as a plasma because the charged particles collide so frequently with neutral atoms that their motion is controlled by ordinary hydrodynamic forces rather than by electromagnetic forces. If ω is the frequency of typical plasma oscillations and τ is the mean time between collisions with neutral atoms, we require $\omega\tau > 1$ for the gas to behave like a plasma rather than a neutral gas.

1. $\lambda_D \ll L$
2. $N_D \gg 1$
3. $\omega\tau > 1$

Often in lectures people think the condition 1 and 2 are equivalent; they are NOT exactly. The argument would be that condition 1 is valid when the distribution is Maxwellian, which happens to be the equilibrium distribution for large number of particles; condition 2 is the statistical priori for that to be valid.

2.6 Plasma Frequency

Besides the length scale, we are also interested in time scale. The plasma frequency,

$$\omega_p = \sqrt{\frac{n e^2}{\epsilon_0 m}} \quad (2.7)$$

is the most fundamental time-scale in plasma physics. Clearly, there is a different plasma frequency for each species. However, the relatively fast electron frequency is, by far, the most important, and references to “the plasma frequency” in most textbooks invariably mean the electron plasma frequency.

ω_p corresponds to the typical electrostatic oscillation frequency of a given species in response to a small charge separation. For instance, consider a one-dimensional situation in which a slab consisting entirely of one charge species is displaced from its quasi-neutral position by an infinitesimal distance δx . The resulting charge density which develops on the leading face of the slab is $\sigma = e n \delta x$. An equal and opposite charge density develops on the opposite face. The x -directed electric field generated inside the slab is of magnitude $E_x = -\sigma/\epsilon_0 = -e n \delta x/\epsilon_0$. Thus, Newton’s law applied to an individual particle inside the slab yields

$$m \frac{d^2 \delta x}{dt^2} = e E_x = -m \omega_p^2 \delta x$$

giving $\delta x = \delta x_0 \cos \omega_p t$.

Plasma frequency is closely associated with plasma oscillation and waves. A more thorough derivation is given in Section 10.3. Note that plasma oscillations will only be observed if the plasma system is studied over time periods τ longer than the plasma period $\tau_p \equiv 1/\omega_p$, and if external actions change the system at a rate no faster than ω_p . In the opposite case, one is clearly studying something other than plasma physics (e.g., nuclear reactions), and the system cannot not usefully be considered to be a plasma. Likewise, observations over length-scales L shorter than the distance $v_t \tau_p$ traveled by a typical plasma particle during a plasma period will also not detect plasma behaviour. In this case, particles will exit the system before completing a plasma oscillation. This distance, which is the spatial equivalent to τ_p , is called the *Debye length*, and takes the form

$$\lambda_D \equiv \sqrt{T/m} \omega_p^{-1}.$$

Note that

$$\lambda_D = \sqrt{\frac{\epsilon_0 T}{n e^2}}$$

is independent of mass, and therefore generally comparable for different species. Clearly, our idealized system can only usefully be considered to be a plasma provided that

$$\frac{\lambda_D}{L} \ll 1 \quad (2.8)$$

and

$$\frac{\tau_p}{\tau} \ll 1$$

Here, τ and L represent the typical time-scale and length-scale of the process under investigation. It should be noted that, despite the conventional requirement Equation 2.8, plasma physics is capable of considering structures on the Debye scale. The most important example of this is the Debye sheath: i.e., the boundary layer which surrounds a plasma confined by a material surface.

Let $V_A = B/\sqrt{\mu_0\rho}$ be the Alfvén speed and $\omega_c = qB/m$ be the ion gyro-frequency. We have this convenient relation

$$\frac{\omega_p}{\omega_c} = \frac{c}{V_A} \quad (2.9)$$

2.7 Spatial Scales

- *Ion inertial length*: the scale at which ions decouple from electrons and the magnetic field becomes frozen into the electron fluid rather than the bulk plasma.

$$d_i = \frac{c}{\omega_{pi}} = \frac{V_{Ai}}{\omega_{ci}} \quad (2.10)$$

where V_{Ai} is the ion Alfvén speed.

- *Electron inertial length (plasma skin depth)* : the scale over which the electron distribution function can change substantially, or the depth in a plasma to which electromagnetic radiation can penetrate.

$$d_e = \frac{c}{\omega_{pe}} = \frac{V_{Ae}}{\omega_{ce}} \quad (2.11)$$

where V_{Ae} is the electron Alfvén speed.

- *Ion gyroradius*: also known as Larmor radius, the scale of ion gyration around the magnetic field

$$r_L = \frac{v_{\perp,ti}}{\omega_c} = \frac{mv_{\perp,ti}}{qB} \quad (2.12)$$

where $v_{\perp,ti} = \sqrt{p_{\perp,i}/\rho}$ is the perpendicular ion thermal speed.

- *Electron gyroradius*: the scale of electron gyration around the magnetic field

$$r_L = \frac{v_{\perp,te}}{\omega_c} = \frac{mv_{\perp,te}}{qB} \quad (2.13)$$

where $v_{\perp,te} = \sqrt{p_{\perp e}/\rho}$ is the perpendicular ion thermal speed.

One convenient relation between the inertial length and the gyroradius is

$$\frac{r_L}{d_i} = \frac{v_{ti}}{v_A} = \sqrt{\beta/2} \quad (2.14)$$

In the relativistic case, the gyroradius is given as

$$r_L = \frac{p}{qB} = \frac{\gamma mv_{\perp}}{qB} \quad (2.15)$$

where γ is the Lorentz factor. For astroparticle studies, people love to use [GeV] for particle energy and normalize the velocity and charge to c and e, respectively. Therefore,

$$\begin{aligned} \frac{r_L}{[m]} &= \frac{\gamma mc^2/[J] \cdot v_{\perp}/c}{q/e \cdot B/[T]} \cdot \frac{1}{ce} \\ &= \frac{\gamma mc^2/[GeV] \cdot v_{\perp}/c}{q/e \cdot B/[T]} \cdot \frac{10^9}{c} \\ &= 3.3 \frac{\gamma mc^2/[GeV] \cdot v_{\perp}/c}{q/e \cdot B/[T]} \end{aligned} \quad (2.16)$$

- *Debye length*: the scale over which electric fields are screened out by a redistribution of the electrons:

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T^*}{e^2 n_e}} = \frac{v_{th}^*}{\omega_{pe}} \quad (2.17)$$

where $T^* = \left(\frac{1}{T_e} + \frac{1}{T_i}\right)^{-1}$, and v_{th}^* is the corresponding thermal speed. Even in quasineutral cold plasma, where ion contribution virtually seems to be larger due to lower ion temperature, the ion term is actually often dropped, giving

$$\lambda_D = \sqrt{\frac{\epsilon_0 k_B T_e}{e^2 n}} = \frac{m_e}{q_e} \frac{\sqrt{\epsilon_0 P_e}}{\rho_e} = \sqrt{\frac{\epsilon_0 v_{te}}{e^2 n_e}} = \frac{v_{te}}{\omega_{pe}} \quad (2.18)$$

where $v_{te} = \sqrt{P_e/\rho_e} = \sqrt{k_B T_e/m_e}$, which is only valid when the mobility of ions is negligible compared to the process's timescale.

Comparing Equation 2.11 and Equation 2.18, we come up with a useful relation between the electron inertial length and the Debye length

$$\frac{\lambda_D}{d_e} = \frac{v_{te}}{c} \quad (2.19)$$

2.8 How to Study Plasma Physics

There is no universal theory for plasma physics due to the multi-scale problems. All useful equations are derived under certain scaling approximations. The key is to understand when it is proper to apply a theory for a given problem.

2.9 Common Acronyms

- **HD:** hydrodynamics, hydrodynamic
- **MHD:** magnetohydrodynamics, magnetohydrodynamic
- **EM:** electromagnetic, electromagnetism
- **PIC:** particle-in-cell
- **CR:** cosmic ray
- **GC:** guiding center
- **GK:** gyrokinetics, gyrokinetic
- **FLR:** finite-Larmor-radius, field line resonance
- **RD:** rotational discontinuity
- **TD:** tangential discontinuity
- **CD:** contact discontinuity
- **RH:** Rankine-Hugoniot
- **dHT:** de Hoffmann-Teller

3 Math

3.1 Vector Identities

Some useful vector identities are listed below:

$$\begin{aligned}\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) + \mathbf{C}(\mathbf{A} \cdot \mathbf{B}) \\ (\mathbf{A} \times \mathbf{B}) \times \mathbf{C} &= \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) + \mathbf{A}(\mathbf{B} \cdot \mathbf{C}) \\ \nabla \times \nabla f &= 0 \\ \nabla \cdot (\nabla \times \mathbf{A}) &= 0 \\ \nabla \cdot (f\mathbf{A}) &= (\nabla f) \cdot \mathbf{A} + f\nabla \cdot \mathbf{A} \\ \nabla \times (f\mathbf{A}) &= (\nabla f) \times \mathbf{A} + f\nabla \times \mathbf{A} \\ \nabla \cdot (\mathbf{A} \times \mathbf{B}) &= \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \\ \nabla(\mathbf{A} \cdot \mathbf{B}) &= (\mathbf{B} \cdot \nabla)\mathbf{A} + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + \mathbf{A} \times (\nabla \times \mathbf{B}) \\ \nabla \cdot (\mathbf{AB}) &= (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B}(\nabla \cdot \mathbf{A}) \\ \nabla \times (\mathbf{A} \times \mathbf{B}) &= (\mathbf{B} \cdot \nabla)\mathbf{A} - (\mathbf{A} \cdot \nabla)\mathbf{B} - \mathbf{B}(\nabla \cdot \mathbf{A}) + \mathbf{A}(\nabla \cdot \mathbf{B}) \\ \nabla \times (\nabla \times \mathbf{A}) &= \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A}\end{aligned}$$

In most cases Einstein notation shall be used to derive all the identities:

$$y = \sum_{i=1}^3 c_i x^i = c_1 x^1 + c_2 x^2 + c_3 x^3$$

is simplified by the convention to

$$y = c_i x^i$$

The upper indices are not exponents but are indices of coordinates, coefficients or basis vectors. That is, in this context x^2 should be understood as the second component of x rather than the square of x (this can occasionally lead to ambiguity). The upper index position in x^i is because, typically, an index occurs once in an upper (superscript) and once in a lower (subscript) position in a term.

A basis gives such a form (via the dual basis), hence when working on \mathbf{R}^n with a Euclidean metric and a fixed orthonormal basis, one has the option to work with only subscripts. However,

if one changes coordinates, the way that coefficients change depends on the variance of the object, and one cannot ignore the distinction. (See this [wiki link](#).)

Commonly used identities:

$$\begin{aligned}\mathbf{A} \cdot \mathbf{B} &= A_i B_i \\ \mathbf{A} \times \mathbf{B} &= \epsilon_{ijk} A_j B_k \\ \nabla \cdot \mathbf{A} &= \partial_i A_i \\ \nabla \times \mathbf{A} &= \epsilon_{ijk} \partial_j A_k \\ \epsilon_{ijk} \epsilon_{imn} &= \delta_{jm} \delta_{kn} - \delta_{jn} \delta_{km}\end{aligned}$$

3.2 Differential of Line Segments

The differential of line segments in a fluid is:

$$\frac{d\mathbf{l}}{dt} = \mathbf{l} \cdot \nabla \mathbf{u} \quad (3.1)$$

Proof. Let the two line segments \mathbf{l}_1 and \mathbf{l}_2 be expressed as starting point vectors and end point vectors:

$$\begin{aligned}\mathbf{l}_1 &= \mathbf{r}_2 - \mathbf{r}_1 \\ \mathbf{l}_2 &= \mathbf{r}'_2 - \mathbf{r}'_1\end{aligned}$$

After time Δt , $\mathbf{l}_1 \rightarrow \mathbf{l}_2$,

$$\mathbf{l}_2 = \mathbf{r}'_2 - \mathbf{r}'_1 = (\mathbf{r}_2 + \mathbf{v}(r_2)\Delta t) - (\mathbf{r}_1 + \mathbf{v}(r_1)\Delta t) = \mathbf{l}_1 + (\mathbf{v}(\mathbf{r}_2) - \mathbf{v}(\mathbf{r}_1))\Delta t$$

The velocity difference can be written as

$$\mathbf{v}(x_2, y_2, z_2) - \mathbf{v}(x_1, y_1, z_1) = \frac{\partial \mathbf{v}}{\partial x}(x_2 - x_1) + \frac{\partial \mathbf{v}}{\partial y}(y_2 - y_1) + \frac{\partial \mathbf{v}}{\partial z}(z_2 - z_1) = (\mathbf{l} \cdot \nabla) \mathbf{v}$$

So

$$\mathbf{l}_2 = \mathbf{l}_1 + \mathbf{l} \cdot \nabla \mathbf{v} \cdot \Delta t \Rightarrow \frac{\mathbf{l}_2 - \mathbf{l}_1}{\Delta t} = \mathbf{l} \cdot \nabla \mathbf{v}$$

3.3 Lagrangian vs Eulerian Descriptions

The Lagrangian operator d/dt is defined to be the *convective derivative*

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \quad (3.2)$$

which characterizes the temporal rate of change seen by an observer moving with the velocity \mathbf{u} . An everyday example of the convective term would be the apparent temporal increase in density of automobiles seen by a motorcyclist who enters a traffic jam of stationary vehicles and is not impeded by the traffic jam.

The convective derivative is sometimes written as D/Dt , which is used to emphasize the fact that this specific convective derivative is defined using the center-of-mass velocity. Note that $(\mathbf{u} \cdot \nabla)$ is a *scalar* differential operator.

3.4 Complex Analysis

In complex analysis, the following statements are equivalent:

1. $f(z)$ is an analytic function of z in some neighbourhood of z_0 .
2. $\oint_C f(z) dz = 0$ for every closed contour C that lies in that neighbourhood of z_0 .
3. $df(z)/dz$ exists at $z = z_0$.
4. $f(z)$ has a convergent Taylor expansion about $z = z_0$.
5. The n^{th} derivative $d^n f(z)/dz^n$ exists at $z = z_0$ for all n .
6. The *Cauchy-Riemann condition* is satisfied at $z = z_0$. Take u and v to be the real and imaginary parts respectively of a complex-valued function of a single complex variable $z = x + iy$, $f(x + iy) = u(x, y) + iv(x, y)$.

$$\begin{aligned}\frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}\end{aligned}$$

An idea of *analytic continuation* is introduced here. In practice, an analytic function is usually defined by means of some mathematical expression — such as a polynomial, an infinite series, or an integral. Ordinarily, there is some region within which the expression is meaningful and does yield an analytic function. Outside this region, the expression may cease to be meaningful, and the question then arises as to whether or not there is any way of extending the definition of the function so that this “extended” function is analytic over a larger region. A simple example is given as follows.

Ex. 1 Polynomial series

$$f(z) = \sum_{n=0}^{\infty} z^n$$

which describes an analytic function for $|z| < 1$ but which diverges for $|z| > 1$. However, the function

$$g(z) = \frac{1}{1-z}$$

is analytic over the whole plane (except at $z = 1$), and it coincides with $f(z)$ inside the unit circle.

Such a function $g(z)$, which coincides with a given analytic $f(z)$ over that region for which $f(z)$ is defined by which also is analytic over some extension of that region, is said to be an *analytic continuation* of $f(z)$. It is useful to think of $f(z)$ and $g(z)$ as being the same function and to consider that the formula defining $f(z)$ failed to provide the values of $f(z)$ in the extended region because of some defect in the mode of description rather than because of some limitation inherent in $f(z)$ itself. [c.f. G.F.Carrier, M.Krook and C.E.Pearson, Functions of a Complex Variable, McGraw-Hill (1966), p.63]

Ex.2 Laplace transform

$$\mathcal{L}[1] = \int_0^{\infty} dt 1 \cdot e^{i\omega t} = -\frac{1}{i\omega}, \text{ if } \Im(\omega) > 0$$

If you have a pure real frequency ω , then when you integrate v over the real axis, at $v = \omega/k$ you will encounter a singular point. Actually, this integration is convergent if and only if $\Im(\omega) > 0$. $-\frac{1}{i\omega}$ is the *analytic continuation* of $f(\omega)$ for all complex ω except $\omega = 0$.

To calculate the integral around singular points, we may take advantage of the Cauchy integral formula and the residual theorem.

Theorem 2.1 Cauchy integral

Let C_ϵ be a circular arc of radius ϵ , centered at α , with subtended angle θ_0 in counterclockwise direction. Let $f(z)$ be an analytic function on $C_\epsilon +$ inside C_ϵ . Then

$$\lim_{\epsilon \rightarrow 0} \int_{C_\epsilon} \frac{f(z) dz}{z - \alpha} = i\theta_0 f(\alpha)$$

Proof: On C_ϵ , $z = \alpha + \epsilon e^{i\theta}$, $dz = i\epsilon e^{i\theta} d\theta$.

$$LHS = \lim_{\epsilon \rightarrow 0} \int_{C_\epsilon} \frac{f(\alpha + \epsilon e^{i\theta}) i\epsilon e^{i\theta} d\theta}{\epsilon e^{i\theta}} = i\theta_0 f(\alpha).$$

Theorem 2.2 Residue

Let $f(z)$ be an analytic function on a closed contour $C + \text{inside } C$. If point α is inside C , we have

$$f(\alpha) = \frac{1}{2\pi i} \oint_C \frac{f(z)dz}{z - \alpha}$$

Proof:

$\frac{f(z)}{z - \alpha}$ is analytic within region bounded by $C + L_1 - C_\epsilon + L_2$, where L_1 and L_2 are two paths that connects/breaks C and C_ϵ . Therefore

$$\begin{aligned} & \int_C + \int_{L_1} - \int_{C_\epsilon} + \int_{L_2} \frac{f(z)}{z - \alpha} dz = 0 \\ & \Rightarrow \oint_C \frac{f(z)}{z - \alpha} dz = \oint_{C_\epsilon} \frac{f(z)}{z - \alpha} dz = 2\pi i f(\alpha) \text{ as } L_1 \rightarrow -L_2, \epsilon \rightarrow 0. \end{aligned}$$

There is also a purely algebraic proof available.

Note that the value of $f(z)$ on C determines value of $f(\alpha)$ for all α within C . This has a close relation to the potential theory. Actually, what Cauchy-Riemann condition says physically is that the potential flow is both irrotational and incompressible!

Theorem 2.3 Residual theorem

Let $f(z)$ be an analytic function on $C + \text{inside } C$. If point α is inside C , we have

$$\oint_C \frac{f(z)dz}{z - \alpha} = 2\pi i f(\alpha) \equiv 2\pi i \text{Res}\left[\frac{f(z)}{z - \alpha}; z = \alpha\right] \quad (3.3)$$

Khan Academy has a [nice video](#) on this. Applying this powerful theorem, we can calculate many integrals analytically which contain singular points.

Ex.3

$$f(\omega) = \int_{-\infty}^{\infty} dv \frac{e^{iv}}{v - \omega} = 2\pi i e^{i\omega}, \Im(\omega) > 0$$

Pick a semi-circle contour C_R in the upper plane of complex v . Let C be a closed contour of a line along the real axis $\Re(v)$ and the semi-circle C_R . e^{iv} is analytic along and inside C , so

$$f(\omega) = \left(\oint_C - \int_{C_R} \right) \frac{dv e^{iv}}{v - \omega} = 2\pi i e^{i\omega} - \int_{C_R} \frac{dv e^{iv}}{v - \omega} = 2\pi i e^{i\omega} \text{ as } R \rightarrow \infty$$

$$\left(e^{iv} = e^{i(v_r + iv_i)} = e^{iv_r} e^{-v_i}, v_i > 0; v - \omega \rightarrow \infty \right)$$

$2\pi i e^{i\omega}$ is the *analytic continuation* of $f(\omega)$ for all ω . Analytic continuation is achieved if we deform the contour integration in the complex v plane.

Ex.4

$$\frac{\epsilon(\omega)}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{k^2} \chi(\omega)$$

where

$$\begin{aligned} \chi(\omega) &= \int_{-\infty}^{\infty} dv \frac{\partial g/\partial v}{v - \omega/k}, \text{ Im}(\omega) > 0 \\ &= \int_L \frac{\partial g(v)/\partial v}{v - \omega/k}, \text{ for all complex } \omega, \text{ as long as } L \text{ lies below } \omega \end{aligned}$$

Landau integral: pick a trajectory *under* the singular point in the complex plane to achieve the integration.

FIGURE NEEDED!

Let C_ϵ be a small semi-circle under ω/k . Then

$$\begin{aligned} \frac{\epsilon(\omega)}{\epsilon_0} &= 1 - \frac{\omega_{pe}^2}{k^2} \left[P \int_{-\infty}^{\infty} dv \frac{\partial g(v)\partial v}{v - \omega/k} + \int_{C_\epsilon} dv \frac{\partial g(v)\partial v}{v - \omega/k} \right] \\ &= 1 - \frac{\omega_{pe}^2}{k^2} \left[P \int_{-\infty}^{\infty} dv \frac{\partial g(v)\partial v}{v - \omega/k} + i\pi \frac{\partial g(v)}{\partial v} \Big|_{v=\frac{\omega}{k}} \right] \end{aligned}$$

where P denotes the principle value integral. This is the same as Equation 12.31 that will be discussed in Section 12.5.

3.5 Electromagnetic Four-Potential

An **electromagnetic four-potential** is a relativistic vector function from which the electromagnetic field can be derived. It combines both an electric scalar potential and a magnetic vector potential into a single four-vector.

As measured in a given frame of reference, and for a given gauge, the first component of the electromagnetic four-potential is conventionally taken to be the electric scalar potential, and the other three components make up the magnetic vector potential. While both the scalar and vector potential depend upon the frame, the electromagnetic four-potential is Lorentz covariant.

The electromagnetic four-potential can be defined as:

Table 3.1: Electromagnetic four-potential

SI units	Gaussian units
$A^\alpha = \left(\frac{\phi}{c}, \mathbf{A} \right)$	$A^\alpha = (\phi, \mathbf{A})$

in which ϕ is the electric potential, and \mathbf{A} is the magnetic potential (a vector potential).

The electric and magnetic fields associated with these four-potentials are:

Table 3.2: Electromagnetic fields from the four-potential

SI units	Gaussian units
$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}$ $\mathbf{B} = \nabla \times \mathbf{A}$	$\mathbf{E} = -\nabla\phi - \frac{1}{c} \frac{\partial\mathbf{A}}{\partial t}$ $\mathbf{B} = \nabla \times \mathbf{A}$

3.5.1 Multipolar gauge

A common problem is how to obtain the four-potential from EM field. This can be achieved in a simple way using the multipolar gauge (also known as the line gauge, point gauge or Poincaré gauge (named after Henri Poincaré))

$$\mathbf{r} \cdot \mathbf{A} = 0 \quad (3.4)$$

which says that the potential vector is perpendicular to the position vector \mathbf{r} . With this we have

$$\begin{aligned} \mathbf{A}(\mathbf{r}, t) &= -\mathbf{r} \times \int_0^1 \mathbf{B}(u\mathbf{r}, t) u \, du \\ \varphi(\mathbf{r}, t) &= -\mathbf{r} \cdot \int_0^1 \mathbf{E}(u\mathbf{r}, t) u \, du \end{aligned} \quad (3.5)$$

3.6 Helmholtz's Theorem

Helmholtz's theorem, also known as the fundamental theorem of vector calculus, states that any sufficiently smooth, rapidly decaying vector field in three dimensions can be resolved into the sum of an irrotational (curl-free) vector field and a solenoidal (divergence-free) vector field.

Let \mathbf{F} be a vector field on a bounded domain $V \subseteq \mathbf{R}^3$, which is twice continuously differentiable, and let S be the surface that encloses the domain V . Then \mathbf{F} can be decomposed into a curl-free component and a divergence-free component:

$$\mathbf{F} = \nabla\Phi + \nabla \times \mathbf{A}$$

where

$$\begin{aligned}\Phi(\mathbf{r}) &= \frac{1}{4\pi} \int_V \frac{\nabla' \cdot \mathbf{F}'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' - \frac{1}{4\pi} \oint_S \hat{\mathbf{n}}' \cdot \frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS' \\ \mathbf{A} &= \frac{1}{4\pi} \int_V \frac{\nabla' \times \mathbf{F}'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' - \frac{1}{4\pi} \oint_S \hat{\mathbf{n}}' \times \frac{\mathbf{F}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS'\end{aligned}$$

and ∇' is the nabla operator with respect to \mathbf{r}' , not \mathbf{r} .

If $V = \mathbf{R}^3$ and is therefore unbounded, and \mathbf{F} vanishes faster than $1/r$ as $r \rightarrow \infty$, then the second component of both scalar and vector potential are zero. That is,

$$\begin{aligned}\Phi(\mathbf{r}) &= \frac{1}{4\pi} \int_V \frac{\nabla' \cdot \mathbf{F}'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \\ \mathbf{A} &= \frac{1}{4\pi} \int_V \frac{\nabla' \times \mathbf{F}'(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV'\end{aligned}$$

3.7 Toroidal and Poloidal Decomposition

The earliest use of these terms cited by the Oxford English Dictionary (OED) is by Walter M. Elsasser (1946) in the context of the generation of the Earth's magnetic field by currents in the core, with “toroidal” being parallel to lines of latitude and “poloidal” being in the direction of the magnetic field (i.e. towards the poles).

The OED also records the later usage of these terms in the context of toroidally confined plasmas, as encountered in magnetic confinement fusion. In the plasma context, the toroidal direction is the long way around the torus, the corresponding coordinate being denoted by z in the slab approximation or ζ or ϕ in magnetic coordinates; the poloidal direction is the short way around the torus, the corresponding coordinate being denoted by y in the slab approximation



Figure 3.1: Toroidal-Poloidal Decomposition.

or θ in magnetic coordinates. (The third direction, normal to the magnetic surfaces, is often called the “radial direction”, denoted by x in the slab approximation and variously ψ , χ , r , ρ , or s in magnetic coordinates.)

In vector calculus, a topic in pure and applied mathematics, a poloidal-toroidal decomposition is a restricted form of the Helmholtz decomposition. It is often used in the spherical coordinates analysis of solenoidal vector fields, for example, magnetic fields and incompressible fluids.

For a three-dimensional vector field \mathbf{F} with zero divergence

$$\nabla \cdot \mathbf{F} = 0$$

This \mathbf{F} can be expressed as the sum of a toroidal field \mathbf{T} and poloidal vector field \mathbf{P}

$$\mathbf{F} = \mathbf{T} + \mathbf{P}$$

where \hat{r} is a radial vector in spherical coordinates (r, θ, ϕ) . The toroidal field is obtained from a scalar field, $\psi(r, \theta, \phi)$ as the following curl,

$$\mathbf{T} = \nabla \times (\hat{r}\Psi(\mathbf{r}))$$

and the poloidal field is derived from another scalar field $\phi(r, \theta, \phi)$ as a twice-iterated curl,

$$\mathbf{P} = \nabla \times \nabla \times (\hat{r}\Phi(\mathbf{r}))$$

This decomposition is symmetric in that the curl of a toroidal field is poloidal, and the curl of a poloidal field is toroidal. The poloidal-toroidal decomposition is unique if it is required that the average of the scalar fields Ψ and Φ vanishes on every sphere of radius r .

3.8 Magnetic Dipole Field

If there's no current,

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = 0 \Rightarrow B = -\nabla V$$

The divergence free condition for magnetic field then yields a laplace equation

$$\Delta V = 0$$

for the Gauss potential V . The general complex solution of Laplace equation in spherical coordinates is

$$\begin{aligned} V(\mathbf{r}) &= \sum_{l=0}^{\infty} \sum_{m=-l}^l (A_l r^l + B_l r^{-l-1}) P_l^m(\cos \theta) e^{-im\phi} \\ &= \sum_{l=0}^{\infty} \sum_{m=-l}^l (A_l r^l + B_l r^{-l-1}) Y_l^m(\theta, \phi) \end{aligned}$$

where

$$Y_l^m(\theta, \phi) = P_l^m(\cos \theta) e^{-im\phi}$$

are the (complex) spherical harmonics. The indices l and m indicate degree and order of the function. The general real solution is

$$V(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=0}^l (A_l r^l + B_l r^{-l-1}) P_l^m(\cos \theta) [S_m \sin m\phi + C_m \cos m\phi]$$

P_l^m are the [associated Legendre polynomials](#). The first three terms of $P_l^m(\cos \theta)$ are

$$\begin{aligned} P_0^0(\cos \theta) &= 1 \\ P_1^0(\cos \theta) &= \cos \theta \\ P_1^1(\cos \theta) &= -\sin \theta \end{aligned}$$

There are two parts in the real solution. The first part, starting with coefficients A_l , represents the potential from an exterior source, which is analytic at $r = 0$ but diverges as $r \rightarrow \infty$. The second part, starting with coefficients B_l , represents the potential from an interior source, which is singular at $r = 0$ and finite as $r \rightarrow \infty$.

Ex 1. Planet dipole field

For a planetary dipole field, we only consider the contribution from the internal source. The Gauss potential for the dipole reads

$$\begin{aligned} V &= \frac{1}{r^2} [B_1 C_0 \cos \theta + B_1 C_1 \cos \phi \sin \theta + B_1 S_1 \sin \phi \sin \theta] \\ &= \frac{1}{r^2} [G_{10} \cos \theta + G_{11} \cos \phi \sin \theta + H_{11} \sin \phi \sin \theta] \end{aligned}$$

If r is defined in the unit of planet radius, then the dipole moment $\mathbf{m} = (G_{11}, H_{11}, G_{10})$ in Cartesian coordinates would have the SI unit of $[\text{Tm}^3]$. We then have

$$\mathbf{B} = -\nabla V = -\nabla \left[\frac{\mathbf{m} \cdot \mathbf{r}}{r^3} \right] = \frac{3(\mathbf{m} \cdot \hat{r})\hat{r} - \mathbf{m}}{r^3}$$

Often in literature we see dipole moments in the unit of $[\text{nT}] [\text{R}]^3$: this indicates that this dipole strength is equivalent to the equatorial field strength B_{eq} at the planet's surface, and the field strength is scaled as r^{-3} .

In SI units, we need to append additional coefficients

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{3(\mathbf{m} \cdot \hat{r})\hat{r} - \mathbf{m}}{r^3}$$

such that the dipole moment \mathbf{m} has the unit of [A] [m]², consistent with its physical meaning originated from a closed current loop.

In spherical coordinates (r, θ, φ) aligned with the dipole moment,

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{m}{r^3} (-2 \cos \theta, -\sin \theta, 0) \quad (3.6)$$

It is usually convenient to work in terms of the latitude, $\vartheta = \pi/2 - \theta$, rather than the polar angle, θ . An individual magnetic field-line satisfies the equation

$$r = r_{\text{eq}} \cos^2 \vartheta$$

where r_{eq} is the radial distance to the field line in the equatorial plane ($\vartheta = 0^\circ$). It is conventional to label field lines using the L-shell parameter, $L = r_{\text{eq}}/R_E$. Here, $R_E = 6.37 \times 10^6$ m is the Earth's radius. Thus, the variation of the magnetic field-strength along a field line characterized by a given L -value is

$$B = \frac{B_E}{L^3} \frac{(1 + 3 \sin^2 \vartheta)^{1/2}}{\cos^6 \vartheta}$$

where $B_E = \mu_0 M_E / (4\pi R_E^3) = 3.11 \times 10^{-5}$ T is the equatorial magnetic field-strength on the Earth's surface.

In Cartesian representation,

$$\mathbf{B} = \frac{\mu_0}{4\pi} \frac{1}{r^5} \begin{pmatrix} (3x^2 - r^2) & 3xy & 3xz \\ 3yx & (3y^2 - r^2) & 3yz \\ 3zx & 3zy & (3z^2 - r^2) \end{pmatrix} \mathbf{m} \quad (3.7)$$

Note that (x, y, z) and r are all normalized to the radius of the planet R .

Ex 2. 2D dipole field

Let us consider the solution to the Laplace equation in polar coordinates (r, θ) . Via separation of variable it can be shown that the general real solution takes the form

$$V = c_0 \ln r + D_0 + \sum_{n=1}^{\infty} (A_n \cos n\theta + B_n \sin n\theta) (C_n r^n + D_n r^{-n})$$

For a dipole field, we only consider $n = 1$ from the internal source. If we absorb D_1 into A_1 and B_1 , we have

$$V = (A_1 \cos \theta + B_1 \sin \theta) \frac{1}{r} = \frac{\mathbf{m} \cdot \mathbf{r}}{r^2}$$

where the 2D magnetic moment is $\mathbf{m} = (A_1, B_1)$.

Thus

$$\mathbf{B}_{\text{dipole},2D} = -\nabla V = \frac{2(\mathbf{m} \cdot \hat{r})\hat{r} - \mathbf{m}}{r^2} \quad (3.8)$$

where the minus sign has been absorbed into the coefficients. 2D dipole scales as r^{-2} .

Ex 3. Magnetic monopoles

Even though magnetic monopoles have not yet been observed in nature, there are some nice math tricks that allow us to approximate the magnetic dipole field as two opposite charged magnetic monopoles sitting close to each other. In 3D, the magnetic monopole would give a magnetic field

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{g}{r^2} \hat{r}$$

where g is the strength of the magnetic monopole in units of magnetic charge. In 2D, it would be

$$\mathbf{B}(\mathbf{r}) = C \frac{g}{r} \hat{r}$$

The analogy can be made from Coulomb's law: an electric dipole field in 3D scales as r^{-3} when the electric monopole field scales as r^{-2} .

3.9 Green's Function

The Vlasov theory and the 2-fluid theory only tell us when instability will happen, but neither tell us the severity of instability. To know the magnitude of instability mathematically, we can introduce Green's function

$$G(x, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx - i\omega(k)t} dk$$

where t is the beam's pulse length and x is the propagation distance. At $t = 0$, we have

$$\begin{aligned} G(x, 0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} dk = \frac{1}{2i\pi x} e^{ikx} \Big|_{k=-\infty}^{k=\infty} \\ &= \lim_{k \rightarrow \infty} \frac{1}{2i\pi x} [\cos kx + i \sin kx - \cos kx - i \sin kx] \\ &= \lim_{k \rightarrow \infty} \frac{1}{\pi} \frac{\sin kx}{x} = \delta(x) \end{aligned}$$

where $\delta(x)$ is the δ -function.

The integral $\int_0^\infty \frac{\sin x}{x} dx$ is called the Dirichlet integral. It is a very famous and important generalized improper integral. Here at least you can see that

$$\int_{-\infty}^{\infty} G(x, 0) = \frac{1}{\pi} \int_0^\infty \frac{\sin kx}{x} dx = 1 \quad \text{and} \quad G(0, 0) = \infty$$

3.10 Linearization

In linear wave analysis, we need to check how small perturbations evolve in the PDEs. For plasma physics, it is usually very useful to transform into a magnetic field guided coordinates, with the notation \parallel being parallel to the background magnetic field and \perp being perpendicular to the background magnetic field \mathbf{B} . Besides the perturbation of plasma moments (i.e. density, velocity, pressure, etc.), we also need the perturbations to the magnitude of the magnetic field B and the unit vector $\hat{\mathbf{b}}$. Linearizing $B^2 = \mathbf{B} \cdot \mathbf{B}$, we find

$$\delta B = \hat{\mathbf{b}} \cdot \delta \mathbf{B} = \delta B_{\parallel} \quad (3.9)$$

Linearizing $\hat{\mathbf{b}} = \mathbf{B}/B$, we obtain

$$\delta \hat{\mathbf{b}} = \delta \left(\frac{\mathbf{B}}{B} \right) = \frac{\delta \mathbf{B} B - \delta B \mathbf{B}}{B^2} = \frac{\delta \mathbf{B}}{B} - \frac{\delta \mathbf{B}_{\parallel}}{B} = \frac{\delta \mathbf{B}_{\perp}}{B} \quad (3.10)$$

The divergence-free of magnetic field gives

$$\mathbf{k} \cdot \delta \mathbf{B} = k_{\parallel} \delta B_{\parallel} + k_{\perp} \delta B_{\perp} = 0$$

Thus

$$\mathbf{k} \cdot \hat{\mathbf{b}} = \mathbf{k} \cdot \frac{\delta \mathbf{B}_{\perp}}{B} = \frac{k_{\perp} \delta B_{\perp}}{B} = -k_{\parallel} \frac{\delta B_{\parallel}}{B} = -k_{\parallel} \frac{\delta B}{B} \quad (3.11)$$

These seemingly trivial relations have profound implications in physics. Equation 3.9 tells us that the perturbation of magnetic field magnitude has only contribution from the parallel component, which is why in satellite observations people only look at parallel perturbations for compressional wave modes. Equation 3.10 tells us that the perturbation in the unit vector is only related to the perpendicular fluctuations.

3.11 Wave Equations

The waves in plasma physics is governed by second order ODEs. Here we list some second order ODEs that has been studied mostly in plasma physics.

- Schrödinger Equation:

$$\frac{d^2\varphi}{dx^2} + \frac{2m}{\hbar^2} [E - U(x)] \varphi = 0$$

Schrödinger Equation appears in the nonlinear wave studies (Chapter 17).

- Shear Alfvén wave:

$$\frac{d}{dx} \left\{ \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A(x))^2] \frac{dE}{dx} \right\} - k^2 \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2 - g \frac{1}{\rho_0} \frac{d\rho_0}{dx}] E = 0$$

The Shear Alfvén wave equation appears in nonuniform ideal MHD (Equation 10.30, Equation 16.8).

- EM waves in non-magnetized plasma, O mode:

$$\frac{d^2 E}{dx^2} + \frac{\omega^2}{c^2} \left[1 - \frac{\omega_{pe}(x)^2}{\omega^2} \right] E = 0 \quad (3.12)$$

- Electron cyclotron resonance heating (ECRH):

$$\frac{d^2 E}{dx^2} + \frac{\omega^2}{c^2} \left[1 - \frac{\omega_{pe}(x)^2}{\omega(\omega - \Omega_e(x))} \right] E = 0$$

In general, a second order ODE

$$\frac{d^2 u(x)}{dx^2} + a_1(x) \frac{du(x)}{dx} + a_2(x) u(x) = 0$$

can be rewritten to get rid of the first derivative. Let

$$u(x) = E(x) e^{-\frac{1}{2} \int^x a_1(x) dx}$$

we have

$$\frac{d^2 E(x)}{dx^2} + k^2(x) E(x) = 0 \quad (3.13)$$

where

$$k^2(x) = a_2(x) - \frac{a_1^2}{4} - \frac{1}{2} \frac{da_1(x)}{dx}$$

Note that the lower bound of the integral is left on purpose to account for a constant. We will concentrate at special points, i.e. zeros (cutoff) and poles (resonance) of $k^2(x) \equiv \frac{\omega^2}{c^2} n^2(x)$.

First, we will introduce *Wentzel-Kramers-Brillouin-Jeffreys* (WKBJ) solution to Equation 3.13:

$$E(x) \sim \frac{1}{\sqrt{k(x)}} e^{\pm i \int^x k(x') dx'}.$$

Proof.

For simplicity, here we use a simple harmonic oscillation analogy. Consider

$$\frac{d^2x(t)}{dt^2} + \Omega^2(t)x(t) = 0$$

Assume $\Omega \gg 1$, and T is the time scale over which Ω changes appreciably. We would anticipate solutions like

$$\begin{aligned} x(t) &\sim e^{\pm i\Phi(t)} \\ \dot{x}(t) &\sim \pm i\dot{\Phi}(t)x(t) \\ \ddot{x}(t) &\sim -\dot{\Phi}^2(t)x(t) \cancel{+ i\ddot{\Phi}(t)\overline{x(t)}} \\ \Rightarrow \dot{\Phi}(t) &= \Omega(t), \text{ or } \Phi(t) = \int^t \Omega(t') dt' + \text{const.} \end{aligned}$$

Write $x(t) = A(t) \sin[\Phi(t)]$, where $A(t)$ is slowly varying in time, $\dot{A}(t) \ll \Omega A$, which is almost a periodic motion. From adiabatic theory in classical mechanics, $\oint pdq \simeq \text{const.}$, we have

$$\begin{aligned} \oint mv_x dx &\simeq \text{const.} \\ \oint m\dot{x}dx &\simeq \text{const.} \end{aligned}$$

Then in a period $2\pi/\Omega$,

$$\frac{2\pi}{\Omega} \oint m\dot{x}^2 dt = \frac{2\pi}{\Omega} \oint mA^2\Omega^2 \cos^2 \Phi dt \simeq \text{const.}$$

which leads to

$$mA^2\Omega \simeq \text{const.}, \quad A \sim \frac{1}{\sqrt{\Omega}}$$

Thus the general form of solution can be written as

$$x(t) \sim \frac{1}{\sqrt{\Omega}} e^{\pm i\Phi(t)} \sim \frac{1}{\sqrt{\Omega}} e^{\pm i \int^t \Omega(t') dt' + \text{const.}}$$

Note:

1. There is no lower limit to the integral, because it is like adding a constant.
2. This solution is valid if $k \cdot L > O(3)$, where L is the length scale over which $k^2(x)$ changes appreciably. (???) Formally the condition should be $kL \gg 1$. Apparently near resonance ($k \rightarrow \infty$), the condition breaks down, then how do we reconcile the solution? There are more discussions coming up and applications in Chapter 10 and Chapter 16.

3.11.1 Airy Function

We want to develop a general method for cut-off and resonance. Away from the turning point x_t ,

$$E_{\text{WKB}}(x) \sim \frac{1}{\sqrt{k(x)}} e^{\pm i \int^x k(x) dx}$$

Near $x = x_t$, we can use a linear approximation for $k^2(x)$ (first term in the Taylor expansion),

$$k^2(x) \approx -k_0^2 a(x - x_t)$$

Then

$$\begin{aligned} \frac{d^2 E}{dx^2} + k^2(x)E &= 0 \\ \frac{d^2 E}{dx^2} - k_0^2 a(x - x_t)E &= 0 \end{aligned}$$

Let $\frac{x-x_t}{l} = \zeta$, we have

$$\frac{d^2 E}{d\zeta^2} - l^2 k_0^2 a l \zeta E(\zeta) = 0$$

If we choose l s.t. $l^3 k_0^2 a = 1$ (non-dimensional treatment), we obtain

$$\frac{d^2 E}{d\zeta^2} - \zeta E(\zeta) = 0 \quad (3.14)$$

which is equivalent to the case where $k^2(\zeta) = -\zeta$ that shows the linear approximation. Equation 3.14, known as the *Airy equation* or *Stokes equation*, is the simplest second-order linear differential equation with a turning point (a point where the character of the solutions changes from oscillatory to exponential). From the WKB theory, we get the solution

$$E_{\text{WKB}} \sim \frac{1}{\sqrt{k(x)}} e^{\pm i \int^x k(x') dx'} = \begin{cases} \zeta^{-1/4} e^{\mp \frac{2}{3} \zeta^{3/2}} & \text{if } \zeta > 0 \\ (-\zeta)^{-1/4} e^{\pm i \frac{2}{3} (-\zeta)^{3/2}} & \text{if } \zeta < 0 \end{cases}$$

Note that the solution blows up at $\zeta = 0$ (turning point) miserably. For $\zeta > 0$, one solution is exponentially decay while the other is exponentially growing; for $\zeta < 0$, the two solutions are oscillatory. Solutions can also be found as series in ascending powers of ζ by the standard method. Assume that a solution is $E = a_0 + a_1 \zeta + a_2 \zeta^2 + \dots$. Substitute this in Stokes Equation and equate powers of ζ will give relations between the constants a_0, a_1, a_2 , etc., and lead finally to

$$\begin{aligned} E &= a_0 \left\{ 1 + \frac{\zeta^3}{3 \cdot 2} + \frac{\zeta^6}{6 \cdot 5 \cdot 3 \cdot 2} + \frac{\zeta^9}{9 \cdot 8 \cdot 6 \cdot 5 \cdot 3 \cdot 2} + \dots \right\} \\ &\quad + a_1 \left\{ \zeta + \frac{\zeta^4}{4 \cdot 3} + \frac{\zeta^7}{7 \cdot 6 \cdot 4 \cdot 3} + \frac{\zeta^{10}}{10 \cdot 9 \cdot 7 \cdot 6 \cdot 4 \cdot 3} + \dots \right\} \end{aligned}$$

which contains the two arbitrary constants a_0 and a_1 , and is therefore the most general solution. The series are convergent for all ζ , which confirms that every solution of Stokes Equation is finite, continuous and single valued.

This form is usually not easy to interpret in physical sense. Besides this, we can find a more useful solution to Equation 3.14 using the integral representation. An equivalent but maybe more intuitive approach is to solve Equation 3.14 with Fourier transform; a coefficient 2π naturally appears. Both approaches reach the same results.

Write

$$E(\zeta) = \int_a^b dt e^{t\zeta} f(t)$$

where the integral represents a path in the complex t plane from a to b . Then

$$\begin{aligned} \frac{d^2 E}{d\zeta^2} &= \int_a^b dt t^2 e^{t\zeta} f(t) \\ \zeta E(\zeta) &= \int_a^b dt \zeta e^{t\zeta} f(t) = \int_a^b dt \frac{d}{dt} (e^{t\zeta}) f(t) \\ &= e^{t\zeta} f(t) \Big|_a^b - \int_a^b dt e^{t\zeta} f'(t) \end{aligned}$$

The limits a, b are chosen so that the first term vanishes at both limits. Then Equation 3.14 is satisfied if

$$\begin{cases} e^{t\zeta} f(t) \Big|_a^b = 0 \\ t^2 f(t) = -\frac{df(t)}{dt} \Rightarrow f(t) = Ae^{-\frac{1}{3}t^3} \end{cases}$$

where A is a constant. The solution is now written as

$$E(\zeta) = \int_a^b dt A e^{t\zeta - \frac{1}{3}t^3}$$

The limits a and b must therefore be chosen so that $e^{t\zeta - \frac{1}{3}t^3}$ is zero for both. (Note that ζ is a constant.) This is only possible if $t = \infty$ and the real part of t^3 is positive:

$$\begin{aligned} \Re(t^3) > 0 &\Leftrightarrow 2\pi n - \frac{1}{2}\pi < \arg t^3 < 2\pi n + \frac{1}{2}\pi \\ &\Leftrightarrow 2\pi n - \frac{1}{2}\pi < 3\arg t < 2\pi n + \frac{1}{2}\pi \end{aligned}$$

where n is an integer. Figure 3.2 is a diagram of the complex t -plane, and a and b must each be at ∞ in one of the shaded sectors. They cannot both be in the same sector, for then the integral would be zero. Hence the contour may be chosen in three ways, as shown by the three curves C_1, C_2, C_3 , and the solution would be

$$E(\zeta) = \int_{C_1, C_2, C_3} dt e^{t\zeta - \frac{1}{3}t^3}$$

This might appear at first to give three independent solutions, but the contour C_1 and be distorted so as to coincide with the two contours $C_2 + C_3$, so that

$$\int_{C_1} = \int_{C_2} + \int_{C_3}$$

and therefore there are only two independent solutions.

Proof.

Note that:

1. $e^A \neq 0 \forall |A| < \infty$;
2. $e^A = 0 \Leftrightarrow A \rightarrow \infty$ and $\Re(A) < 0$.

Therefore we have

$$e^{-\frac{1}{3}t^3} \rightarrow 0 \text{ as } t \rightarrow \infty \text{ and } \Re\left(\frac{1}{3}t^3\right) > 0$$

In polar coordinates, let $t = |t|e^{i\theta}$. Then

$$t^3 = |t|^3 e^{i3\theta} = |t|^3 (\cos 3\theta + i \sin 3\theta)$$

$$\Re\left(\frac{1}{3}t^3\right) > 0 \Leftrightarrow \cos 3\theta > 0 \Leftrightarrow 3\theta \in (-\frac{\pi}{2}, \frac{\pi}{2}), (\frac{3\pi}{2}, \frac{5\pi}{2}), (-\frac{3\pi}{2}, -\frac{5\pi}{2})$$

Jeffreys (1956) defines two special [Airy functions](#) $Ai(x), Bi(x)$ as follows

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i} \int_{C_1} dt e^{-\frac{1}{3}t^3 + \zeta t} \\ Bi(\zeta) &= \frac{1}{2\pi} \left[\int_{C_2} + \int_{C_3} \right] dt e^{-\frac{1}{3}t^3 + \zeta t} \end{aligned} \tag{3.15}$$



Figure 3.2: Solution to Stokes Equation in complex t-plane.

Obviously he took the Fourier transform such that a coefficient 2π naturally appears. In Equation 3.15, the contour C_1 can be distorted so as to coincide with the imaginary t-axis for almost its whole length. It must be displaced slightly to the left of this axis at its ends to remain in the shaded region at infinity. Let $t = is$. Then the Airy function of the first kind in Equation 3.15 becomes

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i(\zeta s + \frac{1}{3}s^3)} ds \\ &= \frac{1}{\pi} \int_0^{\infty} \cos\left(\zeta s + \frac{1}{3}s^3\right) ds \end{aligned}$$

It is known as the Airy integral, which is the solution for $y \rightarrow 0$ as $x \rightarrow \infty$. The other linearly independent solution, the Airy function of the second kind, denoted $Bi(x)$, is defined as the solution with the same amplitude of oscillation as $Ai(x)$ as $x \rightarrow -\infty$ which differs in phase by $\pi/2$:

$$Bi(\zeta) = \frac{1}{\pi} \int_0^{\infty} \left[e^{-\frac{s^3}{3} + s\zeta} + \sin\left(\zeta s + \frac{1}{3}s^3\right) \right] ds$$

$Ai(x)$ and $Bi(x)$ are shown in ?@fig-airy.

```
KeyNotes.plot_airy()
```

As an interesting experiment, we can check if $Ai(x)$ is recovered from solving the second order ODE numerically:

```
@sco KeyNotes.plot_airy_ode()
```

Here we start from the right boundary and move towards the left.

Now, to find approximations for Airy functions, we use the method of *steepest descent*. This approximation is based on the assumption that major contribution to the integral is from near the **saddle point**. As an example of saddle point, consider a complex function $\exp f(z) = e^{-z^2}$, where $z = x + iy$.

- If $z = \text{real} = x$, $e^{-z^2} = e^{-x^2}$;
- If $z = \text{imag} = iy$, $e^{-z^2} = e^{y^2}$.

The procedure goes as follows:

1. Detour path of integral s.t. it passes through the saddle point of the integral, along the direction of steepest descent.

2. Obtain major contribution by integrating the Gaussian function.

$$I = \int_C dt e^{f(t)}, \quad f(t) = f(t_s) + (t - t_s)f'(t_s) + \frac{1}{2}(t - t_s)^2 f''(t_s) + \dots$$

where $f'(t_s) = 0$ at the saddle point $t = t_s$ simplifies I to the integral of a Gaussian function.

For $Ai(\zeta)$,

$$\begin{aligned} f(t) &= t\zeta - \frac{1}{3}t^3 \\ f'(t_s) &= \zeta - t_s^2 = 0 \\ f''(t_s) &= -2t_s \end{aligned}$$

Consider $\zeta > 0$ (where the solution is either exponentially decaying or growing.) Then

$$\begin{aligned} t_{s1} &= -\sqrt{\zeta}, \quad t_{s2} = \sqrt{\zeta} \\ f(t_{s1}) &= -\frac{2}{3}\zeta^{3/2}, \quad f''(t_{s1}) = 2\zeta^{1/2} \end{aligned}$$

So

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i} \int_{C_1} dt e^{-\frac{2}{3}\zeta^{3/2} + \zeta^{1/2}(t - t_s)^2} + \dots \\ &\approx \frac{1}{2\pi i} e^{-\frac{2}{3}\zeta^{3/2}} \int_{C_1} dt e^{\sqrt{\zeta}(t - t_s)^2} \end{aligned}$$

Let $e^{\sqrt{\zeta}(t - t_s)^2} = e^{-\rho^2}$, where $\rho = \text{real}$ is the direction of steepest descent, and $\rho = \text{imag}$ is the direction of steepest ascent.

$$\begin{aligned} i\rho &= \pm\zeta^{1/4}(t - t_s) \\ \frac{d\rho}{dt} &= \frac{i}{\zeta^{1/4}} \Rightarrow dt \text{ is purely imaginary along steepest descent.} \end{aligned}$$

Then for $\zeta > 0$

$$\begin{aligned} Ai(\zeta) &= \frac{1}{2\pi i} e^{-\frac{2}{3}\zeta^{3/2}} \int_{-\infty}^{\infty} \frac{id\rho}{\zeta^{1/4}} e^{-\rho^2} \\ &= \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{-\frac{2}{3}\zeta^{3/2}} \text{ as } \zeta \rightarrow \infty \end{aligned}$$

The ratio of accuracy is shown in the following table (LABEL???). In practice, it is already a very good approximation when $\zeta > 3$. (think of $kL \gg 3$ for WKBJ solution! ???)

ζ	1	2	3	6	∞
ratio	0.934	0.972	0.983	0.993	1

Now, consider $\zeta < 0$. When $\zeta \rightarrow -\infty$, we anticipate oscillating behavior of $Ai(\zeta)$. $\zeta = -|\zeta|$,

$$\begin{cases} f(t) = \zeta t - \frac{1}{3}t^3 = -t|\zeta| - \frac{1}{3}t^3 \\ f'(t) = -|\zeta| - t^2 = 0 \\ f''(t) = -2t \end{cases}$$

$$\Rightarrow f(t_{s1}) = i\frac{2}{3}|\zeta|^{3/2}$$

$$f''(t_{s1}) = 2i\sqrt{|\zeta|}$$

so the contribution from t_{s1} is

$$\frac{1}{2\pi i} e^{i\frac{2}{3}|\zeta|^{3/2}} \int_{t_{s1}} e^{\frac{1}{2}(2i\sqrt{|\zeta|})(t-t_s)^2 + \dots}$$

Let $-\rho^2 = i\sqrt{|\zeta|}(t - t_{s1})^2$, differentiate on both sides to get

$$dt = \pm \frac{e^{i\pi/4}}{|\zeta|^{1/4}} d\rho$$

Again, $\rho = \text{real}$ is the direction of steepest descent at $t = t_{s1}$. Do the same to t_{s2} , then by summing them up we have for $\zeta < 0$,

$$\begin{aligned} Ai(\zeta) &\approx \frac{1}{2\pi i} \left[e^{i\frac{2}{3}|\zeta|^{3/2}} \int_{-\infty}^{\infty} \frac{e^{i\pi/4}}{|\zeta|^{1/4}} d\rho e^{-\rho^2} - e^{-i\frac{2}{3}|\zeta|^{3/2}} \int_{-\infty}^{\infty} \frac{e^{-i\pi/4}}{|\zeta|^{1/4}} d\rho e^{-\rho^2} \right] \\ &= \frac{1}{2\pi i} \left[\frac{e^{i\frac{2}{3}|\zeta|^{3/2}+i\pi/4}}{|\zeta|^{1/4}} \sqrt{\pi} - \frac{e^{-i\frac{2}{3}|\zeta|^{3/2}-i\pi/4}}{|\zeta|^{1/4}} \sqrt{\pi} \right] \end{aligned} \quad (3.16)$$

An interesting application of the gradient descent method is to find [Stirling's approximation](#). Mathematically, we can proof that

$$n! = \int_0^\infty dt e^{-t} t^n = \int_0^\infty dt e^{-t+n \ln t} \equiv \int_0^\infty dt e^{f(t)}$$

and by following the steepest descent method,

$$\begin{aligned} f(t) &= -t + n \ln t \\ f'(t) &= -1 + n \frac{1}{t} = 0 \\ f''(t) &= -n \frac{1}{t^2} \end{aligned}$$

we can find the approximation Stirling formula

$$\begin{aligned} n! &\approx \int_0^\infty dt e^{-n+n \ln n - \frac{1}{2n}(t-t_s)^2} \\ &= e^{-n} n^n \int_0^\infty dt e^{-\frac{1}{2n}(t-t_s)^2} \\ &= \sqrt{2n\pi} e^{-n} n^n \end{aligned}$$

The following table (LABEL???) show the goodness of approximation of Stirling formula. In practice, it is already a very good approximation when $n > 3$.

ζ	1	2	3	10	∞
ratio	0.922	0.9595	0.973	0.9917	1

See (Budden 1961) *Chapter 15: The Airy Integral Function, And The Stokes Phenomenon* for more details.

3.11.2 Uniformly Valid WKBJ Solution Across the Turning Point

In this section, we present the WKBJ solution that is uniformly valid everywhere, even at the turning point.

Consider the standard equation,

$$\frac{d^2 E(x)}{dx^2} + k^2(x)E(x) = 0$$

Let $x = 0$ be the turning point, i.e. we assume that $k^2(x)$ is a monotonically decreasing function of x with $k(0) = 0$. For the region $x > 0$, we first identify the exponentially decaying factor of the Airy function, $Ai(\zeta)$, to be the phase integral in the WKBJ solution:

$$-\frac{2}{3}\zeta^{3/2} = i \int_0^x k(x') dx'$$

Note that this yields $\zeta = \zeta(x)$, a known function of x (I confuse myself later about ζ and x ...). The branch for $\zeta(x)$ is to be chosen so that for $x > 0$, ζ is real and positive.

We next verify that the uniformly valid solution to the standard equation is simply

$$E(x) = \frac{1}{\sqrt{\zeta'(x)}} Ai(\zeta) \tag{3.17}$$

where the prime denotes a derivative. For large values of ζ , we can use the asymptotic formula of Ai Equation 3.16, and notice that

$$\begin{aligned}-\zeta^{1/2}\zeta' &= ik(x) \\ \zeta'(x) &= -i \frac{k(x)}{\zeta^{1/2}}\end{aligned}$$

we can see that Equation 3.17 reduces to the standard WKBJ solutions for large values of ζ

$$E(x) = \frac{1}{\sqrt{\zeta'(x)}} Ai(\zeta) \sim \frac{1}{\sqrt{k(x)}} e^{-\frac{2}{3}\zeta^{3/2}} = \frac{1}{\sqrt{k(x)}} e^{i \int_0^x k(x') dx'}$$

We can also show that Equation 3.17 is valid for small values of ζ , i.e. near the turning point at $x = 0$. (Hint: Near $x = 0$, $k^2(x)$ may be approximated by a linear function of x . This linear approximation then yields $\zeta(x)$ as a linear function of x according to Equation 3.17.)

Ex. Choose a smooth plasma density profile which monotonically increases with x s.t.

$$\frac{\omega_{pe}^2(x)}{\omega^2} = 1 + \tanh x$$

and launch a wave of frequency ω from $x = -\infty$, the vacuum region, toward the positive x -direction with $\frac{\omega^2}{c^2} = 10 \text{ m}^{-2}$. (This is like launching a wave from low B region into high B region.) Numerically integrate the wave equation Equation 3.12,

$$\frac{d^2 E}{dx^2} + \frac{\omega^2}{c^2} \left[1 - \frac{\omega_{pe}(x)^2}{\omega^2} \right] E = 0$$

from some large positive values of x , we get the results in Figure 3.3.

We know that away from x_t , WKBJ solution works. To the left of x_t (with $\zeta < -3$), Equation 3.16 gives

$$Ai(\zeta) = \frac{1}{2i\sqrt{\pi}|\zeta|^{1/4}} \left[e^{i\frac{2}{3}|\zeta|^{3/2} + i\frac{\pi}{4}} - e^{-\frac{2}{3}|\zeta|^{3/2} - i\frac{\pi}{4}} \right]$$

Choose the branch s.t. $\zeta x > 0$ if $x > x_t$; $\zeta(x) < 0$ if $x < x_t$.

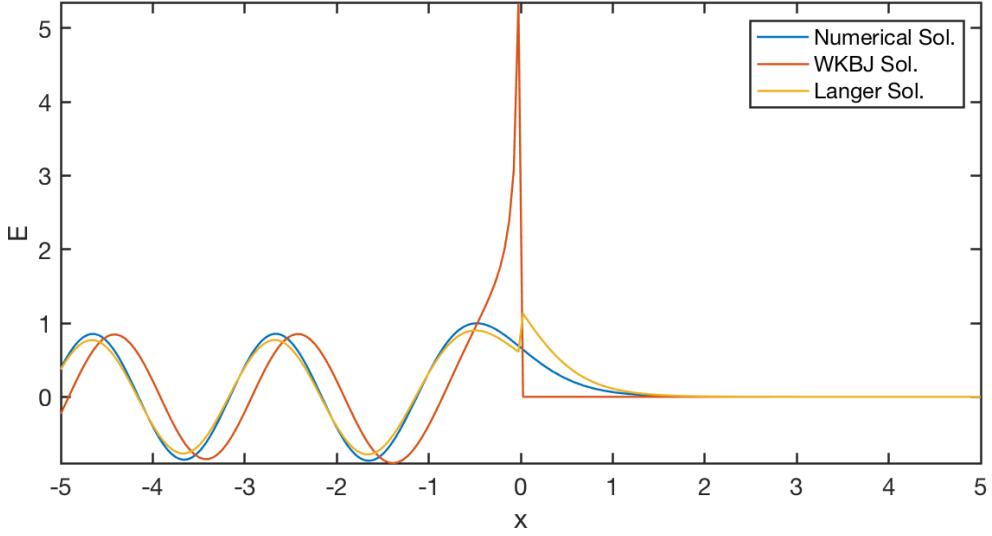


Figure 3.3: Comparison between the WKBJ solution, uniformly valid solution and numerical integral solution for the O-mode with monotonically increasing density with x .

$$\begin{aligned}
E(x) &= \frac{1}{\sqrt{\zeta'(x)}} Ai(\zeta) \\
&= \frac{C_3}{\sqrt{k(x)/|\zeta|^{1/2}}} \frac{1}{|\zeta|^{1/4}} \left\{ e^{i\frac{2}{3}|\zeta|^{3/2} + i\frac{\pi}{4}} - c.c \right\} \\
&= \frac{C_3}{\sqrt{k(x)}} \left\{ e^{i(-\int_{x_t}^x k(x') dx'') + i\frac{\pi}{4}} - c.c \right\} \\
&= \frac{C_3}{\sqrt{k(x)}} \left\{ e^{-i\int_{x_t}^a k(x') dx' - i\int_a^x k(x') dx'' + i\frac{\pi}{4}} - c.c \right\} \\
&= \frac{C_4}{\sqrt{k(x)}} \left\{ e^{-i\int_a^x k(x') dx'} - e^{i\int_{x_t}^a k(x') dx'} e^{2i\int_{x_t}^a k(x') dx'' - i\frac{\pi}{2}} \right\} \\
&= \frac{C_4}{\sqrt{k(x)}} \left\{ e^{-i\int_a^x k dx'} + R \cdot e^{i\int_a^x k dx'} \right\},
\end{aligned}$$

where

$$C_4 = C_3 \cdot e^{-i\int_{x_t}^a k dx' + i\frac{\pi}{4}}$$

and

$$R = -e^{2i\int_{x_t}^a k dx' - i\frac{\pi}{2}} = ie^{2i\int_{x_t}^a k dx'} = ie^{-2i\int_a^{x_t} k dx'}$$

is the reflection coefficient at $x = a$.

3.11.3 Stokes Phenomenon

In complex analysis the **Stokes phenomenon** is that the asymptotic behavior of functions can differ in different regions of the complex plane, and that these differences can be described in a quantitative way.

Ex. For the simple wave equation

$$\frac{d^2E}{dz^2} - E = 0$$

the solution can be given in various forms

$$\begin{aligned} E &= e^z, e^{-z}, \cosh z, \sinh z \\ E = \cosh z &= \frac{1}{2}(e^z + e^{-z}) \Rightarrow \begin{cases} E \sim \frac{1}{2}e^z z \rightarrow \infty \\ E \sim \frac{1}{2}e^{-z} z \rightarrow -\infty \end{cases} \\ E = \sinh z &= \frac{1}{2}(e^z - e^{-z}) \Rightarrow \begin{cases} E \sim \frac{1}{2}e^z z \rightarrow \infty \\ E \sim \frac{-1}{2}e^{-z} z \rightarrow -\infty \end{cases} \end{aligned}$$

Note that if a solution is exponentially growing in one direction, its asymptotic solution can contain an arbitrary amount of the exponentially decaying solution; that is, specifying an asymptotic growing solution in one direction cannot completely specify the solution in the entire complex plane.

The two linearly independent approximate solutions to Airy Equation 3.14 are the Airy function approximations from WKBJ method:

$$\begin{aligned} \frac{d^2E}{d\zeta^2} - \zeta E &= 0 \\ \Rightarrow E(\zeta) &= \begin{cases} Ai(\zeta) \sim \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{-\frac{2}{3}\zeta^{3/2}}, & \zeta > 0 \\ Bi(\zeta) \sim \frac{1}{2\sqrt{\pi}\zeta^{1/4}} e^{\frac{2}{3}\zeta^{3/2}}, & \zeta > 0 \end{cases} \end{aligned}$$

which is very accurate for $\zeta > 3$ (see previous section).

Stokes found that you can add an arbitrary amount of $Ai(\zeta)$ to $Bi(\zeta)$ without changing the behaviour of solution. ($Ai(\zeta) < O(\zeta^{-1})$???)

We want to find ζ s.t. E_{WKBJ} is purely growing/decaying exponentially:

$$\zeta^{3/2} \text{ is purely real} \iff (|\zeta|e^{i\theta})^{3/2} \text{ is purely real} \iff \sin \frac{3}{2}\theta = 0, \theta = 0, \pm \frac{2}{3}\pi, \pm \frac{4}{3}\pi$$

The lines in the complex plane on which WKBJ solution is purely growing/decaying exponentially are called *Stokes lines* (Figure 3.4). It is accompanied with *anti-Stokes lines* (in the opposite direction to Stokes lines), on which WKBJ solution is purely oscillatory. The exponentially growing solution on Stokes lines is called the *dominant solution*; the decaying solution on Stokes lines is called the *sub-dominant/recessive solution*. The sub-dominant solution will always becomes dominant in a neighboring Stokes line. However, the inverse is not true. (It may contain an amount of sub-dominant solution.) Each term changes from dominant to subdominant, or the reverse, when ζ crosses an anti-Stokes line???

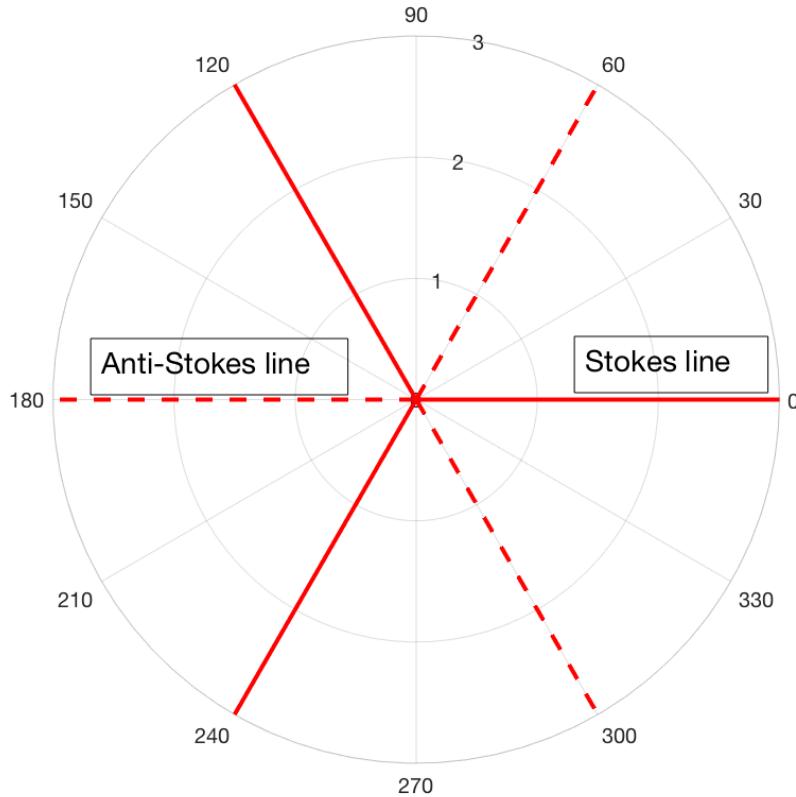


Figure 3.4: Stokes lines and anti-Stokes lines for WKBJ solution of Airy Equation.

Even though we say “arbitrary”, the analytic solution in the whole complex plane possess a limit on that amount. The next question would be: how much exponentially decaying solution can you add to the exponentially growing solution? (Note that the asymptotic series is also a divergent series: more terms don’t lead to high resolution accuracy! I have questions on this part???)

For the asymptotic approximation solution of Airy Equation 3.14, we need to define $\arg(\zeta)$

properly to make it single value. Let us choose the branch cut $-\pi \leq \arg \zeta < \pi$. The complex plane is demonstrated in Figure 3.5.

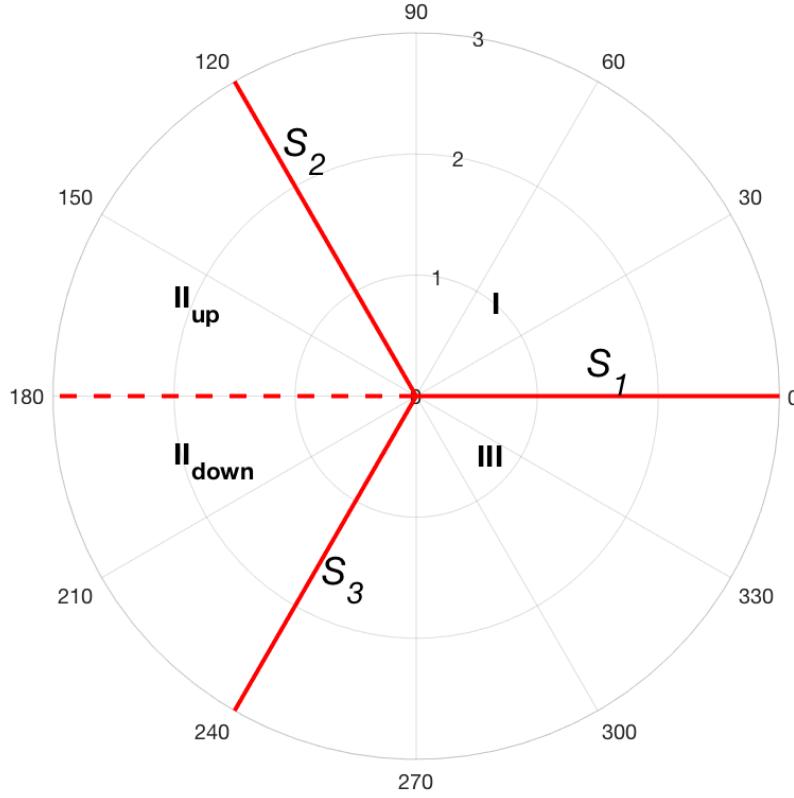


Figure 3.5: Marked Stokes lines and regions for Airy solution.

Let us start from region I. The solution in region I is

$$E_I \sim \zeta^{-1/4} [A_1 e^{-\frac{2}{3}\zeta^{3/2}} + B_1 e^{\frac{2}{3}\zeta^{3/2}}]$$

On S_1 , $e^{-\frac{2}{3}\zeta^{3/2}}$ is sub-dominant, $e^{\frac{2}{3}\zeta^{3/2}}$ is dominant. (The former would be dominant on neighboring Stokes lines S_2 and S_3 .) Crossing S_2 into region II_{up} , we have

$$E_{\text{II}_{\text{up}}} \sim \zeta^{-1/4} \left[\underbrace{A_1 e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{(\lambda_2 A_1 + B_1) e^{\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}} \right]$$

where λ_2 is the Stokes constant on S_2 . The constant in the subdominant term has changed by $\lambda_2 A_1$ because the differential equation is linear, and it cannot depend on B_1 . Otherwise it would

be unaltered if we added to the solution in region I any multiple of the solution in which $A_1 = 0$. (See (Budden 1961) 15.13.) Crossing the branch cut to region II_{down}, $\zeta^{\text{up}} = \zeta^{\text{down}} e^{i2\pi}$.

$$\begin{aligned}\zeta_{\text{up}}^{-1/4} &= \zeta_{\text{down}}^{-1/4} e^{i2\pi(-1/4)} = i\zeta_{\text{down}}^{1/4} \\ \zeta_{\text{up}}^{3/2} &= \zeta_{\text{down}}^{3/2} e^{i2\pi(3/2)} = -\zeta_{\text{down}}^{3/2}\end{aligned}$$

so

$$E_{\text{II}_{\text{down}}} = -i\zeta_{\text{II}_{\text{down}}}^{1/4} \left[A_1 e^{\frac{2}{3}\zeta_{\text{II}_{\text{down}}}^{3/2}} + (\lambda_2 A_1 + B_1) e^{-\frac{2}{3}\zeta_{\text{II}_{\text{down}}}^{3/2}} \right]$$

We can also go clockwise from I to III, crossing S_1 ,

$$E_{\text{III}} \sim \zeta^{-1/4} \left[\underbrace{B_1 e^{\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{(-\lambda_1 B_1 + A_1) e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}} \right]$$

where λ_1 is the Stokes constant on S_1 , and the minus sign indicates the clockwise direction.

Crossing S_3 from III to II_{down},

$$E_{\text{II}_{\text{down}}} \sim \zeta^{-1/4} \left[\underbrace{(-\lambda_1 B_1 + A_1) e^{-\frac{2}{3}\zeta^{3/2}}}_{\text{dominant}} + \underbrace{[(-\lambda_3)(-\lambda_1 B_1 + A_1) + B_1] e^{\frac{2}{3}\zeta^{3/2}}}_{\text{sub-dominant}} \right]$$

where λ_3 is the Stokes constant on S_3 , and the minus sign indicates the clockwise direction.

Since the solution to Airy Equation is analytic, the solutions in region II_{down} obtained from two directions must be equal. Therefore

$$\begin{aligned}-\lambda_1 B_1 + A_1 &= -i(\lambda_2 A_1 + B_1) \\ -\lambda_3(-\lambda_1 B_1 + A_1) + B_1 &= -iA_1 \\ \begin{pmatrix} i\lambda_2 + 1 & -\lambda_1 + i \\ -\lambda_3 + i & \lambda_3\lambda_1 + 1 \end{pmatrix} \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} &= 0 \\ \Rightarrow \lambda_1 &= \lambda_2 = \lambda_3 = i\end{aligned}$$

All three Stokes constants are i . For the Airy equation, the amount of exponentially decaying solution you can have going from one Stokes line to another is restricted to i in counter-clockwise direction.

3.11.4 Application of Stokes Lines

Ex.1 Reflection Coefficient from Stokes Constant ???

Ex.2 O-mode

For EM O-mode wave in a non-magnetized plasma, the governing Equation 3.12 is rewritten here

$$\frac{d^2E}{dz^2} + k^2 E = 0, \quad k^2 = \frac{\omega^2}{c^2} \left[1 - \frac{\omega_p^2(z)}{\omega^2} \right]$$

Let the turning point be $z = z_t$. Near z_t , let $k^2(z) = -p^2(z)(z - z_t)$ (Figure 3.3), where $p(z)$ is real and positive for all real z . Then

$$k = \pm ip(z)(z - z_t)^{1/2}$$

In the following we need to choose sign s.t. k is real and positive on the anti-Stokes line AS_1 . (The figure of complex z plane is the same as Figure 3.5 except that the center point is $z = z_t$. AS_1 is the dashed line.)

To make z single value, define $-\pi \leq \arg(z - z_t) < \pi$. On AS_1 ,

$$\begin{aligned} k &= \pm ip(z)|z - z_t|^{1/2} e^{i\frac{\pi}{2}} \\ &= \mp |p(z)(z - z_t)|^{1/2} \\ &= \mp |k| \end{aligned}$$

We choose the second sign for right propagating wave, which has

$$k = -ip(z)(z - z_t)^{1/2}$$

The sign is important here because it determines the propagation direction.

For z on S_1 , $\arg(z - z_t) = 0$.

$$k = -ip(z)|z - z_t|^{1/2} e^{i\frac{1}{2}0} = -i|k|$$

In region I (between S_1 and S_2)

$$E_I \sim \frac{1}{\sqrt{k}} e^{-i\omega t \pm i \int_{z_t}^z k dz'} \sim \frac{1}{\sqrt{k}} e^{-i\omega t \pm \int_{z_t}^z |k| dz'}$$

For a decaying (subdominant) solution we must use minus (lower) sign, so

$$E_I \sim E_{S_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t - \int_{z_t}^z k dz'}$$

Crossing S_2 into region II_{up} ,

$$E_{S_2} \sim \frac{1}{\sqrt{k}} e^{-i\omega t} \left[\underbrace{e^{-i \int_{z_t}^z k dz'}}_{\text{dominant on } S_2} + \underbrace{i e^{i \int_{z_t}^z k dz'}}_{\text{sub-dominant on } S_2} \right] = E_{AS_1}$$

where the first term represents the reflected wave, and the second term represents the incident wave. Since the magnitude of the incident and reflected wave is the same, the reflection coefficient should be 1, and the absorption coefficient should be 0. (In fact, there is a transmitted wave that is exponentially decay. WKBJ method cannot resolve this exponentially small value.)

Ex.3 Bohr-Sommerfeld Quantization Rule

This is the classical potential well problem, where the two boundaries are $z = z_1$ and $z = z_2$.

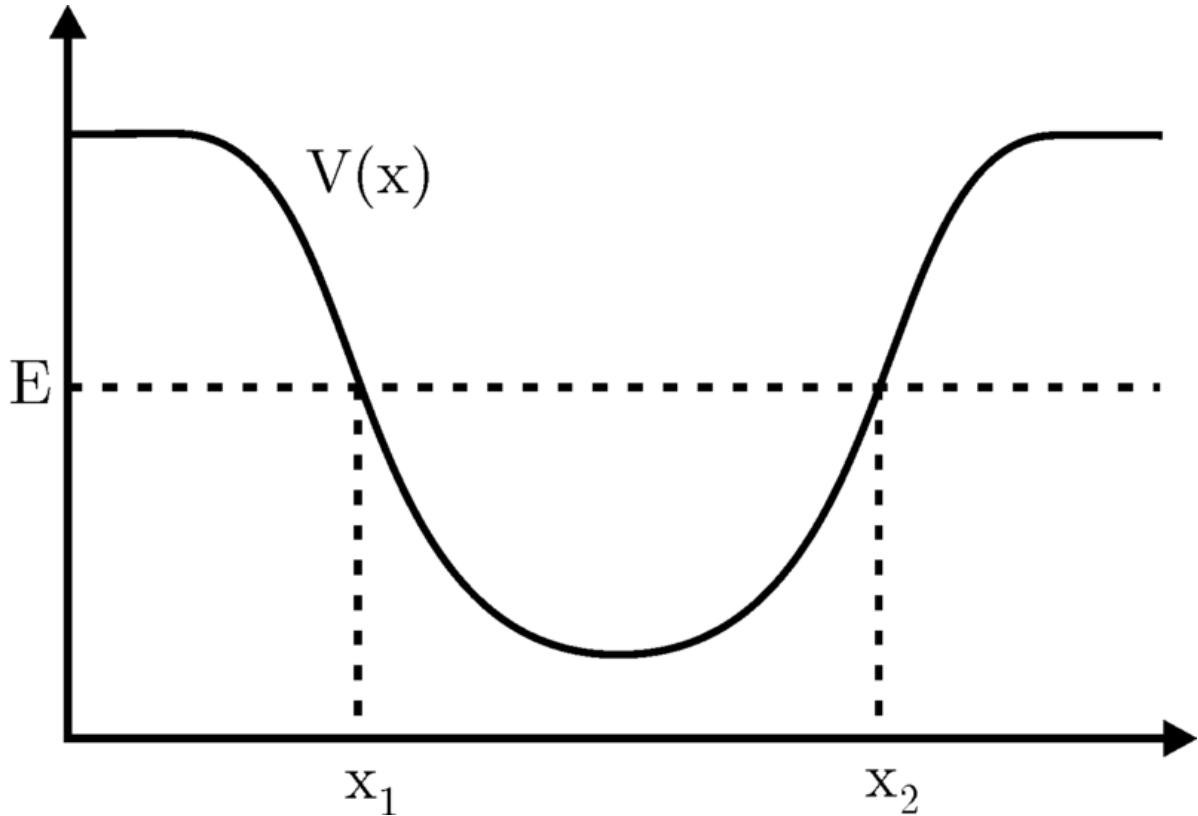


Figure 3.6: Potential well.

Schrödinger equation reads

$$\frac{d^2\Psi}{dz^2} - \frac{2m}{\hbar^2} [E - V(z)] \Psi = 0$$

Imagine there is a potential well between z_1 and z_2 shown in Figure 3.6: what are the allowable energy states?

In this case, $k^2 = -\frac{2m}{\hbar^2} [E - V(z)]$. Wave travelling to the right has positive $k(z)$, while wave travelling to the left has negative $k(z)$. Following the discussion of reflection coefficient,

$$R = ie^{-2i \int_0^{z_2} k(z') dz'} \\ R' = ie^{+2i \int_0^{z_1} k(z') dz'}$$

For waves bouncing back and forth inside the well, we must have

$$RR' = 1 \\ \Rightarrow (-1)e^{-2i \int_{z_1}^{z_2} k(z') dz'} = 1 \\ \Rightarrow \int_{z_1}^{z_2} k(z') dz' = \pi(n + \frac{1}{2}), \quad n = 0, 1, 2, 3, \dots$$

which is called the *Bohr-Sommerfield quantization rule*.

Another way to do this problem is by recognizing the Stokes and anti-Stokes lines in the complex z plane and match the solution in the whole domain. Figure needed!!! As the opposite in HW8.1%k2 .vs. z , Stokes line

$$k^2 = -\frac{2m}{\hbar^2} [E - V(z)] \equiv -p^2(z)(z - z_1)(z - z_2)$$

where for large z, we must choose a minus sign in front of $p(z)$ for $p(z)$ to be real and positive.
So

$$k = \pm ip(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

Then for a single-value solution, pick

$$-\pi \leq \arg(z - z_1) < \pi \\ -\frac{\pi}{2} \leq \arg(z - z_2) < \frac{3\pi}{2}$$

For z on AS_1 ,

$$k(z) = \pm ip(z)|z - z_1|^{1/2}e^{i\frac{1}{2}\cdot 0} \cdot |z - z_2|^{1/2}e^{i\frac{1}{2}\cdot \pi} = \pm i|k|e^{i\frac{1}{2}\pi} = \mp|k|$$

For a real and positive k (right-traveling wave), we must choose the lower sign +, so

$$k = -ip(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

For z on S_1 ,

$$k = -ip(z)|z - z_1|^{1/2}|z - z_2|^{1/2} = -i|k|$$

For z on S_4 ,

$$k = -ip(z)|z - z_1|^{1/2}e^{i\frac{1}{2}\pi}|z - z_2|^{1/2}e^{i\frac{1}{2}\pi} = i|k|$$

So

$$k(z) = \begin{cases} -i|k| & z > z_2 \\ |k| & z_1 < z < z_2 \\ i|k| & z < z_1 \end{cases} \quad (3.18)$$

On S_1 ,

$$\Psi_{E_{S_1}} \sim \frac{1}{\sqrt{k}} e^{\pm i \int_{z_2}^z k dz'} = \frac{1}{\sqrt{k}} e^{-i \int_{z_2}^z k dz'}$$

such that this is sub-dominant on S_1 ($k = -i|k|$).

Crossing S_2 into region I, we have

$$\Psi_I \sim \frac{1}{\sqrt{k}} \left\{ \underbrace{e^{-i \int_{z_2}^z k dz'}}_{\text{dominant on } S_2} + i \underbrace{e^{+i \int_{z_2}^z k dz'}}_{\text{subdominant on } S_2} \right\} \quad (3.19)$$

On S_4 ,

$$\Psi_{S_4} \sim \frac{1}{\sqrt{k}} e^{\pm i \int_{z_1}^z k dz'} \sim \frac{1}{\sqrt{k}} e^{-i \int_{z_1}^z k dz'},$$

where minus sign is chosen s.t. it is subdominant on S_4 .

Crossing S_3 into region I in the clockwise direction, we have

$$\begin{aligned}\Psi_I &\sim \left\{ \begin{array}{ll} \underbrace{e^{-i \int_{z_1}^z k dz'}}_{\text{dominant on } S_3} & -i \underbrace{e^{+i \int_{z_1}^z k dz'}}_{\text{subdominant on } S_3} \\ \end{array} \right\} \\ &= \frac{1}{\sqrt{k}} \left\{ e^{-i \int_{z_2}^z k dz'} \cdot e^{-i \int_{z_1}^{z_2} k dz'} - ie^{i \int_{z_2}^z k dz'} \cdot e^{i \int_{z_1}^{z_2} k dz'} \right\}\end{aligned}\quad (3.20)$$

Because the solution is analytic, Equation 3.19 and Equation 3.20 should be equal, therefore

$$\begin{aligned}\frac{1}{e^{-i \int_{z_1}^{z_2} k dz'}} &= \frac{i}{-ie^{i \int_{z_1}^{z_2} k dz'}} \\ \Rightarrow e^{2i \int_{z_1}^{z_2} k dz'} &= -1 = e^{i(\pi + 2n\pi)}, \quad n = 0, 1, 2, 3, \dots \\ \Rightarrow \int_{z_1}^{z_2} k dz' &= (\frac{1}{2} + n)\pi, \quad n = 0, 1, 2, 3, \dots\end{aligned}$$

Ex.4 Tunneling Problem

Instead of a potential well, now we consider another classical tunneling problem, where $E > V$ if $z_1 < z < z_2$.

$$k^2(z) = \pm p^2(z)(z - z_1)(z - z_2) = p^2(z)(z - z_1)(z - z_2)$$

where the plus sign is chosen s.t. $p(z)$ is real and positive for all real z , and then

$$k = p(z)(z - z_1)^{1/2}(z - z_2)^{1/2}$$

Define arguments as follows

$$0 \leq \arg z - z_2 < 2\pi, \quad -\pi < \arg z - z_1 \leq \pi$$

to remove the ambiguity in the branch in the expression of k .

Since $z = z_1, z = z_2$ are simple turning points (with different slope in linear approximation), the Stokes and anti-Stokes lines are shown in Figure needed!!!

The branch cut defined as above give the following values of k on the real z axis:

$$k(z) = \begin{cases} |k| & z > z_2 \\ i|k| & z_1 < z < z_2 \\ -|k| & z < z_1 \end{cases} \quad (3.21)$$

On AS_1 , the solution

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i \int_{z_2}^z k dz'} \quad (3.22)$$

represents the outgoing wave propagating to $z = \infty$.

On S_2 , solution Equation 3.22 behaves like $E \sim e^{\int_b^z |k| dz'}$ which is dominant on S_2 with respect to $z = z_2$. Thus, on S_1 , it is subdominant.

On crossing S_1 from AS_1 , solution Equation 3.22 remains valid. In fact, it is valid in the region bounded by S_1 , S_2 and S_3 . Note that it is subdominant on S_2 with respect to $z = z_1$ (because we just showed that it is dominant on S_2 with respect to $z = z_2$). Thus, we rewrite the solution as

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i \int_{z_2}^{z_1} k dz'} . \underbrace{e^{i \int_{z_1}^z k dz'}}_{\text{subdominant on } S_2 \text{ w.r.t } z=a}$$

which becomes dominant on S_3 . Upon crossing S_3 , this solution becomes

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i \int_{z_2}^{z_1} k dz'} \left[\underbrace{e^{i \int_{z_1}^z k dz'}}_{\text{dominant on } S_3} + \underbrace{i e^{-i \int_{z_1}^z k dz'}}_{\text{subdominant on } S_3} \right]$$

where we pick up the Stokes constant i and the subdominant solution, represented by the last term of the solution.

Now, this is the solution on AS_4 ($z < z_1$). Referring to Equation 3.21, we see that the first term of the solution represents the reflected wave, and the second term the incident wave from $z = -\infty$. Since the incident wave and reflected wave have equal amplitude, we have

$$|R|^2 = 1$$

The transmitted wave has unit amplitude (see Eq.(3.22)). The incident wave has amplitude $e^{i \int_a^b k dz'}(i)$. Thus,

$$|T|^2 = \frac{1}{|e^{i \int_{z_1}^{z_2} k dz'}(i)|^2} = \frac{1}{|e^{i \int_{z_1}^{z_2} i|k| dz'}(i)|^2} = e^{-2 \int_{z_1}^{z_2} |k| dz}$$

where z_1, z_2 are far apart, i.e. $|T|^2$ is exponentially small.

Note: $|R|$ and $|T|$ above are not valid if z_1, z_2 are close to each other. The correct transmission coefficient, for general values of z_1, z_2 is

$$|T|^2 = \frac{e^{-2 \int_{z_1}^{z_2} |k| dz}}{1 + e^{-2 \int_{z_1}^{z_2} |k| dz}}$$

and the reflection coefficient is

$$|R|^2 = 1 - |T|^2$$

Ex.5 Two waves: one is launched from $z = +\infty$, incident onto a resonant layer at $z = a$ ($a > 0$), and the other launched from $z = -\infty$, incident onto a absorption layer at $z = 0$, with the model equation

$$\frac{d^2 E(z)}{dz^2} + k_0^2 \left(\frac{z}{z-a} \right) E(z) = 0$$

This looks like a similar case for ECRH.

$$k^2(z) = k_0^2 \left(\frac{z}{z-a} \right) \Rightarrow k(z) = k_0 z^{1/2} (z-a)^{-1/2}$$

where k_0 is real and positive.

Define arguments as follows

$$-\frac{3\pi}{2} \leq \arg z < \frac{\pi}{2}, \quad -\frac{3\pi}{2} < \arg z - a \leq \frac{\pi}{2}$$

to remove the ambiguity in the branch in the expression of k .

For z on AS_1 ,

$$k = k_0 |z|^{1/2} e^{-\frac{1}{2}0} |z-a|^{-1/2} e^{i\frac{1}{2}0} = |k|$$

For z on S_1 ,

$$k = k_0 |z|^{1/2} e^{-\frac{1}{2}0} |z-a|^{-1/2} e^{i\frac{1}{2}\pi} = i|k|$$

For z on AS_2 ,

$$k = k_0 |z|^{1/2} e^{i\frac{1}{2}(-\pi)} |z-a|^{-1/2} e^{i\frac{-1}{2}(-\pi)} = |k|$$

Then the branch cut defined as above give the following values of k on the real z axis:

$$k(z) = \begin{cases} |k| & z > a \\ i|k| & 0 < z < a \\ |k| & z < 0 \end{cases} \quad (3.23)$$

For wave launched from large magnetic field side $z = +\infty$ going to $z = -\infty$ on AS_3 , the solution on AS_3 is (Note that even though z is negative, it is the upper limit of the integral and k is positive. It took me a long time to get the right sign here for a left propagating wave.)

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i \int_0^z k dz'}$$

This solution behaves like

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i \int_0^z i|k| dz'} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + \int_0^z |k| dz'} \quad (3.24)$$

on S_1 . Thus, it is dominant on S_1 with respect to $z = 0$; it is subdominant on the neighboring Stokes line S_2 . Thus solution Equation 3.24 remains valid upon crossing Stokes line S_2 . It is the solution in region I, bounded by S_1 and S_2 . Since it is dominant on S_1 with respect to $z = 0$, it is subdominant on S_1 with respect to $z = a$. So we can write it as

$$E \sim \frac{1}{\sqrt{k}} e^{-i\omega t - i \int_0^a k dz'} . \underbrace{e^{-i \int_a^z k dz'}}_{\text{subdominant on } S_1 \text{ w.r.t. } z=a}$$

which is also the incident wave from $z = 0$. This is the solution in region I as well as on AS_1 ($k = |k|$). So it is the incident wave solution from $z = \infty$. The reflection coefficient must be 0!

The transmission coefficient is then

$$\begin{aligned} |T|^2 &= \left| \frac{E(z = -\infty)}{E(z = \infty)} \right|^2 = \frac{1}{|e^{-i \int_0^a k dz}|^2} \\ &= e^{-2 \int_0^a |k| dz} = e^{-2 \int_0^a k_0 \sqrt{\frac{z}{a-z}} dz} \end{aligned}$$

Let $z = a \sin^2 \theta$, we can finally have

$$|T|^2 = e^{-k_0 a \pi}$$

The fraction of power absorbed by resonance is

$$1 - |R|^2 - |T|^2 = 1 - e^{-k_0 a \pi}$$

Note: these expressions of $|T|^2$, $|R|^2$ are valid even if a is small.

For wave launched from low magnetic field side $z = -\infty$,

$$E_{AS_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i \int_a^z k dz'}$$

On S_1 , a subdominant (decaying) solution with respect to $z = 0$ is

$$E_{S_1} \sim \frac{1}{\sqrt{k}} e^{-i\omega t + i \int_0^z i|k| dz'}$$

In regions I bounded by S_1 , S_2 and AS_1 ,

$$E_I \sim E_{S_1} \frac{1}{\sqrt{k}} e^{i \int_a^0 k dz'} \cdot e^{i \int_0^z k dz'}$$

E_{S_1} is dominant on S_1 with respect to $z = a$, but is subdominant on S_1 with respect to $z = 0$.

Crossing S_2 into AS_2 ,

$$E_{AS_2} \sim \frac{1}{\sqrt{k}} e^{-i\omega t} \cdot e^{i \int_a^0 k dz'} \left[\underbrace{e^{i \int_0^z k dz'}}_{\text{dominant on } S_3} - i \underbrace{e^{-i \int_0^z k dz'}}_{\text{subdominant on } S_3} \right],$$

where the first term represents the reflected wave, and the second term represents the incident wave.

The transmission coefficient is

$$\begin{aligned} T &= \left| \frac{E(z = \infty)}{E(z = -\infty)} \right| \\ &= \left| \frac{\sqrt{k(-\infty)}}{\sqrt{k(\infty)}} \right| \left\| \frac{e^{i \int_a^\infty k dz}}{e^{i \int_a^0 k dz} \cdot e^{i \int_0^{-\infty} k dz}} \right\| \\ &= \frac{1}{\left| e^{i \int_a^0 i|k| dz} \right|} = e^{-\int_0^a |k| dz} \end{aligned}$$

Then the transmission coefficient in power is

$$T^2 = e^{-2 \int_0^a k \sqrt{|\frac{z}{z-a}|} dz'} = e^{-k_0 a \pi}$$

where the last equivalence requires some transformation tricks. If $a \sim \lambda$, $k_0 = 2\pi/\lambda$, the transmitted power fraction is on the order of $e^{-20} \sim 10^{-9}$.

The reflection coefficient is zero and that the transmission coefficient (in power) is $T = e^{-k_0 a \pi}$. Thus, the fraction $1 - T$ of the incident power is absorbed by that resonant layer.

3.11.5 “Exact” WKBJ Solution

- Shielding or tunneling

If you have a high density region between x_1 and x_2 ,

$$k^2(x) = \frac{\omega^2}{c^2} \left[1 - \frac{\omega_p^2(x)}{\omega^2} \right]$$

The transmission coefficient in power is

$$|T|^2 = \frac{e^{-2 \int_{x_1}^{x_2} |k| dx}}{1 + e^{-2 \int_{x_1}^{x_2} |k| dx}}$$

and the reflection coefficient in power is

$$|R|^2 = \frac{1}{1 + e^{-2 \int_{x_1}^{x_2} |k| dx}}$$

Note:

- $|R|^2 + |T|^2 = 1$
- When $x_1 = x_2$, i.e. $\omega^2 = \omega_p^2$ (peak), $|R|^2 = |T|^2 = \frac{1}{2}$.
- ECRH

$$\frac{d^2 E}{dz^2} + k_0^2 \left(\frac{z}{z-a} \right) E = 0$$

where

$$k^2(z) = k_0^2 \left(\frac{z}{z-a} \right)$$

(I) Incident wave from $z = -\infty$ (from low B_0 side)

$$\begin{aligned}|R| &= 1 - e^{-\pi k_0 a} \\ |T| &= e^{-\frac{1}{2}\pi k_0 a} \\ |R|^2 + |T|^2 &= 1 - e^{-\pi k_0 a} + e^{-2\pi k_0 a} < 1\end{aligned}$$

(II) Incident wave from $z = \infty$

$$\begin{aligned}|R| &= 0 \\ |T| &= e^{-\pi k_0 a/2} \\ |R|^2 + |T|^2 &= e^{-\pi k_0 a} < 1\end{aligned}$$

which means almost perfect absorption at $z = a$.

Note: These formulas are valid even if $a \rightarrow 0$, in which limit $|R| \rightarrow 0$, $|T| \rightarrow 1$, for both (I) and (II), as expected from the model equation.

Singularity in refractive index n can lead to absorption, even without collision.

3.12 Normalization

In physics we frequently have to deal with unit conversions. A solid understanding of the unit system and normalization methods paves the path for understanding the scales behind systems.

3.13 Wavelet Analysis

Wavelet Transform (WT) decomposes a function into a set of wavelets. A Wavelet is a wave-like oscillation that is localized in time. Two basic properties of a wavelet are *scale* and *location*.

Wavelet is a predecessor of Fourier Tranform (FT), which provides better results when dealing with changing background. It is the de-facto method for modern wave analysis. The key advantages of WT compared with FT are:

- Fewer “hard” parameters to tune in WT. For example, if you want to make a spectrogram with FT, you need to specify the size of local DFTs. In WT you do not need to worry about this; instead the validity of the result can be shown by the cone of influence. In a sense, FT makes immediate conversion from time to frequency domain, while WT let you choose the intermediate steps you wish for.
- Flexible forms of wavelets to choose from. In practice, if you have any prior knowledge to the signal you want to identify, you can find for an appropriate wavelet that is close to that shape, which gives better fitting compared to the sinuisoidal functions in FT.

Check the [notes from my blog](#) and the references therein.

4 Theoretical Mechanics

Newtonian mechanics is great, but theoretical mechanics is marvellous. It does not tell you anything new, but let you think of the same thing from a different prospective. Theoretical mechanics takes advantage of a system's constraints to solve problems. The constraints limit the *degrees of freedom* the system can have, and can be used to reduce the number of coordinates needed to solve for the motion. The formalism is well suited to arbitrary choices of coordinates, known in the context as *generalized coordinates*. The kinetic and potential energies of the system are expressed using these generalized coordinates or momenta, and the equations of motion can be readily set up, thus analytical mechanics allows numerous mechanical problems to be solved with greater efficiency than fully vectorial methods. *It does not always work for non-conservative forces or dissipative forces like friction, in which case one may revert to Newtonian mechanics.*

4.1 Generalized Coordinates

In Newtonian mechanics, one customarily uses all three Cartesian coordinates, or other 3D coordinate system, to refer to a body's position during its motion. In physical systems, however, some structure or other system usually constrains the body's motion from taking certain directions and pathways. So a full set of Cartesian coordinates is often unneeded, as the constraints determine the evolving relations among the coordinates, which relations can be modeled by equations corresponding to the constraints. In the Lagrangian and Hamiltonian formalisms, the constraints are incorporated into the motion's geometry, reducing the number of coordinates to the minimum needed to model the motion. These are known as generalized coordinates, denoted q_i ($i = 1, 2, 3\dots$).

4.2 D'Alembert's Principle

This principle is the analogy of Newton's second law in theoretical mechanics. It states that infinitesimal virtual work done by a force across reversible displacements is zero, which is the work done by a force consistent with ideal constraints of the system. The idea of a constraint is useful — since this limits what the system can do, and can provide steps to solving for the motion of the system. The equation for D'Alembert's principle is:

$$\delta W = \mathbf{Q} \cdot d\mathbf{q} = 0$$

where

$$\mathbf{Q} = (Q_1, Q_2, \dots, Q_N)$$

are the generalized forces (script Q instead of ordinary Q is used here to prevent conflict with canonical transformations below) and \mathbf{q} are the generalized coordinates. This leads to the generalized form of Newton's laws in the language of theoretical mechanics:

$$\mathbf{Q} = \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{\mathbf{q}}} \right) - \frac{\partial T}{\partial \mathbf{q}}$$

where T is the total kinetic energy of the system, and the notation

$$\frac{\partial}{\partial \mathbf{q}} = \left(\frac{\partial}{\partial q_1}, \frac{\partial}{\partial q_2}, \dots, \frac{\partial}{\partial q_N} \right)$$

4.3 Lagrangian Mechanics

Lagrangian and Euler–Lagrange equations

The introduction of generalized coordinates and the fundamental Lagrangian function:

$$L(\mathbf{q}, \dot{\mathbf{q}}, t) = T(\mathbf{q}, \dot{\mathbf{q}}, t) - V(\mathbf{q}, \dot{\mathbf{q}}, t)$$

where T is the total kinetic energy and V is the total potential energy of the entire system, then either following the calculus of variations or using the above formula — lead to the Euler-Lagrange equations;

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{q}}} \right) = \frac{\partial L}{\partial \mathbf{q}}$$

which are a set of N second-order ordinary differential equations, one for each $q_i(t)$. This formulation identifies the actual path followed by the motion as a selection of the path over which the time integral of kinetic energy is least, assuming the total energy to be fixed, and imposing no conditions on the time of transit.

4.4 Hamiltonian Mechanics

Hamiltonian and Hamilton's equations

The Legendre transformation of the Lagrangian replaces the generalized coordinates and velocities $(\mathbf{q}, \dot{\mathbf{q}})$ with (\mathbf{q}, \mathbf{p}) ; the generalized coordinates and the generalized momenta conjugate to the generalized coordinates:

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = \left(\frac{\partial L}{\partial \dot{q}_1}, \frac{\partial L}{\partial \dot{q}_2}, \dots, \frac{\partial L}{\partial \dot{q}_N} \right) = (p_1, p_2, \dots, p_N)$$

and introduces the Hamiltonian (which is in terms of generalized coordinates and momenta):

$$H(\mathbf{q}, \mathbf{p}, t) = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, t)$$

This leads to the Hamilton's equations:

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}$$

which are now a set of $2N$ first-order ordinary differential equations, one for each $q_i(t)$ and $p_i(t)$. Another result from the Legendre transformation relates the time derivatives of the Lagrangian and Hamiltonian:

$$\frac{dH}{dt} = -\frac{\partial L}{\partial t}$$

which is often considered one of Hamilton's equations of motion additionally to the others. The generalized momenta can be written in terms of the generalized forces in the same way as Newton's second law:

$$\dot{\mathbf{p}} = \mathbf{Q}$$

One interesting application of the Hamilton's equation in plasma physics is the prove of Vlasov equation (See Giulia's notes)

$$\frac{df(\mathbf{q}, \mathbf{p}, t)}{dt} = 0$$

You may also find the application in deriving the gyrokinetic equations.

4.5 Hamilton-Lagrange Formalism of Lorentz Equation

Two mathematically equivalent formalisms describe charged particle dynamics, namely

1. the Lorentz equation

$$m \frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

2. Hamiltonian-Lagrangian dynamics.

The two formalisms are complementary: the Lorentz equation is intuitive and suitable for approximate methods, whereas the more abstract Hamiltonian-Lagrangian formalism exploits time and space symmetries. A brief review of the Hamiltonian-Lagrangian formalism follows, emphasizing aspects relevant to dynamics of charged particles.

The starting point is to postulate the existence of a function L , called the Lagrangian, which:

1. contains *all* information about the particle dynamics for a given situation;

2. depends only on generalized coordinates $Q_i(t), \dot{Q}_i(t)$ appropriate to the problem;
3. possibly has an explicit dependence on time t .

If such a function $L(Q_i(t), \dot{Q}_i(t), t)$ exists, then information on particle dynamics is retrieved by manipulation of the *action integral*

$$S = \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t) dt$$

This manipulation is based on d'Alembert's principle of least action. According to this principle, one considers the infinity of possible trajectories a particle could follow to get from its initial position $Q_i(t_1)$ to its final position $Q_i(t_2)$, and postulates that the trajectory actually followed is the one that results in the lowest value of S . Thus, the value of S must be minimized (note that S here is action, and not entropy). Minimizing S does not give the actual trajectory directly, but rather gives equations of motion, which can be solved to give the actual trajectory.

Minimizing S is accomplished by considering an arbitrary nearby alternative trajectory $Q_i(t) + \delta Q_i(t)$ having the same beginning and end points as the actual trajectory, i.e., $\delta Q_i(t_1) = Q_i(t_2) = 0$. In order to make the variational argument more precise, δQ_i is expressed as

$$\delta Q_i(t) = \epsilon \eta_i(t)$$

where ϵ is an arbitrarily adjustable scalar assumed to be small so that $\epsilon^2 < \epsilon$ and $\eta_i(t)$ is a function of t that vanishes when $t = t_1$ or $t = t_2$ but is otherwise arbitrary. Calculating δS to second order in ϵ gives

$$\begin{aligned} \delta S &= \int_{t_1}^{t_2} L(Q_i + \delta Q_i, \dot{Q}_i + \delta \dot{Q}_i, t) dt - \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t) dt \\ &= \int_{t_1}^{t_2} L(Q_i + \epsilon \eta_i, \dot{Q}_i + \epsilon \dot{\eta}_i, t) dt - \int_{t_1}^{t_2} L(Q_i(t), \dot{Q}_i(t), t) dt \\ &= \int_{t_1}^{t_2} \left(\epsilon \eta_i \frac{\partial L}{\partial Q_i} + \frac{(\epsilon \eta_i)^2}{2} \frac{\partial^2 L}{\partial Q_i^2} + \epsilon \dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} + \frac{(\epsilon \dot{\eta}_i)^2}{2} \frac{\partial^2 L}{\partial \dot{Q}_i^2} \right) dt \end{aligned}$$

Suppose the trajectory $Q_i(t)$ is the one that minimizes S . Any other trajectory must lead to a higher value of S and so δS must be positive for any finite value of ϵ . If ϵ is chosen to be sufficiently small, then the absolute values of the terms of order ϵ^2 above will be smaller than the absolute values of the terms linear in ϵ . The sign of ϵ could then be chosen to make δS negative, but this would violate the requirement that δS must be positive. The only way out of this dilemma is to insist that the sum of the terms linear in ϵ vanishes so $\delta S \sim \epsilon^2$ and is therefore always positive as required. Insisting that the sum of terms linear in ϵ vanishes implies

$$0 = \int_{t_1}^{t_2} \left(\eta_i \frac{\partial L}{\partial Q_i} + \dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} \right) dt$$

Using $\dot{\eta}_i = d\eta_i/dt$ the above expression may be integrated by parts to obtain

$$\begin{aligned} 0 &= \int_{t_1}^{t_2} \left(\eta_i \frac{\partial L}{\partial Q_i} + \dot{\eta}_i \frac{\partial L}{\partial \dot{Q}_i} \right) dt \\ &= \left[\eta_i \frac{\partial L}{\partial \dot{Q}_i} \right]_{t_1}^{t_2} + \int_{t_1}^{t_2} \left[\eta_i \frac{\partial L}{\partial Q_i} - \eta_i \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{Q}_i} \right) \right] dt \end{aligned}$$

Since $\eta_i(t_{1,2}) = 0$, the integrated term vanishes and since η_i was an arbitrary function of t , the coefficient of i in the integrand must vanish, yielding Lagrange's equation

$$\frac{dP_i}{dt} = \frac{\partial L}{\partial Q_i} \quad (4.1)$$

where the *canonical momentum* P_i is defined as

$$P_i \equiv \frac{\partial L}{\partial \dot{Q}_i} \quad (4.2)$$

Lagrange's equation shows that if L does not depend on a particular generalized coordinate Q_j , then $dP_j/dt = 0$, in which case the canonical momentum P_j is a *constant of motion*; the coordinate Q_j is called a *cyclic* or *ignorable* coordinate. This is a very powerful and profound result. Saying that the Lagrangian function does not depend on a coordinate is equivalent to saying that the system is *symmetric* in that coordinate or translationally invariant with respect to that coordinate. The quantities P_j and Q_j are called conjugate and action has the dimensions of the product of these quantities.

Hamilton extended this formalism by introducing a new function related to the Lagrangian. This new function, called the Hamiltonian, provides further useful information and is defined as

$$H \equiv \left(\sum_i P_i \dot{Q}_i \right) - L \quad (4.3)$$

Partial derivative of H with respect to P_i and to Q_i give Hamilton's equations

$$\dot{Q}_i = \frac{\partial H}{\partial P_i}, \quad \dot{P}_i = -\frac{\partial H}{\partial Q_i}$$

which are equations of motion having a close relation to phase-space concepts. The time derivative of the Hamiltonian is

$$\frac{dH}{dt} = \sum_i \frac{dP_i}{dt} \dot{Q}_i + \sum_i P_i \frac{d\dot{Q}_i}{dt} - \left(\sum_i \frac{\partial L}{\partial Q_i} \dot{Q}_i + \sum_i \frac{\partial L}{\partial \dot{Q}_i} \frac{d\dot{Q}_i}{dt} + \frac{\partial L}{\partial t} \right) = -\frac{\partial L}{\partial t} \quad (4.4)$$

This shows that if L does not explicitly depend on time, the Hamiltonian H is a *constant of the motion*. As will be shown later, H corresponds to the energy of the system, so if $\partial L/\partial t = 0$, the energy is a constant of the motion. Thus, energy is conjugate to time in analogy to canonical momentum being conjugate to position (note that energy \times time also has the units of action). If the Lagrangian does not explicitly depend on time, then the system can be thought of as being “symmetric” with respect to time, or “translationally” invariant with respect to time.

The Lagrangian for a charged particle in an electromagnetic field is

$$L = \frac{mv^2}{2} + q\mathbf{v} \cdot \mathbf{A}(\mathbf{x}, t) - q\phi(\mathbf{x}, t) \quad (4.5)$$

The validity of Equation 4.5 will now be established by showing that it generates the Lorentz equation when inserted into Lagrange's equation. Since no symmetry is assumed, there is no reason to use any special coordinate system and so ordinary Cartesian coordinates will be used as the canonical coordinates, in which case Equation 4.2 gives the canonical momentum as

$$\mathbf{P} = mv\mathbf{\hat{v}} + q\mathbf{A}(\mathbf{x}, t)$$

The left-hand side of Equation 4.1 becomes

$$\frac{d\mathbf{P}}{dt} = m\frac{d\mathbf{v}}{dt} + q\left(\frac{\partial \mathbf{A}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{A}\right)$$

while the right-hand side of Equation 4.1 becomes

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{x}} &= q\nabla(\mathbf{v} \cdot \mathbf{A}) - q\nabla\phi = q(\mathbf{v} \cdot \nabla \mathbf{A} + \mathbf{v} \times \nabla \times \mathbf{A}) - q\nabla\phi \\ &= q(\mathbf{v} \cdot \nabla \mathbf{A} + \mathbf{v} \times \mathbf{B}) - q\nabla\phi \end{aligned}$$

Equating the above two expressions gives the Lorentz equation, where the electric field is defined as $\mathbf{E} = -\partial \mathbf{A}/\partial t - \nabla\phi$ in accordance with Faraday's law. This proves that Equation 4.5 is mathematically equivalent to the Lorentz equation when used with the principle of least action.

The Hamiltonian associated with this Lagrangian is, in Cartesian coordinates,

$$\begin{aligned} H &= \mathbf{P} \cdot \mathbf{v} - L = \frac{mv^2}{2} + q\phi \\ &= \frac{(\mathbf{P} - q\mathbf{A}(\mathbf{x}, t))^2}{2m} + q\phi(\mathbf{x}, t) \end{aligned} \quad (4.6)$$

where the last line is the form more suitable for use with Hamilton's equations, i.e., $H = H(\mathbf{x}, \mathbf{P}, t)$. Equation 4.6 also shows that H is, as promised, the particle energy. If generalized coordinates are used, the energy can be written in a general form as $E = H(Q, P, t)$. Equation 4.4 showed that even though both Q and P depend on time, the energy depends on time

only if H explicitly depends on time. Thus, in a situation where H does not explicitly depend on time, the energy would have the form $E = H(Q(t), P(t)) = \text{const.}$

It is important to realize that both canonical momentum and energy depend on the reference frame. For example, a bullet fired in an airplane in the direction opposite to the airplane motion, and with a speed equal to the airplane's speed, has a large energy as measured in the airplane frame, but zero energy as measured by an observer on the ground. A more subtle example (of importance to later analysis of waves and Landau damping) occurs when \mathbf{A} and/or ϕ has a wave-like dependence, e.g., $\phi(\mathbf{x}, t) = \phi(\mathbf{x} - \mathbf{v}_{\text{ph}}t)$, where \mathbf{v}_{ph} is the wave phase velocity. This potential is time-dependent in the lab frame and so the associated Lagrangian has an explicit dependence on time in the lab frame, which implies that *energy is not a constant of the motion in the lab frame*. In contrast, ϕ is time-independent in the wave frame and so the energy is a *constant of the motion* in the wave frame. Existence of a constant of the motion reduces the complexity of the system of equations and typically makes it possible to integrate at least one equation in closed form. Thus, it is advantageous to analyze the system in the frame having the most constants of the motion.

5 Relativity

5.1 Velocity

The path of an object in three-dimensional space (in an inertial frame) may be expressed in terms of three spatial coordinate functions $x^i(t)$ of time t , where i is an index which takes values 1, 2, 3.

The three coordinates form the 3d position vector, written as a column vector

$$\mathbf{x}(t) = \begin{bmatrix} x^1(t) \\ x^2(t) \\ x^3(t) \end{bmatrix}$$

The components of the velocity \mathbf{u} (tangent to the curve) at any point on the world line are

$$\mathbf{u} = \begin{bmatrix} u^1 \\ u^2 \\ u^3 \end{bmatrix} = \frac{d\mathbf{x}}{dt} = \begin{bmatrix} \frac{dx^1}{dt} \\ \frac{dx^2}{dt} \\ \frac{dx^3}{dt} \end{bmatrix} \quad (5.1)$$

Each component is simply written $u^i = dx^i/dt$.

5.2 Theory of Relativity

In Einstein's theory of relativity, the path of an object moving relative to a particular frame of reference is defined by four coordinate functions $x_\mu(\tau)$, where μ is a spacetime index which takes the value 0 for the timelike component, and 1, 2, 3 for the spacelike coordinates. The zeroth component is defined as the time coordinate multiplied by c ,

$$x^0 = ct \quad (5.2)$$

Each function depends on one parameter called its *proper time*. As a column vector,

$$\mathbf{x} = \begin{bmatrix} x^0(\tau) \\ x^1(\tau) \\ x^2(\tau) \\ x^3(\tau) \end{bmatrix}$$

Time dilation

From time dilation, the differentials in coordinate time t and proper time τ are related by

$$dt = \gamma(u)d\tau \quad (5.3)$$

where the Lorentz factor,

$$\gamma(u) = \frac{1}{\sqrt{1 - \frac{u^2}{c^2}}} \quad (5.4)$$

is a function of the Euclidean norm u of the 3d velocity vector \mathbf{u} :

$$u = \| \mathbf{u} \| = \sqrt{(u^1)^2 + (u^2)^2 + (u^3)^2}$$

5.2.1 Four-velocity

A *four-velocity* is a *four-vector* in four-dimensional spacetime that represents the relativistic counterpart of velocity, which is a three-dimensional vector in space.

Physical events correspond to mathematical points in time and space, the set of all of them together forming a mathematical model of physical four-dimensional spacetime. The history of an object traces a curve in spacetime, called its world line. If the object has mass, so that its speed is necessarily less than the speed of light, the world line may be parametrized by the proper time of the object. The four-velocity is the rate of change of four-position with respect to the proper time along the curve. The velocity, in contrast, is the rate of change of the position in (three-dimensional) space of the object, as seen by an observer, with respect to the observer's time.

The four-velocity is the tangent four-vector of a timelike world line. The four-velocity \mathbf{U} at any point of world line $\mathbf{X}(\tau)$ is defined as

$$\mathbf{U} = \frac{d\mathbf{X}}{d\tau} \quad (5.5)$$

where \mathbf{X} is the four-position and τ is the proper time.¹

The relationship between the time t and the coordinate time x^0 is defined by

$$x^0 = ct \quad (5.6)$$

Taking the derivative of this with respect to the proper time τ , we find the U_μ velocity component for $\mu = 0$:

$$U^0 = \frac{dx^0}{d\tau} = \frac{d(ct)}{d\tau} = c \frac{dt}{d\tau} = c\gamma(u) \quad (5.7)$$

¹The four-velocity defined here using the proper time of an object does not exist for world lines for massless objects such as photons travelling at the speed of light.

and for the other 3 components to proper time we get the U_μ velocity component for $\mu = 1, 2, 3$:

$$U^i = \frac{dx^i}{d\tau} = \frac{dx^i}{dt} \frac{dt}{d\tau} = \frac{dx^i}{dt} \gamma(u) = \gamma(u) u^i \quad (5.8)$$

where we have used the chain rule from Equation 5.1 and Equation 5.3.

Thus, we find for the four-velocity \mathbf{U} :

$$\mathbf{U} = \gamma \begin{bmatrix} c \\ \mathbf{u} \end{bmatrix}$$

Written in standard four-vector notation this is

$$\mathbf{U} = \gamma(c, \mathbf{u}) = (\gamma c, \gamma \mathbf{u}) \quad (5.9)$$

where γc is the temporal component and $\gamma \mathbf{u}$ is the spatial component. Unlike most other four-vectors, the four-velocity has only 3 independent components u_x, u_y, u_z instead of 4. The γ factor is a function of the three-dimensional velocity \mathbf{u} .

Using the differential of the four-position in the rest frame, the magnitude of the four-velocity can be obtained by the Minkowski metric with signature $(-, +, +, +)$:

$$\|\mathbf{U}\|^2 = \eta_{\mu\nu} U^\mu U^\nu = \eta_{\mu\nu} \frac{dX^\mu}{d\tau} \frac{dX^\nu}{d\tau} = -c^2 \quad (5.10)$$

which is always a fixed constant.

6 Unit Conversion

6.1 Formula Conversion

Part of the materials in this chapter are referenced from the [NRL Plasma Formulary](#) and notes by Dana Longcope.

To derive a dimensionally correct SI formula from one expressed in Gaussian (CGS) units, substitute for each quantity according to $\bar{Q} = \bar{k}Q$, where \bar{k} is the coefficient in the second column of Table 6.1 corresponding to Q (overbars denote variables expressed in Gaussian units). To go from SI to natural units in which $\hbar = c = 1$, use $Q = \hat{k}^{-1}\hat{Q}$, where \hat{k} is the coefficient corresponding to the third column.

Here

$$\begin{aligned}\alpha &= 10^2 \text{ cm m}^{-1} \\ \beta &= 10^7 \text{ erg J}^{-1} \\ \epsilon_0 &= 8.8542 \times 10^{-12} \text{ F m}^{-1} \\ \mu_0 &= 4\pi \times 10^7 \text{ H m}^{-1} \\ c &= (\epsilon_0 \mu_0)^{-1/2} = 2.9979 \times 10^8 \text{ m s}^{-1} \\ \hbar &= 1.0546 \times 10^{-34} \text{ J s}\end{aligned}$$

Table 6.1: Formula conversion coefficients

Physical Quantity	CGS Units to SI	Natural Units to SI
Capacitance	$\alpha/4\pi\epsilon_0$	ϵ_0^{-1}
Charge	$(\alpha\beta/4\pi\epsilon_0)^{1/2}$	$(\epsilon_0\hbar c)^{-1/2}$
Charge density	$(\beta/4\pi\alpha^5\epsilon_0)^{1/2}$	$(\epsilon_0\hbar c)^{-1/2}$
Current	$(\alpha\beta/4\pi\epsilon_0)^{1/2}$	$(\mu_0/\hbar c)^{1/2}$
Current density	$(\beta/4\pi\alpha^3\epsilon_0)^{1/2}$	$(\mu_0/\hbar c)^{1/2}$
Electric field	$(4\pi\beta\epsilon_0/\alpha^3)^{1/2}$	$(\epsilon_0/\hbar c)^{1/2}$
Electric potential	$(4\pi\beta\epsilon_0/\alpha)^{1/2}$	$(\epsilon_0/\hbar c)^{1/2}$
Electric conductivity	$(4\pi\epsilon_0)^{-1}$	ϵ_0^{-1}
Energy	β	$(\hbar c)^{-1}$
Energy density	β/α^3	$(\hbar c)^{-1}$
Force	β/α	$(\hbar c)^{-1}$

Physical Quantity	CGS Units to SI	Natural Units to SI
Frequency	1	c^{-1}
Inductance	$4\pi\epsilon_0/\alpha$	μ_0^{-1}
Length	α	1
Magnetic induction	$(4\pi\beta/\alpha^3\mu_0)^{1/2}$	$(\mu_0\hbar c)^{-1/2}$
Magnetic intensity	$(4\pi\mu_0\beta/\alpha^3)^{1/2}$	$(\mu_0\hbar c)^{-1/2}$
Mass	β/α^2	c/\hbar
Momentum	β/α	\hbar^{-1}
Power	β	$(\hbar c^2)^{-1}$
Pressure	β/α^3	$(\hbar c)^{-1}$
Resistance	$4\pi\epsilon_0/\alpha$	$(\epsilon_0/\mu_0)^{1/2}$
Time	1	c
Velocity	α	c^{-1}

- Bohr radius in CGS: $\bar{a}_0 = \bar{\hbar}^2/\bar{m}\bar{e}^2$

In SI: $\alpha a_0 = (\hbar\beta)^2/[(m\beta/\alpha^2)(e^2\alpha\beta/4\pi\epsilon_0)]$, or $a_0 = \epsilon_0 h^2/\pi me$.

6.1.1 Maxwell's Equations

- Faraday's law
 - SI: $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$
 - CGS: $\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$
- Ampère's law
 - SI: $\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$
 - CGS: $\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c} \mathbf{J}$
- Poisson equation
 - SI: $\nabla \cdot \mathbf{D} = \rho$
 - CGS: $\nabla \cdot \mathbf{D} = 4\pi\rho$
- Absence of magnetic monopoles
 - SI: $\nabla \cdot \mathbf{B} = 0$
 - CGS: $\nabla \cdot \mathbf{B} = 0$
- Lorentz force on charge q
 - SI: $q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$
 - CGS: $q(\mathbf{E} + \frac{1}{c}\mathbf{v} \times \mathbf{B})$
- Constitutive relations

- SI & CGS: $\mathbf{D} = \epsilon \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H}$

The electromagnetic energy in volume V is given by

$$\begin{aligned} W &= \frac{1}{2} \int_V dV (\mathbf{H} \cdot \mathbf{B} + \mathbf{E} \cdot \mathbf{D}) \quad \text{SI} \\ &= \frac{1}{8\pi} \int_V dV (\mathbf{H} \cdot \mathbf{B} + \mathbf{E} \cdot \mathbf{D}) \quad \text{CGS} \end{aligned} \quad (6.1)$$

Poynting's theorem is

$$\frac{\partial W}{\partial t} + \int_S \mathbf{N} \cdot d\mathbf{S} = - \int_V dV \mathbf{J} \cdot \mathbf{E} \quad (6.2)$$

where S is the closed surface bounding V and the Poynting vector (energy flux across S) is given by

$$\begin{aligned} \mathbf{N} &= \mathbf{E} \times \mathbf{H} \quad (\text{SI}) \\ &= c \mathbf{E} \times \mathbf{H} / 4\pi \quad (\text{CGS}) \end{aligned} \quad (6.3)$$

EM formulas can be converted to the corresponding formula from SI to CGS using the substitutions in the left column of Table 6.2. The first three are fundamental relations between the unit systems. The next three are obtained using the equations listed, and the last three are related variables.

Table 6.2: EM formula conversion

Physical Quantity	SI	CGS
Magnetic field	\mathbf{B}	\mathbf{B}
Vacuum permeability	μ_0	4π
Vacuum permittivity	ϵ_0	$1/4\pi c^2$ ¹
Electric field	\mathbf{E}	$c\mathbf{E}$
Current density	\mathbf{J}	\mathbf{J}/c
Charge density	ρ	ρ/c
Charge	q	q/c
Current	I	I/c
Resistivity	η	$c^2 \eta$

Some common expression conversions are listed in Table 6.3.

¹From the unit conversion we can see that modifying the speed of light in numerical simulations only modify ϵ_0 , but not μ_0 .

Table 6.3: EM formula conversion examples

SI	CGS
$q\mathbf{v} \times \mathbf{B}$	$q\frac{\mathbf{v}}{c} \times \mathbf{B}$
$\mathbf{J} \times \mathbf{B}$	$\frac{1}{c}\mathbf{J} \times \mathbf{B}$
$\frac{q_1 q_2}{4\pi\epsilon_0}$	$q_1 q_2$
$\frac{B}{\sqrt{\mu_0\rho}}$	$\frac{B}{\sqrt{4\pi\rho}}$
$\frac{1}{2\mu_0}B^2$	$\frac{1}{8\pi}B^2$
$\frac{\epsilon_0}{2}E^2$	$\frac{1}{8\pi}E^2$
$\omega_p^2 = \frac{q^2 n}{\epsilon_0 m}$	$\omega_p^2 = \frac{4\pi q^2 n}{m}$
$\Omega_c = \frac{qB}{m}$	$\Omega_c = \frac{qB}{mc}$

6.2 Unit System Conversion

A rule of thumb is to check the number of constants in the system: making one constant equal to 1 requires to shrink the number of basic variables by 1. In ideal MHD Equation 8.49, there are 4 basic quantities (length, mass, time, and current density) and 1 physical constant μ_0 . If we include temperature, then there will be 5 basic quantities and 2 physical constants μ_0 , R . This indicates that we need 3 reference quantities in order to make the 2 physical constants equal to 1 in the normalized units. For example, in the Earth's magnetosphere, we select a reference length, e.g. $l_0 = 1 R_E$, a reference mass density, e.g. $\rho_0 = 1.67 \times 10^{-17} \text{ kg} \cdot \text{m}^{-3}$, and a reference magnetic field, e.g. $B_0 = 3.12 \times 10^{-5} \text{ T}$ (Earth's magnetic field strength at the equator). All the rest conversion factors can be derived from these together with the physical constants μ_0 , R , m_i expressed SI units. (R is needed for temperature and m_i is needed for number density.) Inserting the initially chosen values, we get a full set of conversion factors from variables in normalized units n' , B' , u' , p' , T' , \mathcal{E}' to SI units n , B , u , p , T , \mathcal{E} :

$$U_{\text{SI}} = U' * U_0$$

where each conversion factor is summarized in Table 6.4. Note that in the definition of number density, the denominator is m_i instead of m_0 and $R = 2k_B/m_i$ only appears in temperature. This is because mass does not appear in ideal MHD: for a given ρ , the same results are obtained for a heavy species with small number density and a light species with large number density, e.g. $\rho = 2m_i n_i = m_i 2n_i$. In principle we can choose the reference mass arbitrarily, but here in order to make the values in SI units "look good" we choose m_i . If you need to compare with a hybrid model, an easy way to make self-consistent conversion is to make mass $m_0 = m_i$ as a substitute for any of the basic variables. In this case, the dimensionless $\rho = n$. Also note that in Table 6.4 the MHD pressure and temperature are the sum of electron's and ion's pressure and temperature, which is the reason a factor of "2" pops out in the definition of temperature. If we are comparing to a hybrid model which assumes massless electrons, a proper modification would be $p_0 = \rho_0 v_0^2/2$ and $T_0 = p_0 * m_i/(k_B \rho_0) = m_i v_0^2/k_B$.

Table 6.4: MHD unit conversion.

Table 6.5: Example basic quantities in MHD

Basic variable	Notation	Definition	Value
Length	l_0	$l_0 = R_E$	$6.371 \times 10^6 \text{ m}$
Magnetic field	B_0	$B_0 = B_E$	$3.12 \times 10^{-5} \text{ T}$
Mass density	ρ_0	$\rho_0 = \rho_E$	$1.67 \times 10^{-17} \text{ kg} \cdot \text{m}^{-3}$

Table 6.6: Example derived quantities in MHD

Derived variable	Notation	Definition	Value
Mass	m_0	$m_0 = \rho_0 l_0^3$	$4.32 \times 10^3 \text{ kg}$
Velocity	v_0	$v_0 = B_0 / \sqrt{\mu_0 \rho_0}$	$6.81 \times 10^6 \text{ m} \cdot \text{s}^{-1}$
Time	t_0	$t_0 = l_0 / v_0$	0.94 s
Number density	n_0	$n_0 = \rho_0 / m_i$	10^{10} m^{-3}
Pressure	p_0	$p_0 = \rho_0 v_0^2$	$7.75 \times 10^{-4} \text{ N} \cdot \text{m}^{-2}$
Temperature	T_0	$T_0 = m_i v_0^2 / (2k_B)$	$2.81 \times 10^9 \text{ K}$
Energy density	\mathcal{E}_0	$\mathcal{E}_0 = p_0$	$7.75 \times 10^{-4} \text{ J} \cdot \text{m}^{-3}$

6.2.1 Example: $\rho, \rho\mathbf{u}, \mathcal{E}, \mathbf{B}$ to n, u, T

When dealing with velocity distribution functions, one easy way is to write $f(\mathbf{v}) = f(\mathbf{v}; n, \mathbf{u}, T)$. For instance, a 3D Maxwellian is given as

$$f(\mathbf{v}) = n \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{m(\mathbf{v} - \mathbf{u})^2}{2k_B T} \right)$$

From dimensionless conserved variables $(\rho, \rho\mathbf{u}, \mathcal{E}, \mathbf{B})$ to SI units $(n, \mathbf{u}, T, \mathbf{B})$, we have

$$\begin{aligned} n_{\text{SI}} &= n_0 n = \rho_{\text{SI}} / m_i = \frac{\rho_0}{m_i} \rho \\ u_{\text{SI}} &= u_0 \left(\frac{\rho u}{\rho} \right) \\ T_{\text{SI}} &= T_0 \frac{(\gamma - 1)}{\rho} \left[\mathcal{E} - \frac{1}{2} \rho u^2 - \frac{1}{2} B^2 \right] \\ B_{\text{SI}} &= B_0 B \end{aligned}$$

7 Single-Particle Motions

What makes plasmas particularly difficult to analyze is the fact that the densities fall in an intermediate range. Fluids like water are so dense that the motions of individual molecules do not have to be considered. Collisions dominate, and the simple equations of ordinary fluid dynamics suffice. At the other extreme in very low-density devices, only single-particle trajectories need to be considered; collective effects are often unimportant. Plasma behaves sometimes like fluids, and sometimes like a collection of individual particles. The first step in learning how to deal with this schizophrenic personality is to understand how single particles behave in electric and magnetic fields.

Single particle motion in neutral gases is trivial – particles move in straight lines until they hit other particles or the wall. Because of this simplicity, there is no point in keeping track of the details of single particle motion in a neutral gas and instead a statistical averaging of this motion suffices; this averaging shows that neutral gases have Maxwellian velocity distributions and are in a local thermodynamic equilibrium. In contrast, plasma particles are nearly collisionless and typically have complex trajectories that are strongly affected by both electric and magnetic fields.

As what will be shown in Section 12.2.5, the velocity distribution in a plasma will become Maxwellian when enough collisions have occurred to maximize the entropy. However, since collisions occur infrequently in hot plasmas, many important phenomena have time scales shorter than the time required for the plasma velocity distribution to become Maxwellian. A collisionless model is thus required to characterize these fast phenomena. In these situations randomization does not occur, entropy is conserved, the distribution function need not be Maxwellian, and the plasma is not in thermodynamic equilibrium. Thermodynamic concepts therefore do not apply, and the plasma is instead characterized by concepts from classical mechanics such as momentum or energy conservation of individual particles. In these collisionless situations the complex details of single particle dynamics are not washed out by collisions but instead persist and influence the macroscopic scale. As an example, the cyclotron resonance of a single particle can be important at the macroscopic scale in a collisionless plasma. This chapter examines various aspects of single particle motion and shows how these aspects can influence the macroscopic properties of a plasma.

Furthermore, study of single particle dynamics has a very direct relevance to Vlasov theory because, as shown in Section 12.3.1, any function constructed from constants of single particle motion is a valid solution of the collisionless Vlasov equation. Thus, knowledge of single particle

dynamics provides a “repertoire” of constants of the motions from which solutions to the Vlasov equation suitable for various situations can be constructed.

Finally, the study of single particle motion develops valuable intuition regarding wave–particle interactions and identifies certain unusual situations, such as stochastic or non-adiabatic particle motion, that are beyond the descriptive capability of fluid models.

Here we assume that the EM fields are prescribed and not affected by the charged particles. The associated approach is often called *test particle*. Alternatively, we can say that it is possible to deduce intuitive and quite accurate analytic solutions for the drift of charged particles in arbitrarily complicated electric and magnetic fields provided the field are *slowly changing in both space and time* (this requirement is essentially the slowness requirement for adiabatic invariance introduced in Section 7.8). Drift solutions are obtained by solving the Lorentz equation

$$m \frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (7.1)$$

iteratively, taking advantage of the assumed separation of scales between fast and slow motions.

When relativity is considered, the Lorentz equation becomes

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (7.2)$$

where $\mathbf{p} = \gamma m\mathbf{v}$ is the relativistic momentum and $\gamma = (1 - v^2/c^2)^{-1/2}$ the *Lorentz factor*. The rest mass of an electron is $m_e = 511 \text{ keV}/c^2$ and of a proton $m_p = 938 \text{ MeV}/c^2$.

7.1 Uniform E and B Fields

The convention here is that velocity \mathbf{v} represents single particle velocity as you will see through most of this chapter.

7.1.1 E=0

In this case, a charged particle has a simple cyclotron gyration. The equation of motion is

$$m \frac{d\mathbf{v}}{dt} = q\mathbf{v} \times \mathbf{B}$$

Taking \hat{z} to be the direction of \mathbf{B} ($\mathbf{B} = B\hat{z}$), we have

$$\begin{aligned} m\dot{v}_x &= qBv_y, & m\dot{v}_y &= -qBv_x, & m\dot{v}_z &= 0, \\ \ddot{v}_x &= \frac{qB}{m}\dot{v}_y = -\left(\frac{qB}{m}\right)^2 v_x \\ \ddot{v}_y &= \frac{qB}{m}\dot{v}_x = -\left(\frac{qB}{m}\right)^2 v_y \end{aligned}$$

This describes a simple harmonic oscillator at the *cyclotron frequency*, which we define to be

$$\omega_c \equiv \frac{|q|B}{m}$$

By the convention we have chosen, ω_c is always nonnegative. The solution of velocity is then

$$v_{x,y} = v_\perp \exp(\pm i\omega_c t + i\delta_{x,y})$$

The \pm denote the sign of q . We may choose the phase δ so that

$$v_x = v_\perp e^{i\omega_c t} = \dot{x}$$

where v_\perp is a positive constant denoting the speed in the plane perpendicular to \mathbf{B} . Then

$$v_y = \frac{m}{qB} \dot{v}_x = \pm \frac{1}{\omega_c} \dot{v}_x = \pm i v_\perp e^{i\omega_c t} = \dot{y}$$

Integrating once again, we have

$$\begin{aligned} x - x_0 &= -i \frac{v_\perp}{\omega_c} e^{i\omega_c t} \\ y - y_0 &= \pm i \frac{v_\perp}{\omega_c} e^{i\omega_c t} \end{aligned}$$

We define the *Larmor radius* to be

$$r_L \equiv \frac{v_\perp}{\omega_c} = \frac{mv_\perp}{|q|B} \quad (7.3)$$

Note that 2π does not appear in Equation 7.3!

Taking the real part of the positions, we have

$$\begin{aligned} x - x_0 &= r_L \sin \omega_c t \\ y - y_0 &= \pm r_L \cos \omega_c t \end{aligned}$$

This describes a circular orbit about a *guiding center* (x_0, y_0) which is fixed. The direction of the gyration is always such that the magnetic field generated by the charged particle is opposite to the externally imposed field. Plasma particles, therefore, tend to *reduce* the magnetic field, and plasmas are *diamagnetic*. In addition to this motion, there is an arbitrary velocity v_z along \mathbf{B} which is not affected by \mathbf{B} . The trajectory of a charged particle in space is, in general, a helix. Looking along (against) the magnetic field, the particle rotating clockwise (anticlockwise) has a negative charge. In plasma physics this is the convention of *right-handedness*.

The unit of ω_c in SI units is rad/s. To convert it to hertz, we need to divide it by 2π . The corresponding *oscillation frequencies* $f_{c\alpha} = \omega_{c\alpha}/(2\pi)$ of electrons and protons are

$$\begin{aligned} f_{ce} [\text{Hz}] &= 28B[\text{nT}] \\ f_{cp} [\text{Hz}] &= 0.015B[\text{nT}] \end{aligned}$$

7.1.2 Finite \mathbf{E}

If now we allow an electric field to be present, the motion will be found to be the sum of two motions: the usual circular Larmor gyration plus a drift of the guiding center. The assumed spatial uniformity and time-independence of the fields represent the extreme limit of assuming that the fields are slowly changing in space and time. We may choose \mathbf{E} to lie in the x-z plane so that $E_y = 0$. As before, the z component of velocity is unrelated to the transverse components and can be treated separately. The equations of motion is now

$$m\dot{\mathbf{v}} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

whose z component is

$$\dot{v}_z = \frac{q}{m} E_z$$

or

$$v_z = \frac{qE_z}{m} t + v_{z0}$$

This is a straightforward acceleration along \mathbf{B} . The transverse components are

$$\begin{aligned}\dot{v}_x &= \frac{q}{m} E_x \pm \omega_c v_y \\ \dot{v}_y &= 0 \mp \omega_c v_x\end{aligned}$$

Differentiating, we have (for constant \mathbf{E})

$$\begin{aligned}\ddot{v}_x &= -\omega_c^2 v_x \\ \ddot{v}_y &= \mp \omega_c \left(\frac{q}{m} E_x \pm \omega_c v_y \right) = -\omega_c^2 \left(v_y + \frac{E_x}{B} \right)\end{aligned}$$

We can write this as

$$\frac{d^2}{dt^2} \left(v_y + \frac{E_x}{B} \right) = -\omega_c^2 \left(v_y + \frac{E_x}{B} \right)$$

so that it reduces to the previous case if we replace v_y by $v_y + (E_x/B)$. The velocity solution is then replaced by

$$\begin{aligned}v_x &= v_\perp e^{i\omega_c t} \\ v_y &= \pm v_\perp e^{i\omega_c t} - \frac{E_x}{B}\end{aligned}$$

The Larmor motion is the same as before, but there is superimposed a drift \mathbf{v}_{gc} of the guiding center in the $-y$ direction (for $E_x > 0$).

To obtain a general formula for \mathbf{v}_{gc} , we can solve the momentum equation in vector form. We may omit the $m\mathbf{d}\mathbf{v}/dt$ term, since this term gives only the circular motion at ω_c , which we already know about. Then the momentum equation becomes

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0$$

Taking the cross product with \mathbf{B} , we have

$$\mathbf{E} \times \mathbf{B} = \mathbf{B} \times (\mathbf{v} \times \mathbf{B}) = \mathbf{v}B^2 - \mathbf{B}(\mathbf{v} \cdot \mathbf{B})$$

The transverse components with respect to \mathbf{B} of this equation are

$$\mathbf{v}_{gc} = \mathbf{E} \times \mathbf{B}/B^2 \equiv \mathbf{v}_E$$

We define this to be \mathbf{v}_E , the electric field drift of the guiding center. In magnitude, this drift is

$$v_E = \frac{E(\text{V/m})}{B(\text{tesla})} \frac{\text{m}}{\text{sec}}$$

It is important to note that \mathbf{v}_E is independent of q , m , and v_\perp . The reason is obvious from the following physical picture. In the first half-cycle of the ion's orbit, it gains energy from the electric field and increases in v_\perp and, hence, in r_L . In the second half-cycle, it loses energy and decreases in r_L . This difference in r_L on the left and right sides of the orbit causes the drift v_E . A negative electron gyrates in the opposite direction but also gains energy in the opposite direction; it ends up drifting in the same direction as an ion. For particles of the same velocity but different mass, the lighter one will have smaller r_L and hence drift less per cycle. However, its gyration frequency is also larger, and the two effects exactly cancel. Two particles of the same mass but different energy would have the same ω_c . The slower one will have smaller r_L and hence gain less energy from \mathbf{E} in a half-cycle. However, for less energetic particles the fractional change in r_L for a given change in energy is larger, and these two effects cancel. Another way of interpreting this behavior is to recall that according to the theory of special relativity the electric field \mathbf{E}' observed in a frame moving with velocity \mathbf{u} is $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$ and so \mathbf{v}_E is simply a statement that a particle drifts in such a way to ensure that the electric field seen in its own frame vanishes.

The three-dimensional orbit in space for a charged particle with a nonzero initial velocity is therefore a slanted helix with changing pitch. The pitch angle is defined as

$$\alpha = \arctan \frac{v_\perp}{v_\parallel} = \arcsin \frac{v_\perp}{v} = \arccos \frac{v_\parallel}{v} \quad (7.4)$$

7.1.3 Gravitational Field

The foregoing result can be applied to other forces by replacing $q\mathbf{E}$ in the equation of motion by a general force \mathbf{F} . The guiding center drift caused by \mathbf{F} is then

$$\mathbf{v}_f = \frac{1}{q} \frac{\mathbf{F} \times \mathbf{B}}{B^2}$$

In particular, if \mathbf{F} is the force of gravity $m\mathbf{g}$, there is a drift

$$\mathbf{v}_g = \frac{m}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2} \quad (7.5)$$

This is similar to the drift \mathbf{v}_E in that it is perpendicular to both the force and \mathbf{B} , but it differs in one important respect. The drift \mathbf{v}_g changes sign with the particle's charge. Under a gravitational force, ions and electrons drift in opposite directions, so there is a net current density in the plasma given by

$$\mathbf{j} = n(m_i + m_e) \frac{\mathbf{g} \times \mathbf{B}}{B^2}$$

The physical reason for this drift is again the change in Larmor radius as the particle gains and loses energy in the gravitational field. Now the electrons gyrate in the opposite sense to the ions, but the force on them is in the same direction, so the drift is in the opposite direction. The magnitude of \mathbf{v}_g is usually negligible, but when the lines of force (i.e. magnetic field lines) are curved, there is an effective gravitational force due to centrifugal force. This force, which is not negligible, is independent of mass; this is why we did not stress the m dependence of the drift here. Centrifugal force is the basis of a plasma instability called the "gravitational" instability, which has nothing to do with real gravity.

And now it shall be clear that the $\mathbf{E} \times \mathbf{B}$ drift analysis can be easily generalized to describe the effect on a charged particle of any force orthogonal to \mathbf{B} by simply making the replacement $\mathbf{E} \rightarrow \mathbf{F}/q$ in the Lorentz equation. Thus, any spatially uniform, temporally constant force orthogonal to \mathbf{B} will cause a drift

$$\mathbf{v}_F = \frac{\mathbf{F} \times \mathbf{B}}{qB^2}$$

which leads to two counter-intuitive and important conclusions:

1. A steady-state electric field perpendicular to a magnetic field does not drive currents in a plasma, but instead causes a bulk motion of the entire plasma across the magnetic field with the velocity \mathbf{v}_E .
2. A steady-state force (e.g., gravity, centrifugal force, etc.) perpendicular to the magnetic field causes oppositely directed motions for electrons and ions and so drives a cross-field current

$$\mathbf{j}_F = \sum_{\sigma} n_{\sigma} \frac{\mathbf{F} \times \mathbf{B}}{B^2}$$

7.2 Nonuniform \mathbf{B} Field

Now that the concept of a guiding center drift is firmly established, we can discuss the motion of particles in inhomogeneous fields — \mathbf{E} and \mathbf{B} fields which vary in space or time. For uniform fields we were able to obtain exact expressions for the guiding center drifts. As soon as we

introduce inhomogeneity, the problem becomes too complicated to solve exactly. To get an approximate answer, it is customary to expand in the small ratio r_L/L , where L is the scale length of the inhomogeneity. This type of theory, called *orbit theory*, can become extremely involved. We shall examine only the simplest cases, where only one inhomogeneity occurs at a time.

7.2.1 B B: Grad-B Drift

Here the magnetic field lines are straight, but their density increases, say, in the y direction. We can anticipate the result by using our simple physical picture. The gradient in $|B|$ causes the Larmor radius to be larger at the bottom of the orbit than at the top, and this should lead to a drift, in opposite directions for ions and electrons, perpendicular to both \mathbf{B} and ∇B . The drift velocity should obviously be proportional to r_L/L and to v_\perp .

Consider the Lorentz force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$, averaged over a gyration. Clearly, $\bar{F}_x = 0$, since the particle spends as much time moving up as down. We wish to calculate \bar{F}_y , in an approximate fashion, by using the *undisturbed orbit* of the particle to find the average. The undisturbed orbit is given by the solution in the first section for a uniform \mathbf{B} field. Taking the real part of the solution for v_x and y , we have

$$F_y = -qv_x B_z(y) = -qv_\perp (\cos \omega_c t) \left[B_0 \pm r_L (\cos \omega_c t) \frac{\partial B}{\partial y} \right]$$

where we have made a Taylor expansion of \mathbf{B} field about the point $x_0 = 0, y_0 = 0$

$$\begin{aligned} \mathbf{B} &= \mathbf{B}_0 + (\mathbf{r} \cdot \nabla) \mathbf{B} + \dots \\ Bz &= B_0 + y(\partial B_z / \partial y) + \dots \end{aligned}$$

This expansion of course requires $r_L/L \ll 1$, where L is the length scale of $\partial B_z / \partial y$. The first term above averages to zero in a gyration, and the average of $\cos^2 \omega_c t$ is 1/2, so that

$$\bar{F}_y = \mp qv_\perp r_L \frac{1}{2} \frac{\partial B}{\partial y}$$

The guiding center drift velocity is then

$$\mathbf{v}_{gc} = \frac{1}{q} \frac{\mathbf{F} \times \mathbf{B}}{B^2} = \frac{1}{q} \frac{\bar{F}_y}{|B|} \hat{x} = \mp \frac{v_\perp r_L}{B} \frac{1}{2} \frac{\partial B}{\partial y} \hat{x}$$

where we have used the formula shown previously. Since the choice of the y axis was arbitrary, this can be generalized to

$$\mathbf{v}_{\nabla B} = \pm \frac{1}{2} v_\perp r_L \frac{\mathbf{B} \times \nabla B}{B^2}$$

This has all the dependences we expected from the physical picture; only the factor $\frac{1}{2}$ (arising from the averaging) was not predicted. Note that the \pm stands for the sign of the charge, and lightface B stands for $|B|$. The quantity $\mathbf{v}_{\nabla B}$ is called the *grad-B drift*; it is in opposite directions for ions and electrons and causes a current transverse to \mathbf{B} . An exact calculation of $\mathbf{v}_{\nabla B}$ would require using the exact orbit, including the drift, in the averaging process.

7.2.2 Curved B: Curvature Drift

Here we assume the magnetic field lines to be curved with a constant radius of curvature R_c , and we take $|B|$ to be constant. Such a field does not obey Maxwell's equations in a vacuum, so in practice the grad-B drift will always be added to the effect derived here. A guiding center drift arises from the centrifugal force felt by the particles as they move along the field lines in their thermal motion. If v_{\parallel}^2 denotes the average square of the component of random velocity along \mathbf{B} , and \mathbf{R}_c denotes the radius of curvature vector from the center to the curve, the average centrifugal force is

$$\mathbf{F}_{cf} = \frac{mv_{\parallel}^2}{R_c} \hat{r} = mv_{\parallel}^2 \frac{\mathbf{R}_c}{R_c^2}$$

According to the guiding center drift formula, this gives rise to a drift

$$\mathbf{v}_R = \frac{1}{q} \frac{\mathbf{F}_{cf} \times \mathbf{B}}{B^2} = \frac{mv_{\parallel}^2}{qB^2} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2}$$

The drift \mathbf{v}_c is called the *curvature drift*.

We must now compute the grad-B drift which accompanies this when the decrease of $|B|$ with radius is taken into account. In a vacuum, we have $\nabla \times \mathbf{B} = 0$ (current-free). In the cylindrical coordinates, $\nabla \times \mathbf{B}$ only has a z component, since \mathbf{B} has only a θ component and ∇B only an r component. We then have

$$(\nabla \times \mathbf{B})_z = \frac{1}{r} \frac{\partial}{\partial r} (r B_{\theta}) = 0, \quad B \propto \frac{1}{r}$$

Thus

$$|B| \propto \frac{1}{R_c}, \quad \frac{\nabla B}{B} = -\frac{\mathbf{R}_c}{R_c^2} \quad (7.6)$$

Using the expression of the grad-B drift, we have

$$\mathbf{v}_{\nabla B} = \mp \frac{1}{2} \frac{v_{\perp} r_L}{B^2} \mathbf{B} \times \frac{\mathbf{R}_c}{R_c^2} = \pm \frac{1}{2} \frac{v_{\perp}^2}{\omega_c} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B} = \frac{1}{2} \frac{m}{q} v_{\perp}^2 \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2}$$

Adding this to \mathbf{v}_c , we have the total drift in a curved vacuum field

$$\mathbf{v}_c + \mathbf{v}_{\nabla B} = \frac{m}{q} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2} \left(v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2 \right)$$

or with just the magnetic field

$$\mathbf{v}_c + \mathbf{v}_{\nabla B} = \frac{m}{q} \frac{\mathbf{B} \times \nabla B}{B^3} \left(v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2 \right) \quad (7.7)$$

or if the kinetic energy W and pitch angle θ are involved

$$\mathbf{v}_c + \mathbf{v}_{\nabla B} = \frac{\mathbf{B} \times \nabla B}{qB^3} \left(2W \cos^2 \theta + W \sin^2 \theta \right) = \frac{\mathbf{B} \times \nabla B}{qB^3} (1 + W \cos^2 \theta) \quad (7.8)$$

It is unfortunate that these drifts add. This means that if one bends a magnetic field into a torus for the purpose of confining a thermonuclear plasma, the particles will drift out of the torus no matter how one juggles the temperatures and magnetic fields.

For a Maxwellian distribution, v_{\parallel}^2 and $\frac{1}{2}v_{\perp}^2$ are each equal to $k_B T/m$, since v_{\perp} involves two degrees of freedom. Then the average curved-field drift can be written as

$$\bar{\mathbf{v}}_{c+\nabla B} = \pm \frac{v_{th}^2}{R_c \omega_c} \hat{y} = \pm \frac{\bar{r}_L}{R_c} v_{th} \hat{y}$$

where \hat{y} here is the direction of $\widehat{\mathbf{R}}_c \times \mathbf{B}$. This shows that $\bar{\mathbf{v}}_{c+\nabla B}$ depends on the charge of the species but not on its mass.

If current exists and we cannot assume $\nabla \times \mathbf{B} = 0$, we rely on the original definition of curvature radius

$$\frac{\hat{r}}{R_c} = \frac{\mathbf{R}_c}{R_c^2} = -\hat{b} \cdot \nabla \hat{b} \quad (7.9)$$

Then the curvature drift can be written as

$$\mathbf{v}_R = \frac{mv_{\parallel}^2}{qB^2} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2} = \frac{mv_{\parallel}^2}{qB^4} \mathbf{B} \times [\mathbf{B} \cdot \nabla \mathbf{B}] \quad (7.10)$$

The equivalence of Equation 7.7 and Equation 7.10 can be verified by using the equality¹

$$\begin{aligned} \mathbf{B} \times (\nabla \times \mathbf{B}) &= (\nabla \mathbf{B}) \cdot \mathbf{B} - (\mathbf{B} \cdot \nabla) \mathbf{B} \\ (\mathbf{B} \cdot \nabla) \mathbf{B} &= (\nabla \mathbf{B}) \cdot \mathbf{B} = B \nabla B \end{aligned}$$

¹The last equality can be checked by writing out each component of \mathbf{B} .

7.2.3 B B: Magnetic Mirrors

Now we consider a magnetic field which is pointed primarily in the z direction and whose magnitude varies in the z direction. Let the field be axisymmetric, with $B_\theta = 0$ and $\partial/\partial\theta = 0$. Since the magnetic field lines converge and diverge, there is necessarily a component B_r . We wish to show that this gives rise to a force which can trap a particle in a magnetic field.

We can obtain B_r from $\nabla \cdot \mathbf{B} = 0$:

$$\frac{1}{r} \frac{\partial}{\partial r} (r B_r) + \frac{\partial B_z}{\partial z} = 0$$

If $\partial\mathbf{B}_z/\partial z$ is given at $r = 0$ and does not vary much with r , we have approximately

$$r B_r = - \int_0^r r \frac{\partial B_z}{\partial z} dr \simeq -\frac{1}{2} r^2 \left[\frac{\partial \mathbf{B}_z}{\partial z} \right]_{r=0}$$

$$B_r = -\frac{1}{2} r \left[\frac{\partial \mathbf{B}_z}{\partial z} \right]_{r=0}$$

The variation of $|B|$ with r causes a grad-B drift of guiding centers about the axis of symmetry, but there is no radial grad-B drift, because $\partial B/\partial\theta = 0$. The components of the Lorentz force are

$$F_r = q \underbrace{(v_\theta B_z - v_z B_\theta)}_1$$

$$F_\theta = q \underbrace{(-v_r B_z + v_z B_r)}_2$$

$$F_z = q(v_r B_\theta - \underbrace{v_\theta B_r}_4)$$

Two terms vanish if $B_\theta = 0$, and terms 1 and 2 give rise to the usual Larmor gyration. Term 3 vanishes on the axis; when it does not vanish, this azimuthal force causes a drift in the radial direction. This drift merely makes the guiding centers follow the magnetic field lines. Term 4 is the one we are interested in. Using the expression of B_r , we have

$$F_z = \frac{1}{2} q v_\theta r_L \frac{\partial B_z}{\partial z}$$

We must now average over one gyration. For simplicity, consider a particle whose guiding center lies on the axis. Then v_θ is a constant during a gyration; depending on the sign of q , v_θ is $\mp v_\perp$. Since $r = r_L$, the average force is

$$\bar{F}_z = \mp \frac{1}{2} q v_\perp r_L \frac{\partial B_z}{\partial z} = \mp \frac{1}{2} q \frac{v_\perp^2}{\omega_c} \frac{\partial B_z}{\partial z} = -\frac{1}{2} \frac{m v_\perp^2}{B} \frac{\partial B_z}{\partial z}$$

We define the *magnetic moment* of the gyrating particle to be

$$\mu \equiv \frac{1}{2}mv_{\perp}^2/B$$

so that

$$\bar{F}_z = -\mu \frac{\partial B_z}{\partial z}$$

This is a specific example of the force on a diamagnetic particle, which in general can be written

$$\mathbf{F}_{\parallel} = -\mu \frac{\partial B}{\partial s} = -\mu \nabla_{\parallel} B$$

where ds is a line element along \mathbf{B} . Note that the definition of magnetic moment here is the same as the usual definition for the magnetic moment of a current loop with area A and current I : $\mu = IA$. In the case of a singly charged ion, I is generated by a charge e coming around $\omega_c/2\pi$ times a second: $I = e\omega_c/2\pi$. The area A is $\pi r_L^2 = \pi r_L^2/\omega_c^2$. Thus

$$\mu = \frac{e\omega_c}{2\pi} \frac{\pi r_L^2}{\omega_c^2} = \frac{1}{2} \frac{v_{\perp}^2 e}{\omega_c} = \frac{1}{2} \frac{mv_{\perp}^2}{B}$$

As the particle moves into regions of stronger or weaker B , its Larmor radius changes, but *remains invariant*. To prove this, consider the component of the equation of motion along \mathbf{B} :

$$m \frac{dv_{\parallel}}{dt} = -\mu \frac{\partial B}{\partial s}$$

Multiplying by v_{\parallel} on the left and its equivalent ds/dt on the right, we have

$$mv_{\parallel} \frac{dv_{\parallel}}{dt} = \frac{d}{dt} \left(\frac{1}{2} mv_{\parallel}^2 \right) = -\mu \frac{\partial B}{\partial s} \frac{ds}{dt} = -\mu \frac{dB}{dt}$$

Here dB/dt is the variation of B as seen by the particle; B itself is constant. The particle's energy must be conserved, so we have

$$\frac{d}{dt} \left(\frac{1}{2} mv_{\parallel}^2 + \frac{1}{2} mv_{\perp}^2 \right) = \frac{d}{dt} \left(\frac{1}{2} mv_{\parallel}^2 + \mu B \right) = 0$$

With the previous equation this becomes

$$-\mu \frac{dB}{dt} + \frac{d}{dt}(\mu B) = 0$$

so that

$$\frac{d\mu}{dt} = 0$$

The invariance of μ is the basis for one of the primary schemes for plasma confinement: the *magnetic mirror*. As a particle moves from a weak-field region to a strong-field region in the course of its thermal motion, it sees an increasing B , and therefore its v_{\perp} must increase in order to keep μ constant. Since its total energy must remain constant, v_{\parallel} must necessarily decrease. If B is high enough in the “throat” of the mirror, v_{\parallel} eventually becomes zero; and the particle is “reflected” back to the weak-field region. It is, of course, the force \mathbf{F}_{\parallel} which causes the reflection. The nonuniform field of a simple pair of coils form two magnetic mirrors between which a plasma can be trapped. This effect works on both ions and electrons.

The trapping is not perfect, however. For instance, a particle with $v_{\perp} = 0$ will have no magnetic moment and will not feel any force along \mathbf{B} . A particle with small v_{\perp}/v_{\parallel} at the midplane ($B = B_0$) will also escape if the maximum field B_m is not large enough. For given B_0 and B_m , which particles will escape? A particle with $v_{\perp} = v_{\perp 0}$ and $v_{\parallel} = v_{\parallel 0}$ at the midplane will have $v_{\perp} = v'_{\perp}$ and $v_{\parallel} = 0$ at its turning point. Let the field be B' there. Then the invariants of μ yields

$$\frac{1}{2} \frac{mv_{\perp 0}^2}{B_0} = \frac{1}{2} \frac{mv'_{\perp 0}^2}{B'}$$

Conservation of energy requires

$$v'_{\perp 0}^2 = v_{\perp 0}^2 + v_{\parallel 0}^2 \equiv v_0^2$$

Combining the above two equations, we find

$$\frac{B_0}{B'} = \frac{v_{\perp 0}^2}{v'_{\perp 0}^2} \equiv \sin^2 \theta$$

where θ is the pitch angle of the orbit in the weak-field region. Particles with smaller θ will mirror in regions of higher B . If θ is too small, B' exceeds B_m ; and the particles does not mirror at all. Replacing B' by B_m , we see that the smallest θ of a confined particle is given by

$$\sin^2 \theta_m = \frac{B_0}{B_m} \equiv \frac{1}{R_m}$$

where R_m is the *mirror ratio*. It defines the boundary of a region in velocity space in the shape of a cone, called a *loss cone*. Particles lying within the loss cone are not confined. Consequently, a mirror-confined plasma is never isotropic. Note that the loss cone is independent of q or m . Without collisions, both ions and electrons are equally well confined. When collisions occur, particles are lost when they change their pitch angle in a collision and are scattered into the loss cone. Generally, electrons are lost more easily because they have a higher collision frequency.

The magnetic mirror was first proposed by Enrico Fermi as a mechanism for the acceleration of cosmic rays. Protons bouncing between magnetic mirrors approaching each other at high velocity could gain energy at each bounce. How such mirrors could arise is another story. A further example of the mirror effect is the confinement of particles in the Van Allen belts. The magnetic field of the earth, being strong at the poles and weak at the equator, forms a natural mirror with rather large R_m .

7.3 Nonuniform E Field

Now we let the magnetic field be uniform and the electric field be nonuniform. For simplicity, we assume \mathbf{E} to be in the x direction and to vary sinusoidally in the x direction:

$$\mathbf{E} \equiv E_0(\cos kx)\hat{x}$$

This field distribution has a wavelength $\lambda = 2\pi/k$, and is the result of a sinusoidal distribution of charges, which we need not specify. In practice, such a charge distribution can arise in a plasma during a wave motion. The equation of motion is

$$m \frac{d\mathbf{v}}{dt} = q[\mathbf{E}(x) + \mathbf{v} \times \mathbf{B}]$$

whose transverse components are

$$\dot{v}_x = \frac{qB}{m}v_y + \frac{q}{m}E_x(x)$$

$$\dot{v}_y = -\frac{qB}{m}v_x$$

$$\ddot{v}_x = -\omega_c^2 v_x \pm \omega_c \frac{\dot{E}_x}{B}$$

$$\ddot{v}_y = -\omega_c^2 v_y - \omega_c^2 \frac{E_x(x)}{B}$$

Here $E_x(x)$ is the electric field at the position of the particle. To evaluate this, we need to know the particle's orbit, which we are trying to solve for in the first place. If the electric field is weak, we may, as an approximation, use the *undisturbed orbit* to evaluate $E_x(x)$. The orbit in the absence of the E field was given before

$$x = x_0 + r_L \sin \omega_c t$$

so we have

$$\ddot{v}_y = -\omega_c^2 v_y - \omega_c^2 \frac{E_0}{B} \cos k(x_0 + r_L \sin \omega_c t)$$

Anticipating the result, we look for a solution which is the sum of a gyration at ω_c and a steady drift \mathbf{v}_E . Since we are interested in finding an expression for v_E , we take out the gyroscopic motion by averaging over a cycle. The v_x component then gives $\bar{v}_x = 0$.² In the v_y component, the oscillating term \ddot{v}_y clearly averages to zero, and we have

²I am kind of lost here. Probably the logic here is that first we assume the solution has an additional steady drift, so we let $\ddot{v}_y = 0$? Check Y.Y's note.

$$\ddot{\bar{v}}_y = 0 = -\omega_c^2 \bar{v}_y - \omega_c^2 \frac{E_0}{B} \cos k(x_0 + \bar{r}_L \sin \omega_c t)$$

Expanding the cosine, we have

$$\cos k(x_0 + r_L \sin \omega_c t) = \cos(kx_0) \cos(kr_L \sin \omega_c t) - \sin(kx_0) \sin(kr_L \sin \omega_c t)$$

It will suffice to treat the small Larmor radius case, $kr_L \ll 1$. The Taylor expansions allow us to write

$$\cos k(x_0 + r_L \sin \omega_c t) \approx \cos(kx_0) \left(1 - \frac{1}{2}k^2 r_L^2 \sin^2 \omega_c t\right) - \sin(kx_0) kr_L \sin \omega_c t$$

The last term vanishes upon averaging over time, and it gives

$$\bar{v}_y = -\frac{E_0}{B} \cos(kx_0) \left(1 - \frac{1}{4}k^2 r_L^2\right) = -\frac{E_x(x_0)}{B} \left(1 - \frac{1}{4}k^2 r_L^2\right)$$

Thus the usual $\mathbf{E} \times \mathbf{B}$ drift is modified by the inhomogeneity to read

$$\mathbf{v}_E = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \left(1 - \frac{1}{4}k^2 r_L^2\right)$$

The physical reason for this is easy to see. An ion with its guiding center at a maximum of \mathbf{E} actually spends a good deal of its time in regions of weaker \mathbf{E} . Its average drift, therefore, is less than E/B evaluated at the guiding center. In a linearly varying \mathbf{E} field, the ion would be in a stronger field on one side of the orbit and in a field weaker by the same amount on the other side; the correction to \mathbf{v}_E then cancels out. From this it is clear that the correction term depends on the *second derivative* of \mathbf{E} . For the sinusoidal distribution we assumed, the second derivative is always negative with respect to \mathbf{E} . For an arbitrary variation of \mathbf{E} , we need only replace ik by ∇ and write the drift as

$$\mathbf{v}_E = \left(1 + \frac{1}{4}r_L^2 \nabla^2\right) \frac{\mathbf{E} \times \mathbf{B}}{B^2} \quad (7.11)$$

The second term is called the *finite-Larmor-radius effect*. What is the significance of this correction? Since r_L is much larger for ions than for electrons, \mathbf{v}_E is no longer independent of species. If a density clump occurs in a plasma, an electric field can cause the ions and electrons to separate, generating another electric field. If there is a feedback mechanism that causes the second electric field to enhance the first one, \mathbf{E} grows indefinitely, and the plasma is unstable. Such an instability, called a *drift instability*, is one type of plasma instabilities. The grad-B drift, of course, is also a finite-Larmor-radius effect and also causes charges to separate. However, $\mathbf{v}_{\nabla B} \propto kr_L$ whereas the correction term above is proportional to $k^2 r_L^2$. The nonuniform-E-field effect, therefore, is important at relatively large k , or small scale lengths of the inhomogeneity. For this reason, drift instabilities belong to a more general class called *microinstabilities*.

7.4 Time-Varying E Field

Let us now take \mathbf{E} and \mathbf{B} to be uniform in space but varying in time. First, consider the case in which \mathbf{E} alone varies sinusoidally in time, and let it lie along the x axis:

$$\mathbf{E} = E_0 e^{i\omega t} \hat{x}$$

Since $\dot{E}_x = i\omega E_x$, we can write the velocity x-component equation as

$$\ddot{v}_x = -\omega_c^2 \left(v_x \mp \frac{i\omega}{\omega_c} \frac{\tilde{E}_x}{B} \right)$$

Let us define

$$\begin{aligned}\tilde{v}_p &\equiv \pm \frac{i\omega}{\omega_c} \frac{\tilde{E}_x}{B} \\ \tilde{v}_E &\equiv \frac{\tilde{E}_x}{B}\end{aligned}$$

where the tilde has been added merely to emphasize that the drift is oscillating. The plus (minus) sign, as usual, denotes positive (negative) q . Now the x and y velocity component equations can be written as

$$\begin{aligned}\ddot{v}_x &= -\omega_c^2 (v_x - \tilde{v}_p) \\ \ddot{v}_y &= -\omega_c^2 (v_y - \tilde{v}_E)\end{aligned}$$

By analogy with the derivation in the case of a uniform EM field, we try a solution which is the sum of a drift and a gyroscopic motion:

$$\begin{aligned}v_x &= v_\perp e^{i\omega_c t} + \tilde{v}_p \\ v_y &= \pm i v_\perp e^{i\omega_c t} + \tilde{v}_E\end{aligned}$$

If we now differentiate twice with respect to time, we find

$$\begin{aligned}\ddot{v}_x &= -\omega_c^2 v_x + (\omega_c^2 - \omega^2) \tilde{v}_p \\ \ddot{v}_y &= -\omega_c^2 v_y + (\omega_c^2 - \omega^2) \tilde{v}_E\end{aligned}$$

This is **not** the same as the previous expressions right above unless $\omega^2 \ll \omega_c^2$. If we now make the assumption that \mathbf{E} varies slowly, so that $\omega^2 \ll \omega_c^2$, then there we have the approximate solution.

The solution of velocities in x and y tells us that the guiding center motion has two components. The y component, perpendicular to \mathbf{B} and \mathbf{E} , is the usual $\mathbf{E} \times \mathbf{B}$ drift, except that v_E now oscillates slowly at the frequency ω . The x component, a new drift *along the direction of E*, is

called the *polarization drift*. By replacing $i\omega$ with $\partial/\partial t$, we can generalize the expression of v_p and define the polarization drift as

$$\mathbf{v}_p = \pm \frac{1}{\omega_c B} \frac{d\mathbf{E}}{dt}$$

Since \mathbf{v}_p is in opposite directions for ions and electrons, there is a *polarization current*; for $Z = 1$, this is

$$\mathbf{j}_p = ne(v_{ip} - v_{ep}) = \frac{ne}{eB^2} (M + m) \frac{d\mathbf{E}}{dt} = \frac{\rho}{B^2} \frac{d\mathbf{E}}{dt} \quad (7.12)$$

where ρ is the mass density.

The physical reason for the polarization current is simple. Consider an ion at rest in a magnetic field. If a field \mathbf{E} is suddenly applied, the first thing the ion does is to move in the direction of \mathbf{E} . Only after picking up a velocity \mathbf{v} does the ion feel a Lorentz force $e\mathbf{v} \times \mathbf{B}$ and begin to move perpendicular to both fields. If \mathbf{E} is now kept constant, there is no further \mathbf{v}_p drift but only a \mathbf{v}_E drift. However, if \mathbf{E} is reversed, there is again a momentary drift, this time to the left. Thus \mathbf{v}_p is a startup drift due to inertia and occurs only in the first half-cycle of each gyration during which \mathbf{E} changes. Consequently, \mathbf{v}_p goes to zero when $\omega/\omega_c \ll 1$.

The polarization effect in a plasma is similar to that in a solid dielectric, where $\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$. The dipoles in a plasma are ions and electrons separated by a distance r_L . But since ions and electrons can move around to preserve quasineutrality, the application of a steady \mathbf{E} field does not result in a polarization field \mathbf{P} . However, if \mathbf{E} oscillates, an oscillating current \mathbf{j}_p results from the lag due to the ion inertia.

It is obvious that in this case the $\mathbf{E} \times \mathbf{B}$ drift speed depends on time and the guiding center coordinate system is non-inertial. The polarization drift arises from the fictitious (inertial) force $-md\mathbf{v}_E/dt$. It means that the guiding center locations will not change due to the polarization drift!

Let us demonstrate this in a single configuration. Assume $\mathbf{B} = B\hat{z}$, $\mathbf{E} = Et\hat{y}$. Let θ be the angle in the perpendicular plane:

$$\mathbf{v}_\perp = v_\perp \hat{e}_\perp = v_\perp (\cos \theta \hat{x} + \sin \theta \hat{y})$$

The guiding center is defined as

$$\mathbf{X} = \mathbf{x} - \mathbf{r}_L = \mathbf{x} - \mathbf{v}_\perp / \omega_c = \mathbf{x} - \frac{v_\perp}{\omega_c} \hat{\theta}$$

where $\hat{\theta} = \mathbf{b} \times \hat{e}_\perp$. The equation of motion

$$\begin{aligned} m \frac{d\mathbf{v}}{dt} &= q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \\ \frac{d\mathbf{x}}{dt} &= \mathbf{v} \end{aligned}$$

in the guiding center coordinates the two selected scalar equations can be written as

$$\begin{aligned} m \frac{dv_x}{dt} &= qE_x + qv_y B \\ \frac{dX_y}{dt} &= v_y - \frac{1}{\omega_c} \frac{dv_x}{dt} \end{aligned}$$

such that

$$\frac{dX_y}{dt} = 0$$

7.5 Time-Varying **B** Field

Finally, we allow the magnetic field to vary in time. Since the Lorentz force is always perpendicular to **v**, a magnetic field itself cannot impart energy to a charged particle. However, associated with the time-varying **B** is an electric field given by

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$$

and this can accelerate the particles. We can no longer assume the fields to be completely uniform. Let $\mathbf{v}_\perp = d\mathbf{l}/dt$ be the transverse velocity, \mathbf{l} being the element of the path along a particle trajectory (with v_\parallel neglected). Taking the scalar product of the equation of motion with \mathbf{v}_\perp , we have

$$\frac{d}{dt} \left(\frac{1}{2} mv_\perp^2 \right) = q\mathbf{E} \cdot \mathbf{v}_\perp = q\mathbf{E} \cdot \frac{d\mathbf{l}}{dt}$$

The change in one gyration is obtained by integrating over one period:

$$\delta \left(\frac{1}{2} mv_\perp^2 \right) = \int_0^{2\pi/\omega_c} q\mathbf{E} \cdot \frac{d\mathbf{l}}{dt} dt$$

If the field changes slowly, we can replace the time integral by a line integral over the unperturbed orbit:

$$\delta \left(\frac{1}{2} mv_\perp^2 \right) = \oint q\mathbf{E} \cdot d\mathbf{l} = q \int_s (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -q \int_s \dot{\mathbf{B}} \cdot d\mathbf{S}$$

Here **S** is the surface enclosed by the Larmor orbit and has a direction given by the right-hand rule when the fingers point in the direction of **v**. Since the plasma is diamagnetic, we have $\mathbf{B} \cdot d\mathbf{S} < 0$ for ions and $\mathbf{B} \cdot d\mathbf{S} > 0$ for electrons. Then

$$\delta \left(\frac{1}{2} mv_\perp^2 \right) = \pm q\dot{B}\pi r_L^2 = \pm q\pi\dot{B} \frac{v_\perp^2}{\omega_c \pm qB} \frac{m}{\pm qB} = \frac{\frac{1}{2} mv_\perp^2}{B} \cdot \frac{2\pi\dot{B}}{\omega_c}$$

The quantity $\frac{2\pi\dot{B}}{\omega_c} = \frac{\dot{B}}{\omega_c}$ is just the change δB during one period of gyration. Thus

$$\delta\left(\frac{1}{2}mv_{\perp}^2\right) = \mu\delta B$$

Since the left-hand-side is $\delta(\mu B)$, we have the desired result

$$\delta\mu = 0$$

The magnetic moment is invariant in slowly varying magnetic fields.

As the B field varies in strength, the Larmor orbits expand and contract, and the particles lose and gain transverse energy. This exchange of energy between the particles and the field is described very simply by the invariant of magnetic moment. The invariance of μ allows us to prove easily the following well-known theorem: *The magnetic flux through a Larmor orbit is constant.*

The flux Φ is given by BS , with $S = \pi r_L^2$. Thus

$$\Phi = B\pi\frac{v_{\perp}^2}{\omega_c^2} = B\pi\frac{v_{\perp}^2 m^2}{q^2 B^2} = \frac{2\pi m}{q^2} \frac{\frac{1}{2}mv_{\perp}^2}{B} = \frac{2\pi m}{q^2} \mu$$

Therefore, Φ is constant if μ is constant.

This property is used in a method of plasma heating known as *adiabatic compression*. A plasma is injected into a sequence of magnetic mirrors and by keep increasing the magnetic field in subsequent mirrors we can increase the plasma velocities.

7.6 Drifts in slowly changing arbitrary fields

Most of the drift motion we have considered before can be described under one framework. Consider in arbitrarily complicated but slowly changing fields subject to the following restrictions:

1. The time variation is so slow that the fields can be considered as approximately constant during each cyclotron period of the motion.
2. The fields vary so *gradually* in space that they are nearly uniform over the spatial extent of any single complete cyclotron drift.
3. The electric and magnetic fields are related by Faraday's law $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$.
4. $E/B \ll c$ so that relativistic effects are unimportant (otherwise there would be a problem with v_E becoming faster than c .)

These are known as the *guiding center approximation*. The center of the gyro motion is the *guiding center* (GC) and we call the frame of reference where $v_{\parallel} = 0$ is the *guiding center system* (GCS). In this more general situation a charged particle will gyrate about \mathbf{B} , stream parallel to \mathbf{B} , have $\mathbf{E} \times \mathbf{B}$ drifts across \mathbf{B} , and may also have force-based drifts. The analysis is based on the assumption that all these various motions are well-separated; this assumption is closely related to the requirement that the fields vary slowly and also to the concept of adiabatic invariant.

The assumed separation of scales is expressed by decomposing the particle motion into a fast, oscillatory component — the gyro-motion — and a slow component obtained by averaging out the gyromotion. As sketched in Figure 7.1, the particle's position and velocity are each decomposed into two terms

$$\mathbf{x}(t) = \mathbf{x}_{\text{gc}}(t) + \mathbf{r}_L(t), \quad \mathbf{v}(t) = \frac{d\mathbf{x}}{dt} = \mathbf{v}_{\text{gc}}(t) + \mathbf{v}_L(t)$$

where $\mathbf{r}_L(t), \mathbf{v}_L(t)$ give the *fast* gyration of the particle in a cyclotron orbit and $\mathbf{x}_{\text{gc}}(t), \mathbf{v}_{\text{gc}}(t)$ are the *slowly* changing motion of the *guiding center* obtained after averaging out the cyclotron motion. Ignoring any time dependence of the fields for now, the magnetic field seen by the particle can be written as

$$\begin{aligned} \mathbf{B}(\mathbf{x}(t)) &= \mathbf{B}(\mathbf{x}_{\text{gc}}(t) + \mathbf{r}_L(t)) \\ &= \mathbf{B}(\mathbf{x}_{\text{gc}}(t)) + (\mathbf{r}_L(t) \cdot \nabla) \mathbf{B} \end{aligned}$$

Because \mathbf{B} was assumed to be nearly uniform over the cyclotron orbit, it is sufficient to keep only the first term in the Taylor expansion of the magnetic field. The electric field may be expanded in a similar fashion.

After insertion of these Taylor expansions for the non-uniform electric and magnetic fields, the Lorentz equation becomes

$$\begin{aligned} m \frac{d[\mathbf{v}_{\text{gc}}(t) + \mathbf{v}_L(t)]}{dt} &= q [\mathbf{E}(\mathbf{x}_{\text{gc}}(t)) + (\mathbf{r}_L(t) \cdot \nabla) \mathbf{E}] \\ &\quad + q[\mathbf{v}_{\text{gc}}(t) + \mathbf{v}_L(t)] \times [\mathbf{B}(\mathbf{x}_{\text{gc}}(t)) + (\mathbf{r}_L(t) \cdot \nabla) \mathbf{B}] \end{aligned}$$

The gyromotion (i.e. the fast cyclotron motion) is defined to be the solution of the equation

$$m \frac{d\mathbf{v}_L(t)}{dt} = q\mathbf{v}_L(t) \times \mathbf{B}(\mathbf{x}_{\text{gc}}(t))$$

Substracting this fast motion equation from the prior one leaves

$$\begin{aligned} m \frac{d\mathbf{v}_{\text{gc}}(t)}{dt} &= q [\mathbf{E}(\mathbf{x}_{\text{gc}}(t)) + (\mathbf{r}_L(t) \cdot \nabla) \mathbf{E}] \\ &\quad + q \{ \mathbf{v}_{\text{gc}}(t) \times [\mathbf{B}(\mathbf{x}_{\text{gc}}(t)) + (\mathbf{r}_L(t) \cdot \nabla) \mathbf{B}] + \mathbf{v}_L(t) \times (\mathbf{r}_L(t) \cdot \nabla) \mathbf{B} \} \end{aligned} \tag{7.13}$$

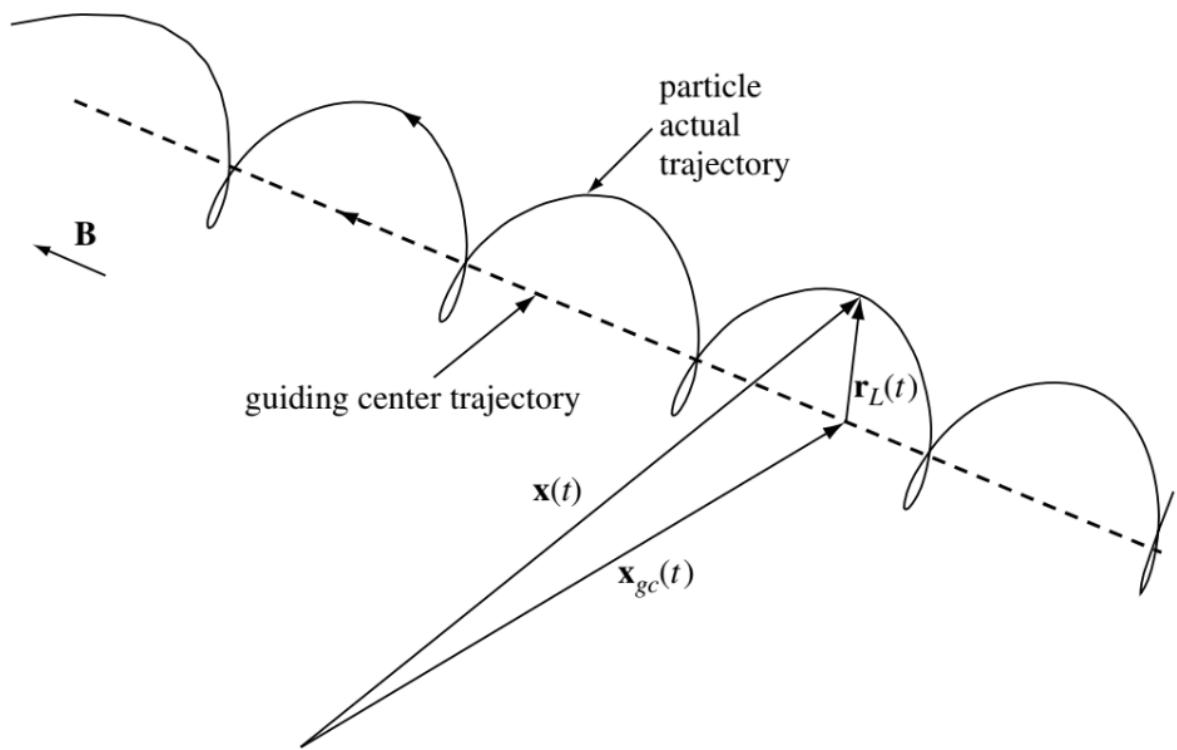


Figure 7.1: Drift in an arbitrarily complicated field.

Let us now average Equation 7.13 over one gyroperiod in which case terms *linear* in gyromotion average to zero. What remains is an equation describing the slow quantities, namely

$$m \frac{d\mathbf{v}_{gc}(t)}{dt} = q \{ \mathbf{E}(\mathbf{x}_{gc}(t)) + \mathbf{v}_{gc}(t) \times \mathbf{B}(\mathbf{x}_{gc}(t)) + \langle \mathbf{v}_L(t) \times (\mathbf{r}_L \cdot \nabla) \mathbf{B} \rangle \} \quad (7.14)$$

where $\langle \rangle$ means averaged over a cyclotron period. The guiding center velocity can now be decomposed into components perpendicular and parallel to \mathbf{B} ,

$$\mathbf{v}_{gc}(t) = \mathbf{v}_{\perp gc}(t) + v_{\parallel gc}(t)\hat{b}$$

so that

$$\frac{d\mathbf{v}_{gc}(t)}{dt} = \frac{d\mathbf{v}_{\perp gc}(t)}{dt} + \frac{d(\mathbf{v}_{\parallel gc}(t)\hat{b})}{dt} = \frac{d\mathbf{v}_{\perp gc}(t)}{dt} + \frac{d\mathbf{v}_{\parallel gc}(t)}{dt}\hat{b} + \mathbf{v}_{\parallel gc}(t)\frac{d\hat{b}}{dt}$$

Denoting the distance along the magnetic field by s , the derivative of the magnetic field unit vector can be written, to lowest order, as (???)

$$\frac{d\hat{b}}{dt} = \frac{\partial \hat{b}}{\partial s} \frac{ds}{dt} = v_{\parallel gc}\hat{b} \cdot \nabla \hat{b}$$

so Equation 7.14 becomes

$$\begin{aligned} m \left[\frac{d\mathbf{v}_{\perp gc}(t)}{dt} + \frac{d\mathbf{v}_{\parallel gc}(t)}{dt}\hat{b} + \mathbf{v}_{\parallel gc}^2(t)\hat{b} \cdot \nabla \hat{b} \right] &= q\mathbf{E}(\mathbf{x}_{gc}(t)) \\ &\quad + q\mathbf{v}_{gc}(t) \times \mathbf{B}(\mathbf{x}_{gc}(t)) \\ &\quad + q\langle \mathbf{v}_L(t) \times (\mathbf{r}_L \cdot \nabla) \mathbf{B} \rangle \end{aligned}$$

The component of this equation along \mathbf{B} is

$$m \frac{d\mathbf{v}_{\parallel gc}(t)}{dt} = q \left[E_{\parallel}(\mathbf{x}_{gc}(t)) + \langle \mathbf{v}_L(t) \times (\mathbf{r}_L \cdot \nabla) \mathbf{B} \rangle_{\parallel} \right]$$

while the component perpendicular to \mathbf{B} is

$$\begin{aligned} m \left[\frac{d\mathbf{v}_{\perp gc}(t)}{dt} + \mathbf{v}_{\parallel gc}^2(t)\hat{b} \cdot \nabla \hat{b} \right] &= q\mathbf{E}_{\perp}(\mathbf{x}_{gc}(t)) \\ &\quad + q\mathbf{v}_{gc}(t) \times \mathbf{B}(\mathbf{x}_{gc}(t)) \\ &\quad + q\langle \mathbf{v}_L(t) \times (\mathbf{r}_L \cdot \nabla) \mathbf{B} \rangle_{\perp} \end{aligned} \quad (7.15)$$

Equation 7.15 is of the generic form

$$m \frac{d\mathbf{v}_{\perp gc}}{dt} = \mathbf{F}_{\perp} + q\mathbf{v}_{gc} \times \mathbf{B} \quad (7.16)$$

where

$$\mathbf{F}_\perp = q \left[\mathbf{E}_\perp(\mathbf{x}_{gc}(t)) + \langle \mathbf{v}_L(t) \times (\mathbf{r}_L \cdot \nabla) \mathbf{B} \rangle_\perp \right] - m \mathbf{v}_{\parallel gc}^2(t) \hat{b} \cdot \nabla \hat{b} \quad (7.17)$$

Equation 7.16 is solved iteratively based on the assumption that \mathbf{v}_{gc} has a slow time dependence. In the first iteration, the time dependence is neglected altogether so that the left-hand side of Equation 7.16 is set to zero to obtain the “first guess” for the perpendicular drift to be

$$\mathbf{v}_{\perp gc} \simeq \mathbf{v}_F \equiv \frac{\mathbf{F}_\perp \times \mathbf{B}}{qB^2}$$

Next, \mathbf{v}_p is defined to be a correction to this first guess, where \mathbf{v}_p is assumed small and incorporates effects due to any time dependence of $\mathbf{v}_{\perp gc}$. To determine \mathbf{v}_p , we write $\mathbf{v}_{\perp gc} = \mathbf{v}_F + \mathbf{v}_p$ so, to second order Equation 7.16 becomes

$$m \frac{d(\mathbf{v}_F + \mathbf{v}_p)}{dt} = \mathbf{F}_\perp + q(\mathbf{v}_F + \mathbf{v}_p) \times \mathbf{B} \quad (7.18)$$

In accordance with the slowness condition, it is assumed that $|d\mathbf{v}_p/dt| \ll |d\mathbf{v}_F/dt|$ so Equation 7.18 becomes

$$0 = -m \frac{d\mathbf{v}_F}{dt} + q\mathbf{v}_p \times \mathbf{B}$$

Crossing this equation with \mathbf{B} gives the general *polarization drift*

$$\mathbf{v}_p = -\frac{m}{qB^2} \frac{d\mathbf{v}_F}{dt} \times \mathbf{B}$$

The most important example of the polarization drift is when \mathbf{v}_F is the $\mathbf{E} \times \mathbf{B}$ drift in a uniform, constant magnetic field so that ($\mathbf{E} \perp \mathbf{B}$?)

$$\begin{aligned} \mathbf{v}_p &= -\frac{m}{qB^2} \frac{d}{dt} \left(\frac{\mathbf{E} \times \mathbf{B}}{B^2} \right) \times \mathbf{B} \\ &= \frac{m}{qB^2} \frac{d\mathbf{E}}{dt} \end{aligned}$$

To calculate the middle term on the RHS of Equation 7.18, it is necessary to average over cyclotron orbits. This middle term is defined as the “grad B” force

$$\mathbf{F}_{\nabla B} = q \langle \mathbf{v}_L(t) \times (\mathbf{r}_L(t) \cdot \nabla) \mathbf{B} \rangle \quad (7.19)$$

To simplify the algebra for the averaging, a local Cartesian coordinate system is used with x axis in the direction of the gyrovelocity at $t = 0$ and z axis in the direction of the magnetic field at the *gyrocenter*. Thus, the Larmor orbit velocity has the form

$$\mathbf{v}_L(t) = v_{L0} [\hat{x} \cos \omega_c t - \hat{y} \sin \omega_c t]$$

where

$$\omega_c = \frac{qB}{m}$$

is the cyclotron frequency we know before and the Larmor orbit position has the form

$$\mathbf{r}_L(t) = \frac{v_{L0}}{\omega_c} [\hat{x} \sin \omega_c t + \hat{y} \cos \omega_c t]$$

Inserting the above two expressions in Equation 7.19 gives

$$\mathbf{F}_{\nabla B} = q \frac{v_{L0}^2}{\omega_c} \langle [\hat{x} \cos \omega_c t - \hat{y} \sin \omega_c t] \times ([\hat{x} \sin \omega_c t + \hat{y} \cos \omega_c t] \cdot \nabla) \mathbf{B} \rangle$$

Noting that $\langle \sin^2 \omega_c t \rangle = \langle \cos^2 \omega_c t \rangle = 1/2$ while $\langle \sin \omega_c t \cos \omega_c t \rangle = 0$, this reduces to

$$\begin{aligned} \mathbf{F}_{\nabla B} &= \frac{qv_{L0}^2}{2\omega_c} \left[\hat{x} \times \frac{\partial \mathbf{B}}{\partial y} - \hat{y} \times \frac{\partial \mathbf{B}}{\partial x} \right] \\ &= \frac{mv_{L0}^2}{2B} \left[\hat{x} \times \frac{\partial(B_y \hat{y} + B_z \hat{z})}{\partial y} - \hat{y} \times \frac{\partial(B_x \hat{x} + B_z \hat{z})}{\partial x} \right] \\ &= \frac{mv_{L0}^2}{2B} \left[\hat{z} \left(\frac{\partial B_y}{\partial y} + \frac{\partial B_x}{\partial x} \right) - \hat{y} \frac{\partial B_z}{\partial y} - \hat{x} \frac{\partial B_z}{\partial x} \right] \end{aligned}$$

But from $\nabla \cdot \mathbf{B} = 0$, it is seen that $\frac{\partial B_y}{\partial y} + \frac{\partial B_x}{\partial x} = -\frac{\partial B_z}{\partial z}$ so the “grad B” force is

$$\mathbf{F}_{\nabla B} = -\frac{mv_{L0}^2}{2B} \nabla B$$

where the approximation $B_z \simeq B$ has been used since the magnetic field direction is mainly in the \hat{z} direction.

By applying the guiding center approximation, we reduce the number of independent variables from 6 (x, y, z, v_x, v_y, v_z) to 5 ($x, y, z, v_{\perp}, v_{\parallel}$).

7.7 Summary of Guiding Center Drifts

General force:

$$\mathbf{v}_f = \frac{1}{q} \frac{\mathbf{F} \times \mathbf{B}}{B^2} \quad (7.20)$$

Electric field:

$$\mathbf{v}_E = \frac{\mathbf{E} \times \mathbf{B}}{B^2} \quad (7.21)$$

Gravitational field:

$$\mathbf{v}_g = \frac{m}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2} \quad (7.22)$$

Nonuniform electric field:

$$\mathbf{v}_E = \left(1 + \frac{1}{4} r_L^2 \nabla^2\right) \frac{\mathbf{E} \times \mathbf{B}}{B^2} \quad (7.23)$$

Nonuniform magnetic field:

Grad-B:

$$\begin{aligned} \mathbf{v}_{\nabla B} &= \pm \frac{1}{2} v_{\perp} r_L \frac{\mathbf{B} \times \nabla B}{B^2} \\ &= \frac{mv_{\perp}^2}{2qB^3} \mathbf{B} \times \nabla B \end{aligned} \quad (7.24)$$

Curvature drift:

$$\begin{aligned} \mathbf{v}_c &= \frac{mv_{\parallel}^2}{q} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2} \\ &= \frac{mv_{\parallel}^2}{qB^4} \mathbf{B} \times [\mathbf{B} \cdot \nabla \mathbf{B}] \\ &= \frac{mv_{\parallel}^2}{qB^3} \mathbf{B} \times \nabla B \text{ (current-free)} \end{aligned} \quad (7.25)$$

Curved vacuum field:

$$\begin{aligned} \mathbf{v}_c + \mathbf{v}_{\nabla B} &= \frac{m}{q} \left(v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2 \right) \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2} \\ &= \frac{m}{q} \frac{\mathbf{B} \times \nabla B}{B^3} \left(v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2 \right) \end{aligned} \quad (7.26)$$

Polarization drift:

$$\mathbf{v}_p = \pm \frac{1}{\omega_c B} \frac{d\mathbf{E}}{dt} \quad (7.27)$$

7.8 Adiabatic Invariants

It is well known in classical mechanics that whenever a system has a periodic motion, the action integral $\oint pdq$ taken over a period is a constant of the motion. Here p and q are the generalized momentum and coordinate which repeat themselves in the motion. If a slow change is made in the system, so that the motion is not quite periodic, the constant of the motion does not change

and is then called an *adiabatic invariant*.³ By slow here we mean slow compared with the period of motion, so that the integral $\oint pdq$ is well defined even though it is strictly no longer an integral over a closed path. Adiabatic invariants play an important role in plasma physics; they allow us to obtain simple answers in many instances involving complicated motions. There are three adiabatic invariants, each corresponding to a different type of periodic motion.

A review of adiabatic charged-particle motion is given by Northrop (1963).

7.8.1 Adiabatic Invariant of a Pendulum

Perfect symmetry is never attained in reality. This leads to the practical question of how constants of the motion behave when space and/or time symmetries are “good”, but not perfect. Does the utility of constants of the motion collapse abruptly when the slightest non-symmetrical blemish rears its ugly head, does the utility decay gracefully, or does something completely different happen? To answer these questions, we begin by considering the problem of a small-amplitude pendulum having a time-dependent, but slowly changing resonant frequency $\omega(t)$. Since $\omega^2 = g/l$, the time-dependence of the frequency might result from either a slow change in the gravitational acceleration g or else from a slow change in the pendulum length l . In both cases the pendulum equation of motion will be

$$\frac{d^2x}{dt^2} + \omega^2(t)x = 0 \quad (7.28)$$

This equation cannot be solved exactly for arbitrary $\omega(t)$ but if a modest restriction is put on $\omega(t)$ the equation can be solved *approximately* using the WKB method (Wentzel 1926, Kramers 1926, Brillouin 1926). This method is based on the hypothesis that the solution for a time-dependent frequency is likely to be a generalization of the constant-frequency solution

$$x = \text{Re}[A \exp(i\omega t)]$$

where this generalization is postulated to be of the form

$$x = \text{Re}\left[A(t)e^{i\int^t \omega(t')dt'}\right] \quad (7.29)$$

Here $A(t)$ is an amplitude function determined as follows: calculate the first derivative we get

$$\frac{dx}{dt} = \text{Re}\left[i\omega A e^{i\int^t \omega(t')dt'} + \frac{dA}{dt} e^{i\int^t \omega(t')dt'}\right]$$

³This invariant has the name adiabatic associated with it because historically it was firstly studied systematically by Paul Ehrenfest in the irreversible adiabatic process in thermal dynamics. Later the word “irreversible” is neglected and it was called “adiabatic invariant”, although this conclusion is valid in a general system that changes slowly that may have nothing to do with thermal dynamics. This also has deep root with quantum mechanics.

then the second derivative

$$\frac{d^2x}{dt^2} = \operatorname{Re} \left[\left(i \frac{d\omega}{dt} A + 2i\omega \frac{dA}{dt} - \omega^2 A + \frac{d^2A}{dt^2} \right) e^{i \int^t \omega(t') dt'} \right]$$

and insert this last result into Equation 7.28 which reduces to

$$i \frac{d\omega}{dt} A + 2i\omega \frac{dA}{dt} + \frac{d^2A}{dt^2} = 0 \quad (7.30)$$

since the terms involving ω^2 cancel exactly. To proceed further, we make an assumption — the validity of which is to be checked later — that the time dependence of dA/dt is *sufficiently* slow to allow dropping the last term in Equation 7.30 relative to the middle term. The two terms that remain in Equation 7.30 can then be rearranged as

$$\frac{1}{\omega} \frac{d\omega}{dt} = -\frac{2}{A} \frac{dA}{dt}$$

which has the exact solution

$$A(t) \sim \frac{1}{\sqrt{\omega(t)}} \quad (7.31)$$

The assumption of slowness is thus at least self-consistent, for if t is indeed slowly changing, Equation 7.31 shows that A will also be slowly changing and the dropping of the last term in Equation 7.30 is justified. The slowness requirement can be quantified by assuming that the frequency has an exponential dependence

$$\omega(t) = \omega_0 e^{\alpha t}$$

Thus,

$$\alpha = \frac{1}{\omega} \frac{d\omega}{dt}$$

is a measure of how fast the frequency is changing compared to the frequency itself. Hence, dropping the last term in Equation 7.30 is legitimate if

$$\alpha = \frac{1}{\omega} \frac{d\omega}{dt} \ll 4\omega_0 \quad (7.32)$$

In other words, if Equation 7.32 is satisfied, then the fractional change of the pendulum period per period is small.

Equation 7.31 indicates that when ω is time-dependent, the pendulum amplitude is not constant and so the pendulum energy is *not* conserved. It turns out that what is conserved is the *action integral*

$$S = \oint v dx \quad (7.33)$$

where the integration is over one period of oscillation. This integral can also be written in terms of time as

$$S = \int_{t_0}^{t_0+\tau} v \frac{dx}{dt} dt$$

where t_0 is a time when x is at an instantaneous maximum and τ , the period of a complete cycle, is defined as the interval between two successive times when $dx/dt = 0$ and d^2x/dt^2 has the same sign (e.g., for a pendulum, t_0 would be a time when the pendulum has swung all the way to the right and so is reversing its velocity while τ is the time one has to wait for this to happen again). To show that action is conserved, Equation 7.33 can be integrated by parts as

$$\begin{aligned} S &= \int_{t_0}^{t_0+\tau} \left[\frac{d}{dt} \left(x \frac{dx}{dt} \right) - x \frac{d^2x}{dt^2} \right] dt \\ &= \left[x \frac{dx}{dt} \right]_{t_0}^{t_0+\tau} - \int_{t_0}^{t_0+\tau} x \frac{d^2x}{dt^2} dt \\ &= \int_{t_0}^{t_0+\tau} \omega^2 x^2 dt \end{aligned} \quad (7.34)$$

where the integrated term has vanished by virtue of the definitions of t_0 and τ , and Equation 7.28 has been used to substitute d^2x/dt^2 . Equation 7.29 and Equation 7.31 can be combined to give

$$x(t) = x(t_0) \sqrt{\frac{\omega(t_0)}{\omega(t)}} \cos \left(\int_{t_0}^t \omega(t') dt' \right)$$

so Equation 7.34 becomes

$$\begin{aligned} S &= \int_{t_0}^{t_0+\tau} \omega(t')^2 \left[x(t_0) \sqrt{\frac{\omega(t_0)}{\omega(t')}} \cos \left(\int_{t_0}^{t'} \omega(t'') dt'' \right) \right] dt' \\ &= [x(t_0)]^2 \omega(t_0) \int_{t_0}^{t_0+\tau} \omega(t') \cos^2 \left(\int_{t_0}^{t'} \omega(t'') dt'' \right) dt' \\ &= [x(t_0)]^2 \omega(t_0) \int_0^{2\pi} d\xi \cos^2 \xi = \pi [x(t_0)]^2 \omega(t_0) = \text{const.} \end{aligned} \quad (7.35)$$

where $\xi = \int_{t_0}^{t'} \omega(t'') dt''$ and $d\xi = \omega(t') dt'$. Equation 7.33 shows that S is the area in phase-space enclosed by the trajectory $[x(t), v(t)]$ and Equation 7.35 shows that for a slowly changing pendulum frequency, *this area is a constant of the motion*. Since the average energy of the pendulum scales as $\sim [\omega(t)x(t)]^2$, we see from Equation 7.31 that the ratio

$$\frac{\text{energy}}{\text{frequency}} \sim \omega(t)x^2(t) \sim S \sim \text{const.} \quad (7.36)$$

The ratio in Equation 7.36 is the classical equivalent of the quantum number N of a simple harmonic oscillator because in quantum mechanics the energy E of a simple harmonic oscillator is related to the frequency by the relation $E/\hbar\omega = N + 1/2$.

This analysis clearly applies to any dynamical system having an equation of motion of the form of Equation 7.28. Hence, if the dynamics of plasma particles happens to be of this form, then S can be added to our repertoire of constants of the motion.

7.8.2 Extension of WKB method to general adiabatic invariant

Action has the dimensions of (canonical momentum) \times (canonical coordinate) so we may anticipate that for general Hamiltonian systems, the action integral given in Equation 7.33 is not an invariant because v is not, in general, proportional to P . We postulate that the general form for the action integral is

$$S = \oint P dQ \quad (7.37)$$

where the integral is over one period of the periodic motion and P, Q are the relevant canonical momentum-coordinate conjugate pair. The proof of adiabatic invariance used for Equation 7.33 does not work directly for Equation 7.37; we now present a slightly more involved proof to show that Equation 7.37 is indeed the more general form of adiabatic invariant.

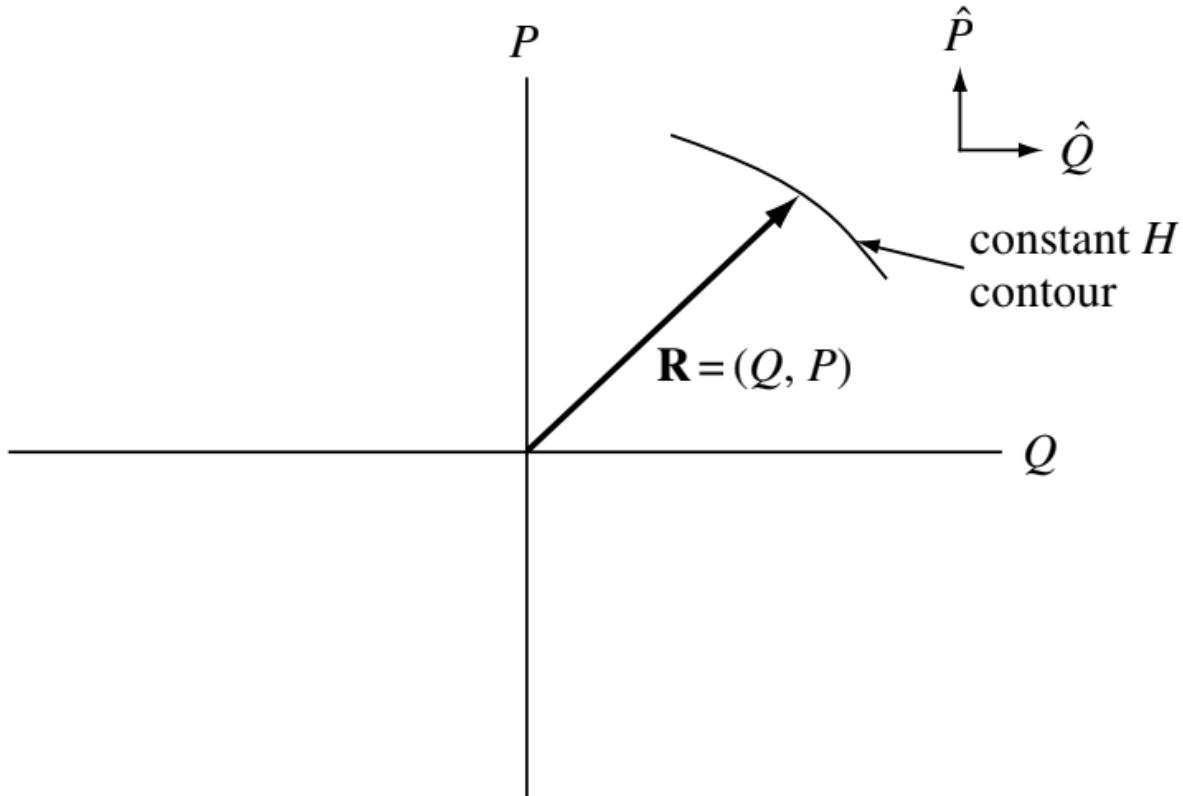


Figure 7.2: Canonical coordinate-momentum plane.

Let us define the radius vector in the $Q - P$ plane to be $\mathbf{R} = (Q, P)$ and define unit vectors in the Q and P directions by \hat{Q} and \hat{P} ; these definitions are shown in Figure 7.2. Furthermore, we define the z direction as being normal to the $Q - P$ plane; thus, the unit vector $\hat{z} = \hat{Q} \times \hat{P}$. Hamilton's equations ($\dot{P} = -\partial H / \partial Q$, $\dot{Q} = \partial H / \partial P$) may be written in vector form as

$$\frac{d\mathbf{R}}{dt} = -\hat{z} \times \nabla H \quad (7.38)$$

where

$$\nabla = \hat{Q} \frac{\partial}{\partial Q} + \hat{P} \frac{\partial}{\partial P}$$

is the gradient operator in the $Q - P$ plane. Equation 7.38 shows that the phase-space “velocity” $d\mathbf{R}/dt$ is *orthogonal* to ∇H . Hence, \mathbf{R} *stays on a level contour of H* . If H is a constant, then, in order for the motion to be periodic, the path along this level contour must circle around and join itself, like a road of constant elevation around the rim of a mountain (or a crater). If H is not constant, but slowly changing in time, the contour will circle around and nearly join itself.

Equation 7.38 can be inverted by crossing it with \hat{z} to give

$$\nabla H = \hat{z} \times \frac{d\mathbf{R}}{dt} \quad (7.39)$$

For periodic and near-periodic motions, $d\mathbf{R}/dt$ is always in the same sense (always clockwise or always counterclockwise). Thus, Equation 7.39 shows that an observer following the path would always see that H is increasing on the left-hand side of the path and decreasing on the right-hand side (or vice versa). For clarity, the origin of the $Q - P$ plane is redefined to be at a local maximum or minimum of H . Hence, near the extremum H must have the Taylor expansion

$$H(P, Q) = H_{\text{extremum}} + \frac{P^2}{2} \left[\frac{\partial^2 H}{\partial P^2} \right]_{P=0, Q=0} + \frac{Q^2}{2} \left[\frac{\partial^2 H}{\partial Q^2} \right]_{P=0, Q=0}$$

where the second order derivatives are either both positive (valley) or both negative (hill). Since H is assumed to have a slow dependence on time, these second derivatives will be time-dependent so that H has the form

$$H = \alpha(t) \frac{P^2}{2} + \beta(t) \frac{Q^2}{2} \quad (7.40)$$

where $\alpha(t)$ and $\beta(t)$ have the same sign. The term H_{extremum} has been dropped because it is just an additive constant to the energy and does not affect Hamilton's equations. From Equation 7.38 the direction of rotation of \mathbf{R} is seen to be counterclockwise if the extremum of H is a hill, and clockwise if a valley.

Hamilton's equations operating on Equation 7.40 give

$$\frac{dP}{dt} = -\beta Q, \quad \frac{dQ}{dt} = \alpha P \quad (7.41)$$

These equations do not directly generate the simple harmonic oscillator equation because of the time dependence of α, β . However, if we define the auxiliary variable

$$\tau = \int^t \beta(t') dt'$$

then

$$\frac{d}{dt} = \frac{d\tau}{dt} \frac{d}{d\tau} = \beta \frac{d}{d\tau}$$

so Equation 7.41 becomes

$$\frac{dP}{d\tau} = -Q, \quad \frac{dQ}{d\tau} = \frac{\alpha}{\beta} P \quad (7.42)$$

Substituting for Q in the right-hand equation using the left-hand equation gives

$$\frac{d^2P}{d\tau^2} + \frac{\alpha}{\beta} P = 0$$

which is a simple harmonic oscillator with $\omega^2(\tau) = \alpha(\tau)/\beta(\tau)$. The action integral may be rewritten as

$$S = \oint P \frac{dQ}{d\tau} d\tau$$

where the integral is over one period of the motion. Using Equation 7.42 and following the same procedure as was used with Equation 7.35, this becomes

$$S = \oint P^2 \frac{\alpha}{\beta} d\tau = \lambda^2 \int \left[\left(\frac{\alpha(\tau')}{\beta(\tau')} \right)^{1/2} \cos^2 \left(\int^{\tau'} (\alpha/\beta)^{1/2} d\tau'' \right) \right] d\tau'$$

where λ is a constant dependent on initial conditions. By introducing the orbit phase $\phi = \int^{\tau} (\alpha/\beta)^{1/2} d\tau$, the above becomes

$$S = \lambda^2 \int_0^{2\pi} d\phi \cos^2 \phi = \text{const.}$$

Thus, the general action integral is indeed an adiabatic invariant. This proof is of course only valid in the vicinity of an extremum of H , i.e., only where H can be adequately represented by Equation 7.40.

7.8.3 Proof for the general adiabatic invariant

We now develop a proof for the general adiabatic invariant. This proof is not restricted to small oscillations (i.e., being near an extremum of H) as was the previous discussion. Let the

Hamiltonian depend on time via a slowly changing parameter $\lambda(t)$, so that $H = H(P, Q, \lambda(t))$. From Equation 4.6 the energy is given by

$$E(t) = H(P, Q, \lambda(t)) \quad (7.43)$$

and, in principle, this relation can be inverted to give $P = P(E(t), Q, \lambda(t))$. Suppose a particle is executing nearly periodic motion in the $Q - P$ plane. We define the turning point Q_{tp} as a position where $dQ/dt = 0$. Since Q is oscillating, there will be a turning point associated with Q having its maximum value and a turning point associated with Q having its minimum value. From now on let us only consider turning points where Q has its maximum value, that is, we only consider the turning points on the right-hand side of the nearly periodic trajectories in the $Q - P$ plane shown in Figure 7.3.

If the motion is periodic, then the turning point for the $N + 1^{\text{th}}$ period will be the same as the turning point for the N^{th} period, but if the motion is only nearly periodic, there will be a slight difference as shown in Figure 7.3. This difference can be characterized by making the turning point a function of time so $Q_{tp} = Q_{tp}(t)$. This function is only defined for the times when $dQ/dt = 0$. When the motion is not exactly periodic, this turning point is such that $Q_{tp}(t + \tau) \neq Q_{tp}(t)$, where τ is the time interval required for the particle to go from the first turning point to the next turning point. The action integral is over one entire period of oscillation starting from a right-hand turning point and then going to the next right-hand turning point and so can be written as

$$S = \oint P dQ = \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} P dQ$$

From Equation 4.6 it is seen that P/m is not, in general, the velocity and so the velocity dQ/dt is not, in general, proportional to P (???). Thus, the turning points are not necessarily at the locations where P vanishes, and in fact P need not change sign during a period. However, S still corresponds to the area of phase-space enclosed by one period of the phase-space trajectory.

We can now calculate

$$\begin{aligned} \frac{dS}{dt} &= \frac{d}{dt} \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} P(E(t), Q, \lambda(t)) dQ \\ &= \left[P \frac{dQ}{dt} \right]_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} + \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} \left(\frac{\partial P}{\partial t} \right)_Q dQ \\ &= \int_{Q_{tp}(t)}^{Q_{tp}(t+\tau)} \left[\left(\frac{\partial P}{\partial E} \right)_{Q,\lambda} \frac{dE}{dt} + \left(\frac{\partial P}{\partial \lambda} \right)_{Q,E} \frac{d\lambda}{dt} \right] dQ \end{aligned} \quad (7.44)$$

Because $dQ/dt = 0$ at the turning point, the integrated term vanishes and so there is no contribution from motion of the turning point. From Equation 7.43 we have

$$1 = \frac{\partial H}{\partial P} \left(\frac{\partial P}{\partial E} \right)_{Q,\lambda}$$

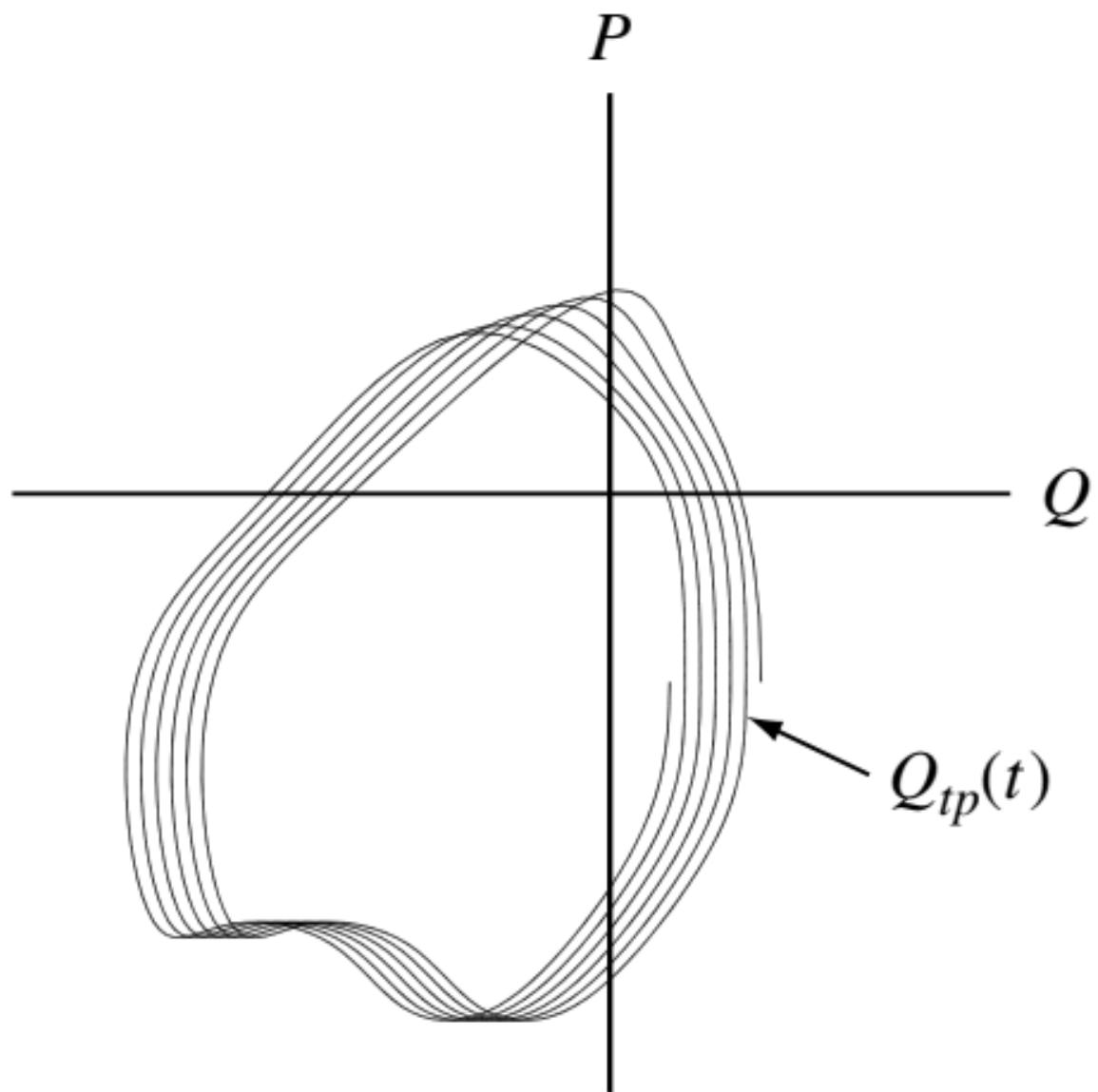


Figure 7.3: Nearly periodic-phase trajectory for slowly changing Hamiltonian. The turning point $Q_{tp}(t)$ is where Q is at its maximum.

and

$$0 = \frac{\partial H}{\partial P} \left(\frac{\partial P}{\partial \lambda} \right)_{Q,E} + \frac{\partial H}{\partial \lambda}$$

so Equation 7.44 becomes

$$\frac{dS}{dt} = \oint \left(\frac{\partial H}{\partial P} \right)^{-1} \left[\frac{dE}{dt} - \frac{\partial H}{\partial \lambda} \frac{d\lambda}{dt} \right] dQ \quad (7.45)$$

From Equation 7.43 we have

$$\frac{dE}{dt} = \frac{\partial H}{\partial P} \frac{dP}{dt} + \frac{\partial H}{\partial Q} \frac{dQ}{dt} + \frac{\partial H}{\partial \lambda} \frac{d\lambda}{dt} = \frac{\partial H}{\partial \lambda} \frac{d\lambda}{dt} \quad (7.46)$$

since the first two terms canceled due to Hamilton's equations. Substitution of Equation 7.46 into Equation 7.45 gives $dS/dt = 0$, completing the proof of adiabatic invariance. No assumption has been made here that P, Q are close to the values associated with an extremum of H .

This proof seems too neat, because it has established adiabatic invariance simply by careful use of the chain rule, and by taking partial derivatives. However, this observation reveals the underlying essence of adiabaticity, namely it is the differentiability of H, P with respect to λ from one period to the next and the Hamilton nature of the system, which together provide the conditions for the adiabatic invariant to exist. If the motion had been such that after one cycle the motion had changed so drastically that taking a derivative of H or P with respect to λ would not make sense, then the adiabatic invariant would not exist.

7.8.4 The First Adiabatic Invariant

We have already met the quantity

$$\mu = mv_\perp^2/2B$$

and have proved its invariance in spatially and temporally varying \mathbf{B} fields. The periodic motion involved, of course, is the Larmor gyration. If we take p to be angular momentum $mv_\perp r$ and dq to be the coordinate $d\theta$, the action integral becomes

$$\oint pdq = \oint mv_\perp r d\theta = 2\pi r_L mv_\perp = 2\pi \frac{mv_\perp^2}{\omega_c} = 4\pi \frac{m}{|q|} \mu$$

Thus μ is a constant of the motion as long as q/m is not changed. We have proved the invariance of μ only with the implicit assumption $\omega/\omega_c \ll 1$, where ω is a frequency characterizing the rate of change of \mathbf{B} as seen by the particle. A proof exists, however, that μ is invariant even when $\omega/\omega_c \leq q$. In theorists' language, μ is invariant "to all orders in an expansion in ω/ω_c ." What this means in practice is that μ remains much more nearly constant than \mathbf{B} does during one period of gyration.

The first invariant is derived from the action integral of the fast gyromotion. By expanding the magnetic vector potential \mathbf{A} about the guiding center and integrating the canonical momentum perpendicular to the field over one gyro-period, we obtain the first adiabatic invariant:

$$I_1 = \oint \mathbf{P}_\perp \cdot d\rho \approx \frac{c}{e} \frac{p_\perp^2}{2B} = \frac{cm}{e} \mu$$

This yields the magnetic moment μ , which is conserved in the guiding-center limit:

$$\mu = \frac{p_\perp^2}{2mB}$$

In the relativistic case, the magnetic moment is related to the adiabatic invariant by $M = \mu/\gamma$.

It is just as important to know when an adiabatic invariant does *not* exist as to know when it does. Adiabatic invariance of μ is violated when ω_c is not small compared with ω . We give three examples of this.

- *Magnetic Pumping.* If the strength of \mathbf{B} in a mirror confinement system is varied sinusoidally, the particles' v_\perp would oscillate; but there would be no gain of energy in the long run. However, if the particles make collisions, the invariance of μ is violated, and the plasma can be heated. In particular, a particle making a collision during the compression phase can transfer part of its gyration energy into v_\parallel energy, and this is not taken out again in the expansion phase.
- *Cyclotron Heating.* Now imagine that the \mathbf{B} field is oscillated at the frequency ω_c . The induced electric field will then rotate in phase with some of the particles and accelerate their Larmor motion continuously. The condition $\omega \ll \omega_c$ is violated, μ is not conserved, and the plasma can be heated.
- *Magnetic Cusps.* If the current in one of the coils in a simple magnetic mirror system is reversed, a magnetic cusp is formed. This configuration has, in addition to the usual mirrors, a spindle-cusp mirror extending over 360° in azimuth. A plasma confined in a cusp device is supposed to have better stability properties than that in an ordinary mirror. Unfortunately, the loss-cone losses are larger because of the additional loss region; and the particle motion is nonadiabatic. Since the \mathbf{B} field vanishes at the center of symmetry, ω_c is zero there; and μ is not preserved. The local Larmor radius near the center is larger than the device. Because of this, the adiabatic invariant μ does not guarantee that particles outside a loss cone will stay outside after passing through the nonadiabatic region. Fortunately, there is in this case another invariant: the canonical angular momentum $p_\theta = mr v_\theta - er A_\theta$. This ensures that there will be a population of particles trapped indefinitely until they make a collision.

7.8.5 The Second Adiabatic Invariant

Once the fast gyromotion is averaged out, the remaining motion consists of bouncing and drifting.

Second Invariant (I_2 or J): Derived by integrating the parallel momentum along the field line between bounce points s_m and s'_m

$$J = \oint p_{\parallel} ds = 2p \int_{s_m}^{s'_m} \sqrt{1 - \frac{B(s)}{B_m}} ds$$

Conservation of J implies that as a particle drifts across field lines, it adjusts its bounce path to keep this integral constant.

Consider a particle trapped between two magnetic mirrors: it bounces between them and therefore has a periodic motion at the “bounce frequency”. A constant of this motion is given by $\oint mv_{\parallel} ds$, where ds is an element of path length (of the guiding center) along a field line. However, since the guiding center drifts across field lines, the motion is not exactly periodic, and the constant of the motion becomes an adiabatic invariant. This is called the *longitudinal invariant J* and is defined for a half-cycle between the two turning points

$$J = \int_a^b v_{\parallel} ds$$

We shall prove that J is invariant in a static, nonuniform B field; the result is also true for a slowly time-varying B field.

Before embarking on this somewhat lengthy proof, let us consider an example of the type of problem in which a theorem on the invariance of J would be useful. As we have already seen, the earth’s magnetic field mirror-traps charged particles, which slowly drift in longitude around the earth. If the magnetic field were perfectly symmetric, the particle would eventually drift back to the same magnetic field line. However, the actual field is distorted by such effects as the solar wind. In that case, will a particle ever come back to the same magnetic field line? Since the particle’s energy is conserved and is equal to $\frac{1}{2}mv_{\perp}^2$ at the turning point, the invariance of J indicates that $|B|$ remains the same at the turning point. However, upon drifting back to the same longitude, a particle may find itself on another magnetic field line at a different altitude. This cannot happen if J is conserved. J determines the length of the field lines between turning points, and no two lines have the same length between points with the same $|B|$. Consequently, the particle returns to the same magnetic field line even in a slightly asymmetric field.

To prove the invariance of J , we first consider the invariance of $v_{\parallel}\delta s$, where δs is a segment of the path along \mathbf{B} . Because of guiding center drifts, a particle on s will find itself on another magnetic field line $\delta s'$ after a time Δt . The length of $\delta s'$ is defined by passing planes

perpendicular to \mathbf{B} through the end points of δs . The length of δs is obviously proportional to the radius of curvature:

$$\frac{\delta s}{R_c} = \frac{\delta s'}{R'_c}$$

so that

$$\frac{\delta s' - \delta s}{\Delta t \delta s} = \frac{R'_c - R_c}{\Delta t R_c}$$

The “radial” component of \mathbf{v}_{gc} is just

$$\mathbf{v}_{gc} \cdot \frac{\mathbf{R}_c}{R_c} = \frac{R'_c - R_c}{\Delta t}$$

The guiding center drift in curved vacuum field is

$$\mathbf{v}_{gc} = \mathbf{v}_c + \mathbf{v}_{\nabla B} = \pm \frac{1}{2} v_{\perp} r_L \frac{\mathbf{B} \times \nabla B}{B^2} + \frac{mv_{\parallel}^2}{q} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2}$$

The curvature drift has no component along \mathbf{R}_c . Using the above three equations, we have

$$\frac{1}{\delta s} \frac{d}{dt} \delta s = \mathbf{v}_{gs} \cdot \frac{\mathbf{R}_c}{R_c^2} = \frac{1}{2} \frac{m}{q} \frac{v_{\perp}^2}{B^3} (\mathbf{B} \times \nabla B) \cdot \frac{\mathbf{R}_c}{R_c^2}$$

This is the rate of change of δs as seen by the particle. We must now get the rate of change of v_{\parallel} as seen by the particle. The parallel and perpendicular energies are defined by

$$W \equiv \frac{1}{2} mv_{\parallel}^2 + \frac{1}{2} mv_{\perp}^2 = \frac{1}{2} mv_{\parallel}^2 + \mu B \equiv W_{\parallel} + W_{\perp}$$

Thus v_{\parallel} can be written

$$v_{\parallel} = \left[\frac{2}{m} (W - \mu B) \right]^{1/2}$$

Here W and μB are constant, and only B varies. Therefore,

$$\frac{\dot{v}_{\parallel}}{v_{\parallel}} = -\frac{1}{2} \frac{\mu \dot{B}}{W - \mu B} = -\frac{1}{2} \frac{\mu \dot{B}}{W_{\parallel}} = -\frac{\mu \dot{B}}{mv_{\parallel}^2}$$

Since B was assumed static, \dot{B} is not zero only because of the guiding center motion:

$$\dot{B} = \frac{dB}{dr} \cdot \frac{dr}{dt} = \mathbf{v}_{gc} \cdot \nabla B = \frac{mv_{\parallel}^2}{q} \frac{\mathbf{R}_c \times \mathbf{B}}{R_c^2 B^2} \cdot \nabla B$$

Now we have

$$\frac{\dot{v}_{\parallel}}{v_{\parallel}} = -\frac{\mu}{q} \frac{(\mathbf{R}_c \times \mathbf{B}) \cdot \nabla B}{R_c^2 B^2} = -\frac{1}{2} \frac{m}{q} \frac{v_{\perp}^2}{B} \frac{(\mathbf{R}_c \times \mathbf{B}) \cdot \nabla B}{R_c^2 B^2}$$

The fractional change in $v_{\parallel}\delta s$ is

$$\frac{1}{v_{\parallel}\delta s} \frac{d}{dt}(v_{\parallel}\delta s) = \frac{1}{\delta s} \frac{d\delta s}{dt} + \frac{1}{v_{\parallel}} \frac{dv_{\parallel}}{dt}$$

From the previous derivations we see that the these two terms cancel, so that

$$v_{\parallel}\delta s = \text{constant}$$

This is not exactly the same as saying that J is constant, however. In taking the integral of $v_{\parallel}\delta s$, between the turning points, it may be that the turning points on $\delta s'$ do not coincide with the intersections of the perpendicular planes. However, any error in J arising from such a discrepancy is negligible because near the turning points, v_{\parallel} is nearly zero. Consequently, we have proved

$$J = \int_a^b v_{\parallel} ds = \text{constant}$$

An example of the violation of J invariance is given by a plasma heating scheme called *transit-time magnetic pumping*. Suppose an oscillating current is applied to the coils of a mirror system so that the mirrors alternately approach and withdraw from each other near the bounce frequency. Those particles that have the right bounce frequency will always see an approaching mirror and will therefore gain v_{\parallel} . J is not conserved in this case because the change of \mathbf{B} occurs on a time scale not long compared with the bounce time.

7.8.6 The Third Adiabatic Invariant

Referring again to the earth dipole case, we see that the slow drift of a guiding center around the earth constitutes a third type of periodic motion. The adiabatic invariant connected with this turns out to be the total magnetic flux Φ enclosed by the drift surface,

$$\Phi = \int \mathbf{B} \cdot d\mathbf{S} \tag{7.47}$$

It is almost obvious that, as \mathbf{B} varies, the particle will stay on a surface such that the total number of magnetic field lines enclosed remains constant. This invariant, Φ , has few applications because most fluctuations of \mathbf{B} occur on a time scale short compared with the drift period. As an example of the violation of Φ invariance, we can cite some recent work on the excitation of ultra-low frequency magnetohydrodynamic waves in the ionosphere. These waves have a long

period comparable to the drift time of a particle around the earth. The particles can therefore encounter the wave in the same phase each time around. If the phase is right, the wave can be excited by the conversion of particle drift energy to wave energy.

Here are some time scales to remember for a 1 MeV electron/ion in the Earth's radiation belt at L shell 4 from Solène Lejosne's [calculator](#):

Type	Gyroperiod	Bounce period	Drift period ⁴
1 MeV electron	0.22 ms	0.22 s	16 mins
1 MeV ion	135 ms	5.5 s	11 mins

In radiation belt studies, it is typical to consider a quantity that is inversely proportional to the magnetic flux:

$$L^* = \frac{2\pi M}{|\phi|R_E} = \frac{2\pi B_{eq}R_E^2}{|\phi|} \quad (7.48)$$

where M is the magnetic moment of the Earth's dipole field, R_E is the radius of the Earth, and B_{eq} is the magnetic field strength at the equatorial surface. This presentation of the third invariant is termed L^* , or Roederer's L parameter, which is the radial distance to the equatorial point where the particle would be found if all nondipolar perturbations in the geomagnetic field were adiabatically turned off leaving only the dipole field (Roederer, 1970). It should be noted that in a nondipolar field L^* does *not* represent a spatial coordinate, instead L^* is a property of a stably trapped particle (Roederer and Lejosne, 2018). That is, in absence of acceleration or loss processes, L^* defines a surface with constant phase space density. If a particle encounter a boundary of the radiation belt system (i.e. the magnetopause or atmosphere) during its drift leading to its loss from the system, the particle does not have an L^* value. The particle is said to be in the drift loss cone.

Therefore, the L and L^* parameters coincide in a dipole field, but in a nondipolar field, these parameters are different. Typically, L is used with particle fluxes whereas L^* is better suited for phase space density. An interesting phenomenon from a nondipolar field is the *drift-shell splitting*, as been proposed as one of the mechanisms for causing the pressure anisotropy in the magnetosphere (Chapter 22).

7.9 Ring Current

One typical application of the single particle motion is the study of ring current. Beyond the lowest order of motion (i.e. gyration combined with parallel drift) of charged particles in the magnetosphere, let us examine the higher order corrections. For the case of non-time-varying

⁴Note that for the same energy at the same L shell, the electron drifts slower than the ion!

fields, and a weak electric field, these corrections consist of a combination of $\mathbf{E} \times \mathbf{B}$ drift, grad- B drift, and curvature drift:

$$\mathbf{v}_{1\perp} = \frac{\mathbf{E} \times \mathbf{B}}{B^2} + \frac{\mu}{m\Omega} \hat{b} \times \nabla B + \frac{v_{\parallel}^2}{\Omega} \hat{b} \times (\hat{b} \cdot \nabla) \hat{b}$$

Let us neglect the $\mathbf{E} \times \mathbf{B}$ drift, since this motion merely gives rise to the convection of plasma within the magnetosphere, without generating a current. By contrast, there is a net current associated with grad- B drift and curvature drift. In the limit in which this current does not strongly modify the ambient magnetic field (i.e., $\nabla \times \mathbf{B} \simeq \mathbf{0}$), which is certainly the situation in the Earth's magnetosphere, we can write

$$(\mathbf{B} \cdot \nabla) \mathbf{B} = -\mathbf{B} \times (\nabla \times \mathbf{B}) \simeq \frac{\nabla_{\perp} B}{B}$$

It follows that the higher order drifts can be combined to give

$$\mathbf{v}_{1\perp} = \frac{(v_{\perp}^2/2 + v_{\parallel}^2)}{\Omega B} \mathbf{B} \times \nabla B$$

For the dipole field Equation 3.6, the above expression yields

$$\mathbf{v}_{1\perp} \simeq -\text{sgn}(\Omega) \frac{6\mathcal{E}L^2}{eB_E R_E} (1 - B/2B_m) \frac{\cos^5 \vartheta (1 + \sin^2 \vartheta)}{(1 + 3\sin^2 \vartheta)^2} \hat{\varphi} \quad (7.49)$$

where \mathcal{E} is the energy of the particle, B_E is the equatorial magnetic field-strength on the Earth's surface, and B_m is the magnetic field-strength at the mirror points. Note that the drift is in the azimuthal direction. A positive drift velocity corresponds to eastward motion, whereas a negative velocity corresponds to westward motion. It is clear that, in addition to their gyromotion and periodic bouncing motion along field lines, charged particles trapped in the magnetosphere also slowly precess around the Earth. The ions drift westwards and the electrons drift eastwards, giving rise to a net westward current circulating around the Earth. This current is known as the *ring current*.

Although the perturbations to the Earth's magnetic field induced by the ring current are small, they are still detectable. In fact, the ring current causes a slight reduction in the Earth's magnetic field in equatorial regions. The size of this reduction is a good measure of the number of charged particles contained in the Van Allen belts (Chapter 24). During the development of so-called geomagnetic storms, charged particles are injected into the Van Allen belts from the outer magnetosphere, giving rise to a sharp increase in the ring current, and a corresponding decrease in the Earth's equatorial magnetic field. These particles eventually precipitate out of the magnetosphere into the upper atmosphere at high latitudes, giving rise to intense auroral activity, serious interference in electromagnetic communications, and, in extreme cases, disruption of electric power grids. The ring current induced reduction in the Earth's magnetic field is measured by the so-called *Dst* index, which is based on hourly

averages of the northward horizontal component of the terrestrial magnetic field recorded at four low-latitude observatories; Honolulu (Hawaii), San Juan (Puerto Rico), Hermanus (South Africa), and Kakioka (Japan). A reduction in the Dst index by 600 nT corresponds to a 2% reduction in the terrestrial magnetic field at the equator.

According to Equation 7.49, the precessional drift velocity of charged particles in the magnetosphere is a rapidly decreasing function of increasing latitude (i.e., most of the ring current is concentrated in the equatorial plane). Since particles typically complete many bounce orbits during a full rotation around the Earth, it is convenient to average Equation 7.49 over a bounce period to obtain the *average* drift velocity. This averaging can only be performed numerically. The final answer is well approximated by

$$\langle v_d \rangle \simeq \frac{6 \mathcal{E} L^2}{e B_E R_E} (0.35 + 0.15 \sin \alpha_{\text{eq}}) \quad (7.50)$$

The average drift period (i.e., the time required to perform a complete rotation around the Earth) is simply

$$\langle \tau_d \rangle = \frac{2\pi L R_E}{\langle v_d \rangle} \simeq \frac{\pi e B_E R_E^2}{3 \mathcal{E} L} (0.35 + 0.15 \sin \alpha_{\text{eq}})^{-1}$$

Thus, the drift period for protons and electrons is

$$\langle \tau_d \rangle_p = \langle \tau_d \rangle_e \simeq \frac{1.05}{\mathcal{E}(\text{MeV}) L} (1 + 0.43 \sin \alpha_{\text{eq}})^{-1} \text{ hours}$$

Note that MeV energy electrons and ions precess around the Earth with about the same velocity, only in opposite directions, because there is no explicit mass dependence in Equation 7.50. It typically takes ~ 20 mins to perform a full rotation. The drift period only depends weakly on the equatorial pitch angle α_{eq} , as is the case for the bounce period. Somewhat paradoxically, the drift period is shorter on more distant L -shells. Note, of course, that particles only get a chance to complete a full rotation around the Earth if the inner magnetosphere remains quiescent on time-scales of order an hour, which is, by no means, always the case.

Note that since the rest mass of an electron is $0.511 \text{ MeV}/c^2$, most of the above formulae require relativistic correction when applied to MeV energy electrons. Fortunately, however, there is no such problem for protons, whose rest mass energy is $0.938 \text{ GeV}/c^2$.

Current theory suggests that the lower latitude (compared with the region I current) region II current is closely linked to the ring current, or more precisely, the partial ring current (substorm current wedge?) on the nightside. The ring current is generated from supra-thermal electrons and ions but not the thermal portion, thus requires extra care on the high energy tail distribution which may not be present in a model that only resolves the thermal part.

7.10 Particle Acceleration

All particle accelerations essentially come from the work of electric field. Classical theory considers adiabatic mechanisms; non-adiabatic mechanisms like interactions with turbulence or wave-particle interactions are added later. One criterion for checking the validity of adiabatic assumption is given as

$$\kappa = \sqrt{R_C/R_L} > \kappa_{\text{set}}$$

where R_C is the radius of curvature, and R_L is the particle gyro-radius. If κ is large, it means that the characteristic length of the changing magnetic field structure is large compared to the gyromotion. In practice $\kappa_{\text{set}} \simeq 3$ is usually used.

There are three basic mechanisms of adiabatic acceleration:

- *Betatron acceleration* refers to situations in which the magnetic field strength increases slowly in time (compared with a gyroperiod, i.e. adiabatic assumption), so that μ remains constant, but the particle kinetic energy is not constant due to the presence of electric fields (associated with the time-varying magnetic field). Then, the perpendicular energy is increased due to constancy of μ . One simple model to test betatron acceleration is to determine the change in energy based on the change in magnetic field.
- *Fermi acceleration*, is the acceleration that charged particles undergo when streaming along and drifting in relaxing curved magnetic field lines. It takes place as particles are repeatedly reflected, usually by a magnetic mirror.⁵ Fermi acceleration is a result of the conservation of the second adiabatic invariant and accelerates particles as the length of flux tubes decrease. Another form of Fermi acceleration seen in the tail comes in “reflections”, which is analogous to the classical system of a ball gaining energy as it bounces off two walls moving towards each other. There are two types of Fermi acceleration: *first-order Fermi acceleration* (in shocks) and *second-order Fermi acceleration* (in the environment of moving magnetized gas clouds) (Axford, Leer, and Skadron 1977). The first-order Fermi acceleration involves particles gaining energy through head-on collisions with moving magnetic structures, like shock waves. The energy gain per collision is proportional to the shock velocity. The second-order Fermi acceleration involves particles gaining energy through random scattering off of moving magnetized clouds. The energy gain per collision is proportional to the square of the cloud velocity, making it less efficient than first-order acceleration. In both cases the environment has to be collisionless with pre-existing medium energy particles in order for the mechanism to be effective. This is because Fermi acceleration only applies to particles with energies exceeding the thermal energies, and frequent collisions with surrounding particles will cause severe energy loss (heating) and as a result no acceleration will occur. See more in [wiki](#) and (Fermi 1949).

⁵Mirror structures trap electrons, which gives further time and space for the Fermi acceleration process. There are observation evidence that high electron acceleration can be associated with the favorable occurrence of mirror modes. Likely, the merging of flux ropes is also favorable for Fermi acceleration and direct reconnected electric field acceleration near the diffusion region.

- *localized electric fields* E_{\parallel} (parallel to the magnetic field), which directly accelerate particles. It is found that direct acceleration by parallel electric fields in structures like dipolarization fronts is smaller than observation error.

It is important to point out the difference between acceleration and heating. When we consider the motion of single particles, there is no such concept as temperature, thus no heating exists. However, acceleration of a bunch of particles can lead to an extension of the velocity distributions, therefore increasing the temperature. In practice, people use the term “acceleration” in the cases of high-energy tails, and “heating” in the cases of the thermal populations.

In the shock scenario, we have the following acceleration types:

- *diffusive shock acceleration* (DSA). DSA is actually considered a subclass of Fermi acceleration, specifically first-order Fermi acceleration. The “diffusive” in DSA refers to the scattering of particles by magnetic turbulence, which allows them to repeatedly cross the shock front.
- *Shock drift acceleration* (SDA). Gradients within the shock transition region may give rise to the phenomenon of Shock Drift Acceleration (SDA), where ions travel along the shock front and gyrate through the ramp multiple times before either passing into the downstream or escaping into the upstream region. This drift is perpendicular to both the magnetic field and the shock normal. The particle gains energy from the electric field as it drifts along the shock front, where the electric field can be generated from a sudden compression of the plasma.
- *Shock surfing acceleration* (SSA). SSA occurs when a charged particle is “trapped” in the electric potential of the shock front. The particle is reflected back and forth across the shock front by the magnetic field, effectively “surfing” on the shock wave. The particle gains energy from the electric field during each reflection.
- *Stochastic shock drift acceleration* (SSDA). SSDA adds to the general SDA theory by introducing a stochastic process where the accelerated electrons are pitch-angle scattered by whistler waves in the shock transition layer, allowing electrons to undergo the SDA process longer and hence gain more energy.

Both SDA and SSA are important mechanisms for pre-accelerating particles, meaning they can inject particles into further acceleration processes like DSA. DSA is the dominant mechanism for producing high-energy cosmic rays. Pitch angle distributions (PADs) are an important tool to distinguish between betatron and Fermi types of acceleration. Since betatron acceleration acts on perpendicular particles, it produces PADs peaked around 90° , sometimes called a “pancake distribution”. On the other hand, Fermi acceleration acts on parallel particles, so it produces a field-aligned and anti-field-aligned distribution, known as a “cigar distribution”.

Under adiabatic assumptions and guiding center approximations, the evolution of a particle energy \mathcal{E} can be written as (Northrop (1963), Dahlin, Drake, and Swisdak (2014))

$$\frac{d\mathcal{E}}{dt} = \frac{\mu}{\gamma} \frac{\partial B}{\partial t} + q(\mathbf{v}_{\parallel} + \mathbf{v}_c + \mathbf{v}_g) \cdot \mathbf{E} \quad (7.51)$$

where $\mu = m\gamma^2 v_{\perp}^2 / (2B)$ is the magnetic moment, γ is the relativistic Lorentz factor, and \mathbf{v}_c and \mathbf{v}_g are the curvature and gradient drift velocities

$$\begin{aligned}\mathbf{v}_c &= \frac{v_{\parallel}^2 \mathbf{b}}{\Omega_c} \times \boldsymbol{\kappa} \\ \mathbf{v}_g &= \frac{v_{\perp}^2 \mathbf{b}}{2\Omega_c} \times \frac{\nabla B}{B}\end{aligned}$$

$\Omega_c = eB/(\gamma mc)$ is the cyclotron frequency, and $\boldsymbol{\kappa} = \mathbf{b} \cdot \nabla \mathbf{b}$ is the curvature. Equation 7.51 can equivalently be written as

$$\frac{d\mathcal{E}}{dt} = qE_{\parallel}v_{\parallel} + \frac{\mu}{\gamma} \left(\frac{\partial B}{\partial t} + \mathbf{u}_E \cdot \nabla B \right) + \gamma m v_{\parallel}^2 (\mathbf{u}_E \cdot \boldsymbol{\kappa}) \quad (7.52)$$

Equation 7.52 describes the acceleration process in terms of ExB drift, where particles accelerate when drifting towards stronger B field or B curvature direction. The first term is the acceleration by the parallel electric field. The second term corresponds to perpendicular heating or cooling due to the conservation of μ . The third term drives parallel acceleration and arises from the first-order Fermi mechanism. For example, newly reconnected field lines downstream from a reconnecting X-line accelerate as a result of the tension force ($\sim B^2 \boldsymbol{\kappa}$) and “straighten”. Particles that reflect from this moving field line receive a Fermi “kick” and thereby gain energy. A particle trapped in an elongated magnetic island whose ends are contracting due to these tension forces will repeatedly gain energy as it reflects from the ends of the island. More generally, any particle reflecting from this moving bent field line will gain energy.

If we integrate all the particles in the phase space from Equation 7.52, the total particle energy change can be written as (non-relativistic case)

$$\frac{d\mathcal{E}_{\text{tot}}}{dt} = E_{\parallel} J_{\parallel} + \frac{p_{\perp}}{B} \left(\frac{\partial B}{\partial t} + \mathbf{u}_E \cdot \nabla B \right) + (p_{\parallel} + mn u_{\parallel}^2) \mathbf{u}_E \cdot \boldsymbol{\kappa} \quad (7.53)$$

The terms on the RHS are the parallel electric field acceleration, betatron acceleration, and Fermi acceleration, respectively. Depending on the drift terms on the RHS of Equation 7.51, additional terms can be added, such as the polarization drift as shown below.

Particle accelerations are important consequences of magnetic reconnection and shocks. In magnetic reconnection processes, the energy is rapidly converted from magnetic field to particles in the forms of flows, bulk (“thermal”) heating, and high-energy (“nonthermal”) tails. (Ji et al. 2022) describes the current understanding of electron acceleration during magnetic reconnection

in the non-relativistic and relativistic regimes. First-principles numerical simulations suggest power-law nonthermal particle populations are more prevalent in systems with low β and moderate guide (perpendicular to the reconnection plane) magnetic fields, with acceleration efficiency enhanced in 3D compared to simplified 2D calculations. Multiple acceleration mechanisms have been studied. In the localized diffusion region near an X-line without a guide field, particles may become demagnetized, prone to be directly accelerated by the reconnection electric field as they undergo so-called *Speiser* or *meandering orbits*. There is recent laboratory evidence for electron acceleration by this mechanism in action during magnetically driven reconnection at low- β using laser-powered capacitor coils. There is also numerical evidence that a substantial fraction of dissipated magnetic energy is carried by particles accelerated by this mechanism when they are demagnetized and escape from plasmoids in 3D.

Over larger scales when the particles remain magnetized, the three basic particle acceleration processes dominate in collisionless reconnection. The energization rate in the non-relativistic limit is given approximately by the following equation:

$$\frac{d\mathcal{E}}{dt} = qE_{\parallel}v_{\parallel} + \mu\frac{dB}{dt} + q\mathbf{E} \cdot \mathbf{u}_c + \frac{1}{2}m\frac{d}{dt}|\mathbf{u}_E|^2 \quad (7.54)$$

where \mathcal{E} is the energy of a particle, \mathbf{u}_E is the $\mathbf{E} \times \mathbf{B}$ drift velocity, $\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}_E \cdot \nabla$, E_{\parallel} is the parallel electric field, and v_{\parallel} is the drift-corrected guiding center parallel velocity. For slowly varying fields, \mathbf{u}_c is the curvature drift of particles and reduces to $\mathbf{u}_c \sim (mv_{\parallel}^2/qB)(\hat{b} \times \boldsymbol{\kappa})$, where $\boldsymbol{\kappa} = \hat{b} \cdot \nabla \hat{b}$ is the magnetic field curvature and \hat{b} is the unit vector in the direction of \mathbf{B} . The terms on the right-hand side represent energy gain by the parallel electric field, betatron heating, Fermi acceleration, and the polarization drift respectively. The first three mechanisms are sketched in [?@fig-particle-acceleration-reconnection](#). The polarization drift is particularly important for lower-energy ions, and this term accounts for the kinetic energy gain in the large-scale reconnection outflow jets. Although Equation 7.54 is valid in the non-relativistic limit, the same basic acceleration mechanisms are responsible for energizing particles in more extreme relativistic contexts relevant to astrophysical systems.

Particles gain energy when their drift induced by the curvature of the magnetic field lines is aligned with the reconnection electric field, which is a form of Fermi heating. Fermi acceleration may resolve two major challenges for heating models. First, it may operate over large volumes in the presence of many magnetic islands or plasmoids, which may be sampled by electrons in the presence of 3D ergodic field lines. There are several space observations of energetic electrons near magnetic islands or plasmoids. Second, the Fermi process naturally generates power laws because a particle undergoing Fermi acceleration gains energy at a rate proportional to its energy, $\delta\mathcal{E} \propto \mathcal{E}$. In the non-relativistic limit, adiabatic particle energization by Fermi acceleration along with the other heating mechanisms for magnetized particles leads to a Parker-like transport equation for the distribution $f(v_{\parallel}, v_{\perp})$ of particles within a magnetic flux tube including time-varying plasma density n and magnetic field \mathbf{B} :

$$\frac{\partial f}{\partial t} + \frac{\dot{B}}{B}v_{\perp}^2\frac{\partial f}{\partial v_{\perp}^2} + \left(\frac{\dot{n}}{n} - \frac{\dot{B}}{B}\right)v_{\parallel}\frac{\partial f}{\partial v_{\parallel}} = \frac{D\{f\}}{\tau_{\text{diff}}} - \frac{f}{\tau_{\text{esc}}} + \frac{f_{\text{inj}}}{\tau_{\text{inj}}} \quad (7.55)$$

where $\frac{D\{f\}}{\tau_{\text{diff}}}$ includes diffusive processes such as pitch-angle or energy scattering, τ_{esc} is a typical time for particles to escape the reconnection geometry, and $f_{\text{inj}}/\tau_{\text{inj}}$ is a source of newly injected particles. This transport equation leads to power-law solutions under certain assumptions, though the Fermi process becomes much less effective at generating power-law nonthermal tails in the presence of a significant guide magnetic field component. With certain assumptions, it is possible to solve the Parker equation using the velocity and magnetic fields from MHD simulations to obtain predictions of the particle acceleration, but this approach does not include the self-consistent feedback.

7.11 Test Particle Model

The test particle method is not self-consistent, because we only describe the effect of the fields onto particles, but not vice versa. This reduces the PDEs to ODEs which is much simpler to solve. Despite of this, we use the trajectories of test particles to infer approximate kinetic properties of the system. You will see many examples of understanding more complicated scenario using the drifts derived from test particle motions, especially in terms of stability analysis.

There are four methods in test particle modelling:

1. Trajectory sampling. It solves individual particle trajectories, but it may not be trivial to select the samples.
2. Forward Monte Carlo. We inject particles in source regions where f is known, and follow them until they reach the regions of interest. This is similar to PIC, except that the fields are not self-consistent.
3. *Forward Liouville*.
4. *Backward Liouville*. It also makes use of the Liouville's theorem for f , but there is no sampling, neither in \mathbf{x} nor in \mathbf{v} space, implying no statistical errors. The procedure starts by choosing a given point \mathbf{x} in space ...

8 Plasmas as Fluid

In a plasma the situation is much more complicated than that single particle orbits; the **E** and **B** fields are not prescribed but are determined by the positions and motions of the charges themselves. One must solve a self-consistent problem; that is, find a set of particle trajectories and field patterns such that the particles will generate the fields as they move along their orbits and the fields will cause the particles to move in those exact orbits. And this must be done in a time-varying situation. It sounds very hard, but it is not.

We have seen that a typical plasma density might be 10^{18} ion–electron pairs per m³. If each of these particles follows a complicated trajectory and it is necessary to follow each of these, predicting the plasma’s behavior would be a hopeless task. Fortunately, this is not usually necessary because, surprisingly, the majority—perhaps as much as 80%—of plasma phenomena observed in real experiments can be explained by a rather crude model. This model is that used in fluid mechanics, in which the identity of the individual particle is neglected, and only the motion of fluid elements is taken into account. Of course, in the case of plasmas, the fluid contains electrical charges. In an ordinary fluid, frequent collisions between particles keep the particles in a fluid element moving together. It is surprising that such a model works for plasmas, which generally have infrequent collisions. But we shall see that there is a reason for this.

A more refined treatment—the kinetic theory of plasmas—requires more mathematical calculation. An introduction to kinetic theory is given in Chapter 12.

In some plasma problems, neither fluid theory nor kinetic theory is sufficient to describe the plasma’s behavior. Then one has to fall back on the tedious process of following the individual trajectories. Brute-force computer simulation can play an important role in filling the gap between theory and experiment in those instances where even kinetic theory cannot come close to explaining what is observed.

8.1 Definitions

Variable	Description
m_s	mass
q_s	charge
n_s	number density

Variable	Description
ρ_s	mass density
σ_s, ρ^*	charge density ($= n_s q_s$)
T_s	temperature
p_s	scalar pressure
\mathbf{u}_s	flow velocity
\mathbf{J}_s	current density ($= \rho_s^* \mathbf{u}_s$)
e_s	total energy density ($= \frac{\rho_s \mathbf{u}_s^2}{2} + \frac{p_s}{\gamma-1}$)

where s denotes the species (e.g. H^+ , O^+). Do not confuse σ here with conductivity. Then the total quantities without subscripts can be written as

$$\begin{aligned}
n &= \sum_s n_s \\
\rho &= \sum_s \rho_s = \sum_s m_s n_s \\
p &= \sum_s p_s \\
T &= \sum_s \frac{n_s}{n} T_s \\
\mathbf{u} &= \frac{1}{\rho} \sum_s \rho_s \mathbf{u}_s \\
\mathbf{v}_s &= \mathbf{u}_s - \mathbf{u}, \quad \text{relative velocity of the } s^{\text{th}} \text{ species} \\
\sigma &= \sum_s \sigma_s = \sum_s q_s n_s \\
\mathbf{J} &= \sum_s \mathbf{J}_s = \sum_s \sigma_s \mathbf{u}_s = \sum_s \sigma_s \mathbf{v}_s + \mathbf{u} \sum_s \sigma_s = \sum_s \sigma_s \mathbf{v}_s + \sigma \mathbf{u} = \mathbf{J}_{cd} + \mathbf{J}_{cv} \\
&\text{where } \mathbf{J}_{cd} = \sum_s \sigma_s \mathbf{v}_s \text{ is the conduction current density} \\
&\mathbf{J}_{cv} = \sigma \mathbf{u} \text{ is the convection current density} \\
e &= \frac{1}{\rho} \sum_s \rho_s e_s + e_f = \frac{\rho u^2}{2} + \frac{p}{\gamma-1} + e_f \text{ (kinetic + internal + field energy)}
\end{aligned}$$

It can be easily verified that

$$\sum_s \rho_s \mathbf{v}_s = 0$$

In this note we tried to use \mathbf{v} for all species/particle-based velocities, and \mathbf{u} for all bulk velocities.

We have 5 independent unknown quantities for each species: n_j, \mathbf{u}_j, p_j . The pressure can be replaced by temperature T_j . Additionally, we have the 6 EM field quantities: \mathbf{B}, \mathbf{E} . For a

2-fluid despcription with ions and electrons, altogether we have $5 * 2 + 6 = 16$ unknowns, so we need 16 equations to determine the system.

8.2 Relation of Plasma to Ordinary Electromagnetics

8.2.1 Maxwell's Equations

In vacuum:

$$\begin{aligned}\epsilon_0 \nabla \cdot \mathbf{E} &= \sigma \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{B} &= \mu_0(\mathbf{J} + \epsilon_0 \dot{\mathbf{E}})\end{aligned}\tag{8.1}$$

In a medium:

$$\begin{aligned}\nabla \cdot \mathbf{D} &= \sigma \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{H} &= \mathbf{J} + \dot{\mathbf{D}} \\ \mathbf{D} &= \epsilon \mathbf{E} \\ \mathbf{B} &= \mu \mathbf{H}\end{aligned}\tag{8.2}$$

σ and \mathbf{J} stand for the “free” charge and current densities. The “bound” charge and current densities arising from polarization and magnetization of the medium are included in the definition of the quantities \mathbf{D} and \mathbf{H} in terms of ϵ and μ . In a plasma, the ions and electrons comprising the plasma are the equivalent of the “bound” charges and currents. Since these charges move in a complicated way, it is impractical to try to lump their effects into two constants ϵ and μ . Consequently, in plasma physics, one generally works with the vacuum equations, in which σ and \mathbf{J} include *all* the charges and currents, both external and internal.

Note that we have used \mathbf{E} and \mathbf{B} in the vacuum equations rather than their counterparts \mathbf{D} and \mathbf{H} , which are related by the constants ϵ_0 and μ_0 . This is because the forces $q\mathbf{E}$ and $\mathbf{J} \times \mathbf{B}$ depend on \mathbf{E} and \mathbf{B} rather than \mathbf{D} and \mathbf{H} , and it is not necessary to introduce the latter quantities as long as one is dealing with the vacuum equations.

8.2.2 Classical Treatment of Magnetic Materials

Since each gyrating particle has a magnetic moment, it would seem that the logical thing to do would be to consider a plasma as a magnetic material with a permeability μ_m .¹ To see why this is *not* done in practice, let us review the way magnetic materials are usually treated.

¹We have put a subscript m on the permeability to distinguish it from the adiabatic invariant μ .

The ferromagnetic domains, say, of a piece of iron have magnetic moments μ_i , giving rise to a bulk magnetization

$$\mathbf{M} = \frac{1}{V} \sum_i \mu_i$$

per unit volume. This has the same effect as a bound current density equal to

$$\mathbf{J}_b = \nabla \times \mathbf{M}$$

In the vacuum Ampère's law, we must include in \mathbf{J} both this current and the "free", or externally applied, current \mathbf{J}_f :

$$\mu_0^{-1} \nabla \times \mathbf{B} = \mathbf{J}_f + \mathbf{J}_b + \epsilon_0 \dot{\mathbf{E}}$$

We wish to write this equation in the simple form

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \epsilon_0 \dot{\mathbf{E}}$$

by including \mathbf{J}_b in the definition of \mathbf{H} . This can be done if we let

$$\mathbf{H} = \mu_0^{-1} \mathbf{B} - \mathbf{M}$$

To get a simple relation between \mathbf{B} and \mathbf{H} , we assume \mathbf{M} to be proportional to \mathbf{B} or \mathbf{H} :

$$\mathbf{M} = \chi_m \mathbf{H}$$

The constant χ_m is the magnetic susceptibility. We now have

$$\mathbf{B} = \mu_0 (1 + \chi_m) \mathbf{H} \equiv \mu_m \mathbf{H}$$

This simple relation between \mathbf{B} and \mathbf{H} is possible because of the linear relation between \mathbf{M} and \mathbf{H} .

In a plasma with a magnetic field, each particle has a magnetic moment μ_α , and the quantity \mathbf{M} is the sum of all these μ_α 's in 1 m^3 . But now we have

$$\mu_\alpha = \frac{mv_{\perp\alpha}^2}{2B} \propto \frac{1}{B} \quad \mathbf{M} \propto \frac{1}{B}$$

The relation between \mathbf{M} and \mathbf{H} (or \mathbf{B}) is no longer linear, and we cannot write $\mathbf{B} = \mu_m \mathbf{H}$ with μ_m constant. It is therefore not useful to consider a plasma as a magnetic medium.

8.2.3 Classical Treatment of Dielectrics

The polarization \mathbf{P} per unit volume is the sum over all the individual moments \mathbf{p}_i of the electric dipoles. This gives rise to a bound charge density

$$\sigma_b = -\nabla \cdot \mathbf{P} \quad (8.3)$$

In the vacuum equation, we must include both the bound charge and the free charge:

$$\epsilon_0 \nabla \cdot \mathbf{E} = \sigma_f + \sigma_b$$

We wish to write this in the simple form

$$\nabla \cdot \mathbf{D} = \sigma_f$$

by including σ_b in the definition of \mathbf{D} . This can be done by letting

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \equiv \epsilon \mathbf{E}$$

If \mathbf{P} is linearly proportional to \mathbf{E} ,

$$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$$

then ϵ is a constant given by

$$\epsilon = (1 + \chi_e) \epsilon_0$$

There is no a priori reason why a relation like the above cannot be valid in a plasma, so we may proceed to try to get an expression for ϵ in a plasma.

8.2.4 The Dielectric Constant of a Plasma

We have seen in Section 7.4 that a fluctuating \mathbf{E} field gives rise to a polarization current \mathbf{J}_p . This leads, in turn, to a polarization charge given by the equation of continuity:

$$\frac{\partial \sigma_p}{\partial t} + \nabla \cdot \mathbf{J}_p = 0$$

This is the equivalent of Equation 8.3, except that, as we noted before, a polarization effect does not arise in a plasma unless the electric field is time varying. Since we have an explicit expression for \mathbf{J}_p but not for σ_p , it is easier to work with the Ampère's law:

$$\nabla \times \mathbf{B} = \mu_0 (\mathbf{J}_f + \mathbf{J}_p + \epsilon \dot{\mathbf{E}})$$

We wish to write this in the form

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{J}_f + \epsilon \dot{\mathbf{E}})$$

This can be done if we let

$$\epsilon = \epsilon_0 + \frac{j_p}{\dot{E}}$$

From Equation 7.12 for \mathbf{J}_p , we have

$$\epsilon = \epsilon_0 + \frac{\rho}{B^2} \quad \text{or} \quad \epsilon_R \equiv \frac{\epsilon}{\epsilon_0} = 1 + \frac{\mu_0 \rho c^2}{B^2} \quad (8.4)$$

This is the *low-frequency plasma dielectric constant for transverse motions*. The qualifications are necessary because our expression for \mathbf{J}_p is valid only for $\omega^2 \ll \omega_c^2$ and for \mathbf{E} perpendicular to \mathbf{B} . The general expression for ϵ , of course, is very complicated and hardly fits on one page.

Note that as $\rho \rightarrow 0$, ϵ_R approaches its vacuum value, unity, as it should. As $B \rightarrow \infty$, ϵ_R also approaches unity. This is because the polarization drift \mathbf{v}_p then vanishes, and the particles do not move in response to the transverse electric field. In a usual laboratory plasma, the second term in Equation 8.4 is large compared with unity. For instance, if $n = 10^{16} \text{ m}^{-3}$ and $B = 0.1 \text{ T}$ we have (for hydrogen)

$$\frac{\mu_0 \rho c^2}{B^2} = \frac{(4\pi \times 10^{-7})(10^{16})(1.67 \times 10^{-27})(3 \times 10^8)^2}{(0.1)^2} = 189$$

This means that the electric fields due to the particles in the plasma greatly alter the fields applied externally. A plasma with large ϵ shields out alternating fields, just as a plasma with small λ_D shields out dc fields.

8.3 Fluid Equations

In (F. F. Chen 2016), this is introduced in the diffusion chapter 5.7, which is a bit weird. In (Bellan 2008), this is introduced in chapter 2 where Vlasov equation is first derived and then follows the simplifications which lead to 2-fluid and MHD.

8.3.1 Equation of Continuity

The integral form of mass conservation for each species is

$$\frac{d}{dt} \int_V \rho_s dx^3 = 0$$

The conservation of matter requires that the total number of particles N_s in a volume V can change only if there is a net flux of particles across the surface S bounding that volume. Since the particle flux density is $n_s \mathbf{u}_s$, we have, by the divergence theorem,

$$\frac{\partial N_s}{\partial t} = \int_V \frac{\partial n_s}{\partial t} dV = - \oint n \mathbf{u}_s \cdot d\mathbf{S} = - \int_V \nabla \cdot (n_s \mathbf{u}_s) dV$$

Since this must hold for any volume V, the integrands must be equal:

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{u}_s) = 0 \quad (8.5)$$

There is one such *equation of continuity* for each species. Any sources or sinks of particles are to be added to the right-hand side.

8.3.2 Momentum Equation

Maxwell's equations tell us what \mathbf{E} and \mathbf{B} are for a given state of the plasma. To solve the self-consistent problem, we must also have an equation giving the plasma's response to given \mathbf{E} and \mathbf{B} . In the fluid approximation, we consider the plasma to be composed of two or more interpenetrating fluids, one for each species. In the simplest case, when there is only one species of ion, we shall need two equations of motion, one for the positively charged ion fluid and one for the negatively charged electron fluid. In a partially ionized gas, we shall also need an equation for the fluid of neutral atoms. The neutral fluid will interact with the ions and electrons only through collisions. The ion and electron fluids will interact with each other even in the absence of collisions, because of the \mathbf{E} and \mathbf{B} fields they generate.

The equation of motion for a single particle is

$$m \frac{d\mathbf{v}}{dt} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (8.6)$$

Assume first that there are no collisions and no thermal motions. Then all the particles in a fluid element move together, and the average velocity \mathbf{u} of the particles in the element is the

same as the individual particle velocity \mathbf{v} . The fluid equation is obtained simply by multiplying Equation 8.6 by the density n :

$$mn \frac{d\mathbf{u}}{dt} = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \quad (8.7)$$

This is, however, not a convenient form to use. In Equation 8.6, the time derivative is to be taken *at the position of the particles*. On the other hand, we wish to have an equation for fluid elements *fixed in space*, because it would be impractical to do otherwise. Consider a drop of cream in a cup of coffee as a fluid element. As the coffee is stirred, the drop distorts into a filament and finally disperses all over the cup, losing its identity. A fluid element at a fixed spot in the cup, however, retains its identity although particles continually go in and out of it.

To make the transformation to variables in a fixed frame, consider $\mathbf{G}(x, t)$ to be any property of a fluid in one-dimensional x space. The change of \mathbf{G} with time *in a frame moving with the fluid* is the sum of two terms:

$$\frac{d\mathbf{G}(x, t)}{dt} = \frac{\partial \mathbf{G}}{\partial t} + \frac{\partial \mathbf{G}}{\partial x} \frac{\partial x}{\partial t} = \frac{\partial \mathbf{G}}{\partial t} + u_x \frac{\partial \mathbf{G}}{\partial x}$$

The first term on the right represents the change of \mathbf{G} at a fixed point in space, and the second term represents the change of \mathbf{G} as the observer moves with the fluid into a region in which \mathbf{G} is different. In three dimensions, this generalizes to

$$\frac{d\mathbf{G}}{dt} = \frac{\partial \mathbf{G}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{G}$$

This is called the *convective derivative* (Equation 3.2).

In the case of a plasma, we take \mathbf{G} to be the fluid velocity \mathbf{u} and write Equation 8.7 as

$$mn \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

where $\partial \mathbf{u} / \partial t$ is the time derivative in a fixed frame.

Stress Tensor \rightarrow scalar pressure

$$mn \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \nabla p \quad (8.8)$$

What we have shown here is only a special case: the transfer of x momentum by motion in the x direction; and we have assumed that the fluid is isotropic, so that the same result holds in the y and z directions. But it is also possible to transfer y momentum by motion in the x direction, for instance. This shear stress cannot be represented by a scalar p but must be given by a tensor \mathbf{P} , the stress tensor, whose components $P_{ij} = mn\bar{v}_i\bar{v}_j$ specify both the direction of

motion and the component of momentum involved. In the general case the term ∇p is replaced by $-\nabla \cdot \mathbf{P}$.

When the distribution function is an isotropic Maxwellian, \mathbf{P} is written

$$\mathbf{P} = \begin{pmatrix} p & 0 & 0 \\ 0 & p & 0 \\ 0 & 0 & p \end{pmatrix}$$

$-\nabla \cdot \mathbf{P} = \nabla p$. A plasma could have two temperatures T_{\perp} and T_{\parallel} in the presence of a magnetic field. In that case, there would be two pressures p_{\perp} and p_{\parallel} in the presence of a magnetic field. In that case, there would be two pressure $p_{\perp} = nk_B T_{\perp}$ and $p_{\parallel} = nk_B T_{\parallel}$. The stress tensor is then

$$\mathbf{P} = \begin{pmatrix} p_{\perp} & 0 & 0 \\ 0 & p_{\perp} & 0 \\ 0 & 0 & p_{\parallel} \end{pmatrix}$$

where the coordinate of the third row or column is the direction of \mathbf{B} . This is still diagonal and shows isotropy in a plane perpendicular to \mathbf{B} .

In an ordinary fluid, the off-diagonal elements of \mathbf{P} are usually associated with viscosity. When particles make collisions, they come off with an average velocity in the direction of the fluid velocity \mathbf{u} at the point where they made their last collision. This momentum is transferred to another fluid element upon the next collision. This tends to equalize \mathbf{u} at different points, and the resulting resistance to shear flow is what we intuitively think of as viscosity. The longer the mean free path, the farther momentum is carried, and the larger is the viscosity. In a plasma there is a similar effect which occurs even in the absence of collisions. The Larmor gyration of particles (particularly ions) brings them into different parts of the plasma and tends to equalize the fluid velocities there. The Larmor radius rather than the mean free path sets the scale of this kind of collisionless viscosity. It is a *finite-Larmor-radius effect* which occurs in addition to collisional viscosity and is closely related to the \mathbf{v}_E drift in a nonuniform \mathbf{E} field (Equation 7.11).

8.3.2.1 Collisions

If there is a neutral gas, the charged fluid will exchange momentum with it through collisions. The momentum lost per collision will be proportional to the relative velocity $\mathbf{u} - \mathbf{u}_0$, where \mathbf{u}_0 is the velocity of the neutral fluid. If τ , the mean free time between collisions, is approximately constant, the resulting force term can be roughly written as $-mn(\mathbf{u} - \mathbf{u}_0)/\tau$. The equation of motion can be generalized to include anisotropic pressure and neutral collisions as follows:

$$mn \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = qn(\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \nabla \cdot \vec{P} - \frac{mn(\mathbf{u} - \mathbf{u}_0)}{\tau} \quad (8.9)$$

This can also be written as (including the pressure term and other forces)

$$\rho \frac{d\mathbf{u}}{dt} = (\rho^* \mathbf{E} + \mathbf{J} \times \mathbf{B}) - \nabla \cdot \vec{P} + \mathbf{f}_n$$

or, in an equivalent conservative form,

$$\frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla \cdot (\rho\mathbf{u}\mathbf{u}) = (\rho^* \mathbf{E} + \mathbf{J} \times \mathbf{B}) - \nabla \cdot \vec{P} + \mathbf{f}_n$$

Collisions between charged particles have not been included; these will be discussed in section ADD IT!.

8.3.2.2 Comparison with Ordinary Hydrodynamics

Ordinary fluids obey the Navier–Stokes equation

$$\rho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} \right] = -\nabla p + \rho \nu \nabla^2 \mathbf{u} \quad (8.10)$$

This is the same as Equation 8.9 except for the absence of electromagnetic forces and collisions between species (there being only one species). The viscosity term $\rho \nu \nabla^2 \mathbf{u}$, where ν is the kinematic viscosity coefficient, is just the collisional part of $\nabla \cdot \vec{P} - \nabla p$ in the absence of magnetic fields. Equation 8.10 describes a fluid in which there are frequent collisions between particles. Equation 8.9, on the other hand, was derived without any explicit statement of the collision rate. Since the two equations are identical except for the \mathbf{E} and \mathbf{B} terms, can Equation 8.9 really describe a plasma species? The answer is a guarded yes, and the reasons for this will tell us the limitations of the fluid theory. This is extremely important to clarify — there are still quite many people think that collision is assumed for the fluid theory, therefore do not believe that MHD can be used to describe plasmas.

In the derivation of Equation 8.9, we did actually assume implicitly that there were many collisions. This assumption came in the derivation of the pressure tensor (Section 12.4) when we took the velocity distribution to be Maxwellian. Such a distribution generally comes about as the result of frequent collisions. However, this assumption was used only to take the average of v^2 . Any other distribution with the same average would give us the same answer. The fluid theory, therefore, is not very sensitive to deviations from the Maxwellian distribution, but there are instances in which these deviations are important. Kinetic theory must then be used.

There is also an empirical observation by Irving Langmuir which helps the fluid theory. In working with the electrostatic probes which bear his name, Langmuir discovered that the electron distribution function was far more nearly Maxwellian than could be accounted for by the collision rate. This phenomenon, called *Langmuir's paradox*, has been attributed at times

to high-frequency oscillations. There has been no satisfactory resolution of the paradox, but this seems to be one of the few instances in plasma physics where nature works in our favor.

Another reason the fluid model works for plasmas is that the magnetic field, when there is one, can play the role of collisions in a certain sense. When a particle is accelerated, say by an \mathbf{E} field, it would continuously increase in velocity if it were allowed to free-stream. When there are frequent collisions, the particle comes to a limiting velocity proportional to \mathbf{E} . The electrons in a copper wire, for instance, drift together with a velocity $\mathbf{v} = \mu\mathbf{E}$, where μ is the mobility. A magnetic field also limits free-streaming by forcing particles to gyrate in Larmor orbits. The electrons in a plasma also drift together with a velocity proportional to \mathbf{E} , namely, $\mathbf{v}_E = \mathbf{E} \times \mathbf{B}/B^2$. In this sense, a collisionless plasma behaves like a collisional fluid. Of course, particles do free-stream *along* the magnetic field, and the fluid picture is not particularly suitable for motions in that direction. *For motions perpendicular to \mathbf{B} , the fluid theory is a good approximation.*

8.3.3 Equation of State

One more relation is needed to close the system of equations. A complete description requires the equation of energy. However, the simplest way is to use the thermodynamic equation of state relating p to n :

$$p = C\rho^\gamma \quad (8.11)$$

where C is a constant and γ is the ratio of specific heats C_p/C_ν . The term ∇p is therefore given by

$$\frac{\nabla p}{p} = \gamma \frac{\nabla n}{n} \quad (8.12)$$

For isothermal compression, we have

$$\nabla p = \nabla(nk_B T) = k_B T \nabla n$$

so that, clearly, $\gamma = 1$. For adiabatic compression, $k_B T$ will also change, giving γ a value larger than one. If N is the number of degrees of freedom, γ is given by

$$\gamma = (2 + N)/N$$

The validity of the equation of state requires that heat flow be negligible; that is, that thermal conductivity be low. Again, this is more likely to be true in directions perpendicular to \mathbf{B} than parallel to it. Fortunately, most basic phenomena can be described adequately by the crude assumption of the equation of state.

8.4 Two-Fluid Model

Besides the Vlasov theory (Chapter 12), we can apply the simpler but equally powerful 2-fluid model, in which electrons and ions are treated as two different fluids. Depending on the form of the pressure term, we have the 5/6/10-moment equations.

8.4.1 Five-moment

Five-moment refers to (ρ, \mathbf{u}, p) . Under either adiabatic or isothermal assumptions, we can get rid of the pressure/energy equation and replace it with an equation of state. For simplicity, let the plasma have only two species: ions and electrons; extension to more species is trivial. The charge and current densities are then given by

$$\begin{aligned}\sigma &= n_i q_i + n_e q_e \\ \mathbf{J} &= n_i q_i \mathbf{v}_i + n_e q_e \mathbf{v}_e\end{aligned}$$

We shall neglect collisions and viscosity. The complete equation set is:

$$\begin{aligned}\epsilon_0 \nabla \cdot \mathbf{E} &= n_i q_i + n_e q_e \\ \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \nabla \cdot \mathbf{B} &= 0 \\ \mu_0^{-1} \nabla \times \mathbf{B} &= n_i q_i \mathbf{u}_i + n_e q_e \mathbf{u}_e + \epsilon_0 \dot{\mathbf{E}} \\ m_j n_j \left[\frac{\partial \mathbf{u}_j}{\partial t} + (\mathbf{u}_j \cdot \nabla) \mathbf{u}_j \right] &= q_j n_j (\mathbf{E} + \mathbf{u}_j \times \mathbf{B}) - \nabla p_j \quad j = i, e \\ \frac{\partial n_j}{\partial t} + \nabla \cdot (n_j \mathbf{u}_j) &= 0 \quad j = i, e \\ p_j &= C_j n_j^\gamma \quad j = i, e\end{aligned}\tag{8.13}$$

where C_j is a constant and γ is the adiabatic index with the approximations in Table 8.2.

Table 8.2: Equation of state approximations.

Regime	Equation of state	Name
$v_{\text{ph}} \gg v_{ts}$	$p_s \sim n_s^\gamma$	adiabatic
$v_{\text{ph}} \ll v_{ts}$	$p_s = n_s k_B T_s, T_s = \text{const.}$	isothermal

There are 16 scalar unknowns: $n_i, n_e, p_i, p_e, \mathbf{u}_i, \mathbf{u}_e, \mathbf{E}$, and \mathbf{B} . There are apparently 18 scalar equations if we count each vector equation as three scalar equations. However, two of Maxwell's equations are superfluous, since the two of the equations can be recovered from the divergences of the other two. The simultaneous solution of this set of 16 equations in 16 unknowns gives a self-consistent set of fields and motions in the fluid approximation.

8.4.2 Six-moment

For monatomic gases, the six-moment equations for all charged fluids (indexed by s) can be written as:

$$\begin{aligned} \frac{\partial \rho_s}{\partial t} + \nabla \cdot (\rho_s \mathbf{u}_s) &= 0 \\ \frac{\partial(\rho_s \mathbf{u}_s)}{\partial t} + \nabla \cdot [\rho_s \mathbf{u}_s \mathbf{u}_s + p_{s\perp} \mathbf{I} + (p_{s\parallel} - p_{s\perp}) \hat{b} \hat{b}] &= \frac{q_s}{m_s} \rho_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) \\ \frac{\partial p_{s\parallel}}{\partial t} + \nabla \cdot (p_{s\parallel} \mathbf{u}_s) &= -2p_{s\parallel} \hat{b} \cdot (\hat{b} \cdot \nabla) \mathbf{u}_s \\ \frac{\partial p_{s\perp}}{\partial t} + \nabla \cdot (p_{s\perp} \mathbf{u}_s) &= -p_{s\perp} (\nabla \cdot \mathbf{u}_s) + p_{s\perp} \hat{b} \cdot (\hat{b} \cdot \nabla) \mathbf{u}_s \end{aligned} \quad (8.14)$$

The two pressure equations in Equation 8.14 can be combined to give the equation for the average pressure $p = (2p_{\perp} + p_{\parallel})/3$:

$$\frac{\partial p_s}{\partial t} + \nabla \cdot (p_s \mathbf{u}_s) = (p_s - p_{s\parallel}) \hat{b} \cdot (\hat{b} \cdot \nabla) \mathbf{u}_s - \left(p_s - \frac{p_{s\parallel}}{3} \right) \nabla \cdot \mathbf{u}_s \quad (8.15)$$

Alternatively, we can solve for the hydrodynamic energy density $e = \frac{\rho_s \mathbf{u}^2}{2} + \frac{p_s}{\gamma-1}$ for each species:

$$\frac{\partial e_s}{\partial t} + \nabla \cdot [\mathbf{u}_s (e_s + p_s) + \mathbf{u}_s \cdot (p_{s\parallel} - p_{s\perp}) \hat{b} \hat{b}] = \frac{q_s}{m_s} \rho_s \mathbf{u}_s \cdot \mathbf{E}$$

which is more of a conservative form and can be beneficial to get better jump conditions across shock waves. Note, however, that the parallel pressure equation is still solved with the adiabatic assumption, so non-adiabatic heating is not properly captured. In addition, the magnetic energy is not included into the energy density, so the jump conditions are only approximate (I think this is more of a numerical consideration). In general, there can be many more source terms on the right hand sides of the above equations corresponding to gravity, charge exchange, chemical reactions, collisions, etc.

The electric field \mathbf{E} and magnetic field \mathbf{B} are obtained from Maxwell's equations:

$$\begin{aligned} \frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} &= 0 \\ \frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} &= -c^2 \mu_0 \mathbf{J} \\ \nabla \cdot \mathbf{E} &= \frac{\rho_c}{\epsilon_0} \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned}$$

where $\rho_c = \sum_s (q_s/m_s) \rho_s$ is the total charge density and $\mathbf{J} = \sum_s (q_s/m_s) \rho_s \mathbf{u}_s$ is the current density.

The last two equations are constraints on the initial conditions; these are not guaranteed to hold numerically. Classical tricks involves using hyperbolic/parabolic cleaning or a facially-collocated Yee-type mesh.

8.4.3 Five/Ten-moment

(L. Wang et al. 2020)

- Continuity equation for each species

$$\frac{\partial(m_s n_s)}{\partial t} + \frac{\partial(m_s n_s u_{j,s})}{\partial x_j} = 0$$

- Momentum equation for each species

$$n_s m_s \frac{d\mathbf{u}_s}{dt} = n_s q_s (\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \nabla \cdot \vec{P}_s - \mathbf{R}_{s\alpha}$$

or in Einstein's notation,

$$\frac{\partial(m_s n_s u_{i,s})}{\partial t} = n_s q_s (E_i + \epsilon_{ijk} u_{j,s} B_k) - \frac{\partial \mathcal{P}_{ij,s}}{\partial x_j} - R_{i,s\alpha}$$

where ϵ_{ijk} is the Levi-Civita symbol. The moments are defined as

$$\begin{aligned} n_s(\mathbf{x}) &\equiv \int f_s d\mathbf{v} \\ u_{i,s}(\mathbf{x}) &\equiv \frac{1}{n_s(\mathbf{x})} \int v_i f_s d\mathbf{v} \\ \mathcal{P}_{ij,s}(\mathbf{x}) &\equiv m_s \int v_i v_j f_s d\mathbf{v} \end{aligned}$$

with $f_s(\mathbf{x}, \mathbf{v}, t)$ being the phase space distribution function. We will neglect the subscript s hereinafter for convenience. For completeness, \mathcal{P}_{ij} relates to the more familiar thermal pressure tensor

$$P_{ij} \equiv m \int (v_i - u_i)(v_j - u_j) f d\mathbf{v}$$

by

$$\mathcal{P}_{ij} = P_{ij} + nm u_i u_j$$

For simplicity, non-ideal effects like viscous dissipation are neglected. The electric and magnetic fields \mathbf{E} and \mathbf{B} are evolved using Maxwell equations

$$\begin{aligned}\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} &= 0 \\ \frac{\partial \mathbf{E}}{\partial t} - c^2 \nabla \times \mathbf{B} &= -\frac{1}{\epsilon_0} \sum_s q_s n_s \mathbf{u}_s \\ \nabla \cdot \mathbf{B} &= 0 \\ \nabla \cdot \mathbf{E} &= \epsilon_0^{-1} \sum_s n_s q_s\end{aligned}$$

with $c = 1/\sqrt{\mu_0 \epsilon_0}$ being the speed of light.

To close the system, the second order moment \mathcal{P}_{ij} or P_{ij} must be specified. For example, a cold fluid closure simply sets $P_{ij} = 0$, while an isothermal equation of state (EOS) assumes that the temperature is constant. Or, assuming zero heat flux and that the pressure tensor is isotropic, we can write an adiabatic EOS for $P_{ij} = p I_{ij}$,

$$\frac{\partial \epsilon}{\partial t} + \nabla \cdot [(p + \epsilon) \mathbf{u}] = n q \mathbf{u} \cdot \mathbf{E}$$

where

$$\epsilon \equiv \frac{p}{\gamma - 1} + \frac{1}{2} m n \mathbf{u}^2$$

is the total fluid (thermal + kinetic) energy and γ is the adiabatic index, setting to 5/3 for a fully ionized plasma. For a plasma with S species ($s = 1, \dots, S$) this system is closed and has a total of $5S + 6$ equations, and are here referred to as the *five-moment* model. More general models can be obtained by retaining the evolution equations for all six components of the pressure tensor in the so-called *ten-moment* model

$$\frac{\partial \mathcal{P}_{ij,s}}{\partial t} + \frac{\partial \mathcal{Q}_{ijm,s}}{\partial x_m} = n_s q_s u_{i,s} E_j + \frac{q_s}{m_s} \epsilon_{iml} \mathcal{P}_{mj,s} B_l \quad (8.16)$$

where the third moment

$$\mathcal{Q}_{ijm,s}(\mathbf{x}) \equiv m_s \int v_i v_j v_m f_s d\mathbf{v}$$

relates to the heat flux tensor defined in the fluid frame

$$Q_{ijm} \equiv m \int (v_i - u_i)(v_j - u_j)(v_m - u_m) d\mathbf{v}$$

by

$$\mathcal{Q}_{ijm} = Q_{ijm} + u_i \mathcal{P}_{jm} - 2 n m u_i u_j u_m$$

Again, the equations here must be closed by some approximation for the heat-flux tensor. Another option is to include evolution equations for even higher order moments, e.g., the ten independent components of the heat-flux tensor.

Theoretically, the multifluid-Maxwell equations approach the Hall magnetohydrodynamics (MHD) under asymptotic limits of vanishing electron mass ($m_e \rightarrow 0$) and infinite speed of light ($c \rightarrow \infty$). All waves and effects within the two-fluid picture are retained, for example, the light wave, electron and ion inertial effects like the ion cyclotron wave and whistler wave. Particularly, through properly devised heat-flux closures, the ten-moment model could partially capture nonlocal kinetic effects like Landau damping, in a manner similar to the gyrokinetic models.

Although multifluid-Maxwell models provide a more complete description of the plasma than reduced, asymptotic models like MHD, they are less frequently used. The reason for this is the fast kinetic scales involved. Retaining the electron inertia adds plasma-frequency and cyclotron time-scale, while non-neutrality adds Debye length spatial-scales. Further, inclusion of the displacement currents means that EM waves must be resolved when using an explicit scheme. Fortunately, the restrictions due to kinetic scales are introduced *only through the non-hyperbolic source terms* of the momentum equation, the Ampère's law, and the pressure equation. Therefore we may eliminate these restrictions by *updating the source term separately either exactly or using an implicit algorithm* (Note: BATSRUS applies the point-implicit scheme.). This allows larger time steps and leads to significant speedup, especially with realistic electron/ion mass ratios. The speed of light constraint still exists, however, can be greatly relaxed, using reduced values for the speed of light and/or sub-cycling Maxwell equations. Of course, an implicit Maxwell solver, or a reduced set of electromagnetic equations like the Darwin approximation², can also relax the time-step restrictions.

8.4.4 Characteristic wave speeds

The fastest wave speed in the 5/6/10-moment equations is the speed of light c . Theoretically, the characteristic wave speeds shall be consistent and fallback to MHD in the isotropic 5-moment case.

Numerically, however, using c in the numerical fluxes makes the scheme rather diffusive (?). The discretization time step is limited by the Courant-Friedrichs-Lowy (CFL) condition based on the speed of light. In practice, we can reduce the speed of light to a value that is a factor of 2-3 faster than the fastest flow and fast wave speed to speed up the simulation.

8.5 Generalized Ohm's Law

The generalized Ohm's law can be derived from two-fluid equations. Landau once pointed out that the validity of describing plasma behaviors with two-fluid equations is based on the fact that the time scale of ions and electrons to reach an equilibrium Maxwellian distribution is much smaller than that of the heat exchange between species. The Ohm's law is established

²the Darwin approximation ignores light waves by neglecting $\partial\mathbf{E}/\partial t$ in Amperè's law.

upon the collisions between electrons, ions and neutral species, and is used to describe the relation between \mathbf{J} and \mathbf{E} , \mathbf{B} , \mathbf{u} .

8.5.1 Basic definitions and assumptions

Assume plasma consists of electrons and ionized H^+ . We use e , i , a as subscripts for electron, ion and neutral species respectively.

$$\begin{aligned} q_e &= -q_i = -e \\ m_e &\ll m_i \approx m_a \end{aligned}$$

With the definition of ionization degree in Equation 2.1,

$$\alpha = \frac{n_i}{n_i + n_a} \quad (8.17)$$

we have

$$\begin{aligned} \frac{n_i}{n_a} &= \frac{\alpha}{1 - \alpha} \\ \frac{n_i}{2n_i + n_a} &= \frac{\alpha}{1 + \alpha} \\ \frac{n_a}{2n_i + n_a} &= \frac{1 - \alpha}{1 + \alpha} \end{aligned}$$

The quasi-neutrality assumption gives

$$\begin{aligned} \rho_e^* + \rho_i^* &= 0 \Rightarrow n_e = n_i \\ \rho_e = m_e n_e &\ll \rho_i = m_i n_i \end{aligned}$$

Assume a system of ideal gas in thermal equilibrium,

$$\begin{aligned} T_e &= T_i = T_a \\ p &= p_e + p_i + p_a \end{aligned}$$

Using the ionization degree defined in Equation 8.17, we have

$$\begin{aligned} p_e = p_i &= \frac{n_i}{2n_i + n_a} p = \frac{\alpha}{1 + \alpha} p \\ p_a &= \frac{n_a}{2n_i + n_a} p = \frac{1 - \alpha}{1 + \alpha} p \end{aligned}$$

Also, we have

$$v_{e,th} \gg v_{i,th}$$

Assume that the flow velocity is much smaller than the thermal speed,

$$|\mathbf{v}_s| \ll |\mathbf{v}_{s,th}|, \quad s = e, i, a.$$

Typical time scale must satisfy $\tau_0 \gg \tau_{\text{collision}}$.

Finally, we ignore viscosity and non-EM force in the derivation.

The average velocity of protons with respect to bulk velocity of the plasma, the velocity difference between electron and ion, and the average velocity of neutral particles are

$$\begin{aligned}\mathbf{v}_i &= \mathbf{u}_i - \mathbf{u} \\ \mathbf{v}_{ei} &= \mathbf{u}_e - \mathbf{u}_i \\ \mathbf{u}_a &= \mathbf{u} - \frac{n_i}{n_a} \left(\mathbf{v}_i + \frac{m_e}{m_i} \mathbf{v}_{ei} \right)\end{aligned}$$

The average momenta of electrons relative to protons and neutral particles are

$$\begin{aligned}\mathbf{I}_{ei} &= m_e \mathbf{v}_{ei} \\ \mathbf{I}_{ea} &= m_e \left[\left(1 + \frac{n_i m_e}{n_a m_i} \right) \mathbf{v}_{ei} + \left(1 + \frac{n_i}{n_a} \right) \mathbf{v}_i \right] \approx m_e \left[\mathbf{v}_{ei} + \left(\frac{1}{1-\alpha} \right) \mathbf{v}_i \right] \\ \mathbf{I}_{ia} &= m_i (\mathbf{u}_i - \mathbf{u}_a) = m_i (\mathbf{v}_i + \mathbf{u} - \mathbf{u}_a) = \frac{1}{1-\alpha} m_i \mathbf{v}_i + \frac{\alpha}{1-\alpha} m_e \mathbf{v}_{ei}\end{aligned}$$

Note that from Newton's third law, $\Delta \mathbf{I}_{sk} = -\Delta \mathbf{I}_{ks}$.

8.5.2 Equation of motion for each species

8.5.2.1 For electrons

$$\begin{aligned}m_e n_e \frac{d\mathbf{u}_e}{dt} &= m_e n_i \frac{d}{dt} (\mathbf{v}_{ei} + \mathbf{v}_i + \mathbf{u}) \\ &= -\nabla p_e - n_i e [\mathbf{E} + (\mathbf{v}_{ei} + \mathbf{v}_i + \mathbf{u}) \times \mathbf{B}] + \mathbf{f}_e^c\end{aligned} \tag{8.18}$$

where the collisional force

$$\mathbf{f}_e^c = -n_i m_e \mathbf{v}_{ei} (\tau_{ei}^{-1} + \tau_{ea}^{-1}) - \frac{1}{1-\alpha} n_i m_e \mathbf{v}_i \tau_{ea}^{-1} \tag{8.19}$$

Proof:

$\mathbf{f}_{sk}^c \cdot \tau_{sk} = n_s \Delta \mathbf{I}_{sk}$, where τ_{sk} is the average collision interval.

$$\Delta \mathbf{I}_{sk} = -\Delta \mathbf{I}_{ks}$$

Assuming elastic collision, we have

$$\begin{aligned}\Delta \mathbf{I}_{sk} &= -\frac{m_s m_k}{m_s + m_k} \mathbf{v}_{sk} \\ \Rightarrow \Delta \mathbf{I}_{ek} &= -\frac{m_k}{m_e + m_k} m_e \mathbf{v}_{ek} \approx -\mathbf{I}_{ek} \text{(ignore electron mass)} \\ \Delta \mathbf{I}_{ia} &= -\frac{m_a}{m_a + m_i} m_i \mathbf{v}_{ia} \approx -\frac{1}{2} \mathbf{I}_{ia} \text{(ion mass=neutron mass)}\end{aligned}$$

Therefore the collision force of electron is

$$\mathbf{f}_e^c = -n_i \mathbf{I}_{ei} \tau_{ei}^{-1} - n_i \mathbf{I}_{ea} \tau_{ea}^{-1} = -n_i m_e \mathbf{v}_{ei} (\tau_{ei}^{-1} + \tau_{ea}^{-1}) - \frac{1}{1-\alpha} n_i m_e \mathbf{v}_i \tau_{ea}^{-1}$$

The collision force between ion and neutral is

$$\mathbf{f}_{ia}^c = -\frac{1}{2} n_i \mathbf{I}_{ia} \tau_{ia}^{-1}$$

8.5.2.2 For protons

$$m_i n_i \frac{d\mathbf{u}_i}{dt} = m_i n_i \frac{d}{dt}(\mathbf{v}_i + \mathbf{u}) = -\nabla p_i + n_i e [\mathbf{E} + (\mathbf{v}_i + \mathbf{u} \times \mathbf{B})] + \mathbf{f}_i^c \quad (8.20)$$

where the collision force is

$$\mathbf{f}_i^c = \mathbf{f}_{ia}^c + \mathbf{f}_{ie}^c = n_i m_e \mathbf{v}_{ei} \tau_{ei}^{-1} - \frac{\alpha}{2(1-\alpha)} n_i m_e \mathbf{v}_{ei} \tau_{ia}^{-1} - \frac{1}{2(1-\alpha)} n_i m_i \mathbf{v}_i \tau_{ia}^{-1} \quad (8.21)$$

8.5.2.3 For the whole plasma

If we do not consider the effect of other forces on plasma and ignore electron mass ($m_e n_e \ll m_i n_i$),

$$m_i (n_i + n_a) \frac{d\mathbf{u}}{dt} = -\nabla p + \mathbf{J} \times \mathbf{B} = -\nabla p - n_i e \mathbf{v}_{ei} \times \mathbf{B} \quad (8.22)$$

where the total current density is

$$\mathbf{J} = \sum_s n_s q_s \mathbf{u}_s = -n_i e (\mathbf{v}_{ei} + \mathbf{v}_i + \mathbf{u}) + n_i e (\mathbf{v}_i + \mathbf{u}) = -n_i e \mathbf{v}_{ei}$$

Physically we usually think of current in plasma as the velocity difference between electrons and ions.

8.5.3 Preconditions

If an external B-field is present and the conductor is not at rest but moving at velocity \mathbf{u} , then an extra term must be added to account for the current induced by the Lorentz force on the charge carriers:

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

where σ is the electric conductivity.

In the rest frame of the moving conductor this term drops out because $\mathbf{u} = 0$. There is no contradiction because the electric field in the rest frame differs from the E-field in the lab frame: $\mathbf{E}' = \mathbf{E} + \mathbf{u} \times \mathbf{B}$. Electric and magnetic fields are relative and related by the Lorentz transformation.

If the current \mathbf{J} is alternating because the applied voltage or E-field varies in time, then reactance must be added to resistance to account for self-inductance (see electrical impedance). The reactance may be strong if the frequency is high or the conductor is coiled. (hyzhou: This is related to the Ponderomotive force. I have some questions here because the time derivative of velocities are all neglected here, which in turn makes the derivation only applicable to a steady system. The obvious problem then is that the generalized Ohm's law may not hold in a time-dependent system!)

8.5.4 Derivation

Negligible terms:

- $\frac{d\mathbf{v}_e}{dt}$ in Equation 8.18 and $\frac{d\mathbf{v}_i}{dt}$ in Equation 8.20, because $|\mathbf{v}| \ll |\mathbf{v}_{th}|$
- $\frac{d\mathbf{u}}{dt}$ in Equation 8.18. From Equation 8.22 and $m_e \ll m_i$,

$$m_e n_i \frac{d\mathbf{u}}{dt} = \frac{m_e}{m_i} \frac{n_i}{n_i + n_a} (-\nabla p - n_i e \mathbf{v}_{ei} \times \mathbf{B}) \ll 1$$

From Equation 8.18,

$$m_e n_i \frac{d\mathbf{u}}{dt} = -\nabla p_e - n_i e \mathbf{v}_{ei} \times \mathbf{B} - n_i e [\mathbf{E} + (\mathbf{v}_i + \mathbf{u}) \times \mathbf{B}] + \mathbf{f}_e^c$$

The above two equations give

$$\begin{aligned} 0 = & -\nabla p_e - n_i e \mathbf{v}_{ei} \times \mathbf{B} - n_i e [\mathbf{E} + (\mathbf{v}_i + \mathbf{u}) \times \mathbf{B}] \\ & - n_i m_e \mathbf{v}_{ei} (\tau_{ei}^{-1} + \tau_{ea}^{-1}) - \frac{1}{1-\alpha} n_i m_e \mathbf{v}_i \tau_{ea}^{-1} \end{aligned} \tag{8.23}$$

Equation 8.20, Equation 8.22 and Equation 8.23 are the three main equations we need to derive the generalized Ohm's law.

If we define the ratios of gyro-period and collision time as

$$\begin{aligned}\kappa_{ei} &\equiv \frac{1}{\Omega_e \tau_{ei}} = \frac{m_e}{eB} \tau_{ei}^{-1} \\ \kappa_{ea} &\equiv \frac{1}{\Omega_e \tau_{ea}} = \frac{m_e}{eB} \tau_{ea}^{-1} \\ \kappa_{ia} &\equiv \frac{1}{2\Omega_i \tau_{ia}} = \frac{m_i}{2eB} \tau_{ia}^{-1}\end{aligned}$$

and ion current density as

$$\mathbf{J}_i = n_i e \mathbf{v}_i$$

we can write the three main equations above as

$$0 = -\nabla p_e - n_i e(\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \mathbf{J} \times \mathbf{B} - \mathbf{J}_i \times \mathbf{B} + (\kappa_{ei} + \kappa_{ea})B\mathbf{J} - \frac{1}{1-\alpha}\kappa_{ea}B\mathbf{J}_i \quad (8.24)$$

$$\begin{aligned}m_i n_i \frac{d\mathbf{u}}{dt} &= -\nabla p_i + n_i e(\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \mathbf{J}_i \times \mathbf{B} \\ &\quad - (\kappa_{ei} - \frac{\alpha}{1-\alpha} \frac{m_e}{m_i} \kappa_{ia})B\mathbf{J} - \frac{1}{1-\alpha}\kappa_{ia}B\mathbf{J}_i\end{aligned} \quad (8.25)$$

$$\frac{1}{\alpha} m_i n_i \frac{d\mathbf{u}}{dt} = -\nabla p + \mathbf{J} \times \mathbf{B} \quad (8.26)$$

Subtracting Equation 8.25 from Equation 8.26 and Equation 8.24, we get the ion current density as

$$\begin{aligned}\mathbf{J}_i &= \frac{1-\alpha}{(\kappa_{ea} + \kappa_{ia})B} \left[\alpha \nabla p - \nabla(p_e + p_i) + (1-\alpha)\mathbf{J} \times \mathbf{B} + (\kappa_{ea} + \frac{\alpha}{1-\alpha} \frac{m_e}{m_i} \kappa_{ia})B\mathbf{J} \right] \\ &= \frac{1-\alpha}{(\kappa_{ea} + \kappa_{ia})B} \left[\frac{\alpha(\alpha-1)}{\alpha+1} \nabla p + (1-\alpha)\mathbf{J} \times \mathbf{B} + (\kappa_{ea} + \frac{\alpha}{1-\alpha} \frac{m_e}{m_i} \kappa_{ia})B\mathbf{J} \right]\end{aligned}$$

Then we can eliminate $\frac{d\mathbf{u}}{dt}$ and \mathbf{J}_i from the three main equations and get the generalized Ohm's law

$$\begin{aligned}\nabla p_e + \frac{\alpha(\alpha-1)}{\alpha+1} r_e \nabla p &= -n_i e(\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \left[1 - 2r_e(1-\alpha) - \alpha \frac{m_e}{m_i} r_i \right] \mathbf{J} \times \mathbf{B} \\ &\quad + \left[\kappa_{ei} - (1-r_e)\kappa - \frac{\alpha}{1-\alpha} \frac{m_e}{m_i} r_e \kappa_{ia} \right] B\mathbf{J} \\ &\quad + \frac{(1-\alpha)^2}{(\kappa_{ea} + \kappa_{ia})B} \left[\frac{\alpha}{\alpha+1} \nabla p \times \mathbf{B} - \mathbf{J} \times \mathbf{B} \times \mathbf{B} \right]\end{aligned}$$

where

$$r_e = \frac{\kappa_{ea}}{\kappa_{ea} + \kappa_{ia}}, \quad r_i = \frac{\kappa_{ia}}{\kappa_{ea} + \kappa_{ia}}$$

8.5.5 Simplification of Generalized Ohm's Law

- If the mean free path of protons is much smaller than that of electrons, i.e. the cross section of proton is much smaller than that of electron,

$$\begin{aligned}\frac{\tau_{ia}}{\tau_{ea}} &= \frac{l_{ia} \cdot v_{e,th}}{l_{ea} \cdot v_{i,th}} \propto \left(\frac{m_i}{m_e}\right)^{1/2} \cdot \frac{l_{ia}}{l_e a} \ll \left(\frac{m_i}{m_e}\right)^{1/2} \\ \Rightarrow \frac{\kappa_{ea}}{\kappa_{ia}} &= 2 \frac{\Omega_i \tau_{ia}}{\Omega_e \tau_{ea}} \propto 2 \frac{m_e}{m_i} \cdot \frac{\tau_{ia}}{\tau_{ea}} \propto 2 \left(\frac{m_e}{m_i}\right)^{1/2} \cdot \frac{l_{ia}}{l_{ea}} \ll 1 \\ \Rightarrow r_e &= \frac{\kappa_{ea}}{\kappa_{ea} + \kappa_{ia}} \ll 1 \quad r_i = \frac{\kappa_{ia}}{\kappa_{ea} + \kappa_{ia}} \approx 1\end{aligned}$$

Therefore we have

$$\begin{aligned}\nabla p_e &= -n_i e (\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \mathbf{J} \times \mathbf{B} + (\kappa_{ei} + \kappa_{ea}) B \mathbf{J} + \frac{(1-\alpha)^2}{\kappa_{ia} B} \left[\frac{\alpha}{\alpha+1} \nabla p \times \mathbf{B} - \mathbf{J} \times \mathbf{B} \times \mathbf{B} \right] \\ \mathbf{J} &= \frac{n_i e}{(\kappa_{ei} + \kappa_{ea}) B} (\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \frac{1}{(\kappa_{ei} + \kappa_{ea}) B} (\nabla p_e - \mathbf{J} \times \mathbf{B}) \\ &\quad - \frac{(1-\alpha)^2}{\kappa_{ia} (\kappa_{ei} + \kappa_{ea}) B} \left[\frac{\alpha}{\alpha+1} \nabla p \times \mathbf{B} - \mathbf{J} \times \mathbf{B} \times \mathbf{B} \right]\end{aligned}$$

- If EM field is dominant, i.e. EM force \gg thermal pressure gradient, $|\mathbf{J} \times \mathbf{B}| \gg |\nabla p| \sim |\nabla p_e|$, then

$$\mathbf{J} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \frac{\sigma}{n_i e} \mathbf{J} \times \mathbf{B} + \frac{\sigma}{n_i e} \frac{2(1-\alpha)^2 \Omega_i \tau_{ia}}{B} \mathbf{J} \times \mathbf{B} \times \mathbf{B}$$

where σ is the electrical conductivity,

$$\sigma = \frac{n_i e^2}{m_e (\tau_{ei}^{-1} + \tau_{ea}^{-1})}$$

- For a fully ionized plasma, $\alpha = 1 \Rightarrow \kappa_{ea} = \kappa_{ia} = 0$, we have

$$\mathbf{J} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) + \frac{\sigma}{n_i e} (\nabla p_e - \mathbf{J} \times \mathbf{B})$$

where

$$\sigma = \frac{n_i e^2 \tau_{ei}}{m_e}$$

Note that the electron pressure gradient term here shows an amazing effect: it can generate EM field out of nothing! This is one theory in astrophysics which describes the origin of EM field.

- For a fully ionized plasma, if EM field is dominant, we have

$$\mathbf{J} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \frac{\sigma}{n_i e} \mathbf{J} \times \mathbf{B} = \sigma (\mathbf{E} + \mathbf{u} \times \mathbf{B}) - \Omega_e \tau_{ei} \mathbf{J} \times \mathbf{B}$$

5. If there is weak magnetic field and dense plasma, i.e. $n_i \gg 1$, $|\mathbf{B}| \ll 1$,

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B})$$

8.5.6 Discussions

In Section 8.6.1, we will see that a relatively simple form can be obtained by assuming that the electron inertia is much smaller than the Lorentz force, which is a good approximation when considering only the perpendicular velocity to the magnetic field.

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B} + \eta \mathbf{J} + \frac{1}{en} \mathbf{J} \times \mathbf{B} - \frac{1}{en} \nabla \cdot \vec{P}_e + \frac{m_e}{ne^2} \left[\frac{\partial \mathbf{J}}{\partial t} + \nabla \cdot (\mathbf{J} \mathbf{u} + \mathbf{u} \mathbf{J}) \right] \quad (8.27)$$

where the first term on the right-hand side is the *convection term*, the second term is the *resistivity term (conductive term)*, the third term is called the *Hall term*, the fourth term is the *electron pressure term*, and the fifth term is called the *electron inertia term*, since it is proportional to the mass of electrons.

Note that both \mathbf{u} and \mathbf{J} are the first-order moments, with \mathbf{u} being the (weighted) sum of the first-order moment of electrons and ions while \mathbf{J} being the difference between them. The generalized Ohm's law is actually the difference between the electrons' and ions' first-order moment equations. *The generalized Ohm's law is an equation that governs the time evolution of \mathbf{J} .* Also note that Ampère's law, with the displacement current retained, is an equation governing the time evolution of \mathbf{E} . However, in the approximation of the resistive MHD, the time derivative terms $\partial \mathbf{E} / \partial t$ and $\partial \mathbf{J} / \partial t$ are ignored in Ampère's law and Ohm's law, respectively. In this approximation, Ohm's law is directly solved to determine \mathbf{E} and Ampère's law is directly solved to determine \mathbf{J} . [Introduction to plasma physics: with space and laboratory applications, D. A. Gurnett and A. Bhattacharjee.]

$$\underbrace{\mathbf{E}}_a + \underbrace{\mathbf{u} \times \mathbf{B}}_b - \underbrace{\frac{1}{\sigma} \mathbf{J}}_c - \underbrace{\frac{\mathbf{J} \times \mathbf{B}}{en_e}}_d + \underbrace{\frac{1}{n_e e} \nabla (n_e k_B T_e)}_e = 0$$

Denote each term above with a to e. From the single-fluid MHD momentum equation, let ω be the oscillation frequency of the perturbed velocity, and variable in scalar form be the characteristic magnitude of that quantity,

$$\begin{aligned} \rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) &= \mathbf{J} \times \mathbf{B} - \nabla p + \rho \mathbf{g} \\ &\Rightarrow \omega \rho U \sim J B \end{aligned} \quad (8.28)$$

Using and the assumptions of MHD, the relations between each term in generalized Ohm's law are

$$\begin{aligned}\frac{b}{d} &= \frac{UB}{JB/en_e} = \frac{en_e UB}{\omega \rho U} = \frac{en_e B}{\omega n_e m_i} = \frac{\Omega_i}{\omega} \gg 1 \\ \frac{c}{d} &= \frac{J/\sigma}{JB/en_e} = \frac{en_e}{B\sigma} = \frac{en_e}{Be^2 n_e / \nu_{ei} m_e} = \frac{\nu_{ei}}{\Omega_e} \ll 1 \\ \frac{b}{d} &= \frac{U}{J/en_e} \ll 1, \text{ if } \mathbf{J} \text{ is carried by } \mathbf{u}_e\end{aligned}$$

which generates direct contradiction with the MHD assumption that $\mathbf{u} \ll \mathbf{u}_e$ if currents are mostly carried by electrons. (WHAT ABOUT THE PRESSURE TERM SCALING? The gradient implies that it is related to the system size; the pressure implies that it is also related to thermal motion?)

There is an interesting point about the electron pressure term. If we assume an isotropic plasma with $n_i = n_e$ and an adiabatic process $P/n^\gamma = C$ where C is a constant and γ is the adiabatic index, we have

$$\begin{aligned}\mathbf{E}_{\nabla P_e} &= \frac{\nabla P_e}{n} = \frac{\nabla P_i}{n} = \frac{\nabla(Cn^\gamma)}{n} \\ &= C\gamma n^{\gamma-2} \nabla n = C \frac{\gamma}{\gamma-1} \nabla n^{\gamma-1}\end{aligned}\tag{8.29}$$

Equation 8.29 indicates that the associated electric field is a potential field that only relates to density.

8.5.7 Ohm's Law in a Multi-Ion System

$$\begin{aligned}\mathbf{E} &= - \sum_i \mathbf{u}_i \times \mathbf{B} + \eta \mathbf{J} + \frac{1}{en_e} \mathbf{J} \times \mathbf{B} - \frac{1}{en_e} \nabla \cdot \vec{P}_e \\ &\quad + \frac{m_e}{n_e e^2} \left[\frac{\partial \mathbf{J}}{\partial t} + \nabla \cdot (\mathbf{J} \mathbf{u} + \mathbf{u} \mathbf{J}) \right]\end{aligned}\tag{8.30}$$

where the quasi-neutrality holds

$$en_e = \sum_i q_i n_i$$

8.6 Magnetohydrodynamics

Particle motion in the two-fluid system was described by the individual species' mean velocities $\mathbf{u}_e, \mathbf{u}_i$ and by the pressures \vec{P}_e, \vec{P}_i , which provide an accounting for the mean square of the random deviation of the velocity from its average value. Magnetohydrodynamics is an alternate description of the plasma where, instead of using $\mathbf{u}_e, \mathbf{u}_i$ to describe mean motion, two new

velocity variables that are a linear combination of \mathbf{u}_e , \mathbf{u}_i are used. As will be seen below, this means a slightly different definition for pressure must also be used.

The new velocity-like variables are

1. the current density

$$\mathbf{J} = \sum_s n_s q_s \mathbf{u}_s$$

which is essentially the relative velocity between ions and electrons, and

2. the center-of-mass velocity

$$\mathbf{u} = \frac{1}{\rho} \sum_s m_s n_s \mathbf{u}_s$$

where

$$\rho = \sum_s m_s n_s$$

is the total mass density.

Magnetohydrodynamics is primarily concerned with *low-frequency, long-wavelength, magnetic* behavior of the plasma. There are three basic assumptions in MHD:

$$\begin{aligned} \epsilon\omega/4\pi\sigma &\ll 1, \\ (v/c)^2 &\ll 1, \\ \lambda/L &\ll 1, \end{aligned}$$

where ω is the plasma frequency, v is the plasma bulk speed, λ is the average free distance, and L is the system characteristic length.

Standard orderings of ideal MHD can also be written as

$$\epsilon \sim \omega/\Omega_c \sim k\rho$$

where the plasma varies on frequency scales ω small compared to the gyrofrequency Ω_c , and varies on spatial scales $1/k$ long compared to the gyroradius ρ . (Here, ϵ means a “small” value.) Thus it covers phenomenon related to compressional and shear Alfvén waves and instabilities, ion acoustic waves, and ion and electron kinetic effects such as Landau damping. However, it does not include drift-waves or other micro-instabilities because they result from finite Larmor radius (FLR) effects which vanish in the usual MHD ordering.

Single fluid MHD is somehow inconsistent since there is only one velocity. The definition of current using velocity cannot be applied, and the current can only be given by Ampère’s law (without the displacement current), $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$.

Multiplying Equation 8.5 by m_s and summing over species gives the MHD continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{8.31}$$

To obtain an MHD equation of motion we take the first moment of the Vlasov equation and multiply the single species momentum equation by m_s and sum over species,

$$\frac{\partial}{\partial t} \sum_s m_s \int \mathbf{v} f_s d\mathbf{v} + \frac{\partial}{\partial \mathbf{x}} \cdot \sum_s \int m_s \mathbf{v} \mathbf{v} f_s d\mathbf{v} + \sum_s q_s \int \mathbf{v} \frac{\partial}{\partial \mathbf{v}} \cdot [(\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s] = 0 \quad (8.32)$$

The right-hand side is zero since $\mathbf{R}_{ei} + \mathbf{R}_{ie} = 0$, i.e., the total plasma cannot exert drag on itself. We now define random velocities relative to \mathbf{u} (rather than to \mathbf{u}_s as was the case for the two-fluid equations) so that the second term can be written as

$$\sum_s \int m_s \mathbf{v} \mathbf{v} f_s d\mathbf{v} = \sum_s \int m_s (\mathbf{v}' + \mathbf{u})(\mathbf{v}' + \mathbf{u}) f_s d\mathbf{v} = \sum_s \int m_s \mathbf{v}' \mathbf{v}' f_s d\mathbf{v} + \rho \mathbf{u} \mathbf{u} \quad (8.33)$$

where $\sum_s \int m_s \mathbf{v}' f_s d\mathbf{v} = 0$ has been used to eliminate terms linear in \mathbf{v}' . The MHD pressure tensor is now defined in terms of the random velocities relative to \mathbf{u} and is given by

$$\vec{P}^{\text{MHD}} = \sum_s \int m_s \mathbf{v}' \mathbf{v}' f_s d\mathbf{v} \quad (8.34)$$

We insert Equation 8.33 and Equation 8.34 in Equation 8.32, integrate by parts on the acceleration term, and perform the summation over species to obtain the MHD equation of motion

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) = \left(\sum_s n_s q_s \right) \mathbf{E} + \mathbf{J} \times \mathbf{B} - \nabla \cdot \vec{P}^{\text{MHD}} \quad (8.35)$$

MHD is typically used to describe phenomena having spatial scales large enough for the plasma to be essentially neutral, i.e., $\sum_s n_s q_s = 0$ so that the \mathbf{E} term can be dropped. Just as in the two-fluid situation, the left-hand side of Equation 8.35 contains a factor times the MHD continuity equation, since the left-hand side can be expanded as

$$\begin{aligned} \frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u}) &= \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) \right] \mathbf{u} + \rho \frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} \\ &= \rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) \end{aligned}$$

This leads to the standard form for the MHD equation of motion,

$$\rho \frac{D \mathbf{u}}{Dt} = \mathbf{J} \times \mathbf{B} - \nabla \cdot \vec{P}^{\text{MHD}} \quad (8.36)$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$$

is the convective derivative defined using the center-of-mass velocity. This notation is used to emphasize the difference from single species treatment. Scalar approximations of the MHD pressure tensor will be postponed until after discussing the MHD Ohm's law and its implications.

8.6.1 MHD Ohm's law

Equation 8.36 provides one equation relating \mathbf{J} and \mathbf{u} ; let us now find the other one. In order to do this, consider the two-fluid electron equation of motion,

$$m_e \frac{d\mathbf{u}_e}{dt} = -e(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - \frac{1}{n_e} \nabla(n_e k_B T_e) - \nu_{ei} m_e (\mathbf{u}_e - \mathbf{u}_i) \quad (8.37)$$

In MHD we are interested in low-frequency phenomena with large spatial scales. If the characteristic time scale of the phenomenon is long compared to the electron cyclotron motion, then the electron inertia term $m_e d\mathbf{u}_e/dt$ can be dropped since it is small compared to the magnetic force term $-e(\mathbf{u}_e \times \mathbf{B})$. This assumption is reasonable for velocities perpendicular to \mathbf{B} , but can be a poor approximation for the velocity component parallel to \mathbf{B} , since parallel velocities do not provide a magnetic force. Since $\mathbf{u}_e - \mathbf{u}_i = -\mathbf{J}/n_e e$ and $\mathbf{u}_i \simeq \mathbf{u}$, Equation 8.37 reduces to the generalized Ohm's law

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} - \frac{1}{n_e e} \mathbf{J} \times \mathbf{B} + \frac{1}{n_e e} \nabla(n_e k_B T_e) = \eta \mathbf{J} \quad (8.38)$$

(What if we consider ion species with charge larger than 1?) The term $-\mathbf{J} \times \mathbf{B}/n_e e$ on the left-hand side is called the *Hall term* and can be neglected in either of the following two cases:

1. The pressure term in the MHD equation of motion, Equation 8.36 is negligible compared to the other two terms, which therefore must balance, giving

$$|\mathbf{J}| \sim \omega \rho |\mathbf{u}| / |\mathbf{B}|$$

Here $\omega \sim D/Dt$ is the characteristic frequency of the phenomenon. In this case comparison of the Hall term with the $\mathbf{U} \times \mathbf{B}$ term shows that the Hall term is small by a factor $\sim \omega/\omega_{ci}$, where $\omega_{ci} = q_i B/m_i$ is the ion cyclotron frequency. Thus dropping the Hall term is justified for phenomena having characteristic frequencies small compared to ω_{ci} .

2. The electron-ion collision frequency is large compared to the electron cyclotron frequency $\omega_{ce} = q_e B/m_e$, in which case the Hall term may be dropped since it is small by a factor ω_{ce}/ν_{ei} compared to the right-hand side resistive term $\eta \mathbf{J} = (m_e \nu_{ei}/n_e e^2) \mathbf{J}$.

8.6.2 Ideal MHD

If the Hall term is dropped, the system is called ideal MHD. Typically, Equation 8.38 will not be used directly; instead its curl will be used to provide the induction equation

$$-\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}) - \frac{1}{n_e e} \nabla n_e \times \nabla k_B T_e = \nabla \times \left(\frac{\eta}{\mu_0} \nabla \times \mathbf{B} \right)$$

Usually the density gradient is parallel to the temperature gradient so that the thermal electromotive force term $(n_e e)^{-1} \nabla n_e \times \nabla k_B T_e$ can be dropped, or the thermal term is often simply ignored, in which case the induction equation reduces to

$$-\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{u} \times \mathbf{B}) = \nabla \times \left(\frac{\eta}{\mu_0} \nabla \times \mathbf{B} \right) \quad (8.39)$$

or written in the form of Ohm's law

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \eta \mathbf{J} \quad (8.40)$$

If the resistive term $\eta \mathbf{J}$ is so small as to be negligible compared to the other terms in Equation 8.40, then the plasma is said to be *ideal* or *perfectly conducting*. From the Lorentz transformation of electromagnetic theory we realize that

$$\mathbf{E} + \mathbf{u} \times \mathbf{B} = \mathbf{E}'$$

where \mathbf{E}' is the electric field observed in the frame moving with velocity \mathbf{u} . This implies that the magnetic flux in ideal plasmas is time-invariant in the frame moving with velocity \mathbf{u} , because otherwise Faraday's law would imply the existence of an electric field in the moving frame. The frozen-in flux concept is the essential "bed-rock" concept underlying ideal MHD. While this concept is often an excellent approximation, it must be kept in mind that the concept becomes invalid in situations when any one of the electron inertia, electron pressure, or Hall terms becomes important and leads to different, more complex behavior.

The frozen-in flux concept is frequently expressed in a slightly different form as a frozen-in field concept, i.e., it is often said that magnetic field lines are frozen into the plasma so that the plasma and magnetic field lines move together as an ensemble. While this point of view can be quite intuitive and useful, it contains some ambiguity because an ideal plasma can actually move across magnetic field lines in certain situations. These situations are such that magnetic flux is preserved within the plasma even though it is moving across field lines.

A formal proof of the frozen-in flux property will now be established by direct calculation of the rate of change of the magnetic flux through a surface $S(t)$ bounded by a material line $C(t)$, i.e., a closed contour that moves with the plasma. This magnetic flux is

$$\Phi(t) = \int_{S(t)} \mathbf{B}(\mathbf{x}, t) \cdot d\mathbf{s}$$

and the flux changes with respect to time due to either (i) the explicit time dependence of $\mathbf{B}(t)$ or (ii) changes in the surface $S(t)$ resulting from plasma motion. The rate of change of flux is thus

$$\frac{D\Phi}{Dt} = \lim_{\delta t \rightarrow 0} \left(\frac{\int_{S(t+\delta t)} \mathbf{B}(\mathbf{x}, t + \delta t) \cdot d\mathbf{s} - \int_{S(t)} \mathbf{B}(\mathbf{x}, t) \cdot d\mathbf{s}}{\delta t} \right)$$

The displacement of a segment $d\mathbf{l}$ of the bounding contour C is $\mathbf{u}\delta t$, where \mathbf{u} is the velocity of this segment. The incremental change in surface area due to this displacement is $\Delta S = \mathbf{u}\delta t \times d\mathbf{l}$. The rate of change of flux can thus be expressed as

$$\begin{aligned}\frac{D\Phi}{Dt} &= \lim_{\delta t \rightarrow 0} \left(\frac{\int_{S(t+\delta t)} (\mathbf{B} + \delta t \frac{\partial \mathbf{B}}{\partial t}) \cdot d\mathbf{s} - \int_{S(t)} \mathbf{B} \cdot d\mathbf{s}}{\delta t} \right) \\ &= \lim_{\delta t \rightarrow 0} \left(\frac{\int_{S(t)} (\mathbf{B} + \delta t \frac{\partial \mathbf{B}}{\partial t}) \cdot d\mathbf{s} + \oint_C \mathbf{B} \cdot \mathbf{u} \delta t \times d\mathbf{l} - \int_{S(t)} \mathbf{B} \cdot d\mathbf{s}}{\delta t} \right) \\ &= \int_{S(t)} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{s} + \oint_C \mathbf{B} \cdot \mathbf{u} \times d\mathbf{l} \\ &= \int_{S(t)} \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{s} + \oint_C d\mathbf{l} \cdot (\mathbf{B} \cdot \mathbf{u}) \\ &= \int_{S(t)} \left[\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{B} \times \mathbf{u}) \right] \cdot d\mathbf{S}\end{aligned}$$

Thus, if

$$\frac{\partial \mathbf{B}}{\partial t} + \nabla \times (\mathbf{B} \times \mathbf{u}) = 0 \quad (8.41)$$

then

$$\frac{D\Phi}{Dt} = 0$$

so that the magnetic flux linked by any closed material line is constant. Therefore, magnetic flux is frozen into an ideal plasma because Equation 8.39 reduces to Equation 8.41 if $\eta = 0$. Equation 8.41 is called the ideal MHD induction equation.

8.6.2.1 Equation of state

Double adiabatic laws

A procedure analogous to that which led to Equation 8.11 or Equation 12.28 gives the MHD adiabatic relation

$$\frac{p^{\text{MHD}}}{\rho^\gamma} = \text{const.}$$

where again $\gamma = (N + 2)/N$ and N is the number of dimensions of the system. It was shown in the previous section that magnetic flux is conserved in the plasma frame. This means that, as shown in Figure 8.1, a tube of plasma initially occupying the same volume as a magnetic flux tube is constrained to evolve in such a way that $\int \mathbf{B} \cdot d\mathbf{s}$ stays constant over the plasma tube cross-section. For a flux tube of infinitesimal cross-section, the magnetic field is approximately uniform over the cross-section and we may write this as $BA = \text{const}$, where A is the cross-sectional area.

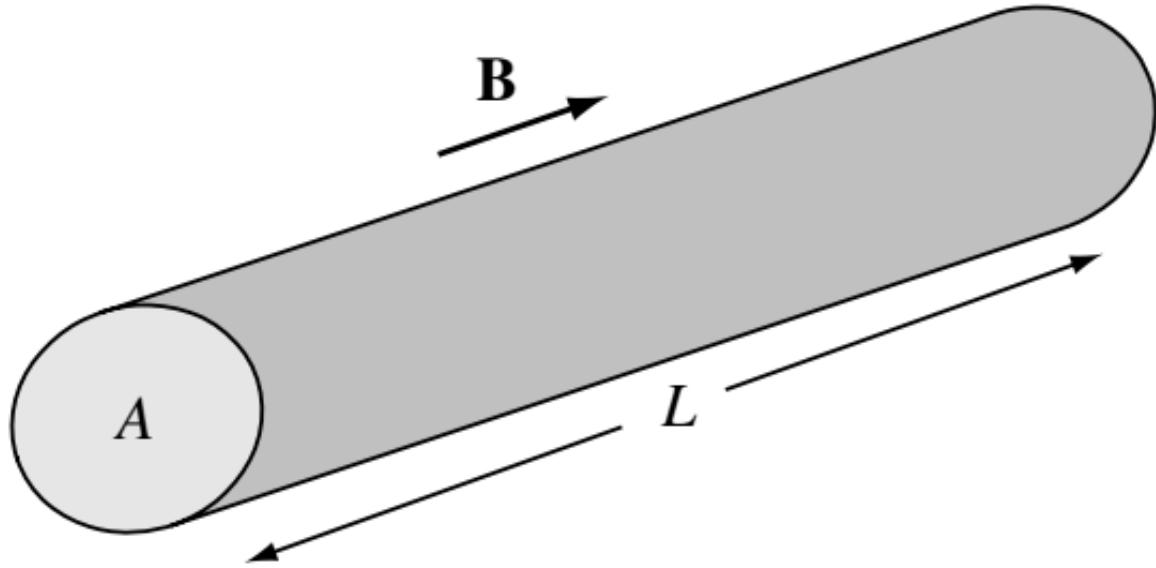


Figure 8.1: Magnetic flux tube with flux $\Phi = BA$.

Let us define two temperatures for this magnetized plasma, namely T_{\perp} , the temperature corresponding to motions perpendicular to the magnetic field, and T_{\parallel} , the temperature corresponding to motions parallel to the magnetic field. If for some reason (e.g., anisotropic heating or compression) the temperature develops an anisotropy such that $T_{\perp} \neq T_{\parallel}$ and if collisions are infrequent, this anisotropy will persist for a long time, since collisions are the means by which the two temperatures equilibrate. Thus, rather than assuming that the MHD pressure is fully isotropic, we consider the less restrictive situation where the MHD pressure tensor is given by

$$\vec{P}^{\text{MHD}} = \begin{bmatrix} p_{\perp} & 0 & 0 \\ 0 & p_{\perp} & 0 \\ 0 & 0 & p_{\parallel} \end{bmatrix} = p_{\perp} \mathbf{I} + (p_{\parallel} - p_{\perp}) \hat{b} \hat{b}$$

The first two coordinates (x, y -like) in the above matrix refer to the directions perpendicular to the local magnetic field \mathbf{B} and the third coordinate (z -like) refers to the direction parallel to \mathbf{B} . The tensor expression on the right-hand side is equivalent (here \mathbf{I} is the unit tensor) but allows for arbitrary, curvilinear geometry. We now develop separate adiabatic relations for the perpendicular and parallel directions:

- Parallel direction: here the number of dimensions is $N = 1$ so that $\gamma = 3$ and so the adiabatic law gives

$$\frac{p_{\parallel}^{1D}}{\rho_{1D}^3} = \text{const.} \quad (8.42)$$

where ρ_{1D} is the *one-dimensional* mass density; i.e., $\rho_{1D} \sim 1/L$, where L is the length along the flux tube in Figure 8.1. The three-dimensional mass density, which has been used implicitly until now, has the proportionality $\rho \sim 1/LA$, where A is the cross-section of the flux tube; similarly the three-dimensional pressure has the proportionality $p_{\parallel} \sim \rho T_{\parallel}$. However, we must be careful to realize that $p_{\parallel}^{1D} \sim \rho_{1D} T_{\parallel}$ so, using $BA = \text{const.}$, Equation 8.42 can be recast as

$$\text{const.} = \frac{p_{\parallel}^{1D}}{\rho_{1D}^3} \sim \frac{\rho_{1D} T_{\parallel}}{\rho_{1D}^3} \sim T_{\parallel} \rho_{1D}^2 \sim \left(\frac{1}{LA} \right) T_{\parallel} (LA)^3 B^2$$

or

$$\frac{p_{\parallel} B^2}{\rho^3} = \text{const.} \quad (8.43)$$

- Perpendicular direction: here the number of dimensions is $N = 2$ so that $\gamma = 2$ and the adiabatic law gives

$$\frac{p_{\perp}^{2D}}{\rho_{2D}^2} = \text{const.} \quad (8.44)$$

where p_{\perp}^{2D} is the 2-D perpendicular pressure, and has dimensions of energy per unit area, while ρ_{2D} is the 2-D mass density and has dimensions of mass per unit area. Thus, $\rho_{2D} \sim 1/A$ so $p_{\perp}^{2D} \sim \rho_{2D} T_{\perp} \sim T_{\perp}/A$, in which case Equation 8.44 can be recast as

$$\text{const.} = \frac{p_{\parallel} B^2}{\rho^3} = T_{\perp} A \sim \left(\frac{1}{LA} \right) T_{\perp} \frac{LA}{B}$$

or

$$\frac{p_{\perp}}{\rho B} = \text{const.} \quad (8.45)$$

Equation 8.43 and Equation 8.45 are called the double adiabatic or CGL laws after Chew, Goldberger, and Low (Chew, Goldberger, and Low 1956) who first developed these laws.

Single adiabatic law

If collisions are sufficiently frequent to equilibrate the perpendicular and parallel temperatures, then the pressure tensor becomes fully isotropic and the dimensionality of the system is $N = 3$ so that $\gamma = 5/3$. There is now just one pressure and one temperature and the adiabatic relation becomes

$$\frac{p}{\rho^{5/3}} = \text{const.} \quad (8.46)$$

8.6.2.2 MHD approximations for Maxwell's equations

The various assumptions contained in MHD lead to a simplifying approximation of Maxwell's equations. In particular, the assumption of charge neutrality in MHD makes Poisson's equation superfluous because Poisson's equation prescribes the relationship between non-neutrality and the electrostatic component of the electric field. The assumption of charge neutrality also has implications for the current density. To see this, the two-fluid continuity equation is multiplied by q_s and then summed over species to obtain the charge conservation equation

$$\frac{\partial}{\partial t} \left(\sum_s n_s q_s \right) + \nabla \cdot \mathbf{J} = 0 \quad (8.47)$$

Thus, charge neutrality implies

$$\nabla \cdot \mathbf{J} = 0$$

Let us now consider Ampère's law

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \dot{\mathbf{E}}$$

Taking the divergence gives

$$\nabla \cdot$$

which is equivalent to Equation 8.47 if Poisson's equation is invoked.

Finally, MHD is restricted to phenomena having characteristic velocities V_{ph} slow compared to the speed of light in vacuum, $c = (\mu_0 \epsilon_0)^{-1/2}$. Again, t_{char} is assumed to represent the characteristic time scale for a given phenomenon and l_{char} is assumed to represent the corresponding characteristic length scale so that $V_{\text{ph}} \sim l_{\text{char}}/t_{\text{char}}$. Faraday's equation gives the scaling

$$\nabla \times \mathbf{E} = -\dot{\mathbf{B}} \rightarrow E \sim Bl_{\text{char}}/t_{\text{char}}$$

On comparing the magnitude of the displacement current term in Ampère's law to the left-hand side it is seen that

$$\frac{\mu_0 \epsilon_0 |\dot{\mathbf{E}}|}{|\nabla \times \mathbf{B}|} \sim \frac{c^{-2} E / t_{\text{char}}}{B / l_{\text{char}}} \sim \left(\frac{V_{\text{ph}}}{c} \right)^2$$

Thus, if $V_{\text{ph}} \ll c$ the displacement current term can be dropped from Ampère's law resulting in the so-called “pre-Maxwell” form (i.e. Darwin approximation)

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \quad (8.48)$$

The divergence of Equation 8.48 gives $\nabla \cdot \mathbf{J} = 0$ so it is unnecessary to specify it separately.

8.6.2.3 Complete ideal MHD equation set

In the most common sense, ideal MHD involves equations of compressible, adiabatic and inviscid fluid:

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) &= 0 \\ \frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot \left(\rho \mathbf{u} \mathbf{u} - \frac{1}{\mu_0} \mathbf{B} \mathbf{B} + \frac{\mathbf{B}^2}{2\mu_0} + p \right) &= 0 \\ \frac{\partial e}{\partial t} + \nabla \cdot \left[(e + p_{\text{tot}}) \mathbf{u} - \frac{1}{\mu_0} \mathbf{B} (\mathbf{B} \cdot \mathbf{u}) \right] &= 0 \\ \frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) &= 0 \\ e &= \frac{\rho \mathbf{u}^2}{2} + \frac{p}{\gamma - 1} + \frac{\mathbf{B}^2}{2\mu_0} \end{aligned} \quad (8.49)$$

where ρ is the mass density, \mathbf{u} the velocity, e the total energy density, \mathbf{B} the magnetic field, p the thermal pressure, $p_{\text{tot}} = p + p_B = p + \mathbf{B}^2/(2\mu_0)$ the total pressure and γ the adiabatic index (ratio of specific heats). Microscopic dissipations of any kind (viscosity, resistivity, or conduction) are not included in ideal MHD. Note that there is only one constant μ_0 appeared in Equation 8.49. The introduction of temperature comes with a new constant $R \equiv 2k_B/m$:

$$p = \rho R T$$

Many often in simulations, the dimensionless units are applied. Equation 8.49 under the dimensionless units can be written as

$$\begin{aligned} \frac{\partial \rho_*}{\partial t_*} + \nabla_* \cdot (\rho_* \mathbf{u}_*) &= 0 \\ \frac{\partial \rho_* \mathbf{u}_*}{\partial t_*} + \nabla_* \cdot \left(\rho_* \mathbf{u}_* \mathbf{u}_* - \mathbf{B}_* \mathbf{B}_* + \frac{\mathbf{B}_*^2}{2} + p_* \right) &= 0 \\ \frac{\partial e_*}{\partial t_*} + \nabla_* \cdot [(e_* + p_*) \mathbf{u}_* - \mathbf{B}_* (\mathbf{B}_* \cdot \mathbf{u}_*)] &= 0 \\ \frac{\partial \mathbf{B}_*}{\partial t_*} - \nabla_* \times (\mathbf{u}_* \times \mathbf{B}_*) &= 0 \\ \mathcal{E}_* &= \frac{\rho_* \mathbf{u}_* \cdot \mathbf{u}_*}{2} + \frac{p_*}{\gamma - 1} + \frac{\mathbf{B}_*^2}{2} \end{aligned} \quad (8.50)$$

where the subscripts $*$ denotes the variables in dimensionless units, and the adiabatic equation of state is given as

$$p_* = \rho_* T_*$$

In more theoretical derivations, for simplicity the energy equation is substituted with either single adiabatic law Equation 8.46 or double adiabatic law Equation 8.43 and Equation 8.45.³

³In the case of discontinuities, adiabatic laws fail to describe the jump, and the energy equation needs to be used due to its conservation property.

If we drop the isotropic assumption and instead use separate pressure equations for the parallel and perpendicular components as in Equation 8.14, we get the anisotropic MHD equations:⁴

$$\begin{aligned} \frac{\partial \rho_s}{\partial t} + \nabla \cdot (\rho_s \mathbf{u}_s) &= 0 \\ \frac{\partial(\rho_s \mathbf{u}_s)}{\partial t} + \nabla \cdot \left[\rho_s \mathbf{u}_s \mathbf{u}_s + p_{s\perp} \mathbf{I} + (p_{s\parallel} - p_{s\perp}) \hat{b} \hat{b} - \frac{1}{\mu_0} \left(\mathbf{B} \mathbf{B} - \frac{B^2}{2} \mathbf{I} \right) \right] &= 0 \\ \frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u}_s \times \mathbf{B}) &= 0 \\ \frac{\partial p_{s\parallel}}{\partial t} + \nabla \cdot (p_{s\parallel} \mathbf{u}_s) + 2p_{s\parallel} \hat{b} \cdot (\hat{b} \cdot \nabla) \mathbf{u}_s &= \frac{\delta p_{s\parallel}}{\delta t} \\ \frac{\partial p_s}{\partial t} + \nabla \cdot (p_s \mathbf{u}_s) + \frac{2}{3} p_{s\perp} (\nabla \cdot \mathbf{u}_s) + \frac{2}{3} (p_{\parallel} - p_{\perp}) \hat{b} \cdot (\hat{b} \cdot \nabla) \mathbf{u}_s &= 0 \end{aligned} \quad (8.51)$$

For capturing jump conditions across a discontinuity, the conservative energy equation is needed to substitute the total pressure equation:

$$\frac{\partial e}{\partial t} + \nabla \cdot \left[\mathbf{u} \left(e + p_{\perp} + \frac{\mathbf{B}^2}{2\mu_0} + \mathbf{u} \cdot \left((p_{\parallel} - p_{\perp}) \hat{b} \hat{b} - \frac{\mathbf{B} \mathbf{B}}{2\mu_0} \right) \right) \right] \quad (8.52)$$

with total energy density

$$e = \frac{\rho \mathbf{u}^2}{2} + \frac{p}{\gamma - 1} + \frac{\mathbf{B}^2}{2\mu_0} \quad (8.53)$$

The plasma instabilities associated with anisotropic pressure limit, i.e. firehose, mirror, and ion cyclotron instability may contain kinetic effects that cannot be fully described by MHD (Chapter 14). Moreover, the grid resolution in normal global MHD simulations may not be fine enough to resolve even hydromagnetic instabilities.

- Criterion for the firehose instability

$$\frac{p_{\parallel}}{p_{\perp}} > 1 + \frac{\mathbf{B}^2}{\mu_0 p_{\perp}} \quad (8.54)$$

- Criterion for the mirror instability

$$\frac{p_{\perp}}{p_{\parallel}} > 1 + \frac{\mathbf{B}^2}{2\mu_0 p_{\parallel}} \quad (8.55)$$

⁴This is slightly different in some terms compared to Equation 8.14. Meng Xing mentioned in the BATSRUS anisotropic MHD paper that the polytropic index is taken to be 5/3. Also note that here the unknown is the total pressure p instead of perpendicular pressure p_{\perp} .

- Criterion for the ion cyclotron instability

$$\frac{p_{\perp}}{p_{\parallel}} > 1 + C_1 \left(\frac{\mathbf{B}^2}{2\mu_0 p_{\parallel}} \right)^{C_2} \quad (8.56)$$

where from observations in the magnetosphere [Anderson+ 1996; Gary+ 1995] we can take the average values $C_1 = 0.3$ and $C_2 = 0.5$.

The effect of these instabilities is represented by the collision term of the parallel pressure on the RHS of the fourth equation in Equation 8.51:

$$\frac{\delta p_{\parallel}}{\delta t} = \frac{\bar{p}_{\parallel} - p_{\parallel}}{\tau} \quad (8.57)$$

where \bar{p}_{\parallel} is the marginally stable parallel pressure obtained from Equation 8.54, Equation 8.55, and Equation 8.56. Equation 8.57 is called the pressure relaxation term due to the instabilities, with the relaxation time τ related to their growth rates.

These equations provide a self-consistent description of phenomena that satisfy all the various assumptions we have made, namely:

1. The plasma is charge-neutral since characteristic lengths are much longer than a Debye length;
2. The characteristic velocity of the phenomenon under consideration is slow compared to the speed of light;
3. The pressure and density gradients are parallel, so there is no electrothermal EMF;
4. The time scale is long compared to both the electron and ion cyclotron periods.

Even though these assumptions are self-consistent, they may not accurately portray a real plasma and so MHD models, while intuitively appealing, must be used with caution.

Finally, it is worth mentioning that MHD plasmas can be categorized yet another way, namely according to the relative importance of the magnetic force $\mathbf{J} \times \mathbf{B}$ compared to the hydrodynamic force $\nabla \cdot \tilde{\mathbf{P}}$. If the magnetic force is negligible compared to the hydrodynamic force, then there is not much point in using MHD because in this case the system of equations is simply classical hydrodynamics. Thus, the only non-trivial MHD situations are where the magnetic and hydrodynamic forces are of comparable importance or where the magnetic force is much more important than the hydrodynamic force. Using Equation 8.48, the nominal ratio of the hydrodynamic force to the magnetic force is defined as

$$\beta = \frac{p}{B^2/2\mu_0} \sim \frac{\nabla \cdot \tilde{\mathbf{P}}}{\mathbf{J} \times \mathbf{B}} \sim \frac{p/L}{B^2/2\mu_0 L}$$

The characteristic gradient scale length L is assumed to be comparable for both types of forces and so cancels out in the comparison. Low- β plasmas are those where $B^2/2\mu_0$ is much

larger than p so the hydrodynamic force is negligible compared to the magnetic force, whereas $\beta = \mathcal{O}(1)$ plasmas are those where the magnetic and hydrodynamic forces are comparable. The different β regime has important implications on the validity of many approximations seen in later chapters.

8.6.3 Hall MHD

As an extension to the ideal/resistive MHD model, Hall MHD decouples the electron and ion motions by retrieving the Hall term in the generalized Ohm's law

$$\mathbf{E} = -\mathbf{u} \times \mathbf{B} + \eta \mathbf{J} + \frac{1}{en} \mathbf{J} \times \mathbf{B} - \frac{\nabla p_e}{ne} \quad (8.58)$$

where the first term on the right-hand-side is the convective term, the second is the resistive term, the third is the Hall term, and the fourth is the electron pressure gradient term. The electron pressure scalar is simplified from the electron pressure tensor, which is obtained from an independent electron pressure equation.

The Hall MHD equations are

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= -\nabla \cdot (\rho \mathbf{u}) \\ \frac{\partial(\rho \mathbf{u})}{\partial t} &= -\nabla \cdot \left(\rho \mathbf{u} \mathbf{u} + (p + p_e) \bar{\bar{I}} + \frac{B^2}{2\mu_0} \bar{\bar{I}} - \frac{\mathbf{B} \mathbf{B}}{\mu_0} \right) \\ \frac{\partial e}{\partial t} &= -\nabla \cdot \left[(\epsilon + p) \mathbf{u} + (\epsilon_e + p_e) \mathbf{u}_e + \mathbf{u}_e \cdot \left(\frac{\mathbf{B}^2}{\mu_0} \bar{\bar{I}} - \frac{\mathbf{B} \mathbf{B}}{\mu_0} \right) - \mathbf{B} \times \eta \mathbf{J} \right] \quad (8.59) \\ \frac{\partial \mathbf{B}}{\partial t} &= -\nabla \times \left[\mathbf{u}_e \times \mathbf{B} + \eta \mathbf{J} - \frac{\nabla p_e}{ne} \right] \\ \frac{\partial p_e}{\partial t} + \nabla \cdot (p_e \mathbf{u}_e) &= -(\gamma - 1)p_e \nabla \cdot \mathbf{u}_e \end{aligned}$$

where $\bar{\bar{I}}$ is the identity matrix, ρ is the mass density, \mathbf{u} is the plasma bulk velocity, \mathbf{B} is the magnetic field, p_e is the electron pressure, p is the ion thermal pressure, $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ is the current density, $\epsilon = \rho \mathbf{u}^2/2 + p/(\gamma - 1)$ is the hydrodynamic energy density, and $\epsilon_e = \rho \mathbf{u}_e^2/2 + p_e/(\gamma - 1)$ is the electron hydrodynamic energy density. The Hall velocity is defined as

$$\mathbf{v}_H = -\frac{\mathbf{J}}{ne} \quad (8.60)$$

and the electron bulk velocity is given by

$$\mathbf{u}_e = \mathbf{u} + \mathbf{v}_H \quad (8.61)$$

The total energy density is

$$e = \epsilon + \epsilon_e + \frac{\mathbf{B}^2}{2\mu_0} = \frac{1}{2}\rho \mathbf{u}^2 + \frac{1}{\gamma - 1}(p + p_e) + \frac{\mathbf{B}^2}{2\mu_0}$$

where γ is the adiabatic index. Note that in Hall MHD only $(\rho, \mathbf{u}, \mathbf{B}, p, p_e)$ are unknowns; all others are derived quantities.

To get back to the ideal MHD equations, we simply ignore the Hall velocity Equation 8.60 and replace \mathbf{u}_e in the energy conservation and induction equation with \mathbf{u} . *Note that this does not mean the current density in ideal MHD is zero: it is just a term that we neglect when deriving the electric field in the Poynting flux as well as the induction equation.*

One important thing to recognize is that while ideal MHD has no intrinsic scales, Hall MHD introduces a new characteristic length scale into the system – the ion skin depth d_i . In ideal MHD, when the plasma beta is low, the speed of the shear Alfvén mode and the fast magnetosonic mode become approximately equal. This is known as *degeneracy*, which means it is difficult to distinguish between the two modes, as they essentially travel at the same speed. In Hall MHD, when the wavelength of the waves becomes comparable to d_i , the distinction between the shear Alfvén and the fast magnetosonic modes becomes clear:

1. Fast magnetosonic waves (whistler waves) are affected by electron dynamics and propagate at higher frequencies within the Hall-MHD framework.
2. Shear Alfvén waves (kinetic Alfvén waves, or Alfvén-ion-cyclotron waves) are predominantly driven by ion dynamics and tend to propagate at lower frequencies.

8.7 Fluid Drifts Perpendicular to \mathbf{B}

Since a fluid element is composed of many individual particles, one would expect the fluid to have drifts perpendicular to \mathbf{B} if the individual guiding centers have such drifts. However, since the ∇p term appears only in the fluid equations, there is a drift associated with it which the fluid elements have but the particles do not have. For each species, we have an equation of motion

$$mn \left[\underbrace{\frac{\partial \mathbf{u}}{\partial t}}_1 + \underbrace{(\mathbf{u} \cdot \nabla) \mathbf{u}}_2 \right] = qn(\mathbf{E} + \underbrace{\mathbf{u} \times \mathbf{B}}_3) - \nabla p \quad (8.62)$$

Consider the ratio of term 1 to term 3:

$$\frac{\text{term 1}}{\text{term 3}} \approx \left| \frac{mni\omega v_{\perp}}{qnv_{\perp}B} \right| \approx \frac{\omega}{\omega_c}$$

Here we have taken $\partial/\partial t = i\omega$ and are concerned only with \mathbf{v}_{\perp} . For drifts slow compared with the time scale associated with ω_c , we may neglect term 1. We shall also neglect the $(\mathbf{u} \cdot \nabla) \mathbf{u}$ term and show a *posteriori* that this is all right. Let \mathbf{E} and \mathbf{B} be uniform, but let n and p have a gradient. This is the usual situation in a magnetically confined plasma column (Fig. 3.4 ADD FIGURE!). Taking the cross product of Equation 8.62 with \mathbf{B} , we have (neglecting the left-hand side)

$$\begin{aligned}
0 &= qn[\mathbf{E} \times \mathbf{B} + (\mathbf{u}_\perp \times \mathbf{B}) \times \mathbf{B}] - \nabla p \times \mathbf{B} \\
&= qn[\mathbf{E} \times \mathbf{B} + \mathbf{B}(\mathbf{u}_\perp \cdot \mathbf{B}) - \mathbf{u}_\perp \mathbf{B}^2] - \nabla p \times \mathbf{B} \\
&= qn[\mathbf{E} \times \mathbf{B} - \mathbf{u}_\perp \mathbf{B}^2] - \nabla p \times \mathbf{B}
\end{aligned}$$

Therefore,

$$\mathbf{u}_\perp = \frac{\mathbf{E} \times \mathbf{B}}{B^2} - \frac{\nabla p \times \mathbf{B}}{qnB^2} \equiv \mathbf{u}_E + \mathbf{u}_D$$

where

$$\begin{aligned}
\mathbf{u}_E &\equiv \frac{\mathbf{E} \times \mathbf{B}}{B^2} \quad \mathbf{E} \times \mathbf{B} \text{ drift} \\
\mathbf{u}_D &\equiv -\frac{\nabla p \times \mathbf{B}}{qnB^2} \quad \text{Diamagnetic drift}
\end{aligned} \tag{8.63}$$

The drift \mathbf{u}_E is the same as for guiding centers, but there is now a new drift \mathbf{u}_D , called the diamagnetic drift. Since \mathbf{u}_D is perpendicular to the direction of the gradient, our neglect of $(\mathbf{u} \cdot \nabla)\mathbf{u}$ is justified if $\mathbf{E} = 0$. If $\mathbf{E} = -\nabla\phi \neq 0$ (i.e. a potential field), $(\mathbf{u} \cdot \nabla)\mathbf{u}$ is still zero if $\nabla\phi$ and ∇p are in the same direction; otherwise, there could be a more complicated solution involving $(\mathbf{u} \cdot \nabla)\mathbf{u}$.

With the help of Equation 8.12, we can write the diamagnetic drift as

$$\mathbf{u}_D = \pm \frac{\gamma k_B T}{eB} \frac{\hat{z} \times \nabla n}{n} \tag{8.64}$$

In particular, for an isothermal plasma in the geometry of Fig.3.4 (ADD IT!), in which $\nabla n = \partial n / \partial r \hat{r} = n' \hat{r}$ ($n' < 0$), we have the following formulas familiar to experimentalists who have worked with Q-machines

$$\begin{aligned}
\mathbf{u}_{Di} &= \frac{k_B T_i}{eB} \frac{n'}{n} \hat{\theta} \\
\mathbf{u}_{De} &= -\frac{k_B T_e}{eB} \frac{n'}{n} \hat{\theta}
\end{aligned}$$

The physical reason for this drift can be seen from Figure 8.2. Here we have drawn the orbits of ions gyrating in a magnetic field. There is a density gradient toward the left, as indicated by the density of orbits. Through any fixed volume element there are more ions moving downward than upward, since the downward-moving ions come from a region of higher density. There is, therefore, a fluid drift perpendicular to ∇n and \mathbf{B} , *even though the guiding centers are stationary*. The diamagnetic drift reverses sign with q because the direction of gyration reverses. The magnitude of \mathbf{u}_D does not depend on mass because the $m^{-1/2}$ dependence of the velocity is cancelled by the $m^{1/2}$ dependence of the Larmor radius—less of the density gradient is sampled during a gyration if the mass is small.

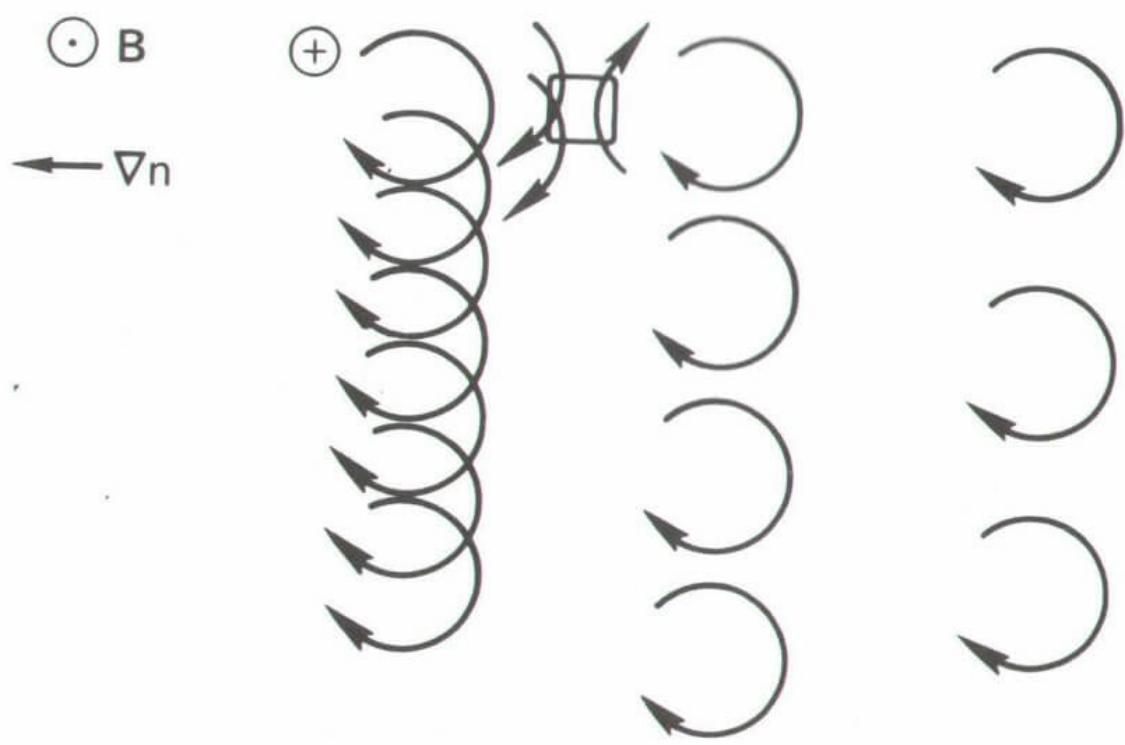


Figure 8.2: Origin of the diamagnetic drift.

Since ions and electrons drift in opposite directions, there is a diamagnetic current. For $\gamma = Z = 1$, this is given by

$$\mathbf{J}_D = ne(\mathbf{u}_{Di} + \mathbf{u}_{De}) = \frac{\mathbf{B}}{B^2} \times \nabla(p_i + p_e) = (k_B T_i + k_B T_e) \frac{\mathbf{B} \times \nabla n}{B^2} \quad (8.65)$$

In the particle picture, one would not expect to measure a current if the guiding centers do not drift. In the fluid picture, the current \mathbf{J}_D flows wherever there is a pressure gradient. These two viewpoints can be reconciled if one considers that all experiments must be carried out in a finite-sized plasma. Suppose the plasma were in a rigid box (Fig. 3.6 ADD IT!). If one were to calculate the current from the single-particle picture, one would have to take into account the particles at the edges which have cycloidal paths. Since there are more particles on the left than on the right, there is a net current downward, in agreement with the fluid picture.

The reader may not be satisfied with this explanation because it was necessary to specify reflecting walls. If the walls were absorbing or if they were removed, one would find that electric fields would develop because more of one species—the one with larger Larmor radius—would be collected than the other. Then the guiding centers would drift, and the simplicity of the model would be lost. Alternatively, one could imagine trying to measure the diamagnetic current with a current probe (Fig. 3.7 ADD IT!). This is just a transformer with a core of magnetic material. The primary winding is the plasma current threading the core, and the secondary is a multturn winding all around the core. Let the whole thing be infinitesimally thin, so it does not intercept any particles. It is clear from Fig. 3.7 that a net upward current would be measured, there being higher density on the left than on the right, so that the diamagnetic current is a real current. From this example, one can see that it can be quite tricky to work with the single-particle picture. The fluid theory usually gives the right results when applied straightforwardly, even though it contains “fictitious” drifts like the diamagnetic drift.

What about the grad-B and curvature drifts which appeared in the single-particle picture? The curvature drift also exists in the fluid picture, since the centrifugal force is felt by all the particles in a fluid element as they move around a bend in the magnetic field. A term $\bar{F}_{cf} = nmv_{\parallel}^2/R_c = nk_B T_{\parallel}/R_c$ has to be added to the right-hand side of the fluid equation of motion. This is equivalent to a gravitational force mng , with $g = k_B T_{\parallel}/mR_c$, and leads to a drift $\mathbf{u}_g = (m/q)(\mathbf{g} \times \mathbf{B})/B^2$, as in the particle picture (Equation 7.5).

The grad-B drift, however, does not exist for fluids. It can be shown on thermodynamic grounds that a magnetic field does not affect a Maxwellian distribution. This is because the Lorentz force is perpendicular to \mathbf{v} and cannot change the energy of any particle. The most probable distribution $f(\mathbf{v})$ in the absence of \mathbf{B} is also the most probable distribution in the presence of \mathbf{B} . If $f(\mathbf{v})$ remains Maxwellian in a nonuniform \mathbf{B} field, and there is no density gradient, then the net momentum carried into any fixed fluid element is zero. There is no fluid drift even though the individual guiding centers have drifts; the particle drift in any fixed fluid element cancel out. To see this pictorially, consider the orbits of two particles moving through a fluid element in a nonuniform \mathbf{B} field (Figure 8.3). Since there is no \mathbf{E} field, the Larmor radius

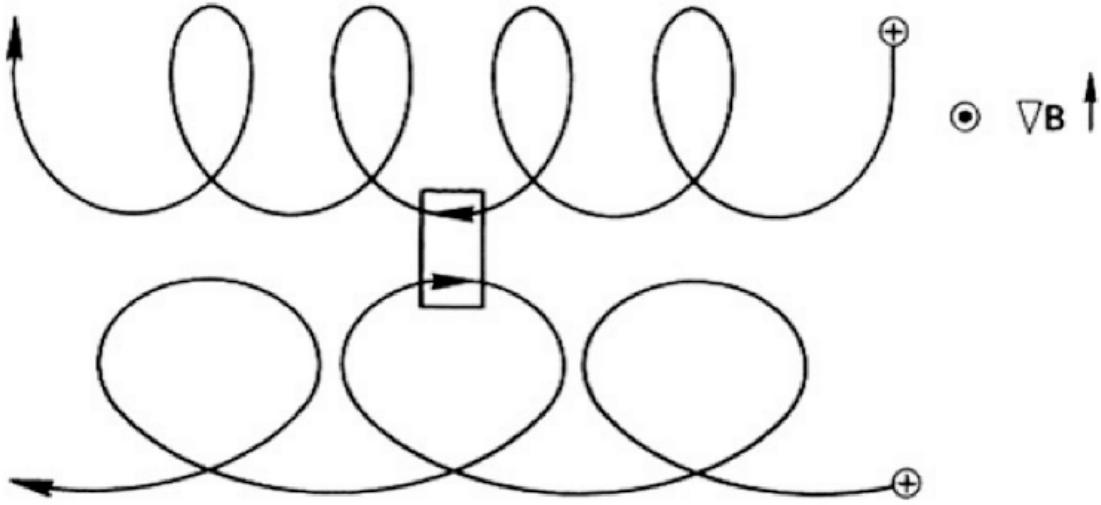


Figure 8.3: In a nonuniform \mathbf{B} field the guiding centers drift but the fluid elements do not.

changes only because of the gradient in \mathbf{B} ; there is no acceleration, and the particle energy remains constant during the motion. If the two particles have the same energy, they will have the same velocity and Larmor radius while inside the fluid element. There is thus a perfect cancellation between particle pairs when their velocities are added to give the fluid velocity.

When there is a nonuniform \mathbf{E} field, it is not easy to reconcile the fluid and particle pictures. Then the finite-Larmor-radius effect of Section 7.3 causes both a guiding center drift and a fluid drift, but these are not the same; in fact, they have opposite signs! The particle drift was calculated in Chapter 7, and the fluid drift can be calculated from the off-diagonal elements of \mathbf{P} . It is extremely difficult to explain how the finite-Larmor-radius effects differ. A simple picture like Fig. 3.6 will not work because one has to take into account subtle points like the following: in the absence of a density gradient, the density of guiding centers is not the same as the density of particles! (???)

(I need to think carefully about these pictures.)

8.8 Fluid Drifts Parallel to \mathbf{B}

The z component of the fluid equation of motions is

$$mn \left[\frac{\partial u_z}{\partial t} + (\mathbf{u} \cdot \nabla) u_z \right] = qnE_z - \frac{\partial \phi}{\partial z} \quad (8.66)$$

The convective term can often be neglected because it is much smaller than the $\partial u_z / \partial t$ term. We shall avoid complicated arguments here and simply consider cases in which u_z is spatially

uniform. Using Equation 8.66, we have

$$\frac{\partial u_z}{\partial t} = \frac{q}{m} E_z - \frac{\gamma k_B T}{mn} \frac{\partial n}{\partial z} \quad (8.67)$$

This shows that the fluid is accelerated along \mathbf{B} under the combined electrostatic and pressure gradient forces. A particularly important result is obtained by applying Equation 8.67 to massless electrons. Taking the limit $m \rightarrow 0$ and specifying $q = -e$ and $\mathbf{E} = -\nabla\phi$, we have

$$qE_z = e \frac{\partial \phi}{\partial z} = \frac{\gamma k_B T_e}{n} \frac{\partial n}{\partial z}$$

Electrons are so mobile that their heat conductivity is almost infinite. We may then assume isothermal electrons and take $\gamma = 1$. Integrating, we have

$$e\phi = k_B T_e \ln n + C$$

or

$$n = n_0 \exp(e\phi/k_B T_e) \quad (8.68)$$

This is just the *Boltzmann relation* for electrons.

What this means physically is that electrons, being light, are very mobile and would be accelerated to high energies very quickly if there were a net force on them. Since electrons cannot leave a region en masse without leaving behind a large ion charge, the electrostatic and pressure gradient forces on the electrons must be closely in balance. This condition leads to the Boltzmann relation. Note that Equation 8.68 applies to each magnetic field lines separately. Different field lines may be charged to different potentials arbitrarily unless a mechanism is provided for the electrons to move across \mathbf{B} . The conductors on which field lines terminate can provide such a mechanism, and the experimentalist has to take these end effects into account carefully.

Figure 8.4 shows graphically what occurs when there is a local density clump in the plasma. Let the density gradient be toward the center of the diagram, and suppose $k_B T$ is constant. There is then a pressure gradient toward the center. Since the plasma is quasineutral, the gradient exists for both the electron and ion fluids. Consider the pressure gradient force \mathbf{F}_p on the electron fluid. It drives the mobile electrons away from the center, leaving the ions behind. The resulting positive charge generates a field \mathbf{E} whose force \mathbf{F}_E on the electrons opposes \mathbf{F}_p . Only when $\mathbf{F}_E = -\mathbf{F}_p$ is a steady state achieved. If \mathbf{B} is a constant, \mathbf{E} is an electrostatic field $\mathbf{E} = -\nabla\phi$, and ϕ must be large at the center, where n is large. This is just what Equation 8.68 tells us. The deviation from strict neutrality adjusts itself so that there is just enough charge to set up the \mathbf{E} field required to balance the forces on the electrons.

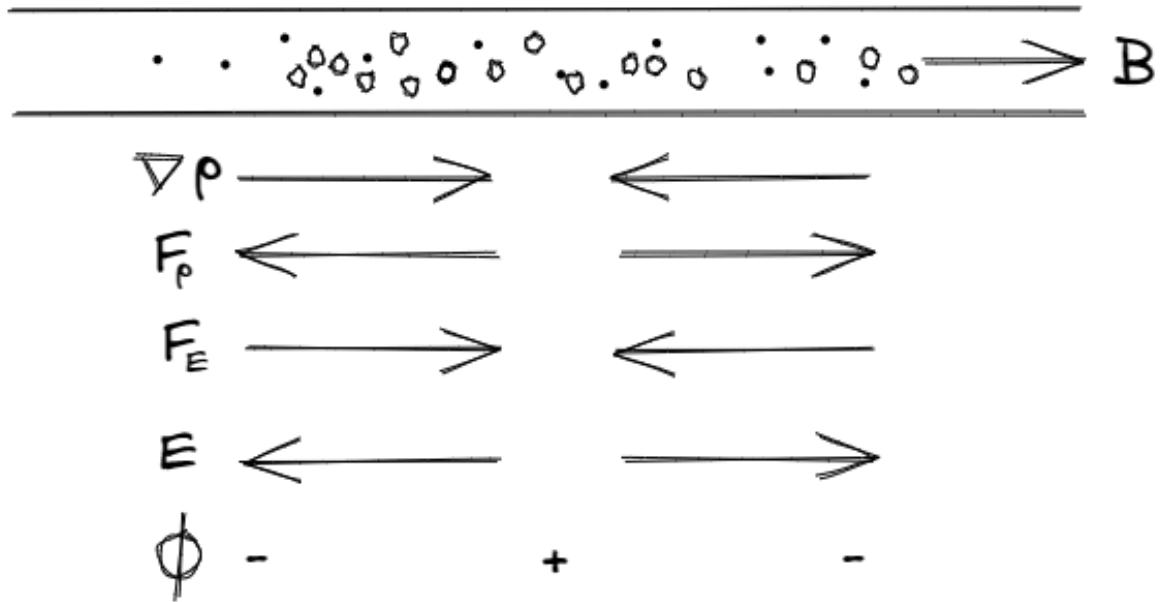


Figure 8.4: Physical reason for the Boltzmann relation between density and potential.

8.9 The Plasma Approximation

The previous example reveals an important characteristic of plasmas that has wide application. We are used to solving for \mathbf{E} from Poisson's equation when we are given the charge density σ . In a plasma, the opposite procedure is generally used. \mathbf{E} is found from the equations of motion, and Poisson's equation is used only to find σ . The reason is that a plasma has an overriding tendency to remain neutral. If the ions move, the electrons will follow. \mathbf{E} must adjust itself so that the orbits of the electrons and ions preserve neutrality. The charge density is of secondary importance; it will adjust itself so that Poisson's equation is satisfied. This is true, of course, only for low-frequency motions in which the electron inertia is not a factor.

In a plasma, it is usually possible to assume $n_i = n_e$ and $\nabla \cdot \mathbf{E} \neq 0$ at the same time. We shall call this the *plasma approximation*. It is a fundamental trait of plasmas, one which is difficult for the novice to understand. *Do not use Poisson's equation to obtain \mathbf{E} unless it is unavoidable!* In the set of fluid equations Equation 8.13, we may now eliminate Poisson's equation and also eliminate one of the unknowns by setting $n_i = n_e = n$.

The *plasma approximation* is almost the same as the condition of quasineutrality discussed earlier but has a more exact meaning. Whereas quasineutrality refers to a general tendency for a plasma to be neutral in its state of rest, the plasma approximation is a mathematical shortcut that one can use even for wave motions. As long as these motions are slow enough that both ions and electrons have time to move, it is a good approximation to replace Poisson's equation by the equation $n_i = n_e$. Of course, if only one species can move and the other cannot follow,

such as in high-frequency electron waves, then the plasma approximation is not valid, and \mathbf{E} must be found from Maxwell's Equation 8.1 rather than from the ion and electron equations of motion. Ampère's law with the displacement current retained gives the time evolution of \mathbf{E} :

$$\frac{d\mathbf{E}}{dt} = \frac{1}{\epsilon_0} \left(\frac{\nabla \times \mathbf{B}}{\mu_0} - \mathbf{J} \right)$$

We shall return to the question of the validity of the plasma approximation when we come to the theory of ion waves. At that time, it will become clear why we had to use Poisson's equation in the derivation of Debye shielding. For electron waves, we can also show that with a simple "modified" electron test particle model that updates the electric field while keeping the magnetic field constant, we can retain the plasma oscillation and electron Bernstein wave.

9 Diffusion and Resistivity

9.1 Diffusion and Mobility in Weakly Ionized Gases

The infinite, homogeneous plasmas assumed in the previous chapter for the equilibrium conditions are, of course, highly idealized. Any realistic plasma will have a density gradient, and the plasma will tend to diffuse toward regions of low density. For example, the central problem in controlled thermonuclear reactions is to impede the rate of diffusion by using a magnetic field. Before tackling the magnetic field problem, however, we shall consider the case of diffusion in the absence of magnetic fields. A further simplification results if we assume that the plasma is weakly ionized, so that charge particles collide primarily with neutral atoms rather than with one another. The case of a fully ionized plasma is deferred to a later section, since it results in a nonlinear equation for which there are few simple illustrative solutions. In any case, partially ionized gases are not rare: High-pressure arcs and ionospheric plasmas fall into this category, and most of the early work on gas discharges involved fractional ionizations between 10^3 and 10^6 , when collisions with neutral atoms are dominant.

The picture, then, is of a nonuniform distribution of ions and electrons in a dense background of neutrals (Fig. 5.1 ADD IT!). As the plasma spreads out as a result of pressure-gradient and electric field forces, the individual particles undergo a random walk, colliding frequently with the neutral atoms. We begin with a brief review of definitions from atomic theory.

9.2 Collision Parameters

When an electron, say, collides with a neutral atom, it may lose any fraction of its initial momentum, depending on the angle at which it rebounds. In a head-on collision with a heavy atom, the electron can lose twice its initial momentum, since its velocity reverses sign after the collision. The probability of momentum loss can be expressed in terms of the equivalent cross section σ that the atoms would have if they were perfect absorbers of momentum.

In Fig. 5.2, electrons are incident upon a slab of area A and thickness dx containing n_n neutral atoms per m^3 . The atoms are imagined to be opaque spheres of cross-sectional area σ ; that is, every time an electron comes within the area blocked by the atom, the electron loses all of its momentum. The number of atoms in the slab is $n_n Adx$. The fraction of the slab blocked by atoms is $n_n A\sigma dx/A = n_n \sigma dx$. If a flux Γ of electrons is incident on the slab, the flux emerging

on the other side is $\Gamma' = \Gamma(1 - n_n \sigma dx)$. Thus the change of Γ with distance is $d\Gamma/dx = -n_n \sigma \Gamma$, or

$$\Gamma = \Gamma_0 e^{-n_n \sigma x} \equiv \Gamma_0 e^{-x/\lambda_m}$$

In a distance λ_m , the flux would be decreased to $1/e$ of its initial value. The quantity λ_m is the *mean free path* for collisions:

$$\lambda_m = 1/n_n \sigma \quad (9.1)$$

After traveling a distance λ_m , a particle will have had a good probability of making a collision. The mean time between collisions, for particles of velocity v , is given by

$$\tau = \lambda_m/v$$

and the mean frequency of collisions is

$$\tau^{-1} = v/\lambda_m = n_n \sigma v \quad (9.2)$$

If we now average over particles of all velocities v in a Maxwellian distribution, we have what is generally called the collision frequency ν :

$$\nu = n_n \bar{\sigma v} \quad (9.3)$$

9.3 Diffusion Parameters

The fluid equation of motion including collisions is, for any species s ,

$$mn \frac{d\mathbf{v}}{dt} = mn \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = q_s n \mathbf{E} - \nabla p - mn\nu \mathbf{v}$$

The averaging process used to compute ν is such as to make the last equation correct; we need not be concerned with the details of this computation. The quantity ν must, however, be assumed to be a constant in order for the equation to be useful. We shall consider a steady state in which $\partial \mathbf{v} / \partial t = 0$. If \mathbf{v} is sufficiently small (or ν sufficiently large), a fluid element will not move into regions of different \mathbf{E} and ∇p in a collision time, and the convective derivative $d\mathbf{v}/dt$ will also vanish. Setting the left-hand side to zero, we have, for an isothermal plasma,

$$\mathbf{v} = \frac{1}{mn\nu} (q_s n \mathbf{E} - k_B T \nabla n) = \frac{q_s}{m\nu} \mathbf{E} - \frac{k_B T \nabla n}{mv} \frac{n}{n}$$

The coefficients above are called the *mobility* and the *diffusion coefficient*:

$$\mu \equiv |q|/m\nu \quad \text{Mobility} \quad (9.4)$$

$$D \equiv k_B T / m \nu \quad \text{Diffusion coefficient} \quad (9.5)$$

These will be different for each species. Note that D is measured in m^2/s . The transport coefficients μ and D are connected by the *Einstein relation*:

$$\mu = |q| D / k_B T \quad (9.6)$$

With the help of these definitions, the flux Γ_s of the s th species can be written

$$\Gamma_s = n \mathbf{v}_s = \pm \mu_s n \mathbf{E} - D_s \nabla n \quad (9.7)$$

Fick's law of diffusion is a special case of this, occurring when either $\mathbf{E} = 0$ or the particles are uncharged, so that $\mu = 0$:

$$\Gamma = -D \nabla n \quad \text{Fick's law}$$

This equation merely expresses the fact that diffusion is a random-walk process, in which a net flux from dense regions to less dense regions occurs simply because more particles start in the dense region. This flux is obviously proportional to the gradient of the density. In plasmas, Fick's law is not necessarily obeyed. Because of the possibility of organized motions (plasma waves), a plasma may spread out in a manner which is not truly random.

9.4 Decay of a Plasma by Diffusion

9.4.1 Ambipolar Diffusion

Ambipolar diffusion is related to the concept of quasineutrality. In most plasmas, the forces acting on the ions are different from those acting on the electrons, so naively one would expect one species to be transported faster than the other, whether by diffusion or convection or some other process. If such differential transport has a divergence, then it results in a change of the charge density. The latter will create an electric field that can alter the transport of one or both species in such a way that they become equal.

The simplest example is a plasma localized in an unmagnetized vacuum (e.g. inertial confinement fusion). Both electrons and ions will stream outward with their respective thermal velocity. If the ions are relatively cold, their thermal velocity will be small. The thermal velocity of the electrons will be fast due to their high temperature and low mass: $v_{th,e} = \sqrt{k_B T_e / m_e}$. As the electrons leave the initial volume, they will leave behind a positive charge density of ions, resulting in an outwardly directed electric field. This field will act on the electrons to slow them down and on the ions to speed them up. The net result is that both ions and electrons stream outward at the speed of sound, $c_s \approx \sqrt{k_B T_e / m_i}$, which is much smaller than the electron thermal velocity, but usually much larger than the ion thermal velocity.

We now consider how a plasma created in a container decays by diffusion to the walls. Once ions and electrons reach the wall, they recombine there. The density near the wall, therefore, is essentially zero. The fluid equations of motion and continuity govern the plasma behavior; but if the decay is slow, we need only keep the time derivative in the continuity equation. The time derivative in the equation of motion, will be negligible if the collision frequency ν is large. We thus have

$$\frac{\partial n}{\partial t} + \nabla \cdot \Gamma_s = 0 \quad (9.8)$$

with Γ_s given by Equation 9.7. It is clear that if Γ_i and Γ_e were not equal, a serious charge imbalance would soon arise. If the plasma is much larger than a Debye length, it must be quasineutral; and one would expect that the rates of diffusion of ions and electrons would somehow adjust themselves so that the two species leave at the same rate. How this happens is easy to see. The electrons, being lighter, have higher thermal velocities and tend to leave the plasma first. A positive charge is left behind, and an electric field is set up of such a polarity as to retard the loss of electrons and accelerate the loss of ions. The required \mathbf{E} field is found by setting $\Gamma_i = \Gamma_e = \Gamma$. From @Equation 9.7, we can write

$$\Gamma = \mu_i n \mathbf{E} - D_i \nabla n = -\mu_e n \mathbf{E} - D_e \nabla n$$

$$\mathbf{E} = \frac{D_i - D_e}{\mu_i + \mu_e} \frac{\nabla n}{n}$$

The common flux Γ is then given by

$$\begin{aligned} \Gamma &= \mu_i \frac{D_i - D_e}{\mu_i + \mu_e} \nabla n - D_i \nabla n \\ &= \frac{\mu_i D_i - \mu_i D_e - \mu_i D_i - \mu_e D_i}{\mu_i + \mu_e} \nabla n \\ &= -\frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e} \nabla n \end{aligned}$$

This is Fick's law with a new diffusion coefficient

$$D_a \equiv \frac{\mu_i D_e + \mu_e D_i}{\mu_i + \mu_e} \quad (9.9)$$

called the *ambipolar diffusion coefficient*. If this is constant, the continuity equation above becomes simply

$$\frac{\partial n}{\partial t} = D_a \nabla^2 n \quad (9.10)$$

The magnitude of Da can be estimated if we take $\mu_e \gg \mu_i$. That this is true can be seen from Equation 9.4. Since ν is proportional to the thermal velocity, which is proportional to $m^{-1/2}$, μ is proportional to $m^{-1/2}$. Equation 9.4 and Equation 9.9 then give

$$D_a \approx D_i + \frac{\mu_i}{\mu_e} D_e = D_i + \frac{T_e}{T_i} D_i$$

For $T_e = T_i$, we have

$$D_a \approx 2D_i$$

The effect of the ambipolar electric field is to enhance the diffusion of ions by a factor of two, and the diffusion rate of the two species together is primarily controlled by the slower species.

9.4.2 Diffusion in a Slab

The diffusion Equation 9.10 can easily be solved by the method of separation of variables. We let

$$n(\mathbf{r}, t) = T(t)S(\mathbf{r})$$

whereupon Equation 9.10, with the subscript on D_a understood, becomes

$$\begin{aligned} S \frac{dT}{dt} &= DT \nabla^2 S \\ \frac{1}{T} \frac{dT}{dt} &= \frac{D}{S} \nabla^2 S \end{aligned}$$

Since the left side is a function of time alone and the right side a function of space alone, they must both be equal to the same constant, which we shall call $-1/\tau$. The function T then obeys the equation

$$\frac{dT}{dt} = -\frac{T}{\tau} \quad (9.11)$$

with the solution

$$T = T_0 e^{-t/\tau}$$

The spatial part S obeys the equation

$$\nabla^2 S = -\frac{1}{D\tau} S \quad (9.12)$$

In slab geometry, this becomes

$$\frac{d^2 S}{dx^2} = -\frac{1}{D\tau} S \quad (9.13)$$

with the solution

$$S = A \cos \frac{x}{(D\tau)^{1/2}} + B \sin \frac{x}{(D\tau)^{1/2}} \quad (9.14)$$

We would expect the density to be nearly zero at the walls (Fig. 5.3 ADD IT!) and to have one or more peaks in between. The simplest solution is that with a single maximum. By symmetry,

we can reject the odd (sine) term in Equation 9.14. The boundary conditions $S = 0$ at $x = \pm L$ then requires

$$\frac{L}{(D\tau)^{1/2}} = \frac{\pi}{2}$$

or

$$\tau = \left(\frac{2L}{\pi}\right)^2 \frac{1}{D}$$

Combining all the equations above, we have

$$n = n_0 e^{-t/\tau} \cos \frac{\pi x}{2L}$$

This is called the *lowest diffusion mode*. The density distribution is a cosine, and the peak density decays exponentially with time. The time constant τ increases with L and varies inversely with D , as one would expect.

There are, of course, higher diffusion modes with more than one peak. Suppose the initial density distribution is as shown by the top curve in Fig. 5.4 ADD IT. Such an arbitrary distribution can be expanded in a Fourier series:

$$n = n_0 \left(\sum_l a_l \cos \frac{(l + \frac{1}{2})\pi x}{L} \sum_m b_m \sin \frac{m\pi x}{L} \right)$$

We have chosen the indices so that the boundary condition at $x = \pm L$ is automatically satisfied. To treat the time dependence, we can try a solution of the form

$$n = n_0 \left(\sum_l a_l e^{-t/\tau_l} \cos \frac{(l + \frac{1}{2})\pi x}{L} \sum_m b_m e^{-t/\tau_m} \sin \frac{m\pi x}{L} \right)$$

Substituting this into the diffusion Equation 9.10, we see that each cosine term yields a relation of the form

$$-\frac{1}{\tau_l} = -D \left[\left(l + \frac{\pi}{2} \right) \frac{\pi}{L} \right]^2$$

and similarly for the sine terms. Thus the decay time constant for the l th mode is given by

$$\tau_l = \left[\frac{L}{(l + \frac{1}{2})\pi} \right]^2 \frac{1}{D}$$

The fine-grained structure of the density distribution, corresponding to large l numbers, decays faster, with a smaller time constant τ_l . The plasma decay will proceed as indicated in Fig. 5.4 ADD IT. First, the fine structure will be washed out by diffusion. Then the lowest diffusion mode, the simple cosine distribution of Fig. 5.3 ADD IT, will be reached. Finally, the peak density continues to decay while the plasma density profile retains the same shape.

I WANT TO DO A SIMULATION ON THIS! TRY METHODOFLINES.JL?

9.4.3 Diffusion in a Cylinder

The spatial part of the diffusion equation, eq-diffusion_spatial, reads, in cylindrical geometry,

$$\frac{d^2S}{dr^2} + \frac{1}{r} \frac{dS}{dr} + \frac{1}{D\tau} S = 0 \quad (9.15)$$

This differs from Equation 9.13 by the addition of the middle term, which merely accounts for the change in coordinates. The need for the extra term is illustrated simply in Fig. 5.5 ADD IT. If a slice of plasma in (a) is moved toward larger x without being allowed to expand, the density would remain constant. On the other hand, if a shell of plasma in (b) is moved toward larger r with the shell thickness kept constant, the density would necessarily decrease as $1/r$. Consequently, one would expect the solution to Equation 9.15 to be like a damped cosine (Fig. 5.6 ADD IT). This function is called a *Bessel function of order zero*, and eq-diffusion_spatial_cylinder is called Bessel's equation (of order zero). Instead of the symbol \cos , it is given the symbol J_0 . The function $J_0(r/[D\tau]^{1/2})$ is a solution to eq-diffusion_spatial_cylinder, just as $\cos[x/(d\tau)^{1/2}]$ is a solution to eq-diffusion_spatial. Both $\cos kx$ and $J_0(kr)$ are expressible in terms of infinite series.

FIGURE: Motion of a plasma slab in rectilinear and cylindrical geometry, illustrating the difference between a cosine and a Bessel function.

To satisfy the boundary condition $n = 0$ at $r = a$, we must set $a/(D\tau)^{1/2}$ equal to the first zero of J_0 ; namely, 2.4. This yields the decay time constant τ . The plasma again decays exponentially, since the temporal part of the diffusion equation, Equation 9.11, is unchanged. We have described the lowest diffusion mode in a cylinder. Higher diffusion modes, with more than one maximum in the cylinder, will be given in terms of Bessel functions of higher order, in direct analogy to the case of slab geometry.

9.5 Steady State Solutions

In many experiments, a plasma is maintained in a steady state by continuous ionization or injection of plasma to offset the losses. To calculate the density profile in this case, we must add a source term to the equation of continuity:

$$\frac{\partial n}{\partial t} - D\nabla^2 n = Q(\mathbf{r})$$

The sign is chosen so that when Q is positive, it represents a source and contributes to positive $\partial n / \partial t$. In steady state, we set $\partial n / \partial t = 0$ and are left with a Poisson-type equation for $n(\mathbf{r})$.

9.5.1 Constant Ionization Function

In many weakly ionized gases, ionization is produced by energetic electrons in the tail of the Maxwellian distribution. In this case, the source term Q is proportional to the electron density n . Setting $Q = Zn$, where Z is the “ionization function”, we have

$$\nabla^2 n = -(Z/D)n$$

This is the same equation as that for S, Equation 9.12. Consequently, the density profile is a cosine or Bessel function, as in the case of a decaying plasma, only in this case the density remains constant. The plasma is maintained against diffusion losses by whatever heat source keeps the electron temperature at its constant value and by a small influx of neutral atoms to replenish those that are ionized.

9.5.2 Plane Source

We next consider what profile would be obtained in slab geometry if there is a localized source on the plane $x = 0$. Such a source might be, for instance, a slit-collimated beam of ultraviolet light strong enough to ionize the neutral gas. The steady state diffusion equation is then

$$\frac{d^2 n}{dx^2} = -\frac{Q}{D}\delta(0)$$

Except at $x = 0$, the density must satisfy $\partial^2 n / \partial x^2 = 0$. This obviously has the solution (Fig. 5.7 ADD IT)

$$n = n_0 \left(1 - \frac{|x|}{L}\right)$$

The plasma has a linear profile. The discontinuity in slope at the source is characteristic of δ -function sources.

9.5.3 Line Source

Finally, we consider a cylindrical plasma with a source located on the axis. Such a source might, for instance, be a beam of energetic electrons producing ionization along the axis. Except at $r = 0$, the density must satisfy

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial n}{\partial r}\right) = 0$$

The solution that vanishes at $r = a$ is

$$n = n_0 \ln(a/r)$$

The density becomes infinite at $r = 0$ (Fig. 5.8); it is not possible to determine the density near the axis accurately without considering the finite width of the source.

9.6 Recombination

When an ion and an electron collide, particularly at low relative velocity, they have a finite probability of recombining into a neutral atom. To conserve momentum, a third body must be present. If this third body is an emitted photon, the process is called *radiative recombination*. If it is a particle, the process is called *three-body recombination*. The loss of plasma by recombination can be represented by a negative source term in the equation of continuity. It is clear that this term will be proportional to $n_e n_i = n^2$. In the absence of the diffusion terms, the equation of continuity then becomes

$$\frac{\partial n}{\partial t} = -\alpha n^2$$

The constant of proportionality α is called the recombination coefficient and has units of m^3/s . This is a nonlinear equation for n . This means that the straightforward method for satisfying initial and boundary conditions by linear superposition of solutions is not available. Fortunately, it is such a simple nonlinear equation that the solution can be found by inspection. It is

$$\frac{1}{n(\mathbf{r}, t)} = \frac{1}{n_0(\mathbf{r})} + \alpha t$$

where $n_0(\mathbf{r})$ is the initial density distribution. After the density has fallen far below its initial value, it decays *reciprocally* with time, $n \propto 1/\alpha t$. This is a fundamentally different behavior from the case of diffusion, in which the time variation is exponential.

Figure 5.9 (ADD IT!) shows the results of measurements of the density decay in the afterglow of a weakly ionized H plasma. When the density is high, recombination, which is proportional to n^2 , is dominant, and the density decays reciprocally. After the density has reached a low value, diffusion becomes dominant, and the decay is thenceforth exponential.

9.7 Diffusion Across a Magnetic Field

The rate of plasma loss by diffusion can be decreased by a magnetic field; this is the problem of confinement in controlled fusion research. Consider a weakly ionized plasma in a magnetic field (fig-gyration-collision ADD IT!). Charged particles will move along \mathbf{B} by diffusion and mobility according to Equation 9.7, since \mathbf{B} does not affect motion in the parallel direction. Thus we have, for each species,

$$\Gamma_z = \pm \mu n E_z - D \frac{\partial n}{\partial z}$$

If there were no collisions, particles would not diffuse at all in the perpendicular direction — they would continue to gyrate about the same magnetic field line. There are, of course, particle drifts across \mathbf{B} because of electric fields or gradients in \mathbf{B} , but these can be arranged to be

parallel to the walls. For instance, in a perfectly symmetric cylinder (fig-drift-cylinder ADD IT!), the gradients are all in the radial direction, so that the guiding center drifts are in the azimuthal direction. The drifts would then be harmless.

FIGURE: A charged particle in a magnetic field will gyrate about the same line of force until it makes a collision. #fig-gyration-collision

FIGURE: Particle drifts in a cylindrically symmetric plasma column do not lead to losses.#fig-drift-cylinder

When there are collisions, particles migrate across \mathbf{B} to the walls along the gradients. They do this by a random-walk process (fig-diffusion-collision-neutral ADD IT!). When an ion, say, collides with a neutral atom, the ion leaves the collision traveling in a different direction. It continues to gyrate about the magnetic field in the same direction, but its phase of gyration is changed discontinuously. (The Larmor radius may also change, but let us suppose that the ion does not gain or lose energy on the average.)

FIGURE: Diffusion of gyrating particles by collisions with neutral atoms.#fig-diffusion-collision-neutral

The guiding center, therefore, shifts position in a collision and undergoes a random walk. The particles will diffuse in the direction opposite ∇n . The step length in the random walk is no longer λ_m , as in magnetic-field-free diffusion, but has instead the magnitude of the Larmor radius r_L . Diffusion across \mathbf{B} can therefore be slowed down by decreasing r_L ; that is, by increasing B .

To see how this comes about, we write the perpendicular component of the fluid equation of motion for either species as follows:

9.7.1 Ambipolar Diffusion Across \mathbf{B}

Because the diffusion and mobility coefficients are anisotropic in the presence of a magnetic field, the problem of ambipolar diffusion is not as straightforward as in the $B = 0$ case. Consider the particle fluxes perpendicular to \mathbf{B} (Fig. 5.13 ADD IT!). Ordinarily, since $\Gamma_{e\perp}$ is smaller than $\Gamma_{i\perp}$, a transverse electric field would be set up so as to aid electron diffusion and retard ion diffusion. However, this electric field can be short-circuited by an imbalance of the fluxes along \mathbf{B} . That is, the negative charge resulting from $\Gamma_{e\perp} < \Gamma_{i\perp}$ can be dissipated by electrons escaping along the field lines. Although the total diffusion must be ambipolar, the perpendicular part of the losses need not be ambipolar. The ions can diffuse out primarily radially, while the electrons diffuse out primarily along \mathbf{B} . Whether or not this in fact happens depends on the particular experiment. In short plasma columns with the field lines terminating on conducting plates, one would expect the ambipolar electric field to be short-circuited out. Each species then diffuses radially at a different rate. In long, thin plasma columns terminated by insulating plates, one would expect the radial diffusion to be ambipolar because escape along \mathbf{B} is arduous.

Mathematically, the problem is to solve simultaneously the continuity Equation 9.8 for ions and electrons. It is not the fluxes j , but the divergence $\nabla \cdot j$ which must be set equal to each other. Separating $\nabla \cdot j$ into perpendicular and parallel components, we have

$$\begin{aligned}\nabla \cdot i &= \nabla_{\perp} \cdot (\mu_i n \mathbf{E}_{\perp} - D_{i\perp} \nabla n) + \frac{\partial}{\partial z} \left(\mu_i n E_z - D_i \frac{\partial n}{\partial z} \right) \\ \nabla \cdot e &= \nabla_{\perp} \cdot (\mu_e n \mathbf{E}_{\perp} - D_{e\perp} \nabla n) + \frac{\partial}{\partial z} \left(-\mu_e n E_z - D_e \frac{\partial n}{\partial z} \right)\end{aligned}$$

The equation resulting from setting $\nabla \cdot i = \nabla \cdot e$ cannot easily be separated into one-dimensional equations. Furthermore, the answer depends sensitively on the boundary conditions at the ends of the field lines. Unless the plasma is so long that parallel diffusion can be neglected altogether, there is no simple answer to the problem of ambipolar diffusion across a magnetic field.

9.8 Collisions in Fully Ionized Plasmas

When the plasma is composed of ions and electrons alone, all collisions are Coulomb collisions between charged particles. However, there is a distinct difference between

- collisions between like particles (ion–ion or electron–electron collisions) and
- collisions between unlike particles (ion–electron or electron–ion collisions).

Consider two identical particles colliding (Fig. 5.16 ADD IT!). If it is a head-on collision, the particles emerge with their velocities reversed; they simply interchange their orbits, and the two guiding centers remain in the same places. The result is the same as in a glancing collision, in which the trajectories are hardly disturbed. The worst that can happen is a 90° collision, in which the velocities are changed 90° in direction. The orbits after collision will then be the dashed circles, and the guiding centers will have shifted. However, it is clear that the “center of mass” of the two guiding centers remains stationary. For this reason, collisions between like particles give rise to very little diffusion. This situation is to be contrasted with the case of ions colliding with neutral atoms. In that case, the final velocity of the neutral is of no concern, and the ion random-walks away from its initial position. In the case of ion–ion collisions, however, there is a detailed balance in each collision; for each ion that moves outward, there is another that moves inward as a result of the collision.

When two particles of opposite charge collide, however, the situation is entirely different (Fig. 5.17 ADD IT!). The worst case is now the 180° collision, in which the particles emerge with their velocities reversed. Since they must continue to gyrate about the magnetic field lines in the proper sense, both guiding centers will move in the same direction. *Unlike-particle collisions give rise to diffusion.* The physical picture is somewhat different for ions and electrons because of the disparity in mass. The electrons bounce off the nearly stationary ions and random-walk in the usual fashion. The ions are slightly jostled in each collision and move about as a result

of frequent bombardment by electrons. Nonetheless, because of the conservation of momentum in each collision, the rates of diffusion are the same for ions and electrons, as we shall show.

9.8.1 Plasma Resistivity

The fluid equations of motion including the effects of charged-particle collisions may be written as follows

$$\begin{aligned} m_i n \frac{d\mathbf{v}_i}{dt} &= en(\mathbf{E} + \mathbf{v}_i \times \mathbf{B}) - \nabla p_i - \nabla \cdot \boldsymbol{\pi}_i + \mathbf{P}_{ie} \\ m_e n \frac{d\mathbf{v}_e}{dt} &= -en(\mathbf{E} + \mathbf{v}_e \times \mathbf{B}) - \nabla p_e - \nabla \cdot \boldsymbol{\pi}_e + \mathbf{P}_{ei} \end{aligned} \quad (9.16)$$

The terms \mathbf{P}_{ie} and \mathbf{P}_{ei} represent, respectively, the momentum gain of the ion fluid caused by collisions with electrons, and vice versa. The stress tensor \mathbf{P}_j has been split into the isotropic part p_j and the anisotropic viscosity tensor $\boldsymbol{\pi}_j$. Like-particle collisions, which give rise to stresses within each fluid individually, are contained in $\boldsymbol{\pi}_j$. Since these collisions do not give rise to much diffusion, we shall ignore the terms $\nabla \cdot \boldsymbol{\pi}_j$. As for the terms \mathbf{P}_{ie} and \mathbf{P}_{ei} , which represent the friction between the two fluids, the conservation of momentum requires

$$\mathbf{P}_{ie} = -\mathbf{P}_{ei}$$

We can write \mathbf{P}_{ei} in terms of the collision frequency in the usual manner:

$$\mathbf{P}_{ei} = mn(\mathbf{v}_i - \mathbf{v}_e)\nu_{ei}$$

and similarly for \mathbf{P}_{ie} . Since the collisions are Coulomb collisions, one would expect \mathbf{P}_{ei} to be proportional to the Coulomb force, which is proportional to e^2 (for singly-charged ions). Furthermore, \mathbf{P}_{ei} must be proportional to the density of electrons n_e and to the density of scattering centers n_i , which, of course, is equal to n_e . Finally, \mathbf{P}_{ei} should be proportional to the relative velocity of the two fluids. On physical grounds, then, we can write \mathbf{P}_{ei} as

$$\mathbf{P}_{ei} = \eta e^2 n^2 (\mathbf{v}_i - \mathbf{v}_e) \quad (9.17)$$

where η is a constant of proportionality. Comparing this with the last equation, we see that

$$\nu_{ei} = \frac{ne^2}{m}\eta \quad (9.18)$$

The constant η is the *specific resistivity* of the plasma; that this jibes with the usual meaning of resistivity will become clear shortly.

9.8.2 Mechanics of Coulomb Collisions

When an electron collides with a neutral atom, no force is felt until the electron is close to the atom on the scale of atomic dimensions; the collisions are like billiard-ball collisions. When an electron collides with an ion, the electron is gradually deflected by the long-range Coulomb field of the ion. Nonetheless, one can derive an effective cross section for this kind of collision. It will suffice for our purposes to give an order-of-magnitude estimate of the cross section. In Fig. 5.18 ADD IT!, an electron of velocity \mathbf{v} approaches a fixed ion of charge e . In the absence of Coulomb forces, the electron would have a distance of closest approach r_0 , called the impact parameter. In the presence of a Coulomb attraction, the electron will be deflected by an angle χ , which is related to r_0 . The Coulomb force is

FIGURE:Orbit of an electron making a Coulomb collision with an ion.

$$F = -\frac{e^2}{4\pi\epsilon_0 r^2}$$

This force is felt during the time the electron is in the vicinity of the ion; this time is roughly

$$T \approx r_0/v$$

The change in the electron's momentum is therefore approximately

$$\Delta(mv) = |FT| \approx \frac{e^2}{4\pi\epsilon_0 r_0 v}$$

We wish to estimate the cross section for large-angle collisions, in which $\chi \geq 90^\circ$. For a 90° collision, the change in mv is of the order of mv itself. Thus

$$\Delta(mv) \simeq mv \simeq e^2/4\pi\epsilon_0 r_0 v, \quad r_0 = e^2/4\pi\epsilon_0 mv^2 \quad (9.19)$$

The cross section is then

$$\sigma = \pi r_0^2 = e^4/16\pi\epsilon_0^2 m^2 v^4$$

The collision frequency is, therefore,

$$\nu_{ei} = n\sigma v = ne^4/16\pi\epsilon_0^2 m^2 v^3 \quad (9.20)$$

and the resistivity is

$$\eta = \frac{m}{ne^2} \nu_{ei} = \frac{e^2}{16\pi\epsilon_0^2 m v^3} \quad (9.21)$$

For a Maxwellian distribution of electrons, we may replace v^2 by $k_B T_e/m$ for our order-of-magnitude estimate:

$$\eta \approx \frac{\pi e^2 m^{1/2}}{(4\pi\epsilon_0)^2 (k_B T_e)^{3/2}} \quad (9.22)$$

This is the resistivity based on large-angle collisions alone. In practice, because of the long range of the Coulomb force, small-angle collisions are much more frequent, and the cumulative effect of many small-angle deflections turns out to be larger than the effect of large-angle collisions. It was shown by Spitzer that Equation 9.22 should be multiplied by a factor $\ln \Lambda$:

$$\eta \approx \frac{\pi e^2 m^{1/2}}{(4\pi\epsilon_0)^2 (k_B T_e)^{3/2}} \ln \Lambda \quad (9.23)$$

where

$$\Lambda = \overline{\lambda_D/r_0} = 12\pi n \lambda_D^3$$

This factor represents the maximum impact parameter, in units of r_0 as given by Equation 9.19, averaged over a Maxwellian distribution. The maximum impact parameter is taken to be λ_D because Debye shielding suppresses the Coulomb field at larger distances. Although Λ depends on n and $k_B T_e$, its logarithm is insensitive to the exact values of the plasma parameters. Typical values of $\ln \Lambda$ are given below (λ_D defined in Section 2.4).

$k_B T_e$ [eV]	n [m^{-3}]	$\ln \Lambda$	
0.2	10^{15}	10.7	(Q-machine)
2	10^{17}	11.8	(lab plasma)
100	10^{19}	15.4	(typical torus)
10^4	10^{21}	20.0	(fusion reactor)
10^3	10^{27}	9.6	(laser plasma)

It is evident that $\ln \Lambda$ varies only a factor of two as the plasma parameters range over many orders of magnitude. For most purposes, it will be sufficiently accurate to let $\ln \Lambda = 10$ regardless of the type of plasma involved.

9.8.3 Physical Meaning of

Let us suppose that an electric field \mathbf{E} exists in a plasma and that the current that it drives is all carried by the electrons, which are much more mobile than the ions. Let $B = 0$ and $k_B T_e = 0$, so that $\nabla \cdot \mathbf{P}_e = 0$. Then, in steady state, the electron equation of motion from Equation 9.16 reduces to

$$en\mathbf{E} = \mathbf{P}_{ei} \quad (9.24)$$

Since $\mathbf{j} = en(\mathbf{v}_i - \mathbf{v}_e)$, Equation 9.17 can be written

$$\mathbf{P}_{ei} = \eta en\mathbf{j}$$

so that Equation 9.24 becomes

$$\mathbf{E} = \eta\mathbf{j} \quad (9.25)$$

This is simply Ohm's law, and the constant η is just the specific resistivity. The expression for η in a plasma, as given by Equation 9.22 or Equation 9.23, has several features which should be pointed out.

1. In Equation 9.23, we see that η is *independent of density* (except for the weak dependence in $\ln \Lambda$). This is a rather surprising result, since it means that if a field \mathbf{E} is applied to a plasma, the current \mathbf{j} , as given by Equation 9.25, is independent of the number of charge carriers. The reason is that although j increases with n_e , the frictional drag against the ions increases with n_i . Since $n_e = n_i$ these two effects cancel. This cancellation can be seen in Equation 9.20 and Equation 9.21. The collision frequency ν_{ei} is indeed proportional to n , but the factor n cancels out in η . A fully ionized plasma behaves quite differently from a weakly ionized one in this respect. In a weakly ionized plasma, we have $\mathbf{j} = -nev\mathbf{v}_e$, $\mathbf{v}_e = -\mu_e\mathbf{E}$, so that $\mathbf{j} = ne\mu_e\mathbf{E}$. Since the electron mobility μ_e depends only on the density of neutrals, the current is proportional to the plasma density n .
2. Equation 9.23 shows that η is proportional to $T_e^{-3/2}$. As a plasma is heated, the Coulomb cross section decreases, and the resistivity drops rather rapidly with increasing temperature. Plasmas at thermonuclear temperatures (tens of keV) are essentially collisionless; this is the reason so much theoretical research is done on collisionless plasmas. Of course, there must always be some collisions; otherwise, there wouldn't be any fusion reactions either. An easy way to heat a plasma is simply to pass a current through it. The I^2R or ηj^2 losses then turn up as an increase in electron temperature. This is called *ohmic heating*. The $(k_B T_e)^{-3/2}$ dependence of η , however, does not allow this method to be used up to thermonuclear temperatures. The plasma becomes such a good conductor at temperatures above 1 keV that ohmic heating is a very slow process in that range.
3. Equation 9.20 shows that ν_{ei} varies as v^{-3} . The fast electrons in the tail of the velocity distribution make very few collisions. The current is therefore carried mainly by these electrons rather than by the bulk of the electrons in the main body of the distribution. The strong dependence on v has another interesting consequence. If an electric field is suddenly applied to a plasma, a phenomenon known as *electron runaway* can occur. A few electrons which happen to be moving fast in the direction of $-\mathbf{E}$ when the field is applied will have gained so much energy before encountering an ion that they can make only a glancing collision. This allows them to pick up more energy from the electric field and decrease their collision cross section even further. If E is large enough, the cross section falls so fast that these runaway electrons never make a collision. They form an accelerated electron beam detached from the main body of the distribution.

9.8.4 Numerical Values of

Exact computations of η which take into account the ion recoil in each collision and are properly averaged over the electron distribution were first given by Spitzer. The following result for hydrogen is sometimes called the Spitzer resistivity:

$$\eta_{\parallel} = 5.2 \times 10^{-5} \frac{Z \ln \Lambda}{T^{3/2} [\text{eV}]} [\text{ohm} \cdot \text{m}] \quad (9.26)$$

Here Z is the ion charge number, which we have taken to be 1 elsewhere. Since the dependence on m_i is weak, these values can also be used for other gases. The subscript \parallel means that this value of η is to be used for motions parallel to \mathbf{B} . For motions perpendicular to \mathbf{B} , one should use η_{\perp} given by

$$\eta_{\perp} = 2.0 \eta_{\parallel}$$

This does not mean that conductivity along \mathbf{B} is only two times better than conductivity across \mathbf{B} . A factor like $\omega_c^2 \tau^2$ still has to be taken into account. The factor 2.0 comes from a difference in weighting of the various velocities in the electron distribution. In perpendicular motions, the slow electrons, which have small Larmor radii, contribute more to the resistivity than in parallel motions.

For $k_B T_e = 100 \text{ eV}$, Equation 9.26 yields

$$\eta = 5 \times 10^{-7} [\text{ohm} \cdot \text{m}]$$

This is to be compared with various metallic conductors:

material	η [ohm · m]
copper	2×10^{-8}
stainless steel	7×10^{-7}
mercury	1×10^6

A 100-eV plasma, therefore, has a conductivity like that of stainless steel.

9.8.5 Pulsed Currents

When a steady-state current is drawn between two electrodes aligned along the magnetic field, electrons are the dominant current carrier, and sheaths are set up at the cathode to limit the current to the set value. When the current is pulsed, however, it takes time to set up the current distribution. It was shown by Stenzel and Urrutia that this time is controlled by whistler waves (R-waves), which must travel the length of the device to communicate the voltage information.

9.8.6 Collisions Between Species

Y.Y talked about this during his lecture that the collision rate among ions or electrons is much larger than the collision rate between ions and electrons. I need to go back to the notes. See also [StackExchange Q&A](#).

9.8.7 Conductivity Tensor

Resistivity is the inverse of conductivity. Consider the effect of collisions as friction in the equation of motion for electrons:

$$m \frac{d\mathbf{u}_e}{dt} = q(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - m\nu_c(\mathbf{u}_e - \mathbf{u}')$$

where \mathbf{u}' is the velocity of the collision targets.

- In unmagnetized plasma, let $\mathbf{B} = 0$.

Assume a steady state with cold electrons ($\mathbf{u}_e = \mathbf{v}_e$) and non-moving targets $\mathbf{u}' = 0$, we can easily have

$$\mathbf{E} = -\frac{m_e \nu_c}{e} \mathbf{u}_e$$

Since the electron current is defined as

$$\mathbf{j} = -en_e \mathbf{u}_e$$

we have from the last two equations

$$\mathbf{j} = \frac{n_e e^2}{m_e \nu_c} \mathbf{E} \equiv \sigma \mathbf{E} \quad (9.27)$$

where σ is the classical conductivity.

- In magnetized plasma, let $\mathbf{B} = b\hat{z}$.

Similar to the steady unmagnetized case,

$$0 = -e(\mathbf{E} + \mathbf{u}_e \times \mathbf{B}) - \nu_c m_e \mathbf{u}_e$$

In order to avoid confusion, let σ_0 denote the classical conductivity from Equation 9.27. Using the definition of current density and writing in scalar forms, we now have

$$\begin{aligned} j_x &= \sigma_0 E_x - \frac{\Omega_e}{\nu_c} j_y \\ j_y &= \sigma_0 E_y - \frac{\Omega_e}{\nu_c} j_x \\ j_z &= \sigma_0 E_z \end{aligned}$$

the solution of which gives

$$\begin{aligned} j_x &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2} \sigma_0 E_x - \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2} E_y \\ j_y &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2} \sigma_0 E_y - \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2} E_x \\ j_z &= \sigma_0 E_z \end{aligned}$$

or in matrix form,

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_P & -\sigma_H & 0 \\ \sigma_H & \sigma_P & 0 \\ 0 & 0 & \sigma_{\parallel} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$$

where

$$\begin{aligned} \sigma_P &= \frac{\nu_c^2}{\nu_c^2 + \Omega_e^2} \sigma_0 && \text{Pedersen conductivity} && (\perp \mathbf{B} \& \parallel \mathbf{E}) \\ \sigma_H &= \frac{\Omega_e \nu_c}{\nu_c^2 + \Omega_e^2} \sigma_0 && \text{Hall conductivity} && (\perp \mathbf{B} \& \perp \mathbf{E}) \\ \sigma_{\parallel} &= \sigma_0 = \frac{n_e e^2}{m_e \nu_c} && \text{parallel conductivity} && (\parallel \mathbf{B}) \end{aligned}$$

Magnetic field-aligned currents (FACs) are very important in plasma physics. From [?@fig-conductivity](#), we see that $\Omega_e \simeq \nu_c$ is the most complicated regime where the three conductivities are comparable, and $\Omega_e \gg \nu_c$ is the most anisotropic regime as particles are tightly bound in the perpendicular direction but free to move along \mathbf{B} .

```
KeyNotes.plot_conductivity()
```

The study of conductivity is most relevant in the Earth's ionosphere, where $\Omega_e \simeq \nu_c$ happens at about 100 km above from the ground, and σ_0 is on the order of 100 S/m. Ionosphere forms within the atmosphere through the mechanisms of electron impact ionization and photoionization and in turn contribute in a crucial way to the plasma interactions. The ionosphere is electrically conductive and thus carry part of a current system that is continued outside the ionosphere by plasma currents. The properties of the ionospheric currents are usually described in terms of the *electrical conductivity* of the ionosphere. Because a strong background magnetic field threads the ionosphere, the conductivities are strongly anisotropic. In the direction of the magnetic field, the conductivity is referred to as the *parallel* or *Birkeland* conductivity. The Birkeland conductivity σ_{\parallel} is so high that it short-circuits the field-aligned component of the ionospheric electric field, and normally electrical equipotentials align with the ionospheric magnetic field lines for quasi-static (> 1 min) large-scale electrodynamic conditions. Perpendicular to the magnetic field direction, the conductivity has components parallel to the electric field σ_P and perpendicular to the electric field σ_H .

Because the electric field vanishes in the field-aligned direction through the ionosphere ($\mathbf{E}_{\parallel} = 0$), it is meaningful to integrate the transverse σ_P and σ_H along the magnetic field direction to evaluate the Pedersen and Hall conductances Σ_P and Σ_H , respectively. For any \mathbf{E} ,

$$\boldsymbol{\Sigma} \cdot \mathbf{E} = \Sigma_P \mathbf{E} + \Sigma_H \hat{b} \times \mathbf{E}$$

The continuity of currents gives

$$\begin{aligned} \nabla \cdot \mathbf{j} + \frac{\partial \rho^*}{\partial t} &= 0 \\ \nabla \cdot (\boldsymbol{\sigma} \cdot \mathbf{E}) &= -\frac{\partial \rho^*}{\partial t} \\ \int \nabla \cdot (\boldsymbol{\sigma} \cdot \mathbf{E}) dz &= - \int \frac{\partial \rho^*}{\partial t} dz \quad (9.28) \\ \nabla \cdot \left(\int \boldsymbol{\sigma} dz \cdot \mathbf{E} \right) &= -\rho^* v_{\parallel} \\ \nabla \cdot (\boldsymbol{\Sigma} \cdot \mathbf{E}) &= -j_{\parallel} \\ \nabla \cdot (\boldsymbol{\Sigma} \cdot \nabla \Phi) &= j_{\parallel} \end{aligned}$$

$\mathbf{E} = -\nabla\Phi$, and j_{\parallel} is defined such that upward ($+z$) is considered positive. Here we assume that \mathbf{B} is aligned with \hat{z} ; a more realistic geometry requires the consideration of inclination angle. See Equation 23.1 and note the sign difference between the definitions of j_{\parallel} .

The forces that act on a flux tube in the ionosphere can be expressed in terms of these conductances and the force balance expressions can consequently be used to derive other properties of the ionospheric flow such as its velocity. (???)

Mass loading contributes to the electrodynamics of the plasma interaction in ways that are analogous to the contributions of the ionospheric conductivities. Pickup ionization conserves momentum. An increase of the ion density results in a reduction of the flow velocity and its associated electric field. Thus in developing the theory of the interaction, pickup effects can be directly incorporated into the above conductances and the interaction region extends well above the atmospheric exobase.

The Pedersen current (aligned with the electric field) exerts forces that slow the flow in the ionosphere. The Hall current flows perpendicular to the electric and magnetic field, thereby breaking the symmetry of the interaction. The Hall current results in a rotation of the flow away from the corotation direction and produces ionospheric asymmetries.

9.9 Diffusion of Fully Ionized Plasmas

The magnetic induction equation can be derived from Faraday's law and Ohm's law. If magnetic diffusivity η is constant, we can write it in the form

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}$$

We define *magnetic Reynold number* R_m as the ratio of the convection term and the diffusion term:

$$R_m \equiv \frac{\nabla \times (\mathbf{u} \times \mathbf{B})}{\eta \nabla^2 \mathbf{B}} \sim \frac{uBL_0^2}{\eta L_0 B} = \frac{uL_0}{\eta}$$

There are two extreme cases of R_m . One thing that you should keep in mind is that this is no longer ideal MHD because $\mathbf{E} + \mathbf{u} \times \mathbf{B} \neq 0$.

9.9.1 Small R_m condition

If $R_m \ll 1$, the convection part can be ignored, (e.g. $\mathbf{u} \approx 0$)

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B}$$

This is a diffusion equation, meaning that unlike in ideal MHD, plasma no longer ties to the magnetic field. The diffusion time scale τ_0 can be derived as follows:

$$\frac{B}{\tau_0} = \eta \frac{B}{L_0^2} \Rightarrow \tau_0 = \frac{L_0^2}{\eta} = \mu \sigma L_0^2$$

σ is the electrical conductivity, which is a derived parameter from Ohm's law. For fully ionized plasma, $\sigma = \frac{n_i e^2 \tau_{ei}}{m_e}$.

The magnetic diffusivity can be written as

$$\eta = \frac{1}{\mu \sigma} = \frac{m_e}{\mu n_i e^2 \tau_{ei}} = \frac{m_e n_e \ln \Lambda}{\mu n_i e^2 \cdot 5.45 \times 10^5 T^{3/2}} = 5.2 \times 10^7 \frac{\ln \Lambda}{T^{3/2}}$$

where $\ln \Lambda$ is the Coulomb logarithm (generally $10 \sim 20$ which is insensitive to n and T , Section 9.8.2).

Examples:

- Evolution of solar coronal magnetic field and the energy release of solar flares in solar corona.

$$T = 10^7 K, n = 10^{14} m^{-3}, \Rightarrow \eta = 1 m^2 s^{-1}$$

For supergranules, $L_0 = 10^7 m = 10 Mm$, ($R_s = 700 Mm$), the diffusion time scale $\tau_d = 10^{14} s \approx 3 \times 10^6$ yr, which is very very long.

For flare, the observed diffusion time scale is $\tau_d = 1 s \Rightarrow L_0 = 1 m$. The magnetic reconnection (in solar corona) diffusion time scale is about $\tau_d = 100 s \Rightarrow L_0 = 10 m$. The viewing angle at 1AU to the sun is about 2000 arcsec, where 1 arcsecond corresponds to 700km on the sun. Right now the best resolution we have is about $0.2 \sim 0.5$ arcsecond, which is about 100km. This means that we are three orders of magnitude beyond the resolution we need to observe electron diffusion region on the sun!

- Requirement of dynamo

If there is no such thing as a magnetic dynamo, the geomagnetic field will have been already gone. In the core of Earth, $\eta \approx 1 m^2 s^{-1}$, radius of Earth $L_0 = 6 \times 10^6 m \Rightarrow \tau_d \approx 4 \times 10^{12} s \approx 10^6$ yr. However, the age of our Earth is about 10^9 yr $\gg 10^6$ yr. This indicates that something is generating magnetic field inside the core.

3. Current sheet

Consider an ideal configuration of magnetic field,

$$\mathbf{B} = B(x, t)\hat{y},$$

where

$$B(x, 0) = \begin{cases} +B_0 & x > 0 \\ -B_0 & x < 0 \end{cases}$$

Then the magnetic diffusion equation can be written as

$$\frac{\partial B}{\partial t} = \eta \frac{\partial^2 B}{\partial x^2}$$

There is an analytic solution

$$B(x, t) = B_0 \operatorname{erf}(\xi)$$

where

$$\xi = \frac{x}{\sqrt{4\eta t}}$$

and

$$\operatorname{erf}(\xi) = \frac{2}{\sqrt{\pi}} \int_0^\xi e^{-z^2} dz$$

is the *error function*. It has some basic properties as follows:

$$\operatorname{erf}(\xi) \Rightarrow \begin{cases} 1 & , \xi \rightarrow +\infty \\ -1 & , \xi \rightarrow -\infty \end{cases}$$

$$\operatorname{erf}(\xi) = \frac{2}{\sqrt{\pi}} \left(\xi - \frac{\xi^3}{3} + \frac{\xi^5}{10} - \frac{\xi^7}{42} + \dots \right) \approx \frac{2\xi}{\sqrt{\pi}}, |\xi| \ll 1$$

The shape is shown in Figure 9.1.

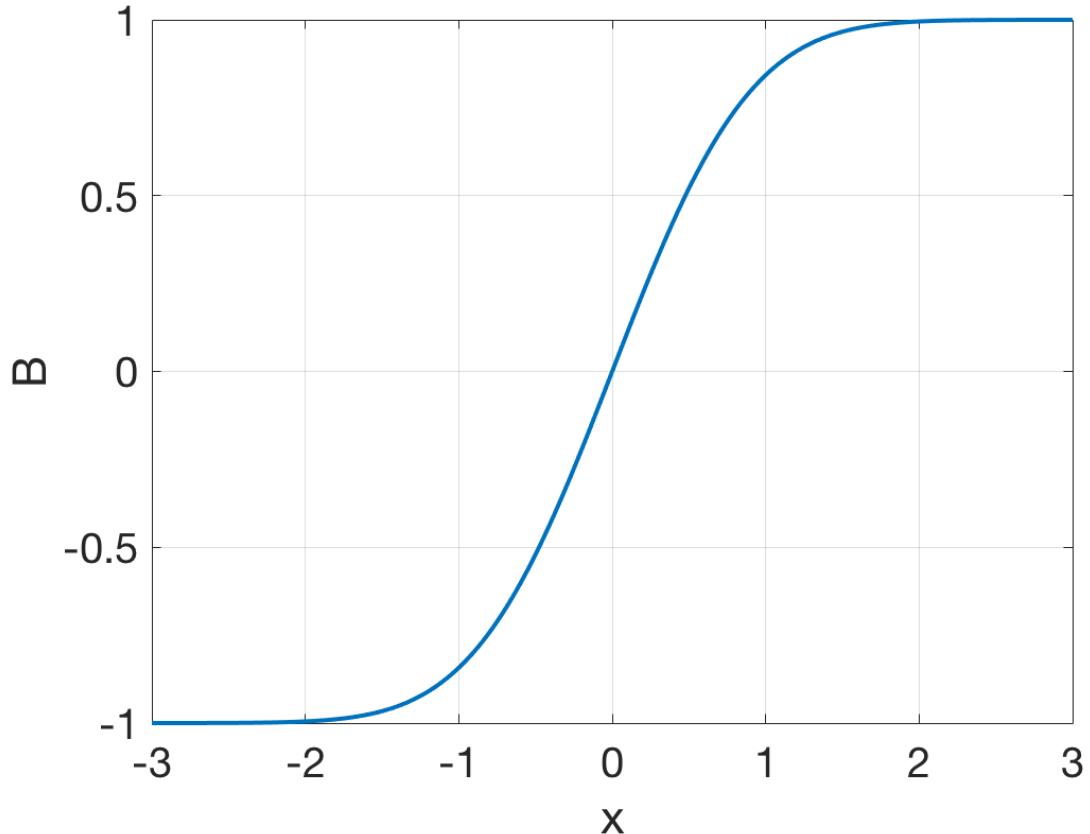


Figure 9.1: Error function.

$$|B(x, t)| \approx \begin{cases} \frac{x}{\sqrt{\pi\eta t}} B_0 & , |\xi| \ll 1 \text{ or } |x| \ll \sqrt{\pi\eta t} \\ B_0 & , |\xi| \gg 1 \text{ or } |x| \gg \sqrt{\pi\eta t} \\ 0.995 B_0 & , |\xi| = 2 \text{ or } |x| = 2\sqrt{\pi\eta t} \end{cases}$$

If you want to draw this time-variant magnetic field at two different times t_1 and t_2 ($t_1 < t_2$), you should have less dense lines at t_2 than t_1 . An example plot is present in Figure 9.2.

Conclusion:

1. In a certain amount of time, B decreases in the near region and stays the same in the far region.
- 2.

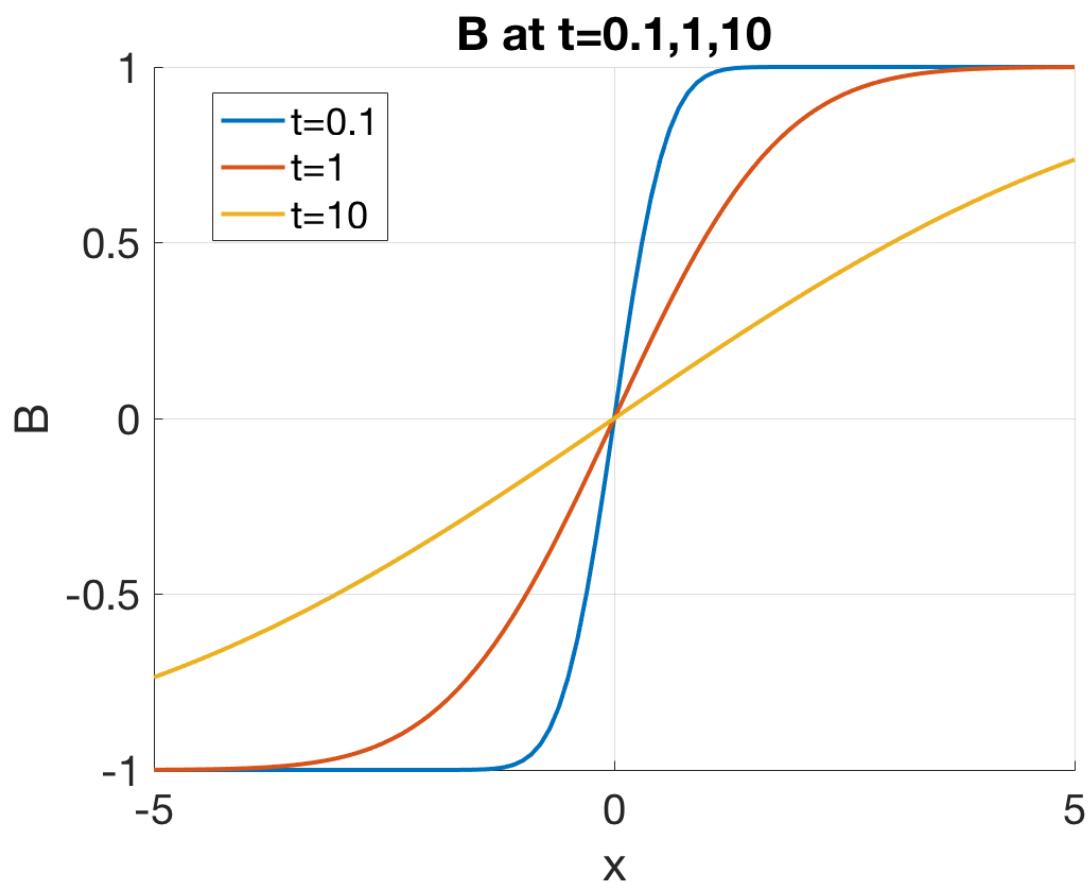


Figure 9.2: Magnetic field at $t = 0.1, 1, 10$, respectively.

$$j_z = \frac{1}{\mu} \frac{dB}{dx},$$

$$\delta = 2\sqrt{4\eta t}, \text{ (characteristic breadth)}$$

$$\frac{d\delta}{dt} = 2\sqrt{\frac{\eta}{t}}, \text{ (characteristic speed)}$$

3.

$$j = \int_{-\infty}^{\infty} j_z dx = \frac{1}{\mu} \int_{-\infty}^{\infty} dB = \frac{2B_0}{\mu} = \text{const.}$$

The current sheet gets thickened with time, but the total current in z remains the same.

4. Dissipation of magnetic energy, $W_B = \frac{B^2}{2\mu}$???

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} (W_0 - W_B) dx = -\frac{\partial}{\partial t} \int_{-\infty}^{\infty} \frac{B(x)^2}{2\mu} dx$$

9.9.2 Large R_m condition

If $R_m \ll 1$, the diffusion part can be ignored,

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B})$$

When the conductivity is large, the diffusion coefficient becomes very small, and frozen-in flux phenomenon happens.

1. Frozen-flux theorem

The magnetic field lines are frozen in the plasma flow. The magnetic flux through a surface moving with the plasma is conserved,

$$\frac{d\Phi}{dt} = 0,$$

where

$$\Phi = \int \mathbf{B} \cdot d\mathbf{A}.$$

ADD FIGURES!!!

Proof.

Consider a surface S with boundary Γ . the magnetic flux through S is

$$\Phi = \iint_S \mathbf{B} \cdot d\mathbf{A}$$

Then the time derivative of flux can be written as

$$\begin{aligned}\dot{\Phi} &= \iint_S \dot{\mathbf{B}} \cdot d\mathbf{A} + \iint_S \mathbf{B} \cdot d\dot{\mathbf{A}} \\ &= \iint_S (-\nabla \times \mathbf{E}) \cdot d\mathbf{A} + \iint_S \mathbf{B} \cdot d\dot{\mathbf{A}} \\ &= \iint_S \nabla \times (\mathbf{u} \times \mathbf{B}) \cdot dl + \iint_S \mathbf{B} \cdot d\dot{\mathbf{A}} \\ &= \oint_{\Gamma} \mathbf{u} \times \mathbf{B} \cdot dl + \iint_S \mathbf{B} \cdot d\dot{\mathbf{A}}\end{aligned}$$

The change of area can be expressed as

$$\begin{aligned}\Delta \mathbf{A} &= (\mathbf{u} \Delta t) \times dl, \\ d\dot{\mathbf{A}} &= \mathbf{u} \times dl.\end{aligned}$$

Substituting into the above equation, we have

$$\dot{\Phi} = \oint_{\Gamma} \mathbf{u} \times \mathbf{B} \cdot dl + \oint_{\Gamma} \mathbf{B} \cdot (\mathbf{u} \times dl) = 0$$

Keep in mind we are still under MHD approximation

$$\frac{\partial |\mathbf{B}|}{\partial t} \ll \Omega_i |\mathbf{B}|$$

This looks like the conservation of magnetic moment.

2. Stretching of magnetic field lines

$$\frac{d\mathbf{B}}{dt} = \frac{\partial \mathbf{B}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{B} = \nabla \times (\mathbf{u} \times \mathbf{B}) + (\mathbf{u} \cdot \nabla) \mathbf{B} = (\mathbf{B} \cdot \nabla) \mathbf{u} - \mathbf{B}(\nabla \cdot \mathbf{u})$$

where the first term on the right is the shear motion and the second term is the expansion related change.

Combined with the continuity equation,

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{u} = 0$$

we can get

$$\frac{d}{dt} \left(\frac{\mathbf{B}}{\rho} \right) = \left(\frac{\mathbf{B}}{\rho} \cdot \nabla \right) \mathbf{u} \quad (9.29)$$

Proof.

$$\begin{aligned} \frac{d\mathbf{B}}{dt} &= (\mathbf{B} \cdot \nabla) \mathbf{u} + \frac{\mathbf{B}}{\rho} \frac{d\rho}{dt} \\ \frac{1}{\rho} \frac{d\mathbf{B}}{dt} - \frac{\mathbf{B}}{\rho^2} \frac{d\rho}{dt} &= \left(\frac{\mathbf{B}}{\rho} \cdot \nabla \right) \mathbf{u} \end{aligned}$$

Compare Equation 9.29 and Equation 3.1, we have

$$\frac{\mathbf{B}}{\rho} \propto \mathbf{l}$$

i.e. \mathbf{B}/ρ and \mathbf{l} has the same form of variation. If density does not change much, the magnetic field strength will be proportional to l .

9.10 Pitch Angle Scattering

This section first follows Tautz et al. (2013). This paper gives the definition of pitch angle scattering in terms of diffusion coefficient, typical behavior of pitch angle scattering, and consequent pitch angle distribution.

Strongly pitch-angle-anisotropic phase space distribution functions cannot be properly accounted for by the diffusion approximation. Examples include solar energetic particles (SEPs), heliospheric particle populations such as Jovian electron jets and suprathermal ion species accelerated at interplanetary traveling shocks.

The key to this problem is the determination of the pitch-angle diffusion coefficient $D_{\mu\mu}$ that occurs in the Fokker-Planck transport equation. In general, one can distinguish at least three different methods of addressing this problem.

First, the wave number k -dependence of the turbulent power spectrum $G(k)$ can be specified to derive analytical approximations for $D_{\mu\mu}$. The quasi-linear theory (QLT) derived by Jokipii (1966) has been the standard theory, until it was realized that QLT is not only inaccurate but, in fact, invalid for some scenarios. For the example of isotropic turbulence, it has been known that QLT cannot properly describe pitch-angle scattering, because it neglects 90° scattering

and leads to infinitely large mean-free paths. This problem was remedied by the application of the second-order QLT (Shalchi 2005; Tautz et al. 2008), which considers deviations from the unperturbed spiral orbits that were assumed in QLT.

Second, to allow for more complex turbulence properties and to validate the permissibility of the analytical perturbation theories, one can resort to test particle simulations in specified turbulent magnetic fields. By tracing particle trajectories, the mean square displacements and the associated diffusion parameters can be obtained.

Third, rather than entirely specifying the turbulent magnetic fields, one can perform direct numerical simulations to compute solutions to the magnetohydrodynamic equations, while the test-particle trajectories are still integrated as in the previous method. Such computations do not require assumptions regarding the turbulence spectrum that is seen by the energetic particles. They are, however, limited regarding the extent of the inertial range of the turbulence spectrum, owing to computational constraints.

Perhaps the most important transport process of high-energy particles is represented by pitch-angle scattering, i.e., by stochastic variations in $\mu = v_{\parallel}/v = \cos \theta$ with range $[-1, 1]$, where θ is the angle between \mathbf{B}_0 the mean magnetic field and \mathbf{v} the particle velocity. This process is related to diffusion along the mean magnetic field, which is described by the parallel diffusion coefficient, κ_{\parallel} , or the parallel mean free path, $\lambda_{\parallel} = (3/v)\kappa_{\parallel}$, which are also related to the cosmic ray anisotropy.

The time evolution of the pitch angle is shown in Figure 9.3 for a sample of typical single-particle trajectories. It is indeed confirmed that particles with $\mu \approx [0, \pm 1]$ almost retain their original pitch angle. However, scattering through 90° can occur, a fact that is not included in QLT.

9.10.1 Definition of Pitch Angle Scattering

The usual definition of pitch-angle scattering can be found in the so-called Taylor-Green-Kubo (TGK) formalism (Taylor 1922; Green 1951; Kubo 1957; Shalchi 2011) as

$$\begin{aligned} D_{\mu\mu}(\mu) &= \int_0^\infty dt \langle \dot{\mu}(t) \dot{\mu}(0) \rangle \\ &= \frac{1}{2} \frac{d}{dt} \langle \Delta\mu(t)^2 \rangle \end{aligned} \tag{9.30}$$

where the second line employs the definition $\Delta\mu(t) = \mu(t) - \mu(0)$. It can be shown that both versions agree with each other, if t is high enough that the expression becomes asymptotically time-independent.

The combination of diffusion (Fick 1855) and random walk (Chandrasekhar 1943) motivated the usual definition of the diffusion in terms of the mean-square displacement (e.g., Tautz 2012) $\kappa = \langle \Delta x^2 \rangle / (2t)$, which can also be used for a third expression for $D_{\mu\mu}$, namely

$$D_{\mu\mu} = \frac{1}{2t} \langle \Delta\mu(t)^2 \rangle \tag{9.31}$$

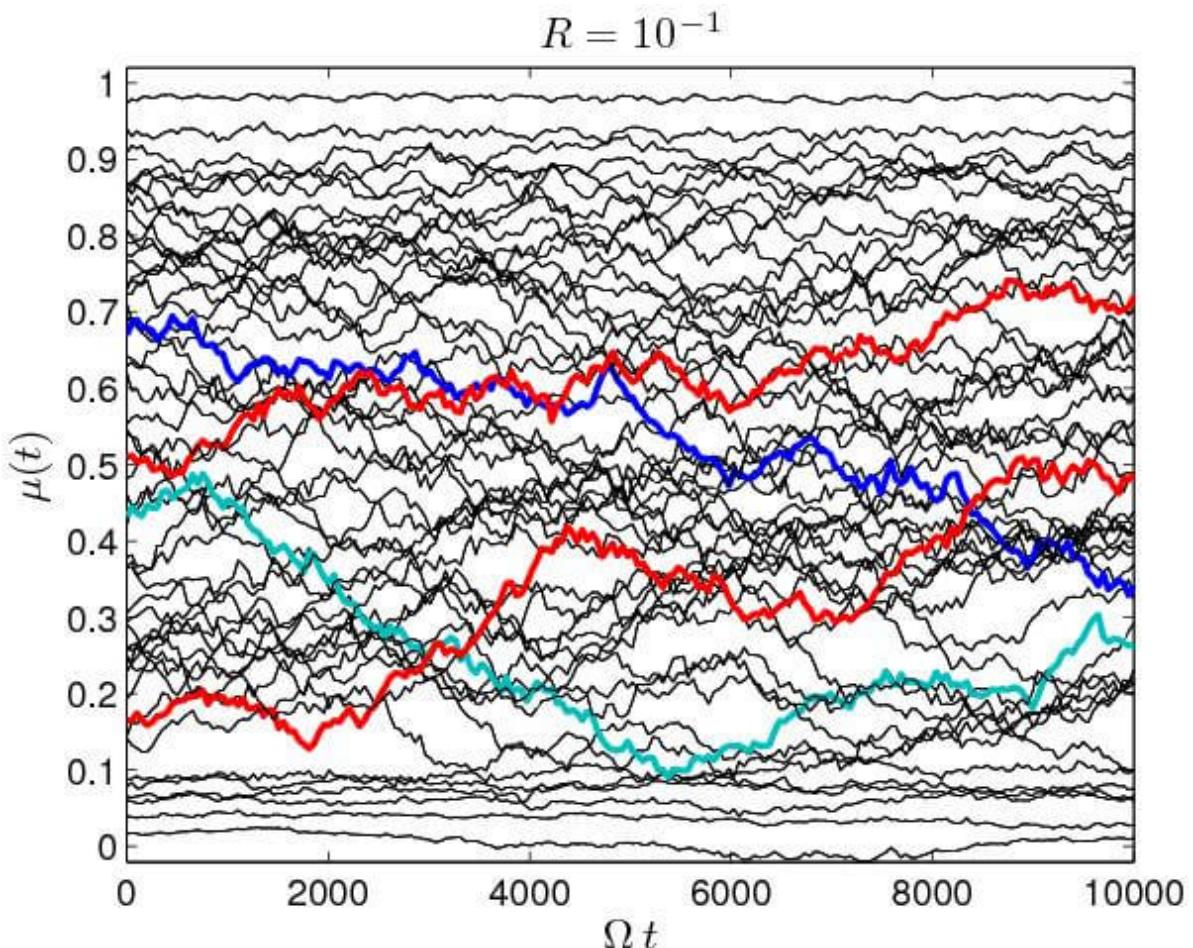


Figure 9.3: Pitch-angle cosine, $\mu = v_{\parallel}/v$ as a function of the normalized time, $\tau = \Omega t$ for a relative slab turbulence strength $\delta B/B_0 = 10^{-2}$. Four particles with initial pitch angles in the range $0.1 < \mu_0 < 0.8$ are highlighted.

which is again valid if t is high enough. However, the formal limit $t \rightarrow \infty$ is forbidden since $|\Delta\mu|$ cannot exceed a value of 2. For high enough times, $D_{\mu\mu}$ will always be dominated by the $1/t$ dependence, independent of the choice of the formula. Therefore, a meaningful, time-independent value for $D_{\mu\mu}$ can be obtained if and only if

1. t is long enough that the initial conditions become insignificant;
2. t is short enough that the behavior of $D_{\mu\mu}$ is not already dominated by the $1/t$ proportionality.

9.11 Cosmic Ray Diffusion

Cosmic-ray (CR) diffusion in turbulent and magnetized media is a fundamental problem in space physics and astrophysics. CR diffusion strongly depends on the properties of magnetohydrodynamic (MHD) turbulence that they interact with. In the presence of magnetic fields, the CR diffusion perpendicular and parallel to the magnetic field needs to be treated separately. Within the inertial range of MHD turbulence, the superdiffusion of turbulent magnetic fields causes the perpendicular superdiffusion of CRs. In the direction parallel to the magnetic field, most earlier studies are focused on the diffusion induced by gyroresonant scattering. Due to the scale-dependent anisotropy of Alfvén and slow modes in MHD turbulence, Alfvén and slow modes are inefficient in scattering the CRs with gyroradii much smaller than the turbulence injection scale, while fast modes are identified as the more efficient agent of scattering. In 2020s, strong scattering of CRs by sharp intermittent magnetic field bends in MHD turbulence is proposed to be important for affecting CR parallel diffusion. The enhanced local scattering can be associated with the intermittent and fractal structure of MHD turbulence.

The resonant scattering faces its long-standing 90° problem in the framework of the quasi-linear theory (QLT, Jokipii 1966). To resolve this issue, nonresonant interactions such as magnetic mirroring was explored. This consideration of the mirroring effect naturally solves the 90° problem and limits the pitch-angle range for gyroresonant scattering. Magnetic mirrors can naturally form in MHD turbulence due to compressions of magnetic fields, which are induced by pseudo-Alfvénic modes in an incompressible medium and slow and fast modes in a compressible medium. Unlike the trapping effect of magnetic mirrors considered for compressible MHD waves, in MHD turbulence CRs do not experience trapping due to the perpendicular superdiffusion of turbulent magnetic fields. Instead, they bounce among different magnetic mirrors and move diffusively along the local magnetic field, which leads to a new diffusion mechanism termed *mirror diffusion*. The mirror diffusion accounts for the suppressed diffusion of CRs when the mirroring condition is satisfied.

9.11.1 Perpendicular Superdiffusion of Mirroring CRs

The perpendicular superdiffusion of CRs is caused by that of turbulent magnetic fields and takes place within the inertial range of MHD turbulence (Lazarian & Yan 2014). To simulate the CR scenario, each CR can be represented by a beam of charged particles initially close to each other with the same pitch-angle cosine μ_0 and r_L . The superdiffusion is obtained by 1. measuring the perpendicular separation of each pair of particle trajectories and obtain its rms value for all pairs; 2. averaging over all beams as the measured perpendicular displacement $\langle \sqrt{\langle \delta x_\perp \rangle} \rangle$.

9.11.2 Parallel Mirror Diffusion

In the presence of a magnetic mirror with a longitudinal magnetic gradient, CRs with r_L smaller than the mirror size l_\parallel and μ satisfying

$$\mu < \sqrt{\frac{\delta B_\parallel}{B_0 + \delta B_\parallel}} \quad (9.32)$$

can be reflected by the magnetic mirror under the mirror force, where δB_\parallel is the parallel magnetic fluctuation over l_\parallel and B_0 is the mean magnetic field strength. The mirror force is

$$F_\parallel = -M \frac{\partial B_\parallel}{\partial l_\parallel}$$

where M is the magnetic moment

$$M = \frac{\gamma m u_\perp^2}{2B}$$

For mirroring CRs with their motion along the magnetic field dominated by magnetic mirroring, there is no stochastic change of μ , and M is a constant, known as the first adiabatic invariant. The particle is reflected at different mirror points (i.e., regions with stronger B) multiple times and move back and forth in the direction parallel to the magnetic field. However, in the direction perpendicular to the \mathbf{B}_0 , the particle is not trapped due to the perpendicular superdiffusion of turbulent magnetic fields (see previous subsection). As expected, CRs are not trapped between two magnetic mirror points, but diffusively move along the magnetic field in MHD turbulence due to the perpendicular superdiffusion.

9.11.3 Transition between Mirror Diffusion and Scattering Diffusion

CRs with sufficiently small μ (Equation 9.32) are subject to the nonresonant mirroring interaction and undergo the mirror diffusion. For CRs with larger μ , their parallel diffusion is dominated by the resonant pitch-angle scattering and undergo the scattering diffusion. The

transition between mirror diffusion and scattering diffusion takes place as μ stochastically changes due to the gyroresonant scattering. The transition between the two diffusion regimes is expected to occur at a critical μ as μ_c .

In compressible MHD turbulence, under the consideration that fast modes are mainly responsible for mirroring and scattering, c is given by (Lazarian & Xu, 2021)

$$\mu_c = \min\{\mu_{\max}, \mu_{\text{eq}}\}$$

where

$$\mu_{\max} = \sqrt{\frac{\delta B_f}{B_0 + \delta B_f}}$$

and

$$\mu_{\text{eq}} \approx \left[\frac{14}{\pi} \left(\frac{\delta B_f}{B_0} \right)^2 \left(\frac{r_L}{L_{\text{inj}}} \right)^{\frac{1}{2}} \right]^{\frac{2}{11}}$$

corresponds to the pitch-angle cosine where mirroring and scattering are in balance, with δB_f as the magnetic fluctuation at L_{inj} the injection length of fast modes.

For incompressible MHD turbulence, pseudo-Alfvén modes are responsible for mirroring. Under the consideration of inefficient scattering by Alfvén and pseudo-Alfvén modes for CRs with $r_L \ll L_{\text{inj}}$, we have

$$\mu_{\max} = \sqrt{\frac{\delta B_s}{B_0 + \delta B_s}}$$

where δB_s is the magnetic fluctuation of pseudo-Alfvén modes at L_{inj} . As pseudo-Alfvén modes are slaved to Alfvén modes (Beresnyak & Lazarian 2019), $\delta B_s/B_0$ can be approximately by MA. Therefore, we have

$$\mu_c = \sqrt{\frac{MA}{1 + MA}}$$

At $\mu \mu < \mu_c$, mirroring dominates scattering, and CRs are expected to undergo the mirror diffusion. At $\mu > \mu_c$, scattering becomes dominant, and CRs are expected to undergo the scattering diffusion.

9.12 Solutions of the Diffusion Equation

9.12.1 Time Dependence

9.12.2 Time-Independent Solutions

9.13 Bohm Diffusion and Neoclassical Diffusion

10 Waves

Waves are generated by instabilities (Chapter 13). In this chapter, we mainly focus on the wave propagation.

10.1 Basic Properties

10.2 Polarization

- [Electromagnetic Waves and Polarization](#)

In plasma physics, wave polarization is defined with respect to the background magnetic field \mathbf{B}_0 , not the wave propagation direction \mathbf{k} .¹

Wave Polarization

- Compressibility: Certain waves can modify plasma densities, while others can't.

10.2.1 Dispersion Relation

Waves are a very general phenomenon of most media. In order for a wave to propagate in the medium a number of conditions need to be satisfied, however. The first is that the medium allows for a particular range of frequencies ω and wave-vectors \mathbf{k} to exist in the medium; i.e., it allows for eigenmodes. These ranges are specified by the dispersion relation $D(\omega, \mathbf{k}, \dots) = 0$ which formulates the condition that the dynamical equations of the medium possess small-amplitude solutions. This dispersion relation is usually derived in the linear infinitesimally small amplitude approximation. However, nonlinear dispersion relations can sometimes also be formulated in which case $D(\omega, \mathbf{k}, |\mathbf{a}|)$ depends on the fluctuation amplitude $|\mathbf{a}|$ as well.

Dispersion relations describe the effect of dispersion on the properties of waves in a medium. A dispersion relation relates the wavelength or wavenumber of a wave to its frequency. Given the dispersion relation, one can calculate the phase velocity and group velocity of waves in the

¹This is different from the definition in optics, which makes it easier to analyze the interaction with particles since the particle gyration handedness around the magnetic field is also defined in the same way.

medium, as a function of frequency. Therefore, obtaining the dispersion relation is the key of describing the wave propagation.

Often when we say a wave mode is *dispersive*, it means that waves at different frequencies ω (or wavelength λ) travel at different phase velocities v_{ph} . Common dispersive waves include water waves in the ocean, light waves in a prism, and sound waves in bars and plates; non-dispersive waves include sound waves in air, and light waves in a vacuum.

10.2.2 Damping/Growth Rate

The solutions of the dispersion relation are in most cases complex, and for real wave vector \mathbf{k} can be written as $\omega(\mathbf{k}) = \omega_r(\mathbf{k}) + i\gamma(\omega_r, \mathbf{k})$, where the index r indicates the real part, and γ is the imaginary part of the frequency which itself is a function of the real frequency and wave number, because each mode of given frequency can behave differently in time, and the wave under normal conditions will be dispersive, i.e. it will not be a linear function of wave number. In most cases the amplitude of a given wave will change slowly in time, which means that the imaginary part of the frequency is small compared to the real frequency. If this is granted, then γ can be determined by a simple procedure directly from the dispersion relation $D(\omega, \mathbf{k}) = D_r(\omega, \mathbf{k}) + iD_i(\omega, \mathbf{k})$, which can be written as the sum of its real D_r and imaginary D_i parts because a small imaginary part γ in the frequency changes the dispersion relation only weakly, and it can be expanded with respect to this imaginary part. Up to first order in γ/ω one then obtains²

$$\begin{aligned} D_r(\omega_r, \mathbf{k}) &= 0 \\ \gamma(\omega_r, \mathbf{k}) &= -\frac{D_i(\omega, \mathbf{k})}{\partial D_r(\omega_r, \mathbf{k})/\partial\omega|_{\gamma=0}} \end{aligned} \quad (10.1)$$

The first of these expressions determines the real frequency as function of wave number $\omega_r(\mathbf{k})$ which can be calculated directly from the real part of the dispersion relation. The second equation is a prescription to determine the imaginary part of the frequency, i.e. the damping or growth rate of the wave.

10.2.3 Dielectric Function

Usually when the permittivity of a material is function of space or frequency, it is call dielectric function. The dielectric constant ϵ is a quantity which appears in electrostatic when people describe how a material screens an external time-independent electric field. When they begin to study how a material screens an external time-dependent electric field $\mathbf{E} \propto e^{-i\omega t}$ in electrodynamic sense they found that the number ϵ depends on frequency, so one gets $\epsilon(\omega)$. It would be stupid to call a quantity, which essentially depends on frequency, just “dielectric constant”, therefore one calls it “dielectric function”. Further studies showed that ϵ depends

²HOW TO DERIVE THIS?

not only on the frequency but also on the wave-vector of the field, $\mathbf{E} \propto e^{-i\omega t + ikx}$, so one gets the dielectric function $\epsilon = \epsilon(k, \omega)$.

10.3 Plasma Oscillations

If the electrons in a plasma are displaced from a uniform background of ions, electric fields will be built up in such a direction as to restore the neutrality of the plasma by pulling the electrons back to their original positions. Because of their inertia, the electrons will overshoot and oscillate around their equilibrium positions with a characteristic frequency known as the *plasma frequency*. This oscillation, also known as the Langmuir oscillation, is so fast that the massive ions do not have time to respond to the oscillating field and may be considered as fixed. In Fig. 4.2 (ADD IT!), the open rectangles represent typical elements of the ion fluid, and the darkened rectangles the alternately displaced elements of the electron fluid. The resulting charge bunching causes a spatially periodic \mathbf{E} field, which tends to restore the electrons to their neutral positions.

We shall derive an expression for the plasma frequency ω_p in the simplest case, making the following assumptions:

1. There is no magnetic field;
2. there are no thermal motions ($k_B T = 0$);
3. the ions are fixed in space in a uniform distribution;
4. the plasma is infinite in extent; and
5. the electron motions occur only in the x direction. As a consequence of the last assumption, we have

$$\nabla = \hat{x}\partial_x, \mathbf{E} = E\hat{x}, \nabla \times \mathbf{E} = 0, \mathbf{E} = -\nabla\phi$$

There is, therefore, no fluctuating magnetic field; this is an electrostatic oscillation.

The electron equations of continuity and motion are

$$\begin{aligned} \frac{\partial n_e}{\partial t} + \nabla \cdot (n_e \mathbf{u}_e) &= 0 \\ m_e n_e \left[\frac{\partial \mathbf{u}_e}{\partial t} + (\mathbf{u}_e \cdot \nabla) \mathbf{u}_e \right] &= -en_e \mathbf{E} \end{aligned}$$

The only Maxwell equation we shall need is the one that does not involve \mathbf{B} : Poisson's equation. This case is an exception to the general rule of Section 8.9 that Poisson's equation cannot be used to find \mathbf{E} . This is a high-frequency oscillation; electron inertia is important, and the deviation from neutrality is the main effect in this particular case. Consequently, we write

$$\epsilon_0 \nabla \cdot \mathbf{E} = \epsilon_0 \partial \mathbf{E} / \partial x = e(n_i - n_e)$$

The last three equations together can be easily solved by the procedure of *linearization*. By this we mean that the amplitude of oscillation is small, and terms containing higher powers of amplitude factors can be neglected. We first separate the dependent variables into two parts: an “equilibrium” part indicated by a subscript 0, and a “perturbation” part indicated by a subscript 1:

$$n_e = n_0 + n_1 \quad \mathbf{u}_e = \mathbf{u}_0 + \mathbf{u}_1 \quad \mathbf{E}_e = \mathbf{E}_0 + \mathbf{E}_1$$

The equilibrium quantities express the state of the plasma in the absence of the oscillation. Since we have assumed a uniform neutral plasma at rest before the electrons are displaced, we have

$$\begin{aligned} \nabla n_0 &= \mathbf{u}_0 = \mathbf{E}_0 = 0 \\ \frac{\partial n_0}{\partial t} &= \frac{\partial \mathbf{u}_0}{\partial t} = \frac{\partial \mathbf{E}_0}{\partial t} = 0 \end{aligned}$$

The momentum equation now becomes

$$m_e \frac{\partial \mathbf{u}_1}{\partial t} = -e \mathbf{E}$$

The term $(\mathbf{u}_1 \cdot \nabla) \mathbf{u}_1$ is seen to be quadratic in an amplitude quantity, and we shall linearize by neglecting it. The *linear theory* is valid as long as $|u_1|$ is small enough that such quadratic terms are indeed negligible. Similarly, the continuity equation becomes

$$\frac{\partial n_1}{\partial t} + n_0 \nabla \cdot \mathbf{u}_1 = 0$$

In Poisson’s equation, we note that $n_{i0} = n_{e0}$ in equilibrium and that $n_{i1} = 0$ by the assumption of fixed ions, so we have

$$\epsilon_0 \partial \mathbf{E} / \partial x = -e n_1$$

The oscillating quantities are assumed to behave sinusoidally:

$$\begin{aligned} \mathbf{n}_1 &= n_1 e^{i(kx - \omega t)} \\ \mathbf{u}_1 &= \mathbf{u}_1 e^{i(kx - \omega t)} \hat{x} \\ \mathbf{E}_1 &= \mathbf{E}_1 e^{i(kx - \omega t)} \hat{x} \end{aligned}$$

The time derivative $\partial/\partial t$ can therefore be replaced by $-i\omega$, and the gradient ∇ by $ik\hat{x}$. Now the linearized equations

$$\begin{aligned} \frac{\partial n_1}{\partial t} &= -n_0 \frac{\partial u_1}{\partial x} \\ \frac{\partial u_1}{\partial t} &= -\frac{e}{m_e} E_1 \\ \frac{\partial E_1}{\partial x} &= -\frac{e}{\epsilon_0} n_1 \end{aligned} \tag{10.2}$$

become

$$\begin{aligned}-i\omega n_1 &= -n_0 iku_1 \\ -im\omega u_1 &= -eE_1 \\ ik\epsilon_0 E_1 &= -en_1\end{aligned}\tag{10.3}$$

Eliminating u_1 and E_1 , we have

$$n_1 = \frac{n_0 e^2}{\epsilon_0 m_e \omega^2} n_1$$

If n_1 does not vanish, we must have

$$\omega^2 = \frac{n_0 e^2}{\epsilon_0 m_e}$$

The *plasma frequency* is therefore

$$\omega_p = \sqrt{\frac{n_0 e^2}{\epsilon_0 m_e}} \text{ rad/s}\tag{10.4}$$

Numerically, one can use the approximate formula

$$\omega_p / 2\pi = f_p \approx 9\sqrt{n} \text{ m}^{-3}$$

This frequency, depending only on the plasma density, is one of the fundamental parameters of a plasma. Because of the smallness of m , the plasma frequency is usually very high. For instance, in a plasma of density $n = 10^{18} \text{ m}^{-3}$, we have

$$f_p \approx 9(10^{18})^{1/2} = 9 \times 10^9 \text{ s}^{-1} = 9 \text{ GHz}$$

Radiation at f_p normally lies in the microwave range. We can compare this with another electron frequency: ω_c . A useful numerical formula is

$$f_{ce} \simeq 28 \text{ GHz/T}$$

Thus if $B = 0.32 \text{ T}$ and $n = 10^{18} \text{ m}^{-3}$, the cyclotron frequency is approximately equal to the plasma frequency for electrons.

Equation 10.4 tells us that if a plasma oscillation is to occur at all, it must have a frequency depending only on n . In particular, ω does not depend on k , so the group velocity $d\omega/dk$ is zero. The disturbance does not propagate. How this can happen can be made clear with a mechanical analogy (Fig. 4.3 fig-independent-springs). Imagine a number of heavy balls suspended by springs equally spaced in a line. If all the springs are identical, each ball will oscillate vertically with the same frequency. If the balls are started in the proper phases relative to one another, they can be made to form a wave propagating in either direction. The frequency will be fixed

by the springs, but the wavelength can be chosen arbitrarily. The two undisturbed balls at the ends will not be affected, and the initial disturbance does not propagate. Either traveling waves or standing waves can be created, as in the case of a stretched rope. Waves on a rope, however, must propagate because each segment is connected to neighboring segments.

FIGURE: Synthesis of a wave from an assembly of independent oscillators.

This analogy is not quite accurate, because plasma oscillations have motions in the direction of \mathbf{k} rather than transverse to \mathbf{k} . However, as long as electrons do not collide with ions or with each other, they can still be pictured as independent oscillators moving horizontally (in fig-independent-springs). But what about the electric field? Won't that extend past the region of initial disturbance and set neighboring layers of plasma into oscillation? In our simple example, it will not, because the electric field due to equal numbers of positive and negative infinite plane charge sheets is zero. In any finite system, however, plasma oscillations will propagate. In Fig. 4.4 ADD IT!, the positive and negative (shaded) regions of a plane plasma oscillation are confined in a cylindrical tube. The fringing electric field causes a coupling of the disturbance to adjacent layers, and the oscillation does not stay localized.

10.4 Classification of EM Waves in Uniform Plasma

$$\begin{aligned} \left\{ \begin{array}{ll} \mathbf{k} \parallel \mathbf{B}_0 & \text{Parallel Propagation,} \\ \mathbf{k} \perp \mathbf{B}_0 & \text{Perpendicular Propagation} \end{array} \right. \\ \left\{ \begin{array}{ll} \mathbf{k} \parallel \mathbf{E}_1 & \text{Longitudinal Waves,} \\ \mathbf{k} \perp \mathbf{E}_1 & \text{Transverse Waves} \end{array} \right. \\ \left\{ \begin{array}{ll} \mathbf{B}_1 = 0 & \text{Electrostatic Waves,} \\ \mathbf{B}_1 \neq 0 & \text{Electromagnetic Waves} \end{array} \right. \end{aligned}$$

Note:

1. Wave is longitudinal \iff Wave is electrostatic
2. Wave is transverse \implies Wave is electromagnetic
3. Wave is electromagnetic $\not\implies$ Wave is transverse. You can always add a component of \mathbf{E}_1 parallel to \mathbf{k} without changing \mathbf{B}_1 .

10.5 ES vs. EM Waves

A practical way to distinguish ES and EM waves is to check $\nabla \times \mathbf{E}$ and $\nabla \cdot \mathbf{E}$, where \mathbf{E} is the electric field of the wave: * If the curvature is relatively small and the divergence is relatively large, then it is likely to be ES. * Otherwise it is likely to be EM.

As we will see in Section 10.6, the dielectric function is defined in Equation 10.7. From other perspectives, the dielectric function shows up in the Ampère's law as well as the Poisson's equation

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \equiv \mu_0 \boldsymbol{\epsilon} \cdot \frac{\partial \mathbf{E}}{\partial t}$$

$$\nabla \cdot (\epsilon_0 \mathbf{E}_1) + q_j n_j \equiv \nabla \cdot (\boldsymbol{\epsilon} \cdot \mathbf{E}_1) = 0$$

Let us consider waves in an *isotropic* plasma. For isotropic plasmas, the dielectric tensor $\boldsymbol{\epsilon}$ shrinks to a scalar ϵ . For cold plasma (static ion background), the dielectric function is

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2}$$

For electrostatic (ES) waves, let $\epsilon = 0$, we have

$$\omega = \pm \omega_{pe}$$

For electromagnetic (EM) waves, from Maxwell's equations we have

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t},$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \equiv \mu_0 \epsilon \frac{\partial \mathbf{E}}{\partial t}.$$

With $\nabla \rightarrow i\mathbf{k}$, $\partial/\partial t \rightarrow -i\omega$, we can get the dispersion relation

$$i\mathbf{k} \times \mathbf{E} = i\omega \mathbf{B}$$

$$i\mathbf{k} \times \mathbf{B} = -i\mu_0 \epsilon \omega \mathbf{E}$$

$$\Rightarrow k^2 \mathbf{E} - (\mathbf{k} \cdot \mathbf{E}) \mathbf{k} = \omega^2 \mu_0 \epsilon \mathbf{E}.$$

If $\mathbf{k} \perp \mathbf{E}$, by substituting the dielectric function inside we have

$$k^2 = \omega^2 \epsilon \mu_0 = \omega^2 \epsilon_0 \mu_0 \left[1 - \frac{\omega_{pe}^2}{\omega^2} \right]$$

$$\Rightarrow \omega^2 = k^2 c^2 + \omega_{pe}^2.$$

For both waves, $\nabla \cdot (\boldsymbol{\epsilon} \mathbf{E}_1) = 0 \Rightarrow i\epsilon(\mathbf{k} \cdot \mathbf{E}_1) = 0$ is always valid. However, for electrostatic wave, $\mathbf{E}_1 = -\nabla \phi_1 = -i\mathbf{k}\phi_1 \Rightarrow \mathbf{k} \parallel \mathbf{E}_1 \Rightarrow \epsilon = 0$, while for EM wave, usually $\mathbf{k} \perp \mathbf{E}_1$ ($\mathbf{k} \perp \mathbf{E}_1 \Rightarrow$ EM wave, but EM waves do not necessarily need to be transverse. You can always add a component of \mathbf{E}_1 parallel to \mathbf{k} without changing \mathbf{B}_1), ϵ does not need to be zero. Therefore, *getting the dispersion relation by setting ϵ to 0 is only valid for isotropic ES waves*. For EM waves, there's a systematic way to get all the dispersion relations starting from dielectric function, explained in detail in Section 10.6. Here we just have a simple summary of the steps.

From Maxwell's equation for the perturbed field,

$$\begin{aligned}\nabla \times \mathbf{E}_1 &= -\mu_0 \frac{\partial \mathbf{H}_1}{\partial t} \\ \nabla \times \mathbf{H}_1 &= \mathbf{J}_1 + \epsilon_0 \frac{\partial \mathbf{E}_1}{\partial t}\end{aligned}$$

where we have assumed

$$\left\{ \begin{array}{l} \mathbf{E}_1(\mathbf{x}, t) \\ \mathbf{H}_1(\mathbf{x}, t) \end{array} \right\} = \Re \left\{ \begin{array}{l} \tilde{\mathbf{E}}_1 e^{i\mathbf{k} \cdot \mathbf{x} - i\omega t} \\ \tilde{\mathbf{H}}_1 e^{i\mathbf{k} \cdot \mathbf{x} - i\omega t} \end{array} \right\}$$

It quickly follows that

$$\begin{aligned}\mathbf{k} \times \mathbf{E}_1 &= \mu_0 \omega \mathbf{H}_1 \\ i\mathbf{k} \times \mathbf{H}_1 &= i\mathbf{k} \times \left(\frac{\mathbf{k} \times \mathbf{E}_1}{\mu_0 \omega} \right) = \mathbf{J}_1 - \epsilon_0 i\omega \mathbf{E}_1\end{aligned}$$

Then there comes the wave equation

$$\mathbf{k} \times (\mathbf{k} \times \mathbf{E}_1) = \mathbf{k}(\mathbf{k} \cdot \mathbf{E}_1) - k^2 \mathbf{E}_1 = -i\omega \mu_0 \mathbf{J}_1 - \frac{\omega^2}{c^2} \mathbf{E}_1 \equiv -\frac{\omega^2}{c^2} \frac{\epsilon}{\epsilon_0} \mathbf{E}_1$$

If we can express the total current density as a function of perturbed electric field, $\mathbf{J}_1 = \mathbf{J}_1(\mathbf{E}_1)$, from MHD, 2-fluid, or Vlasov model combining with the property of the media, we can obtain the expression for the dielectric function ϵ . With some effort, we get

$$\mathbf{A} \begin{pmatrix} E_{1x} \\ E_{1y} \\ E_{1z} \end{pmatrix} = 0$$

from which the condition for non-trivial solutions leads to

$$\det A = 0 \Rightarrow \begin{cases} \text{eigenvalue for } \omega = \omega(\mathbf{k}) \\ \text{eigenvectors } \Rightarrow \text{polarization of E field} \end{cases}$$

10.6 Cold Uniform Plasma

As long as $T_e = T_i = 0$, the linear plasma waves can easily be generalized to an arbitrary number of charged particle species and an arbitrary angle of propagation θ relative to the magnetic field. Waves that depend on finite T , such as ion acoustic waves, are not included in this treatment. The derivations go back to late 1920s when Appleton and Wilhelm Altar first calculated the cold plasma dispersion relation (CPDR).

First, we define the dielectric tensor of a plasma as follows. The fourth Maxwell equation is

$$\nabla \times \mathbf{B} = \mu_0(\mathbf{j} + \epsilon_0 \dot{\mathbf{E}})$$

where \mathbf{j} is the plasma current due to the motion of the various charged particle species s , with density n_s , charge q_s , and velocity \mathbf{v}_s :

$$\mathbf{j} = \sum_s n_s q_s \mathbf{v}_s \quad (10.5)$$

Considering the plasma to be a dielectric with internal currents \mathbf{j} , we may write Equation 10.5 as

$$\nabla \times \mathbf{B} = \mu_0 \dot{\mathbf{D}}$$

where

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \frac{i}{\omega} \mathbf{j} \quad (10.6)$$

is the **electric displacement field** or **electric induction**. It accounts for the effects of *bound charge within materials* (i.e. plasma). Here we have assumed an $\exp(-i\omega t)$ dependence for all plasma motions. Let the current \mathbf{j} be proportional to \mathbf{E} but not necessarily in the same direction (because of the magnetic field $B_0 \hat{\mathbf{z}}$); we may then define a conductivity tensor $\boldsymbol{\sigma}$ by the relation

$$\mathbf{j} = \boldsymbol{\sigma} \cdot \mathbf{E}$$

Equation 10.6 becomes

$$\mathbf{D} = \epsilon \left(\mathbf{I} + \frac{i}{\epsilon_0 \omega} \boldsymbol{\sigma} \right) \cdot \mathbf{E} = \boldsymbol{\epsilon} \cdot \mathbf{E} \quad (10.7)$$

Thus the effective dielectric constant of the plasma is the tensor

$$\boldsymbol{\epsilon} = \epsilon_0 (\mathbf{I} + i \boldsymbol{\sigma} / \epsilon_0 \omega)$$

where \mathbf{I} is the unit tensor. In electromagnetism, a **dielectric** is an electrical insulator that can be polarised by an applied electric field. When a dielectric material is placed in an electric field, electric charges do not flow through the material as they do in an electrical conductor, because they have no loosely bound, or free, electrons that may drift through the material, but instead they shift, only slightly, from their average equilibrium positions, causing dielectric polarisation.

To evaluate $\boldsymbol{\sigma}$, we use the linearized fluid equation of motion for species s , neglecting the collision and pressure terms:

$$m_s \frac{\partial \mathbf{v}_s}{\partial t} = q_s (\mathbf{E} + \mathbf{v}_s \times \mathbf{B}_0) \quad (10.8)$$

Defining the cyclotron and plasma frequencies for each species as

$$\omega_{cs} \equiv \left| \frac{q_s B_0}{m_s} \right|, \quad \omega_{ps}^2 \equiv \left| \frac{n_0 q_s^2}{\epsilon_0 m_s} \right|$$

We can separate Equation 10.8 into x, y, and z components and solve for \mathbf{v}_s , obtaining

$$\begin{aligned} v_{xs} &= \frac{i q_s}{m_s \omega} \frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1 - (\omega_{cs}/\omega)^2} \\ v_{ys} &= \frac{i q_s}{m_s \omega} \frac{E_y \mp i(\omega_{cs}/\omega)E_x}{1 - (\omega_{cs}/\omega)^2} \\ v_{zs} &= \frac{i q_s}{m_s \omega} E_z \end{aligned} \quad (10.9)$$

where \pm stands for the sign of q_s . The plasma current is

$$\mathbf{j} = \sum_s n_{0s} q_s \mathbf{v}_s$$

so that

$$\begin{aligned} \frac{i}{\epsilon_0 \omega} j_x &= \sum_s \frac{i n_{0s}}{\epsilon_0 \omega} \frac{i q_s^2}{m_s \omega} \frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1 - (\omega_{cs}/\omega)^2} \\ &= \sum_s -\frac{\omega_{ps}^2}{\omega^2} \frac{E_x \pm i(\omega_{cs}/\omega)E_y}{1 - (\omega_{cs}/\omega)^2} \end{aligned} \quad (10.10)$$

Using the identities

$$\begin{aligned} \frac{1}{1 - (\omega_{cs}/\omega)^2} &= \frac{1}{2} \left[\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}} \right] \\ \pm \frac{\omega_{cs}/\omega}{1 - (\omega_{cs}/\omega)^2} &= \frac{1}{2} \left[\frac{\omega}{\omega \mp \omega_{cs}} - \frac{\omega}{\omega \pm \omega_{cs}} \right], \end{aligned}$$

we can write Equation 10.10 as follows:

$$\begin{aligned} \frac{1}{\epsilon_0 \omega} j_x &= -\frac{1}{2} \sum_s \frac{\omega_{ps}^2}{\omega^2} \left[\left(\frac{\omega}{\omega \pm \omega_{cs}} + \frac{\omega}{\omega \mp \omega_{cs}} \right) E_x \right. \\ &\quad \left. + \left(\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}} \right) i E_y \right] \end{aligned} \quad (10.11)$$

Similarly, the y and z components are

$$\begin{aligned} \frac{1}{\epsilon_0 \omega} j_y &= -\frac{1}{2} \sum_s \frac{\omega_{ps}^2}{\omega^2} \left[\left(\frac{\omega}{\omega \pm \omega_{cs}} + \frac{\omega}{\omega \mp \omega_{cs}} \right) i E_x \right. \\ &\quad \left. + \left(\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}} \right) E_y \right] \end{aligned} \quad (10.12)$$

$$\frac{i}{\epsilon_0 \omega} j_z = - \sum_s \frac{\omega_{ps}^2}{\omega^2} E_z \quad (10.13)$$

Use of Equation 10.11 in Equation 10.6 gives

$$\begin{aligned} \frac{1}{\epsilon_0} D_x &= E_x - \frac{1}{2} \sum_s \frac{\omega_{ps}^2}{\omega^2} \left[\left(\frac{\omega}{\omega \pm \omega_{cs}} + \frac{\omega}{\omega \mp \omega_{cs}} \right) E_x \right. \\ &\quad \left. + \left(\frac{\omega}{\omega \mp \omega_{cs}} + \frac{\omega}{\omega \pm \omega_{cs}} \right) i E_y \right] \end{aligned} \quad (10.14)$$

We define the convenient abbreviations

$$\begin{aligned} R &\equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2} \left(\frac{\omega}{\omega \pm \omega_{cs}} \right) \\ L &\equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2} \left(\frac{\omega}{\omega \mp \omega_{cs}} \right) \\ S &\equiv \frac{1}{2}(R + L) \quad D \equiv \frac{1}{2}(R - L)^* \\ P &\equiv 1 - \sum_s \frac{\omega_{ps}^2}{\omega^2} \end{aligned} \quad (10.15)$$

where “R” stands for right, “L” stands for left, “S” stands for sum, “D” stands for difference, and “P” stands for plasma. Do not confuse D with the electric displacement field \mathbf{D} . Using these in Equation 10.14 and proceeding similarly with the y and z components, we obtain

$$\begin{aligned} \epsilon_0^{-1} D_x &= S E_x - i D E_y \\ \epsilon_0^{-1} D_y &= i D E_x + S E_y \\ \epsilon_0^{-1} D_z &= P E_z \end{aligned}$$

Comparing with Equation 10.7, we see that

$$\boldsymbol{\epsilon} = \epsilon_0 \begin{pmatrix} S & -iD & 0 \\ iD & S & 0 \\ 0 & 0 & P \end{pmatrix} \equiv \epsilon_0 \boldsymbol{\epsilon}_R \quad (10.16)$$

We next derive the wave equation by taking the curl of the equation $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}$ and substituting $\nabla \times \mathbf{B} = \mu_0 \boldsymbol{\epsilon} \cdot \ddot{\mathbf{E}}$, obtaining

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \epsilon_0 (\boldsymbol{\epsilon}_R \cdot \ddot{\mathbf{E}}) = -\frac{1}{c^2} \boldsymbol{\epsilon}_R \cdot \ddot{\mathbf{E}} \quad (10.17)$$

Assuming an $\exp(i\mathbf{k} \cdot \mathbf{r})$ spatial dependence of \mathbf{E} and defining a vector index of refraction

$$\mathbf{n} = \frac{c}{\omega} \mathbf{k}$$

We can write Equation 10.17 as

$$\mathbf{n} \times (\mathbf{n} \times \mathbf{E}) + \boldsymbol{\epsilon}_R \cdot \mathbf{E} = 0 \quad (10.18)$$

The uniform plasma is isotropic in the $x - y$ plane, so we may choose the y axis so that $k_y = 0$, without loss of generality. If θ is the angle between \mathbf{k} and \mathbf{B}_0 , we then have

$$n_x = n \sin \theta \quad n_z = n \cos \theta \quad n_y = 0$$

The next step is to separate Equation 10.18 into components, using the elements of $\boldsymbol{\epsilon}_R$ given in Equation 10.16. This procedure readily yields

$$\mathbf{R} \cdot \mathbf{E} \equiv \begin{pmatrix} S - n^2 \cos \theta & -iD & n^2 \sin \theta \cos \theta \\ iD & S - n^2 & 0 \\ n^2 \sin \theta \cos \theta & 0 & P - n^2 \sin^2 \theta \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = 0 \quad (10.19)$$

From this it is clear that the E_x , E_y components are coupled to E_z only if one deviates from the principal angles $\theta = 0, 90^\circ$.

Equation 10.19 is a set of three simultaneous, homogeneous equations; the condition for the existence of a solution is that the determinant of \mathbf{R} vanish: $||\mathbf{R}|| = 0$. We then obtain

$$(iD)^2(P - n^2 \sin^2 \theta) + (S - n^2) \times [(S - n^2 \cos^2 \theta)(P - n^2 \sin^2 \theta) - n^4 \sin^2 \theta \cos^2 \theta] = 0 \quad (10.20)$$

By replacing $\cos^2 \theta$ by $1 - \sin^2 \theta$, we can solve for $\sin^2 \theta$, obtaining

$$\sin^2 \theta = \frac{-P(n^4 - 2Sn^2 + RL)}{n^4(S - P) + n^2(PS - RL)}$$

We have used the identity $S^2 - D^2 = RL$. Similarly,

$$\cos^2 \theta = \frac{Sn^4 - (PS + RL)n^2 + PRL}{n^4(S - P) + n^2(PS - RL)}$$

Dividing the last two equations, we obtain

$$\tan^2 \theta = \frac{-P(n^4 - 2Sn^2 + RL)}{Sn^4 - (PS + RL)n^2 + PRL}$$

Since $2S = R + L$, the numerator and denominator can be factored to give the cold-plasma dispersion relation

$$\tan^2 \theta = \frac{-P(n^2 - R)(n^2 - L)}{(Sn^2 - RL)(n^2 - P)} \quad (10.21)$$

10.6.1 Wave Modes

The principal modes of cold plasma waves can be recovered by setting $\theta = 0^\circ$ and 90° . When $\theta = 0^\circ$,

$$P(n^2 - R)(n^2 - L) = 0$$

There are three roots:

- $P = 0$ (Langmuir wave)
- $n^2 = R$ (R wave)
- $n^2 = L$ (L wave)

When $\theta = 90^\circ$,

$$(Sn^2 - RL)(n^2 - P) = 0$$

There are two roots:

- $n^2 = RL/S$ (extraordinary wave)
- $n^2 = P$ (ordinary wave)

By inserting the definitions of Equation 10.15, one can verify that these are identical to the dispersion relations given in separate derivations, with the addition of corrections due to ion motions.

10.6.2 Resonances

The resonances can be found by letting n go to ∞ . We then have

$$\tan^2 \theta_{res} = -P/S$$

This shows that the resonance frequencies depend on angle θ .

- If $\theta = 0^\circ$, the possible solutions are $P = 0$ and $S = \infty$. The former is the plasma resonance $\omega = \omega_p$, while the latter occurs when either $R = \infty$ (i.e. $\omega = \omega_{ce}$, electron cyclotron resonance) or $L = \infty$ (i.e. $\omega = \omega_{ci}$, ion cyclotron resonance).
- If $\theta = 90^\circ$, the possible solutions are $P = \infty$ or $S = 0$. The former cannot occur for finite ω_p and ω , and the latter yields the upper and lower hybrid frequencies, as well as the two-ion hybrid frequency when there is more than one ion species.

10.6.3 Cutoffs

The cutoffs can be found by setting $n = 0$ in Equation 10.21. Again using $S^2 - D^2 = RL$, we find that the condition for cutoff is independent of θ :

$$PRL = 0$$

- The conditions $R = 0$ and $L = 0$ yield the ω_R and ω_L cutoff frequencies, with the addition of ion corrections.

For R-waves, since $\omega_{pi}^2 \ll \omega_{pe}^2, \omega_{ci} \ll \omega_{ce}$, the cutoff frequency can be approximated by

$$\begin{aligned} 1 - \frac{\omega_{pe}^2}{\omega(\omega - \omega_{ce})} - \frac{\omega_{pi}^2}{\omega(\omega + \omega_{ci})} &= 0 \\ 1 = \frac{\omega_{pe}^2 \left[\omega \left(1 + \frac{\omega_{pi}^2}{\omega_{pe}^2} \right) + \omega_{ci} - \frac{\omega_{pi}^2}{\omega_{pe}^2} \omega_{ce} \right]}{\omega_{ce} \omega (\omega - \omega_{ce}) \left(\frac{\omega}{\omega_{ce}} + \frac{\omega_{ci}}{\omega_{ce}} \right)} \\ 1 = \frac{\omega_{pe}^2 (\omega + \omega_{ci})}{\omega^2 (\omega - \omega_{ce})} \\ \omega^3 - \omega_{ce} \omega^2 - \omega_{pe}^2 \omega - \omega_{pe}^2 \omega_{ci} &= 0 \end{aligned}$$

Here somehow we can ignore $\omega_{pe}^2 \omega_{ci}$ (I DON'T KNOW WHY???) and obtain the positive solution

$$\omega_{R=0} \approx \frac{\omega_{ce}}{2} \left[1 + \sqrt{1 + 4\omega_{pe}^2/\omega_{ce}^2} \right] \quad (10.22)$$

In the low density limit, $\omega_p \ll \omega_c$, $(1+x)^{1/2} \approx 1 + x/2$ when $x \rightarrow 0$,

$$\omega_{R=0} \approx \omega_{ce} (1 + \omega_{pe}^2/\omega_{ce}^2)$$

In the high density limit, $\omega_p \gg \omega_c$,

$$\omega_{R=0} \approx \omega_{pe} + \omega_{ce}/2$$

Similarly for L-waves, the cutoff frequency can be approximated by

$$\omega_{L=0} \approx \frac{\omega_{ce}}{2} \left[-1 + \sqrt{1 + 4\omega_{pe}^2/\omega_{ce}^2} \right] \quad (10.23)$$

In the low density limit, $\omega_p \ll \omega_c$,

$$\omega_{L=0} \approx \omega_{pe}^2 / \omega_{ce}$$

In the high density limit, $\omega_p \gg \omega_c$,

$$\omega_{L=0} \approx \omega_{pe} - \omega_{ce}/2$$

- The condition $P = 0$ is seen to correspond to cutoff as well as to resonance. This degeneracy is due to our neglect of thermal motions. Actually, $P = 0$ (or $\omega = \omega_p$) is a resonance for longitudinal waves and a cutoff for transverse waves.

10.6.4 Polarizations

The information contained in Equation 10.21 is summarized in the Clemmow–Mullaly–Allis (CMA) diagram³. One further result, not in the diagram, can be obtained easily from this formulation. The middle line of Equation 10.19 reads

$$iDE_x + (S - n^2)E_y = 0$$

Thus the polarization in the plane perpendicular to \mathbf{B}_0 is given by

$$\frac{iE_x}{E_y} = \frac{n^2 - S}{D}$$

From this it is easily seen that waves are linearly polarized at resonance ($n^2 = \infty$) and circularly polarized at cutoff ($n^2 = 0$, $R = 0$ or $L = 0$; thus $S = \pm D$).

10.6.5 Low Frequency Limit

It is very useful to obtain the circularly polarized wave dispersion relation in the low frequency regime.

The R-wave corresponds to electron. When $\omega \ll \omega_{ce}$,

$$\begin{aligned} n^2 &= R = 1 - \frac{\omega_{pi}^2}{\omega(\omega + \omega_{ci})} - \frac{\omega_{pe}^2}{\omega(\omega - \omega_{ce})} \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &\simeq 1 + \frac{\omega_{pe}^2}{\omega \omega_{ce}} - \frac{\omega_{pi}^2}{\omega \omega_{ci}} \frac{\omega_{ci}}{\omega_{ci} + \omega} \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &= 1 + \frac{\omega_{pi}^2}{\omega \omega_{ci}} - \frac{\omega_{pi}^2}{\omega \omega_{ci}} \frac{1}{1 + \omega/\omega_{ci}} \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &= 1 + \frac{\omega_{pi}^2}{\omega \omega_{ci}} \left(1 - \frac{1}{1 + \omega/\omega_{ci}} \right) \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &= 1 + \frac{\omega_{pi}^2}{\omega_{ci}^2} \frac{1}{1 + \omega/\omega_{ci}} \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &= 1 + \frac{c^2}{V_A^2} \frac{1}{1 + \omega/\omega_{ci}} \\ \frac{k_{\parallel}^2 c^2}{\omega^2} &\simeq \frac{c^2}{V_A^2} \frac{1}{1 + \omega/\omega_{ci}} \\ \frac{k_{\parallel}^2}{\omega^2} &= \frac{1}{V_A^2} \frac{1}{1 + \omega/\omega_{ci}} \end{aligned} \tag{10.24}$$

³A [kinetic CMA diagram](#) is also generated.

Equation 10.24 can be arranged into a quasi-quadratic equation for $v_{\text{ph}} = \omega/k_{\parallel}$

$$\frac{\omega^2}{k_{\parallel}^2} = v_A^2 + v_A^2 \frac{\omega}{\omega_{ci}} = v_A^2 + W \frac{\omega}{k_{\parallel}}$$

where we denote $W = \frac{v_A^2 k_{\parallel}}{\omega_{ci}}$. The solution is then

$$v_w = \frac{W}{2} \pm \sqrt{\frac{W^2}{4} + v_A^2} \quad (10.25)$$

When $v_A \ll W$, i.e. $\frac{\omega_{ci}}{k_{\parallel}} \ll v_A$, we have $v_w = W$.

From Gary (1993), under the assumption of $\omega \ll \omega_{ci}$, $n^2 = R$ yields

$$\omega^2 = k^2 v_A^2 \left(1 + \frac{kc}{\omega_{pi}} \right)$$

The derivation can be found in the appendix of the paper Ion beam instability model for the Mercury upstream waves. The derivation was neglected in Gary's book.

For $\omega_{ci} \ll \omega \ll \omega_{ce}$, we can make further simplification:

$$k^2 c^2 = \omega^2 \left(1 + \frac{\omega_{pe}^2}{\omega \omega_{ce}} \right)$$

This is the *whistler wave*, with group velocity $v_g = \partial\omega/\partial k \propto \sqrt{\omega}$. It means that high frequency waves transpose energy faster than low frequency waves. In other words, one will hear high frequency components earlier than low frequency components, creating a “whistler effect”. This was discovered during the first world war, and the theoretical explanation came out in the 1950s. Also note that since whistler wave travels along the field line, in near-Earth space we have signals traveling from the south hemisphere to the north hemisphere within this frequency regime. [Here](#) is an observation example from Palmer station, Antarctica. For $\omega \ll \omega_{ci}$, Alfvén wave is recovered.

See more in Section 10.16.

The L-wave corresponds to ion. When $\omega < \omega_{ci}$, $c \gg V_A$,

$$\begin{aligned} \frac{k^2 c^2}{\omega^2} &= \omega^2 \left(1 + \frac{c^2}{v_A^2} \frac{\omega_{ci}}{\omega_{ci} - \omega} \right) \\ \frac{\omega^2}{k^2} &= V_A^2 \left(1 - \frac{\omega}{\omega_{ci}} \right) \end{aligned}$$

For $\omega \lesssim \omega_{ci}$, we get the ion cyclotron wave; for $\omega \ll \omega_{ci}$, Alfvén wave is recovered.

?@fig-dispersion-parallel shows the dispersion relations for L/R waves in a rough scale (ACTUALLY THE SCALES ARE SO BAD...). Above the cut-off frequencies ($\omega_{R=0}$ and $\omega_{L=0}$) the solution to the wave dispersion equation is called the *free-space mode*. Below electron and ion cyclotron frequencies the waves are called the *cyclotron modes*. At low frequencies ($\omega \rightarrow 0$) L- and R-modes merge and the dispersion becomes that of the shear Alfvén wave $n^2 \rightarrow c^2/v_A^2$.

```
KeyNotes.plot_dispersion_parallel()
```

The dispersion curve for a R-wave propagating parallel to the equilibrium magnetic field is sketched in ?@fig-dispersion-R-wave. The continuation of the Alfvén wave above the ion cyclotron frequency is called the *electron cyclotron wave*, or sometimes the *whistler wave*. The latter terminology is prevalent in ionospheric and space plasma physics contexts. The phase speed is mostly super-Alfvénic except near the electron gyrofrequency. The wave which propagates above the cutoff frequency, ω_1 , is a standard right-handed circularly polarized electromagnetic wave, somewhat modified by the presence of the plasma. Note that the low-frequency branch of the dispersion curve differs fundamentally from the high-frequency branch, because the former branch corresponds to a wave which can only propagate through the plasma in the presence of an equilibrium magnetic field, whereas the high-frequency branch corresponds to a wave which can propagate in the absence of an equilibrium field.

For a L-wave, similar considerations to the above give a dispersion curve of the form sketched in ?@fig-dispersion-L-wave. In this case, n^2 goes to infinity at the ion cyclotron frequency, Ω_i , corresponding to the so-called *ion cyclotron resonance* (at $L \rightarrow \infty$). At this resonance, the rotating electric field associated with a left-handed wave resonates with the gyromotion of the ions, allowing wave energy to be converted into perpendicular kinetic energy of the ions. There is a band of frequencies, lying above the ion cyclotron frequency, in which the left-handed wave does not propagate. At very high frequencies a propagating mode exists, which is basically a standard left-handed circularly polarized electromagnetic wave, somewhat modified by the presence of the plasma.

As before, the lower branch in ?@fig-dispersion-L-wave describes a wave that can only propagate in the presence of an equilibrium magnetic field, whereas the upper branch describes a wave that can propagate in the absence an equilibrium field. The continuation of the Alfvén wave to just below the ion cyclotron frequency is generally called the *ion cyclotron wave*. Note that the phase speed is always sub-Alfvénic.

10.6.6 Faraday Rotation

A linearly polarized plane wave can be expressed as a sum of left- and right-hand circularly polarized waves (R- and L-modes having equal amplitudes, E_0). If we assume that at $z = 0$,

the wave is linearly polarized along the x -axis, and that the wave vector \mathbf{k} and the background magnetic field \mathbf{B}_0 are along the z -axis, we can write

$$\mathbf{E} = E_0[(e^{ik_R z} + e^{ik_L z})\hat{x} + (e^{ik_R z} - e^{ik_L z})\hat{y}]e^{-i\omega t}$$

The ratio of the E_x and E_y components is

$$\frac{E_x}{E_y} = \cot\left(\frac{k_L - k_R}{2}z\right)$$

Hence, due to different phase speeds of R- and L-modes the linearly polarized wave that is travelling along a magnetic field will experience the rotation of its plane of polarization. This is called *Faraday rotation*. The magnitude of the rotation depends on the density and magnetic field of the plasma. Considering frequencies above the plasma frequency one can show that the rate of change in the rotation angle ϕ with the distance travelled (assumed here to be in the z -direction) is

$$\frac{d\phi}{dz} = \frac{-e^3}{2m_e^2\epsilon_0 c\omega^2} n_e B_0$$

and the total rotation from the source to the observer is

$$\phi = \frac{-e^3}{2m_e^2\epsilon_0 c\omega^2} \int_0^d n_e \mathbf{B} \cdot d\mathbf{s}$$

where $d\mathbf{s}$ is along the wave propagation path. The total rotation thus depends on both the density and magnetic field of the medium.

Faraday rotation is an important diagnostic tool both in laboratories and in astronomy. It can be used to obtain information of the magnetic field of the cosmic plasma. Note that density has to be known using other methods. On the other hand, if the magnetic field is known, Faraday rotation can give information of the density.

10.6.7 Perpendicular Wave Propagation

Let us now consider wave propagation, at arbitrary frequencies, perpendicular to the equilibrium magnetic field, i.e. $\theta = 90^\circ$.

The cutoff frequencies, at which n^2 goes to zero, are the roots of $R = 0$ and $L = 0$ according to $n^2 = LR/S$. In fact, we have already solved these equations in the previous sections (recall that cutoff frequencies do not depend on θ). There are two cutoff frequencies, $\omega_{R=0}$ and $\omega_{L=0}$, which are specified by Equation 10.22 and Equation 10.23, respectively.

Let us, next, search for the resonant frequencies, at which n^2 goes to infinity. According to the previous discussions, the resonant frequencies are solutions of

$$S = 1 - \frac{\omega_{pe}^2}{\omega^2 - \Omega_e^2} - \frac{\omega_{pi}^2}{\omega^2 - \Omega_i^2} = 0 \quad (10.26)$$

The roots of this equation can be obtained as follows. First, we note that if the first two terms are equated to zero, we obtain $\omega = \omega_{\text{UH}}$, where

$$\omega_{\text{UH}} \equiv \sqrt{\omega_{pe}^2 + \Omega_e^2} \quad (10.27)$$

If this frequency is substituted into the third term, the result is far less than unity. We conclude that ω_{UH} is a good approximation to one of the roots of Equation 10.26. To obtain the second root, we make use of the fact that the product of the square of the roots is

$$\Omega_e^2 \Omega_i^2 + \omega_{pe}^2 \Omega_i^2 + \omega_{pi}^2 \Omega_e^2 \simeq \Omega_e^2 \Omega_i^2 + \omega_{pi}^2 \Omega_e^2$$

We, thus, obtain $\omega = \omega_{\text{LH}}$, where

$$\omega_{\text{LH}} \equiv \sqrt{\frac{\Omega_e^2 \Omega_i^2 + \omega_{pi}^2 \Omega_e^2}{\omega_{pe}^2 + \Omega_e^2}} \quad (10.28)$$

The first resonant frequency, ω_{UH} , is greater than the electron cyclotron or plasma frequencies, and is called the *upper hybrid frequency*. The second resonant frequency, ω_{LH} , lies between the electron and ion cyclotron frequencies, and is called the *lower hybrid frequency*. (F. F. Chen 2016) gave some nice explanations of the physical origins of these frequencies by looking at the electrostatic electron/ion waves perpendicular to \mathbf{B} . At low frequencies, the mode in question reverts to the compressional-Alfvén wave discussed previously. Note that the shear-Alfvén wave does not propagate perpendicular to the magnetic field.

Using the above information, and the easily demonstrated fact that

$$\omega_{\text{LH}} < \omega_{L=0} < \omega_{\text{UH}} < \omega_{R=0}$$

we can deduce that the dispersion curve for the mode in question takes the form sketched in **?@fig-dispersion-perp-wave**. The lowest frequency branch corresponds to the compressional-Alfvén wave. The other two branches constitute the *extraordinary*, or *X*-, wave. The upper branch is basically a linearly polarized (in the y -direction) electromagnetic wave, somewhat modified by the presence of the plasma. This branch corresponds to a wave which propagates in the absence of an equilibrium magnetic field. The lowest branch corresponds to a wave which does not propagate in the absence of an equilibrium field. Finally, the middle branch corresponds to a wave which converts into an electrostatic plasma wave in the absence of an equilibrium magnetic field.

Wave propagation at oblique angles is generally more complicated than propagation parallel or perpendicular to the equilibrium magnetic field, but does not involve any new physical effects.

10.7 MHD Waves

10.7.1 Cold MHD

By ignoring pressure, gravity, viscosity and rotation, we have

$$\begin{aligned}\rho \frac{\partial \mathbf{u}}{\partial t} &= \mathbf{j} \times \mathbf{B}_0 \\ \mathbf{E} &= -\mathbf{u} \times \mathbf{B}_0 \\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}_1}{\partial t} \\ \nabla \cdot \mathbf{B}_1 &= 0 \\ \nabla \times \mathbf{B}_1 &= \mu_0 \mathbf{j}\end{aligned}\tag{10.29}$$

As usual in wave analysis, $\mathbf{u}, \mathbf{j}, \mathbf{E}$ are treated as perturbations. The MHD wave equation for the electric field can then be obtained,

$$\begin{aligned}\dot{\mathbf{E}} &= -\dot{\mathbf{u}} \times \mathbf{B}_0 = -\frac{1}{\rho}(\mathbf{j} \times \mathbf{B}) \times \mathbf{B}_0 = -\frac{1}{\mu_0 \rho}[(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0] \times \mathbf{B}_0 \\ \ddot{\mathbf{E}} &= [(\nabla \times (\nabla \times \mathbf{E})) \times \mathbf{V}_A] \times \mathbf{V}_A\end{aligned}$$

where $\mathbf{V}_A = \mathbf{B}_0 / \sqrt{\mu_0 \rho}$ is the Alfvén velocity, or if we mutate the triad cross terms,

$$\ddot{\mathbf{E}} = \mathbf{V}_A \times [\mathbf{V}_A \times \nabla \times (\nabla \times \mathbf{E})]\tag{10.30}$$

Alternatively, we can also get the MHD wave equation for the magnetic field:

$$\begin{cases} \dot{\mathbf{B}}_1 = \nabla \times (\mathbf{u} \times \mathbf{B}_0) \\ (\nabla \times \mathbf{B}_1) \times \mathbf{B}_0 = \mu_0 \mathbf{j} \times \mathbf{B}_0 = \mu_0 \rho \dot{\mathbf{u}} \end{cases} \Rightarrow \ddot{\mathbf{B}}_1 = \nabla \times \left[\left(\frac{1}{\mu_0 \rho} (\nabla \times \mathbf{B}_1) \times \mathbf{B}_0 \right) \times \mathbf{B}_0 \right]$$

or

$$\ddot{\mathbf{B}}_1 = \nabla \times \left[((\nabla \times \mathbf{B}_1) \times \mathbf{V}_A) \times \mathbf{V}_A \right]\tag{10.31}$$

We will see soon that in cold MHD the slow mode ceases to exist, and the fast mode moves at Alfvén speed, such that along the magnetic field line, we only have a single wave mode.

10.7.2 Hot MHD

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0 \\
\rho \frac{d\mathbf{v}}{dt} &= -\nabla p + \mathbf{j} \times \mathbf{B} \\
\mathbf{j} &= \frac{1}{\mu_0} \nabla \times \mathbf{B} \\
\frac{d}{dt} (p \rho^{-\gamma}) &= 0 \\
\frac{\partial \mathbf{B}}{\partial t} &= -\nabla \times \mathbf{E} \\
\mathbf{E} &= -\mathbf{v} \times \mathbf{B}
\end{aligned}$$

$\dot{\mathbf{E}}$ is ignored because we only consider low frequency waves. We assume no background flow, $\mathbf{u}_0 = 0$, so the current is purely caused by perturbed velocity \mathbf{u}_1 . Performing linearization and plane wave decomposition:

$$\begin{aligned}
-i\omega \rho_1 + i\rho_0 \mathbf{k} \cdot \mathbf{v} &= 0 \\
-i\omega \rho_0 \mathbf{v} &= -i\mathbf{k} p_1 + \mathbf{j} \times \mathbf{B}_0 \\
\mathbf{j} &= \frac{1}{\mu_0} i\mathbf{k} \times \mathbf{B}_1 \\
p_1/p_0 - \gamma \rho_1/\rho_0 &= 0 \\
-i\omega \mathbf{B}_1 &= i\mathbf{k} \times (\mathbf{v} \times \mathbf{B}_0)
\end{aligned}$$

Let $\mathbf{B}_0 = B_0 \hat{z}$. The linearized equations can be further simplified:

$$\begin{aligned}
-i\omega \rho_0 \mathbf{v} &= -i\mathbf{k} \left(\gamma p_0 \frac{\mathbf{k} \cdot \mathbf{v}}{\omega} \right) + \left[\frac{1}{\mu_0} i\mathbf{k} \times \left(-\frac{\mathbf{k} \times (\mathbf{v} \times \mathbf{B}_0)}{\omega} \right) \right] \times \mathbf{B}_0 \\
\omega^2 \mathbf{v} - v_s^2 \mathbf{k}(\mathbf{k} \cdot \mathbf{v}) - v_A^2 [\mathbf{k} \times (\mathbf{k} \times (\mathbf{v} \times \hat{z}))] \times \hat{z} &= 0
\end{aligned}$$

where $v_s = \sqrt{\frac{\gamma p_0}{\rho_0}}$ is the sound speed, and $v_A = \sqrt{\frac{B_0^2}{\mu_0 \rho_0}}$ is the Alfvén speed. If we write $\mathbf{V}_A = \mathbf{B}_0 / \sqrt{\mu_0 \rho_0}$, this can also be written as

$$\omega^2 \mathbf{v} - v_s^2 \mathbf{k}(\mathbf{k} \cdot \mathbf{v}) - \mathbf{k} \times \mathbf{k} \times (\mathbf{v} \times \mathbf{V}_A) \times \mathbf{V}_A = 0$$

Due to the symmetry in the perpendicular x-y plane, for simplicity, we assume the wave vector \mathbf{k} lies in the x-z plane with an angle w.r.t. the z axis θ :

$$\mathbf{k} = k_x \hat{x} + k_z \hat{z} = k_x \hat{x} + k_{\parallel} \hat{z} = k \sin \theta \hat{x} + k \cos \theta \hat{z}$$

Now it can be written as

$$\begin{pmatrix}
-\omega^2/k^2 + v_A^2 + v_s^2 \sin^2 \theta & 0 & v_s^2 \sin \theta \cos \theta \\
0 & -\omega^2/k^2 + v_A^2 \cos^2 \theta & 0 \\
v_s^2 \sin \theta \cos \theta & 0 & -\omega^2/k^2 + v_s^2 \cos^2 \theta
\end{pmatrix}
\begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} = 0 \quad (10.32)$$

10.7.3 Alfvén Wave

For any nonzero v_y , the y -component of Equation 10.32 gives

$$\omega^2 = k^2 v_A^2 \cos^2 \theta = k_{\parallel}^2 v_A^2$$

which is known as the *Alfvén wave*, in a uniform plasma immersed in a uniform background magnetic field with phase speed

$$v_p = v_A \cos \theta$$

The group velocity and hence energy propagation is always parallel to \mathbf{B} regardless of the direction of \mathbf{k} , and for this reason this mode is also known as the *guided mode*. This property, of course, has the direct bearing on the feature of Alfvén wave resonant absorption.

Given the velocity perturbation $\mathbf{v}_1 = (0, v_y, 0)$, $-i\omega\rho_1 + \rho_0\mathbf{k} \cdot \mathbf{v} = 0$, $\omega\mathbf{B}_1 + \mathbf{k} \times (\mathbf{v} \times \mathbf{B}_0) = 0$, the other perturbations are given as

$$\begin{aligned} \rho_1 &= 0 \\ p_1 &= 0 \\ \mathbf{E} &= -B_0 v_y \hat{x} \\ \mathbf{B}_1 &= \frac{\mathbf{k}}{\omega} \times \mathbf{E} = -\frac{k_z B_0 v_y}{\omega} \hat{y} = -\frac{\mathbf{v}}{\omega/k_{\parallel}} B_0 \\ \mathbf{j} &= \frac{1}{\mu_0} \nabla \times \mathbf{B}_1 = \frac{i\mathbf{k} \times \mathbf{B}_1}{\mu_0} \end{aligned} \tag{10.33}$$

SAW has a wave vector \mathbf{k} in the XZ-plane. \mathbf{E} shall oscillate in the X-direction; \mathbf{B} shall oscillate in the Y-direction. The electric current of the wave \mathbf{j} lies in the XZ-plane. The timescale of the variations of the wave fields is much longer than the ion gyroperiod Ω_i^{-1} . In both the perpendicular and parallel directions, the spatial scale of the waves $1/k$ are much larger than ion motion scale r_{iL} . The wave carries a Poynting flux $\mathbf{S} = \mathbf{E} \times \mathbf{B}_1$ strictly parallel to \mathbf{B}_0 . The ratio of the wave electric field to the wave magnetic field $|\mathbf{E}|/|\mathbf{B}_1|$ is exactly one Alfvén speed V_A .

\mathbf{E} (or \mathbf{B}_1) in Equation 10.33 shows that the Alfvén wave in a uniform plasma is a linearly polarized wave if $\mathbf{v}_1 = (0, v_y, 0)$. If instead we set $\mathbf{k} = (0, 0, k)$ ($\theta = 0^\circ$) and $\mathbf{v}_1 = (v_x, v_y, 0)$, then $\mathbf{k} \cdot \mathbf{v} = 0 \Rightarrow \rho_1 = 0$ but $\mathbf{E} = -\mathbf{v} \times \mathbf{B}_0 = kv_y \hat{x} - kv_x \hat{y}$ can be a circularly polarized wave or else. Correlated \mathbf{B}_1 and \mathbf{v} corresponds to waves propagating anti-parallel to the \mathbf{B}_0 ($\mathbf{k} \cdot \mathbf{B}_0 < 0$), and anti-correlated \mathbf{B}_1 and \mathbf{v} corresponds to waves propagating parallel to the \mathbf{B}_0 ($\mathbf{k} \cdot \mathbf{B}_0 > 0$). (This is the same as expressed by the Alfvénicity condition Equation 18.4.) The resultant magnetic field then exhibits shear, thus the Alfvén wave is called the *shear Alfvén wave* (SAW). An animation of SAW is shown in [?@fig-alfven-wave](#).

To understand what happens physically in an Alfvén wave, recall that this is an electromagnetic wave with a fluctuating magnetic field \mathbf{B}_1 given by

$$\nabla \times \mathbf{E}_1 = -\dot{\mathbf{B}}_1 \quad E_x = (\omega/k)B_y \quad (10.34)$$

The small component B_y , when added to \mathbf{B}_0 , gives the magnetic field lines a sinusoidal ripple, shown exaggerated in Figure 10.1. At the point shown, B_y is in the positive y direction, so, according to Equation 10.34, E_x is in the positive x direction if ω/k is in the z direction. The electric field E_x gives the plasma an $\mathbf{E}_1 \times \mathbf{B}_0$ drift in the negative y direction. Since we have taken the limit $\omega^2 \ll \Omega_c^2$, both ions and electrons will have the same drift v_y , obtained from Section 10.6 the component v_y under $T_i = 0$:

$$\begin{aligned} v_{ix} &= \frac{iq}{m\omega} \left(1 - \frac{\Omega_c^2}{\omega^2} \right)^{-1} E_1 \\ v_{iy} &= \frac{q}{m\omega} \frac{\Omega_c}{\omega} \left(1 - \frac{\Omega_c^2}{\omega^2} \right)^{-1} E_1 \end{aligned} \quad (10.35)$$

Thus, the fluid moves up and down in the y direction. The magnitude of this velocity is $|E_x/B_0|$. Since the ripple in the field is moving by at the phase velocity ω/k , the magnetic field is also moving downward at the point indicated in Figure 10.1. The downward velocity of the magnetic field lines is $(\omega/k)|B_y/B_0|$, which, according Equation 10.34, is just equal to the fluid velocity $|E_x/B_0|$. Thus, the fluid and the field lines oscillate together as if the particles were stuck to the lines. The magnetic field lines act as if they were mass-loaded strings under tension, and an Alfvén wave can be regarded as the propagating disturbance occurring when the strings are plucked. This concept of plasma frozen to the field lines and moving with them is a useful one for understanding many low-frequency plasma phenomena. It can be shown that this notion is an accurate one as long as there is no electric field along \mathbf{B} .

It remains for us to see what sustains the electric field E_x which we presupposed was there. As \mathbf{E}_1 fluctuates, the ions' inertia causes them to lag behind the electrons, and there is a polarization drift \mathbf{v}_p in the direction of \mathbf{E}_1 . This drift v_{ix} is given by Equation 10.35 and causes a current \mathbf{j}_1 to flow in the x direction. The resulting $\mathbf{j}_1 \times \mathbf{B}_0$ force on the fluid is in the y direction and is 90° out of phase with the velocity \mathbf{v}_1 . This force perpetuates the oscillation in the same way as in any oscillator where the force is out of phase with the velocity. It is, of course, always the ion inertia that causes an overshoot and a sustained oscillation, but in a plasma the momentum is transferred in a complicated way via the electromagnetic forces.

In a more realistic geometry for experiments, \mathbf{E}_1 would be in the radial direction and \mathbf{v}_1 in the azimuthal direction. The motion of the plasma is then incompressible. This is the reason the ∇p term in the equation of motion could be neglected.

In a non-uniform plasma, SAW attains the interesting property of a continuous spectrum. To illustrate this feature, let us consider the simplified slab model of a cold plasma with a non-uniform density, $\rho = \rho(x)$, and a uniform $\mathbf{B}_0 = B_0 \hat{z}$. Assuming at $t = 0$ a localized initial

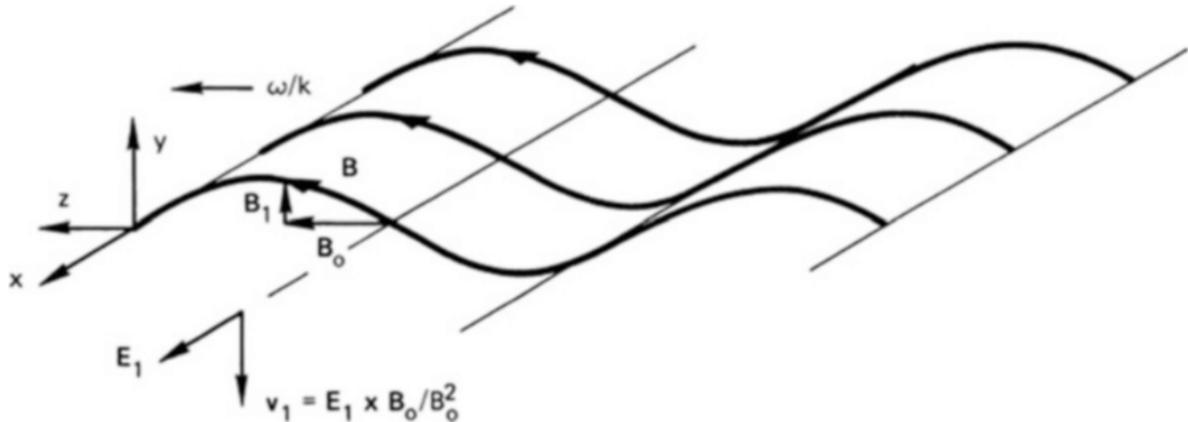


Figure 10.1: Relation among the oscillating quantities in an Alfvén wave and the (exaggerated) distortion of the lines of force.

perturbation $\mathbf{B}_{1y}(x, t = 0) = \exp(-x^2/\Delta x^2)$, $|k_y \Delta_x| \ll 1$, and $\partial \mathbf{B}_{1y}/\partial t = 0$, the perturbation then evolves according to the following wave equation (Equation 10.31, $B_{1z} = 0$ so no coupling between the fast mode and Alfvén mode):

$$[\partial_t^2 + \omega_A^2(x)]B_{1y}(x, t) = 0$$

Here $\omega_A^2(x) = k_z^2 v_A^2(x)$ and the solution is

$$B_{1y}(x, t) = \hat{B}_{1y}(x, 0) \cos[\omega_A(x)t] \quad (10.36)$$

Equation 10.36 shows that every point in x oscillates at a different frequency, $\omega_A(x)$. With a continuously varying $\omega_A(x)$; the wave frequency, thus, constitutes a continuous spectrum. While the above result is based on a model with a 1D non-uniformity in x , this general feature of SAW continuous spectrum also holds in magnetized plasmas with 2D or 3D non-uniformities. A good example is geomagnetic pulsations in the Earth's magnetosphere observed by Engebretson shown in Figure 1 of (L. Chen, Zonca, and Lin 2021).

Equation 10.36 also indicates an unique and important property of SAW continuous spectrum: the spatial structure evolves with time. Specifically, the wave number in the non-uniformity direction is, time asymptotically, given by:

$$|k_x| = \left| \frac{\partial \ln B_{1y}}{\partial x} \right| \simeq \left| \frac{d\omega_A(x)}{dx} \right| t \quad (10.37)$$

That $|k_x|$ increases with t is significant, since it implies that any initially long-scale perturbations will evolve into short scales. This point is illustrated in Figure ??? (CAN I PERFORM THE

SIMULATION?); showing the evolution of a smooth B_{1y} at $t = 0$ to a spatially fast varying B_{1y} at a later t .

Another consequence of $|k_x|$ increasing with t is the temporal decay of B_{1x} . From $\nabla \cdot \mathbf{B}_1 \simeq \nabla_{\perp} \cdot \mathbf{B}_{1\perp} = 0$, we can readily derive that, for $|\omega'_A t| \gg |k_y|$:

$$B_{1x}(x, t) \simeq \frac{k_y}{\omega'_A(x)t} \hat{B}_{1y}(x, 0) e^{-i\omega_A(x)t} \left[1 + \mathcal{O}\left(\frac{k_y}{|\omega'_A t|} + \dots\right) \right]$$

That is, B_{1x} decays temporally due to the phase mixing of increasingly more rapidly varying neighboring perturbations.

Noting that, as $t \rightarrow \infty$, $|k_x| \rightarrow \infty$, it thus suggests that the perturbation will develop singular structures toward the steady state. As we will see in the field line resonance Chapter 16, the singularity is reached at the Alfvén resonant point x_r , where $\omega^2 = \omega_A^2(x_r)$ along with a finite resonant wave-energy absorption rate. Note that at the isolated extrema of the SAW continuum, $|\omega'_A| = 0$, phase mixing vanishes; consequently, perturbation remains regular and experiences no damping via resonant absorption. This feature has important implications to Alfvén instabilities in laboratory plasmas.

Space plasmas support a variety of waves, but for heating the plasma and accelerating the electrons and ions, Alfvén waves are a predominant source. Near the Sun, Alfvén waves are excited and propagate outward. They exchange their energy with particles to accelerate them in the form of solar wind and heat the electrons. When perpendicular wavelength of the wave becomes comparable to the ion gyro-radius ($k_{\perp} r_{Li} \sim 1$) or inertial length ($k_{\perp} d_i \sim 1$) in the case of kinetic Alfvén waves (KAWs, Section 10.8.4) or inertial Alfvén waves (IAWs, Equation 10.71), respectively, then the wave in both the limits has a *nonzero parallel electric field* component which is responsible for the acceleration of the particles via the Landau mechanism. This is also consistent with the generalized Ohm's law Equation 8.27: only when we go beyond Hall MHD can \mathbf{E}_{\parallel} be nonzero. (WHAT ABOUT $\eta \mathbf{j}$?) KAW relates to $\frac{\nabla P_e}{ne}$ term, while IAW relates to $\frac{\partial \mathbf{j}}{\partial t}$ term. The key interest is in E_{\parallel} . KAWs and IAWs have significance not only in space plasmas but also in laboratory plasma such as in fusion reactors.

Alfvén wave has very high saturation level, meaning that it takes a long time for the wave to reach the nonlinear phase. (???)

SAW in a Slab

We now look deeper into the properties of Alfvén waves in a nonuniform magnetized plasma slab that carries a current flowing along an externally imposed magnetic field $B_{0z}\hat{z}$, where B_{0z} is assumed to be a constant. First, we formulate the governing equation for the slab geometry, under the ideal MHD condition. Then we show that Alfvén waves are always neutrally stable, with important indication at the end.

The presence of an equilibrium current density $\mathbf{J}_0 = \hat{z}J_0(x)$ produces a local magnetic field of the form

$$\mathbf{B}_0 = \hat{z}B_{0z} + \hat{y}B_{0y}(x)$$

The Ampère's law gives

$$\nabla \times \mathbf{B}_0 = \mu_0 \mathbf{J}_0 \Rightarrow \frac{\partial B_{0y}}{\partial x} - \frac{\partial B_{0x}}{\partial y} = \frac{\partial B_{0y}}{\partial x} = \mu_0 J_0(x) \quad (10.38)$$

From the force balance equation,

$$\mathbf{J}_0 \times \mathbf{B}_0 = \nabla P_0 \quad (10.39)$$

Substituting Equation 10.38 into Equation 10.39, we get

$$\frac{B_{0y}(x)^2}{2\mu_0} + P_0(x) = \text{const.}$$

Designate all perturbation quantities with a subscript 1, and assume $e^{-i\omega t + ik_y y + ik_z z}$ dependence for all perturbations (nonuniform in the x -direction, thus no sinusoidal wave assumption). From linearized Faraday's law and Ohm's law in ideal MHD,

$$\begin{aligned} -\nabla \times \mathbf{E}_1 &= \nabla \times (\mathbf{v}_1 \times \mathbf{B}_0) = \frac{\partial \mathbf{B}_1}{\partial t} \\ -i\omega \mathbf{B}_1 &= \mathbf{v}_1 (\nabla \cdot \mathbf{B}_0) - \mathbf{B}_0 (\nabla \cdot \mathbf{v}_1) + (\mathbf{B}_0 \cdot \nabla) \mathbf{v}_1 - (\mathbf{v}_1 \cdot \nabla) \mathbf{B}_0 \end{aligned}$$

where we have assumed the plasma is incompressible. Replace ∇ with $i\mathbf{k}$, $\mathbf{v}_1 = i\omega \boldsymbol{\xi}_1$ and take the x -component, we get

$$B_{1x} = i(\mathbf{k} \cdot \mathbf{B}_0) \xi_{1x} \quad (10.40)$$

where $\mathbf{k} = \hat{y}k_y + \hat{z}k_z$.

The MHD force law can be linearized to

$$\rho_0 \frac{\partial \mathbf{v}_1}{\partial t} = -\nabla \left(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0} \right) + \frac{1}{\mu_0} [(\mathbf{B}_0 \cdot \nabla) \mathbf{B}_1 + (\mathbf{B}_1 \cdot \nabla) \mathbf{B}_0] \quad (10.41)$$

Since the plasma is incompressible, $\nabla \cdot \mathbf{v}_1 = 0$, $\dot{\boldsymbol{\xi}} = \mathbf{v}_1 \Rightarrow \nabla \cdot \boldsymbol{\xi} = 0$. In addition, $\nabla \cdot \mathbf{B}_1 = 0$. We then have

$$\begin{aligned} \mathbf{k} \cdot \boldsymbol{\xi}_{1yz} &= i \frac{\partial \xi_{1x}}{\partial x} \\ \mathbf{k} \cdot \mathbf{B}_{1yz} &= i \frac{\partial B_{1x}}{\partial x} \end{aligned}$$

where $\boldsymbol{\xi}_{1yz} = (0, \xi_{1y}, \xi_{1z})$, $\mathbf{B}_{1yz} = (0, B_{1y}, B_{1z})$. The x -component of Equation 10.41 gives

$$-\rho_0 \omega^2 \xi_{1x} = -\frac{\partial}{\partial x} \left(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0} \right) + \frac{1}{\mu_0} [(\mathbf{B}_0 \cdot \nabla) B_{1x}] \quad (10.42)$$

The dot product of Equation 10.41 with \mathbf{k} gives

$$\begin{aligned}-\rho_0 \omega^2 \mathbf{k} \cdot \boldsymbol{\xi}_{1yz} &= ik^2 \left(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0} \right) + \frac{1}{\mu_0} [i(\mathbf{k} \cdot \mathbf{B}_0)(\mathbf{k} \cdot \mathbf{B}_{1yz}) + B_{1x} \frac{\partial}{\partial x}(\mathbf{k} \cdot \mathbf{B}_0)] \\ -\rho_0 \omega^2 i \frac{\partial \xi_{1x}}{\partial x} &= ik^2 \left(p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0} \right) + \frac{1}{\mu_0} [i(\mathbf{k} \cdot \mathbf{B}_0)(i \frac{\partial B_{1x}}{\partial x}) + B_{1x} \frac{\partial}{\partial x}(\mathbf{k} \cdot \mathbf{B}_0)]\end{aligned}\quad (10.43)$$

Finally, canceling $p_1 + \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0}$ from Equation 10.42 and Equation 10.43 and substituting B_{1x} from Equation 10.40, we obtain the governing equation

$$\frac{\partial}{\partial x} \left\{ \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \frac{\partial \xi_{1x}}{\partial x} \right\} - k^2 \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \xi_{1x} = 0 \quad (10.44)$$

where $k^2 = k_y^2 + k_z^2$, $\mathbf{v}_A = v_A \mathbf{B}_0 / B_0$, and $v_A = B_0 / \sqrt{\mu_0 \rho_0}$ is the local Alfvén speed. This is the governing equation of shear Alfvén waves in a slab geometry derived by Hasegawa and Liu Chen in the 1970s, which is readily compared with Eq.(10.33) in (Bellan 2008).

It is easy to show that this governing equation always yields *neutrally stable* solutions of SAWs, i.e. $\omega_i = \Im(\omega) = 0$. Multiply it by ξ_{1x}^* , and integrate the resultant equation to get

$$\int_{-\infty}^{\infty} dx \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \left[\left| \frac{d \xi_{1x}}{dx} \right|^2 + k^2 |\xi_{1x}|^2 \right] = 0$$

where we have assumed that ξ_{1x} vanishes on the boundary. This gives

$$\omega^2 = \frac{\int_{-\infty}^{\infty} dx \rho_0 (\mathbf{k} \cdot \mathbf{v}_A)^2 \left[\left| \frac{d \xi_{1x}}{dx} \right|^2 + k^2 |\xi_{1x}|^2 \right]}{\int_{-\infty}^{\infty} \rho_0 \left[\left| \frac{d \xi_{1x}}{dx} \right|^2 + k^2 |\xi_{1x}|^2 \right] dx} \geq 0$$

SAWs are the dominant low frequency waves in a current carrying plasma. The neutrally stable modes studies above can be destabilized by unfavorable curvature, and such modes are called *ballooning modes* (Section 13.7.4). They may also be destabilized by a finite electrical resistivity, and these are *tearing modes* (Section 13.7.5). Their interaction with fusion-generated alpha particles are a major issue in all magnetic fusion schemes. Finally, since the governing equation exhibits a singularity when $\omega = \mathbf{k} \cdot \mathbf{v}_A$, this singularity represents resonance absorption, which forms the basis of *Alfvén wave heating* (i.e. field line resonance, Chapter 16). This singularity also give rise to the so called “Alfvén continuum spectrum” mentioned above.

Note that the governing equation is valid even if B_{0z} is an arbitrary function of x . If in addition, an external gravity $\mathbf{g} = \hat{x}g$ in the x-direction is present, the governing equation is modified simply by inserting the term $-(g/\rho_0)d\rho_0/dx$ in the second square bracket, and the equation is identical to Eq.(10.33) of Bellan. This is the most general equation which describes the *magneto-Rayleigh-Taylor instability* (MRT) in Cartesian geometry using the incompressible, ideal MHD model.

10.7.4 Fast and Slow Wave

The x - z components of Equation 10.32 give

$$\begin{aligned} (\omega^2 - k^2 v_A^2 - k_x^2 v_s^2) v_x - k_x k_z v_x^2 v_z &= 0 \\ (\omega^2 - k_z^2 v_s^2) v_z - k_x k_z v_s^2 v_x &= 0 \end{aligned}$$

The dispersion relation is given by the determinant being 0,

$$\begin{aligned} \omega^4 - k^2(v_A^2 + v_s^2)\omega^2 + k_z^2 v_s^2 k^2 v_A^2 &= 0 \\ \frac{\omega^2}{k^2} = \frac{1}{2}(v_A^2 + v_s^2) \pm \frac{1}{2}\sqrt{(v_A^2 + v_s^2)^2 - 4v_s^2 v_A^2 \cos^2 \theta} \end{aligned} \quad (10.45)$$

“+” corresponds to the fast mode, or magnetosonic mode, and “-” corresponds to the slow mode. The Friedrich graph ?@fig-mhd-phase-speed-low-beta is very useful in interpreting Equation 10.45. Here we only show the case for $v_A > v_s$; if $v_A < v_S$, then along the background magnetic field direction \hat{z} the fast wave will have a phase speed $v_{ph} = v_s$, and the slow wave will a phase speed $v_{ph} = v_A$ that overlaps with the Alfvén wave. Thus in the high- β case slow wave may have a faster phase speed than the Alfvén wave at a proper angle!

Another thing to be careful about is that when you observe a wave propagating at the Alfvén speed along the field line, you need more information to determine the characteristics of the wave:

- Check the β -regime as well as the compressibility to determine if it has fast/slow wave component.
- Check the polarization to see if it is an cyclotron wave.
- Check the parallel electric field to see if has kinetic features.

Given the velocity perturbation $\mathbf{v}_1 = (v_x, 0, v_z)$, the other perturbations are given as

$$\begin{aligned} \rho_1 &= \frac{\rho_0}{\omega} \mathbf{k} \cdot \mathbf{v} \\ p_1 &= \gamma p_0 \frac{\rho_1}{\rho_0} = \frac{\gamma p_0}{\omega} \mathbf{k} \cdot \mathbf{v} \\ \mathbf{E} &= B_0 v_x \hat{y} \\ \mathbf{B}_1 &= \frac{\mathbf{k}}{\omega} \times \mathbf{E} = \frac{(\mathbf{k} \cdot \mathbf{v}) \mathbf{B}_0 - (\mathbf{k} \cdot \mathbf{B}_0) \mathbf{v}}{\omega} = -\frac{B_0 v_x k_z}{\omega} \hat{x} + \frac{B_0 v_x k_x}{\omega} \hat{z} \\ \mathbf{j} &= \frac{1}{\mu_0} \nabla \times \mathbf{B}_1 = \frac{i \mathbf{k} \times \mathbf{B}_1}{\mu_0} \end{aligned} \quad (10.46)$$

\mathbf{E} shows that the fast/slow modes in a uniform plasma are also linearly polarized. When we have $\theta = 90^\circ$ (?@fig-fast-wave), the magnetic perturbation can be simplified to

$$\mathbf{B}_1 = \frac{v}{\omega/k_\perp} \mathbf{B}_0$$

which means that the perturbed magnetic field is always aligned with the background magnetic field.

The distinction between the fast and slow waves can be further understood by comparing the signs of the wave induced fluctuations in the plasma and magnetic pressures: p and $\mathbf{B}_0 \cdot \mathbf{B}/\mu_0$, respectively. It follows from Equation 10.46 that

$$\frac{\mathbf{B}_0 \cdot \mathbf{B}}{\mu_0} = \frac{\mathbf{k} \cdot \mathbf{v} B_0^2 - (\mathbf{k} \cdot \mathbf{B}_0)(\mathbf{B}_0 \cdot \mathbf{v})}{\mu_0 \omega}$$

The z -component of the perturbed momentum equation yields

$$\omega \rho_0 v_z = k \cos \theta p$$

Combining the above, we have

$$\frac{\mathbf{B}_0 \cdot \mathbf{B}}{\mu_0} = \frac{v_A^2}{v_s^2} \left(1 - \frac{k^2 v_s^2 \cos^2 \theta}{\omega^2} \right) p \quad (10.47)$$

Hence, p and $\mathbf{B}_0 \cdot \mathbf{B}/\mu_0$ have the same sign if $V_{\text{ph}} = \omega/k > v_s \cos \theta$, and the opposite sign if $V_{\text{ph}} < v_s \cos \theta$. It is straightforward to show that $v_+ > v_s \cos \theta$, and $v_- < v_s \cos \theta$. Thus, we conclude that in the fast magnetosonic wave the pressure and magnetic energy fluctuations reinforce one another, whereas the fluctuations oppose one another in the slow magnetosonic wave.

The temperature perturbation, based on the equation of state, can be derived as follows:

$$\begin{aligned} T_1 &= T - T_0 = \frac{p_0 + p_1}{(n_0 + n_1)k_B} - \frac{p_0}{n_0 k_B} \\ &= p_0 \left(1 + \gamma \frac{n_1}{n_0} \right) \frac{1}{(n_0 + n_1)k_B} - \frac{p_0}{n_0 k_B} \\ &= p_0 \frac{1}{k_B} \left[\left(1 + \gamma \frac{n_1}{n_0} \right) \frac{1}{n_0 + n_1} - \frac{1}{n_0} \right] \\ &= p_0 \frac{1}{k_B} \left[\frac{(\gamma - 1)n_1}{n_0(n_0 + n_1)} \right] \\ &= T_0 (\gamma - 1) \frac{\mathbf{k} \cdot \mathbf{v}}{\omega + \mathbf{k} \cdot \mathbf{v}} \end{aligned} \quad (10.48)$$

Equation 10.48 shows that temperature does not follow the same sinusoidal shape as the perturbed velocity.

Special attention is required for the extreme cases. When $\theta = 90^\circ$, the fast wave phase speed $v_{\text{ph}} = \sqrt{v_A^2 + v_s^2}$ while the slow wave phase speed $v_{\text{ph}} = 0$. Slow waves cannot propagate perpendicular to the magnetic field. However, this is still a valid solution with all the slow

mode perturbation properties, and you will see in Section 13.11.2 and Section 14.2 that this is named *mirror mode*. When $\theta = 0^\circ$, the fast wave phase speed $v_{\text{ph}} = \max(v_A, v_s)$ while the slow wave phase speed $v_{\text{ph}} = \min(v_A, v_s)$.

It is important to understand the nature of fast and slow waves under the *cold plasma limit*, i.e. in the low- β regime where $v_A \gg v_s$ (cold, strong B). If $v_s \rightarrow 0$, the slow mode ceases to exit ($V_{\text{ph}} \rightarrow 0$) and the phase speed of the fast mode becomes $v_p \approx v_A$. This is often called the *compressional Alfvén wave*, which is actually a fast wave in the low- β limit. If β is low but we still have a finite sound speed, the dispersion relation for the slow wave reduces to

$$\omega \simeq k v_s \cos \theta$$

Thus, in low- β plasmas the slow wave is a sound wave modified by the presence of the magnetic field.

Sometimes the fast wave dispersion relation is written in alternative ways. Let $a = \omega/(k_{\parallel} v_A) = \omega/(k \cos \theta v_A)$. From Equation 10.45, we have

$$\cos^2 \theta = \frac{a^2 (v_s^2/v_A^2 + 1) - v_s^2/v_A^2}{a^4} \quad (10.49)$$

This expression is useful when considering the wave resonance and mode conversions. For example, Alfvén resonance occurs when

$$\omega^2 = k_{\parallel}^2 v_A^2 \quad (10.50)$$

i.e. $a = 1$ under the low frequency approximation. In higher frequency cases, the Alfvén resonance condition is modified as

$$a = \frac{1}{(1 - \omega^2/\Omega_i^2)^2} \quad (10.51)$$

because of the finite ion Larmor radius effects. However, note that the Alfvén resonance singularity can be removed by including non-MHD effects such as electron inertia or ion Larmor radius corrections!

10.7.5 MHD Waves in CPDR

CPDR is a general treatment of cold plasma waves, while MHD waves are the low-frequency approximation to CPDR's solution. For parallel-propagating cold plasma waves,

- The MHD low-frequency fast mode and Alfvén mode degenerate. They have the same dispersion relation (i.e. same phase/group velocity) so it is impossible to distinguish the two modes.

- At higher frequency in the ion scale and below ω_{ce} , the fast mode becomes the R mode, and the Alfvén mode becomes the L mode, with opposite sense of polarization. Therefore, we name these the Alfvén Ion-Cyclotron mode and the magnetosonic-whistler mode. Some people also call the former incompressible Alfvén wave, and the latter compressible Alfvén wave.
- A typical application is the study of beam-beam instability in foreshocks. Typically, we have $\beta \ll 1$: this is a requirement from the two-fluid theory, and also remember the cold plasma assumption is the extreme case $P = 0$. The beam is a source of free energy that can trigger instability. In the fluid theory, this is known as the two-stream instability; in the kinetic treatment, the name “beam-beam” suggests that kinetic effects kick in.
- Here comes another two fancy words: normal cyclotron resonance and anomalous cyclotron resonance. Cyclotron resonance occurs under two conditions: 1) ions move at the same speed as the wave phase speed and 2) the transverse EM wave is left-handed polarized, same as the gyration direction of ions. Low-frequency magnetosonic waves are right-handed, but they can still resonate with ions. This requires that ions move in the same direction as the fast waves but faster, so that from ions’ perspective, the fast waves are doppler-shifted to be left-handed with the same frequency as the ion gyration. Thus this is termed “anomalous”, accounting for the fact of different handedness. Ion cyclotron waves are left-handed, same as the ion gyration. These waves move in the opposite direction as the ion beam, so they never change handedness due to the doppler effect, but the frequency is up-shifted such that it matches the ion cyclotron frequency. This is the reason why ion cyclotron waves possess rest frame frequency slightly smaller than the cyclotron frequency. From Peter Gary’s definition, the instability corresponds to the magnetosonic mode is called ion-ion right-hand resonance instability, and the instability corresponds to the Alfvén ion-cyclotron wave is called ion-ion left-hand resonance instability.
- Whistler waves in literature span a wide of frequency ranges from the MHD scale to the electron cyclotron frequency. Therefore, electron as well as ion instabilities can both generate whistlers. For strictly parallel waves, whistlers and fast waves do not have density perturbations. However, for oblique waves, they can have density perturbations.

10.7.6 Limitations of the MHD model

The MHD model ignores parallel electron dynamics and so has a shear mode dispersion $\omega^2 = k_z^2 v_A^2$ that has no dependence on k_\perp . Some researchers interpret this as a license to allow arbitrarily large k_\perp in which case a shear mode could be localized to a single field line. However, the two-fluid model of the shear mode does have a dependence on k_\perp which becomes important when either $k_\perp c/\omega_{pe}$ or $k_\perp r_{Ls}$ become of order unity (whether to use $c\omega_{pe}$ or r_{Ls} depends on whether $\beta m_i/m_e$ is small or large compared to unity). Since $c\omega_{pe}$ and r_{Ls} are typically small lengths, the MHD point of view is acceptable provided the characteristic length of perpendicular localization is much larger than $c\omega_{pe}$ or r_{Ls} .

MHD also predicts a sound wave which is identical to the ordinary hydrodynamic sound wave

of an unmagnetized gas. The perpendicular behavior of this sound wave is consistent with the two-fluid model because both two-fluid and MHD perpendicular motions involve compressional behavior associated with having finite B_{z1} . However, the parallel behavior of the MHD sound wave is problematical because E_{z1} is assumed to be identically zero in MHD. According to the two-fluid model, any parallel acceleration requires a parallel electric field. The two-fluid B_{z1} mode is decoupled from the two-fluid E_{z1} mode so that the two-fluid B_{z1} mode is both compressional and has no parallel motion associated with it.

The MHD analysis makes no restriction on the electron to ion temperature ratio and predicts that a sound wave would exist for $T_e = T_i$. In contrast, the two-fluid model shows that sound waves can only exist when $T_e \gg T_i$ because only in this regime is it possible to have $k_B T_i / m_i \ll \omega^2 / k_z^2 \ll k_B T_e / m_e$ and so have inertial behavior for ions and kinetic behavior for electrons.

Various paradoxes develop in the MHD treatment of the shear mode but not in the two-fluid description. These paradoxes illustrate the limitations of the MHD description of a plasma and shows that MHD results must be treated with caution for the shear (slow) mode. MHD provides an adequate description of the fast (compressional) mode.

10.8 Two-fluid model of Alfvén modes

We now examine the MHD modes from a two-fluid point of view. The two-fluid point of view shows that the shear mode occurs as one of two distinct modes, only one of which can exist for given plasma parameters. Which of these shear modes occurs depends upon the ratio of hydrodynamic pressure to magnetic pressure. This ratio is defined for each species s as

$$\beta_s = \frac{n k_B T_s}{B^2 / \mu_0}$$

The subscript s is not used if electrons and ions have the same temperature. β_i measures the ratio of ion thermal velocity to the Alfvén velocity since

$$\frac{v_{Ti}^2}{v_A^2} = \frac{k_B T_i / m_i}{B^2 / n m_i \mu_0} = \beta_i$$

Thus, $v_{Ti} \ll v_A$ corresponds to $\beta_i \ll 1$. Magnetic forces dominate hydrodynamic forces in a low β plasma, whereas in a high β plasma the opposite is true.

The ratio of electron thermal velocity to Alfvén velocity is also of interest and is

$$\frac{v_{Te}^2}{v_A^2} = \frac{k_B T_e / m_e}{B^2 / n m_i \mu_0} = \frac{m_i}{m_e} \beta_e$$

Thus, $v_{Te}^2 \gg v_A^2$ when $\beta_e \gg m_e/m_i$ and $v_{Te}^2 \ll v_A^2$ when $\beta_e \ll m_e/m_i$. Shear Alfvén wave physics is different in the $\beta_e \gg m_e/m_i$ and $\beta_e \ll m_e/m_i$ regimes which therefore must be investigated separately. MHD ignores this β_e dependence, an oversimplification which leads to the paradoxes.

Both Faraday's law and the pre-Maxwell Ampère's law (no displacement current) are fundamental to Alfvén wave dynamics. The system of linearized equations thus is

$$\begin{aligned}\nabla \times \mathbf{E}_1 &= -\frac{\partial \mathbf{B}_1}{\partial t} \\ \nabla \times \mathbf{B}_1 &= \mu_0 \mathbf{j}_1\end{aligned}\tag{10.52}$$

If the dependence of \mathbf{j}_1 on \mathbf{E}_1 can be determined, then the combination of Ampère's law and Faraday's law provides a complete self-consistent description of the coupled fields $\mathbf{E}_1, \mathbf{B}_1$ and hence describes the normal modes. From a mathematical point of view, specifying $\mathbf{j}_1(\mathbf{E}_1)$ means that there are as many equations as dependent variables in the pair of Equation 10.52. The relationship between \mathbf{j}_1 and \mathbf{E}_1 is determined by the Lorentz equation or some generalization thereof (e.g., drift equations, Vlasov equation, fluid equation of motion). The MHD derivation used the polarization drift to give a relationship between $j_{1\perp}$ and $E_{1\perp}$ but leaves ambiguous the relationship between $j_{1\parallel}$ and $E_{1\parallel}$.

The two-fluid equations provide a definite description of the relationship between $j_{1\parallel}$ and $E_{1\parallel}$. At frequencies well below the cyclotron frequency, decoupling of modes also occurs in the two-fluid description, and this decoupling is more clearly defined and more symmetric than in MHD. The decoupling in a uniform plasma results because the dependence of \mathbf{j}_1 on \mathbf{E}_1 has the property that $j_{1z} \sim E_{1z}$ and $\mathbf{j}_{1\perp} \sim \mathbf{E}_{1\perp}$. Thus, for $\omega \ll \omega_{ci}$ there is a simple linear relation between parallel electric field and parallel current and another distinct simple linear relation between perpendicular electric field and perpendicular current; these two linear relations mean that the tensor relating \mathbf{j}_1 to \mathbf{E}_1 is diagonal (at higher frequencies this is not the case). The decoupling can be seen by supposing that all first order quantities have the dependence $\exp(ik_\perp \cdot \mathbf{x} + ik_z z)$ where $\mathbf{k}_\perp = k_x \hat{x} + k_y \hat{y}$. Mode decoupling can be seen by examining Table 10.1 which lists the electric and magnetic field components:

Table 10.1: Parallel and perpendicular mode decoupling

\mathbf{E} components	\mathbf{B} components
$* \hat{k}_\perp \cdot \mathbf{E}_1$	$\hat{k}_\perp \cdot \mathbf{B}_1$
$\hat{z} \times \hat{k}_\perp \cdot \mathbf{E}_1$	$* \hat{z} \times \hat{k}_\perp \cdot \mathbf{B}_1$
$* \hat{z} \cdot \mathbf{E}_1$	$\hat{z} \cdot \mathbf{B}_1$

Because of the property that $j_{1z} \sim E_{1z}$ and $\mathbf{j}_{1\perp} \sim \mathbf{E}_{1\perp}$ the terms starting with an asterisk are decoupled from the rest. Hence, one mode consists of solely interrelationships between the starred terms (this mode is called E_z mode since it has finite E_z) and the other distinct mode

consists solely of interrelationships between the non-starred terms (this mode is called the B_z mode since it has finite B_z). Since the modes are decoupled, it is possible to “turn off” the E_z mode when considering the B_z mode and vice versa. If the plasma is non-uniform, the E_z and B_z modes can become coupled.

The ideal MHD formalism sidesteps discussion of the E_z mode. Instead, two disconnected assumptions are invoked in ideal MHD, namely

1. it is assumed that $E_{z1} = 0$ and
2. the parallel current j_{z1} is assumed to arrange itself spontaneously in such a way as to always satisfy $\nabla \cdot \mathbf{j}_1 = 0$.

This pair of assumptions completes the system of equations, but omits the parallel dynamics associated with the E_z mode and instead replaces this dynamics with an assumption that j_{z1} is determined by some unspecified automatic feedback mechanism. In contrast, the two-fluid equations describe how particle dynamics determines the relationship between j_{z1} and E_{z1} . Thus, while MHD is both simpler and self-consistent, it omits some vital physics.

The two-fluid model is based on the linearized equations of motion

$$m_s n \frac{\partial \mathbf{u}_{s1}}{\partial t} = n q_s (\mathbf{E}_1 + \mathbf{u}_{\sigma 1} \times \mathbf{B}) - \nabla \cdot \vec{P}_{s1} \quad (10.53)$$

Charge neutrality is assumed so that $n_i = n_e = n$. Also, the pressure terms is

$$\nabla \cdot \vec{P}_{s1} = \nabla \cdot \begin{bmatrix} p_{s\perp 1} & 0 & 0 \\ 0 & p_{s\perp 1} & 0 \\ 0 & 0 & p_{sz1} \end{bmatrix} = \nabla_{\perp} p_{s\perp 1} + \hat{z} \frac{\partial p_{sz1}}{\partial z} \quad (10.54)$$

Assuming $\omega \ll \omega_{ci}$ implies $\omega \ll \omega_{ce}$ also and so perpendicular motion can be described by drift theory for both ions and electrons. However, here the drift approximation is used for the fluid equations, rather than for a single particle. Following the drift method of analysis, the left hand side of Equation 10.53 is neglected to first approximation, resulting in

$$\mathbf{u}_{s1} \times \mathbf{B} \simeq -\mathbf{E}_{1\perp} + \nabla_{\perp} p_{s\perp 1} / n q_s \quad (10.55)$$

which may be solved for \mathbf{u}_{s1} to give

$$\mathbf{u}_{s\perp 1} = \frac{\mathbf{E}_1 \times \mathbf{B}}{B^2} - \frac{\nabla p_{s\perp 1} \times \mathbf{B}}{n q_s B^2} \quad (10.56)$$

The first term is the single-particle $\mathbf{E} \times \mathbf{B}$ drift and the second term is called the diamagnetic drift, a fluid effect that does not exist for single-particle motion (Section 8.7). Because $\mathbf{u}_{s\perp 1}$ is time-dependent there is also a polarization drift. Recalling that the form of the single-particle polarization drift for electric field only is $\mathbf{v}_p = m \dot{\mathbf{E}}_{1\perp} / q B^2$ and using $\mathbf{E}_{1\perp} = \nabla_{\perp} p_{s\perp 1} / n q_s$ for

the fluid model instead of just $\mathbf{E}_{1\perp}$ for single particles (Equation 10.55) the fluid polarization drift is obtained. With the inclusion of this higher order correction, the perpendicular fluid motion becomes

$$\mathbf{u}_{s\perp 1} = \frac{\mathbf{E}_1 \times \mathbf{B}}{B^2} - \frac{\nabla p_{s\perp 1} \times \mathbf{B}}{nq_s B^2} + \frac{m_s}{q_s B^2} \dot{\mathbf{E}}_{1\perp} - \frac{m_s}{nq_s^2 B^2} \nabla_{\perp} \dot{p}_{s\sigma 1} \quad (10.57)$$

The last two terms are smaller than the first two terms by the ratio ω/ω_{cs} and so may be ignored when the electron and ion fluid velocities are considered separately. However, when the perpendicular current, i.e. $\mathbf{j}_{1\perp} = \sum nq_s \mathbf{u}_{s\perp 1}$ is considered, the electron and ion $\mathbf{E} \times \mathbf{B}$ drift terms cancel so that the polarization terms becomes the leading terms involving the electric field. Because of the mass in the numerator, the ion polarization drift is much larger than the electron polarization drift. Thus, the perpendicular current comes from ion polarization drift and diamagnetic current

$$\mu_0 \mathbf{j}_{1\perp} = \frac{\mu_0 n m_i \dot{\mathbf{E}}_{1\perp}}{B^2} - \sum_s \frac{\nabla p_{s\perp 1} \times \mathbf{B}}{B^2} = \frac{1}{v_A^2} \dot{\mathbf{E}}_{1\perp} - \frac{\mu_0 \nabla p_{\perp 1} \times \mathbf{B}}{B^2} \quad (10.58)$$

where $p_{\perp 1} = \sum p_{s\perp 1}$. The terms involving $\dot{p}_{\perp 1}$ has been dropped because it is small by ω/ω_c compared to the $p_{\perp 1}$ term.

The center of mass perpendicular motions is

$$\mathbf{U}_{\perp 1} = \frac{\sum m_s n \mathbf{u}_{s\perp 1}}{\sum m_s n} \approx \mathbf{u}_{i\perp 1}$$

An important issue for the perpendicular motion is *whether $\mathbf{u}_{s\perp 1}$ is compressible or incompressible*. Let us temporarily ignore parallel motion and consider the continuity equation

$$\frac{\partial n_1}{\partial t} + n \nabla \cdot \mathbf{u}_{s\perp 1} = 0 \quad (10.59)$$

If $\nabla \cdot \mathbf{u}_{s\perp 1} = 0$, the mode does not involve any density perturbation, i.e. $n_1 = 0$, and is said to be an incompressible mode. On the other hand, if $\nabla \cdot \mathbf{u}_{s\perp 1} \neq 0$ then there are fluctuations in density and the mode is said to be compressible.

To proceed further, consider the vector identity

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \nabla \times \mathbf{F} - \mathbf{F} \cdot \nabla \times \mathbf{G}$$

If \mathbf{G} is spatially uniform, this identity reduces to $\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot \nabla \times \mathbf{F}$ which in turn vanishes if \mathbf{F} is the gradient of a scalar. Taking the divergence of Equation 10.57 and ignoring the polarization terms (they are of order ω/ω_{ci} and are only important when calculating the current which we are not interested in right now) gives

$$\nabla \cdot \mathbf{u}_{s\perp 1} = \frac{1}{B^2} \mathbf{B} \cdot \nabla \times \mathbf{E}_1 = \frac{1}{B} \hat{z} \cdot \nabla \times \mathbf{E}_1 \quad (10.60)$$

to lowest order. Setting $\mathbf{E}_1 = -\nabla\phi$ (i.e. assuming that the electric field is electrostatic) would cause the right hand side of Equation 10.60 to vanish, but such an assumption is overly restrictive because all that matters here is the z -component of $\nabla \times \mathbf{E}_1$. The z -component of $\nabla \times \mathbf{E}_1$ involves only the perpendicular component of the electric field (i.e. only the x and y components of the electric field) and so the least restrictive assumption for the right hand side of Equation 10.60 to vanish is to have $\mathbf{E}_{1\perp} = -\nabla_{\perp}\phi$. Thus, one possibility is to have $\mathbf{E}_{1\perp} = -\nabla_{\perp}\phi$ in which case the perpendicular electric field is electrostatic in nature and the mode is incompressible.

The other possibility is to have $\hat{z} \cdot \nabla \times \mathbf{E}_1 \neq 0$. In this case, invoking Faraday's law reduces Equation 10.60 to

$$\nabla \cdot \mathbf{u}_{s\perp 1} = -\frac{1}{B} \hat{z} \cdot \frac{\partial \mathbf{B}_1}{\partial t} = -\frac{1}{B} \frac{\partial B_{z1}}{\partial t} \quad (10.61)$$

Combining Equation 10.59 and Equation 10.61 and then integrating in time gives

$$\frac{n_1}{n} = \frac{B_{z1}}{B}$$

which shows that compression/rarefaction is associated with having finite B_{z1} (similar to the argument in Equation 14.17).

In summary, there are two general kinds of behavior:

Table 10.2: Incompressible and compressible modes from two-fluid theory.

Quantity	<i>Shear modes</i>	<i>Compressible modes</i>
n_1	0	$\neq 0$
$\nabla \cdot \mathbf{u}_{s\perp 1}$	0	$\neq 0$
$\nabla \times \mathbf{E}_{1\perp}$	$-\nabla_{\perp}\phi$	$\neq 0$
B_{z1}	0	$\neq 0$

Equation 10.58 provides a relationship between the perpendicular electric field and the perpendicular current. A relationship between the parallel electric field and the parallel current is now required. To obtain this, all vectors are decomposed into components parallel and perpendicular to the equilibrium magnetic field, i.e. $\mathbf{E}_1 = \mathbf{E}_{\perp 1} + E_{z1}\hat{z}$, $\nabla = \nabla_{\perp} + \hat{z}\partial_z$ etc. All quantities are assumed to be proportional to $f(x, y) \exp(ik_z z - i\omega t)$. Thus, Faraday's law can be written as

$$\nabla_{\perp} \times \mathbf{E}_{\perp 1} + \nabla_{\perp} \times E_{z1}\hat{z} + \hat{z} \frac{\partial}{\partial z} \times \mathbf{E}_{\perp 1} = -\frac{\partial}{\partial t} (\mathbf{B}_{\perp 1} + B_{z1}\hat{z})$$

which has a parallel component

$$\hat{z} \cdot \nabla_{\perp} \times \mathbf{E}_{\perp 1} = i\omega B_{z1}$$

and a perpendicular component

$$(\nabla_{\perp} E_{z1} - ik_z \mathbf{E}_{\perp 1}) \times \hat{z} = i\omega \mathbf{B}_{\perp 1} \quad (10.62)$$

Similarly Ampère's law can be decomposed into

$$\hat{z} \cdot \nabla_{\perp} \times \mathbf{B}_{\perp 1} = \mu_0 \mathbf{j}_{z1}$$

and

$$(\nabla_{\perp} B_{z1} - ik_z \mathbf{B}_{\perp 1}) \times \hat{z} = \mu_0 \mathbf{j}_{\perp 1} \quad (10.63)$$

Substituting Equation 10.58 into Equation 10.63 gives

$$(\nabla_{\perp} B_{z1} - ik_z \mathbf{B}_{\perp 1}) \times \hat{z} = \frac{i\omega}{v_A^2} \mathbf{E}_{\perp 1} - \frac{\mu_0 \nabla p_1 \times \hat{z}}{B}$$

or, after re-arrangement,

$$\nabla_{\perp} \left(B_{z1} + \frac{\mu_0 p_{\perp 1}}{B} \right) \times \hat{z} - ik_z \mathbf{B}_{\perp 1} \times \hat{z} = -\frac{i\omega}{v_A^2} \mathbf{E}_{\perp 1} \quad (10.64)$$

The shear and compressional modes are now considered separately.

10.8.1 Two-fluid shear modes

As discussed above these modes have $B_{z1} = 0$, $\mathbf{E}_{\perp 1} = -\nabla_{\perp} \phi_1$, and $\nabla \cdot \mathbf{u}_{s\perp 1} = 0$. We first consider the parallel component of the linearized equation of motion, namely

$$nm_s \frac{\partial u_{sz1}}{\partial t} = nq_s E_{z1} - \frac{\partial p_{s1}}{\partial z} \quad (10.65)$$

where $p_{s1} = \gamma_s n_{s1} k_B T_s$ and $\gamma = 1$ if the motion is isothermal and $\gamma_s = 3$ if the motion is adiabatic and the compression is one-dimensional. The isothermal case corresponds to $\omega^2/k_z^2 \ll k_B T_s/m_s$ and vice versa for the adiabatic case.

The continuity equation is

$$\frac{\partial n_1}{\partial t} + \nabla \cdot (n \mathbf{u}_{s1}) = 0$$

Because the shear mode is incompressible in the perpendicular direction, the continuity equation reduces to

$$\frac{\partial n_1}{\partial t} + \frac{\partial}{\partial z} (n_0 u_{sz1}) = 0$$

Taking the time derivative of Equation 10.65 gives (isothermal?)

$$\frac{\partial^2 u_{sz1}}{\partial t^2} - \gamma_s \frac{k_B T_s}{m_s} \frac{\partial^2 u_{sz1}}{\partial z^2} = \frac{q_s}{m_s} \frac{\partial E_{z1}}{\partial t} \quad (10.66)$$

which is similar to electron plasma wave and ion acoustic wave dynamics except it has *not* been assumed that E_{z1} is electrostatic.

Invoking the assumption that all quantities are of the form $f(x, y) \exp(ik_z z - i\omega t)$ Equation 10.66 can be solved to give

$$u_{sz1} = \frac{i\omega q_s}{m_s} \frac{E_{z1}}{\omega^2 - \gamma_s k_z^2 k_B T_s / m_s}$$

and so the relation between parallel current and parallel electric field is

$$\mu_0 j_{z1} = \frac{i\omega}{c^2} E_{z1} \sum_s \frac{\omega_{ps}^2}{\omega^2 - \gamma_s k_z^2 k_B T_s / m_s}$$

Using $\hat{z} \cdot \nabla \times \mathbf{B}_1 = \nabla \cdot (\mathbf{B}_1 \times \hat{z}) = \nabla \cdot (\mathbf{B}_{\perp 1} \times \hat{z})$ the parallel component of Ampère's law becomes for the shear wave

$$\nabla_{\perp} \cdot (\mathbf{B}_{\perp 1} \times \hat{z}) = \frac{i\omega}{c^2} E_{z1} \sum_s \frac{\omega_{ps}^2}{\omega^2 - \gamma_s k_z^2 k_B T_s / m_s} \quad (10.67)$$

Ion acoustic wave physics is embedded in Equation 10.67 as well as shear Alfvén physics. The ion acoustic mode can be retrieved by assuming that the electric field is electrostatic in which case $\mathbf{B}_{\perp 1}$ vanishes (???Why???). For the special case where the electric field is just in the z direction, and assuming that $k_B T_i / m_i \ll \omega^2 / k_z^2 \ll k_B T_e / m_e$ the right hand side of Equation 10.67 becomes

$$\left(\frac{\omega_{pi}^2}{\omega^2} - \frac{1}{k_z^2 \lambda_{De}^2} \right) E_{z1} = 0$$

which gives the ion acoustic wave $\omega^2 = k_z^2 k_B T_e / m_i$ (see (F. F. Chen 2016) and (Bellan 2008) Sec.4.2.1). This shows that the acoustic wave is associated with having finite E_{z1} and also requires $T_e \gg T_i$ in order to exist.

Returning to shear waves, we now assume that the electric field is not electrostatic so $\mathbf{B}_{\perp 1}$ does not vanish and Equation 10.67 has to be considered in its entirety. For shear waves the character of the parallel current changes depending on whether the wave parallel phase velocity is faster or slower than the electron thermal velocity:

1. The $\omega^2 / k_z^2 \gg k_B T_e / m_e$ case is called the *inertial limit* while
2. The $\omega^2 / k_z^2 \ll k_B T_e / m_e$ case is called the *kinetic limit*.

The perpendicular component of Faraday's law is (Equation 10.62)

$$\nabla_{\perp} E_{z1} \times \hat{z} - ik_z \mathbf{E}_{\perp 1} \times \hat{z} = i\omega \mathbf{B}_{\perp 1} \quad (10.68)$$

Substitution of $\mathbf{E}_{\perp 1}$ as determined from Equation 10.64 into Equation 10.68 gives

$$-\frac{i\omega}{v_A^2} \nabla_{\perp} E_{z1} \times \hat{z} - ik_z \left(\frac{\mu_0 \nabla p_{\perp 1}}{B} \times \hat{z} - ik_z \mathbf{B}_{\perp 1} \times \hat{z} \right) \times \hat{z} = \frac{\omega^2}{v_A^2} \mathbf{B}_{\perp 1}$$

which may be solved for $\mathbf{B}_{\perp 1}$ to give

$$\mathbf{B}_{\perp 1} = \frac{1}{\omega^2 - k_z^2 v_A^2} \left(-i\omega \nabla_{\perp} E_{z1} \times \hat{z} + ik_z v_A^2 \frac{\mu_0 \nabla_{\perp} p_{\perp 1}}{B} \right)$$

and

$$\mathbf{B}_{\perp 1} \times \hat{z} = \frac{1}{\omega^2 - k_z^2 v_A^2} \left(i\omega \nabla_{\perp} E_{z1} + ik_z v_A^2 \frac{\mu_0 \nabla_{\perp} p_{\perp 1}}{B} \times \hat{z} \right)$$

Substitution of $\mathbf{B}_{\perp 1} \times \hat{z}$ into Equation 10.67 gives

$$\nabla_{\perp} \cdot \left(\frac{1}{\omega^2 - k_z^2 v_A^2} \left(\nabla_{\perp} E_{z1} + k_z v_A^2 \frac{\mu_0 \nabla_{\perp} p_{\perp 1}}{\omega B} \times \hat{z} \right) \right) = E_{z1} \sum_s \frac{\omega_{ps}^2 / c^2}{\omega^2 - \gamma_s k_z^2 k_B T_s / m_s}$$

However, because $\nabla_{\perp} \cdot (\nabla_{\perp} p_{\perp 1} \times \hat{z}) = \nabla \cdot (\nabla p_{\perp 1} \times \hat{z}) = 0$ (divergence of a curl) the term involving pressure vanishes, leaving an equation involving E_{z1} only, namely

$$\nabla_{\perp} \cdot \left(\frac{1}{\omega^2 - k_z^2 v_A^2} \nabla_{\perp} E_{z1} \right) - E_{z1} \sum_s \frac{\omega_{ps}^2 / c^2}{\omega^2 - \gamma_s k_z^2 k_B T_s / m_s} = 0 \quad (10.69)$$

This is the fundamental equation for shear waves. On replacing $\nabla_{\perp} \rightarrow i\mathbf{k}_{\perp}$, Equation 10.69 becomes

$$\frac{k_{\perp}^2}{\omega^2 - k_z^2 v_A^2} + \frac{\omega_{pe}^2}{c^2} \frac{1}{\omega^2 - \gamma_e k_z^2 k_B T_e / m_e} + \frac{\omega_{pi}^2}{c^2} \frac{1}{\omega^2 - \gamma_i k_z^2 k_B T_i / m_i} = 0 \quad (10.70)$$

In the situation where $\omega^2 / k_z^2 \gg k_B T_e / m_e$, the second term dominates the third term since $\omega_{pe}^2 \gg \omega_{pi}^2$ and so Equation 10.70 can be recast as

$$\omega^2 = \frac{k_z^2 v_A^2}{1 + k_{\perp}^2 c^2 / \omega_{pe}^2} \quad (10.71)$$

which is called the *inertial Alfvén wave* (IAW). If $k_{\perp}^2 c^2 / \omega_{pe}^2$ is not too large, then ω / k_z is of the order of the Alfvén velocity and the condition $\omega^2 \gg k_z^2 k_B T_e / m_e$ corresponds to $v_A^2 \gg k_B T_e / m_e$ or

$$\beta_e = \frac{n k_B T_e}{B^2 / 2\mu_0} \ll \frac{m_e}{m_i}$$

Thus, inertial Alfvén wave shear modes exist only in the ultra-low β regime where $\beta_e \ll m_e / m_i$.

In the situation where $k_B T_i / m_i \ll \omega^2 / k_z^2 \ll k_B T_e / m_e$, Equation 10.69 can be recast as

$$\frac{k_{\perp}^2}{\omega^2 - k_z^2 v_A^2} - \frac{\omega_{pe}^2}{c^2} \frac{1}{k_z^2 k_B T_e / m_e} + \frac{\omega_{pi}^2}{c^2} \frac{1}{\omega^2} = 0 \quad (10.72)$$

Because ω^2 appears in the respective denominators of two distinct terms, Equation 10.72 is fourth order in ω^2 and so describes in two distinct modes. Let us suppose that the mode in question is much faster than the acoustic velocity, i.e. $\omega^2/k_z^2 \gg k_B T_e/m_i$. In this case the ion term can be dropped and the remaining terms can be re-arranged to give

$$\omega^2 = k_z^2 v_A^2 \left(1 + \frac{k_\perp^2}{v_A^2} \frac{k_B T_e}{m_e} \frac{c^2}{\omega_{pe}^2} \right)$$

This is called the *kinetic Alfvén wave* (KAW). Using

$$r_{fL}^2 = \frac{1}{v_A^2} \frac{k_B T_e}{m_e} \frac{c^2}{\omega_{pe}^2} = \frac{1}{\omega_{ci}^2} \frac{k_B T_e}{m_i} \quad (10.73)$$

as a fictitious ion Larmor radius calculated using the electron temperature instead of the ion temperature, the KAW dispersion relation can be expressed more succinctly as

$$\omega^2 = k_z^2 v_A^2 \left(1 + k_\perp^2 r_{fL}^2 \right) \quad (10.74)$$

If $k_\perp^2 r_{fL}^2$ is not too large, then ω/k_z is again of the order of v_A and so the condition $\omega^2 \ll k_z^2 k_B T_e/m_e$ corresponds to having $\beta_e \gg m_e/m_i$. The condition $\omega^2/k_z^2 \gg k_B T_e/m_i$ which was also assumed corresponds to assuming that $\beta_e \ll 1$. Thus, the KAW dispersion relation Equation 10.74 is valid in the regime $m_e/m_i \ll \beta_e \ll 1$.

Let us now consider the situation where $\omega^2/k_z^2 \ll k_B T_i/m_i, k_B T_e/m_e$. In this case Equation 10.72 again reduces to

$$\omega^2 = k_z^2 v_A^2 \left(1 + k_\perp^2 r_{fL}^2 \right) \quad (10.75)$$

but this time

$$r_{fL}^2 = \frac{1}{\omega_{ci}^2} \frac{k_B(T_e + T_i)}{m_i} \quad (10.76)$$

This situation would describe shear modes in a high β plasma (ion thermal velocity faster than Alfvén velocity).

To summarize: the shear mode has $B_{z1} = 0, E_{z1} \neq 0, j_{z1} \neq 0, E_{\perp 1} = -\nabla\phi_1$ and exists in the form of the inertial Alfvén wave for $\beta_e \ll m_e/m_i$ and in the form of the kinetic Alfvén wave for $\beta_e \gg m_e/m_i$. The shear mode involves incompressible perpendicular motion, i.e., $\nabla \cdot \mathbf{u}_{s\perp 1} = i\mathbf{k}_\perp \cdot \mathbf{u}_{s\perp 1} = 0$, which means that \mathbf{k}_\perp is orthogonal to $\mathbf{u}_{s\perp 1}$. For example, in Cartesian geometry, this means that if $\mathbf{u}_{s\perp 1}$ is in the x direction, then \mathbf{k}_\perp must be in the y direction while in cylindrical geometry, this means that if $\mathbf{u}_{s\perp 1}$ is in the θ direction, then \mathbf{k}_\perp must be in the r direction. The inertial Alfvén wave is known as a cold plasma wave because its dispersion relation does not depend on temperature (such a mode would exist even in the limit of a cold plasma). The kinetic Alfvén wave depends on the plasma having finite temperature and is therefore called a warm plasma wave. The shear mode can be coupled to ion acoustic modes since both shear and ion acoustic modes involve finite $E_{\parallel 1}$.

10.8.2 Two-fluid compressional modes

The compressional mode involves assuming that B_{z1} is finite and $E_{z1} = 0$. Having $E_{z1} = 0$ means that there is no parallel motion and, in particular, implies that $j_{z1} = 0$. Thus, for the compressional mode Faraday's law has the form

$$\begin{aligned}\nabla_{\perp} \cdot (\mathbf{E}_{\perp 1} \times \hat{z}) &= i\omega B_{z1} \\ -ik_z \mathbf{E}_{\perp 1} \times \hat{z} &= i\omega \mathbf{B}_{\perp 1}\end{aligned}\quad (10.77)$$

Using Equation 10.77 to substitute for $\mathbf{B}_{\perp 1}$ in Equation 10.64 and then solving for $\mathbf{E}_{\perp 1}$ gives

$$\mathbf{E}_{\perp 1} = \frac{i\omega v_A^2}{\omega^2 - k_z^2 v_A^2} \nabla_{\perp} \left(B_{z1} + \frac{\mu_0 p_{\perp 1}}{B} \right) \times \hat{z}$$

Since

$$\mathbf{E}_{\perp 1} \times \hat{z} = -\frac{i\omega v_A^2}{\omega^2 - k_z^2 v_A^2} \nabla_{\perp} \left(B_{z1} + \frac{\mu_0 p_{\perp 1}}{B} \right)$$

Equation 10.77 becomes

$$\nabla_{\perp} \cdot \left(\frac{v_A^2}{\omega^2 - k_z^2 v_A^2} \nabla_{\perp} \left(B_{z1} + \frac{\mu_0 p_{\perp 1}}{B} \right) \right) + B_{z1} = 0 \quad (10.78)$$

If we assume that the perpendicular motion is adiabatic, then

$$\frac{p_{\perp 1}}{p} = \gamma \frac{n_1}{n} = \gamma \frac{B_{z1}}{B}$$

Substitute for $p_{\perp 1}$ in Equation 10.78 gives (???)

$$\nabla_{\perp} \cdot \left(\frac{v_A^2 + c_s^2}{\omega^2 - k_z^2 v_A^2} \nabla_{\perp} B_{z1} \right) + B_{z1} = 0 \quad (10.79)$$

where

$$c_s^2 = \gamma k_B \frac{T_e + T_i}{m_i}$$

On replacing $\nabla_{\perp} \rightarrow i\mathbf{k}_{\perp}$, Equation 10.79 becomes the dispersion relation

$$\frac{-k_{\perp}^2(v_A^2 + c_s^2)}{\omega^2 - k_z^2 v_A^2} + 1 = 0$$

or

$$\omega^2 = k^2 v_A^2 + k_{\perp}^2 c_s^2$$

where $k^2 = k_z^2 + k_{\perp}^2$. Since $\nabla \cdot \mathbf{u}_{s\perp 1} = i\mathbf{k}_{\perp} \cdot \mathbf{u}_{s\perp 1} \neq 0$, the perpendicular wave vector \mathbf{k}_{\perp} is at least partially co-aligned with the perpendicular velocity.

10.8.3 Differences between the two-fluid and MHD descriptions

The two-fluid description shows that the Alfvén mode (finite E_z) appears as either an inertial or a kinetic Alfvén wave depending on the plasma β ; the MHD description assumes that $E_z = 0$ for this mode and does not distinguish between inertial and kinetic modes. The two-fluid description also shows that finite E_z will give ion acoustic modes in the parallel direction which are decoupled. The MHD description predicts a so-called sound wave which differs from the ion acoustic wave because the MHD sound wave does not have the requirement that $T_e \gg T_i$; the MHD sound wave is an artifact for parallel propagation in a plasma with low collisionality (if the collisions are sufficiently large, then the plasma would behave like a neutral gas). Then MHD description predicts a coupling between oblique sound waves via a square root relation (Equation 10.45) which does not exist in the two-fluid model.

10.8.4 KAW properties

The solution of Equation 10.37 exhibits singularities as $t \rightarrow \infty$ naturally suggests that the microscopic length-scale physics neglected in the ideal MHD fluid description should be included in the long-time-scale dynamics of SAWs. For low-frequency SAWs, one can readily recognize the relevant perpendicular (to \mathbf{B}_0) microscopic scales are either the ion Larmor radius, $r_{iL} = v_{ti}/\omega_{ci}$ with v_{ti} and ω_{ci} being, respectively, the ion thermal speed and ion cyclotron frequency, and/or $r_{eL} = v_{te}/\omega_{ce}$ with v_{te} being the electron thermal speed. Including the effects of finite r_{iL} , r_{eL} and/or T_e in the SAW dynamics then led to the discovery of the so-called *kinetic Alfvén wave* (KAW) [Hasegawa and Chen].

In KAWs, parallel electric field E_{\parallel} can be developed and facilitate particle heating, acceleration, and transport, especially for electrons. However, the inverse mechanism is also possible that electrons moving along the magnetic field in the opposite direction, become retarded by this field component and feed their energy into the KAW. For example, Hasegawa [1979] showed that an electron beam moving along an inhomogeneous magnetic field can excite KAWs.

Excitation of KAWs requires $\beta < 1$. It has been found in the plasma sheet, at the plasma sheet boundary layer (PSBL), and in the inner magnetosphere.

KAW differ from SAW because the short wavelength requires a significant E_{\parallel} to maintain charge neutrality due to ion density perturbations caused by the ion polarization drift. When $v_{te} > v_A$, the parallel electric field counteracts electron pressure that would push the electrons away from the ion density perturbations. When $v_{te} < v_A$, the electric field must overcome the electron inertia that prevents the electrons from responding rapidly to the ion density perturbations. E_{\parallel} associated with small-scale KAWs may efficiently accelerate particles on the magnetic field lines. (Chaston+ 2009) presented observations in the magnetotail from the Cluster spacecraft showing that KAWs radiate outward from the X-line with outward energy fluxes equivalent to that contained in the outstreaming ions. Wave-particle energy exchange

between KAWs and plasmas near the dayside magnetopause has been confirmed by MMS observations (Gershman+ 2017).

We consider a KAW with a wave vector \mathbf{k} in the XZ-plane, same as in the SAW case. In many aspects, KAW is similar to SAW: \mathbf{E} shall oscillate in the X-direction; \mathbf{B} shall oscillate in the Y-direction. The electric current of the wave $i\mathbf{k} \times \mathbf{B}_1/\mu_0$ is still in the XZ-plane. The timescale of the variations of the wave field of KAW is much longer than the ion gyroperiod. However, unlike SAW, KAW has

- a perpendicular scale k_\perp^{-1} that is comparable to the particle kinetic scale r_{iL} .
- a very oblique wave vector $k_\perp \gg k_\parallel$ so that the wave is not strongly affected by the Landau damping.
- $k_\perp^{-1} \ll k_\parallel^{-1}$. The finite-Larmor-radius effect starts to become important as the perpendicular wavelength is comparable to the ion gyromotion: Ions can not follow the $\mathbf{E} \times \mathbf{B}$ drift in the electric fields of KAW, because they encounter significantly different electric field in the different phases of the gyromotion. Electrons are still frozen-in in the presence of the wave field. The difference in the ion and electron motion in the perpendicular direction introduces charge separation and coupling to the electrostatic mode. Because the wave electric field E_x is mainly parallel to the \mathbf{k} , $\nabla \times \mathbf{E} = i\mathbf{k} \times \mathbf{E}$ is small and $\nabla \cdot \mathbf{E} = i\mathbf{k} \cdot \mathbf{E}$ is relatively large for KAW. Accordingly, the perpendicular wave electric field E_x is mostly electrostatic in KAW.
- Because of charge separation in KAW, electrons need to move along the magnetic field to preserve the charge neutrality. Associated with the parallel motion of electron, a small wave electric field E_\parallel is established, the existence of which is a distinct feature of KAW. The parallel motion of electrons creates a field-aligned current j_\parallel of KAW. From Ampère's law $i\mathbf{k} \times \mathbf{B}_1 = \mu_0 \mathbf{j}$, the field-aligned current \mathbf{j}_z produces a wave magnetic field \mathbf{B}_{1y} . As a result, KAW is in fact an EM wave. The ratio of the wave electric field to the wave magnetic field is $V_A \sqrt{1 + k_x^2 r_{iL}^2}$ (Stasiewicz et al. 2000). The kinetic correction $k_x^2 r_{iL}^2$ introduces a deviation of the ratio E_x/B_y from one V_A as in the SAWs.

While SAW satisfy $\omega = k_\parallel v_A$, the dispersion relation of KAW can be written as (Johnson & Cheng, 1997)

$$\omega^2 = k_\parallel^2 v_A^2 \left[\frac{1}{1 - I_0(k_\perp^2 r_{iL}^2) e^{-k_\perp^2 r_{iL}^2}} + \frac{T_e}{T_i} \right] k_\perp^2 r_{iL}^2 \quad (10.80)$$

where I_0 is the modified Bessel function. Using a Padé approximation $I_0(x)e^{-x} \approx 1/(1+x)$ when $x = k_\perp^2 r_{iL}^2 \sim \mathcal{O}(1)$, the relation can be simplified to

$$\omega^2 = k_\parallel^2 v_A^2 \left[1 + \left(1 + \frac{T_e}{T_i} k_\perp^2 r_{iL}^2 \right) \right] = k_\parallel^2 v_A^2 [1 + k_\perp^2 (r_{iL}^2 + r_{fL}^2)] \quad (10.81)$$

where $r_{fL} = \sqrt{k_B T_e / m_i} / \omega_{ci}$. So we recover the two-fluid dispersion relation of KAW Equation 10.74. We can see from Equation 10.81 that the phase speed for KAW is always larger than v_A .

While SAWs do not have ion density perturbations, KAWs do. In compressibility is only required in the perpendicular direction, but not in the parallel direction w.r.t. the magnetic field.

The polarizations of KAW can be expressed as

$$\left| \frac{\delta \mathbf{E}_\perp}{\delta \mathbf{B}_\perp} \right| = v_A (1 + k_\perp^2 r_{iL}^2) [1 + k_\perp^2 (r_{iL}^2 + r_{fL}^2)]^{-1/2}$$

where $\delta \mathbf{E}_\perp \perp \delta \mathbf{B}_\perp$. KAWs are right-hand polarized. The parallel electric field is

$$\left| \frac{\delta E_\parallel}{\delta E_\perp} \right| = k_\parallel k_\perp \frac{r_{fL}^2}{1 + k_\perp^2 r_{iL}^2}$$

It has been shown that KAWs can be generated via magnetic reconnection, mode conversion, and phase mixing.

(Gurram, Egedal, and Daughton 2021) used a 2D VPIC simulation to show a transition from KAWs to SAWs from the immediate vicinity of the reconnection region to the exhaust. They checked two important quantities we have seen above:

- the transition in wave speeds from super Alfvénic near the X-point to order of the Alfvén speed deeper into the exhaust ($\sim 60d_i$ from the X-point);
- the transition of $k_\perp d_i$ from larger than unity to smaller than unity.

It would be very intuitive to compare animations between [MHD Alfvén waves](#) and [kinetic Alfvén waves](#). In a typical MHD Alfvén wave, the particles (yellow) move freely along the magnetic field lines (blue). In a kinetic Alfvén wave, some particles become trapped in the weak spots of the wave's magnetic field and ride along with the wave as it moves through space.

Almost half a century after the discovery of KAW, (L. Chen, Zonca, and Lin 2021) demonstrated that the proper treatment of this wave requires gyrokinetics (Chapter 15). However, under some circumstances ($\beta \gg 1?$) the two-fluid theory can recover most if not all of the KAW physics. I need to go over the derivations!

I am not the only person who is confused by so many different names assigned to Alfvén waves. [The Alfvén Wave Zoo](#) is a nice review of all the existence names related to Alfvén waves. Many wave modes degenerate into the classical Alfvén mode in the small k limitation, but they may show different polarizations.

10.9 Kinetic Slow Modes

The least-damped kinetic slow mode is associated with the ion-acoustic (IA) wave and a nonpropagating (NP) mode, both of which exhibit an anticorrelation between δn and δB . A

comparison of the damping rates of the IA mode and the NP mode suggests that the IA mode is the dominating kinetic slow mode at low β_{\parallel} , while the NP mode is the dominating kinetic slow mode at high β_{\parallel} (Verscharen, Chen, and Wicks 2017). Temperature anisotropies alter the dispersion relations and the damping behavior of slow modes in kinetic plasmas, and can drive the NP mode (i.e., the mirror mode) to be unstable.

10.10 Particle Motions

In Fourier transform space, integrating the velocity equations to obtain the coordinates is done by simply dividing by $-i\omega$. For a simple case where $E_y = E_z = 0$ so that \mathbf{E} has only an x -component, we find from Equation 10.9

$$x_s = -\frac{q_s E_x}{m_s(\omega^2 - \omega_j^2)}$$

$$y_s = \frac{\pm\omega_{cs}}{i\omega} x_s$$

so that in general, the trajectory is elliptical. For $\omega \ll \omega_{cs}$, we find $x_s \ll y_s$, so the motion is principally across both the \mathbf{E} and \mathbf{B}_0 directions. However, for $\omega \gg \omega_{cj}$, $x_j \gg y_j$ the motion is principally parallel to the electric field. In this latter case, we would call the particles *unmagnetized*, since the magnetic influence is small. Since it is possible for the wave frequency to be well above the ion cyclotron frequency at the same time it is well below the electron cyclotron frequency, it is possible for ions to be effectively unmagnetized while electrons are magnetized. From the discussion here, it is also clear that in the MHD low frequency regime the $\mathbf{E} \times \mathbf{B}$ drift is important, whereas for high frequency regimes (i.e. unmagnetized), the electric field influence is more important.

When $\omega \simeq \omega_{cs}$, then the linear solutions exhibit resonance effects with large amplitudes, and at resonance, the radius increases uniformly in time and no steady-state solution exists. In this vicinity, we expect the cold plasma approximation to fail and either thermal, inhomogeneous, or nonlinear effects to dominate the dynamics.

10.11 Cold Nonuniform Plasma

Waves and dispersion relations in a uniform plasma is generally nice and easy. However, more interesting and realistic waves shall be found in nonuniform plasmas.

10.11.1 Simple EM Wave

Let us start off by examining a very simple case. Consider a plane electromagnetic wave, of frequency ω , propagating along the z -axis in an unmagnetized plasma whose refractive index,

n , is a function of z . We assume that the wave normal is initially aligned along the z -axis, and, furthermore, that the wave starts off polarized in the y -direction. It is easily demonstrated that the wave normal subsequently remains aligned along the z -axis, and also that the polarization state of the wave does not change. Thus, the wave is fully described by

$$E_y(z, t) \equiv E_y(z) \exp(-i\omega t)$$

and

$$B_x(z, t) \equiv B_x(z) \exp(-i\omega t)$$

It can easily be shown (???) that $E_y(z)$ and $B_x(z)$ satisfy the differential equations

$$\frac{d^2 E_y}{dz^2} + k_0^2 n^2 E_y = 0 \quad (10.82)$$

and

$$\frac{d c B_x}{dz} = -i k_0 n^2 E_y \quad (10.83)$$

respectively. Here, $k_0 = \omega/c$ is the wave-number in free space. Of course, the actual wave-number is $k = k_0 n$.

The solution to Equation 10.82 for the case of a homogeneous plasma, for which n is constant, is straightforward:

$$E_y = A e^{i\phi(z)} \quad (10.84)$$

where A is a constant, and

$$\phi = \pm k_0 n z \quad (10.85)$$

The solution Equation 10.84 represents a wave of constant amplitude, A , and phase, $\phi(z)$. According to Equation 10.85, there are, in fact, two independent waves which can propagate through the plasma. The upper sign corresponds to a wave which propagates in the $+z$ -direction, whereas the lower sign corresponds to a wave which propagates in the $-z$ -direction. Both waves propagate with the constant phase velocity c/n .

In general, if $n = n(z)$ then the solution of Equation 10.82 does not remotely resemble the wave-like solution Equation 10.84. However, in the limit in which $n(z)$ is a “slowly varying” function of z (exactly how slowly varying is something which will be established later on), we expect to recover wave-like solutions. Let us suppose that $n(z)$ is indeed a “slowly varying” function, and let us try substituting the wave solution Equation 10.84 into Equation 10.82. We obtain

$$\left(\frac{d\phi}{dz} \right)^2 = k_0^2 n^2 + i \frac{d^2 \phi}{dz^2} \quad (10.86)$$

This is a non-linear differential equation which, in general, is very difficult to solve. However, we note that if n is a constant then $d^2\phi/dz^2 = 0$. It is, therefore, reasonable to suppose that

if $n(z)$ is a “slowly varying” function then the last term on the right-hand side of the above equation can be regarded as being small. Thus, to a first approximation Equation 10.86 yields

$$\frac{d\phi}{dz} \simeq \pm k_0 n$$

and

$$\frac{d^2\phi}{dz^2} \simeq \pm k_0 \frac{dn}{dz} \quad (10.87)$$

It is clear from a comparison of Equation 10.86 and Equation 10.87 that $n(z)$ can be regarded as a “slowly varying” function of z as long as its variation length-scale is far longer than the wavelength of the wave. In other words, provided that $(dn/dz)/(k_0 n^2) \ll 1$.

The second approximation to the solution is obtained by substituting Equation 10.87 into the right-hand side of Equation 10.86:

$$\frac{d\phi}{dz} \simeq \pm \left(k_0^2 n^2 \pm i k_0 \frac{dn}{dz} \right)^{1/2}$$

This gives

$$\frac{d\phi}{dz} \simeq \pm k_0 n \left(1 \pm \frac{i}{k_0 n} \right)^{1/2} \simeq \pm k_0 n + \frac{i}{2n} \frac{dn}{dz}$$

where a binomial expansion has been used. The above expression can be integrated to give

$$\phi \sim \pm k_0 \int^z n dz + i \log(n^{1/2}) \quad (10.88)$$

Substitution of Equation 10.88 into Equation 10.84 yields the final result

$$E_y \simeq A n^{-1/2} \exp(\pm i k_0 \int^z n dz) \quad (10.89)$$

It follows from Equation 10.83 that

$$cB_x \simeq \mp A n^{1/2} \exp \left(\pm i k_0 \int^z n dz \right) - \frac{iA}{2k_0 n^{3/2}} \frac{dn}{dz} \exp \left(\pm i k_0 \int^z n dz \right) \quad (10.90)$$

Note that the second term is small compared to the first, and can usually be neglected.

Let us test to what extent Equation 10.89 is a good solution of Equation 10.82 by substituting this expression into the left-hand side of the equation. The result is

$$\frac{A}{n^{1/2}} \left[\frac{3}{4} \left(\frac{1}{n} \frac{dn}{dz} \right)^2 - \frac{1}{2n} \frac{d^2n}{dz^2} \right] \exp \left(\pm i k_0 \int^z n dz \right)$$

This must be small compared with either term on the left-hand side of Equation 10.82. Hence, the condition for Equation 10.89 to be a good solution of Equation 10.82 becomes

$$\frac{1}{k_0^2} \left| \frac{3}{4} \left(\frac{1}{n^2} \frac{dn}{dz} \right)^2 - \frac{1}{2n^3} \frac{d^2n}{dz^2} \right| \ll 1 \quad (10.91)$$

The solutions Equation 10.89 and Equation 10.90 (without the second term) are most commonly referred to as the *WKB* solutions, in honour of G. Wentzel, H.A. Kramers, and L. Brillouin, who are credited with independently discovering these solutions (in a quantum mechanical context) in 1926. Actually, H. Jeffries wrote a paper on the WKB solutions (in a wave propagation context) in 1923. Hence, some people call them the WKBJ solutions (or even the JWKB solutions). To be strictly accurate, the WKB solutions were first discussed by Liouville and Green in 1837, and again by Rayleigh in 1912. The advance in science discovery is always a collective achievement.

Recall, that when a propagating wave is normally incident on an interface, where the refractive index suddenly changes (for instance, when a light wave propagating through air is normally incident on a glass slab), there is generally significant reflection of the wave. However, according to the WKB solutions, when a propagating wave is normally incident on a medium in which the refractive index changes slowly along the direction of propagation of the wave then the wave is not reflected at all. This is true even if the refractive index varies very substantially along the path of propagation of the wave, as long as it varies slowly. The WKB solutions imply that as the wave propagates through the medium its wave-length gradually changes. In fact, the wave-length at position z is approximately $\lambda(z) = 2\pi/k_0 n(z)$. The WKB solutions also imply that the amplitude of the wave gradually changes as it propagates. The amplitude of the electric field component is inversely proportional to $n^{1/2}$, whereas the amplitude of the magnetic field component is directly proportional to $n^{1/2}$. Note, however, that the energy flux in the z -direction, given by the Poynting vector $-(E_y B_x^* + E_y^* B_x)/(4\mu_0)$, remains constant (assuming that n is predominately real).

Of course, the WKB solutions are only approximations. In reality, a wave propagating into a medium in which the refractive index is a slowly varying function of position is subject to a small amount of reflection. However, it is easily demonstrated that the ratio of the reflected amplitude to the incident amplitude is of order $(dn/dz)/(k_0 n^2)$. Thus, as long as the refractive index varies on a much longer length-scale than the wavelength of the radiation, the reflected wave is negligibly small. This conclusion remains valid as long as the inequality Equation 10.91 is satisfied. This inequality obviously breaks down in the vicinity of a point where $n^2 = 0$. We would, therefore, expect strong reflection of the incident wave from such a point. Furthermore, the WKB solutions also break down at a point where $n^2 \rightarrow \infty$, since the amplitude of B_x becomes infinite.

10.11.2 Electron Cyclotron Resonance Heating

Let us look at the problem of electron cyclotron resonance heating. The resonance condition is $R = \infty$; the governing equation has the form

$$\frac{d^2E}{dz^2} + \frac{\omega^2}{c^2} \left[1 - \frac{\omega_{pe}(z)^2}{\omega(\omega - \omega_{ce}(z))} \right] E = 0$$

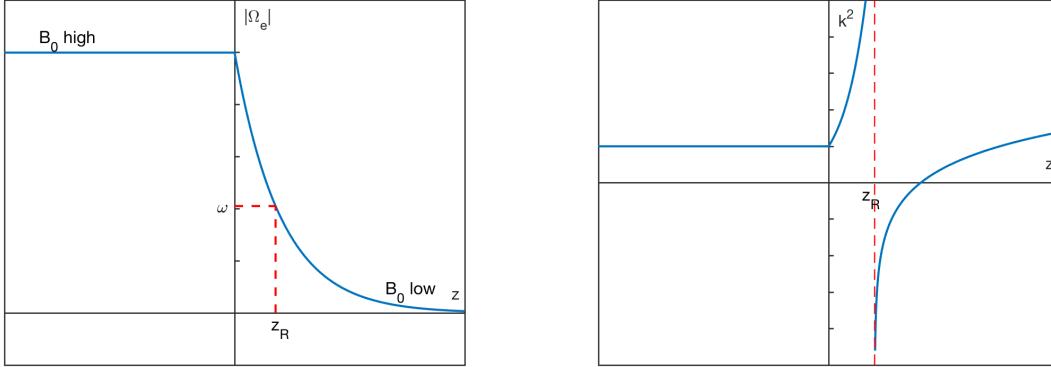


Figure 10.2: Plasma frequency and wave number as functions of z in a non-uniform plasma for the electron cyclotron resonance mode.

We use z here to remind ourselves of the fact that the wave property changes along the field line. Imagine a wave $\sim e^{i(kz-\omega t)}$ incident into a plasma with fixed density n_0 and varying magnetic field $B_0(z)$ as shown in Figure 10.2(a). At $z = z_R$, $\omega_{ce}(z) = \omega_{ce}(z_R) = \omega$. Then we can draw $k^2(z)$ as a function of z as in Figure 10.2(b). There is a pole at $z = z_R$, which indicates resonance since $k^2 \rightarrow \infty$. There is also a zero on the right of z_R .

Close to $z = z_R$, we have $k^2(z) \sim \frac{\text{const.}}{z - z_R}$. If we write

$$\Omega_e(z) = -A'(z - z_R) + \omega, \quad \zeta = z - z_R$$

then (ignore the constants)

$$\frac{d^2E}{dz^2} - \frac{\text{const.}}{z - z_R} E = 0 \Rightarrow \frac{d^2E}{d\zeta^2} - \frac{1}{\zeta} E = 0$$

Now let's stare at Figure 10.2(b) for a few seconds. If there is a wave from left to right, there will be a resonance at $z = z_R$; but if there is a wave from right to left, then the wave will be attenuated before it reaches $z = z_R$ because there is a zero ahead. This means that waves from different origin will have different behaviors!

10.11.3 O-mode

Next consider EM waves in non-magnetized plasma of ordinary O-mode:

$$k^2(x) = \frac{\omega^2}{c^2} n^2 = \frac{\omega^2}{c^2} \left[1 - \frac{\omega_{pe}(x)^2}{\omega^2} \right] \quad \text{if } n \text{ is non-uniform}$$

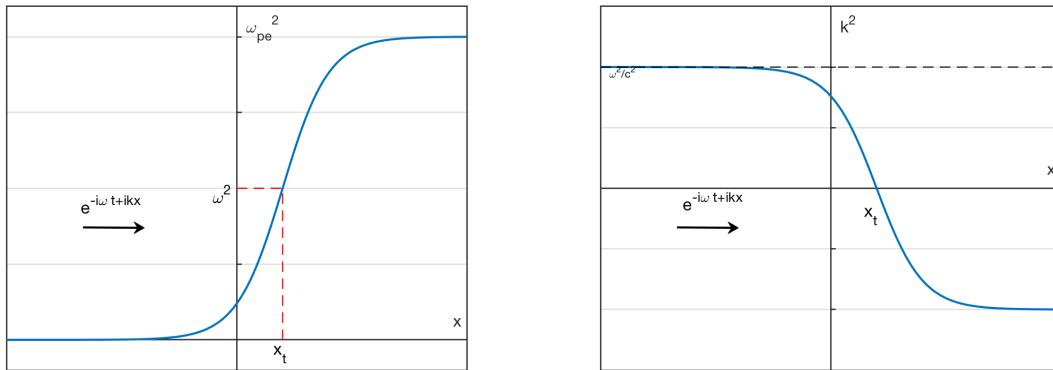


Figure 10.3: Plasma frequency and wave number as functions of x in a non-uniform plasma for O mode.

Note that there is no energy dissipation, because it is collisionless. Imagine a wave $\sim e^{i(kx - \omega t)}$ incident into a plasma with density $n_0(x)$ shown in Figure 10.3. We encounter a cutoff at $x = x_t$, $\omega_{pe}(x_t) = \omega$. Therefore we can draw $k^2(x)$ as a function of x as in Figure 10.3. Then the properties of the wave can be categorized into two regimes:

$$\begin{aligned} k^2 < 0 &\Rightarrow E \sim e^{\pm ikx} \text{ propagating} \\ k^2 > 0 &\Rightarrow E \sim e^{\pm |k|x} \text{ attenuating} \end{aligned}$$

A natural question comes up from this picture: *how does the propagating wave transform into attenuating wave?* Stokes solved this during his honey moon, which is now known as the Stokes phenomenon (Section 3.11.3).

10.12 Warm Uniform Plasma

The analysis of cold plasma waves, although very complicated already, leaves out all of the physics that relates to finite temperature effects. These effects may be included in varying degrees of approximation, and it is instructive to include at this stage only the simplest thermal correction terms through the inclusion of a finite pressure term. For the sake of simplicity, we

only consider the 1D case. The approach is very similar to cold plasma situation, except that we have *pressure* included in the equation, and we also need to specify the relation of pressure and temperature through the equation of state.

10.12.1 Two-fluid approach

Equilibrium:

$$m_i = \infty, T_i = 0, v_{i,0} = 0, n_{e0} = n_{i0} = n_0, E_0 = 0, B_0 = 0 \\ P_0 = \text{const.}, v_{e0} = 0$$

The equation of motion for electron is

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{e}{m_e} \mathbf{E} - \frac{\nabla P}{\rho}$$

Assume

$$P = P_0 \left(\frac{n}{n_0} \right)^\gamma \quad \text{adiabatic} \\ P = n k_B T_e \quad \text{isothermal} (\gamma = 1)$$

Define $P_1 = (\partial P / \partial \rho) \rho_1 \equiv v_{th}^2 \rho_1 = v_{th}^2 n_1 m = k_B T n_1$. Decompose the primitive variables into equilibrium and perturbation components:

$$n = n_0 + n_1 = n_0 + \tilde{n}_1 e^{-i\omega t + ikx} \\ v = \mathbf{v}_0 + \mathbf{v}_1 = \tilde{\mathbf{v}}_1 e^{-i\omega t + ikx} \\ P = P_0 + P_1 = P_0 + \tilde{P}_1 e^{-i\omega t + ikx}$$

Substituting into the equation of motion and keeping only first order terms, we get

$$\frac{\partial v_1}{\partial t} = -\frac{e}{m_e} (E_1) - \frac{\nabla(P_1)}{n_0 m_e} \\ -i\omega v_1 = -\frac{e}{m_e} E_1 - \frac{ik v_{th}^2 n_1}{n_0}$$

The continuity equation gives

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial x} (nv) = 0 \\ \Rightarrow v_1 = \frac{\omega n_1}{k n_0}$$

Substituting this into the linearized equation of motion, we get

$$n_1 = \frac{kn_0eE_1}{im_e(\omega^2 - k^2v_{th}^2)}$$

which is the density perturbation in response to E_1 . Following the same approach as before, we can easily get the dielectric function:

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2 - k^2v_{th}^2}$$

Let $\epsilon = 0$, we get the dispersion relation for warm plasma

$$\omega^2 = k^2v_{th}^2 + \omega_{pe}^2$$

This is called the *Bohm-Gross dispersion relation* (BGDR). (Swanson 2012) has shown that by keeping higher order terms,

$$\omega^2 = \omega_{pe}^2 \left[1 + \frac{3}{2} \left(\frac{kv_{th}}{\omega_{pe}} \right)^2 + \frac{15}{4} \left(\frac{kv_{th}}{\omega_{pe}} \right)^4 \right]$$

so it is apparent that *the moment expansion is an expansion in the ratio of the thermal velocity to the phase velocity*. The BGDR now resolves the ambiguity in cold plasma theory and supports the notion that $P = 0$ in cold plasma is a cutoff rather than a resonance, since this dispersion relation describes a wave with a cutoff at ω_{pe} that propagates near the electron thermal speed for high frequencies.

We see from this example that the fluid equations which are based on moment expansions are valid as long as the phase velocity is large compared to the thermal speed, $v_p \gg v_{th}$.

10.12.2 Vlasov approach

Assume Maxwellian distribution for electrons:

$$g_e(v) = \frac{1}{\sqrt{2\pi}} \frac{1}{v_{th,e}} e^{-v^2/2v_{th,e}^2}$$

Assuming $\frac{\omega}{k} \gg v_{th,e}$, i.e. the phase speed is much larger than the characteristic thermal speed, we can do Taylor series expansion

$$\frac{1}{(v - \omega/k)^2} = \frac{1}{(\omega/k)^2} \frac{1}{(1 - kv/\omega)^2} \approx \frac{k^2}{\omega^2} \left[1 + \frac{2kv}{\omega} + \frac{3k^2v^2}{\omega^2} + \dots \right]$$

Then the dielectric function is

$$\begin{aligned}\frac{\epsilon}{\epsilon_0} &= 1 - \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} dv g(v) \frac{k^2}{\omega^2} \left[1 + \frac{2kv}{\omega} + \frac{3k^2v^2}{\omega^2} + \dots \right] \\ &\doteq 1 - \frac{\omega_{pe}^2}{\omega^2} \left[1 + \frac{3k^2v_{th}^2}{\omega^2} \right].\end{aligned}$$

Let $\epsilon = 0$, we can get the dispersion relation for warm plasma,

$$\begin{aligned}&\text{lowest order: } \omega = \pm \omega_{pe} \\ &\text{first order: } 1 - \frac{\omega_{pe}^2}{\omega^2} \left(1 + \frac{3k^2v_{th}^2}{\omega_{pe}^2} \right) = 0 \Rightarrow \omega^2 = \omega_{pe}^2 + 3k^2v_{th,e}^2\end{aligned}$$

Note here we insert 0^{th} order solution to 1^{st} order equation to get the next level approximation. Comparing with the results from 2-fluid theory, we see that the expression is very similar except a discrepancy in the coefficient. This is owing to the fact that we do not specify γ in the equation of state. Actually, there are still ambiguities and debates about the exact equation of state. I wonder if I can get the exact coefficient under some assumption.

The opposite limit case: $\frac{\omega}{k} \ll v_{th}$. From 2-fluid theory,

$$\frac{\epsilon}{\epsilon_0} \approx 1 + \frac{\omega_{pe}^2}{k^2v_{th,e}^2}$$

From Vlasov theory,

$$\begin{aligned}\frac{\epsilon}{\epsilon_0} &= 1 - \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} dv \frac{\partial g/\partial v}{v - \omega/k} = 1 - \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} \frac{1}{v - \omega/k} \frac{1}{\sqrt{2\pi}} \frac{1}{v_{th,e}} \left(\frac{-v}{v_{th,e}^2} \right) e^{-v^2/2v_{th,e}^2} dv \\ &= 1 + \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} \frac{1}{v_{th,e}^2} \frac{1}{\sqrt{2\pi}} e^{-v^2/2v_{th,e}^2} dv \\ &= 1 + \frac{\omega_{pe}^2}{k^2v_{th,e}^2}\end{aligned}$$

So we can see in the two limit cases that they “almost” give the same results!

If we include ion motion in the 2-fluid theory ($n_{1i} \neq 0, T_i \neq 0, m_i \neq \infty$), applying the linear superposition property, we have

$$\frac{\epsilon}{\epsilon_0} = \frac{1}{\nabla \cdot (\epsilon_0 \mathbf{E}_1)} - \underbrace{\frac{\omega_{pe}^2}{\omega^2 - k^2v_{th,e}^2}}_{en_{1e}} - \underbrace{\frac{\omega_{pi}^2}{\omega^2 - k^2v_{th,i}^2}}_{en_{1i}}$$

10.12.3 Ion-acoustic wave

Assume $v_{th,e} \gg \frac{\omega}{k} \gg v_{th,i}$, we have the simplified dielectric function

$$\frac{\epsilon}{\epsilon_0} = 1 + \frac{\omega_{pe}^2}{k^2 v_{th,e}^2} - \frac{\omega_{pi}^2}{\omega^2}$$

Assume quasi-neutrality condition: $n_{1e} \approx n_{1i}$, s.t.

$$\nabla \cdot (\epsilon_0 \mathbf{E}_1) = e(n_{1e} - n_{1i}) \approx 0$$

so we can ignore the “1” in the dielectric function. Let $\epsilon = 0$, we get

$$\begin{aligned} \frac{\omega_{pe}^2}{k^2 v_{th,e}^2} - \frac{\omega_{pi}^2}{\omega^2} &= 0 \\ \Rightarrow \omega^2 &= k^2 v_{th,e}^2 \frac{m_e}{m_i}, \quad \frac{\omega}{k} = \sqrt{\frac{k_B T_e}{m_i}} \equiv c_s \end{aligned}$$

Physically, electron sees the electric field created by ions due to ion plasma oscillation. Since electrons move much faster than ions, this electric field is nearly electrostatic for electrons. As a result, electron just follows the ion motion.

There are some other ways to get the ion-acoustic wave. If $\omega \ll \omega_{pe}$, we can treat this wave as electrostatic wave for electron. The distribution for electron number density is

$$\begin{aligned} n_e &= n_0 e^{e\phi/k_B T_e} \approx n_{0e} + n_{1e} \approx n_0 [1 + \frac{e\phi}{k_B T_e} + \dots] \\ \Rightarrow n_{1e} &= n_0 \frac{e\phi}{k_B T_e} = \frac{n_0 e E_1}{-ikm_e v_{th,e}^2} \end{aligned}$$

Note that here the tilde signs \sim of the variables are neglected for convenience without ambiguity.

Do we need to calculate n_{1i} ??? Yes. MORE to do here!

Also, from the linearized equation of motion for electron,

$$\begin{aligned} \frac{\partial v_{e1}}{\partial t} &= -\frac{e}{m_e} E_1 - \frac{\nabla P_1}{n_0 m_e} \\ -i\omega v_{e1} &= -\frac{e}{m_e} E_1 - \frac{ikP_1}{n_0 m_e} = -\frac{e}{m_e} E_1 - \frac{ikn_{1e} k_B T_e}{n_0 m_e} \end{aligned}$$

In the $\omega \rightarrow 0$ limit, $LHS \doteq 0$, we get

$$n_{1e} = \frac{n_0 e E_1}{-ikm_e v_{th,e}^2}$$

Then again we get the dielectric function through Poisson's equation.

10.13 Electrostatic Wave in a Magnetized Plasma

Now we continue to discuss the property of electrostatic waves with background magnetic field.

First let us introduce a useful result for continuity equation. Assuming $n_0 = n_0(\mathbf{x})$, $\mathbf{v}_0 = \mathbf{v}_0(\mathbf{x})$ in equilibrium, $\mathbf{x}_1 = \mathbf{x}_1(\mathbf{x}, t)$ is the perturbation in displacement. We can show that the linearized continuity equation has an equivalent form:

$$\frac{\partial n_1(\mathbf{x}, t)}{\partial t} + \nabla \cdot [n_0(\mathbf{x})\mathbf{v}_1(\mathbf{x}, t) + n_1(\mathbf{x}, t)\mathbf{v}_0(\mathbf{x}, t)] = 0 \Leftrightarrow n_1(\mathbf{x}, t) = -\nabla \cdot [n_0(\mathbf{x})\mathbf{x}_1(\mathbf{x}, t)]$$

The proof is related to mass conservation shown as follows. Intuitively, you can think of this as the degree of condensation only depends on displacement, not on how you get there (speed).

In 1D,

$$\begin{aligned} n_0 &= n_0(x_0) = \text{unperturbed density} \\ x &= x_0 + x_1(x_0, t) = \text{instantaneous position} \\ n(x_0, t) &= n_0(x_0) + n_1(x_0, t) = \text{total density} \end{aligned}$$

At time t , $[x_0, x_0 + dx] \rightarrow [x_0 + x_1(x_0, t), x_0 + dx + x_1(x_0 + dx, t)]$. Due to mass conservation, we have (This looks like the derivation of deformation in fluid dynamics.)

$$\begin{aligned} n_0(x_0)dx &= [n_0(x_0 + x_1) + n_1(x_0 + x_1, t)] \cdot [x_0 + dx + x_1(x_0 + dx, t) - (x_0 + x_1(x_0, t))] \\ &\approx \left[n_0(x_0) + x_1 \frac{\partial n_0(x_0)}{\partial x_0} + n_1(x_0, t) \right] dx \left[1 + \frac{\partial x_1(x_0, t)}{\partial x_0} \right] \\ &\Rightarrow n_0(x_0) = n_0(x_0) + n_0(x_0) \frac{\partial x_1(x_0, t)}{\partial x_0} + x_1 \frac{\partial n_0(x_0)}{\partial x_0} + n_1(x_0, t) \\ &\Rightarrow n_1(x_0, t) = -n_0(x_0) \frac{\partial x_1(x_0, t)}{\partial x_0} - x_1 \frac{\partial n_0}{\partial x_0} = -\frac{\partial}{\partial x_0} [n_0(x_0)x_1(x_0, t)] \\ &\Rightarrow n_1(x, t) = -\frac{\partial}{\partial x} [n_0(x)x_1(x, t)] \end{aligned}$$

The simplest equilibrium state in a constant magnetized plasma is

$$\begin{aligned} n_{i0} &= n_{e0} = n_0, \quad \mathbf{E}_0 = 0, \quad \mathbf{B}_0 = B_0 \hat{z} \\ \mathbf{v}_{e0} &= \mathbf{v}_{i0} = 0, \quad T_e = 0, \quad T_i = 0, \quad m_i = \infty \end{aligned}$$

Now introduce an electrostatic perturbation ($\mathbf{E}_1 = -\nabla\phi_1$)

$$\mathbf{E}_1 = \tilde{\mathbf{E}}_1 e^{-i\omega t + i\mathbf{k} \cdot \mathbf{x}} = -i\mathbf{k}\tilde{\phi}_1 e^{-i\omega t + i\mathbf{k} \cdot \mathbf{x}}$$

we can confirm that this is indeed an electrostatic perturbation since $\mathbf{E}_1 \parallel \mathbf{k}$ and $\mathbf{B}_1 = 0$.

Case 1: $\mathbf{k} = k_z \hat{z} \parallel \mathbf{B}_0$, i.e. parallel propagation. Then $\mathbf{E}_1 = \hat{z} E_{1z} e^{-i\omega t + ik_z z}$. This is the same as if there is no magnetic field, so the dielectric function is

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2}$$

Case 2: $\mathbf{k} \perp \mathbf{B}_0$, i.e. perpendicular propagation. Without loss of generality, let $\mathbf{k} = k_x \hat{x}$. Then

$$\mathbf{E}_1 = \hat{x} \tilde{E}_{1x} e^{-i\omega t + ik_x x}$$

The equations of motion (cold plasma) in the x and y direction are

$$\begin{aligned} \ddot{x}_1 &= -\frac{e}{m_e} [E_{1x} + \dot{y}_1 B_0] \\ \ddot{y}_1 &= -\frac{e}{m_e} [-\dot{x}_1 B_0] \\ \Rightarrow \dot{y}_1 &= \frac{e B_0}{m_e} x_1 = |\Omega_e| x_1 \\ \ddot{x}_1 &= -\frac{e}{m_e} [E_{1x} + |\Omega_e| \cdot y_1 B_0] = -\frac{e}{m_e} E_{1x} - \Omega_e^2 x_1, \\ \Rightarrow x_1 &= \frac{-\frac{e}{m_e} E_{1x}}{-\omega^2 + \Omega_e^2} \end{aligned}$$

Then we have the perturbed density in response to the perturbed electric field E_{1x} :

$$n_1 = -n_0 \nabla \cdot \mathbf{x}_1 = -n_0 i k_x x_1 = -n_0 i k_x \frac{-\frac{e}{m_e} E_{1x}}{-\omega^2 + \Omega_e^2}$$

From Poisson's equation, we get the dielectric function (the same method as before):

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2 - \Omega_e^2}$$

Let $\epsilon = 0$, we have

$$\omega = \sqrt{\omega_{pe}^2 + \Omega_e^2} \equiv \omega_{UH}$$

which is called the *upper hybrid* frequency. This is the highest characteristic frequency in plasma. This upper hybrid wave is a havoc to some beam generator devices as it appears near the electron collector.

What if ions are included? Similar to previous derivations and notice that we are still within the range of linear theory, we have

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2 - \Omega_e^2} - \frac{\omega_{pi}^2}{\omega^2 - \Omega_i^2}$$

For $\Omega_i \ll \omega \ll \Omega_e$ with quasi-neutrality condition $n_{1e} \approx n_{1i}$, we can have a simplified dispersion relation by letting $\epsilon = 0$:

$$\begin{aligned} \frac{\omega_{pe}^2}{\Omega_e^2} &= \frac{\omega_{pi}^2}{\omega^2} \\ \Rightarrow \omega &= \sqrt{\omega_{pi}^2 \frac{\Omega_e^2}{\omega_{pe}^2}} = \sqrt{|\Omega_e \Omega_i|} \equiv \omega_{LH} \end{aligned}$$

which gives us the *low hybrid wave* frequency. It equals to the geometric mean of the two cyclotron frequencies. Actually, this can be obtained from pure plasma motion argument. Recall that $\Omega_i \ll \omega \ll \Omega_e$ means that for electrons the plasma seems to be nonmagnetized, so they moves only under the electric field,

$$v_{1ex} = \frac{-eE_{1x}}{m_e(-i\omega)}$$

On the other hand, for ions the magnetic field is strong while electric field still exists, so it experiences polarization drift along the direction of perturbed electric field,

$$v_{1ix} = \frac{1}{\Omega_i} \frac{\partial}{\partial t} \left(\frac{E_{1x}}{B_0} \right)$$

Under quasi-neutrality condition, $v_{1ix} = v_{1ex}$, so we have

$$\frac{-eE_{1x}}{m_e(-i\omega)} = \frac{1}{\Omega_i} \frac{\partial}{\partial t} \left(\frac{E_{1x}}{B_0} \right) \Rightarrow \omega = \sqrt{|\Omega_e \Omega_i|}$$

If we consider warm plasma for 1D, there is an additional pressure term in the momentum equation. The continuity equation together with Poisson's equation give the relation of perturbed displacement and electric field:

$$\begin{aligned} n_1 &= -\nabla \cdot (n_0 \mathbf{x}_1) = -ik_x n_0 x_1 \\ \nabla \cdot (\epsilon_0 \mathbf{E}_1) &= ik_x \epsilon_0 E_1 = -en_{1e} = eik_x n_0 x_1 \\ \Rightarrow E_1 &= \frac{en_0 x_1}{\epsilon_0} \end{aligned}$$

Substituting into the momentum equation, we get

$$\begin{aligned}\ddot{x}_1 + \Omega_e^2 x_1 &= -\frac{e}{m_e} E_1 - \frac{\nabla P_1}{n_0 m_e} \\ &= -\omega_{pe}^2 x_1 - k_x^2 v_{th,e}^2 x_1\end{aligned}$$

where $v_{th,e} = \sqrt{k_B T_e / m_e}$. This gives us (You can gain a sense of the equivalent force law from the dispersion relation.)

$$\omega^2 = \omega_{pe}^2 + \Omega_e^2 + k_x^2 v_{th,e}^2$$

This is also equivalent to the dielectric function

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{\omega^2 - \Omega_e^2 - k_x^2 v_{th,e}^2}$$

10.14 CMA Diagram

The Clemmow-Mullaly-Allis (CMA) diagram classifies all EM + ES waves (including ions) in a cold magnetized plasma. However it is no longer useful in hot plasma waves.

- Resonances and cutoffs
- Mode conversions

10.15 Wave-Particle Interactions

Different wavemodes exchange energy with particles by different mechanisms including

1. Landau resonance which heats in the direction parallel to the magnetic field and can give a parallel beam in the velocity distribution function (VDF) (Figure 10.4 left). Charged particles exchange energy with E_{\parallel} .
2. Cyclotron resonance which heats in the direction perpendicular to the magnetic field and broadens the velocity distribution function in the perpendicular direction (Figure 10.4 middle). Charged particles exchange energy with E_{\perp} . Left-hand polarized ion cyclotron waves can lead to ion heating in the perpendicular direction.
3. Pitch angle scattering which can lead to the formation of a plateau in the VDF (Figure 10.4 right).

The resonance parameters for the Landau and cyclotron resonances are

$$\xi_L = \frac{\omega}{k_{\parallel} v_{th,s}} \quad (10.92)$$

$$\xi_C = \frac{\omega - \Omega_s}{k_{\parallel} v_{th,s}} \quad (10.93)$$

Here the frequency ω is measured in the plasma rest frame. Ω_s denotes the cyclotron frequency of particle species s (ion species and electrons), k_{\parallel} the parallel component of the wavevector, and $v_{th,s}$ the particle thermal speed of species s . In general, the resonance parameter can be defined for arbitrary harmonics of the cyclotron frequency $m = 0, \pm 1, \pm 2, \dots$:

$$\xi^{(m)} = \frac{\omega - m\Omega_s}{k_{\parallel} v_{th,s}} \quad (10.94)$$

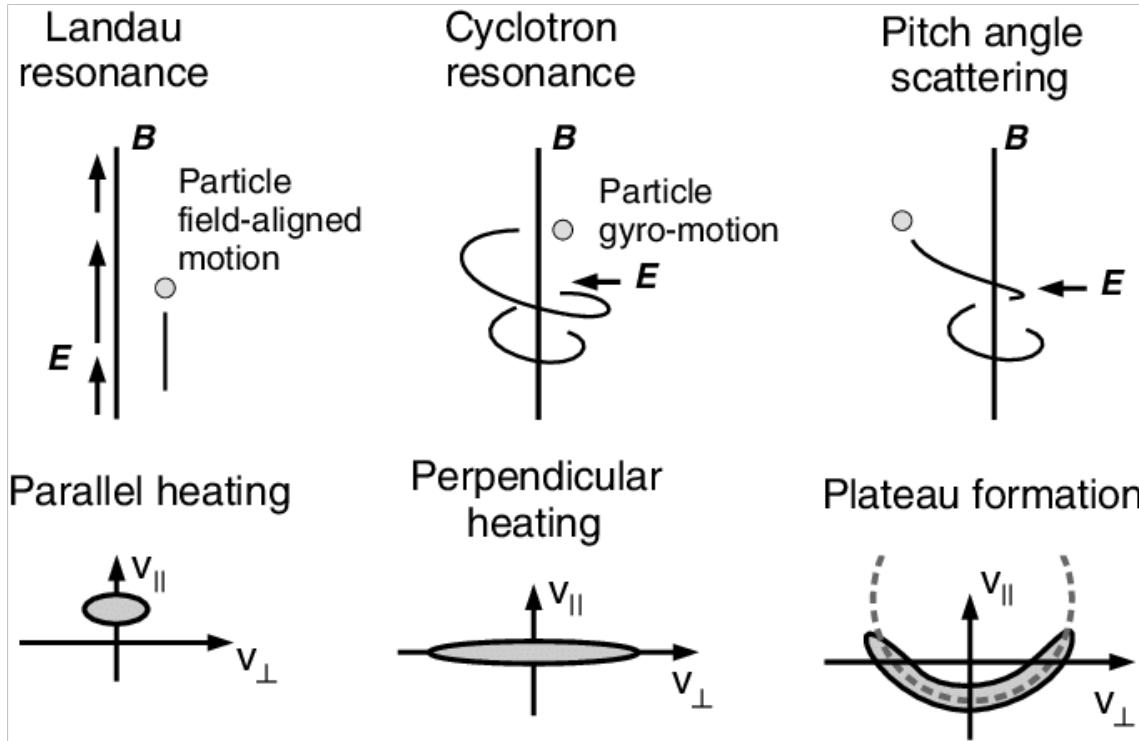


Figure 10.4: Wave-particle interactions and the associated part of the velocity distribution functions. Courtesy of Yasuhito Narita.

Note that the resonance parameters above are defined for a Maxwellian plasma. A correction is needed when treating a non-Maxwellian plasma to find the suitable velocity-space gradient

for the resonance. The resonance is efficient when the parameter ξ_L or ξ_C is on the order of unity. Strictly speaking, the wave damping (or particle acceleration) is most efficient, typically for $1 < \xi < 5$. The upper limit is not exact, but the resonance becomes gradually inefficient at larger values of ξ . For $\xi < 1$ the particle motion is slower than the wave propagation and the particles do not have a sufficient time for exchanging energy with the wave electric field. For $\xi > 5$ there are increasingly fewer particles with higher velocities for the resonance (higher than the thermal speed).

10.15.1 Pitch Angle Scattering

Charged particles can be scattered by the wave electric and magnetic fields incoherently, and the scattering deforms the velocity distribution function along the co-centric contours centered at the wave phase speed (Figure 10.4, right panel). The reason for the deformation is that the particle kinetic energy K_{wv} (per unit mass) does not change in the co-moving frame with the apparent wave phase speed in the parallel direction to the mean magnetic field:

$$K_{wv} = \frac{1}{2} \left[v_\perp^2 + \left(v_\parallel - \frac{\omega}{k_\parallel} \right)^2 \right] = \text{const.}$$

The co-centric deformation of the distribution function achieves a quasi-linear equilibrium in that the velocity-space gradient becomes zero (plateau formation) in the pitch angle directions. For example, in the solar wind ions are found to be resonating with obliquely propagating Alfvén/ion cyclotron waves. Note that the relevant phase speed is ω/k_\parallel , and is different from the true phase speed ω/k . The perpendicular component of the wavevector k_\perp does not play a role in pitch angle scattering.

This process is well explained by [Plasma Zoo: Gyroresonant Scattering](#).

10.16 Whistler Wave

There exist different terminologies for whistler waves in the space and laboratory communities. In space science, a “whistle” is specifically defined as an electromagnetic wave excited by lightning and dispersed while propagating through the ionosphere and magnetosphere. All other excitation mechanisms lead to *whistler-mode wave*. Depending on their sound and spectrograms they are given many exotic names such as hiss, roar, saucers, chorus, risers, hooks, triggered emissions, etc, which are essentially the same thing with slightly different behaviors. In laboratory plasma physics any wave propagating in the whistler mode is simply called a *whistler wave*. However, in bounded laboratory plasmas, in particular solid state plasmas, these waves are often called *helicons*.

Whistler waves are most often referring to electron whistler waves, also known as electron cyclotron waves. Electron whistler waves are VLF emissions observed in the frequency range between the lower hybrid resonance frequency ω_{LH} (Equation 10.28) and electron gyrofrequency $\Omega_e = eB/m_e$.⁴ They are a type of electron-scale plasma waves that contribute to electron scattering, acceleration, and energy transport.⁵ They are driven by the electron temperature anisotropy ($T_{e\perp}/T_{e\parallel} > 1$, and the parallel energy of the resonant electrons exceeds the magnetic energy per electron, $\mathcal{E}_\parallel = \frac{1}{2}m_e v_\parallel^2 > B^2/2\mu_0 n_e$, just like EMIC waves are stimulated by the temperature anisotropy of the ions (C. F. Kennel and Petschek 1966).) or by electron beams or heat fluxes (Gary 1993), and shape the electron velocity distribution function through wave-particle interactions.⁶ (Ren+ 2020)

Whistler modes are *dispersive* waves (i.e., group and phase velocities are frequency-dependent), which can propagate oblique to the static magnetic field \mathbf{B}_0 up to a limiting “resonance cone” angle, given for electrons by (Stenzel 1999)

$$\theta_{\text{phase,max}} = \cos^{-1}(\omega/\Omega_e)$$

Near the oblique resonance the group and phase velocities are essentially orthogonal, and the wave energy is dominated by the electric field. For frequencies $\omega < \Omega_e/2$ there exists an oblique mode at $\theta_{\text{phase}} = \cos^{-1}(2\omega/\Omega_e)$ with parallel group velocity called the Gendrin mode [Gendrin 1961]. Along the direction of background magnetic field \mathbf{B}_0 the wave magnetic field is right-hand circularly polarized.⁷ The electric field of oblique whistlers has both space charge and inductive contributions, $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$. At an oblique density discontinuity the reflection of whistlers produces two reflected and one transmitted mode [Lee, 1969]. Refraction of whistlers in density nonuniformities has been studied carefully [Smith, 1961; Helliwell, 1965; Walker, 1976] since it explains the remarkable *ducting* properties of whistlers. Ducting confines the wave energy in a field-aligned density crest or trough and allows oblique whistlers to propagate undamped over long distances along \mathbf{B}_0 . All ground observations of whistlers, in particular the multiply reflected whistlers, involve ducting.

Plane whistler waves can decay because of

- electron collisions with neutrals and ions,
- collisionless electron cyclotron damping ($\omega - \Omega_e = \mathbf{k} \cdot \mathbf{v}_e$), and
- Landau damping for oblique modes ($\omega = \mathbf{k} \cdot \mathbf{v}_e$).

and it can grow when the distribution function at the resonant velocity has a positive slope because of

⁴[Tsurutani and Smith, 1974] said most often they are with a minimum wave power at $0.5\Omega_e$. Is is observation or theory?

⁵From Equation 10.24 we see the contribution of electron physics in the dispersion relation.

⁶The derivation for the instability criterion involves kinetic theory, similar to the discussion in Chapter 14.

⁷Note that in plasma physics, the wave polarization is defined with respect to the background magnetic field \mathbf{B}_0 , not the propagation direction \mathbf{k} . This is because the polarity contains physical meaning when you think them together with the cyclotron motion of electrons and ions.

- Cherenkov instabilities
- cyclotron resonance
- Landau resonance

Whistler wave can decay into lower hybrid and kinetic Alfvén waves. Parametric instabilities can trap whistler decay waves along the resonance cone. The stability of whistlers in plasmas with non-Maxwellian distribution functions has been studied theoretically in detail [Lee and Crawford, 1970]. The interaction of Langmuir waves and whistlers can produce the observed filamentary fine structure of solar type IV radio bursts [Fomichev and Fainshtein, 1988].

Note that there are also MHD whistler modes in the range of the ion cyclotron frequency Ω_i [Russell+ 1971; Albert, 1980]. Properties of whistler waves can relate closely to electron-scale physics in magnetic reconnection (e.g. (J. Wang et al. 2023) at Mars, THEMIS observation in Earth's tail reconnection exhaust region, and lower-hybrid instability).

We first see the derivations of whistler wave dispersion relation in Section 10.6 at the low frequency MHD limit. Another way to derive the whistler mode dispersion relation, which is probably easier, is to include the Hall term from the generalized Ohm's law:

$$\mathbf{E} = -\mathbf{U} \times \mathbf{B} + \frac{1}{ne} \mathbf{J} \times \mathbf{B}$$

Using Ampère's law and retaining only the Hall term leads to the equation

$$\dot{\mathbf{B}} = -\frac{1}{\mu_0 ne} \nabla \times [(\nabla \times \mathbf{B}) \times \mathbf{B}]$$

Performing linearization and assuming that the magnetic field is parallel to the z-axis and its perturbation is only in x and y, the last equation becomes

$$\begin{aligned}\omega B_{1x} &= -i \frac{k_z^2 B_0}{\mu_0 ne} B_{1y} \\ \omega B_{1y} &= i \frac{k_z^2 B_0}{\mu_0 ne} B_{1x}\end{aligned}$$

which easily yields

$$\omega = \frac{B_0}{\mu_0 ne} k_z^2$$

This is the $v_A \ll W$ case in Equation 10.25 where $v_w = W$. For a more complete derivation, see [Chen Shi's note](#). The dispersion relations of whistler and ion cyclotron wave make a pair, similar to the relation between fast and slow magnetosonic waves.

The dispersion property of whistler waves makes it a problem for hybrid simulations (Section 30.3).

In situ measurements are sensitive to the entire spectrum of oblique whistlers ($\omega < \Omega_e \cos \theta$), hence most of the observed whistlers are oblique, unducted, magnetospherically reflected waves (Stenzel 1999). ???

10.16.1 Ducting of Whistler Waves

Ducting refers to the confinement of whistler waves within channels or “ducts” formed by variations in plasma density along the magnetic field lines. The density can be modulated by instabilities and thereby modifies the whistler wave propagation and amplitude. Most common are density modifications via the ponderomotive force, thermal pressure, and ionization, which can lead to self-focusing and filamentation instabilities.

- Density Enhancements and Depletions: Whistler waves can become trapped within regions of higher or lower plasma density relative to the surrounding environment. These density variations create what are called ducts.
- Refraction: The speed of a whistler wave depends on the plasma density. As whistler waves encounter changes in density along a magnetic field line, they undergo refraction (bending), similar to how light bends when passing between different materials.
- Confinement: The refraction caused by density gradients can keep the whistler waves confined within the duct. The waves essentially bounce back and forth within the duct as they propagate along the magnetic field line.

Types of Ducts:

1. Density Enhancement Ducts: Whistler waves become trapped in regions of higher plasma density.
2. Density Depletion Ducts: Whistler waves become trapped in regions of lower plasma density.

Significance of Ducting:

- Wave Guidance: Ducting allows whistler waves to travel over much longer distances than they otherwise would in a homogeneous plasma.
- Wave Amplification: Ducting can lead to the amplification of whistler wave intensity.
- Interaction with Particles: Ducted whistler waves can interact with charged particles trapped in the Earth’s radiation belts, leading to particle acceleration, scattering, and loss. This plays a crucial role in the dynamics of the radiation belts.

10.16.2 Whistlers in EMHD

While many wave phenomena are thought of in terms of plane monochromatic eigenmodes, a vast class of whistler phenomena arise in the form of temporal transients and spatially bounded phenomena, i.e., wave packets. Such transient phenomena are common in *electron magnetohydrodynamics (EMHD)* physics (Chapter 11). In contrast to ordinary MHD where the plasma behaves like a single magnetized fluid, in EMHD only the electrons are magnetized

while the ions are not. This arises on timescales $f_{ce}^{-1} < t < f_{ci}^{-1}$ and spatial scales $r_{Le} < r < r_{Li}$. The transition region is the Hall MHD. Refer to Stenzel (1999) for more detailed comparisons.

Specifically for whistlers, the whistler wave can be described by EMHD on timescales of $2\pi/\Omega_{ce} < t < 2\pi/\Omega_{LH}$ and spatial scales of $L < \lambda_i$, where the ions are assumed as a motionless neutralizing background [Zhao 2017]. When magnetic field lines are fully frozen into the electron fluid, the whistler wave follows the dispersion relation $\omega = \Omega_{ce}\lambda_e^2 k^2 \cos\theta$. The dispersion relation becomes $\omega = \Omega_{ce}\lambda_e^2 k^2 \cos\theta/(1 + \lambda_e^2 k^2)$ as the electron inertia is contained in Ohm's law. Jinsong Zhao further extended the model by taking the electron thermal pressure and the displacement current into the EMHD model.

10.16.3 Importance of Whistler

1. They play an important role in both the acceleration and loss of electrons in the radiation belt in the inner magnetosphere (See the series of studies by Thorne, R. M).

Lower energy electrons (\sim keV) generate chorus waves -> relax temperature anisotropy -> waves scatter electrons into the loss cone and lead to electron precipitation into the ionosphere.

2. Chorus wave is the driver of pulsating aurora (Nishimura+ 2010) and diffuse aurora (Thorne+ 2010).
3. Chorus waves may be the source of hiss waves in the plasmasphere. Hiss waves play an important role in scattering electrons into the loss cone in the slot region.

10.17 Wave Identification

Measured wave spectra are complex and opaque, i.e. it is very difficult from an inspection of their shape to identify the wave modes that are present in the plasma volume under investigation. In some rare clean cases one can conclude from the observation of a particular maximum peak in the wave power or the observation of only one single field component which wave has been detected. In the general case of broad spectra or mixed spectral peaks and various field components lacking clear dominance of one field component it becomes nearly impossible to decide about the waves. Clearly, when the spectrum is a shapeless power-law one, in most cases, we are dealing with developed turbulence or localised structures in which case it makes no sense to distinguish and search for single modes. Then one must seek shelter among the well developed methods of analysing turbulence.

On the other hand, if the spectra indicate the presence of single waves, one would like to have some quantities at hand which help identifying which modes one is dealing with. It would be helpful if one could measure simultaneously both, the wave frequency and the wave number spectra. This is possible, however, only with sophisticated multi-spacecraft constellations. And

even then only in the rarest cases the determination of the dispersion relation from experiment will be possible.

In application of these theoretical arguments to real observations one therefore has defined some quantities, called “*transport ratios*”, which have turned out to be quite valuable in helping identifying some of the wave modes. Such transport ratios for electromagnetic waves have been given by (Gary 1993).

- Polarization

The polarization of a wave magnetic field with respect to wave number \mathbf{k} is given by

$$P = ib_s/b_A$$

where b_s, b_A are the components of the magnetic fluctuation field \mathbf{b} in the direction \mathbf{S}, \mathbf{A} of magnetosonic and Alfvén waves, respectively, i.e. the vector $\mathbf{A} = \mathbf{k} \times \mathbf{B}_0$ is perpendicular to the wave vector and the ambient magnetic field, while the vector \mathbf{S} is perpendicular to \mathbf{k} (because of the vanishing divergence $\nabla \cdot \mathbf{b} = \mathbf{k} \cdot \mathbf{b} = 0$) in the plane $(\mathbf{k}, \mathbf{B}_0)$. The waves are more magnetosonic or more Alfvénic whether $|P| > 1$ $|P| < 1$, respectively. For $\Re(P) > 0 (< 0)$ the waves are right-hand (left-hand) polarized.

- Compression

The magnetic compression of the wave measures the relative variation in the parallel magnetic fluctuation field

$$C_B = \langle b_{\parallel}^2 \rangle / \langle |\mathbf{b}|^2 \rangle$$

where the fluctuations are taken at a given pair (ω, \mathbf{k}) . The angular brackets $\langle ab \rangle$ mean taking the real part of the correlation function of the two bracketed quantities.

- Parallel compressibility

This ratio together with the compression ratio provides a tool for estimating how compressive a wave is. It is defined for species s as

$$C_{\parallel s} = \frac{B_0^2}{\langle b_{\parallel}^2 \rangle} \frac{\langle b_{\parallel} \Delta n_s \rangle}{n_s B_0}$$

- Non-coplanarity ratio

This ratio measures the fluctuating field component out of the plane $(\mathbf{k}, \mathbf{B}_0)$, and is given by

$$C_c = \langle b_A^2 \rangle / \langle |\mathbf{b}|^2 \rangle$$

- Alfvén ratio: the ratio of velocity to magnetic fluctuations

Defining $\delta\mathbf{v}_A = \mathbf{b}/\sqrt{\mu_0 m_i n}$ (unit V/B), where n is the total plasma density, the Alfvén ratio is defined as

$$R_{As} = \langle |\delta\mathbf{v}_s|^2 \rangle / \langle |\delta\mathbf{v}_A|^2 \rangle$$

where $\delta\mathbf{v}_s$ is the flow velocity of species s . An Alfvén wave has $\delta\mathbf{v}_i = \pm\mathbf{v}_A \mathbf{b}/\mathbf{B}_0$, and its own Alfvén ratio is $R_{Ai} = 1$. The Alfvén ratio thus measures the fraction of Alfvén waves contained in the near-zero frequency fluctuations.

In another form, it can be written as

$$R_A = \mu_0 n m_i \frac{|\delta v_i|^2}{|\delta \mathbf{B}|^2}$$

One property that differs significantly is the Alfvén ratio where velocity fluctuations dominate for the kinetic slow wave and magnetic fluctuations dominate for the KAW.

- Cross helicity

Helicity of a wave is another identifier of the wave mode, it is in particular useful for determining the direction of propagation of the wave by considering its sign. See Section 18.2.9

10.18 Animation of Waves

Visualization is the best way of understanding physics.

11 EMHD

Electron Magnetohydrodynamics (EMHD) is a specialized branch of plasma physics that focuses on the behavior of plasmas at very small spatial scales and very short timescales, where the motion of electrons becomes dominant over the heavier ions. Since electrons are much lighter and more agile than ions, they can exhibit unique behaviors in the presence of magnetic fields, leading to phenomena not captured by standard magnetohydrodynamics (MHD).

11.1 Key Concepts in EMHD

- Fast Electron Motion: Electrons move much faster than ions due to their lower mass. This translates to a separation of timescales in EMHD; you study rapid electron dynamics while assuming the ions form a relatively stationary background.
- Magnetic Field “Freezing”: In standard MHD, the magnetic field lines are considered to be “frozen” into the plasma. In EMHD, this concept breaks down at small scales as electrons can move independently of the magnetic field lines.
- Whistler Waves: EMHD supports a unique type of plasma wave called a whistler wave. These waves are electromagnetic oscillations that propagate along magnetic field lines, driven by the motion of electrons.
- Hall Effect: Electrons, due to their small mass, are readily deflected by magnetic fields. This deflection, the basis of the Hall effect, leads to the generation of electric currents perpendicular to both the magnetic field and the direction of electron flow. These currents play a role in phenomena like magnetic reconnection.

11.2 Equations of EMHD

The core equations of EMHD differ from standard MHD.

- Generalized Ohm’s Law: Describes electron flow, including the Hall effect and pressure gradient terms absent in ideal MHD Ohm’s Law.
- Momentum Equation for Electrons: Governs electron motion under the influence of electric and magnetic fields, as well as pressure gradients.

- Maxwell's Equations: Describe the evolution of the electric and magnetic fields, just as in standard MHD.

12 Kinetic Theory

12.1 Phase Space

Consider a particle moving in a one-dimensional space and let the position of the particle be $x = x(t)$ and the velocity of the particle be $v = v(t)$. A way to visualize the x and v trajectories simultaneously is to plot these trajectories parametrically on a two-dimensional graph, where the horizontal coordinate is given by $x(t)$ and the vertical coordinate is given by $v(t)$. This x - v plane is called *phase-space*. The trajectory (or orbit) of several particles can be represented as a set of curves in phase-space. Examples of a few qualitatively different phase-space orbits are shown in Figure 12.1.

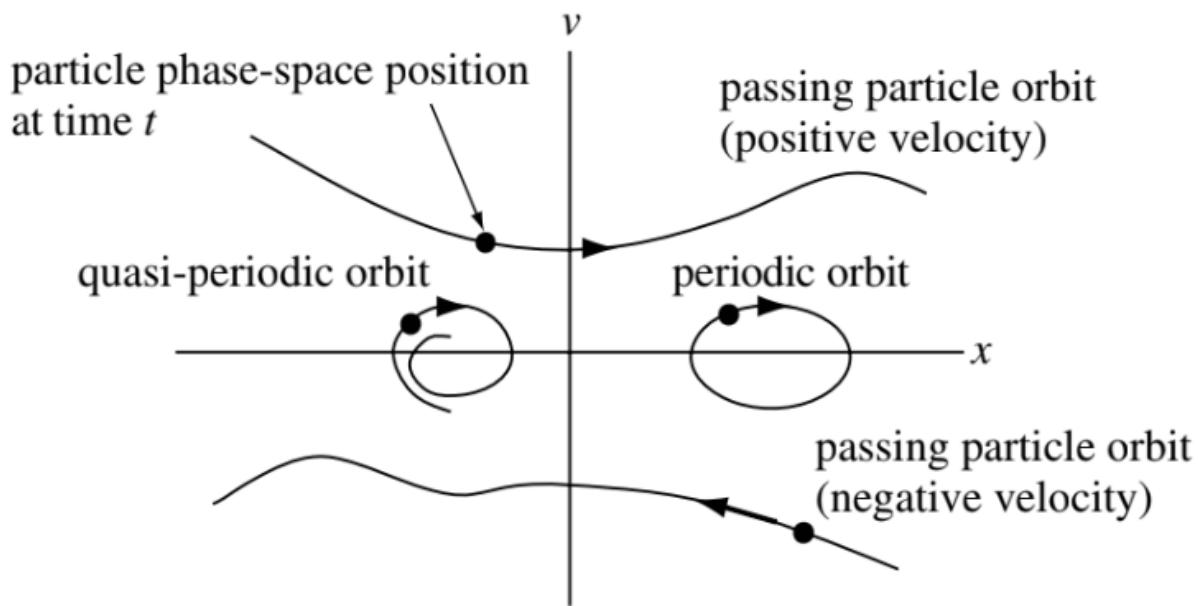


Figure 12.1: Phase-space showing different types of possible particle orbits.

Particles in the upper half-plane always move to the right, since they have a positive velocity, while those in the lower half-plane always move to the left. Particles having exact periodic motion (e.g., $x = A \cos t$, $v = -A \sin t$) alternate between moving to the right and the left and so describe an ellipse in phase-space. Particles with quasi-periodic motions will have near-ellipses or spiral orbits. A particle that does not reverse direction is called a passing

particle, while a particle confined to a certain region of phase-space (e.g., a particle with periodic motion) is called a trapped particle.

12.2 Distribution Function

The fluid theory we have been using so far is the simplest description of a plasma; it is indeed fortunate that this approximation is sufficiently accurate to describe the majority of observed phenomena. There are some phenomena, however, for which a fluid treatment is inadequate. At any given time, each particle has a specific position and velocity. We can therefore characterize the instantaneous configuration of a large number of particles by specifying the density of particles at each point x, v in phase-space. The function prescribing the instantaneous density of particles in phase-space is called the *distribution function* and is denoted by $f(x, v, t)$. Thus, $f(x, v, t)dx dv$ is the number of particles at time t having positions in the range between x and $x + dx$ and velocities in the range between v and $v + dv$. As time progresses, the particle motion and acceleration causes the number of particles in these x and v ranges to change and so f will change. This temporal evolution of f gives a description of the system more detailed than a fluid description, but less detailed than following the trajectory of each individual particle. Using the evolution of f to characterize the system does not keep track of the trajectories of individual particles, but rather characterizes classes of particles having the same x, v .

In fluid theory, the dependent variables are functions of only four independent variables: x, y, z , and t . This is possible because the velocity distribution of each species is assumed to be Maxwellian everywhere and can therefore be uniquely specified by its bulk velocity \vec{U} and the temperature T . Since collisions can be rare in high-temperature plasmas, deviations from thermal equilibrium can be maintained for relatively long times. As an example, consider two velocity distributions $f_1(v_x)$ and $f_2(v_x)$ in a one-dimensional system (Figure 12.2). These two distributions will have entirely different behaviors, but as long as the areas under the curves are the same, fluid theory does not distinguish between them.

When we consider velocity distributions in 3D, we have seven independent variables: $f = f(\mathbf{r}, \mathbf{v}, t)$. By $f(\mathbf{r}, \mathbf{v}, t)$, we mean that the number of particles per meter cubed at position \mathbf{r} and time t with velocity components between v_x and $v_x + dv_x$, v_y and $v_y + dv_y$, and v_z and $v_z + dv_z$ is

$$f(x, y, z, v_x, v_y, v_z, t)dv_x dv_y dv_z$$

12.2.1 Moments of the distribution function

Let us count the particles in the shaded vertical strip in Figure 12.3. The number of particles in this strip is the number of particles lying between x and $x + dx$, where x is the location of the left-hand side of the strip and $x + dx$ is the location of the right-hand side. The number of particles in the strip is equivalently defined as $n(x, t)dx$, where $n(x)$ is the density of particles at x . Thus we see that $\int f(x, v)dv = n(x)$ the transition from a phase-space description (i.e.,

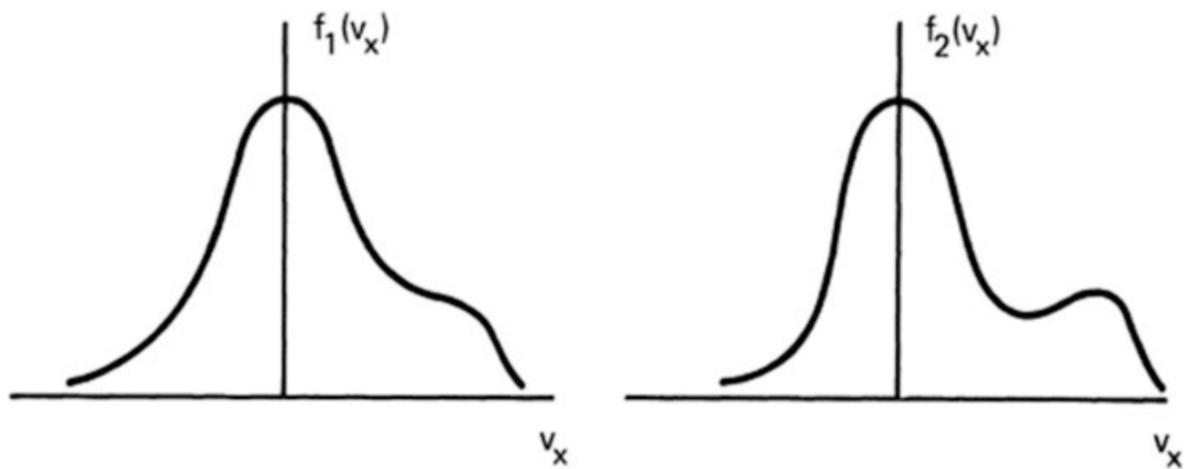


Figure 12.2: Examples of non-Maxwellian distribution functions

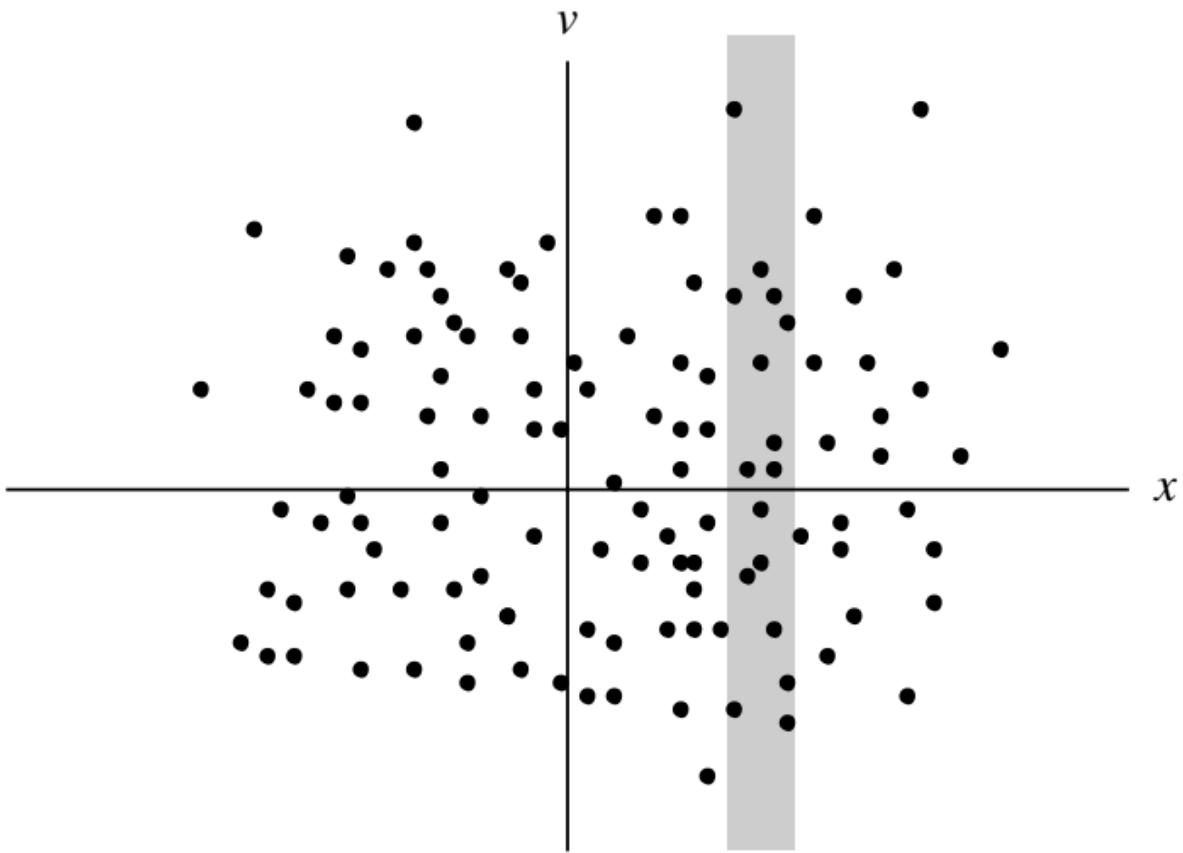


Figure 12.3: Moments give weighted averages of the particles in the shaded vertical strip.

x, v are independent variables) to a conventional description (i.e., only x is an independent variable) involves “integrating out” the velocity dependence to obtain a quantity (e.g., density) depending only on position. Since the number of particles is finite, and since f is a positive quantity, f must vanish as $v \rightarrow \pm\infty$.

In three-dimension, the density is now a function of four scalar variables, $n = n(\mathbf{r}, t)$, which is the integral of the distribution function over the velocity space:

$$\begin{aligned} n(\mathbf{r}, t) &= \int_{-\infty}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z f(\mathbf{r}, \mathbf{v}, t) \\ &= \int_{-\infty}^{\infty} f(\mathbf{r}, \mathbf{v}, t) d^3v \\ &= \int_{-\infty}^{\infty} f(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} \end{aligned} \quad (12.1)$$

Note that $d\mathbf{v}$ is not a vector; it stands for a three-dimensional volume element in velocity space. If f is normalized so that

$$\int_{-\infty}^{\infty} \hat{f}(\mathbf{r}, \mathbf{v}, t) d\mathbf{v} = 1 \quad (12.2)$$

Thus

$$\hat{f}(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{r}, \mathbf{v}, t)/n(\mathbf{r}, t) \quad (12.3)$$

is the probability that a randomly selected particle at position \mathbf{r} has the velocity \mathbf{v} at time t . Using this point of view, we see that averaging over the velocities of all particles at x gives the mean velocity

$$u(\mathbf{x}, t) = \frac{\int \mathbf{v} f(\mathbf{x}, \mathbf{v}, t) d\mathbf{v}}{n(\mathbf{x}, t)} \quad (12.4)$$

Similarly, multiplying \hat{f} by $mv^2/2$ and integrating over velocity will give an expression for the mean kinetic energy of all the particles. This procedure of multiplying f by various powers of \mathbf{v} and then integrating over velocity is called *taking moments of the distribution function*.

Note that \hat{f} is still a function of seven variables, since the shape of the distribution, as well as the density, can change with space and time. From Equation 12.2, it is clear that \hat{f} has the dimensions $(\text{m/s})^{-3}$; and consequently, from Equation 12.3, f has the dimensions $\text{s}^3 \text{m}^{-6}$.

12.2.2 Maxwellian Distribution

A particularly important distribution function is the Maxwellian:

$$\hat{f}_m = \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(-\frac{v^2}{v_{th}^2} \right) \quad (12.5)$$

where

$$v \equiv (v_x^2 + v_y^2 + v_z^2)^{1/2} \quad \text{and} \quad v_{th} \equiv (2k_B T/m)^{1/2}$$

This is the normalized form where \hat{f}_m is equivalent to probability: by using the definite integral

$$\int_{-\infty}^{\infty} \exp(-x^2) dx = \sqrt{\pi}$$

one easily verifies that the integral of \hat{f}_m over $dv_x dv_y dv_z$ is unity.

A common question being asked is:

Why do we see Maxwellian/Gaussian/normal distribution ubiquitously in nature?

Well, this is related to the [central limit theorem](#) in statistics: in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed (e.g. a biased coin which give 95% head and 5% tail). In statistical physics, this is related to the fact that a Maxwellian distribution represents the state of a system with the highest entropy under the constraint of energy conservation.

There are several average velocities of a Maxwellian distribution that are commonly used. The root-mean-square velocity is given by

$$(\bar{v}^2)^{1/2} = (3k_B T/m)^{1/2}$$

The average magnitude of the velocity $|v|$, or simply \bar{v} , is found as follows:

$$\bar{v} = \int_{-\infty}^{\infty} v \hat{f}(\mathbf{v}) d^3v$$

Since \hat{f}_m is isotropic, the integral is most easily done in spherical coordinates in \mathbf{v} space. Since the volume element of each spherical shell is $4\pi v^2 dv$, we have

$$\begin{aligned} \bar{v} &= (m/2\pi k_B T)^{3/2} \int_0^{\infty} v [\exp(-v^2/v_{th}^2)] 4\pi v^2 dv \\ &= (\pi v_{th}^2)^{-3/2} 4\pi v_{th}^4 \int_0^{\infty} [\exp(-y^2)] y^3 dy \end{aligned}$$

The definite integral has a value 1/2, found by integration by parts. Thus

$$\bar{v} = 2\sqrt{\pi} v_{th} = 2(2k_B T/\pi m)^{1/2}$$

The velocity component in a *single direction*, say v_x , has a different average. Of course, \bar{v}_x vanishes for an isotropic distribution; but $|\bar{v}_x|$ does not:

$$|\bar{v}_x| = \int |v_x| \hat{f}_m(\mathbf{v}) d^3 v = \pi^{-1/2} v_{th} = (2k_B T / \pi m)^{1/2}$$

To summarize: for a Maxwellian,

$$\begin{aligned} v_{rms} &= (3k_B T / m)^{1/2} \\ |\bar{v}| &= 2(2k_B T / \pi m)^{1/2} \\ |\bar{v}_x| &= (2k_B T / \pi m)^{1/2} \\ \bar{v}_x &= 0 \end{aligned}$$

For an isotropic distribution like a Maxwellian, we can define another function $g(v)$ which is a function of the scalar magnitude of \mathbf{v} such that

$$\int_0^\infty g(v) dv = \int_{-\infty}^\infty f(\mathbf{v}) dv$$

For a Maxwellian, we see that

$$g(v) = 4\pi n (m/2\pi k_B T)^{3/2} v^2 \exp(-v^2/v_{th}^2) \quad (12.6)$$

(Add figure?) Note the difference between $g(v)$ and a one-dimensional Maxwellian distribution $f(v_x)$. Although $f(v_x)$ is maximum for $v_x = 0$, $g(v)$ is zero for $v = 0$. This is just a consequence of the vanishing of the volume in phase space for $v = 0$. Sometimes $g(v)$ is carelessly denoted by $f(v)$, as distinct from $f(\mathbf{v})$; but $g(v)$ is a different function of its argument than $f(\mathbf{v})$ is of its argument. From Equation 12.6, it is clear that $g(v)$ has dimensions s/m⁴.

(ADD EXAMPLE DISTRIBUTIONS!)

- Isotropic distribution
- Anisotropic (pancake) distribution

$$f(v_\perp, v_\parallel) = \frac{n}{T_\perp T_\parallel^{1/2}} \left(\frac{m}{2\pi k_B} \right)^{3/2} \exp \left(-\frac{mv_\perp^2}{2k_B T_\perp} - \frac{m(v_\parallel - v_{0\parallel})^2}{2k_B T_\parallel} \right)$$

- Beam distribution
- Crescent shape distribution

It is often convenient to present the distribution function as a function of energy instead of velocity. If all energy is kinetic, the energy is simply obtained from $W = mv^2/2$. In the case the particles are in the external electric potential field $U = -q\varphi$ the total energy of particles is $W = mv^2/2 + U$ and the Maxwellian distribution is

$$f(v) = n \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(- \frac{W}{k_B T} \right)$$

This can be written as the energy distribution (???:)

$$g(W) = 4\pi \left[\frac{2(W-U)}{m^3} \right]^{1/2} f(v)$$

The normalization factor is determined by requiring that the integration of the energy distribution over all energies gives the density.

Velocity and energy distribution functions cannot be measured directly. Instead, the observed quantity is *the particle flux to the detector*. Particle flux is defined as the number density of particles multiplied by the velocity component normal to the surface. We define the *differential flux* of particles traversing a unit area per unit time, unit solid angle (in spherical coordinates the differential solid angle is $d\Omega = \sin \theta d\theta d\phi$) and unit energy as $J(W, \Omega, \alpha, \mathbf{r}, t)$. (α is species?) The units of J are normally given as $(\text{m}^2 \text{sr s eV})^{-1}$. Note that in literature cm is often used instead of m and, depending on the actual energy range considered, electron volts are often replaced by keV, MeV, or GeV.

Let us next define how differential particle flux and distribution function are related to each other. We can write the number density in a differential velocity element (in spherical coordinates $d^3v = v^2 dv d\Omega$) as $(dn = f(\alpha, \mathbf{r}, t) v^2 dv d\Omega)$. By multiplying this with v we obtain another expression for the differential flux $f(\alpha, \mathbf{r}, t) v^3 dv d\Omega$. Comparing with our earlier definition of the differential flux we obtain

$$J(W, \Omega, \alpha, \mathbf{r}, t) dW d\Omega = f(\alpha, \mathbf{r}, t) v^3 dv d\Omega$$

Since $dW = mv dv$ we can write the relationship between the differential flux and the distribution function as

$$J(W, \Omega, \alpha, \mathbf{r}, t) = \frac{v^2}{m} f \tag{12.7}$$

One application of the differential flux is the *particle precipitation flux*. With the idea of loss lone, we have a cone of particles that moves along the field lines and can propagate down to the ionosphere, and each shell from v to $v + dv$ corresponds to a specific energy range. This is something we can measure close to the ground and use to infer the plasma properties in the magnetosphere.

12.2.3 Kappa Distribution

The Maxwellian distribution is probably the most studied one theoretically, but may not be the most commonly observed distribution in a collisionless space plasma system. In recent years, another distribution named **Kappa distribution** has gained more attention.

Distribution functions are often nearly Maxwellian at low energies, but they decrease more slowly at high energies. At higher energies the distribution is described better by a power law than by an exponential decay of the Maxwell distribution. Such a behavior is not surprising if we remember that the Coulomb collision frequency decreases with increasing temperature as $T^{-3/2}$ (Equation 9.23). Hence, it takes longer time for fast particles to reach Maxwellian distribution than for slow particles. The kappa distribution has the form

$$f_\kappa(W) = n \left(\frac{m}{2\pi\kappa W_0} \right)^{3/2} \frac{\Gamma(\kappa + 1)}{\Gamma(\kappa - 1/2)} \left(1 + \frac{W}{\kappa W_0} \right)^{-(\kappa+1)}$$

Here W_0 is the energy at the peak of the particle flux and Γ is the gamma function. When $\kappa \gg 1$ the kappa distribution approaches the Maxwellian distribution. When κ is smaller but > 1 the distribution has a high-energy tail. A thorough review is given by (Livadiotis and McComas 2013).

12.2.4 Entropy of a distribution

Collisions cause the distribution function to tend towards a simple final state characterized by having the maximum entropy for the given constraints (e.g. fixed total energy). To see this, we provide a brief discussion of entropy and show how it relates to a distribution function.

Suppose we throw two dice, labeled A and B , and let R denote the result of a throw. Thus R ranges from 2 through 12. The complete set of (A,B) combinations that gives these R 's is listed in Table 12.1:

Table 12.1: All possible combinations of rolling two dices.

R	(A, B)
2	(1,1)
3	(1,2),(2,1)
4	(1,3),(2,2),(3,1)
5	(1,4),(2,3),(3,2),(4,1)
6	(1,5),(2,4),(3,3),(4,2),(5,1)
7	(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)
8	(2,6),(3,5),(4,4),(5,3),(6,2)
9	(3,6),(4,5),(5,4),(6,3)
10	(4,6),(5,5),(6,4)

R	(A, B)
11	$(5,6), (6,5)$
12	$(6,6)$

There are six (A, B) pairs that give $R = 7$, but only one pair for $R = 2$ and only one pair for $R = 12$. Thus, there are six microscopic states (distinct (A, B) pairs) corresponding to $R = 7$ but only one microscopic state corresponding to each of $R = 2$ or $R = 12$. Thus, we know more about the microscopic state of the system if $R = 2$ or $R = 12$ than if $R = 7$. We define the entropy S to be the *natural logarithm of the number of microscopic states corresponding to a given macroscopic state*. Thus, for the dice, the entropy would be the natural logarithm of the number of (A, B) pairs that correspond to a given R . The entropy for $R = 2$ or $R = 12$ would be zero since $S = \ln(1) = 0$, while the entropy for $R = 7$ would be $S = \ln(6)$ since there were six different ways of obtaining $R = 7$.

If the dice were to be thrown a statistically large number of times the most likely result for any throw is $R = 7$ this is the macroscopic state with the largest number of microscopic states. Since any of the possible microscopic states is an equally likely outcome, the most likely macroscopic state after a large number of dice throws is the macroscopic state with the highest entropy.

Now consider a situation more closely related to the concept of a distribution function. In order to do this we first pose the following simple problem: suppose we have a pegboard with \mathcal{N} holes, labeled $h_1, h_2, \dots, h_{\mathcal{N}}$ and we also have \mathcal{N} pegs labeled by $p_1, p_2, \dots, p_{\mathcal{N}}$. What is the number of ways of putting all the pegs in all the holes? Starting with hole h_1 , we have a choice of \mathcal{N} different pegs, but when we get to hole h_2 there are now only $\mathcal{N} - 1$ pegs remaining so that there are now only $\mathcal{N} - 1$ choices. Using this argument for subsequent holes, we see there are $\mathcal{N}!$ ways of putting all the pegs in all the holes.

Let us complicate things further. Suppose that we arrange the holes in \mathcal{M} groups. Say group G_1 has the first 10 holes, group G_2 has the next 19 holes, group G_3 has the next 4 holes and so on, up to group \mathcal{M} . We will use f to denote the number holes in a group, thus $f(1) = 10$, $f(2) = 19$, $f(3) = 4$, etc. The number of ways arranging pegs within a group is just the factorial of the number of pegs in the group, e.g., the number of ways of arranging the pegs within group 1 is just $10!$ and so in general the number of ways of arranging the pegs in the j th group is $[f(j)]!!$.

Let us denote C as the number of ways of putting all the pegs in all the groups *without* caring about the internal arrangement within groups. The number of ways of putting the pegs in all the groups caring about the internal arrangements in all the groups is $C \times f(1)! \times f(2)! \times \dots \times f(4)!$, but this is just the number of ways of putting all the pegs in all the holes, i.e.,

$$C \times f(1)! \times f(2)! \times f(3)! \times \dots \times f(4)! = \mathcal{N}!$$

or

$$C = \frac{\mathcal{N}!}{C \times f(1)! \times f(2) \times f(3)! \times \dots \times f(\mathcal{M})!}$$

Now C is just the number of microscopic states corresponding to the macroscopic state of the prescribed grouping $f(1) = 10$, $f(2) = 19$, $f(3) = 4$, etc. so the entropy is just $S = \ln C$ or

$$\begin{aligned} S &= \ln \left(\frac{\mathcal{N}!}{C \times f(1)! \times f(2) \times f(3)! \times \dots \times f(\mathcal{M})!} \right) \\ &= \ln \mathcal{N}! - \ln f(1)! - \ln f(2)! - \dots - \ln f(\mathcal{M})! \end{aligned}$$

Stirling's formula shows that the large-argument asymptotic limit of the factorial function is

$$\lim_{k \rightarrow \infty} \ln k! = k \ln k - k$$

Noting that $f(1) + f(2) + \dots + f(\mathcal{M}) = \mathcal{N}$, the entropy becomes

$$\begin{aligned} S &= \mathcal{N} \ln \mathcal{N} - f(1) \ln f(1) - f(2) \ln f(2) - \dots - f(\mathcal{M}) \ln f(\mathcal{M}) \\ &= \mathcal{N} \ln \mathcal{N} - \sum_{j=1}^{\mathcal{M}} \ln f(j) \end{aligned}$$

The constant $\mathcal{N} \ln \mathcal{N}$ is often dropped, giving

$$S = - \sum_{j=1}^{\mathcal{M}} f(j) \ln f(j)$$

If j is made into a continuous variable, say $j \rightarrow v$ so that $f(v)dv$ is the number of items in the group labeled by v then the entropy can be written as

$$S = - \int dv f(v) \ln f(v)$$

By now, it is obvious that f could be the velocity distribution function, in which case $f(v)dv$ is just the number of particles in the group having velocity between v and $v + dv$. Since the peg groups correspond to different velocity range coordinates, having more dimensions just means having more groups and so for three dimensions the entropy generalizes to

$$S = - \int d\mathbf{v} f(\mathbf{v}) \ln f(\mathbf{v})$$

If the distribution function depends on position as well, this corresponds to still more peg groups, and so a distribution function that depends on both velocity and position will have the entropy

$$S = - \int d\mathbf{x} \int d\mathbf{v} f(\mathbf{x}, \mathbf{v}) \ln f(\mathbf{x}, \mathbf{v}) \quad (12.8)$$

12.2.5 Effect of collisions on entropy

The highest entropy state is the most likely state of the system because the highest entropy state has the highest number of microscopic states corresponding to the macroscopic state. Collisions (or other forms of randomization) will take some initial prescribed microscopic state and scramble the phase-space positions of the particles, thereby transforming the system to a different microscopic state. This new state could in principle be any microscopic state, but is most likely to be a member of the class of microscopic states belonging to the highest entropy macroscopic state. Thus, any randomization process such as collisions will cause the system to evolve towards the macroscopic state having the maximum entropy.

An important shortcoming of this argument is that it neglects any conservation relations that have to be satisfied. To see this, note that the expression for entropy could be maximized if all the particles are put in one group, in which case $C = \mathcal{N}!$ which is the largest possible value for C . Thus, the maximum entropy configuration of \mathcal{N} plasma particles corresponds to all the particles having the same velocity. However, this would assign a specific energy to the system, which would in general differ from the energy of the initial microstate. This maximum entropy state is therefore not accessible in isolated systems, because energy would not be conserved if the system changed from its initial microstate to the maximum entropy state.

Thus, a qualification must be added to the argument. Randomizing processes will scramble the system to attain the state of maximum entropy consistent with any constraints placed on the system. Examples of such constraints would be the requirements that the total system energy and the total number of particles must be conserved. We therefore reformulate the problem as: given an isolated system with \mathcal{N} particles in a fixed volume V and initial average energy per particle $\langle E \rangle$, what is the maximum entropy state consistent with conservation of energy and conservation of number of particles? This is a variational problem because the goal is to maximize S subject to the constraint that both \mathcal{N} and $\mathcal{N} \langle E \rangle$ are fixed. The method of *Lagrange multipliers* can then be used to take into account these constraints. Using this method the variational problem becomes

$$\delta S - \lambda_1 \delta \mathcal{N} - \lambda_2 \delta (\mathcal{N} \langle E \rangle) = 0$$

where λ_1 and λ_2 are as-yet undetermined Lagrange multipliers. The number of particles is

$$\mathcal{N} = V \int f dv$$

The energy of an individual particle is $E = mv^2/2$, where v is the velocity measured in the rest frame of the center of mass of the entire collection of \mathcal{N} particles. Thus, the total kinetic energy of all the particles in this rest frame is

$$\mathcal{N} \langle E \rangle = V \int \frac{mv^2}{2} f(v) dv$$

and so the variational problem becomes

$$\delta \int dv \left(f \ln f - \lambda_1 V f - \lambda_2 V \frac{mv^2}{2} f \right) = 0$$

Incorporating the volume V into the Lagrange multipliers, and factoring out the coefficient δf , this becomes

$$\int dv \delta f \left(1 + \ln f - \lambda_1 - \lambda_2 \frac{mv^2}{2} \right) = 0$$

Since δf is arbitrary, the integrand must vanish, giving

$$\ln f = \lambda_2 \frac{mv^2}{2} - \lambda_1$$

where the “1” has been incorporated into λ_1 .

The maximum entropy distribution function of an isolated, energy and particle conserving system is therefore

$$f = \lambda_1 \exp(-\lambda_2 mv^2/2)$$

which is the Maxwellian distribution function. We will often assume that the plasma is locally Maxwellian so that $\lambda_1 = \lambda_1(\mathbf{x}, t)$, $\lambda_2 = \lambda_2(\mathbf{x}, t)$. We define the temperature to be

$$k_B T_s(\mathbf{x}, t) = \frac{1}{\lambda_2(\mathbf{x}, t)}$$

The normalization factor is set to be

$$\lambda_1(\mathbf{x}, t) = n_s(\mathbf{x}, t) \left(\frac{m_s}{2\pi k_B T_s(\mathbf{x}, t)} \right)^{N/2}$$

where N is the dimensionality (1, 2 or 3) so that $\int f_s(\mathbf{x}, \mathbf{v}, t) d^N \mathbf{v} = n_s(\mathbf{x}, t)$. Because the kinetic energy of individual particles was defined in terms of velocities measured in the rest frame of the center of mass of the complete system of particles, if this center of mass is moving in the lab frame with a velocity \mathbf{u}_s , then in the lab frame the Maxwellian will have the form

$$f_s(\mathbf{x}, \mathbf{v}, t) = n_s \left(\frac{m_s}{2\pi k_B T_s} \right)^{N/2} \exp \left(-\frac{m_s(\mathbf{v} - \mathbf{u}_s)^2}{2k_B T_s} \right) \quad (12.9)$$

Equation 12.9 is equivalent to Equation 12.5 times number density in 3D.

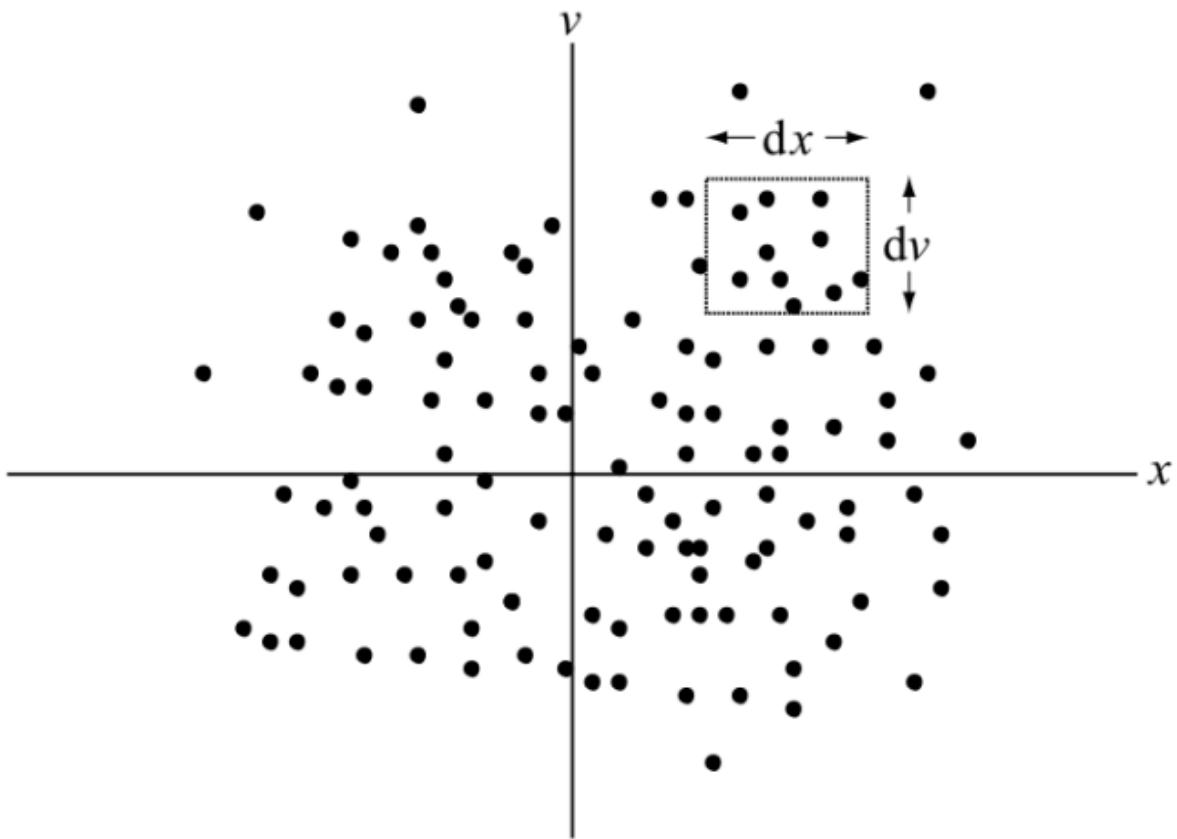


Figure 12.4: A box within phase-space having width dx and height dv .

12.3 Equations of Kinetic Theory

12.3.1 Vlasov equation

Now consider the rate of change of the number of particles inside a small box in phase-space, such as is shown in Figure 12.4. Defining $a(x, v, t)$ to be the acceleration of a particle, it is seen that the particle flux in the horizontal direction is fv and the particle flux in the vertical direction is fa . Thus, the particle fluxes into the four sides of the box are:

1. flux into left side of box is $f(x, v, t)v dv$
2. flux into right side of box is $-f(x + dx, v, t)v dv$
3. flux into bottom of box is $f(x, v, t)a(x, v, t)dx$
4. flux into top of box is $-f(x, v + dv, t)a(x, v + dv, t)dx$

The number of particles in the box is $f(x, v, t)dx dv$ so that the rate of change of the number of particles in the box is

$$\begin{aligned} \frac{\partial f(x, v, t)}{\partial t} dx dv &= -f(x + dx, v, t)v dv + f(x, v, t)v dv \\ &\quad - f(x, v + dv, t)a(x, v + dv, t)dx \\ &\quad + f(x, v, t)a(x, v, t)dx \end{aligned}$$

or, on Taylor expanding the quantities on the right-hand side, we obtain the one-dimensional *Vlasov equation*,

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} + \frac{\partial}{\partial v}(af) = 0 \quad (12.10)$$

It is straightforward to generalize Equation 12.10 to three dimensions and so obtain the three-dimensional Vlasov equation,

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a}f) = 0 \quad (12.11)$$

The symbol ∇ stands, as usual, for the gradient in (x, y, z) space. The symbol $\partial/\partial \mathbf{v}$ or $\nabla_{\mathbf{v}}$ stands for the gradient in velocity space:

$$\nabla_{\mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} = \hat{x} \frac{\partial}{\partial v_x} + \hat{y} \frac{\partial}{\partial v_y} + \hat{z} \frac{\partial}{\partial v_z}$$

Because \mathbf{x}, \mathbf{v} are independent quantities in phase-space, the spatial derivative term has the commutation property,

$$\mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v}f)$$

The particle acceleration is given by the Lorentz force

$$\mathbf{a} = \frac{q}{m}(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

Because $(\mathbf{v} \times \mathbf{B})_i = v_j B_k - v_k B_j$ is independent of v_i , the term $\partial(\mathbf{v} \times \mathbf{B})_i / \partial v_i$ vanishes so that even though the Lorentz acceleration \mathbf{a} is velocity-dependent, it nevertheless commutes with the vector velocity derivative as

$$\mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f)$$

Because of this commutation property the Vlasov equation can also be written as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \quad (12.12)$$

If we “sit on top of” a particle that has a phase-space trajectory $\mathbf{x} = \mathbf{x}(t)$, $\mathbf{v} = \mathbf{v}(t)$ and measure the distribution function as we move along with the particle, the observed rate of change of the distribution function will be $df(\mathbf{x}(t), \mathbf{v}(t), t)/dt$, where the d/dt means that the derivative is measured in the moving frame. Because $d\mathbf{x}/dt = \mathbf{v}$ and $d\mathbf{v}/dt = \mathbf{a}$, this observed rate of change is

$$\left(\frac{df(\mathbf{x}(t), \mathbf{v}(t), t)}{dt} \right)_{\text{orbit}} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0$$

Thus, the distribution function f as measured when moving along a particle trajectory (orbit) is constant. This gives a powerful method for finding solutions to the Vlasov equation. Since f is a constant when measured in a frame following an orbit, we can choose f to depend on any quantity that is constant along the orbit (Jeans 1915, Watson 1956).

For example, if the energy E of particles is constant along their orbits then $f = f(E)$ is a solution to the Vlasov equation. On the other hand, if both the energy and the momentum \mathbf{p} are constant along particle orbits, then any distribution function with the functional dependence $f = f(E, \mathbf{p})$ is a solution to the Vlasov equation. Depending on the situation at hand, the energy and/or momentum may or may not be constant along an orbit and so whether or not $f = f(E, \mathbf{p})$ is a solution to the Vlasov equation depends on the specific problem under consideration. However, there always exists at least one constant of the motion for any trajectory because, just like every human being has an invariant birthday, the initial conditions of a particle trajectory are invariant along this trajectory. As a simple example, consider a situation where there is no electromagnetic field so that $\mathbf{a} = 0$, in which case the particle trajectories are simply $\mathbf{x}(t) = \mathbf{x}_0 + \mathbf{v}_0(t)$, $\mathbf{v}(t) = \mathbf{v}_0$, where $\mathbf{x}_0, \mathbf{v}_0$ are the initial position and velocity. Let us check to see whether $f(\mathbf{x}_0)$ is a solution to the Vlasov equation. By writing $\mathbf{x}_0 = \mathbf{x}(t) - \mathbf{v}_0 t$ so $f(\mathbf{x}_0) = f(\mathbf{x}(t) - \mathbf{v}_0 t)$ we observe that indeed $f = f(\mathbf{x}_0)$ is a solution, since

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = \mathbf{v}_0 \cdot \frac{\partial f}{\partial (\mathbf{x} - \mathbf{x}_0)} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} = \mathbf{v}_0 \cdot \frac{\partial f}{\partial \mathbf{x}} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{x}} = 0$$

12.3.2 Boltzmann equation

It was shown in ??? that the cumulative effect of grazing collisions dominates the cumulative effect of the more infrequently occurring large-angle collisions. In order to see how collisions affect the Vlasov equation, let us now temporarily imagine that the grazing collisions are replaced by an equivalent sequence of abrupt large scattering angle encounters as shown in Figure 12.5. Two particles involved in a collision do not significantly change their positions during the course of a collision, but they do substantially change their velocities. For example, a particle making a head-on collision with an equal mass stationary particle will stop after the collision, while the target particle will assume the velocity of the incident particle. If we draw the detailed phase-space trajectories characterized by a collision between two particles we see that each particle has a sudden change in its vertical coordinate (i.e., velocity) but no change in its horizontal coordinate (i.e., position). The collision-induced velocity jump occurs very fast so that if the phase-space trajectories were recorded with a “movie camera” having insufficient framing rate to catch the details of the jump, the resulting movie would show particles being spontaneously created or annihilated within given volumes of phase-space (e.g., within the boxes shown in Figure 12.5).

The details of these individual jumps in phase-space are complicated and yet of little interest since all we really want to know is the cumulative effect of many collisions. It is therefore both efficient and sufficient to follow the trajectories on the slow time scale while accounting for the apparent “creation” or “annihilation” of particles by inserting a collision operator on the right-hand side of the Vlasov equation. In the example shown here it is seen that when a particle is apparently “created” in one box, another particle must be simultaneously “annihilated” in another box at the same x coordinate but a different v coordinate (of course, what is actually happening is that a single particle is suddenly moving from one box to the other). This coupling of the annihilation and creation rates in different boxes constrains the form of the collision operator. We will not attempt to derive collision operators in this chapter but will simply discuss the constraints on these operators. From a more formal point of view, collisions are characterized by constrained sources and sinks for particles in phase-space and inclusion of collisions in the Vlasov equation causes the Vlasov equation to assume the form

$$\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v} f_s) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f_s) = \sum_{\alpha} C_{s\alpha}(f_s) \quad (12.13)$$

where $C_{s\alpha}(f_s)$ is the rate of change of f_s due to collisions of species s with species α ¹. This is called the *Boltzmann equation*.

The constraints that must be satisfied by the collision operator $C_{s\alpha}(f_s)$ are as follows:

- Conservation of particles - Collisions cannot change the total number of particles at a particular location so

$$\int d\mathbf{v} C_{s\alpha}(f_s) = 0 \quad (12.14)$$

¹For simulations, the effect of numerical diffusion may be treated as one type of “collision”.

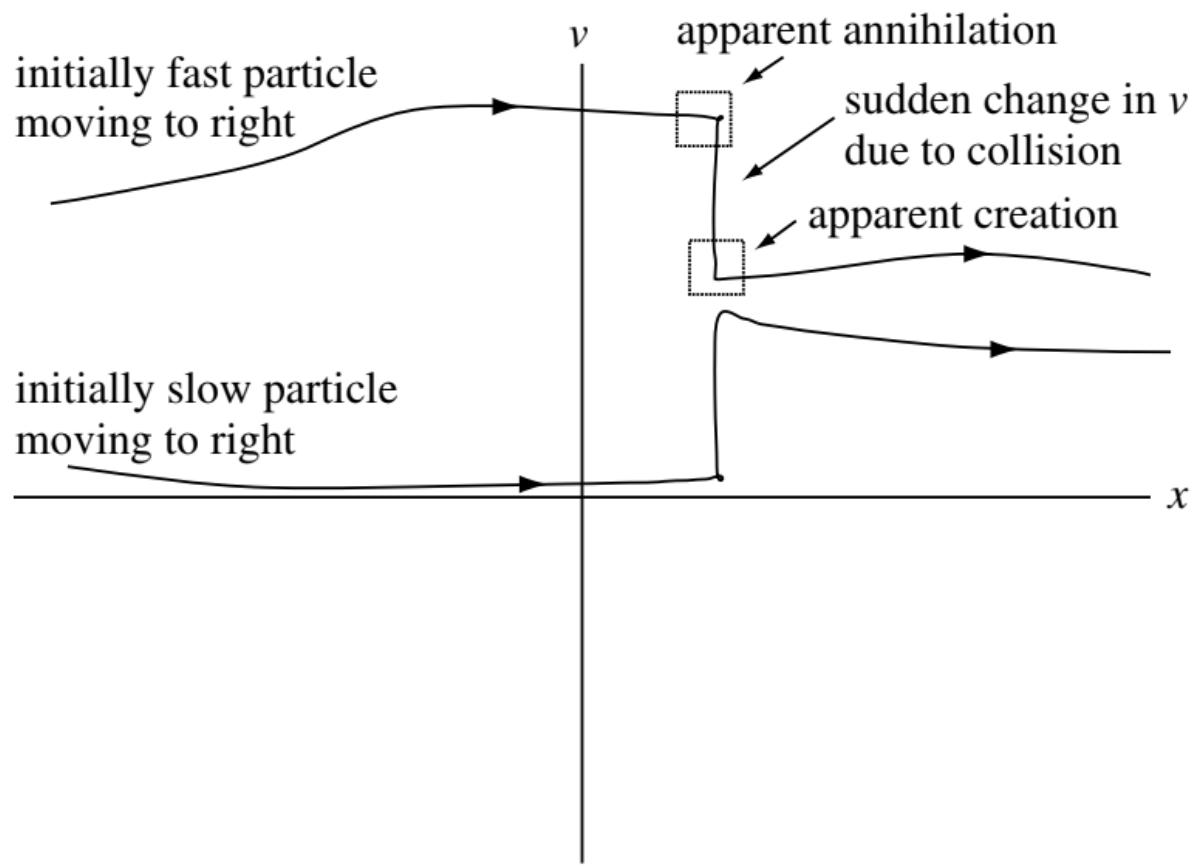


Figure 12.5: Detailed view of collisions causing “jumps” in phase-space.

- Conservation of momentum - Collisions between particles of the same species cannot change the total momentum of that species so

$$\int d\mathbf{v} m_s \mathbf{v} C_{ss}(f_s) = 0 \quad (12.15)$$

while collisions between different species must conserve the total momentum of both species together so

$$\int d\mathbf{v} m_i \mathbf{v} C_{ie}(f_i) + \int d\mathbf{v} m_e \mathbf{v} C_{ei}(f_e) = 0 \quad (12.16)$$

- Conservation of energy - Collisions between particles of the same species cannot change the total energy of that species so

$$\int d\mathbf{v} m_s \mathbf{v}^2 C_{ss}(f_s) = 0 \quad (12.17)$$

while collisions between different species must conserve the total energy of both species together so

$$\int d\mathbf{v} m_i \mathbf{v}^2 C_{ie}(f_i) + \int d\mathbf{v} m_e \mathbf{v}^2 C_{ei}(f_e) = 0 \quad (12.18)$$

In a sufficiently hot plasma, collisions can be neglected. If, furthermore, the force \mathbf{F} is entirely electromagnetic, Equation 12.13 takes the special form of Equation 12.12. Because of its comparative simplicity, this is the equation most commonly studied in kinetic theory. When there are collisions with neutral atoms, the collision term in Equation 12.13 can be approximated by

$$\left(\frac{\partial f}{\partial t} \right)_c = \frac{f_n - f}{\tau}$$

where f_n is the distribution function of the neutral atoms, and τ is a constant collision time. This is called the *Krook collision term*. It is the kinetic generalization of the collision term in Eq. (5.5) in (F. F. Chen 2016). When there are Coulomb collisions, Equation 12.13 can be approximated by

$$\frac{df}{dt} = -\frac{\partial}{\partial \mathbf{v}} \cdot (f \langle \Delta \mathbf{v} \rangle) \frac{1}{2} \frac{\partial^2}{\partial \mathbf{v} \partial \mathbf{v}} : (f \langle \Delta \mathbf{v} \Delta \mathbf{v} \rangle) \quad (12.19)$$

This is called the *Fokker-Planck equation*; it takes into account binary Coulomb collisions only. Here, $\Delta \mathbf{v}$ is the change of velocity in a collision, and Equation 12.19 is a shorthand way of writing a rather complicated expression. The colon operator $\mathbf{ab} : \mathbf{cd} = a_i b_j c_i d_j$.

The fact that df/dt is constant in the absence of collisions means that particles follow the contours of constant f as they move around in phase space. As an example of how these contours can be used, consider the beam-plasma instability of Section ???. In the unperturbed plasma, the electrons all have velocity v_0 , and the contour of constant f is a straight line. The function $f(x, v_x)$ is a wall rising out of the plane of the paper at $v_x = v_0$. The electrons move along the trajectory shown. When a wave develops, the electric field \mathbf{E}_1 causes electrons to

suffer changes in v_x as they stream along. The trajectory then develops a sinusoidal ripple ([?@fig-beam-e-dist-1d B](#)). This ripple travels at the phase velocity, not the particle velocity. Particles stay on the curve as they move relative to the wave. If \mathbf{E}_1 becomes very large as the wave grows, and if there are a few collisions, some electrons will be trapped in the electrostatic potential of the wave. In coordinate space, the wave potential appears as in [?@fig-wave-potential](#). In phase space, $f(x, v_x)$ will have peaks wherever there is a potential trough ([?@fig-contour-dist](#)). Since the contours of f are also electron trajectories, one sees that some electrons move in closed orbits in phase space; these are just the trapped electrons.

Electron trapping is a nonlinear phenomenon which cannot be treated by straightforward solution of the Vlasov equation. However, electron trajectories can be followed on a computer, and the results are often presented in the form of a plot like [?@fig-contour-dist](#).

ADD A TWO STREAM INSTABILITY PHASE ANIMATION!

12.4 Derivation of the Fluid Equations

Instead of just taking moments of the distribution function f itself, moments will now be taken of the entire Vlasov equation to obtain a set of partial differential equations relating the mean quantities $n(\mathbf{x})$, $\mathbf{u}(\mathbf{x})$, etc. We begin by integrating the Vlasov equation, Equation 12.12, over velocity for each species. This first and simplest step in the procedure is called taking the “zeroth” moment, since the operation of multiplying by unity can be considered as multiplying the entire Vlasov equation by \mathbf{v} raised to the power zero. Multiplying the Vlasov equation by unity and then integrating over velocity gives

$$\int \left[\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v} f_s) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f_s) \right] d\mathbf{v} = \sum_{\alpha} \int C_{s\alpha}(f_s) d\mathbf{v}$$

The velocity integral commutes with both the time and space derivatives on the left-hand side because \mathbf{x} , \mathbf{v} and t are independent variables, while the third term on the left-hand side is the volume integral of a divergence in velocity space. Gauss’ theorem (i.e., $\int_V d\mathbf{x} \nabla \cdot \mathbf{Q} = \int_A d\mathbf{s} \cdot \mathbf{Q}$) gives f_s evaluated on a surface at $v = \infty$. However, because $f_s \rightarrow 0$ as $v \rightarrow \infty$, this surface integral in velocity space vanishes. Inserting Equation 12.1, Equation 12.4, and Equation 12.14 into the above, we have the species continuity equation

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{u}_s) = 0 \quad (12.20)$$

Now let us multiply Equation 12.12 by \mathbf{v} and integrate over velocity to take the “first moment”,

$$\int \mathbf{v} \left[\frac{\partial f_s}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (\mathbf{v} f_s) + \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{a} f_s) \right] d\mathbf{v} = \sum_{\alpha} \int \mathbf{v} C_{s\alpha}(f_s) d\mathbf{v}$$

This may be rearranged in a more tractable form by:

1. pulling both the time and space derivatives out of the velocity integral,
2. writing $\mathbf{v} = \mathbf{v}'(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t)$, where $\mathbf{v}'(\mathbf{x}, t)$ is the *random* part of a given velocity, i.e., that part of the velocity that differs from the mean (note that \mathbf{v} is independent of both \mathbf{x} and t but \mathbf{v}' is not; also $d\mathbf{v} = d\mathbf{v}'$ since \mathbf{u} is independent of \mathbf{v}),
3. integrating by parts in 3-D velocity space on the acceleration term and using

$$\partial_i \mathbf{v}_j = \delta_{ij}$$

After performing these manipulations, the first moment of the Vlasov equation becomes

$$\begin{aligned} \frac{\partial(n_s \mathbf{u}_s)}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot \int (\mathbf{v}' \mathbf{v}' + \mathbf{v}' \mathbf{u}_s + \mathbf{u}_s \mathbf{v}' + \mathbf{u}_s \mathbf{u}_s) f_s d\mathbf{v}' \\ - \frac{q_s}{m_s} \int (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s d\mathbf{v}' = -\frac{1}{m_s} \mathbf{R}_{s\alpha} \end{aligned} \quad (12.21)$$

where $\mathbf{R}_{s\alpha}$ is the net frictional drag force due to collisions of species s with species α . Note that $\mathbf{R}_{ss} = 0$, since a species cannot exert a net drag force on itself. The frictional terms have the form

$$\begin{aligned} \mathbf{R}_{ei} &= \nu_{ei} m_e n_e (\mathbf{u}_e - \mathbf{u}_i) \\ \mathbf{R}_{ie} &= \nu_{ie} m_i n_i (\mathbf{u}_i - \mathbf{u}_e) \end{aligned}$$

so that in the ion frame the drag on electrons is simply the total electron momentum $m_e n_e \mathbf{u}_e$ measured in this frame multiplied by the rate ν_{ei} at which this momentum is destroyed by collisions with ions. This form for frictional drag force has the following properties: (i) $\mathbf{R}_{ei} + \mathbf{R}_{ie} = 0$, showing that the plasma cannot exert a frictional drag force on itself, (ii) friction causes the faster species to be slowed down by the slower species, and (iii) there is no friction between species if both have the same mean velocity.

Equation 12.21 can be further simplified by factoring \mathbf{u} out of the velocity integrals and recalling that by definition $\int \mathbf{v}' f_s d\mathbf{v}' = 0$. Thus Equation 12.21 reduces to

$$m_s \left[\frac{\partial(n_s \mathbf{u}_s)}{\partial t} + \frac{\partial}{\partial \mathbf{x}} \cdot (n_s \mathbf{u}_s \mathbf{u}_s) \right] = n_s q_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) - \frac{\partial}{\partial \mathbf{x}} \cdot \vec{P}_s - \mathbf{R}_{s\alpha}$$

where the *pressure tensor* is defined by

$$\vec{P}_s \equiv m_s \int \mathbf{v}' \mathbf{v}' f_s d\mathbf{v}' \quad (12.22)$$

If f_s is an isotropic function of \mathbf{v}' , then the off-diagonal terms in \vec{P}_s vanish and the three diagonal terms are identical. In this case, it is useful to define the diagonal terms to be the *scalar pressure* p_s , i.e.,

$$\begin{aligned} p_s &= m_s \int v'_x v'_x f_s d\mathbf{v}' = m_s \int v'_y v'_y f_s d\mathbf{v}' = m_s \int v'_z v'_z f_s d\mathbf{v}' \\ &= \frac{m_s}{3} \int \mathbf{v}' \cdot \mathbf{v}' f_s d\mathbf{v}' \end{aligned} \quad (12.23)$$

Equation 12.23 defines pressure for a three-dimensional isotropic system. However, we will often deal with systems of reduced dimensionality, i.e., systems with just one or two dimensions. Equation 12.23 can therefore be generalized to these other cases by introducing the general N-dimensional definition for scalar pressure

$$p_s = \frac{m_s}{N} \int \mathbf{v}' \cdot \mathbf{v}' f_s d^N \mathbf{v}' \quad (12.24)$$

where \mathbf{v}' is the N-dimensional random velocity.

The scalar pressure has a simple relation to the generalized Maxwellian as seen by recasting Equation 12.24 as

$$\begin{aligned} p_s &= \frac{m_s}{N} \int \mathbf{v}' \cdot \mathbf{v}' f_s d^N \mathbf{v}' \\ &= \frac{n_s m_s}{N} \left(\frac{m_s}{2\pi k_B T_s} \right)^{N/2} \int v'^2 \exp \left(-\frac{m_s v'^2}{2k_B T_s} \right) d^N \mathbf{v}' \\ &= -\frac{n_s m_s}{N} \left(\frac{\alpha}{\pi} \right)^{N/2} \frac{d}{d\alpha} \int e^{-\alpha v'^2} d^N \mathbf{v}', \quad \alpha \equiv m_s / 2k_B T_s \\ &= -\frac{n_s m_s}{N} \left(\frac{\alpha}{\pi} \right)^{N/2} \frac{d}{d\alpha} \left(\frac{\alpha}{\pi} \right)^{-N/2} \\ &= n_s k_B T_s \end{aligned}$$

which is just the ideal gas law. Thus, the definitions that have been proposed for pressure and temperature are consistent with everyday notions for these quantities.

It is important to emphasize that assuming isotropy is done largely for mathematical convenience and that in real systems the distribution function is often quite anisotropic. Collisions, being randomizing, drive the distribution function towards isotropy, while competing processes simultaneously drive it towards anisotropy. Thus, each situation must be considered individually in order to determine whether there is sufficient collisionality to make f isotropic. Because fully ionized hot plasmas often have insufficient collisions to make f isotropic, the often-used assumption of isotropy is an oversimplification, which may or may not be acceptable depending on the phenomenon under consideration.

On expanding the derivatives on the left-hand side of Equation 12.21, it is seen that two of the terms combine to give \mathbf{u} times Equation 12.20. After removing this embedded continuity equation, E@eq-vlasov-1st-moment reduces to

$$n_s m_s \frac{d\mathbf{u}_s}{dt} = n_s q_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) - \nabla p_s - \mathbf{R}_{s\alpha}$$

where the operator d/dt is the convective derivative (Equation 3.2).

At this point in the procedure it becomes evident that a certain pattern recurs for each successive moment of the Vlasov equation. When we took the zeroth moment, an equation

for the density $\int f d\mathbf{v}$ resulted, but this also introduced a term involving the next higher moment, namely the mean velocity $\sim \int \mathbf{v} f d\mathbf{v}$. Then, when we took the first moment to get an equation for the velocity, an equation was obtained containing a term involving the next higher moment, namely the pressure $\sim \int \mathbf{v} \mathbf{v} f d\mathbf{v}$. Thus, each time we take a moment of the Vlasov equation, an equation for the moment we want is obtained, but because of the $\mathbf{v} \cdot \nabla f$ term in the Vlasov equation, a next higher moment also appears. Thus, moment-taking never leads to a closed system of equations; there will always be a “loose end”, a highest moment for which there is no determining equation. Some sort of ad hoc closure procedure must always be invoked to terminate this chain (as seen below, typical closures involve invoking adiabatic or isothermal assumptions). Another feature of taking moments is that each higher moment embeds terms that contain complete lower moment equations multiplied by some factor. Algebraic manipulation can identify these lower moment equations and eliminate them to give a simplified higher moment equation.

Let us now take the second moment of the Vlasov equation. Unlike the zeroth and first moments, the dimensionality of the system now enters explicitly so the more general pressure definition given by Equation 12.24 will be used. Multiplying the Vlasov equation by $m_s v^2 / 2$ and integrating over velocity gives

$$\left\{ \begin{array}{l} \frac{\partial}{\partial t} \int \frac{m_s v^2}{2} f_s d^N \mathbf{v} \\ + \frac{\partial}{\partial \mathbf{x}} \cdot \int \frac{m_s v^2}{2} \mathbf{v} f_s d^N \mathbf{v} \\ + q_s \int \frac{v^2}{2} \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s d^N \mathbf{v} \end{array} \right\} = \sum_{\alpha} \int m_s \frac{v^2}{2} C_{s\alpha} f_s d^N \mathbf{v} \quad (12.25)$$

We consider each term of this equation separately as follows:

1. The time derivative term becomes

$$\frac{\partial}{\partial t} \int \frac{m_s v^2}{2} f_s d^N \mathbf{v} = \frac{\partial}{\partial t} \int \frac{m_s (\mathbf{v}' + \mathbf{u}_s)^2}{2} f_s d^N \mathbf{v}' = \frac{\partial}{\partial t} \left(\frac{N p_s}{2} + \frac{m_s n_s u_s^2}{2} \right)$$

2. Again using $\mathbf{v} = \mathbf{v}' + \mathbf{u}_s$, the space derivative term becomes

$$\frac{\partial}{\partial \mathbf{x}} \cdot \int \frac{m_s v^2}{2} \mathbf{v} f_s d^N \mathbf{v} = \nabla \cdot \left(\mathbf{Q}_s + \frac{2+N}{2} p_s \mathbf{u}_s + \frac{m_s n_s u_s^2}{2} \mathbf{u}_s \right)$$

where

$$\mathbf{Q}_s = \int \frac{m_s v'^2}{2} \mathbf{v}' f_s d^N \mathbf{v}$$

is called the *heat flux*.

3. On integrating by parts, the acceleration term becomes

$$q_s \int \frac{v^2}{2} \frac{\partial}{\partial \mathbf{v}} \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_s d^N \mathbf{v} = -q_s \int \mathbf{v} \cdot \mathbf{E} f_s d\mathbf{v} = -q_s n_s \mathbf{u}_s \cdot \mathbf{E}$$

4. The collision term becomes (using Equation 12.17)

$$\sum_{\alpha} \int m_s \frac{v^2}{2} C_{s\alpha} f_s d^N \mathbf{v} = \int_{s \neq \alpha} m_s \frac{v^2}{2} C_{s\alpha} f_s d\mathbf{v} = - \left(\frac{\partial W}{\partial t} \right)_{Es}$$

where $(\partial W / \partial t)_{Es}$ is the rate at which species s collisionally transfers energy to species α .

Combining the above four relations, Equation 12.25 becomes

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{N p_s}{2} + \frac{m_s n_s u_s^2}{2} \right) + \nabla \cdot \left(\mathbf{Q}_s + \frac{2+N}{2} p_s \mathbf{u}_s + \frac{m_s n_s u_s^2}{2} \mathbf{u}_s \right) - q_s n_s \mathbf{u}_s \cdot \mathbf{E} \\ = - \left(\frac{\partial W}{\partial t} \right)_{Es} \end{aligned} \quad (12.26)$$

This equation can be simplified by invoking two mathematical identities, the first being

$$\frac{\partial}{\partial t} \left(\frac{m_s n_s u_s^2}{2} \right) + \nabla \cdot \left(\frac{m_s n_s u_s^2}{2} \mathbf{u}_s \right) = n_s \left(\frac{\partial}{\partial t} + \mathbf{u}_s \cdot \nabla \right) \frac{m_s u_s^2}{2} = n_s \frac{d}{dt} \left(\frac{m_s u_s^2}{2} \right)$$

The second identity is obtained by dotting the equation of motion with \mathbf{u}_s :

$$\begin{aligned} n_s m_s \left[\frac{\partial}{\partial t} \left(\frac{u_s^2}{2} \right) + \mathbf{u}_s \cdot \left(\nabla \left(\frac{u_s^2}{2} \right) - \mathbf{u}_s \times \nabla \times \mathbf{u}_s \right) \right] \\ = n_s q_s \mathbf{u}_s \cdot \mathbf{E} - \mathbf{u}_s \cdot \nabla p_s - \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s \end{aligned}$$

or

$$n_s \frac{d}{dt} \left(\frac{m_s u_s^2}{2} \right) = n_s q_s \mathbf{u}_s \cdot \mathbf{E} - \mathbf{u}_s \cdot \nabla p_s - \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s$$

Inserting these two into Equation 12.26 gives the energy evolution equation

$$\frac{N dp_s}{2 dt} + \frac{2+N}{2} p_s \nabla \cdot \mathbf{u}_s = -\nabla \cdot \mathbf{Q}_s + \mathbf{R}_{s\alpha} \cdot \mathbf{u}_s - \left(\frac{\partial W}{\partial t} \right)_{Es} \quad (12.27)$$

The first term on the right-hand side represents the heat flux, the second term gives the frictional heating of species s due to frictional drag on species α , while the last term on the right-hand side gives the rate at which species s collisionally transfers energy to other species. Although Equation 12.27 is complicated, two important limiting situations become evident if we let t_{char} be the characteristic time scale for a given phenomenon and l_{char} be its characteristic length scale. A characteristic velocity $V_{\text{ph}} \sim l_{\text{char}}/t_{\text{char}}$ may then be defined for the phenomenon and so, replacing temporal derivatives by t_{char}^{-1} and spatial derivatives by l_{char}^{-1} in Equation 12.27, it is seen that the two limiting situations are:

1. *Isothermal limit* - The heat flux term dominates all other terms, in which case the temperature becomes spatially uniform. This occurs if (i) $v_{Ts} \gg V_{ph}$ since the ratio of the left-hand side terms to the heat flux term is $\sim V_{ph}/v_{Ts}$ and (ii) the collisional terms are small enough to be ignored.
2. *Adiabatic limit* - The heat flux terms and the collisional terms are small enough to be ignored compared to the left-hand side terms; this occurs when $V_{ph} \gg v_{Ts}$. Adiabatic is a Greek word meaning “impassable”, and is used here to denote that no heat is flowing, i.e., the volume under consideration is thermally isolated from the outside world.

Both of these limits make it possible to avoid solving for \mathbf{Q}_s , which involves the third moment, and so both the adiabatic and isothermal limit provide a closure to the moment equations.

The energy equation may be greatly simplified in the adiabatic limit by re-arranging the continuity equation to give

$$\nabla \cdot \mathbf{u}_s = -\frac{1}{n_s} \frac{dn_s}{dt}$$

and then substituting this expression into the left-hand side of Equation 12.27 to obtain

$$\frac{1}{p_s} \frac{dp_s}{dt} = \frac{\gamma}{n_s} \frac{dn_s}{dt} \quad (12.28)$$

where

$$\gamma = \frac{N+2}{N}$$

Equation 12.28 implies

$$\frac{d}{dt} \left(\frac{p_s}{n_s^\gamma} \right) = 0$$

so p_s/n_s^γ is a constant in the frame of the moving plasma. This constitutes a derivation of adiabaticity based on geometry and statistical mechanics rather than on thermodynamic arguments.

The energy equation derivation can also be found in [An introductory guide to fluid models with anisotropic temperatures](#).

As a short summary, the fluid equations we have been using are simply moments of the Boltzmann equation. One interesting observation from Equation 12.12 is that in order to recover the hydrodynamic equations, the acceleration term $\mathbf{a} \cdot \partial_{\mathbf{v}} f$ is not required: the pressure gradient that drives the flow arises from the thermal motion embedded in the advection term $\mathbf{v} \cdot \partial_{\mathbf{x}} f$.

We must be careful not to become overconfident regarding the descriptive power of the fluid point of view because weaknesses exist in this point of view. For example, as discussed above neither the adiabatic nor the isothermal approximation is appropriate when $V_{ph} \sim v_{Ts}$. The fluid description breaks down in this situation and the Vlasov description must be used in

this situation. Furthermore, the distribution function is Maxwellian only if there are sufficient collisions or some other randomizing process. Because of the weak collisionality of a plasma, this is very often not the case. In particular, since the collision frequency scales as v^{-3} (Equation 9.20), fast particles take much longer to become Maxwellian than slow particles. It is not at all unusual for a real plasma to be in a state where the low-velocity particles have reached a Maxwellian distribution whereas the fast particles form a non-Maxwellian tail.

12.5 Plasma Oscillations and Landau Damping

As an elementary illustration of the use of the Vlasov equation, we shall derive the dispersion relation for electron plasma oscillations, which is originally treated from the fluid point of view. This derivation will require a knowledge of contour integration.

In zeroth order, we assume a uniform plasma with a distribution $f_0(\mathbf{v})$, and we let $\mathbf{B}_0 = \mathbf{E}_0 = 0$. In first order, we denote the perturbation in $f(\mathbf{r}, \mathbf{v}, t)$ by $f_1(\mathbf{r}, \mathbf{v}, t)$:

$$f(\mathbf{r}, \mathbf{v}, t) = f_0(\mathbf{v}) + f_1(\mathbf{r}, \mathbf{v}, t)$$

Since \mathbf{v} is now an independent variable and is not to be linearized, the first-order Vlasov equation for electron is

$$\frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \nabla f_1 - \frac{e}{m} \mathbf{E}_1 \cdot \frac{\partial f_0}{\partial \mathbf{v}} = 0 \quad (12.29)$$

As before, we assume the ions are massive and fixed and that the waves are plane waves in the x direction $f_1 \propto e^{i(kx - \omega t)}$. Then the linearized Vlasov equation becomes

$$\begin{aligned} -i\omega f_1 + ikv_x f_1 x &= \frac{e}{m} E_x \frac{\partial f_0}{\partial v_x} \\ f_1 &= \frac{ieE_x}{m} \frac{\partial f_0 / \partial v_x}{\omega - kv_x} \end{aligned}$$

Poisson's equation gives

$$\epsilon_0 \nabla \cdot \mathbf{E}_1 = ik\epsilon_0 E_x = -en_1 = -e \int f_1 d^3v$$

Substituting for f_1 and dividing by $ik\epsilon_0 E_x$, we have

$$1 = -\frac{e^2}{km\epsilon_0} \int \frac{\partial f_0 / \partial v_x}{\omega - kv_x} d^3v$$

A factor n_0 can be factored out if we replace f_0 by a normalized function \hat{f}_0 :

$$1 = -\frac{\omega_p^2}{k} \int_{-\infty}^{\infty} dv_z \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} \frac{\partial \hat{f}_0(v_x, v_y, v_z)/\partial v_x}{\omega - kv_x} dv_x$$

If f_0 is a Maxwellian or some other factorable distribution, the integration over v_y and v_z can be carried out easily. What remains is the one-dimensional distribution $\hat{f}_0(v_x)$. For instance, a one-dimensional Maxwellian distribution is

$$\hat{f}_m(v_x) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(\frac{-mv_x^2}{2k_B T}\right)$$

Since we are dealing with a one-dimensional problem we may drop the subscript x, begin careful not to confuse v (which is really v_x) with the total velocity v used earlier:

$$1 = \frac{\omega_p^2}{k^2} \int_{-\infty}^{\infty} \frac{\partial \hat{f}_0/\partial v}{v - \omega/k} dv \quad (12.30)$$

Here, \hat{f}_0 is understood to be a one-dimensional distribution function, the integrations over v_y and v_z having been made. This equation holds for any equilibrium distribution $\hat{f}_0(v)$.

The integral in this equation is not straightforward to evaluate because of the singularity at $v = \omega/k$. One might think that the singularity would be of no concern, because in practice ω is almost always never real; waves are usually slightly damped by collisions or are amplified by some instability mechanisms. Since the velocity v is a real quantity, the denominator never vanishes. Landau was the first to treat this equation properly. He found that even though the singularity lies off the path of integration, its presence introduces an important modification to the plasma wave dispersion relation — an effect not predicted by the fluid theory.

Consider an initial value problem in which the plasma is given a sinusoidal perturbation, and therefore k is real. If the perturbation grows or decays, ω will be complex. This integral must be treated as a contour integral in the complex v plane. Possible contours are shown for (a) an unstable wave, with $\Im(\omega) > 0$, and (b) a damped wave, with $\Im(\omega) < 0$. Normally, one would evaluate the line integral along the real v axis by the residue theorem:

$$\int_{C_1} G dv + \int_{C_2} G dv = 2\pi i R(\omega/k)$$

where G is the integrand, C_1 is the path along the real axis, C_2 is the semicircle at infinity, and $R(\omega/k)$ is the residue at ω/k . This works if the integral over C_2 vanishes. Unfortunately, this does not happen for a Maxwellian distribution, which contains the factor

$$\exp(-v^2/v_{th}^2)$$

This factor becomes large for $v \rightarrow \pm i\infty$, and the contribution from C_2 cannot be neglected. Landau showed that when the problem is properly treated as an initial value problem the correct contour to use is the curve C_1 passing below the singularity. This integral must in general be evaluated numerically.

Although an exact analysis of this problem is complicated, we can obtain an approximate dispersion relation for the case of large phase velocity and weak damping. In this case, the pole at ω/k lies near the real v axis. The contour prescribed by Landau is then a straight line along the $\Re(v)$ axis with a small semicircle around the pole. In going around the pole, one obtains $2\pi i$ times half the residue there. Then Equation 12.30 becomes

$$1 = \frac{\omega_p^2}{k^2} \left[P \int_{-\infty}^{\infty} \frac{\partial \hat{f}_0 / \partial v}{v - (\omega/k)} dv + i\pi \frac{\partial \hat{f}_0}{\partial v} \Big|_{v=\omega/k} \right] \quad (12.31)$$

where P stands for the Cauchy principal value. To evaluate this, we integrate along the real v axis but stop just before encountering the pole. If the phase velocity $v_\phi = \omega/k$ is sufficiently large, as we assume, there will not be much contribution from the neglected part of the contour, since both \hat{f}_0 and $\partial \hat{f}_0 / \partial v$ are very small there. The integral above can be evaluated by integration by parts:

$$\int_{-\infty}^{\infty} \frac{\partial \hat{f}_0}{\partial v} \frac{dv}{v - v_\phi} = \left[\frac{\hat{f}_0}{v - v_\phi} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{-\hat{f}_0 dv}{(v - v_\phi)^2} = \int_{-\infty}^{\infty} \frac{\hat{f}_0 dv}{(v - v_\phi)^2}$$

Since this is just an average of $(v - v_\phi)^{-2}$ over the distribution, the real part of the dispersion relation can be written

$$1 = \frac{\omega_p^2}{k^2} \overline{(v - v_\phi)^{-2}}$$

Since $v_\phi \gg v$ has been assumed, we can expand $(v - v_\phi)^{-2}$:

$$(v - v_\phi)^{-2} = v_\phi^{-2} \left(1 - \frac{v}{v_\phi} \right)^{-2} = v_\phi^{-2} \left(1 + \frac{2v}{v_\phi} + \frac{3v^2}{v_\phi^2} + \frac{4v^3}{v_\phi^3} + \dots \right)$$

The odd terms vanish upon taking the average, and we have

$$\overline{(v - v_\phi)^{-2}} \approx v_\phi^{-2} \left(1 + \frac{3\bar{v}^2}{v_\phi^2} \right)$$

We now let \hat{f}_0 be Maxwellian and evaluate \bar{v}^2 . Remembering that v here is an abbreviation for v_x , we can write

$$\frac{1}{2} m \bar{v}_x^2 = \frac{1}{2} k_B T_e$$

there being only one degree of freedom. The dispersion relation then becomes

$$1 = \frac{\omega_p^2}{k^2} \frac{k^2}{\omega^2} \left(1 + 3 \frac{k^2}{\omega^2} \frac{k_B T_e}{m} \right)$$

$$\omega^2 = \omega_p^2 + \frac{\omega_p^2}{\omega^2} \frac{3k_B T_e}{m} k^2$$

If the thermal correction is small (i.e. the second term on the right-hand side is small, such that $\omega \approx \omega_p$), we may replace ω^2 by ω_p^2 in the second term. We then have

$$\omega^2 = \omega_p^2 + \frac{3k_B T_e}{m} k^2$$

which is the same as that been obtained from the fluid equations with $\gamma = 3$.

We now return to the imaginary term in the dispersion relation. In evaluating this small term, it will be sufficiently accurate to neglect the thermal correction to the real part of ω and let $\omega^2 \approx \omega_p^2$. From the evaluation of the real part above we see that the principle value of the integral is approximately k^2/ω^2 . The dispersion relation now becomes

$$1 = \frac{\omega_p^2}{\omega^2} + i\pi \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v} \Big|_{v=v_\phi}$$

$$\omega^2 \left(1 - i\pi \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v} \Big|_{v=v_\phi} \right) = \omega_p^2$$

Treating the imaginary term as small, we can bring it to the right-hand side and take the square root by Taylar series expansion. We then obtain

$$\omega = \omega_p \left(1 + i \frac{\pi}{2} \frac{\omega_p^2}{k^2} \frac{\partial \hat{f}_0}{\partial v} \Big|_{v=v_\phi} \right) \quad (12.32)$$

If \hat{f}_0 is a one-dimensional Maxwellian, we have

$$\frac{\partial \hat{f}_0}{\partial v} = (\pi v_{th}^2)^{-1/2} \left(\frac{-2v}{v_{th}^2} \right) \exp \left(\frac{-v^2}{v_{th}^2} \right) = -\frac{2v}{\sqrt{\pi} v_{th}} \exp \left(\frac{-v^2}{v_{th}^2} \right)$$

We may approximate v_ϕ by ω_p/k in the coefficient, but in the exponent we must keep the thermal correction in the real part of the dispersion relation. The damping is then given by

$$\begin{aligned} \Im(\omega) &= -\frac{\pi}{2} \frac{\omega_p^2}{k^2} \frac{2\omega_p}{k\sqrt{\pi}} \frac{1}{v_{th}^3} \exp \left(\frac{-\omega^2}{k^2 v_{th}^2} \right) \\ &= -\sqrt{\pi} \omega_p \left(\frac{\omega_p}{kv_{th}} \right)^3 \exp \left(\frac{-\omega_p^2}{k^2 v_{th}^2} \right) \exp \left(\frac{-3}{2} \right) \\ \Im \left(\frac{\omega}{\omega_p} \right) &= -0.22\sqrt{\pi} \left(\frac{\omega_p}{kv_{th}} \right)^3 \exp \left(\frac{-1}{2k^2 \lambda_D^2} \right) \end{aligned}$$

Since $\Im(\omega)$ is negative, there is a collisionless damping of plasma waves; this is called *Landau damping*. As is evident from the expression, this damping is extremely small for small $k\lambda_D$, but becomes important for $k\lambda_D = \mathcal{O}(1)$. This effect is connected with f_1 , the distortion of the distribution function caused by the wave.

12.6 The Meaning of Landau Damping

The theoretical discovery of wave damping without energy dissipation by collisions is perhaps the most astounding result of plasma physics research. That this is a real effect has been demonstrated in the laboratory. Although a simple physical explanation for this damping is now available, it is a triumph of applied mathematics that this unexpected effect was first discovered purely mathematically in the course of a careful analysis of a contour integral. Landau damping is a characteristic of collisionless plasmas, but it may also have application in other fields. For instance, in the kinetic treatment of galaxy formation, stars can be considered as atoms of a plasma interacting via gravitational rather than electromagnetic forces. Instabilities of the gas of stars can cause spiral arms to form, but this process is limited by Landau damping.

To see what is responsible for Landau damping, we first notice that $\Im(\omega)$ arises from the pole at $v = v_\phi$. Consequently, the effect is connected with those particles in the distribution that have a velocity nearly equal to the phase velocity — the “resonant particles”. These particles travel along with the wave and do not see a rapidly fluctuating electric field: they can, therefore, exchange energy with the wave effectively. The easiest way to understand this exchange of energy is to picture a surfer trying to catch an ocean wave. (Warning: this picture is only for directing our thinking along the right lines; it does not correctly explain the damping.) If the surfboard is not moving, it merely bobs up and down as the wave goes by and does not gain any energy on the average. Similarly, a boat propelled much faster than the wave cannot exchange much energy with the wave. However, if the surfboard has almost the same velocity as the wave, it can be caught and pushed along by the wave; this is, after all, the main purpose of the exercise. In that case, the surfboard gains energy, and therefore the wave must lose energy and is damped. On the other hand, if the surfboard should be moving slightly faster than the wave, it would push on the wave as it moves uphill; then the wave could gain energy. In a plasma, there are electrons both faster and slower than the wave. A Maxwellian distribution, however, has more slow electrons than fast ones. Consequently, there are more particles taking energy from the wave than vice versa, and the wave is damped. As particles with $v \approx v_\phi$ are trapped in the wave, $f(v)$ is flattened near the phase velocity. This distortion is $f_1(v)$ which we calculated. As seen in Fig (ADD IT!), the perturbed distribution function contains the same number of particles but has gained total energy (at the expense of the wave).

From this discussion, one can surmise that if $f_0(v)$ contained more fast particles than slow particles, a wave can be excited. Indeed, from the expression of ω above, it is apparent that $\Im(\omega)$ is positive if $\partial\hat{f}_0/\partial v$ is positive at $v = v_\phi$. Such a distribution is shown in Fig.7-19

(ADD IT!). Waves with v_ϕ in the region of positive slope will be unstable, gaining energy at the expense of the particles. This is just the finite-temperature analogy of the two stream instability. When there are two cold ($k_B T = 0$) electron streams in motion, $f_0(v)$ consists of two δ -functions. This is clearly unstable because $\partial f_0/\partial v$ is infinite; and indeed, we found the instability from fluid theory. When the streams have finite temperature, kinetic theory tells us that the relative densities and temperatures of the two stream must be such as to have a region of positive $\partial f_0(v)/\partial v$ between them; more precisely, the total distribution function must have a minimum for instability.

The physical picture of a surfer catching waves is very appealing, but it is not precise enough to give us a real understanding of Landau damping. There are actually two kinds of Landau damping. Both kinds are independent of dissipative collisional mechanisms. If a particle is caught in the potential well of a wave, the phenomenon is called “trapping”. As in the case of a surfer, particles can indeed gain or lose energy in trapping. However, trapping does not lie within the purview of the linear theory. That this is true can be seen from the equation of motion

$$m\ddot{x} = qE(x)$$

If one evaluates $E(x)$ by inserting the exact value of x , the equation would be nonlinear, since $E(x)$ is something like $\sin kx$. What is done in linear theory is to use for x the unperturbed orbit; i.e. $x = x_0 + v_0 t$. Then this becomes linear. This approximation, however, is no longer valid when a particle is trapped. When it encounters a potential hill large enough to reflect it, its velocity and position are, of course, greatly affected by the wave and are not close to their unperturbed values. In fluid theory, the equation of motion is

$$m \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = q\mathbf{E}(x)$$

Here, $\mathbf{E}(x)$ is to be evaluated in the laboratory frame, which is easy; but to make up for it, there is the $(\mathbf{v} \cdot \nabla) \mathbf{v}$ term. The neglect of $(\mathbf{v}_1 \cdot \nabla) \mathbf{v}_1$ in linear theory amounts to the same thing as using unperturbed orbits. In kinetic theory, the nonlinear term that is neglected is, from the first-order Vlasov Equation 12.29,

$$\frac{q}{m} E_1 \frac{\partial f_1}{\partial v}$$

When the particles are trapped, they reverse their direction of travel relative to the wave, so the distribution function $f(v)$ is greatly disturbed near $v = \omega/k$. This means that $\partial f_1/\partial v$ is comparable to $\partial f_0/\partial v$, and the term above is not negligible. Hence, trapping is not in the linear theory.

When a wave grows to a large amplitude, collisionless damping with trapping does occur. One then finds that the wave does not decay monotonically; rather, the amplitude fluctuates during the decay as the trapped particles bounce back and forth in the potential wells. This is

nonlinear Landau damping. Since the result before was derived from a *linear* theory, it must arise from a different physical effect. The question is: can untrapped electrons moving close to the phase velocity of the wave exchange energy with the wave? Before giving the answer, let us examine the energy of such electrons.

12.6.1 The Kinetic Energy of a Beam of Electrons

We may divide the electron distribution $f_0(v)$ into a large number of monoenergetic beams. Consider one of these beams: it has unperturbed velocity u and density n_u . The velocity u may lie near v_ϕ , so that this beam may consist of resonant electrons. We now turn on a plasma oscillation $E(x, t)$ and consider the kinetic energy of the beam as it moves through the crests and troughs of the wave. The wave is caused by a self-consistent motion of all the beams together. If n_u is small enough (the number of beams large enough), the beam being examined has a negligible effect on the wave and may be considered as moving in a given field $E(x, t)$. Let

$$E = E_0 \sin(kx - \omega t) = -d\phi/dt$$

$$\phi = (E_0/k) \cos(kx - \omega t)$$

The linearized fluid equation for the beam is

$$m \left(\frac{\partial v_1}{\partial t} + u \frac{\partial v_1}{\partial x} \right) = -eE_0 \sin(kx - \omega t)$$

A possible solution is

$$v_1 = -\frac{eE_0}{m} \frac{\cos(kx - \omega t)}{\omega - ku}$$

This is the velocity modulation caused by the wave as the beam electrons move past. To conserve particle flux, there is a corresponding oscillation in density, given by the linearized continuity equation:

$$\frac{\partial n_1}{\partial t} + u \frac{\partial n_1}{\partial x} = -n_u \frac{\partial v_1}{\partial x}$$

Since v_1 is proportional to $\cos(kx - \omega t)$, we can try $n_1 = \bar{n}_1 \cos(kx - \omega t)$. Substitution of this into the above yields

$$n_1 = -n_u \frac{eE_0 k}{m} \frac{\cos(kx - \omega t)}{(\omega - ku)^2}$$

(n_1 and v_1 can be shown in a series of phase relation plots as in Fig.7-21)(ADD IT!) one wavelength of E and of the potential $-e\phi$ seen by the beam electrons.

We may now compute the kinetic energy W_k of the beam:

$$\begin{aligned} W_k &= \frac{1}{2}m(n_u + n_1)(u + v_1)^2 \\ &= \frac{1}{2}m(n_u u^2 + n_u v_1^2 + 2un_1v_1 + n_1 u^2 + 2n_u uv_1 + n_1 v_1^2) \end{aligned}$$

The last three terms contain odd powers of oscillating quantites, so they will vanish when we average over a wavelength. The change in W_k due to the wave is found by subtracting the first term, which is the original energy. The average energy change is then

$$\langle \Delta W_k \rangle = \frac{1}{2}m \langle n_u v_1^2 + 2un_1v_1 \rangle$$

From the form of v_1 , we have

$$n_u \langle v_1^2 \rangle = \frac{1}{2}n_u \frac{e^2 E_0^2}{m^2(\omega - ku)^2}$$

the factor $\frac{1}{2}$ representing $\langle \cos^2(kx - \omega t) \rangle$. Similarly, from the form of n_1 ,

$$2u \langle n_1 v_1 \rangle = n_u \frac{e^2 E_0^2 k u}{m^2(\omega - ku)^3}$$

Consequently,

$$\begin{aligned} \langle \Delta W_k \rangle &= \frac{1}{4}m n_u \frac{e^2 E_0^2}{m^2(\omega - ku)^2} \left[1 + \frac{2ku}{\omega - ku} \right] \\ &= \frac{n_u}{4} \frac{e^2 E_0^2}{m} \frac{\omega + ku}{m^2(\omega - ku)^3} \end{aligned}$$

This result shows that $\langle \Delta W_k \rangle$ depends on the frame of the observer and that it does not change secularly with time. Consider the picture of a frictionless block sliding over a washboard-like surface. (ADD FIGURE!) In the frame of the washboard, $\langle \Delta W_k \rangle$ is proportional to $-(ku)^{-2}$, as seen by taking $\omega = 0$. It is intuitively clear that (1) $\langle \Delta W_k \rangle$ is negative, since the block spends more time at the peaks than at the valleys, and (2) the block does not gain or lose energy on the average, *once the oscillation is started* (no time-dependence). Now if one goes into a frame in which the washboard is moving with a steady velocity ω/k (a velocity unaffected by the motion of the block, since we have assumed that n_u is negligibly small compared with the density of the whole plasma), it is still true that the block does not gain or lose energy on the average, once the oscillation is started. But the above equation tells us that $\langle \Delta W_k \rangle$ depends on the velocity ω/k , and hence on the frame of the observer. In particular, it shows that a beam has less energy in the presence of the wave than in its absence if $\omega - ku < 0$ or $u > u_\phi$, and it has more energy if $\omega - ku > 0$ or $u < u_\phi$. The reason for this can be traced back to the phase relation between n_1 and v_1 . As Fig.7-23 (ADD IT!) shows, W_k is a parabolic

function of v . As v oscillates between $u - |v_1|$ and $u + |v_1|$, W_k will attain an average value larger than the equilibrium value W_{k0} , provided that the particle spends an equal amount of time in each half of the oscillation. This effect is the meaning of the first term $\frac{1}{2}m \langle n_u v_1^2 \rangle$, which is positive definite. The second term $\frac{1}{2}m \langle 2un_1 v_1 \rangle$ is a correction due to the fact that the particle does not distribution its time equally. In Fig.7-21 (ADD IT!), one sees that both electrons a and b spend more time at the top the potential hill than at the bottom, but electron a reaches that point after a period of deceleration, so that v_1 is negative there, while electron b reaches that point after a period of acceleration (to the right), so that v_1 is positive there. This effect causes $\langle W_k \rangle$ to change sign at $u = v_\phi$.

12.6.2 The Effect of Initial Conditions

The result we have just derived, however, still has nothing to do with linear Landau damping. Damping requires a continuous increase of W_k at the expense of wave energy, but we have found that $\langle \Delta W_k \rangle$ for untrapped particles is contant in time. If neither the untrapped particles nor the trapped particles are responsible for linear Landau damping, what is? The answer can be gleaned form the following observation: if $\langle \Delta W_k \rangle$ is positive, say, there must have been a time when it was increasing. Indeed, there are particles in the original distribution which have velocities so close to v_ϕ that at time t they have not yet gone a half wavelength relative to the wave. For these particles, one cannot take the average $\langle \Delta W_k \rangle$. These particles can absorb energy from the wave and are properly called the “resonant” particles. As time goes on, the number of resonant electrons decreases, since an increasing number will have shifted more than $\frac{1}{2}\lambda$ from their orginal positions. The damping rate, however, can stay constant, since the amplitude is now smaller, and it takes fewer electrons to maintain a constant damping rate.

The effect of the initial conditions is most easily seen form a phase-space diagram (Fig.7-24)(ADD IT with Luxor?).

12.7 A Physical Derivation of Landau Damping

We are now in a position to derive the Landau damping rate recourse to contour integration. Although Landau’s derivation of collisionless damping was short and neat, it was not clear that it concerned a physically observable phenomenon until J. M. Dawson gave the longer, intuitive derivation which is paraphrased in this section. As before, we divide the plasma up into beams of velocity u and density n_u , and examine their motion in a wave

$$E = E_1 \sin(kx - \omega t)$$

From the derivations in the previous section, the velocity of each beam is

$$v_1 = -\frac{eE_1}{m} \frac{\cos(kx - \omega t)}{\omega - ku}$$

This solution satisfies the equation of motion, but it does not satisfy the initial condition $v_1 = 0$ at $t = 0$. It is clear that this initial condition must be imposed; otherwise, v_1 would be very large in the vicinity of $u = \omega/k$, and the plasma would be in a specially prepared state initially. We can fix up the solution to satisfy the initial condition by adding an arbitrary function of $kx - kut$. The composite solution would still satisfy the equation of motion because the operator on the left-hand side, when applied to $f(kx - kut)$, gives zero. Obviously, to get $v_1 = 0$ at $t = 0$, the function $f(kx - kut)$ must be taken to be $-\cos(kx - kut)$. Thus we have, instead of the expression above,

$$v_1 = \frac{-eE_1}{m} \frac{\cos(kx - \omega t) - \cos(kx - kut)}{\omega - ku}$$

Next, we must solve the equation of continuity for n_1 , again subject to the initial condition $n_1 = 0$ at $t = 0$. Since we are now much cleverer than before, we may try a solution of the form

$$n_1 = \bar{n}_1 [\cos(kx - \omega t) - \cos(kx - kut)]$$

Inserting this into the equation of continuity and using the expression for v_1 above, we find

$$\bar{n}_1 \sin(kx - \omega t) = -n_u \frac{eE_1 k}{m} \frac{\sin(kx - \omega t) - \sin(kx - kut)}{(\omega - ku)^2}$$

Apparently, we were not clever enough, since the $\sin(kx - \omega t)$ factor does not cancel. To get a term of the form $\sin(kx - kut)$, which came from the added term in v_1 , we can add a term of the form $A t \sin(kx - kut)$ to n_1 . This term obviously vanishes at $t = 0$, and it will give the $\sin(kx - kut)$ term when the operator on the left-hand side of the equation of continuity operates on the t factor. When the operator operates on the $\sin(kx - kut)$ factor, it yields zero. The coefficient A must be proportional to $(\omega - ku)^{-1}$ in order to match the same factor in $\partial v_1 / \partial x$. Thus we take

$$n_1 = -n_u \frac{eE_1 k}{m} \frac{1}{(\omega - ku)^2} \times [\cos(kx - \omega t) - \cos(kx - kut) - (\omega - ku)t \sin(kx - kut)]$$

This clearly vanishes at $t = 0$, and one can easily verify that it satisfies the equation of continuity.

These expressions for v_1 and n_1 allow us now to calculate the work done by the wave on each beam. The force acting on a unit volume of each beam is

$$F_u = -eE_q \sin(kx - \omega t)(n_u + n_1)$$

and therefore its energy changes at the rate

$$\frac{dW}{dt} = F_u(u + v_1) = -eE_1 \sin(kx - \omega t) \left(\underbrace{n_u u}_1 + \underbrace{n_u v_1}_2 + \underbrace{n_1 u}_3 + \underbrace{n_1 v_1}_4 \right)$$

We now take the spatial average over a wavelength. The first term vanishes because $n_u u$ is a constant. The fourth term can be neglected because it is second order, but in any case it can be shown to have zero average. The second and third terms can be evaluated with the help of identities

$$\begin{aligned} \langle \sin(kx - \omega t) \cos(kx - kut) \rangle &= -\frac{1}{2} \sin(\omega t - kut) \\ \langle \sin(kx - \omega t) \sin(kx - kut) \rangle &= \frac{1}{2} \cos(\omega t - kut) \end{aligned}$$

The result is then

$$\begin{aligned} \left\langle \frac{dW}{dt} \right\rangle_u &= \frac{e^2 E_1^2}{2m} n_u \left[\frac{\sin(\omega t - kut)}{\omega - ku} \right. \\ &\quad \left. + ku \frac{\sin(\omega t - kut) - (\omega - ku)t \cos(\omega t - kut)}{(\omega - ku)^2} \right] \end{aligned}$$

The total work done on the particles is found by summing over all the beams:

$$\sum_u \left\langle \frac{dW}{dt} \right\rangle_u = \int \frac{f_0(u)}{n_u} \left\langle \frac{dW}{dt} \right\rangle_u dy = n_0 \int \frac{\hat{f}_0(u)}{n_u} \left\langle \frac{dW}{dt} \right\rangle_u du$$

Inserting the expression of $\left\langle \frac{dW}{dt} \right\rangle_u$ and using the definition of ω_p , we then find for the rate of change of kinetic energy

$$\begin{aligned}
\left\langle \frac{dW_k}{dt} \right\rangle &= \frac{\epsilon_0 E_1^2}{2} \omega_p^2 \left[\int \hat{f}_0(u) \frac{\sin(\omega t - kut)}{\omega - ku} du \right. \\
&\quad \left. + \int \hat{f}_0(u) \frac{\sin(\omega t - kut) - (\omega - ku)t \cos(\omega t - kut)}{(\omega - ku)^2} ku du \right] \\
&= \frac{1}{2} \epsilon_0 E_1^2 \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \left\{ \frac{\sin(\omega t - kut)}{\omega - ku} + u \frac{d}{du} \left[\frac{\sin(\omega t - kut)}{\omega - ku} \right] \right\} \\
&= \frac{1}{2} \epsilon_0 E_1^2 \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \frac{d}{du} \left[u \frac{\sin(\omega t - kut)}{\omega - ku} \right]
\end{aligned}$$

This is to be set equal to the rate of loss of wave energy density W_w . The wave energy consists of two parts. The first part is the energy density of the electrostatic field:

$$\langle W_E \rangle = \frac{\epsilon \langle E^2 \rangle}{2} = \frac{\epsilon E_1^2}{4}$$

The second part is the kinetic energy of oscillation of the particles. If we again divide the plasma up into beams, the energy per beam is given as before

$$\langle \Delta W_k \rangle_u = \frac{1}{4} \frac{n_u}{m} \frac{e^2 E_1^2}{(\omega - ku)^2} \left[1 + \frac{2ku}{\omega - ku} \right]$$

In deriving this result, we did not use the correct initial conditions, which are important for the resonant particles; however, the latter contribute very little to the total energy of the wave. Summing over the beams, we have

$$\langle \Delta W_k \rangle = \frac{1}{4} \frac{e^2 E_1^2}{m} \int_{-\infty}^{\infty} \frac{f_0(u)}{(\omega - ku)^2} \left[1 + \frac{2ku}{\omega - ku} \right] du$$

The second term in the brackets can be neglected in the limit $\omega/k \gg v_{th}$, which we shall take in order to compare with our previous results. (???) The dispersion relation is found by Poisson's equation:

$$k\epsilon_0 E_1 \cos(kx - \omega t) = -e \sum_u n_1$$

Using the expression for n_1 in the previous section with the wrong initial conditions, we have

$$1 = \frac{e^2}{\epsilon_0 m} \sum_u \frac{n_u}{(\omega - ku)^2} = \frac{e^2}{\epsilon_0 m} \int_{-\infty}^{\infty} \frac{f_0(u) du}{(\omega - ku)^2}$$

Comparing this with the expression of $\langle \Delta W_k \rangle$, we find

$$\langle \Delta W_k \rangle = \frac{1}{4} \frac{e^2 E_1^2}{m} \frac{\epsilon_0 m}{e^2} = \frac{\epsilon_0 E_1^2}{4} = \langle W_E \rangle$$

Thus

$$W_w = \frac{\epsilon E_1^2}{2}$$

The rate of change of wave energy density W_w is given by $-\left\langle \frac{dW_w}{dt} \right\rangle$:

$$\frac{dW_w}{dt} = -W_w \omega_p^2 \int_{-\infty}^{\infty} \hat{f}_0(u) du \frac{d}{du} \left[u \frac{\sin(\omega t - kut)}{\omega - ku} \right]$$

Integration by parts gives

$$\begin{aligned} \frac{dW_w}{dt} &= W_w \omega_p^2 \left\{ \left[u \hat{f}_0(u) \frac{\sin(\omega - ku)t}{\omega - ku} \right]_{-\infty}^{\infty} \right. \\ &\quad \left. - \int_{-\infty}^{\infty} u \frac{d\hat{f}_0}{du} \frac{\sin(\omega - ku)t}{\omega - ku} du \right\} \end{aligned}$$

The integrated part vanishes for well-behaved functions $\hat{f}_0(u)$, and we have

$$\frac{dW_w}{dt} = W_w \frac{\omega}{k} \omega_p^2 \int_{-\infty}^{\infty} \frac{d\hat{f}_0}{du} \left[\frac{\sin(\omega - ku)t}{\omega - ku} \right] du$$

where u has been set equal to ω/k (a constant), since only velocities very close to this will contribute to the integral. In fact, for sufficiently large t , the square bracket can be approximated by a δ -function:

$$\delta\left(u - \frac{\omega}{k}\right) = \frac{k}{\pi} \lim_{t \rightarrow \infty} \left[\frac{\sin(\omega - ku)t}{\omega - ku} \right]$$

(The original form is $\delta(x) = \lim_{\epsilon \rightarrow 0} \frac{\sin(x/\epsilon)}{\pi x}$, where the function on the right is called sinc.)

Thus

$$\frac{dW_w}{dt} = W_w \frac{\omega}{k} \omega_p^2 \frac{\pi}{k} \frac{\omega}{k} \hat{f}_0\left(\frac{\omega}{k}\right) = W_w \pi \omega \frac{\omega_p^2}{k^2} \hat{f}'_0\left(\frac{\omega}{k}\right)$$

Since $\Im(\omega)$ is the growth rate of E_1 , and W_w is proportional to E_1^2 , we must have

$$\frac{dW_w}{dt} = 2\mathfrak{I}(\omega)W_w$$

Hence

$$\mathfrak{I}(\omega) = \frac{\pi}{2}\omega \frac{\omega_p^2}{k^2} \hat{f}'_0\left(\frac{\omega}{k}\right)$$

in agreement with the previous result for $\omega = \omega_p$.

12.7.1 The Resonant Particles

We are now in a position to see precisely which are the resonant particles that contribute to linear Landau damping. Fig.7-25(Sinc function, ADD IT!) gives a plot of the factor multiplying $\hat{f}'_0(u)$ in the integrand. We see that the largest contribution comes from particles with $|\omega - ku| < \pi/t$, or $|v - v_\phi| < \pi/k = \lambda/2$; i.e. those particles in the initial distribution that have not yet traveled a half-wavelength relative to the wave. The width of the central peak narrows with time, as expected. The subsidiary peaks in the “diffraction pattern” of Fig.7-25 come from particles that have traveled into neighboring half-wavelengths of the wave potential. These particles rapidly become spread out in phase, so that they contribute little on the average; the initial distribution is forgotten. Note that the width of the central peak is independent of the initial amplitude of the wave; hence, the resonant particles may include both trapped and untrapped particles. This phenomenon is unrelated to particle trapping.

12.7.2 Two Paradoxes Resolved

Fig.7-25 shows that the integrand sinc function is an even function of $\omega - ku$, so that particles going both faster than the wave and slower than the wave add to Landau damping. This is the physical picture we found in Fig.7-24. On the other hand, the slope of the curve of Fig.7-25, which represents the factor in the integrand of one previous equation (WE NEED NUMBERING!), is an odd function of $\omega - ku$; and one would infer from this that particles traveling faster than the wave give energy to it, while those traveling slower than the wave take energy from it. The two descriptions differ by an integration by parts. Both descriptions are correct; which is to be chosen depends on whether one wishes to have $\hat{f}_0(u)$ or $\hat{f}'_0(u)$ in the integrand.

A second paradox concerns the question of Galilean invariance. If we take the view that damping requires there be fewer particles traveling faster than the wave than slower, there is no problem as long as one is in the frame in which the plasma is at rest. However, if one goes into another frame moving with a velocity V , there would appear to be more particles moving faster than the wave than slower, and one would expect the wave to grow instead of

decay. This paradox is removed by reinserting the second term in Eq.(???) $\frac{2ku}{\omega - ku}$ which we neglected. As shown in the previous section, this term can make $\langle \Delta W_k \rangle$ negative. Indeed, in the moving frame, the second term is not negligible, $\langle \Delta W_k \rangle$ is negative, and the wave appears to have negative energy (that is, there is more energy in the quiescent, drifting Maxwellian distribution than in the presence of an oscillation). The wave “grows”, but adding energy to a negative energy wave makes its amplitude decrease.

12.8 BGK and Van Kampen Modes

We have seen that Landau damping is directly connected to the requirement that $f_0(v)$ be initially uniform in space. (It tends to make negative slopes to zero.) On the other hand, one can generate undamped electron waves if $f(v, t = 0)$ is mode to be constant along the particle trajectories initially. (???) It is easy to from Fig.7-24 that the particles will neither gain nor lose energy, on the average, if the plasma is initially prepared so that the density is constant along each trajectory. Such a wave is called a BGK mode, since it was I. B. Bernstein, J. M. Greene, and M. D. Kruskal who first showed that undamped waves of arbitrary ω , k , amplitude, and waveform were possible. The crucial parameter to adjust in tailoring $f(v, t = 0)$ to form a BGK mode is the relative number of trapped and untrapped particles. If we take the small-amplitude limit of a BGK mode, we obtain what is called a Van Kampen mode. In this limit, only the particles with $v = v_\phi$ are trapped. We can change the number of trapped particles by adding to $f(v, t = 0)$ a term proportional to $\delta(v - v_\phi)$. Examination of Fig.7-24 will show that adding particles along the line $v = v_\phi$ will not cause damping — at a later time, there are just as many particles gaining energy as losing energy. In fact, by choosing distributions with δ -functions at other values of v_ϕ , one can generate undamped Van Kampen modes of arbitrary v_ϕ . Such singular initial conditions are, however, not physical. To get a smoothly varying $f(v, t = 0)$, one must sum over Van Kampen modes with a distribution of v_ϕ s. Although each mode is undamped, the total perturbation will show Landau damping because the various modes get out of phase with one another. (???)

12.9 Ion Landau Damping

Electrons are not the only resonant particles. If a wave has a slow enough phase velocity to match the thermal velocity of ions, ion Landau damping can occur. The ion acoustic wave, for instance, is greatly affected by Landau damping. Recall from the fluid theory that the dispersion relation for ion waves is

$$\frac{\omega}{k} = v_s = \left(\frac{k_B T_e + \gamma_i k_B T_i}{m_i} \right)^{1/2}$$

If $T_e \leq T_i$, the phase velocity lies in the region where $f_{0i}(v)$ has a negative slope, as shown in Fig.7-30(A)(ADD IT!!!). Consequently, ion waves are heavily Landau-damped if $T_e \leq T_i$. Ion

waves are observable only if $T_e \gg T_i$ (Fig. 7-30(B)), so that the phase velocity lies far in the tail of the ion velocity distribution. A clever way to introduce Landau damping in a controlled manner was employed by Alexeff, Jones, and Montgomery. A weakly damped ion wave was created in a heavy-ion plasma (such as xenon) with $T_e \gg T_i$. A small amount of a light atom (helium) was then added. Since the helium had about the temperature as the xenon but had much smaller mass, its distribution function was much broader, as shown by the dashed curve in Fig. 7-30(B). The resonant helium ions then caused the wave to damp.

12.9.1 The Plasma Dispersion Function

To introduce some of the standard terminology of kinetic theory, we now calculate the ion Landau damping of ion acoustic waves in the absence of magnetic fields. Ions and electrons follow the Vlasov equation and have perturbations of the form $f_1 \propto \exp(i kx - i \omega t)$ indicating plane waves propagating in the x direction. The solution for f_1 is given by the linearized momentum equation with appropriate modifications:

$$f_{1j} = -\frac{iq_j E}{m_j} \frac{\partial f_{oj}/\partial v_j}{\omega - kv_j}$$

where E and v_j stand for E_x, v_{xj} ; and the j th species has charge q_j , mass m_j , and particle velocity v_j . The density perturbation of the j th species is given by

$$n_{1j} = \int_{-\infty}^{\infty} f_{1j}(v_j) dv_j = -\frac{iq_j E}{m_j} \int_{-\infty}^{\infty} \frac{\partial f_{oj}/\partial v_j}{\omega - kv_j} dv_j$$

Let the equilibrium distributions f_{0j} be one-dimensional Maxwellians:

$$f_{0j} = \frac{n_{0j}}{v_{thj}\pi^{1/2}} e^{-v_j^2/v_{thj}^2}, \quad v_{thj} \equiv (2k_B T_j/m_j)^{1/2}$$

Introducing the dummy integration variable $s = v_j/v_{thj}$, we can write n_{1j} as

$$n_{1j} = \frac{iq_j E n_{0j}}{km_j v_{thj}^2 \pi^{1/2}} \frac{1}{\zeta_j} \int_{-\infty}^{\infty} \frac{(d/ds)(e^{-s^2})}{s - \zeta_j} ds$$

where

$$\zeta \equiv \omega/kv_{thj}$$

We now define the *plasma dispersion function* $Z(\zeta)$:

$$Z(\zeta) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{e^{-s^2}}{s - \zeta} ds \quad \Im(\zeta) > 0 \tag{12.33}$$

(Why positive imaginary part???) This is a contour integral, as explained in previous sections, and analytic continuation to the lower half plane must be used if $\Im(\zeta) < 0$. $Z(\zeta)$ is a complex function of a complex argument (since ω or k usually has an imaginary part). In cases where $Z(\zeta)$ cannot be approximated by an asymptotic formula, one can do it numerically.

To express n_{1j} in terms of $Z(\zeta)$, we take the derivative with respect to ζ :

$$Z'(\zeta) = \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{e^{-s^2}}{(s - \zeta)^2} ds$$

Integration by parts yields

$$Z'(\zeta) = \frac{1}{\pi^{1/2}} \left[\frac{-e^{-s^2}}{s - \zeta} \right]_{-\infty}^{\infty} + \frac{1}{\pi^{1/2}} \int_{-\infty}^{\infty} \frac{(d/ds)(e^{-s^2})}{s - \zeta} ds$$

The first term vanishes, as it must for any well-behaved distribution function. Now we have

$$n_{1j} = \frac{i q_j E n_{0j}}{k m_j v_{thj}^2} Z'(\zeta_j)$$

Poisson's equation is

$$\epsilon \nabla \cdot \mathbf{E} = ik\epsilon_0 E = \sum_j q_j n_{1j}$$

Combining the last two equations, separating out the electron term explicitly, and defining

$$\Omega_{pj} \equiv (n_{oj} Z_j^2 e^2 / \epsilon_0 m_j)^{1/2}$$

We obtain the dispersion relation

$$k^2 = \frac{\omega_p^2}{v_{the}^2} Z'(\zeta_e) + \sum_j \frac{\Omega_{pj}^2}{v_{thj}^2} Z'(\zeta_j)$$

Electron plasma waves can be obtained by setting $\Omega_{pj} = 0$ (infinitely massive ions). Defining

$$k_D^2 = 2\omega_p^2/v_{the}^2 = \lambda_D^{-2}$$

we then obtain

$$k/k_D^2 = \frac{1}{2} Z'(\zeta_e)$$

which is the same as Equation 12.30 when f_{0e} is Maxwellian.

12.9.2 Ion Waves and Their Damping

To obtain ion waves, go back to the plasma dispersion relation and use the fact that their phase velocity ω/k is much smaller than v_{the} ; hence ζ_e is small, and we can expand $Z(\zeta_e)$ in a power series:

$$Z(\zeta_e) = i\sqrt{\pi}e^{-\zeta_e^2} - 2\zeta_e(1 - \frac{2}{3}\zeta_e^2 + \dots)$$

The imaginary term comes from the residue at a pole lying near the real s axis and represents electron Landau damping. For $\zeta_e \ll 1$, the derivative of the above gives

$$Z'(\zeta_e) = -2i\sqrt{\pi}\zeta_e e^{-\zeta_e^2} - 2 + \dots \simeq -2$$

Electron Landau damping can usually be neglected in ion waves because the slope of $f_e(v)$ is small near its peak. Replacing $Z'(\zeta_e)$ by -2 in the dispersion relation gives

$$\lambda_D \sum_j \frac{\Omega_{pj}^2}{v_{thj}^2} Z'(\zeta_j) = 1 + k^2 \lambda_D^2 \simeq 1$$

The term $k^2 \lambda_D^2$ represents the deviation from quasineutrality. ($1/k \sim L, k^2 \lambda_D^2 \sim \lambda_D^2/L^2 \ll 1$ where L is the system length scale.)

We now specialize to the case of a single ion species. Since $n_{0e} = Z_i n_{0i}$, the coefficient in the equation above is

$$\lambda_D \frac{\Omega_p^2}{v_{thi}^2} = \frac{\epsilon k_B T_e}{n_{0e} e^2} \frac{n_{0i} Z^2 e^2}{\epsilon m_i} \frac{m_i}{2k_B T_i} = \frac{1}{2} \frac{Z T_e}{T_i}$$

For $k^2 \lambda_D^2 \ll 1$, the dispersion relation becomes

$$Z' \left(\frac{\omega}{kv_{thi}} \right) = \frac{2T_i}{ZT_e}$$

Solving this equation is a nontrivial problem. Suppose we take real k and complex ω to study damping in time. Then the real and imaginary parts of ω must be adjusted so that $\Im(Z') = 0$ and $\Re(Z') = 2T_i/ZT_e$. There are in general many possible roots ω that satisfy this, all of them having $\Im(\omega) < 0$. The least damped, dominant root is the one having the smallest $|\Im(\omega)|$. Damping in space is usually treated by taking ω real and k complex. Again we get a series of roots k with $\Im(k) > 0$, representing spatial damping. However, the dominant root does not correspond to the same value of ζ_i as in the complex ω case. It turns out that the spatial problem has to be treated with special attention to the excitation mechanism at the boundaries and with more careful treatment of the electron term $Z'(\zeta_e)$.

To obtain an analytic result, we consider the limit $\zeta_i \gg 1$, corresponding to large temperature ratio $\theta \equiv ZT_e/T_i$. (???) The asymptotic expression for $Z'(\zeta_i)$ is

$$Z'(\zeta_i) = -2i\sqrt{\pi}\zeta_i e^{-\zeta_i^2} + \zeta_i^{-2} + \frac{3}{2}\zeta_i^{-4} + \dots$$

(I think this can be found from the plasma handbook; it can also be found [here](#)) If the damping is small, we can neglect the Landau term in the first approximation. The equation becomes

$$\frac{1}{\zeta_i^2} \left(1 + \frac{3}{2} \frac{1}{\zeta_i^2} \right) = \frac{2}{\theta}$$

Since θ is assumed large, ζ_i^2 is large; and we can approximate ζ_i^2 by $\theta/2$ in the second term. Thus

$$\frac{1}{\zeta_i^2} \left(1 + \frac{3}{\theta} \right) = \frac{2}{\theta}, \quad \zeta_i^2 = \frac{3}{2} + \frac{\theta}{2}$$

or

$$\frac{\omega^2}{k^2} = \frac{2k_B T_i}{m_i} \left(\frac{3}{2} + \frac{ZT_e}{2T_i} \right) = \frac{Zk_B T_e + 3k_B T_i}{m_i}$$

This is the ion wave dispersion relation with $\gamma_i = 3$, generalized to arbitrary Z .

We now substitute the above approximations back into the dispersion relation retaining the Landau term:

$$\begin{aligned} \frac{1}{\zeta_i^2} \left(1 + \frac{3}{\theta} \right) - 2i\sqrt{\pi}\zeta_i e^{-\zeta_i^2} &= \frac{2}{\theta} \\ \frac{1}{\zeta_i^2} \left(1 + \frac{3}{\theta} \right) &= \frac{2}{\theta} (1 + i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2}) \\ \zeta_i^2 &= \left(\frac{3+\theta}{2} \right)^{1/2} (1 + i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2})^{-1} \end{aligned}$$

Expanding the square root, we have

$$\zeta_i \simeq \left(\frac{3+\theta}{2} \right)^{1/2} \left(1 - \frac{1}{2}i\sqrt{\pi}\theta\zeta_i e^{-\zeta_i^2} \right)$$

The approximate damping rate is found by using the above approximation in the imaginary term:

$$-\frac{\Im(\zeta_i)}{\Re(\zeta_i)} = \frac{\Im(\omega)}{\Re(\omega)} = \left(\frac{\pi}{8} \right)^{1/2} \theta(3+\theta)^{1/2} e^{-(3+\theta)/2}$$

This asymptotic expression, accurate for large θ , shows an exponential decrease in damping with increasing θ . When θ falls below 10, this expression becomes inaccurate, and the damping must be computed from the original expression which employs the Z-function. For the experimentally

interesting region $1 < \theta < 10$, the following simple formula is an analytic fit to the exact solution:

$$-\Im(\omega)/\Re(\omega) = 1.1\theta^{7/4} \exp(-\theta^2)$$

What happens when collisions are added to ion Landau damping? Surprisingly little. Ion-electron collisions are weak because the ion and electron fluids move almost in unison, creating little friction between them. Ion-ion collisions (ion viscosity) can damp ion acoustic waves, but we know that sound waves in air can propagate well in spite of the dominance of collisions. Actually, collisions spoil the particle resonances that cause Landau damping, and one finds that the total damping is *less* than the Landau damping unless the collision rate is extremely large. In summary, ion Landau damping is almost always the dominant process with ion waves, and this varies exponentially with the ratio ZT_e/T_i .

12.10 Kinetic Effects in a Magnetic Field

When either the dc magnetic field \mathbf{B}_0 or the oscillating magnetic field \mathbf{B}_1 is finite, the $\mathbf{v} \times \mathbf{B}$ term in the Vlasov equation for a collisionless plasma must be included. The linearized equation is then replaced by

$$\frac{df_1}{dt} = \frac{\partial f_1}{\partial t} + \mathbf{v} \cdot \nabla f_1 + \frac{q}{m} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial f_1}{\partial \mathbf{v}} = -\frac{q}{m} (\mathbf{E}_1 + \mathbf{v} \times \mathbf{B}_1) \cdot \frac{\partial f_0}{\partial \mathbf{v}} \quad (12.34)$$

Resonant particles moving along \mathbf{B}_0 still cause Landau damping if $\omega/k \simeq v_{th}$, but two new kinetic effects now appear which are connected with the velocity component \mathbf{v}_\perp perpendicular to \mathbf{B}_0 . One of these is cyclotron damping, which will be discussed later; the other is the generation of cyclotron harmonics, leading to the possibility of the oscillation commonly called Bernstein waves.

Harmonics of the cyclotron frequency are generated when the particles' circular Larmor orbits are distorted by the wave fields \mathbf{E}_1 and \mathbf{B}_1 . These finite- r_L effects are neglected in ordinary fluid theory but can be taken into account to order $k^2 r_L^2$ by including the viscosity. A kinetic treatment can be accurate even for $k^2 r_L^2 = \mathcal{O}(1)$. To understand how harmonics arise, consider the motion of a particle in an electric field:

$$\mathbf{E} = E_x e^{i(kx - \omega t)} \hat{\mathbf{x}}$$

The equation of motion is

$$\ddot{x} + \omega_c^2 x = \frac{1}{m} E_x e^{i(kx - \omega t)}$$

If kr_L is not small, the exponent varies from one side of the orbit to the other. We can approximate kx by substituting the undisturbed orbit $x = r_L \sin \omega_c t$

$$\ddot{x} + \omega_c^2 x = \frac{1}{m} E_x e^{i(kr_L \sin \omega_c t - \omega t)}$$

The generating function for the Bessel function $J_n(z)$ is

$$e^{z(t-1/t)/2} = \sum_{n=-\infty}^{\infty} t^n J_n(z)$$

Letting $z = kr_L$ and $t = \exp(i\omega_c t)$, we obtain

$$e^{ikr_L \sin \omega_c t} = \sum_{n=-\infty}^{\infty} J_n(kr_L) e^{in\omega_c t}$$

$$\ddot{x} + \omega_c^2 x = \frac{q}{m} E_x \sum_{n=-\infty}^{\infty} J_n(kr_L) e^{i(\omega - n\omega_c)t}$$

The following solution can be verified by direct substitution:

$$x = \frac{q}{m} E_x \sum_{n=-\infty}^{\infty} \frac{J_n(kr_L) e^{i(\omega - n\omega_c)t}}{\omega_c^2 - (\omega - n\omega_c)^2}$$

This shows that the motion has frequency components differing from the driving frequency by multiples of ω_c , and that the amplitudes of these components are proportional to $J_n(kr_L)/[\omega_c^2 - (\omega - n\omega_c)^2]$. When the denominator vanishes, the amplitude becomes large. This happens when $\omega - n\omega_c = \pm\omega_c$, or $\omega = (n \pm 1)\omega_c$, $n = 0, \pm 1, \pm 2, \dots$; that is, when the field $\mathbf{E}(x, t)$ resonates with any harmonic of ω_c . In the fluid limit $kr_L \rightarrow 0$, $J_n(kr_L)$ can be approximated by $(kr_L/2)^n/n!$, which approaches 0 for all n except $n = 0$. For $n = 0$, the coefficient becomes $(\omega_c^2 - \omega^2)^{-1}$, which is the fluid result containing only the fundamental cyclotron frequency.

12.10.1 The Hot Plasma Dielectric Tensor

After Fourier analysis of $f_1(\mathbf{r}, \mathbf{v}, t)$ in space and time, the linearized Vlasov equation can be solved for a Maxwellian distribution $f_0(\mathbf{v})$, and the resulting expression $f_1(\mathbf{k}, \mathbf{v}, \omega)$ can be used to calculate the density and current of each species. The result is usually expressed in the form of an equivalent dielectric tensor $\vec{\epsilon}$, such that the dispersion vector $\mathbf{D} = \vec{\epsilon} \cdot \mathbf{E}$ can be used in the Maxwell's equations to calculate dispersion relations for various waves. The algebra is horrendous and therefore omitted. We quote only a restricted result valid for nonrelativistic plasmas with isotropic pressure $T_\perp = T_\parallel$ and no zero-order drifts \mathbf{v}_{0j} ; these restrictions are easily removed, but the general formulas are too cluttered for our purposes. We further assume $\mathbf{k} = k_x \hat{\mathbf{x}} + k_z \hat{\mathbf{z}}$, with $\hat{\mathbf{z}}$ being the direction of \mathbf{B}_0 ; no generality is lost by setting k_y equal to zero, since the plasma is isotropic in the plane perpendicular to \mathbf{B}_0 . The elements of $\vec{\epsilon}_R = \vec{\epsilon}/\epsilon_0$ are then

$$\begin{aligned}\epsilon_{xx} &= \\ \epsilon_{yy} &= \\ \epsilon_{zz} &= \\ \epsilon_{xz} &= \\ \epsilon_{yz} &= \\ \epsilon_{zx} &= \end{aligned}$$

where $Z(\zeta)$ is the plasma dispersion function, $I_n(b)$ is the n th order Bessel function of imaginary argument, and the other symbols are defined by

$$\begin{aligned}\omega_{ps}^2 &= n_{0s} Z_s^2 e^2 / \epsilon_0 m_s \\ \zeta_s &= (\omega + n\omega_{cs}) / k_z v_{ths} \\ \omega_{cs} &= |Z_s e B_0 / m_s| \\ v_{ths}^2 &= 2k_B T_s / m_s \\ b_s &= \frac{1}{2} k_\perp^2 r_{Ls} = k_x^2 k_B T_s / m_s \omega_{cs}^2\end{aligned}$$

The first sum is over species s , with the understanding that ω_p, b, ζ_0 , and ζ_n all depend on s , and that the \pm stands for the sign of the charge. The second sum is over the harmonic number n . The primes indicate differentiation with respect to the argument.

As foreseen, there appear Bessel functions of finite- r_L parameter b . (The change from $J_n(b)$ to $I_n(b)$ occurs in the integration over velocities.) The elements of $\vec{\epsilon}$ involving motion along $\hat{\mathbf{z}}$ contain $Z'(\zeta_n)$, which gives rise to Landau damping when $n = 0$ and $\omega/k_z \simeq v_{th}$. The $n \neq 0$ terms now make possible another collisionless damping mechanism, cyclotron damping, which occurs when $(\omega_c \pm n\omega_c)/k_z \simeq v_{th}$.

12.10.2 Bernstein Waves

Electrostatic waves propagating at right angles to \mathbf{B}_0 at harmonics of the cyclotron frequency are called Bernstein waves. The dispersion relation can be found by using the dielectric elements give in the previous section in Poisson's equation $\nabla \cdot \vec{\epsilon} \cdot \mathbf{E} = 0$. If we assume electrostatic perturbations such that $\mathbf{E}_1 = -\nabla\phi_0$, and consider waves of the form $\phi_1 = \phi_1 \exp i(\mathbf{k} \cdot \mathbf{r} - \omega t)$, Poisson's equation can be written

$$k_x^2 \epsilon_{xx} + 2k_x k_z \epsilon_{xz} + k_z^2 \epsilon_{zz} = 0$$

Note that we have chosen a coordinate system that has \mathbf{k} lying in the x-z plane, so that $k_y = 0$. We next substitute for ϵ_{xx} , ϵ_{xz} , and ϵ_{zz} from the dielectric tensor expression and express $Z'(\zeta_n)$ in terms of $Z(\zeta_n)$ with the identity

$$Z'(\zeta_n) = -2[1 + \zeta Z(\zeta_n)]$$

via integration by parts. The equation becomes

$$\begin{aligned} k_x^2 + k_z^2 + \sum_s \frac{\omega_p^2}{\omega^2} e^{-b} \zeta_0 \sum_{n=-\infty}^{\infty} I_n(b) \\ \times [k_x^2 \frac{n^2}{b} Z - 2(\frac{2}{b})^{1/2} n k_x k_z (1 + \zeta_n Z) - 2k_z^2 \zeta_n (1 + \zeta_n Z)] = 0 \end{aligned}$$

The expression in the square brackets can be simplified in a few algebraic steps to $2k_z^2[\zeta_{-n} + \zeta_0^2 Z(\zeta_n)]$ by using the definitions $b = k_x^2 v_{th}^2 / 2\omega_c^2$ and $\zeta_n = (\omega + n\omega_c)/k_z v_{th}$. Further noticing that $2k_z^2 \omega_p^2 \zeta_0 / \omega^2 = 2\omega_p^2 / v_{th}^2 \equiv k_D^2$ for each species, we can write the equation as

$$k_x^2 + k_z^2 + \sum_s k_D^2 e^{-b} \sum_{n=-\infty}^{\infty} I_n(b) [\zeta_{-n}/\zeta_0 + \zeta_0 Z(\zeta_n)] = 0$$

The term $\zeta_{-n}/\zeta_0 = 1 - n\omega_c/\omega$. Since $I_n(b) = I_{-n}(b)$, the term $I_n(b)n\omega_c/\omega$ sums to zero when n goes from $-\infty$ to ∞ ; hence, ζ_{-n}/ζ_0 can be replaced by 1. Defining $k^2 = k_x^2 + k_z^2$, we obtain the general dispersion relation for Bernstein waves:

$$1 + \sum_s \frac{k_D^2}{k_{\perp}^2} e^{-b} \sum_{n=-\infty}^{\infty} I_n(b) [1 + \zeta_0 Z(\zeta_n)] = 0$$

(A) *Electron Bernstein Waves.* Let us first consider high-frequency waves in which the ions do not move. These waves are not sensitive to small deviations from perpendicular propagation, and we may set $k_z = 0$, so that $\zeta_n \rightarrow \infty$. There is, therefore, no cyclotron damping; the gaps in the spectrum that we shall find are not caused by such damping. For large ζ_n , we may replace $Z(\zeta_n)$ by $-1/\zeta_n$. (???) The $n = 0$ term in the second sum of the above equation then cancels out, and we can divide the sum into two sums, as follows:

$$k_{\perp}^2 + \sum_s k_D^2 e^{-b} \left[\sum_{n=1}^{\infty} I_n(b) (1 - \zeta_0/\zeta_n) + \sum_{n=1}^{\infty} I_{-n}(b) (1 - \zeta_0/\zeta_{-n}) \right] = 0$$

or

$$k_{\perp}^2 + \sum_s k_D^2 e^{-b} \sum_{n=1}^{\infty} I_n(b) \left[2 - \frac{\omega}{\omega + n\omega_c} - \frac{\omega}{\omega - n\omega_c} \right] = 0$$

The bracket collapses to a single term upon combining over a common denominator:

$$1 = \sum_s \frac{k_D^2}{k_{\perp}^2} e^{-b} \sum_{n=1}^{\infty} I_n(b) \frac{2n^2 \omega_c^2}{\omega^2 - n^2 \omega_c^2}$$

Using the definitions of k_D and b , one obtains the well-known (NOT TO ME!!!) $k_z = 0$ dispersion relation

$$1 = \sum_s \frac{\omega_p^2}{\omega_c^2} \frac{2}{b} e^{-b} \sum_{n=1}^{\infty} \frac{I_n(b)}{(\omega/n\omega_c)^2 - 1}$$

We now specialize to the case of electron oscillations. Dropping the sum over species, we obtain

$$\frac{k_{\perp}^2}{k_D^2} = 2\omega_c^2 \sum_{n=1}^{\infty} \frac{e^{-b} I_n(b)}{\omega^2 - n\omega_c^2} n^2 \equiv \alpha(\omega, b)$$

... (ADD fig.7-33!!!)

To obtain the fluid limit, we repalce $I_n(b)$ by its small- b value $(b/2)^n/n!$. Only the $n = 1$ term remains in the limit $b \rightarrow 0$, and we obtain

$$1 = \frac{\omega_p^2}{\omega_c^2} \frac{2}{b} \frac{b}{2} \left(\frac{\omega^2}{\omega_c^2} - 1 \right)^{-1} = \frac{\omega_p^2}{\omega^2 - \omega_c^2}$$

or $\omega^2 = \omega_p^2 + \omega_c^2 = \omega_h^2$, which is the upper hybrid oscillation. As $k_{\perp} \rightarrow 0$, this frequency must be one of the roots. If ω_h falls between two high harmonics of ω_c , the shape of the $\omega - k$ curves changes near $\omega = \omega_h$ to allow this to occur. ...

- (B) *Ion Bernstein Waves.* In the case of waves at ion cyclotron harmonics, one has to distinguish between *pure* ion Bernstein waves, for which $k_z = 0$, and *neutralized* ion Bernstein waves, for which k_z has a small but finite value. The difference, as we have seen earlier for lower hybrid oscillations, is that finite k_z allows electrons to flow along \mathbf{B}_0 to cancel charge separations. Though the $k_z = 0$ case has already been treated in ...

13 Stability

This chapter discusses plasma equilibrium and stability, or more precisely, instability. We are interested in when, where, and how instabilities arise in plasmas.

Magnetic fields are ubiquitous in cosmological objects. General questions exist about why are they of the form that they are:

- Why does this particular form remain?
- How does it get to this form?

Anything that reorganizes magnetic field is a transport phenomenon therefore instabilities are of great interest. From hydrodynamic instabilities to plasma instabilities, we need to realize the role of magnetic field and kinetic physics in the instability process:

- Existing hydro-instability may be affected by the presence of magnetic field
- Instability may be driven by presence of magnetic fields
- Instability that does not exist in hydrodynamics may arise in the kinetic description.

Because of a multitude of free-energy sources in space plasmas, a very large number of instabilities can develop. The instabilities grow and then saturate in such a way as to reduce the free energy. Generally they can be separated into two groups:

1. Macroinstability, where the involved spatial scale is comparable to the macroscopic size (bulk scale of plasma), and can be treated with fluid theory;
2. Microinstability, where the involved spatial scale is comparable to the microscopic size (gyroradius, inertial length), and should be treated with kinetic theory.

There exists a massive catalogue of plasma instabilities.

Type	Description
Beam acoustic instability	
Bump-on-tail instability	
Buneman instability	
Cherenkov instability	
Chute instability	
Coalescence instability	
Collapse instability	

Type	Description
Counter-streaming instability	
Cyclotron instabilities	Alfvén, electron, electrostatic, ion, magnetoacoustic
Disruptive instability	
Double mission instability	
Drift instability	
Electrothermal instability	
Fan instability	
Filamentation instability	
Firehose instability	Section 14.1
Free electron maser instability	
Gyrotron instability	
Helical instability	Helix
Interchange instability	Rayleigh-Taylor, flute, ballooning, kink, sausage
Ion beam instability	
Lower hybrid drift instability	Section 13.9
Magnetic drift instability	
Modulation instability	
Non-Abelian instability	Chromo-Weibel
Pair instability	
Parker instability	Magnetic buoyancy
Peratt instability	
Pinch instability	
Tearing mode instability	Section 13.7.5
Two stream instability	Kelvin-Helmholtz, Section 13.5
Weak beam instability	
Weibel instability	
Z-pinch instability	Bennett pinck

The basic methodology of examining instabilities

- Take the equations
- Linearize about an equilibrium solution
- Add some perturbations and see what happens
 - Look for *normal mode* solutions $e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}$ and find a relationship between the growth rate $\Im(\omega)$ and the wavenumbers of the disturbance \mathbf{k} and the parameters of the problem (i.e. dispersion relation).
 - Use the *MHD energy principle*: calculate δW , the change in potential energy associated with the disturbance, and look for disturbances with $\delta W < 0$ (the existence of any one means unstable; need $\delta W > 0 \forall$ types of disturbances).

13.1 Analogy of the Energy Principle

First, let us have an intuitive inspection from the energy principle. In classical mechanics, we learn that motion acts to lower the energy in the system. For the simplest Rayleigh-Taylor instability between two fluids with different density, we can easily decide if there is an instability by checking the total potential energy change. In a current-carrying plasma, similar motions can be developed. The energy can be expressed as

$$W = \frac{1}{2}LI^2$$

where L is the inductance and I is the current. The magnetic flux can be written as

$$\Phi = LI$$

so the energy can also be written as

$$W = \frac{\Phi^2}{2L}$$

The relation between magnetic flux and current is similar to charge and voltage, $Q = CV$, where C is the electric capacity. In a system where magnetic flux Φ is conserved (e.g. ideal MHD), plasma lowers its potential by increasing its inductance L . Check (Bellan 2008) P315.

1. current loop, hoop force, increase area;
2. sausage mode;
3. current wire into helix.

13.2 Implication of Single Particle Motion on Instabilities

This section provides some qualitative understanding of instabilities in plasma. Thinking of instabilities from the prospective of single particle motion provides us the physical intuition of the causes and development of instabilities from a very basic level.

13.2.1 Equilibrium stability of a plasma from drifts

A key problem in nuclear fusion is to confine plasma. In the first stage, we need current loops around a torus tube (ADD FIGURE!) to provide a toroidal magnetic field \mathbf{B}_t along the torus. Let us take a look at a torus cut. (ADD FIGURE!) ∇B_0 is pointing towards the inside. \mathbf{B}_0 points outside the plane, so the gradient-curvature drift will lead electrons upward and ions downward, which in turn creates a \mathbf{E} field pointing upwards. $\mathbf{E} \times \mathbf{B}$ drift then will lead both electrons and ions towards the outer boundary. Eventually we will lose the plasma. One way to fix this is to add \mathbf{J}_t , an internal plasma current in the toroidal direction to generate a poloidal

\mathbf{B}_p , so that the magnetic field is stronger on the boundary than that at the center. In this way the total magnetic field becomes a helix.

Consider the poloidal magnetic field generated by \mathbf{J}_t . One step further, our question is: will the plasma tube be stable under infinitesimal perturbations to the ideal cylinder configuration? The answer is no. Imagine a small perturbation shown as in Figure 13.1 b. ∇B points from weak B to strong B regions. On the convex side, the gradient-curvature drift will lead ions to the left and electrons to the right, which in turn generates an electric field pointing left to right. Thereafter the $\mathbf{E} \times \mathbf{B}$ drift will further drag the plasma to the convex region, and the whole system can never return to equilibrium. This the current-carrying plasma instability is called *kink instability*. The situation described here is sometimes referred to as *linear kink instability*. The kink mode can carry currents. Another similar mode is the *sausage instability* as shown in Figure 13.1. Also note that a sufficiently strong B_z (not poloidal/toroidal component) can stabilize these instabilities.

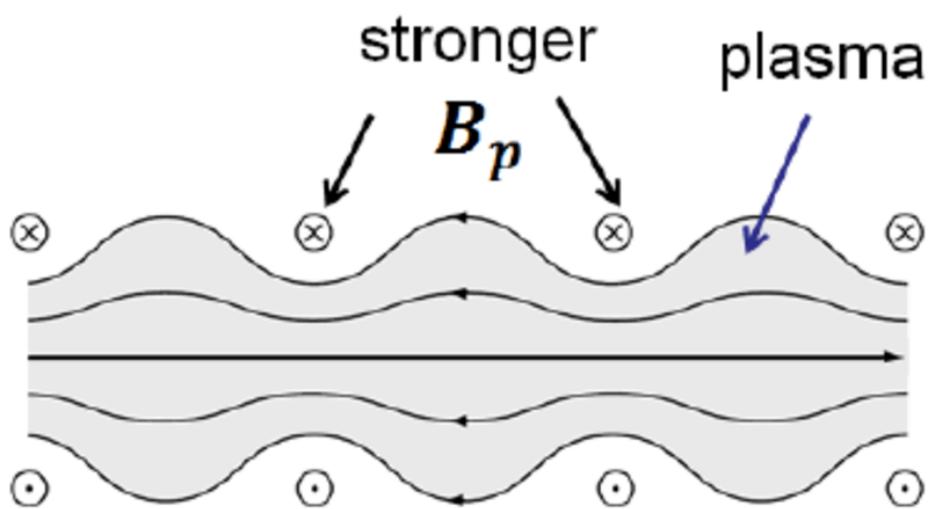
Another famous instability is the Rayleigh-Taylor instability. In fluid dynamics, Rayleigh-Taylor instability happens due to gravity. Here in plasma physics, the role of gravity force is replaced by the electromagnetic force. Imagine a situation where plasma are located at $z > 0$ region, below which is a vacuum space. There is a B field pointing outside the plane while ∇B is pointing upwards ($B_{\text{up}} > B_{\text{down}}$). Due to gradient drift on the boundary, ions will move to the left while electrons will move to the right, where a E field pointing left to right is created. Thus the $\mathbf{E} \times \mathbf{B}$ drift will lead plasma from upper region to lower region, and eventually breaks the interface. (Actually I have some questions for this figure: it seems to me that it is impossible to decide which part of the interface is changing first?)

13.2.2 Stability of magnetic mirror in the scope of single particle motion

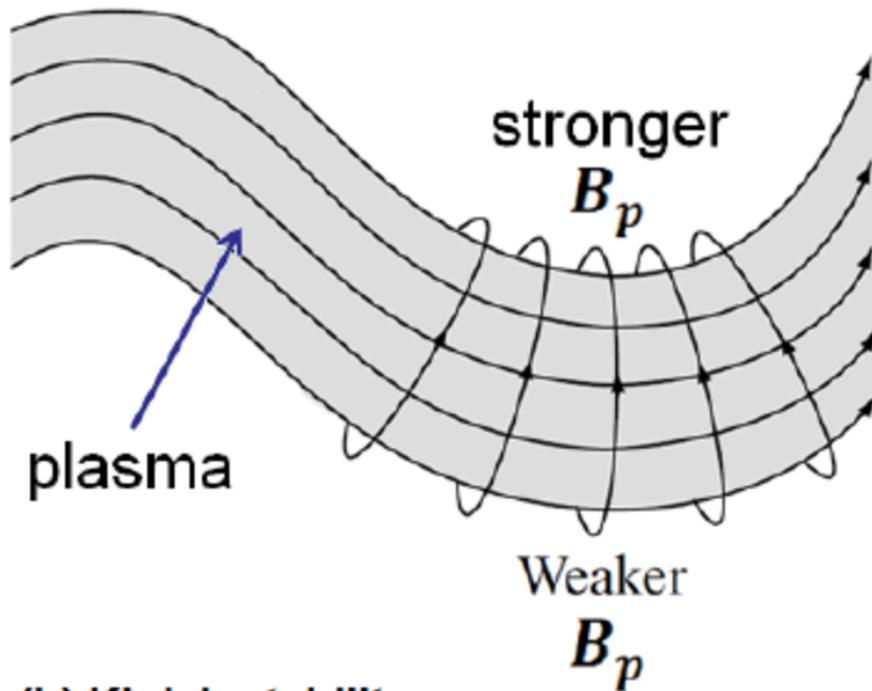
We can deduce the stability of magnetic mirror by assuming a initial small perturbation along the boundary. In the center cross-section cut, first there is a centrifugal force pointing outward, which will cause electrons drifting to one way and ions drifting to the other way. The charge separation will generate an electric field. The $\mathbf{E} \times \mathbf{B}$ drift will then pull the plasma further out if there's a ripple, which will lead to instability. Several names describe the same thing: flute instability, R-T instability, interchangeable instability, gravitational instability and so on. This instability propagates at Alfvén speed.

In general, we can define two configuration categories: the unstable situation, where \mathbf{B} has a “bad” unfavorable curvature, and the stable situation, where \mathbf{B} has a “good” favorable curvature. This depends on whether or not the magnetic pressure in the vacuum is stronger than that inside the plasma. In a basic magnetic mirror, the plasma on the center boundaries are unstable, while those around the coil curvature are stable.

A famous Russian scientist Ioffe introduced conducting bars around the mirror to create an *absolute minimum B-geometry*, where in any point away from the center the B field is stronger. This indeed supresses the R-T instability, but the whole system is, unfortunately, still unstable



(a) **Sausage instability**



(b) **Kink instability**

Figure 13.1: Sausage and kink modes.

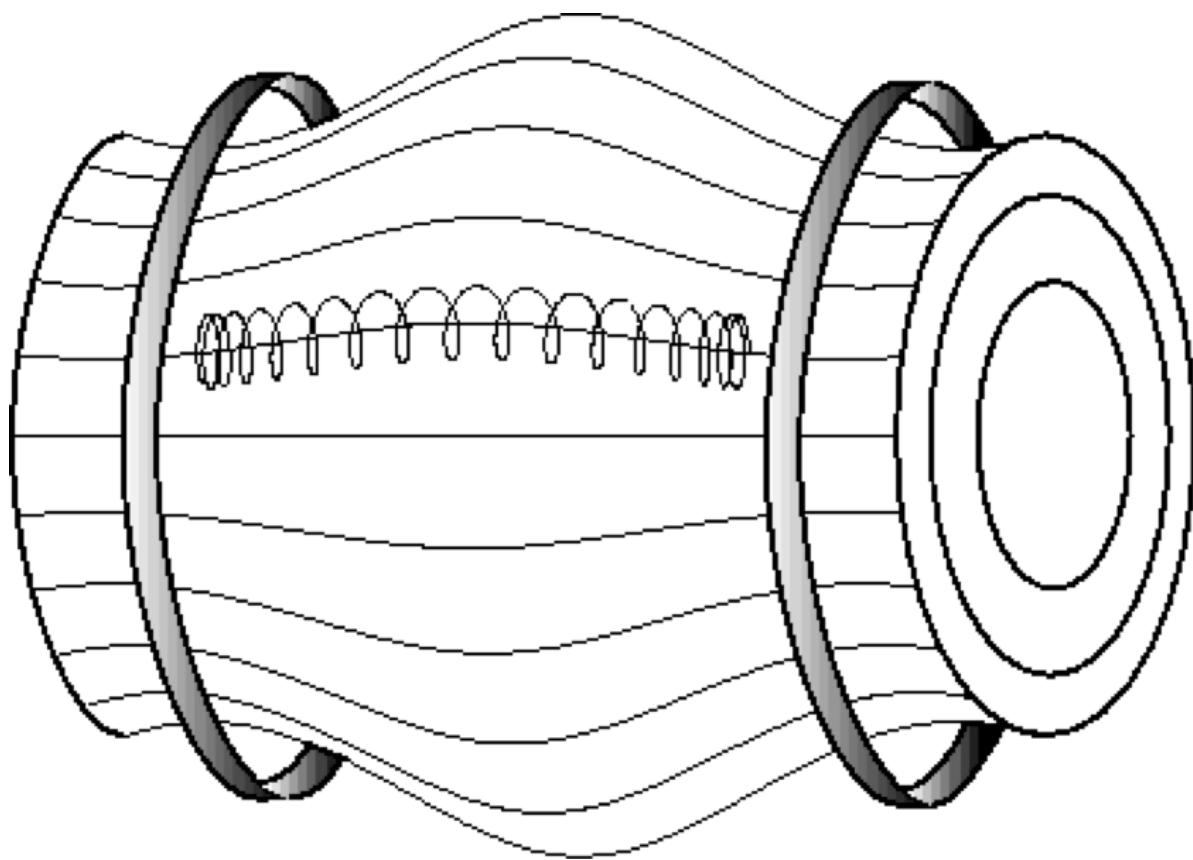


Figure 13.2: Basic magnetic mirror configuration.

due to microinstabilities caused by lost cone distribution. The inverted population (more high speed particles than low speed ones) will lead to instabilities. Later scientists came up with the idea of adding cold gas to modify the distribution, but the cold gas injection procedure eventually kills the mirror configuration.

13.2.3 Stability in the Tokamak

In a classical Tokamak geometry, the poloidal and toroidal magnetic field together created a spiral around the torus.

$$\mathbf{B} = \mathbf{B}_T + \mathbf{B}_P$$

$$B = \sqrt{B_T^2 + B_P^2} \approx B_T$$

The field strength goes down as R increases, which implies that the inner semi-tube is in good curvature and the outer semi-tube is in bad curvature. By plotting mechanical potential $= \mu B$ along the B-line as a function of θ , we can see that there are bumps and valleys. Particles with low v_{\parallel} are trapped, and those with large v_{\parallel} are transit particles. Tokamak is an *average minimum B geometry*, because particles spend longer time on hills (stable region) and less time in valleys (unstable region). This geometry is not as robust as Ioffe bar magnetic mirror, since some particles with small v_{\parallel} are always trapped in bad curvature region.

See more in (F. F. Chen 2016), Third Edition, Chapter Application of Plasmas.

13.3 Two-Stream Instability

The general procedure of obtaining the dispersion relation for electrostatic wave is:

1. Get the linear equations from governing equations.
2. Get the relation between linearized velocity and perturbed electric field.
3. Get the relation between linearized density and perturbed electric field.
4. Get the current response to the perturbed electric field.
5. Get the dielectric tensor from Maxwell's equation. Let $|\epsilon| = 0$. we finally obtain the dispersion relation. For an isotropic case, the dielectric tensor shrinks to a scalar, so we simply have $\epsilon = 0$.

Assume the simple equilibrium state in 1D (static and “cold” ions, “cold” electrons): $m_i = \infty, E_0 = B_0 = 0, p_i = p_e = 0, n_i = n_0, T_e = T_i = 0$. Whenever we say “cold” for plasma, it does not mean that the plasma is at absolute zero degree. This only means that we are

considering a situation where the kinetic pressure plays no roles in the dispersion relation. This is also a *non-magnetized plasma* because $\mathbf{B}_0 = 0$. The variables including perturbations are

$$\begin{aligned} n_e &= n_0 + n_1 \quad n_0 \gg n_1 \\ v_e &= v_0 + v_1 \quad v_0 \gg v_1 \\ E &= \mathcal{E}_0 + E_1 \\ n_1(x, t) &= \tilde{n}_1 e^{-i\omega t + ikx} \end{aligned}$$

The electron continuity and the momentum equations read

$$\begin{aligned} \frac{\partial}{\partial t}(n_0 + n_1) + \frac{\partial}{\partial x}[(n_0 + n_1)(v_0 + v_1)] &= 0 \\ \frac{\partial}{\partial t}(v_0 + v_1) + (v_0 + v_1)\frac{\partial(v_0 + v_1)}{\partial x} &= -\frac{e}{m_e}[(E_0 + E_1) + (\mathbf{v}_0 + \mathbf{v}_1) \times (\mathbf{B}_0 + \mathbf{B}_1)_x] \end{aligned}$$

Again, for electrostatic waves, $\mathbf{B}_1 = 0$. Neglecting high order terms, we get the linearized equations

$$\begin{aligned} \frac{\partial n_1}{\partial t} + \frac{\partial}{\partial x}(n_0 v_1) + \frac{\partial}{\partial x}(n_1 v_0) &= 0 \\ \frac{\partial v_1}{\partial t} + v_0 \frac{\partial v_1}{\partial x} &= -\frac{e}{m_e} E_1 \end{aligned}$$

Assume wave-like perturbations $e^{ikx-i\omega t}$ as in the Vlasov theory, from the linearized momentum equation we have

$$\begin{aligned} (-i\omega + ikv_0)v_1 &= -\frac{e}{m_e} E_1 \\ \Rightarrow v_1 &= \frac{\frac{e}{m_e} E_1}{i(\omega - kv_0)}. \end{aligned}$$

Substituting into the linear continuity equation, we get

$$\begin{aligned} -i\omega n_1 + ikn_0 v_1 + ikn_1 v_0 &= 0 \\ \Rightarrow n_1 &= \frac{kn_0 v_1}{\omega - kv_0} = \frac{kn_0}{\omega - kv_0} \frac{\frac{e}{m_e} E_1}{i(\omega - kv_0)}. \end{aligned}$$

This is the density perturbation in response to electric perturbation E_1 in 2-fluid theory.

Then we can use the Poisson's equation to generate the dielectric function

$$\begin{aligned} \nabla \cdot (\epsilon_0 \mathbf{E}_1) + en_1 &\equiv \nabla \cdot (\epsilon \mathbf{E}_1) = 0 \\ ik\epsilon_0 \tilde{E}_1 + e\tilde{n}_1 &= ik\epsilon \tilde{E}_1 \\ \Rightarrow \epsilon &= \epsilon_0 \left[1 - \frac{\omega_{pe}^2}{(\omega - kv_0)^2} \right] \end{aligned}$$

which is the same as the result given by Vlasov theory (Y.Y. Problem Set #6.2). The advantage of using 2-fluid method is that we do not need to consider velocity space, which simplifies the derivation.

If we have two streams of electron described by $g(v)$ as

$$g(v) = \frac{1}{2}[\delta(v - v_1) + \delta(v - v_2)]$$

with the oscillation frequency ω_{p1}, ω_{p2} and number density $n_{p1,1}, n_{p2,1}$ respectively. In consideration of linear superposition property, we expect the dielectric function to be

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{p1}^2}{(\omega - kv_1)^2} - \frac{\omega_{p2}^2}{(\omega - kv_2)^2}$$

If $g(v)$ is a continuous distribution in general, $g(v) = \sum_j g_j(v)$, then

$$\begin{aligned} \frac{\epsilon}{\epsilon_0} &= 1 - \int_{-\infty}^{\infty} \sum_j \frac{\omega_{p,j}^2 g_j(v) dv}{(\omega - kv)^2} \\ &= 1 - \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} \frac{g(v) dv}{(v - \omega/k)^2}. \end{aligned}$$

Note that $\delta'(x) = x^{-1}\delta(x)$. Here we reconstruct the result of Vlasov theory from 2-fluid theory. The equivalence of the two approaches is explored more thoroughly in later section Fluid Descriptions of Kinetic Modes (ADD LINK!).

Then what happens if ion motion is included? We still have “cold” ions at rest in equilibrium but now with ion perturbations in density. The Poisson’s equation should include ion density perturbation

$$\begin{aligned} \nabla \cdot (\epsilon_0 \mathbf{E}_1) &= \sum_{j=i,e} q_j n_{1j} \\ \Rightarrow \frac{\epsilon}{\epsilon_0} &= 1 - \frac{\omega_{pe}^2}{k^2} \int_{-\infty}^{\infty} dv \frac{\partial g_e(v)/\partial v}{v - \omega/k} - \frac{\omega_{pi}^2}{k^2} \int_{-\infty}^{\infty} dv \frac{\partial g_i(v)/\partial v}{v - \omega/k} \end{aligned}$$

In the simplest equilibrium case, $g_e(v) = \delta(v - v_0)$, $g_i(v) = \delta(v)$

$$\frac{\epsilon}{\epsilon_0} = 1 - \frac{\omega_{pe}^2}{(\omega - kv_0)^2} - \frac{\omega_{pi}^2}{\omega^2}$$

Let $\epsilon = 0$, we get the dispersion relation $\omega = \omega(k)$. An example dispersion relation and dielectric function property are shown in Figure 13.3 and Figure 13.4, respectively. Note that if you have a real wave number k , you will get a pair of conjugate ω , one of which that lies

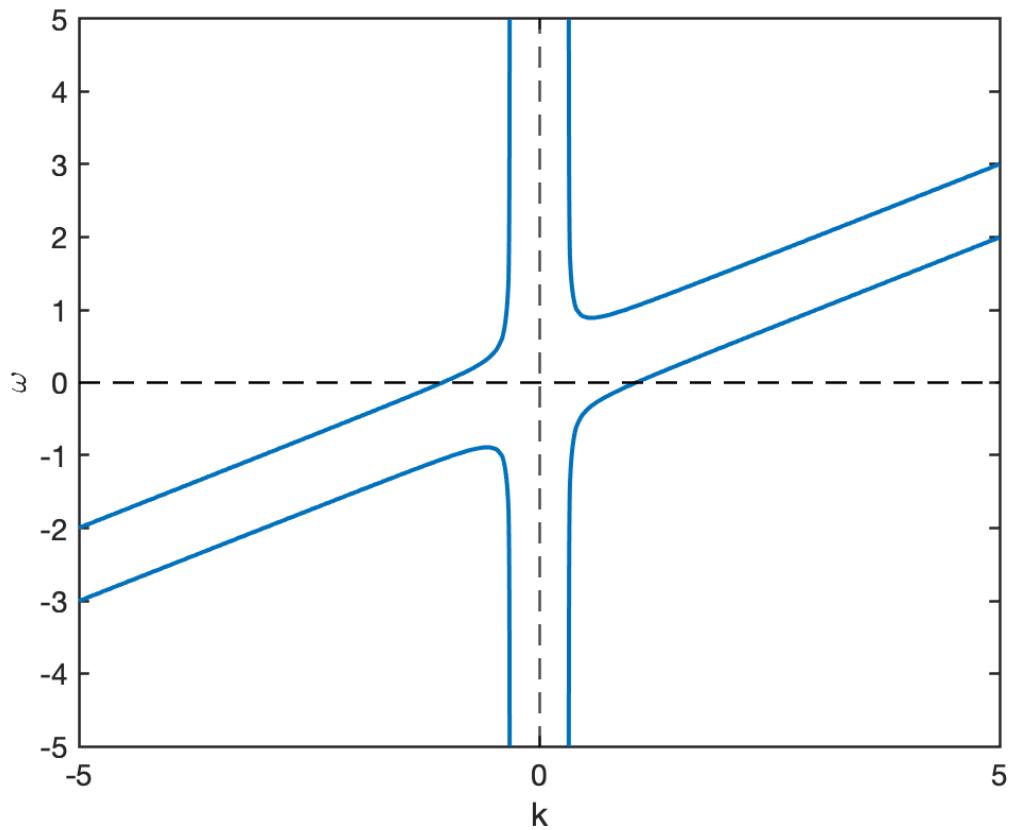


Figure 13.3: 2-stream dispersion relation.

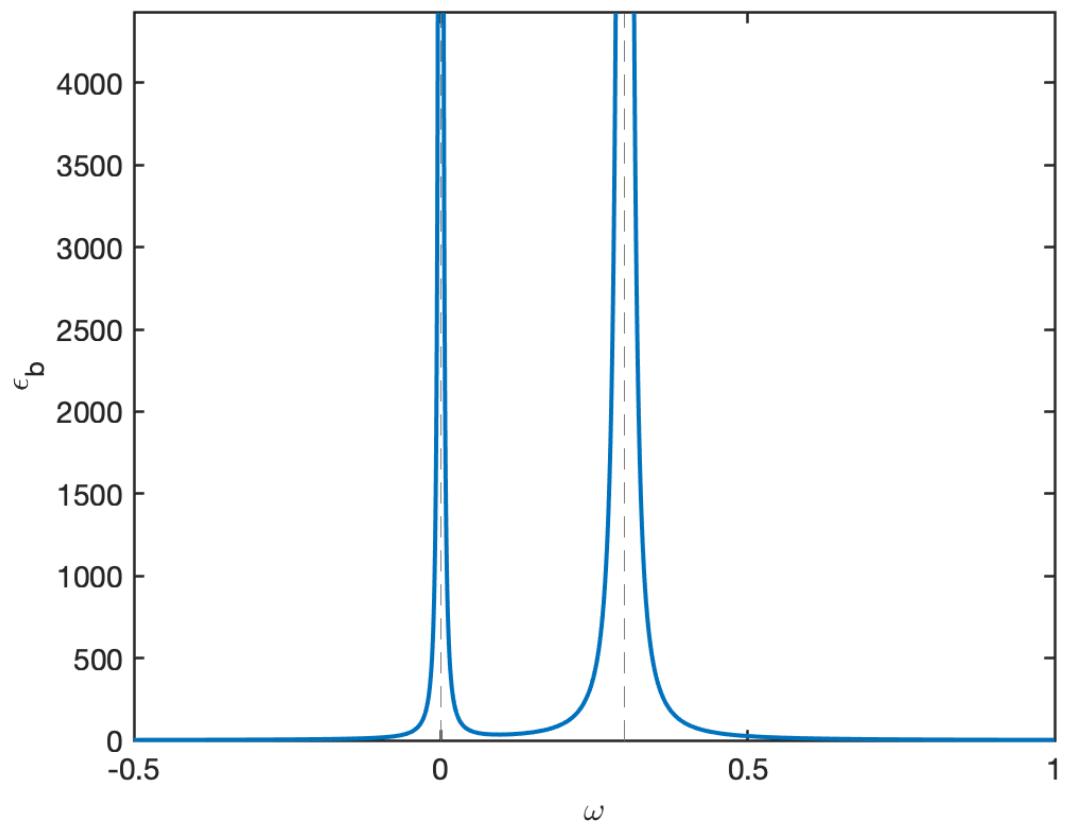


Figure 13.4: 2-stream dielectric function.

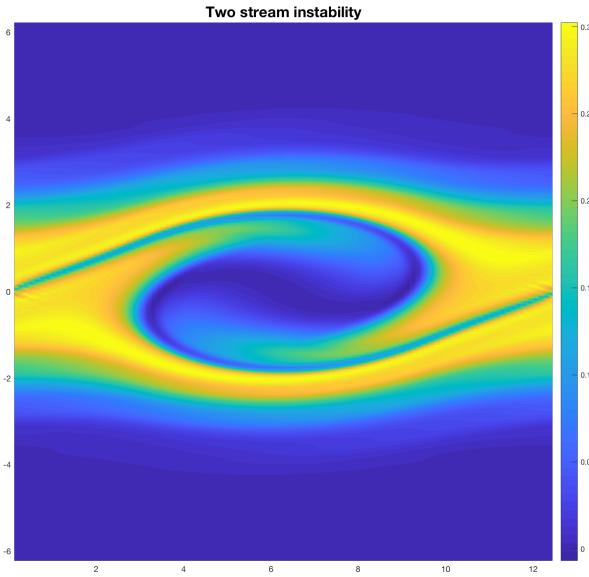


Figure 13.5: 2-stream velocity space distribution.

between 0 and kv_0 is an unstable mode. This will exhibit 2-stream instability as shown by the velocity space distribution in Figure 13.5.

$$\begin{aligned}\epsilon(\omega, k)/\epsilon_0 &= 1 - \frac{\omega_{pe}^2}{(\omega - kv_0)^2} - \frac{\omega_{pi}^2}{\omega^2} = 0 \\ 1 &= \frac{\omega_{pe}^2}{(\omega - kv_0)^2} + \frac{\omega_{pi}^2}{\omega^2}\end{aligned}$$

Let $\omega/\omega_{pe} = z$, $\frac{kv_0}{\omega_{pe}} = \lambda$, such that z and λ are dimensionless numbers. Let the right-hand side be $f(z)$, then

$$f(z) = \frac{1}{(z - \lambda)^2} + \frac{\omega_{pi}^2/\omega_{pe}^2}{z^2} = \frac{z^2 + (z - \lambda)^2 \omega_{pi}^2/\omega_{pe}^2}{(z - \lambda)^2 z^2}$$

We can plot f to find the threshold of k when the instability will happen. See Y.Y's Problem Set 6.3 for details.

13.4 Rayleigh-Taylor Instability

The Rayleigh-Taylor instability is probably the most important MHD instability. It is also called *gravitational instability*, *flute instability* or more generally, *interchange instability*. In ordinary hydrodynamics, a Rayleigh-Taylor instability arises when one attempts to support a heavy fluid on top of a light fluid: the interface becomes “rippled”, allowing the heavy fluid to fall through the light fluid. In plasmas, a Rayleigh-Taylor instability can occur when a dense plasma is supported against gravity by the pressure of a magnetic field.

The situation would not be of much interest or relevance in its own right, since actual gravitational forces are rarely of much importance in plasmas. However, in curved magnetic fields, the centrifugal force on the plasma due to particle motion along the curved field-lines acts like a “gravitational” force (Section 7.1.3). For this reason, the analysis of the Rayleigh-Taylor instability provides useful insight as to the stability properties of plasmas in curved magnetic fields. Rayleigh-Taylor-like instabilities driven by actual field curvature are the most virulent type of MHD instability in non-uniform plasmas.

Here we use a 2-fluid model and a so-called “diffuse boundary” model (F. F. Chen 2016) to describe it mathematically. Recall the structure of magnetic mirror: we have curved magnetic field lines and high density plasma at the center. From the discussion in Section 13.2.2, we know that the central part of magnetic mirror is unstable for Rayleigh-Taylor instability because of centrifugal force. Let us simplify the scenario and study the problem in Cartesian coordinates. The centrifugal force is irrelevant to particle charge and proportional to particle mass, so both ions and electrons have the same acceleration due to it. Let us replace the centrifugal force with gravity \mathbf{g} . In Figure 13.6, there are high density plasma on top and low density plasma on the bottom, with a distribution $\partial n_o / \partial x < 0$.

13.4.1 2-Fluid Diffuse Boundary Model

This section is similar to Section 6.7 in (F. F. Chen 2016).

The continuity and momentum equations are:

$$\begin{aligned}\frac{\partial n_j}{\partial t} + \nabla \cdot (n_j \mathbf{v}_j) &= 0 \\ \frac{\partial \mathbf{v}_j}{\partial t} + \mathbf{v}_j \cdot \nabla \mathbf{v}_j &= \frac{q_j}{m_j} (\mathbf{E} + \mathbf{v}_j \times \mathbf{B}) - \frac{\nabla P_j}{n_j m_j} + \mathbf{g}\end{aligned}$$

where $j = e^-, i^+$ for electrons and ions.

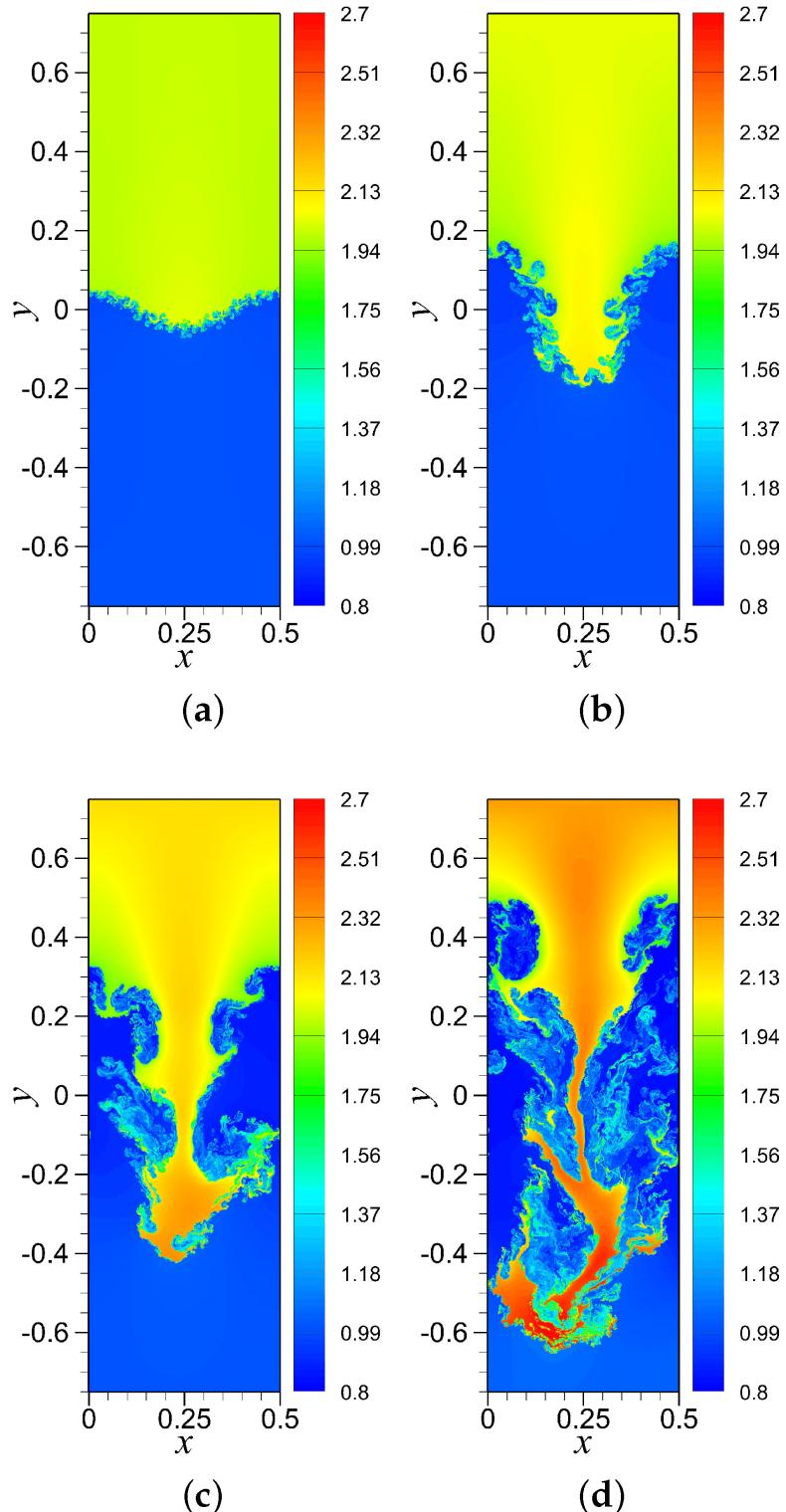


Figure 13.6: Simultaion of Rayleigh-Taylor instability (Rahman and San 2019).

Assume an one-dimensional case

$$\begin{aligned}
T_e = T_i = 0 \Rightarrow P_e = n_e k_B T_e = 0, \quad P_i = n_i k_B T_i = 0 \\
n_0 = n_0(x), \frac{\partial n_0}{\partial x} < 0 \text{ (nonuniform plasma)} \\
\mathbf{g} = g\hat{x}, g = \text{const.} > 0 \\
\mathbf{B}_0 = B_0\hat{z}, \quad B_0 = \text{const.}, \quad \mathbf{E}_0 = 0 \text{ (no gradient/curvature drift)}
\end{aligned}$$

Note that there is no diamagnetic current if $P_e = P_i = 0$ (no electric field so no current along \mathbf{B}_0 ?):

$$\begin{aligned}
\mathbf{J}_{de} \times \mathbf{B}_0 &= -\nabla P_e = 0 \\
\mathbf{J}_{di} \times \mathbf{B}_0 &= -\nabla P_i = 0 \\
\Rightarrow \mathbf{J}_{de} &= \mathbf{J}_{di} = 0
\end{aligned}$$

Instability arises when an equilibrium state is violated. What is the force that balances the gravity? It turns out to be the Lorentz force $\mathbf{v} \times \mathbf{B}$ term: the separation of electrons and ions creates currents, and currents lead to force.

ADD FIGURE!

In equilibrium, $\frac{\partial}{\partial t} = 0$, $\frac{\partial}{\partial y}[n_{0j}(x)v_{0j}] = 0$, $v_{0j} = \text{const.}$,

$$\begin{aligned}
\frac{\partial n_{0j}}{\partial t} + \nabla \cdot (n_{0j} \mathbf{v}_{0j}) &= 0 \\
\frac{q_j}{m_j} \mathbf{v}_j \times \mathbf{B}_0 + \mathbf{g} &= 0 \\
\Rightarrow \begin{cases} \mathbf{v}_i = \frac{gm_i}{q_i B_0}(-\hat{y}) = -\frac{g}{\Omega_i} \hat{y} = -\hat{y} v_{0i} \\ \mathbf{v}_e = \frac{gm_e}{q_e B_0}(\hat{y}) = \frac{g}{\Omega_e} \hat{y} = \hat{y} v_{0e} \approx 0 (v_{0e} \ll v_{0i}) \end{cases}
\end{aligned}$$

where Ω_i , Ω_e are the gyro-frequency for ions and electrons respectively.

Now, we introduce an electrostatic perturbation on this equilibrium state ($\mathbf{B}_1(\mathbf{x}, t) = 0$, $\nabla \times \mathbf{E}_1 = -\frac{\partial \mathbf{B}_1}{\partial t} = i\omega \mathbf{B}_1 = 0$, \mathbf{E}_1 can be written as a gradient of a scalar potential)

$$\mathbf{E}_1(\mathbf{x}, t) = -\nabla \Phi_1 = -\nabla[\phi_1(x)e^{ik_y y - i\omega t}]$$

In addition, we adopt the so-called “local approximation”, i.e. we assume $\partial \phi_1 / \partial x = 0$, $\frac{\partial}{\partial x}[E_1, \mathbf{v}_1, n_1] = 0$. This is a very drastic assumption that greatly simplifies the problem but cannot be justified. This assumption is commonly used in many textbooks, both explicitly and implicitly (e.g. (Bellan 2008) used this to treat universal instability. Remember in solving the Vlasov equations, we integrate along the unperturbed orbits, which also requires this assumption.)

In this case,

$$\mathbf{E}_1 = 0\hat{x} + E_{1y}\hat{y} = \hat{y}\tilde{E}_{1y}e^{ik_y y - i\omega t}$$

where $\tilde{E}_{1y} = -ik_y\phi_1$ is a constant.

Linearization:

$$\begin{aligned} \frac{\partial}{\partial t}(n_0 + n_1) + \nabla \cdot [(n_0 + n_1)(\mathbf{v}_0 + \mathbf{v}_1)] &= 0 \\ \frac{\partial}{\partial t}(\mathbf{v}_0 + \mathbf{v}_1) + (\mathbf{v}_0 + \mathbf{v}_1) \cdot \nabla(\mathbf{v}_0 + \mathbf{v}_1) &= \frac{q}{m}[\mathbf{E}_0 + \mathbf{E}_1 + (\mathbf{v}_0 + \mathbf{v}_1) \times (\mathbf{B}_0 + \mathbf{B}_1)] + \mathbf{g} \end{aligned}$$

Getting rid of the equilibrium and high-order terms, we have (Notice that \mathbf{g} does not even appear here! In MHD, it does, in a very explicit way.)

$$\begin{aligned} i(k_y v_{0y} - \omega)n_1 &= -n_0 ik_y v_{1y} - v_{1x} \frac{\partial n_0}{\partial x} \\ \frac{d}{dt}\mathbf{v}_1 &= i(k_y v_{0y} - \omega)\mathbf{v}_1 = \frac{q}{m}(\mathbf{E}_1 + \mathbf{v}_1 \times \mathbf{B}_0) \end{aligned}$$

Now, from the linearized momentum equation, we can get the x and y components of perturbed velocity; intuitively, you can guess the expression:

$$\begin{aligned} v_{1,ix} &= \frac{E_{1y}}{B_0}, \quad v_{1,ex} = \frac{E_{1y}}{B_0} \\ v_{1,iy} &= \frac{1}{\Omega_i} \frac{d}{dt} \left(\frac{E_{1y}}{B_0} \right) = \frac{i(k_y v_{0i} - \omega)}{\Omega_i} \left(\frac{E_{1y}}{B_0} \right), \quad v_{1,ey} = \frac{1}{\Omega_e} \frac{d}{dt} \left(\frac{E_{1y}}{B_0} \right) \approx 0 \end{aligned}$$

where in the x direction, it is the $\mathbf{E} \times \mathbf{B}$ drift, and in the y direction, it is the polarization drift.

From the linearized continuity equation

$$\begin{aligned} i(k_y v_{0,yi} - \omega)n_{1i} &= -n_0 ik_y v_{1,yi} - v_{1,xi} \frac{\partial n_0}{\partial x} \\ -i\omega n_{1e} &= -v_{1,xe} \frac{\partial n_0}{\partial x} \end{aligned}$$

Then we can get $n_{1e} = n_{1e}(E_{1y})$, $n_{1i} = n_{1i}(E_{1y})$. Setting $n_{1e} = n_{1i}$, we have the dispersion relation

$$\omega(\omega - k_y v_{0i}) = g \frac{1}{n_0} \frac{\partial n_0}{\partial x}$$

When $k_y \rightarrow 0$,

$$\omega^2 = g \frac{1}{n_0} \frac{\partial n_0}{\partial x} < 0 \Rightarrow \text{instability!}$$

Let's think about this 2-fluid approach for a while. Apparently, we cannot treat a sharp boundary, namely $\frac{\partial n_0}{\partial x} = \delta(x)$, with exactly the same equations. However, it's quite a surprise that MHD approach can easily do that, as we will see in the next section.

13.4.2 Single fluid MHD method

In equilibrium, $\mathbf{g} = \hat{x}g$, $\mathbf{B}_0 = \hat{z}B_0(x)$, $\mathbf{U}_0 = 0$, $\mathbf{E}_0 = 0$, $\rho_0(x)$, $p_0(x)$.

$$\frac{\partial \rho_0}{\partial t} + \nabla \cdot (\rho_0 \mathbf{U}_0) = 0$$

$$0 = \frac{1}{\mu_0} \left[-\frac{1}{2} \frac{\partial}{\partial x} (B_0^2) \right] - \frac{\partial}{\partial x} p_0(x) + \rho_0(x)g$$

Note that there's a difference between cases where different pressure is dominant. For example, in z-pinch the magnetic pressure is dominant, while in a laser pulse, the thermal pressure is usually dominant.

Assume perturbations of the form

$$p_1(\mathbf{x}, t) = p_1(x)e^{ik_y y - i\omega t}$$

$$\rho_1(\mathbf{x}, t) = \rho_1(x)e^{ik_y y - i\omega t}$$

$$\mathbf{U}_1 = \frac{\partial \boldsymbol{\xi}_1}{\partial t} = -i\omega \boldsymbol{\xi}_1$$

where $\boldsymbol{\xi}_1$ is the displacement.

We can calculate each linear term:

$$[(\mathbf{B}_0 + \mathbf{B}_1) \cdot \nabla](\mathbf{B}_0 + \mathbf{B}_1) \approx (\mathbf{B}_0 \cdot \nabla)\mathbf{B}_1 + (\mathbf{B}_1 \cdot \nabla)\mathbf{B}_0 = (B_0(x) \frac{\partial}{\partial z})\mathbf{B}_1$$

$$B^2 = (\mathbf{B}_0 + \mathbf{B}_1) \cdot (\mathbf{B}_1 + \mathbf{B}_1) \approx 2\mathbf{B}_0 \cdot \mathbf{B}_1$$

$$\mathbf{U} = \mathbf{U}_0 + \mathbf{U}_1 = \mathbf{U}_1$$

The tension term has no x or y component, so we can just ignore it. Then the linearized momentum equation can be written as

$$\rho_0 \frac{\partial \mathbf{u}_1}{\partial t} = \rho_0 \frac{\partial^2 \boldsymbol{\xi}_1}{\partial t^2} = -\nabla \left(\frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{\mu_0} \right) - \nabla p_1 + \rho_1 \mathbf{g}$$

which can be separated into two scalar equations

$$-\rho_0 \omega^2 \xi_{1x} = -\frac{\partial}{\partial x} \left(\frac{\mathbf{B}_1 \cdot \mathbf{B}_1}{\mu_0} + p_1 \right) + \rho_1 g$$

$$-\rho_0 \omega^2 \xi_{1y} = -ik_y \left(\frac{\mathbf{B}_1 \cdot \mathbf{B}_1}{\mu_0} + p_1 \right)$$

Assume incompressibility

$$\nabla \cdot \mathbf{u} = 0 \Rightarrow \nabla \cdot \mathbf{u}_1 = 0, \quad \nabla \cdot \boldsymbol{\xi}_1 = \frac{\partial \xi_{1x}}{\partial x} + ik_y \xi_{1y} = 0$$

The linearized continuity equation (Section 10.13) yields

$$\rho_1 = -\nabla \cdot (\rho_0 \boldsymbol{\xi}_1) = \xi_{1x} \frac{\partial}{\partial x} \rho_0$$

Combining the last four equations, we have

$$-\rho_0 \omega^2 \xi_{1x} = -\frac{\partial}{\partial x} \left[\rho_0 \omega^2 \frac{1}{k_y^2} \frac{\partial \xi_{1x}}{\partial x} \right] - g \xi_{1x} \frac{\partial \rho_0}{\partial x} \quad (13.1)$$

This is the governing equation for the Rayleigh-Taylor instability, which is the same as Eq.(10.15) in (Bellan 2008). Note that here we have no assumption on the x-dependence; if we simply use the local approximation as before, this immediately gives you the identical result.

To treat the sharp boundary problem, we assume

$$\rho_0 = \begin{cases} \text{const.} & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases}$$

Then for $x < 0$,

$$\begin{aligned} \frac{\partial^2 \xi_{1x}}{\partial x^2} - k_y^2 \xi_{1x} &= 0 \\ \Rightarrow \xi_{1x} &= A e^{k_y x} + B e^{-k_y x} \end{aligned}$$

and for $x > 0$,

$$\begin{aligned} \frac{\partial^2 \xi_{1x}}{\partial x^2} - k_y^2 \xi_{1x} &= 0 \\ \Rightarrow \xi_{1x} &= C e^{k_y x} + D e^{-k_y x} \end{aligned}$$

The coefficient B and C must be zero because of infinite field requirement. Due to continuity at $x = 0$, we set

$$A = D = \xi_0$$

The density profile obeys

$$\frac{\partial \rho_0}{\partial x} = -\rho_0 \delta(x)$$

Integrating the governing Equation 13.1 from $x = 0^-$ to $x = 0^+$ yields

$$\begin{aligned} -\frac{\rho_0 \omega^2}{k_y^2} \xi_{1x} \Big|_{x=0^-}^{x=0^+} - g \xi_{1x}(-\rho_0) &= 0 \\ \Rightarrow \omega^2 &= k_y g \end{aligned}$$

Therefore the growth rate is $\gamma = \Im(\omega) = \sqrt{k_y g}$. You may realize that $\mathbf{k} \cdot \mathbf{B}_0 = 0$ here, so this magnetic stabilizing term vanishes in the dispersion relation.

13.4.3 2-fluid sharp boundary model

Now let's go back and see if we can treat the sharp boundary problem with 2-fluid model. This is actually not easy: it is first solved by S.Chandrasekhar in the view of particle orbit theory. I believe there is a more 'modern' way of doing exactly the same thing, but here I just list the original derivation.

We consider a plasma at uniform temperature lying above a horizontal plane in a uniform gravitational field directed vertically downwards. There is a horizontal magnetic field in x direction uniform in each half column with a jump in field strength produced by a uniform horizontal current sheet at the boundary plane $z = 0$. The gravitational force is balanced by a pressure gradient in the plasma and by the jump in the magnetic pressure at $z = 0$.

We now suppose the boundary of the plasam at $z = 0$ to be rippled by a sinusoidal disturbance as shown in fig-RT_perturb. We may write for the displacement of the interface (ADD FIGURE!)

$$\Delta z = a \sin ky \quad (13.2)$$

where a , the amplitude of the disturbance, is considered small and $k (= 2\pi/\lambda)$ is the wave number of the disturbance in the y-direction. The drift resulting from gravity is given by

$$\mathbf{V}_g = \frac{m}{q} \frac{\mathbf{g} \times \mathbf{B}}{B^2}$$

Since the magnetic field is in the x-direction, the electrons will drift to the right and the positive ions will drift to the left. The gravity drift, therefore, causes a charge separation in the plasma and the resulting boundary has the form shown in ?@fig-RT-displacement. The surface charge $\delta\sigma$ due to the separation ($\delta\Delta z$) of ions and electrons is given by

$$\begin{aligned} \delta\sigma &= Ne\delta\Delta z \\ &= Ne \frac{\partial\Delta z}{\partial y} \delta y \\ &= e \frac{\partial\Delta z}{\partial y} V_g \delta t \end{aligned}$$

Therefore, the time rate of change of the surface charge density is given by

$$\begin{aligned} \frac{\partial\sigma}{\partial t} &= NeV_g \frac{\partial}{\partial y} \Delta z \\ &= -Ne \frac{Mg}{eB} ak \cos ky \\ &= -\frac{NMeg}{B} ak \cos ky \end{aligned} \quad (13.3)$$

where in writing these expressions, we have neglected the electron drift, as being small in the ratio m/M compared to the ion drift. The electric field resulting from the surface charge can

be computed in a straight-forward manner. In the region away from the boundary, we must have

$$\nabla \cdot (\epsilon \mathbf{E}) = 0 \quad (13.4)$$

where $\epsilon = \epsilon_0(1 + \frac{\omega_{pi}^2}{\Omega_i^2}) = \dots$ is the dielectric constant of the plasma. At the interface the electric field is determined by

$$\nabla \cdot (\epsilon \mathbf{E}) = \frac{1}{\mu_0} \frac{\sigma}{dz}$$

where σ is the surface charge density and dz is the infinitesimal thickness of the layer. We now integrate this equation over an element of column $dS dz$. The right-hand side gives the charge within the column element (σdS). Making use of Gauss's theorem to transform the left-hand side, we obtain

$$\epsilon E_z dS = \frac{1}{\mu_0} \sigma dS = \frac{1}{\mu_0} \sigma_0 \cos ky dz$$

Thus the electric field at the interface is given by

$$\epsilon E_z = \frac{\sigma_0}{\mu_0} \cos ky \quad (13.5)$$

The electric field which satisfies Equation 13.4 within the plasma and the boundary condition Equation 13.5 at $z = 0$ has the components

$$E_y = \frac{\sigma_0}{\mu_0 \epsilon} \sin ky e^{-kz}$$

$$E_z = \frac{\sigma_0}{\mu_0 \epsilon} \cos ky e^{-kz}$$

with $E_x = 0$. These electric fields give rise to the drifts which can be computed from the equation

$$\mathbf{V} = \frac{\mathbf{E} \times \mathbf{B}}{B^2}$$

Remembering that \mathbf{B} is in the x-direction, we obtain

$$V_y = \frac{E_z}{B}, \quad V_z = -\frac{E_y}{B}$$

From the foregoing equations we obtain

$$V_y = \frac{\sigma_0}{\mu_0 B} \cos ky e^{-kz}$$

$$V_z = -\frac{\sigma_0}{\mu_0 B} \sin ky e^{-kz}$$

It is clear from the solutions that the velocity field is divergence free and, therefore, does not cause any change in the density of the plasma except at the boundary. We have

$$\frac{\partial}{\partial t} \Delta z(z=0) = V_z(z=0) = -\frac{\sigma}{\mu_0 B} \sin ky \quad (13.6)$$

From Equation 13.2 and Equation 13.6, we obtain the equation of motion for the amplitude a :

$$\frac{da}{dt} = -\frac{\sigma_0}{\mu_0 B}$$

Equation 13.3 and Equation 13.5 yield

$$\frac{d\sigma_0}{dt} = -\frac{NMg}{\mu_0 B} ak$$

From the above two equations, we obtain

$$\begin{aligned} \frac{d^2a}{dt^2} &= \frac{1}{\mu_0 \epsilon B} \frac{NMg}{B} ka \\ &\approx gka, \quad (\text{for } \epsilon \gg 1) \end{aligned}$$

($g = \dots$ from equilibrium). The solution of this equation is given by

$$a(t) = a_0 e^{\pm \sqrt{gk}t}$$

It is interesting to note that the rate of growth of the instability is exactly the same as in the Rayleigh-Taylor instability of a fluid supported against gravity by a second fluid which is weightless. The charge separation is able to overcome exactly the restraining influence of the magnetic field. This exact compensation occurs only in the limit of $\epsilon \gg 1$.

The same result can also be obtained using the rigorous formulation of the Boltzmann transport equation. However, in more complicated cases, the first order orbit theory gives results which agree with those obtained from the Boltzmann equation only in some special cases.

The essential mechanism which gives rise to the instability is the charge separation resulting from the gravity drift — drift arising from a force which does not depend upon the sign of the charge. If we consider a plasma configuration in a torus, the particles experience the centrifugal force $mv_{||}^2/R$ and the gradient B force mv_{\perp}^2/R which are both independent of the sign of the charge. Therefore, we should expect instabilities in a plasma confined to a torus.

13.5 Kelvin-Helmholtz instability

Kelvin-Helmholtz (KH) instability happens due to a velocity shear and produces surface waves at the shear layer that eventually roll up into nonlinear vortices. Typical examples are:

- plane crash
- flapping of flags
- diocotron instability in an electron sheet
- water wave, nonlinear phase
- laser ablation of metal
- low-latitude flanks of the magnetopause during periods of northward interplanetary magnetic field
- accretion discs
- pulsar winds
- jets

It presents a key mechanism for mediating mass, momentum and energy transport at boundary layers with a velocity shear via processes such as viscous interaction, turbulence, mode conversion and magnetic reconnection.

To understand why shear flow can lead to instability, we'll first introduce the Bernoulli theorem in fluid mechanics. From the MHD momentum equation, take $\partial/\partial t = 0$, $\mathbf{J} = 0$, $\mathbf{g} = 0$, we obtain

$$\rho \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p$$

Using the natural coordinates, let $\hat{\mathbf{t}}$ be the unit tangent vector on a streamline, $\hat{\mathbf{n}}$ be the unit normal vector pointing from concave to convex side, and ds be the infinitesimal distance along streamline, we have

$$\begin{aligned} \rho \mathbf{v} \cdot \nabla \mathbf{v} &= (\rho v \frac{\partial}{\partial s}) \mathbf{v} = \rho v_s \frac{\partial}{\partial s} (\hat{\mathbf{t}} v) = \rho v \left(\frac{\partial v}{\partial s} \hat{\mathbf{t}} + \frac{\partial \hat{\mathbf{t}}}{\partial s} v \right) \\ &= -\frac{\partial p}{\partial s} \hat{\mathbf{t}} - \nabla_{\perp} p \\ \Rightarrow \hat{\mathbf{t}} : \quad p v \frac{\partial v}{\partial s} + \frac{\partial p}{\partial s} &= \frac{\partial}{\partial s} \left(\frac{1}{2} \rho v^2 + p \right) = 0 \end{aligned}$$

Therefore, $\frac{1}{2} \rho v^2 + p = \text{const.}$ along a streamline for an incompressible flow. This is the classical Bernoulli equation.

Now, consider two flow layer with velocity shear at the plane interface in fig-two-layer (ADD FIGURE!). Imaging there's a ripple on the layer interface pointing upward at Q . If we examine the cross section at P and Q for the lower layer respectively, we find

$$\text{flux at } P = \rho v A|_P = \text{flux at } Q = \rho v A|_Q$$

Since density along field lines are constant (incompressible) and area $A_P < A_Q$, we have $v_P > v_Q$. From the Bernoulli equation, $p_P < p_Q$. Similar for the upper layer, we get the pressure at P is larger than that at Q . Therefore, the total pressure is pointing away from interface, which let the ripple grow.

Now we are ready to do more careful derivations. The governing equations are

$$\begin{aligned}\nabla \cdot \mathbf{v} &= 0 \\ \rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) &= -\nabla p\end{aligned}$$

In equilibrium, suppose there is a velocity shear in the x direction, and the interface lies along the y direction,

$$\rho_0 = \text{const.}$$

$$p_0 = \text{const.}$$

$$\mathbf{v}_0 = \hat{y}v_{0y}(x)$$

Assume linear perturbations of the form

$$\begin{aligned}v_{1x}(\mathbf{x}, t) &= v_{1x}(x)e^{ik_y y - i\omega t} \\ p_1(\mathbf{x}, t) &= p_1(x)e^{ik_y y - i\omega t}\end{aligned}$$

so the linearized momentum equation is

$$\rho_0 \left(\frac{\partial \mathbf{v}_1}{\partial t} + \mathbf{v}_0 \cdot \nabla \mathbf{v}_1 + \mathbf{v}_1 \cdot \nabla \mathbf{v}_0 \right) = -\nabla p_1$$

where

$$\begin{aligned}\mathbf{v}_0 \cdot \nabla \mathbf{v}_1 &= ik_y v_{0y} \mathbf{v}_1 \\ \mathbf{v}_1 \cdot \nabla \mathbf{v}_0 &= v_{1x} \frac{\partial}{\partial x} v_{0y}(x) \hat{y}\end{aligned}$$

Then the x and y components of the linearized momentum equation give

$$\begin{aligned}-i\omega \rho_0 v_{1x} + \rho_0 ik_y v_{0y} v_{1x} &= -\frac{\partial p_1}{\partial x} \\ -i\omega \rho_0 v_{1y} + \rho_0 ik_y v_{0y} v_{1y} + \rho_0 v_{1x} \frac{\partial}{\partial x} v_{0y}(x) &= -ik_y p_1\end{aligned}$$

Together with the linearized incompressibility condition

$$\frac{\partial v_{1x}}{\partial x} + ik_y v_{1y} = 0$$

by eliminating p_1 and v_{1y} we get

$$\frac{\partial^2 v_{1x}}{\partial x^2} - \left[k_y^2 - \frac{k_y v_{0y}''(x)}{\omega - k_y v_{0y}(x)} \right] v_{1x} = 0 \quad (13.7)$$

Now we are half way from obtaining the dispersion relation. For simplicity, let us assume three layer regions with Region I on the left ($x < 0$), Region II in the middle ($x \in (0, \tau)$), and Region III on the right ($x > \tau$). Let the shear layer II thickness be τ . Set the velocity on the two sides $v_1 = 0$, $v_2 = v_1 + \Delta v \equiv v$, and $k_y = k$. Then

$$v_{0y}''(x) = \left(\frac{v}{\tau}\right)[\delta(x) - \delta(x - \tau)]$$

except at $x = 0$ and $x = \tau$. Equation 13.7 can be simplified to

$$\frac{d^2 v_{1x}}{dx^2} - k^2 v_{1x} = 0$$

In region I ($x < 0$),

$$\begin{aligned} v_{1x} &= \xi_0 e^{kx}, \quad x < 0 \\ \frac{\partial v_{1x}}{\partial x} \Big|_{x=0^-} &= k\xi_0 \end{aligned}$$

In region III ($x > \tau$),

$$\begin{aligned} v_{1x} &= \xi_\tau e^{-k(x-\tau)}, \quad x > \tau \\ \frac{\partial v_{1x}}{\partial x} \Big|_{x=\tau^+} &= -k\xi_\tau \end{aligned}$$

In region II ($x \in (0, \tau)$), we are looking for a solution which is a superposition of the solutions from both sides and is continuous at the boundaries

$$\begin{aligned} v_{1x} &= \xi_\tau \frac{\sinh kx}{\sinh k\tau} + \xi_0 \frac{\sinh k(x-\tau)}{\sinh -k\tau}, \quad 0 < x < \tau \\ \frac{\partial v_{1x}}{\partial x} \Big|_{x=\tau^-} &= k\xi_\tau \frac{\cosh k\tau}{\sinh k\tau} - k\xi_0 \frac{1}{\sinh k\tau} \\ \frac{\partial v_{1x}}{\partial x} \Big|_{x=0^+} &= k\xi_\tau \frac{1}{\sinh k\tau} - k\xi_0 \frac{\cosh k\tau}{\sinh k\tau} \end{aligned}$$

The continuity at $x = 0$ requires $V_{0y} = 0$, $V_{1x} = \xi_0$. Integrating the governing Equation 13.7 from $x = 0^-$ to $x = 0^+$ yields

$$-\xi_0 k \frac{\cosh k\tau}{\sinh k\tau} + \xi_\tau k \frac{1}{\sinh k\tau} - k\xi_0 + \frac{kv}{\omega\tau} \xi_0 = 0$$

Integrating the governing Equation 13.7 from $x = \tau^-$ to $x = \tau^+$ yields

$$-\xi_\tau k + \xi_0 k \frac{1}{\sinh k\tau} - \xi_\tau k \frac{\cosh k\tau}{\sinh k\tau} - \frac{kv}{\omega\tau} \xi_\tau = 0$$

Combining the last two equations, we obtain

$$1 = \left[\sinh k\tau + \cosh k\tau + \frac{kv}{\omega - kv} \frac{\sinh k\tau}{k\tau} \right] \left[\cosh k\tau + \sinh k\tau - \frac{kv \sinh k\tau}{\omega} \frac{1}{k\tau} \right]$$

which is the dispersion relation for the KH instability.

With the identity $\sinh k\tau + \cosh k\tau = e^{k\tau}$, the dispersion relation can be simplified to

$$\begin{aligned} 1 &= \left[e^{k\tau} + \frac{kv}{\omega - kv} \frac{\sinh k\tau}{k\tau} \right] \left[e^{k\tau} - \frac{kv \sinh k\tau}{\omega - k\tau} \right] \\ 1 &= e^{2k\tau} + e^{k\tau} \frac{(k\tau)^2}{\omega(\omega - kv)} \frac{\sinh k\tau}{k\tau} - \frac{(k\tau)^2}{\omega(\omega - k\tau)} \left(\frac{\sinh k\tau}{k\tau} \right)^2 \\ 1 &= e^{2k\tau} + \frac{(k\tau)^2}{\omega(\omega - kv)} \frac{\sinh k\tau}{k\tau} \left[e^{k\tau} - \frac{\sinh k\tau}{k\tau} \right] \end{aligned}$$

Multiplying both sides by $\omega(\omega - kv)$, we get

$$\omega(\omega - k\tau)(1 - e^{2k\tau}) = (kv)^2 \frac{\sinh k\tau}{k\tau} \left[e^{k\tau} - \frac{\sinh k\tau}{k\tau} \right]$$

Assuming $k\tau \ll 1$ (long wavelength approximation), $e^{k\tau} \approx 1 + k\tau$, we obtain

$$\omega(\omega - kv) + \frac{(kv)^2}{2} \approx 0$$

the solution of which is

$$\omega = \frac{1}{2}kv(1 \pm i), \quad k\tau \ll 1$$

In general, the growth rate of KH is

$$\omega_i = \frac{1}{2}|k_y \Delta v|, \quad k_y \tau \ll 1$$

13.5.1 Diocotron instability on electron sheet

HAVEN'T CHECKED!

A diocotron instability is a plasma instability created by two sheets of charge slipping past each other. Energy is dissipated in the form of two surface waves propagating in opposite directions, with one flowing over the other. This instability is the plasma analog of the Kelvin-Helmholtz instability in fluid mechanics.

For the simplest case, we have a uniform electron sheet and a parallel constant magnetic field in the plane of the sheet, as illustrated in fig-electron-sheet. Due to space charge of electron sheet, there is electric field pointing towards the sheet in the upper and lower region. Consider a small perturbation (sinusoidal ripple) on an electron sheet. The coulomb force expels electrons outward, so the electrons will drift, according to the right-hand rule, to the left. The deficit of electrons is equivalent to some positive charge distribution, and thus created an electric field. The $\mathbf{E} \times \mathbf{B}$ drift is pointing outward, so the perturbation is growing. Note that even though

this problem looks innocent, but it is actually not easy. People, even the giants in plasma physics, made a lot of mistakes in the derivation!

If there is also a magnetic field inside the sheet, the $\mathbf{E} \times \mathbf{B}$ drift will form a velocity gradient within the sheet, and lead to K-H instability. Denote σ_0 as the surface charge density and ρ_0 as charge density, we have

$$\sigma_0 = \rho_0\tau = en_0\tau,$$

and the velocity shear across the sheet

$$\Delta v = \frac{E_2}{B_0} + \frac{E_1}{B_0} = -\frac{1}{B_0} \frac{\sigma_0}{\epsilon_0} = -\frac{en_0\tau}{B_0\epsilon_0}.$$

Then from the dispersion relation of K-H mode, we have the growth rate

$$\omega_i = \frac{1}{2} k_y |\Delta v| = \frac{1}{2} k_y \left| \frac{en_0\tau}{B_0\epsilon_0} \right| = \frac{1}{2} k_y \tau \frac{\omega_{pe}^2}{|\Omega_e|},$$

which is valid as long as $k_y\tau \ll 1$, i.e., long wave length limit.

(Bellan 2008) P537.

FIGURE NEEDED from H.W.3.4 Consider the diocotron instability on a MELBA-like annular electron beam which propagates inside a metallic drift tube. Let \$V = \$ beam voltage, $I =$ beam current, $a =$ beam radius, $\tau =$ annular beam thickness ($\tau \ll 1$)m $L =$ length of drift tube, $T =$ beam's pulselength, $B =$ axial magnetic field. Note that the combined self-electric and self-magnetic field of the beam produces a slow rotational $\mathbf{E} \times \mathbf{B}$ drift in the θ -direction. This azimuthal drift velocity, $v_{0\theta}$, is much less than the axial velocity of the beam, but it is sheared.

In equilibrium,

$$\begin{aligned} 0 &= q(\mathbf{v} \times \mathbf{B} + \mathbf{E}) \\ \mathbf{v} &= v_{0\theta}\hat{\theta} + v_{0z}\hat{z} \\ \mathbf{B} &= B_{0\theta}\hat{\theta} + B_{0z}\hat{z} \\ \Rightarrow v_\theta - v_z B_\theta + E_r &= 0 \end{aligned}$$

From Ampère's law,

$$\begin{aligned} B_{0\theta} \cdot 2\pi(a + \tau) &\approx B_\theta \cdot 2\pi a = \mu_0 2\pi a \tau J_z \\ \Rightarrow J_z &= \frac{1}{\mu_0 \tau} B_{0\theta} = -en_0 V_{0z} \end{aligned}$$

From Gauss's law,

$$\begin{aligned} \nabla \cdot \mathbf{E} &= \frac{\rho}{\epsilon_0} = \frac{-en_0}{\epsilon_0} \\ E_r \cdot 2\pi a \Delta &= \frac{-en_0}{\epsilon_0} \cdot 2\pi a \tau \Delta \\ \Rightarrow E_r &= \frac{1}{\epsilon_0} \tau (-e) n_0 \end{aligned}$$

Substituting E_r and J_z into the radial force balance equation, we obtain

$$\begin{aligned} V_\theta|_{r=a+\tau} &= \frac{1}{B_z} [V_{0z} B_\theta - E_r]|_{r=a+\tau} \\ &= \frac{1}{\gamma^2} \frac{E_r}{B_z} \end{aligned}$$

where $\gamma = (1 - \beta)^{-1/2} = 1 + V/(511 \text{ keV})$, and $\beta = V_{0z}/c$.

Let $\nu = \frac{I}{\beta I_A}$ be the Budker parameter, $I_A = 4\pi\epsilon_0 mc^2/e = 17 \text{ kA}$ be the Alfvén-Lawson current, and $\Omega = \frac{eB_{0z}}{m_e}$ be the nonrelativistic cyclotron frequency associated with the axial B field, we have

$$V_\theta|_{r=a+\tau} = \frac{2c^2\nu}{\Omega a \gamma^2}$$

At $r = a$, $V_\theta = 0$ because there's no E field. Therefore the velocity shear in $\hat{\theta}$ is

$$\Delta V_\theta = V_\theta|_{r=a+\tau} - V_\theta|_{r=a} = \frac{2c^2\nu}{\Omega a \gamma^2}$$

Then from the analysis of K-H instability, the temporal growth rate ω_i is given by

$$\begin{aligned} \omega_i &= \frac{1}{2} |k_\theta \Delta V_{0\theta}| \\ &= \frac{1}{2} \frac{m}{a} \frac{2c^2\nu}{\Omega a \gamma^2} \end{aligned}$$

For long wavelength limit, let $m = 1$.

For MELBA-like beam with the following parameters, $V = 700 \text{ keV}$, $I = 1 \text{ kA}$, $a = 5 \text{ cm}$, $\tau = 0.5 \text{ cm}$, $T = 500 \text{ ns}$, $L = 1 \text{ m}$, $B = 2 \text{ kG}$,

$$\omega_i = 1.18 \times 10^7 \text{ s}^{-1}$$

The total number of e-folds of the instability growth during the pulse time T is

$$\omega_i T \approx 5.9$$

Even though this is large, K-H instability will not stay at the initial position and grow in time; instead it will be transported. It is more meaningful to estimate the spatial growth by evaluating the total number of e-folds experienced by a signal of some frequency as it propagates along the machine length L:

$$k_i L = \frac{\omega_i}{V_{0z}} L = \omega_i \frac{L}{\beta c} \approx 0.04$$

Therefore we don't need to worry too much about this instability.

13.6 MHD Stability

$$\mathbf{J} \times \mathbf{B} = \nabla p$$

$$\mathbf{J}_\perp = \frac{\mathbf{B} \times \nabla p}{B^2}$$

The current is often called the *diamagnetic current*. It arises from the plasma pressure gradient.

Using Ampère's law we can write the magnetic force in the form

$$\mathbf{J} \times \mathbf{B} = -\nabla \left(\frac{B^2}{2\mu_0} \right) + \frac{1}{\mu_0} (\mathbf{B} \cdot \nabla) \mathbf{B} \quad (13.8)$$

which separates into the magnetic pressure term and the magnetic tension term.

If we use an anisotropic description of the thermal pressure term, Equation 13.8 can be written as

$$\mathbf{J} \times \mathbf{B} = -\nabla_\perp \left(p_\perp + \frac{B^2}{2\mu_0} \right) + \left(1 + \frac{p_\perp - p_\parallel}{B^2/\mu_0} \right) \mathbf{B} \cdot \nabla \mathbf{B} \quad (13.9)$$

13.6.1 Harris Current Sheet

An example of a MHD equilibrium configuration is the Harris current sheet, in which the variations in the magnetic field and plasma pressure over the current sheet balance each other. The Harris current sheet can be taken as the first approximation of the Earth's magnetotail that can stay stable for long time periods. In a 1D Harris current sheet the magnetic field (assumed here to be in the z -direction) is given by

$$\mathbf{B} = B_0 \tanh \left(\frac{z}{\lambda} \right) \hat{y}$$

The pressure is given by

$$p = p_0 \cosh^{-2} \frac{z}{\lambda}$$

where $p_0 = B_0^2/(2\mu_0)$. The current density is then

$$J_y(z) = \frac{B_0}{\mu_0 \lambda} \operatorname{sech}^2 \left(\frac{z}{\lambda} \right)$$

with a charge density

$$\rho(z) = \rho_0 \operatorname{sech}^2\left(\frac{z}{\lambda}\right)$$

The Harris current sheet is also an equilibrium solution of the Vlasov-Maxwell equations:

$$\begin{aligned} \frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_r f + \frac{q}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \nabla_v f &= 0 \\ f_s(\mathbf{r}_2, \mathbf{v}_2, t_2) &= f_s(\mathbf{r}_1, \mathbf{v}_1, t_1) = f_s(\mathbf{r}_2, \mathbf{v}_2, t_1) \end{aligned}$$

This is known as the Liouville's theorem: the particle's phase space density is conserved along particle trajectories. To satisfy the Vlasov equation, particle distribution functions are written as functions of particle invariants of motion: * Particle energy $W = mv^2/2 + q\varphi$ * Canonical momentum along y direction $P_y = mv_y + qA$

Originally, [Harris 1962] constructed the equilibrium current sheet by linearly combined these two invariants of motion, and assuming exponential form in the particle distribution function:

$$\begin{aligned} f_\alpha &\propto \exp\left[-(W_\alpha - V_{D\alpha}P_{y\alpha})/T_\alpha\right], \quad \alpha = e, i \\ &\propto \exp[-q(\varphi + V_{D\alpha}A)/T_\alpha] \cdot \exp[-m(v - V_{D\alpha}^2)/2T_\alpha] \end{aligned}$$

The first term represents the plasma density dependence on A and φ , i.e., on location, whereas the second term represents a shifted Maxwellian distribution with location-independent V_y shift and temperature. The particle distributions are integrated in velocity space to obtain the charge and current densities

$$\begin{aligned} \rho(A, \varphi) &= \sum_\alpha q_\alpha \int f_\alpha d^3v_\alpha \\ J(A, \varphi) &= \sum_\alpha q_\alpha \int f_\alpha v_\alpha d^3v_\alpha \end{aligned}$$

Finally, from Maxwell's equations, we can obtain the EM fields

$$\begin{aligned} (\nabla \cdot \mathbf{E})/4\pi &= \rho \\ c(\nabla \times \mathbf{B})/4\pi &= J \end{aligned}$$

Even if the particle trajectories within the current sheet are complicated (e.g. gyrating orbits, meandering orbits), their distributions can be very simple, i.e. drifting Maxwellian.

Limitations of the Harris model:

- Harris model is a 1D model with no magnetic field in the current sheet normal direction.
 - People often superimpose a uniform magnetic field (in the normal direction) to the 1D Harris sheet. However, the superimposed model is not a kinetic equilibrium!

- The electrostatic electric field is zero within the Harris model.
 - Corrections can be made to the drifting Maxwellian to generate an electric field.
- The particle distributions are always drifting Maxwellian in the Harris sheet, which often disagrees with observations.
 - Modified forms of distribution

$$\begin{aligned}f_\alpha &\propto \exp [-(W_\alpha - V_{D\alpha} P_{y\alpha})/T_\alpha] \\f_\alpha &\propto \exp [-W_\alpha/T_\alpha] \cdot g(P_{y\alpha}) \\f_\alpha &\propto \exp [-(W_\alpha - V_{D\alpha} P_{y\alpha})/T_\alpha + I_\alpha \omega_\alpha]\end{aligned}$$

where I_α is a new adiabatic invariant associated with periodic motions

$$I = \oint p \, dq$$

Here p is the generalized momentum, and q is the conjugate position coordinate. For example, let $p = mv_z$, $q = z$, we have a so-called current sheet invariant

$$I_z = \oint mv_z \, dz$$

In terms of application, an equilibrium state is important because we may be able to reconstruct the whole structure by simply taking measurements at a few points. Inspired by this simple Harris model, we can also establish a similar equilibrium state for the magnetic cavity. Checkout the nice talk given by Prof. Xuzhi Zhou:

13.6.2 -Pinch and Z-Pinch

-pinch and Z-Pinch are both 1D equilibrium configurations expressed in cylindrical coordinates. In a -pinch cylindrical coils drive an electric current and the magnetic field is axial, while in a Z-pinch the electric current is axial and the magnetic field is toroidal.

13.6.3 Force-Free Field

If $\beta \ll 1$ in MHD equilibrium, the pressure gradient is negligible and thus

$$\mathbf{J} \times \mathbf{B} = 0 \tag{13.10}$$

Such configurations are called *force-free fields* because the magnetic force on the plasma is zero. According to Equation 13.8 in a force-free field the magnetic pressure gradient $\nabla(B^2/2\mu_0)$ is balanced by the magnetic tension force $\mu_0^{-1}(\mathbf{B} \cdot \nabla)\mathbf{B}$. In reality the force-free equilibrium is often

a very good approximation of the momentum equation. It is also evident from Equation 13.10 that in a force-free field the electric current flows along the magnetic field. Such currents are commonly called *field-aligned currents* (FAC).

Using Ampère's law we can write Equation 13.10 as

$$(\nabla \times \mathbf{B}) \times \mathbf{B} = 0$$

From this we see that the innocent-looking equation $\mathbf{J} \times \mathbf{B} = 0$ is in fact nonlinear and thus difficult to solve.

The field-alignment of the electric current can be expressed as

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} = \alpha(\mathbf{r}) \mathbf{B}$$

where α is a function of position. Taking divergence of this we get

$$\mathbf{B} \cdot \nabla \alpha = 0$$

i.e. α is constant along the magnetic field.

In the case α is a *constant in all directions*, the equation

$$\nabla \times \mathbf{B} = \alpha \mathbf{B} \quad (13.11)$$

is linear. Taking a curl Equation 13.11 we get the *Helmholtz equation*:

$$\nabla^2 \mathbf{B} + \alpha^2 \mathbf{B} = 0$$

Solution to the Helmholtz with helical equation in cylindrical symmetry was presented by Lundquist in 1950 in terms of Bessel functions J_0 and J_1 :

$$\begin{aligned} B_R &= 0 \\ B_A &= B_0 J_0\left(\frac{\alpha_0 r}{r_0}\right) \\ B_T &= \pm B_0 J_1\left(\frac{\alpha_0 r}{r_0}\right) \end{aligned}$$

where B_r , B_A , and B_T are radial, axial and tangential magnetic field components, respectively. The solution is a magnetic *flux rope* where magnetic field lines form helixes whose pitch angle increases from the axis ([?@fig-flux-rope](#)). r is the radial distance from the flux rope axis, r_0 is the radius of the flux rope and B_0 is the maximum magnetic field magnitude at the center of the flux rope $r = 0$.

A special case of a force-free magnetic field is the current-free configuration $\nabla \times \mathbf{B} = 0$. Now the magnetic field can be expressed as the gradient of a scalar potential $\mathbf{B} = \nabla \Psi$, and since $\nabla \cdot \mathbf{B} = 0$ it can be found via the *Laplace equation*

$$\nabla^2 \Psi = 0$$

with appropriate boundary conditions and using the methods of potential theory.

For example, the Sun's magnetic field structure is often modeled by the so-called Potential Field Source Surface (**PFSS**) model ([?@fig-PFSS](#)). The magnetic field is computed from the Laplace equation using spherical coordinates from the photosphere to the “source surface”, nominally chosen to be at 2.5 Solar radii. At the source surface the Sun's magnetic field is assumed to be purely radial. The inner boundary conditions are obtained from solar magnetograms. Thus, PFSS assumes that there is no electric current in the corona.

The force-free model is also used to describe the magnetic flux ropes, which often display as bipolar structures. Lately, Prof. Xuzhi Zhou's group has extended the description of magnetic cavity and introduced a new canonical momentum in the axial direction to describe the inverse of magnetic field across a equilibrium flux rope.

13.7 MHD Modes

A simple but representative dispersion relation writes

$$\omega^2 = (\mathbf{k} \cdot \mathbf{V}_A)^2 - \mathbf{k} \cdot \mathbf{g}, \quad \text{where } \mathbf{V}_A = \frac{\mathbf{B}_0}{B_0} \cdot V_A$$

If we treat plasma as a single magnetized fluid,

$$\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \underbrace{\mathbf{u} \cdot \nabla \mathbf{u}}_{\text{K-H inst.}} \right) = \underbrace{-\nabla p}_{\text{ballooning inst.}} + \underbrace{\mathbf{j} \times \mathbf{B}}_{\text{kink, sausage inst.}} + \underbrace{\rho \mathbf{g}}_{\text{R-T inst.}}$$

Qualitatively, we can identify the source for each kind of instability in plasma. We will discuss them separately and in a set of combination below.

13.7.1 Kink Mode

A kink instability, is a *current-driven* plasma instability characterized by transverse displacements of a plasma column's cross-section from its center of mass without any change in the characteristics of the plasma. It typically develops in a thin plasma column carrying a strong axial current which exceeds the [Kruskal–Shafranov limit](#) and is sometimes known as the Kruskal–Shafranov (kink) instability.

The kink instability was first widely explored in fusion power machines with Z-pinch configurations in the 1950s. It is one of the common magnetohydrodynamic instability modes which can develop in a pinch plasma and is sometimes referred to as the $m = 1$ mode.

If a “kink” begins to develop in a column the magnetic forces on the inside of the kink become larger than those on the outside, which leads to growth of the perturbation. As it develops at fixed areas in the plasma, kinks belong to the class of “absolute plasma instabilities”, as opposed to convective processes.

```
KeyNotes.plot_kink()
```

The kink instability is the most dangerous instability in Tokamak. We have discussed this kind of microinstability from the view of single particle motion in Section 13.2; here, we will explore this a little bit further.

String model

First, image a current-carrying plasma column, shown in the x-z plane in fig-kink-column. The metallic wire carries current under tension T , and μ =mass/length is the mass per length. From the basic mechanics, $C_s = \sqrt{T/\mu}$ is the acoustic velocity in the system. Let the background field $\mathbf{B}_0 = \hat{z}B_0$ and the displacement $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x}, t)$. We can show that, if the current \mathbf{I} is sufficiently strong, there will be kink instability.

ADD PLASMA KINK COLUMN FIGURE!

Assume the displacement in x-y plane has the form

$$\boldsymbol{\xi} = (\xi_x, \xi_y) e^{ik_z z - i\omega t}$$

The force law gives (i.e. the basic string model in mechanics textbooks)

$$\mu \frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} = T \frac{\partial^2 \boldsymbol{\xi}}{\partial z^2} + \text{force per unit length}$$

Here, the external force *per length* is the Lorentz force (which is why we say the R-T instability is current-driven),

$$\begin{aligned} \mathbf{I} \times \mathbf{B} &= \left(\hat{x} I \frac{\partial \xi_x}{\partial z} + \hat{y} I \frac{\partial \xi_y}{\partial z} + \hat{z} 0 \right) \times B_0 \hat{z} \\ &= IB_0 \left(\hat{x} \frac{\partial \xi_y}{\partial z} - \hat{y} \frac{\partial \xi_x}{\partial z} \right) \end{aligned}$$

In scalar forms, the force law gives

$$\begin{aligned} \hat{x} : -\mu\omega^2 \xi_x &= T(-k_z^2) \xi_x + IB_0 \frac{\partial \xi_y}{\partial z} \\ \hat{y} : -\mu\omega^2 \xi_y &= T(-k_z^2) \xi_y - IB_0 \frac{\partial \xi_x}{\partial z} \end{aligned}$$

Combining these two equations, we can easily get the dispersion relation

$$\omega^2 = k_z^2 C_s^2 \pm \frac{IB_0}{\mu_0 \mu} k_z$$

The dispersion relation is a representation of the force-law. The first term on the right-hand side is a stabilizing term due to tension; the second term with a minus sign is a destabilizing term due to Lorentz force. Note that the expression is very similar to R-T instability. (Which one?)

We can immediately estimate the scenario in a Tokamak. Take the radius of the column cut as a , wave number $k_z \sim 1/R$ (i.e. wave length is on the order of tokamak radius), $C_s^2 = V_A^2 = B_{0z}^2/(\mu_0 \rho_0)$ (i.e. tension in plasma give rises to Alfvén wave), then the current is

$$I = J_z(\pi a^2) = \frac{B_\theta 2\pi a}{\mu_0} \sim \frac{B_\theta a}{\mu_0}$$

and the mass per unit length is

$$\mu = \rho_0(\pi a^2) \sim \rho_0 a^2$$

The criterion for stability then becomes

$$\begin{aligned} k_z^2 C_s^2 &> \frac{IB_0}{\mu_0} k_z \Rightarrow \frac{1}{R} \frac{B_0 z^2}{\mu_0 \rho_0} > \frac{B_\theta a}{\mu_0} \frac{B_{0z}}{\rho_0 a^2} \\ &\Rightarrow \frac{a}{R} \frac{B_{0z}}{B_{0\theta}} > 1 \end{aligned}$$

which is called the *Kruskal-Shafranov stability criterion*. Usually we define

$$q \equiv \frac{a}{R} \frac{B_{0z}}{B_{0\theta}} = \frac{a}{R} \frac{B_t}{B_p}$$

as the safety factor. A real value for q is about 2 to 3.

Ideal MHD Approach

Now we use a more standard way to treat the kink mode. (Section 10.9 (Bellan 2008)) Assume we have a plasma column with radius a . Inside the column, we assume infinite conductivity, $\sigma = \infty$; outside the column, we assume vacuum so that we can only have current flow on surface $r = a$. Thus, besides the universal background magnetic field in the z direction, we also

have an azimuthal field due to surface current. (You will see later that the decay in θ actually drives the kink instability.)

In equilibrium,

$$\begin{aligned} r < a : \mathbf{B}_0 &= \hat{z}B_0 = \text{const.} \\ p_0 &= \text{const.}, \mathbf{v}_0 = \text{const.}, \mathbf{J}_0 = 0, \rho_0 = \text{const.} \\ r > a : \mathbf{B}_0 &= B_0\hat{z} + B_{0\theta}\frac{a}{r}\hat{\theta} \\ p_0 &= 0, \rho_0 = 0 \end{aligned}$$

The force equation in equilibrium

$$\nabla p_0 = \mathbf{J} \times \mathbf{B}_0$$

is satisfied automatically both for $r > a$ and $r < a$.

Let us introduce a small perturbation

$$\xi_{1r}(\mathbf{x}, t) = \tilde{\xi}_{1r}(r)e^{ikz+im\theta-i\omega t}$$

such that at $r = a$,

$$\xi_{1r}(\mathbf{x}, t)|_{r=a} = \tilde{\xi}_{1a}(r)e^{ikz+im\theta-i\omega t}$$

Before running into linearized equations, we can first take a look at different wave modes. That is, what will the perturbation looks like at a fixed time t with different m ? For simplicity, let us assume $t = 0$. (You can always make a time shift.) The actual displacement is the real part of ξ ,

$$\xi_{1r} = \xi_{1a} \cos(k_z z + m\theta)$$

For $m = 0$,

$$\xi_{1r} = \xi_{1a} \cos(k_z z)$$

which is the sausage mode.

For $m = 1$,

$$\xi_{1r} = \xi_{1a} \cos(k_z z + \theta)$$

If we draw the displacement down for $k_z z = 0, \frac{\pi}{2}, \pi, \frac{3}{2}\pi$, you can see one rotation in a 2π period, which indicates a shape of helix. This is often called the kink mode.

For higher m ,

$$\xi_{1r} = \xi_{1a} \cos(k_z z + m\theta)$$

which looks like m intertwine helices in one axial wavelength.

Now let's return to the perturbed equations. Here we will assume incompressibility as the equation of state,

$$\nabla \cdot \mathbf{v} = 0$$

The procedure to get the dispersion relation goes as follows:

1. Express the perturbed magnetic field as a function of displacement inside and outside the surface.
2. Relate the two regions by total force balance.

(I) $r < a$

$$\mathbf{v}_1 = \frac{\partial \boldsymbol{\xi}_1}{\partial t} \Rightarrow \nabla \cdot \boldsymbol{\xi}_1 = 0$$

The linearized continuity equation gives

$$\rho_1 = -\nabla \cdot (\rho_0 \boldsymbol{\xi}_1) = -\boldsymbol{\xi}_1 \cdot \nabla \rho_0 - \rho_0 \nabla \cdot \boldsymbol{\xi}_1 = 0$$

The linearized force law gives

$$\rho_0 \frac{\partial^2 \boldsymbol{\xi}_1}{\partial t^2} = -\nabla p_1 + \frac{(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0}{\mu_0} + \mathcal{Y}_0 \times \mathbf{B}_1$$

And the Ohm's law gives

$$\begin{aligned} -\frac{\partial \mathbf{B}_1}{\partial t} &= \nabla \times \mathbf{E}_1 = \nabla \times (-\mathbf{v}_1 \times \mathbf{B}_0) \\ \mathbf{B}_1 &= \nabla \times (\boldsymbol{\xi}_1 \times \mathbf{B}_0) = B_0 \nabla \times (\boldsymbol{\xi}_1 \times \hat{z}) = ik_z B_0 \boldsymbol{\xi}_1 \end{aligned}$$

The last equivalence is obtained from the expansion of the second term into four terms and cancellation of zero terms.

In cylindrical coordinates,

$$(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0 = ikB_0^2(\nabla \times \boldsymbol{\xi}_1) \times \hat{z}$$

and

$$\nabla \times \boldsymbol{\xi}_1 = \frac{1}{r} \begin{bmatrix} \hat{r} & r\hat{\theta} & \hat{z} \\ \partial_r & \partial_\theta & \partial_z \\ \xi_{1\theta} & r\xi_{1\theta} & \xi_{1z} \end{bmatrix}$$

so the linearized force law gives

$$\begin{aligned} -\omega^2 \xi_{1r} &= -\frac{1}{\rho_0} \frac{\partial p_1}{\partial r} + ikv_A^2 [ik\xi_{1r} - \frac{\partial \xi_{1z}}{\partial r}] \\ -\omega^2 \xi_{1\theta} &= -\frac{im}{\rho_0 r} p_1 + kv_A^2 [-\frac{im}{r} \xi_{1z} + ik\xi_{1\theta}] \\ -\omega^2 \xi_{1z} &= -ik_z p_0 / \rho_0 \end{aligned}$$

Substituting the expression of ξ_{1z} into the other two equations, we can get

$$\boldsymbol{\xi}_1 = \frac{1}{\omega^2} \nabla \left(\frac{p_1}{\rho_0} \right)$$

From the incompressibility condition, we have

$$\nabla \cdot \boldsymbol{\xi}_1 = 0 \Rightarrow \nabla^2 p_1 = 0$$

or in cylindrical coordinates,

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial p_1}{\partial r} \right) - \frac{m^2 p_1}{r^2} - k^2 p_1 = 0$$

Assume long wavelength limit and $m = 1$ (kink mode),

$$kr < ka \ll 1$$

we have

$$\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial p_1}{\partial r} \right) - \frac{p_1}{r^2} = 0$$

the solution of which from Legendre polynomials (I need to check) is

$$p_1 = Ar + \frac{B}{r} = Ar$$

because p_1 is finite at $r = 0$.

So we have

$$\begin{aligned}\boldsymbol{\xi}_1 &= \frac{1}{\omega^2} \nabla \left(\frac{p_1}{\rho_0} \right) = \hat{r} \frac{A}{\rho_0 \omega} e^{-i\omega t + i\theta + ikz} \\ \Rightarrow \xi_{1ra} &= \frac{A}{\rho_0 \omega^2}\end{aligned}$$

Then the perturbed kinetic pressure on the surface is

$$p_1(r = a^-) = Aa = \rho_0 \omega^2 \xi_{1ra} a,$$

and the perturbed magnetic field is

$$\mathbf{B}_1 = ikB_0 \boldsymbol{\xi}_1 \Rightarrow B_{1z}(r = a^-) = ikB_0 \xi_{1z} = -k^2 B_0 \xi_{1ra}$$

(II) $r > a$

$$\nabla \cdot \mathbf{B}_1 = 0, \quad \nabla \times \mathbf{B}_1 = 0 \Rightarrow \mathbf{B}_1 = \nabla \Psi_1, \quad \nabla^2 \Psi_1 = 0$$

The solution of Laplace equation in cylindrical coordinates is

$$\Psi = \left(\frac{C}{r} + D \right) r e^{-i\omega t + i\theta + ikz}$$

where $D = 0$ because $\Psi < \infty$ when $r \rightarrow \infty$.

The perturbed magnetic field is then

$$\mathbf{B}_1 = \mathbf{B}_{1e} \nabla \Psi = C \left[-\frac{\hat{r}}{r^2} + i \frac{\hat{\theta}}{r^2} + ik \frac{\hat{z}}{r} \right] e^{-i\omega t + i\theta + ikz}$$

and at $r = a$,

$$\mathbf{B}_{1e}(r = a) = B_{1ra} [\hat{r} - i\hat{\theta} - ik\hat{z}] e^{i\theta + ikz}.$$

Now, we want to relate ξ_{1ra} and B_{1ra} by the “frozen-in” law. To first-order approximation, let \hat{n} be the direction normal to the perturbed boundary, we have

$$(\hat{n} \cdot \mathbf{B})_1 = 0$$

The equation for the perturbed boundary (Eq.(10.146) of (Bellan 2008)) gives

$$r - \xi_r - a = 0$$

so

$$\begin{aligned}\hat{n} &= \frac{\nabla(r - \xi_r - a)}{|\nabla(r - \xi_r - a)|} \\ &= \frac{\hat{r} - \frac{i}{r}\xi_{1ra}\hat{\theta} - ik\xi_{1ra}\hat{z}}{|\hat{r} - \frac{i}{r}\xi_{1ra}\hat{\theta} - ik\xi_{1ra}\hat{z}|} \\ &= \hat{r} - \frac{i}{r}\xi_{1ra}\hat{\theta} - ik\xi_{1ra}\hat{z}\end{aligned}$$

where the last equivalence holds because ξ_{1ra}/r and $k\xi_{1ra}$ are both second-order in magnitude.

Therefore we get

$$\begin{aligned}(\hat{n} \cdot \mathbf{B})_1 &= (\hat{r} - \frac{i}{r}\xi_{1ra}\hat{\theta} - ik\xi_{1ra}\hat{z}) \cdot (B_{1r}\hat{r} + B_{0\theta}\hat{\theta} + B_{0z}\hat{z}) = 0 \\ B_{1r} &= B_{0\theta}\frac{i\xi_{1ra}}{a} + ik_z\xi_{1ra}B_{0z} = B_{1ra} = \frac{i\xi_{1ra}}{a}[B_{0\theta} + k_z a B_{0z}] = i\xi_{1ra}(\mathbf{k} \cdot \mathbf{B})\end{aligned}$$

where $\mathbf{k} = k_\theta\hat{\theta} + k_z\hat{z} = \frac{m}{r}\hat{\theta} + k_z\hat{z} = \frac{m}{a}\hat{\theta} + k_z\hat{z}$.

Finally, $p + B^2/2\mu_0$ is continuous across perturbed boundary,

$$\begin{aligned}\left(p + \frac{B^2}{2\mu_0}\right)_{1,interior} &= p_{1i} + \frac{2\mathbf{B}_{oi} \cdot \mathbf{B}_{1i}}{2\mu_0} = \rho_0\omega^2 a \xi_{1ra} - \frac{1}{\mu_0}k^2 B_0^2 a \xi_{1ra} = a\rho_0 \xi_{1ra}(\omega^2 - k^2 v_A^2) \\ \left(p + \frac{B^2}{2\mu_0}\right)_{1,exterior} &= 0 + \frac{1}{2\mu_0}[B_{0e}^2 + 2\mathbf{B}_{oe} \cdot \mathbf{B}_{1e}]_{1,r=a+\xi_r} \\ &= \frac{1}{2\mu_0} \left[B_{0ze}^2 + B_{0\theta a}^2 \left(\frac{a}{a + \xi_{1r}} \right)^2 + 2(B_{0\theta} B_{1e\theta} + B_{0z} B_{1ez}) \right]_1 \\ &= \frac{1}{2\mu_0} \left[-\frac{2\xi_{1ra} B_{0\theta a}^2}{a} + 2[B_{0\theta}(-iB_{1ra}) + B_{0z}(-ikaB_{1ra})] \right] \\ &= \frac{1}{2\mu_0} \left[-\frac{2\xi_{1ra} B_{0\theta a}^2}{a} + \frac{2\xi_{1ra}}{a} [B_{0\theta} + ka B_{0z}]^2 \right]\end{aligned}$$

where in one of the equivalence $\frac{\xi_{1ra}}{a} \rightarrow 0$,

$$\left(\frac{1}{1 + \frac{\xi_{1r}}{a}}\right)^2 \approx -2 \frac{\xi_{1ra}}{a}$$

$$\begin{aligned} \left(p + \frac{B^2}{2\mu_0}\right)_{1,interior} &= \left(p + \frac{B^2}{2\mu_0}\right)_{1,exterior} \\ \omega^2 &= k^2 v_A^2 + \frac{1}{a\mu_0 a^2 \rho_0} [k^2 a^2 B_{0z}^2 + 2kaB_{0\theta}B_{0z}] \\ \omega^2 &= \frac{1}{a\mu_0 a^2 \rho_0} [2k^2 a^2 B_{0z}^2 + 2kaB_{0\theta}B_{0z}] = \frac{2k^2 B_{0z}^2}{\mu_0 \rho_0} \left[1 + \frac{B_{0\theta}}{kaB_{0z}}\right] \end{aligned}$$

For stability, $1 > \left|\frac{B_{0\theta}}{kaB_{0z}}\right| \Rightarrow |ka| > \frac{B_{0\theta}}{B_{0z}}$. Take $|k| = R^{-1}$, where R is the major radius, the stability condition becomes

$$q \equiv \frac{a}{R} \frac{B_{0z}}{B_{0\theta}} > 1$$

and the q the called the safety factor. This is again the Kruskal-Shafranov limit for $m = 1$ kink mode. For sausage mode $m = 0$, the same approach as above can get

$$B_{0\theta} < \sqrt{2}B_{0z}$$

for stability.

Note:

1. This 2-region model can be generalized to 3-region model, which is more realistic compared to experiments. In the liner inertial fusion experiment, there is a mixture of R-T, sausage, kink and many high order modes.
2. In general, the dispersion relation can be written as

$$\omega^2 = (\mathbf{k} \cdot \mathbf{v}_A)^2 - \text{destabilizing term}$$

where the destabilizing term can be gravity $\mathbf{k} \cdot \mathbf{g}$, current I/I_{crit} , tearing mode due to resistivity, etc.

3. If we do the same analysis in Cartesian coordinates (x,y,z), there will be no sausage or kink mode! See HW3.

3-Region Problem

In 572 HW3.1, we have shown that a Cartesian liner is always stable for kink and sausage mode while a cylindrical liner isn't. However, if you also consider the gravity (or equivalently, centrifugal force, etc.) then you can have magnetic Rayleigh-Taylor (MRT) instability.

Here we show the derivation of dispersion relation in a 3-region problem in Cartesian liner. This method can be found in Chandrasekhar's book on instability, P429. Consider Fig.1 shown in Prof.Lau's paper "Anisotropy and feedthrough in magneto-Rayleigh-Taylor instability". The interface of the regions are $x = \alpha$, $x = \beta$. The governing equation including gravity is

$$\frac{d}{dx} \left\{ \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \frac{\partial \xi_{1x}}{\partial x} \right\} - k^2 [\rho_0 (\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2) + g \frac{\partial \rho_0}{\partial x}] \xi_{1x} = 0$$

If $\rho_0 = \text{const.} \Rightarrow g \frac{\partial \rho_0}{\partial x} = 0$. Within region I, II, and III, the governing equation can be simplified to

$$\frac{d^2 \xi_{1x}}{dx^2} - k^2 \xi_{1x} = 0$$

the general solution of which reads

$$\xi_{1x} = A e^{-kx} + B e^{kx} = C \sinh k(x - \alpha) + D \sinh k(x - \beta)$$

where A, B, C, D are coefficients. Cancelling out the unphysical solutions in each region, we have

$$\begin{aligned} \text{I: } & \xi_{1x} = \xi_\beta e^{k(x-\beta)} \\ \text{II: } & \xi_{1x} = \xi_\alpha \frac{\sinh k(x-\beta)}{\sinh k(\alpha-\beta)} + \xi_\beta \frac{\sinh k(x-\alpha)}{\sinh k(\beta-\alpha)} \\ \text{III: } & \xi_{1x} = \xi_\alpha e^{-k(x-\alpha)} \end{aligned}$$

Note that there's a jump for $\rho_0(x)$ at $x = \alpha, \beta$. Now we can integrate the governing equation across the boundaries,

$$\begin{aligned} \int_{x=\alpha^-}^{x=\alpha^+} dx \left[\frac{d}{dx} \left\{ \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \frac{\partial \xi_{1x}}{\partial x} \right\} - k^2 [\rho_0 (\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2) + g \frac{\partial \rho_0}{\partial x}] \xi_{1x} \right] &= 0 \\ \Rightarrow F_1(\xi_\alpha, \xi_\beta) &= 0 \\ \int_{x=\beta^-}^{x=\beta^+} dx \left[\frac{d}{dx} \left\{ \rho_0 [\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2] \frac{\partial \xi_{1x}}{\partial x} \right\} - k^2 [\rho_0 (\omega^2 - (\mathbf{k} \cdot \mathbf{v}_A)^2) + g \frac{\partial \rho_0}{\partial x}] \xi_{1x} \right] &= 0 \\ \Rightarrow F_2(\xi_\alpha, \xi_\beta) &= 0 \end{aligned}$$

Finally, from F_1 and F_2 , we can get the dispersion relation

$$\omega^4 - R\omega^2 + S = 0$$

where R and S are functions of $B_{01}, B_{02}, B_{03}, \rho_{01}, \rho_{02}, \rho_{03}, g, \Delta, k_y$ and k_z .

Next, we can examine the temporal evolution of sinusoidal ripples at interfaces in the form

$$\xi_{\alpha,\beta}(t) = A_1 e^{i\omega_1 t} + A_2 e^{-i\omega_3 t} + A_3 e^{i\omega_3 t} + A_4 e^{-i\omega_3 t}$$

where ω_1 and ω_3 are two eigen mode from the dispersion relation. See the details in the paper. The details of dispersion relation is given in M.Weis, et. al., Phys. Plasmas 21, 122708 (2014), and the coupling of MRT, sausage and kink mode in a cylindrical liner is given by M.Weis, et. al., Phys. Plasma 22, 032706 (2015).

Note: for RT/MRT,

$$\xi_{RT}(t) \sim \xi_0 e^{\gamma t} < \xi_0 e^{\sqrt{gk}t}$$

the growth rate

$$\gamma_{\max} = \sqrt{gk}$$

If we use $s = \frac{1}{2}gt^2 \Rightarrow \sqrt{2s} = \sqrt{|g|}t$,

$$\xi_{RT}(t) < \xi_0 e^{\sqrt{2ks}}$$

which only depends on wavenumber and distance.

13.7.2 Sausage Mode

The $m = 0$ mode is known as the sausage instability. There is no B_z so we have purely toroidal field. The initial equilibrium is established with radially inward Lorentz force and outward pressure gradient. But it is unstable to interchange due to curvature. When a perturbation that causes the rings to shrink, magnetic field in the plasma increases so that the $\mathbf{J} \times \mathbf{B}$ Lorentz force increases. There is nothing to counter this radially inward force increase, which leads to instability.

The dispersion relation is

$$\omega^2 = -2 \frac{p_0}{\rho_0} \frac{k}{R_c^2}$$

Adding B_z to the interior plasma stabilizes sausage instability: the magnetic pressure caused by B_z pushes back to oppose squeezing. The pressure balance at the interface gives

$$p_0 + \frac{B_{0z}^2}{2\mu_0} = \frac{B_\phi^2}{2\mu_0}$$

With the modified dispersion relation

$$\omega^2 = -2 \frac{p_0}{\rho_0 a^2} + \frac{B_{0z}^2}{\mu_0 \rho_0 a^2}$$

we have the condition for stability

$$B_{0z}^2 > \frac{1}{2} B_\phi^2$$

THIS IS PROBABLY DUPLICATE WITH THE PREVIOUS SECTION.

13.7.3 Kink Mode

However, even when the sausage mode is suppressed, the configuration is still unstable to the kink mode. This $m = 1$ mode retains circular cross-section of the tube and the perturbation is a kink of the tube into a helix. Without B_z , the system is unstable for all k ; with B_z , it is unstable for wavelengths long enough such that the pitch of the perturbation follows the pitch of the helix, i.e. the crests/troughs of the perturbations follow the fieldlines of tube

$$B_\phi/R + kB_Z \geq 0$$

In terms of a twist $\Phi = 2LB_\phi/RB_z$, this criterion is equivalent to

$$k \geq -\frac{\Phi}{2L}$$

The perturbation Lorentz force $\mathbf{j}_1 \times \mathbf{B}_0$ is zero. (???)

13.7.4 Ballooning Mode

The ballooning instability is a type of internal pressure-driven plasma instability usually seen in tokamak fusion power reactors or in space plasmas (Hameiri, Laurence, and Mond 1991). It is important in fusion research as it determines a set of criteria for the maximum achievable plasma beta. The name refers to the shape and action of the instability, which acts like the elongations formed in a long balloon when it is squeezed. In literature, the structure of these elongations are commonly referred to as “fingers”.

The narrow fingers of plasma produced by the instability are capable of accelerating and pushing aside the surrounding magnetic field in order to cause a sudden, explosive release of energy. Thus, the instability is also known as the explosive instability.

The interchange instability can be derived from the equations of the ballooning instability as a special case in which the ballooning mode does not perturb the equilibrium magnetic field.(Hameiri, Laurence, and Mond 1991) This special limit is known as the Mercier criterion.

```
KeyNotes.plot_balloon()
```

13.7.5 Tearing Mode

```
KeyNotes.plot_tearing()
```

Tearing mode is closely related to magnetic reconnection (Section 13.8). The reconnection process is very important because it is one of the main way of burst energy transformation. It is known that in collisionless systems current sheets are unstable against tearing instability, a process where the current tends to collapse into filaments. The tearing instability produces magnetic islands that then interact and merge together giving rise to a nonlinear instability phase, where the reconnection process is enhanced. See (Bellan 2008) P413 for more.

Linear Tearing Mode

Consider the interface between two plasmas containing magnetic fields of different orientations. The simplest imaginable field configuration is that illustrated in `?@fig-tearing-config-simple`. Here, the field varies only in the x -direction, and points only in the y -direction. The field is directed in the $-y$ -direction for $x < 0$, and in the $+y$ -direction for $x > 0$. The interface is situated at $x = 0$. The sudden reversal of the field direction across the interface gives rise to a z -directional current sheet at $x = 0$.

With the neglect of plasma resistivity, the field configuration shown in `?@fig-tearing-config-simple` represents a stable equilibrium state, assuming, of course, that we have normal pressure balance across the interface. But, does the field configuration remain stable when we take resistivity into account? If not, we expect an instability to develop which relaxes the

configuration to one possessing lower magnetic energy. As we shall see, this type of relaxation process inevitably entails the breaking and reconnection of magnetic field lines, and is, therefore, termed *magnetic reconnection*. The magnetic energy releases during the reconnection process eventually appears as plasma thermal energy. Thus, magnetic reconnection also involves plasma heating.

In the following, we shall outline the standard method for determining the linear stability of the type of magnetic field configuration shown in [?@fig-tearing-config-simple](#), taking into account the effect of plasma resistivity. We are particularly interested in plasma instabilities which are stable in the absence of resistivity, and only grow when the resistivity is non-zero. Such instabilities are conventionally termed *tearing modes*. Since magnetic reconnection is, in fact, a nonlinear process, we shall then proceed to investigate the nonlinear development of tearing modes.

The equilibrium magnetic field is written

$$\mathbf{B}_0 = B_{0y}(x)\hat{y}$$

where $B_{0y}(-x) = -B_{0y}(x)$. There is assumed to be non-equilibrium plasma flow. The linearized equations of resistive-MHD, assuming incompressible flow, take the form

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}_0) + \frac{\eta}{\mu_0} \nabla^2 \mathbf{B} \quad (13.12)$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla p + \frac{(\nabla \times \mathbf{B}) \times \mathbf{B}_0}{\mu_0} + \frac{(\nabla \times \mathbf{B}_0) \times \mathbf{B}}{\mu_0} \quad (13.13)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (13.14)$$

$$\nabla \cdot \mathbf{v} = 0 \quad (13.15)$$

Here, ρ_0 is the equilibrium plasma density, \mathbf{B} is the perturbed magnetic field, \mathbf{v} the perturbed plasma velocity, and p the perturbed plasma pressure. The assumption of incompressible plasma flow is valid provided that the plasma velocity associated with the instability remains significantly smaller than both the Alfvén velocity and the sonic velocity.

Suppose that all perturbed quantities vary like

$$A(x, y, z, t) = A(x)e^{iky+\gamma t} \quad (13.16)$$

where γ is the instability growth rate. The x -component of Equation 13.12 and the z -component of the curl of Equation 13.13 reduce to ???

$$\gamma B_x = ikB_{0y}v_x + \frac{\eta}{\mu_0} \left(\frac{d^2}{dx^2} - k^2 \right) B_x \quad (13.17)$$

$$\gamma \rho_0 \left(\frac{d^2}{dx^2} - k^2 \right) v_x = \frac{ikB_{0y}}{\mu_0} \left(\frac{d^2}{dx^2} - k^2 - \frac{B''_{0y}}{B_{0y}} \right) B_x \quad (13.18)$$

respectively, where use has been made of Equation 13.14 and Equation 13.15. Here, $'$ denotes d/dx .

It is convenient to normalize Equation 13.17 and Equation 13.18 using a typical magnetic field-strength, B_0 , and a typical scale-length, a . Let us define the *Alfvén time-scale*

$$\tau_A = \frac{a}{V_A}$$

where $V_A = B_0/\sqrt{\mu_0 \rho_0}$ is the Alfvén velocity, and the resistive diffusion time-scale

$$\tau_R = \frac{\mu_0 a^2}{\eta}$$

The ratio of these two time-scale is the *Lundquist number*:

$$S = \frac{\tau_R}{\tau_A}$$

Let $\psi = B_x/B_0$, $\phi = ikV_y/\gamma$, $\bar{x} = x/a$, $F = B_{0y}/B_0$, $F' \equiv dF/d\bar{x}$, $\bar{\gamma} = \gamma \tau_A$, and $\bar{k} = ka$. It follows that

$$\bar{\gamma}(\psi - F\phi) = S^{-1} \left(\frac{d^2}{d\bar{x}^2} - \bar{k}^2 \right) \psi \quad (13.19)$$

$$\bar{\gamma}^2 \left(\frac{d^2}{d\bar{x}^2} - \bar{k}^2 \right) \phi = -\bar{k}^2 F \left(\frac{d^2}{d\bar{x}^2} - \bar{k}^2 - \frac{F''}{F} \right) \psi \quad (13.20)$$

The term on the right-hand side of Equation 13.19 represents plasma resistivity, whilst the term on the left-hand side of Equation 13.20 represents plasma inertia.

It is assumed that *the tearing instability grows on a hybrid time-scale which is much less than τ_R but much greater than τ_A* . It follows that

$$\begin{aligned} \tau_A &\ll \gamma^{-1} \ll \tau_R \\ \gamma \tau_A &\ll 1 \ll \gamma \tau_R \\ \bar{\gamma} &\ll 1 \ll S \bar{\gamma} \end{aligned}$$

Thus, throughout most of the plasma we can neglect the right-hand side of Equation 13.19 and the left-hand side of Equation 13.20, which is equivalent to the neglect of plasma resistivity and inertia. In this case, the two equations reduce to

$$\phi = \frac{\psi}{F} \quad (13.21)$$

$$\frac{d^2\psi}{d\bar{x}^2} - \bar{k}^2\psi - \frac{F''}{F}\psi = 0 \quad (13.22)$$

Equation 13.21 is simply the flux freezing constraint, which requires the plasma to move with the magnetic field. Equation 13.22 is the linearized, static force balance criterion $\nabla \times (\mathbf{j} \times \mathbf{B}) = 0$. These two equations are known collectively as the equations of ideal-MHD, and are valid throughout virtually the whole plasma. However, it is clear that these equations *break down* in the immediate vicinity of the interface, where $F = 0$ (i.e. where the magnetic field reverses direction). Witness, for instance, the fact that the normalized “radial” velocity, ϕ , becomes infinite as $F \rightarrow 0$, according to Equation 13.21.

The ideal-MHD equations break down close to the interface because the neglect of plasma resistivity and inertia becomes untenable as $F \rightarrow 0$. Thus, there is a thin layer, in the immediate vicinity of the interface, $\bar{x} = 0$, where the behaviour of the plasma is governed by the full MHD equations, Equation 13.19 and Equation 13.20. We can simplify these equations, making use of the fact that $\bar{x} \ll 1$ and $d/d\bar{x} \gg 1$ in a thin layer, to obtain the following layer equations:

$$\bar{\gamma}(\psi - \bar{x}\phi) = S^{-1} \frac{d^2\psi}{d\bar{x}^2} \quad (13.23)$$

$$\bar{\gamma}^2 \frac{d^2\phi}{d\bar{x}^2} = -\bar{x} \frac{d^2\psi}{d\bar{x}^2} \quad (13.24)$$

Note that we have redefined the variables ϕ , $\bar{\gamma}$, and S , such that $\phi \rightarrow F'(0)\phi$, $\bar{\gamma} \rightarrow \gamma\tau_H$, and $S \rightarrow \tau_R/\tau_H$. Here,

$$\tau_H = \frac{\tau_A}{k a F'(0)}$$

is the *hydromagnetic time-scale*.

The tearing mode stability problem reduces to solving the non-ideal-MHD layer equations, Equation 13.23 and Equation 13.24, in the immediate vicinity of the interface, $\bar{x} = 0$, solving the ideal-MHD equations, @Equation 13.21 and Equation 13.22, everywhere else in the plasma, matching the two solutions at the edge of the layer, and applying physical boundary conditions as $|\bar{x}| \rightarrow \infty$. This method of solution was first described in a classic paper (Furth, Killeen, and Rosenbluth 1963). The steps are listed in the [farside note](#). The procedure is similar to the 3 layer solution of the K-H instability. After some maths, tearing mode dispersion relation is given as

$$\gamma = \left[\frac{\Gamma(1/4)}{2\pi\Gamma(3/4)} \right]^{4/5} \frac{(\Delta')^{4/5}}{\tau_H^{2/5} \tau_R^{3/5}}$$

where

$$\begin{aligned} \Delta &= 2\pi \frac{\Gamma(3/4)}{\Gamma(1/4)} S^{1/3} Q^{5/4} \\ Q &= \gamma \tau_H^{2/3} \tau_R^{1/3} \end{aligned}$$

The tearing mode is unstable whenever $\Delta' > 0$, and grows on the hybrid time-scale $\tau_H^{2/5} \tau_R^{3/5}$.

13.8 Magnetic Reconnection

Magnetic reconnection is one of the most complex processes known for converting energy from magnetic fields to particle motion. One of the key components in magnetic reconnection is the collision of two magnetic field regions with opposite-directed field lines, embedded in a plasma. The field and plasma combination forms an X-shaped configuration at their closest, and most intense point. Visualization of reconnections can be done from a field perspective or a [particle perspective](#).

There are several key elements for understanding the physics of reconnection at a deeper level (Ji et al. 2022). First, the frozen-in properties in an electron-ion plasma are associated with the electron fluid due to its light weight. This is expressed in terms of the time changing rate of magnetic flux (Equation 7.47) through an arbitrary area, \mathbf{S} (enclosed by loop l), convecting with the electron flow as $d\Phi/dt = \oint(\mathbf{E} + \mathbf{V}_e \times \mathbf{B}) \cdot d\mathbf{l} = 0$. Thus, the frozen-in condition is regulated by the electron momentum equation (i.e. the generalized Ohm's law)

$$\underbrace{\mathbf{E} + \mathbf{V}_e \times \mathbf{B}}_{\text{Ideal}} = \underbrace{\mathbf{R}_{\text{col}}}_{\text{Collisional}} - \underbrace{\frac{\nabla \cdot \mathbf{P}_e}{en}}_{\text{Kinetic}} - \underbrace{\frac{m_e}{e} \frac{d\mathbf{u}_e}{dt}}_{\text{Kinetic}} \quad (13.25)$$

where \mathbf{R}_{col} is the collisional force per electron charge. In a fully ionized collisional plasma, $\mathbf{R}_{\text{col}} \approx \eta \mathbf{j}$ where η is the resistivity due to Coulomb collisions. If the terms on the right-hand side of Equation 13.25 are negligible, then $d\Phi/dt \approx 0$ and the magnetic flux is “frozen-in” to the electron flow. For the generic reconnection layer illustrated in the middle panel of [?@fig-reconnection-configurations](#), deviations from ideal evolution occur within the “diffusion region” (blue), where either finite resistivity or kinetic effects (electron inertia and pressure tensor) are important. Within the diffusion region, field lines converging from opposite sides of the layer can change connectivity.

The next key element for understanding the physics of magnetic reconnection is to grasp the remarkable global consequences of changing field line connectivity within a localized region. In particular, the newly reconnected field lines have a large curvature $\mathbf{B} \cdot \nabla \mathbf{B}$ which produces a tension force — closely analogous to a stretched rubber band. Allowing for the possibility of pressure anisotropy, the MHD momentum equation perpendicular to magnetic field is

$$\rho \frac{d\mathbf{u}_\perp}{dt} = -\nabla_\perp \left(p_\perp + \frac{B^2}{2\mu_0} \right) + \left(1 + \frac{p_\perp - p_\parallel}{B^2/\mu_0} \right) \mathbf{B} \cdot \nabla \mathbf{B} - \mathbf{F}_{\text{col}\perp} \quad (13.26)$$

where \mathbf{F}_{col} is due to collisions between ions (viscous force) and between ions and neutral particles (frictional force). For newly reconnected field lines, the magnetic tension in Equation 13.26 drives an outflow jet approaching the Alfvén speed $V_{\text{out}} \approx V_A \equiv B_0/\sqrt{\mu_0\rho}$ (B_0 is the reconnecting field component). The resulting deficit in magnetic pressure pushes new field lines into the diffusion region with a maximum inflow velocity $V_{\text{in}} \sim (0.01 \rightarrow 0.1)V_A$. As this process continues, the larger stressed region is relaxed, leading to a reconfiguration of the global magnetic field on fast Alfvénic time scales $\sim L/V_A$. Assuming the diffusion regions remain small in comparison to the

global scales, most plasma enters the flow jet across the magnetic separatrices as illustrated in ?@fig-reconnection-configurations. This inflow is a consequence of changing the field line connectivity within the diffusion region, which causes the entire extent of the reconnected field lines to join the outflow. In this limit, the majority of the energy release is associated with the relaxation of field-line tension within outflow jets over long distances. Since the spatial extent of these jets is limited only by the macroscopic configuration, they are one of the most prominent signatures of reconnection in both in-situ and remote-sensing observations. Our primary focus is on situations where the available magnetic energy to drive reconnection is comparable or larger than the initial plasma thermal energy, corresponding to $\beta \equiv p/(B_0^2/2\mu_0) \lesssim 1$. In these regimes, global relaxation of field-line tension is the ultimate “engine” for reconnection and is essentially “ideal”, thus operating in a similar manner for most applications, but with a few important exceptions. Plasma heating preferentially along the magnetic field ($p_{\parallel} \gg p_{\perp}$) can weaken the magnetic tension force in Equation 13.26, whereas in partially ionized regimes the jet formation is more complicated due to interactions with neutrals. In addition, non-ideal kinetic physics may persist along magnetic separatrices (see red lines in ?@fig-reconnection-configurations) to larger distances, while in very large systems shocks may form along these boundaries and play a role in the energy conversion.

In contrast to the ideal physics driving the jet, the non-ideal terms within the diffusion region are intimately dependent upon the plasma conditions and spatial scales, and thus, a variety of different regimes are possible, as illustrated in the various panels of ?@fig-reconnection-configurations. Since the outflow is always energetically limited to V_A in a quasi-steady state, mass conservation implies that the geometry of the diffusion region determines the dimensionless reconnection rate, $R \equiv u_{\text{in}}/V_A \approx \Delta/L$ where Δ is the diffusion region thickness. The current understanding of the diffusion region physics has evolved over more than 60 years in three main stages, yet a full understanding remains elusive for large space and astrophysical problems.

13.8.1 Reconnection Rate

Reconnection rate quantifies the amount of magnetic flux transportation through the X-line. There are two kinds of reconnection rate (Nakamura et al. 2018):

1. The *dimensional* reconnection rate, defined as the out-of-plane electric (usually denoted as E_r or E_y for a 2D X-Z plane configuration, with a unit of [V/m]) at the X-line for steady state reconnection:

$$E_r \equiv E_y \sim U_{\text{in}} B_{\text{in}} \sim U_{\text{out}} B_{\text{out}} \quad (13.27)$$

Steady state is necessary for balancing the inflowing and outflowing flux and a uniform E_r . E_r defines the rate at which the magnetic flux is transferred from the inflow region into the diffusion region to change the field line connectivity. Therefore, the reconnection electric field is also referred to as the dimensional reconnection rate.

2. The *dimensionless* reconnection rate, defined as the ratio between the inflow speed and the ion Alfvén speed based on the background reconnecting field strength B_b and the upstream density:

$$R \equiv \frac{U_{\text{in}}}{V_{Aib}} \quad (13.28)$$

Because of the [magnetic flux conservation in ideal MHD](#), $B_{\text{in}}U_{\text{in}} = B_{\text{out}}U_{\text{out}}$. Assuming $B_{\text{in}} \sim B_b$ and $U_{\text{out}} \sim V_{Aib}$, the normalized reconnection rate can be approximately written by using the unnormalized rate $E_r \sim U_{\text{in}}B_{\text{in}}$ as

$$R = \frac{U_{\text{in}}}{V_{Aib}} \sim \frac{E_r}{V_{Aib}B_b} \sim \frac{E_r}{V_{\text{out}}B_b}$$

The inflow speed shall be taken near the reconnection site. Observationally, this ratio is on the order of $\mathcal{O}(0.1)$. There has been some theoretical arguments based on geometry to explain this universal ratio in recent years. We shall also note that magnetic flux is not conserved close to the center of X-line, in regions typically known as the ion diffusion region and electron diffusion region.

For global magnetosphere studies, a useful concept called *reconnection efficiency* ϵ is defined as ratio of the imposed potential drop to the full possible potential drop across the width of the magnetosphere. The potential drop is calculated from integrating E_r along the X-line. As an example, Ganymede has been observed to possess an unusually high reconnection efficiency about 60% (Zhou et al. 2020), compared to about 10% measured by cross polar cap potential (CPCP) over the full possible potential drop at Earth. High ϵ means more plasmas penetrating inside the magnetosphere instead of diverting around the magnetopause.¹

It is important to note that high reconnection efficiency does not indicate high dimensionless reconnection rate. For instance, in a 2D steady state magnetosphere reconnection, the global reconnection efficiency must be 1, otherwise the system cannot be steady. However, the local dimensionless reconnection rate can still be 0.1, measured from either the speed ratio or the ratio of B_N in the LMN coordinate system to the total magnetic field strength right inside the magnetopause (B. Ö. Sonnerup 1974).

13.8.2 Hall Field

Let X be the inflow direction, Z the outflow direction, and Y the out-of-plane direction. During magnetic reconnection,

- a quadrupolar out-of-plane or *Hall* magnetic field B_y is formed by the field aligned currents j_{\parallel} that flow in the vicinity of the magnetic separatrices when ion and electrons decouple on the length scales less than the ion inertial length (B. U. Ö. Sonnerup 1979).

¹I have a question here. If $\epsilon = 1$, does it mean that all the plasmas hitting the magnetopause will penetrate inside via reconnection? How should I demonstrate this, e.g. using a test particle model?

The quadruple Hall magnetic field was analyzed as a tearing mode eigenmode of the current layer (Terasawa 1983). The Hall effect introduces a coupling between the tearing mode perturbations (2D configuration in-plane \mathbf{B} components) and shear Alfvén mode perturbations (out-of-plane \mathbf{B} component).

- In the context of collisionless plasmas, ion-electron collisions are absent so that ions and electrons behave differently at micro-scale. On the ion scale, ions start to become demagnetized whereas electrons are still coupled to the magnetic field. The difference in the ion and electron motion produces *charge separation*, *Hall electric fields* and the *associated electric current*. (On the even smaller electron-scale, electrons are demagnetized and correspond to the “breaking” of the field line.) The spatial variation of Hall electric fields is mostly in X, normal to and pointing towards the current layer. Relevance of Hall effects in reconnection is considered important at the ion inertial scale, providing a source for non-ideal reconnection electric field. (In observations, the perpendicular scale of the Hall fields and current layer can be as small as electron scales as well?) The parallel scale of the Hall fields are much larger. Equivalently, Hall fields are associated with a wave vector \mathbf{k} so that $k_x \gg k_z$ and $1/k_x$ on the ion-scale.

The Hall fields are continuously produced quasi-steady structures accompanied with a timescale much longer than the ion gyroperiod. Their associated kinetic and magnetic energies propagate away from the reconnection layers carried by Alfvén waves. Numerical simulations suggest that it is primarily kinetic Alfvén waves (KAW, Section 10.8.4) which are excited in the vicinity of reconnection regions (Rogers+, 2001; Shay+, 2011), but further downstream in the exhaust it is primarily shear Alfvén waves (Gurram, Egedal, and Daughton 2021).

In observations, signatures of Hall magnetic fields are used as a working definition of collisionless magnetic reconnection in the Earth’s and planetary magnetosphere. The bipolar/quadrupolar Hall magnetic field B_y can coexist with a guide field (i.e. constant B_y) (for an example at Mars, see (J. Wang et al. 2023)).

(Dai and Wang 2023) emphasized that the physics underlying the Hall effect is similar to that in KAW:

- Both can be described by the two-fluid equations.
- The temporal and spatial scales of Hall fields are exactly the same as those of KAW.
- The polarity of the Hall electric field (in X) and the Hall magnetic field (in Y) is also consistent with that of KAW.
- The ratio E_x/B_y of Hall fields is on the order of V_A , similar to KAW.
- The current system of Hall fields are in the XZ-plane.

As the perpendicular wavelength of Alfvén wave becomes comparable to the ion gyroradius, the ions cannot follow the $\mathbf{E} \times \mathbf{B}$ drift. This is because the wave electric field now is non-uniform in the trajectory of ion gyromotion. As a consequence, the ion transverse motion is modified by the finite-Larmor-radius (FLR) effect. Considering the kinetic correction, the ion actually follows $\mathbf{E} = \mathbf{V}_i \times \mathbf{B} + \frac{1}{n_i q_i} \nabla \cdot \mathbf{P}_i$ in the transverse motion. The FLR effect is manifested in the

ion pressure gradient term: the FLR effect is equivalent to a finite temperature for the ions on the fluid level. At the wavelength of ion-scale wavelength, electrons cannot follow ions in the transverse motion. The difference in the transverse motion of ions and electrons produces a charge separation and coupling to the electrostatic mode, and a quasi-electrostatic electric field is formed in the transverse direction. To remain quasi-neutral, electrons quickly move in the direction parallel to the magnetic field. In the parallel direction, these electrons undergo a force balance between a small parallel electric field and the electron pressure gradient in the case of KAW. The parallel motion of electrons created a field-aligned current (FAC) that is a distinct nature of Alfvén mode. The FAC induces a wave magnetic field perpendicular to the DC (background) magnetic field. A new physical scale length r_{iL} , the ion gyroradius based on the electron temperature, for the current layer and outflow layer in fast reconnection. Through the diffusion region near the X-line, intense KAW turbulence is observed, suggesting that magnetic reconnection is a *source* of KAW. Thus KAW has been invoked as ingredients for fast reconnection.

In the direction parallel to the magnetic field (\hat{z}), B_y reverses but E_x remains its sign. This feature is related to the *superposition* of two KAW mode waves that propagate in the opposite direction and outward from the reconnection region. In the perpendicular direction, the Hall fields appears to exist only for one wavelength along \hat{x} . This is related to the eigenmode (standing mode) structure of KAW in the current layer.

In the literature, the Hall electric field is usually considered to arise from the $\mathbf{j} \times \mathbf{B}/n_e q$ term in the context of the generalized Ohm's law Equation 8.27. The Hall term starts to become important at the ion inertial scale, contributing to the Hall electric field. Interestingly, the $\mathbf{j} \times \mathbf{B}$ also represents a major force term in the one-fluid MHD momentum equation. The $\mathbf{j} \times \mathbf{B}$ force could be important on the large MHD-scale and the argument for the origin of Hall electric field becomes ambiguous; from the MHD momentum equation, $\mathbf{j} \times \mathbf{B}$ term is equivalent to the ion inertial effect, ion pressure effect and the electron pressure effect:

$$\rho \frac{d\mathbf{u}}{dt} + \nabla \cdot (\vec{P}_i + \vec{P}_e) = \mathbf{J} \times \mathbf{B} \quad (13.29)$$

which is derived from the sum of the two-fluid momentum equations multiplied by the mass of each species. The one-fluid MHD momentum equation Equation 13.29 together with the general Ohm's law is nearly equivalent (with some extent of reduction?) to the two-fluid momentum for two species. The Hall electric field is equivalently described by either set of equations.

From the ion momentum equation

$$\mathbf{E} + \mathbf{u}_i \times \mathbf{B} = \frac{1}{n_i q_i} \nabla \cdot \vec{P}_i + \frac{m_i}{q_i} \frac{d\mathbf{v}_i}{dt} \quad (13.30)$$

The perpendicular component of the non-ideal electric field $\mathbf{E} + \mathbf{u}_i \times \mathbf{B}$ is the Hall electric field at the ion kinetic scale; as seen on the RHS, two sources of Hall electric fields are the ion inertial term and the ion pressure gradient. From the KAW perspective, the Hall electric field

is supported by the ion pressure gradient. In the regime of the temporal and spatial scale of KAW, the ion inertial term is small and on the order of $\mathcal{O}(\omega^2/\Omega_i^2)$.

If the ion inertial term is neglected, Equation 13.30 becomes

$$\mathbf{E} + \mathbf{u}_i \times \mathbf{B} = \frac{1}{n_i q_i} \nabla \cdot \vec{P}_i$$

The $\mathbf{u}_i \times \mathbf{B}$ term is kept since the ion flow may not be neglected near the X-line.

13.8.3 Kinetic Signatures of Diffusion Region

(The following part is from [Kinetic signatures of the region surrounding the X-line in asymmetric (magnetopause) reconnection] and [Fluid and kinetics signatures of reconnection at the dawn tail magnetopause-Wind observations].)

Electron diffusion region (EDR) is defined to be the electron-scale region surrounding the X-line in which magnetic connectivity is ultimately broken. Note that this definition is fundamentally non-local in nature.

1. enhanced dissipation [e.g., Zenitani+, 2012]
2. non-gyrotropic particle behavior [e.g., Scudder+, 2008; Aunai+, 2013; Swisdak 2016]
3. electron distribution functions [e.g., Chen+, 2008, Ng+, 2011]
4. In the two inflow regions, we expect a reduction in total pressure, magnetic field B_L and the plasma density n_i relative to ambient upstream conditions.

Some signatures also exist downstream of the diffusion region and along the separatrices. In practice people use a complementary approach for the identification.

Specifically for EDR, it should exhibit a number of properties:

1. the violation of the electron frozen-in condition
2. non-gyrotropic electron distributions [e.g., Scudder+, 2008; Aunai+, 2013; Swisdak 2016]
3. enhanced dissipation [Zenitani+, 2011]

A simple and practical indicator of EDR is the presence of a sunward pointing E_N at the midplane (called the “shoulder”) as this signature coincides with the region of enhanced dissipation, non-gyrotropic electrons at midplane, and counterstreaming electron beams due to electron meandering orbits around the X-line. This EN signature is straightforward to measure experimentally because it is the largest component of the electric field at the magnetopause.

The crescent shape distribution is not as localized as the E_N shoulder especially for ion.

At the dayside magnetopause, reconnection between magnetosheath and geomagnetic field lines is expected to produce

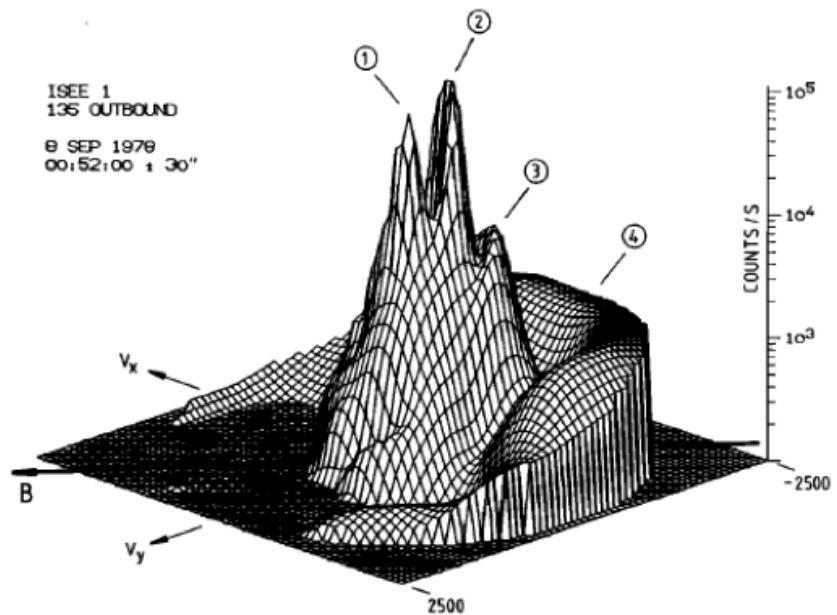


Fig. 12. Relief plot of two-dimensional count rate distributions in the GSE (v_x, v_y) plane for velocities up to 2500 km s $^{-1}$, measured just after the final magnetopause crossing on September 8, 1978. The plot is based on a one-minute accumulation of count rates in 16 velocity and azimuth channels. The three peaks near the center represent (1) the main magnetosheath plasma, (2) ions which have been reflected at the magnetopause, and (3) a small contribution of He $^{+}$ ions, respectively. The crescent-shaped distribution at high energies (4) is interpreted as ions of magnetospheric (ring current) origin leaking out along reconnected field lines. They are seen to move preferentially antiparallel to the projected magnetic field B , as expected for a crossing north of the X line (Figure 1b). The abrupt outer termination of the distribution of energetic ions reflects the high-energy cutoff of the instrument at 40 keV.

Figure 13.7: Ion crescent shape distribution at Earth's magnetopause revealed by ISEE satellite at 1981.

1. a finite magnetic field component normal to the magnetopause, B_N ;
2. Alfvénic plasma flow acceleration associated with a rotational discontinuity at the magnetopause;
3. ion distributions on reconnected field lines consisting of a mixture of magnetosheath and magnetosphere populations where the transmitted magnetosheath population has a “D-shaped” distribution, with a low-energy cutoff at the deHoffmann-Teller velocity [Cowley, 1982];
4. reflected ions in the layers adjacent to the magnetopause;
5. opposite streaming along reconnected field lines of outgoing magnetospheric electrons and incoming magnetosheath electrons, resulting in large parallel electron heat flux;
6. an offset between the ion and electron edges at the inner boundary of the low-latitude boundary layer (LLBL) due to a time-of-flight effect resulting from the fact that entering magnetosheath electrons have much higher parallel speeds than ions while their transverse motions are the same.

13.8.4 D-shape Distribution

The kinetic description of reconnection can also be quantitatively verified. An important kinetic signature is the “D-shaped” ion distribution. Magnetosheath particles can either be reflected at or cross the magnetopause. In the deHoffmann-Teller frame of reference in which the electric field vanishes, only ions traveling toward the magnetopause will cross the magnetopause. Thus, when viewed in the spacecraft frame, only magnetosheath ions with parallel velocity greater than the deHoffmann-Teller velocity can be seen earthward of the magnetopause, resulting in a “D-shaped” distribution. [Cowley, 1982]

The existence of a finite BN at the MP requires field lines on both sides of the MP to move together. In this scenario there must be a reference frame (the HT frame) which slides along the MP with the field-line velocity. In this frame, the convective electric field vanishes ($\mathbf{E}_c = -\mathbf{v} \times \mathbf{B} = 0$), i.e. the flows are field aligned on the two sides of the MP.

In practice, the HT frame for a set of plasma and field measurements across the MP can be found as the reference frame in which the mean square of the convective electric field, $D = \langle |(\mathbf{v} - \mathbf{V}_{HT}) \times \mathbf{B}|^2 \rangle$, is as small as possible. The angle bracket $\langle \dots \rangle$ denotes an average of an enclosed quantity over a set of measurements. The velocity \mathbf{v} for which $D(\mathbf{v})$ is a minimum is the deHoffmann-Teller velocity, \mathbf{V}_{HT} . The ratio D/D_0 , where $D_0 = \langle |(\mathbf{v} \times \mathbf{B})|^2 \rangle$, is used as a measure of the quality of the HT frame. For a good HT frame D/D_0 should be small ($\ll 1$).

In another paper [Characteristics of the flank magnetopause: Cluster observations], \mathbf{V}_{HT} is defined by minimizing $\mathbf{E}' = \mathbf{E} + \mathbf{V}_{HT} \times \mathbf{B}$, and the correlation between $\mathbf{E}_c = -\mathbf{v} \times \mathbf{B}$ and $\mathbf{E}_{HT} = -\mathbf{V}_{HT} \times \mathbf{B}$ describes how well the frame is determined.

13.8.5 Classification of Discontinuity Types

In a fluid description, the magnetopause can be described either as a tangential discontinuity (TD) or as a rotational discontinuity (RD). A TD implies a complete separation of two plasma regimes (in this case the magnetosheath on one side and the magnetosphere on the other side). The boundary as a whole may move, but there is no transport of plasma across the discontinuity, and there is no magnetic field along the boundary normal. An RD-like magnetopause, on the other hand, implies transport across the boundary and a normal magnetic field, and indicates the presence of reconnection. In the vicinity of the X-line, the plasma flow is Alfvénic, i.e., the Walén relation is satisfied:

$$\mathbf{v} - \mathbf{V}_{HT} = \pm \mathbf{V}_A$$

where \mathbf{v} is the plasma velocity, \mathbf{V}_{HT} is the deHoffmann-Teller frame velocity, and \mathbf{V}_A is the local Alfvén velocity.

In some other literature, a more complete form is written as

$$\mathbf{v} - \mathbf{V}_{HT} = \pm(1 - \alpha)^{1/2} \mathbf{V}_A$$

which takes the pressure anisotropy factor $\alpha = (p_{\parallel} - p_{\perp})\mu_0/B^2$ into account.

This is often used to classify the discontinuity type of the magnetopause. For an RD, the flow across the boundary is proportional to the normal magnetic field, i.e., $v_n \propto B_n$. A positive (negative) slope of the regression means that normal magnetic field and flow have the same (opposite) signs. At the magnetopause, we can assume that the flow is inward, i.e., from the magnetosheath into the magnetosphere.

13.8.6 Reconnection Efficiency

This is very confusing. From [Kivelson+ 1997]:

There is an interesting link between flow speeds over the polar cap and reconnection efficiency at the nose of its magnetopause. This follows from the fact that the voltage drop across the magnetosphere is a fraction of the voltage drop across the same distance in the corotating flow upstream. Thus, the convection electric field within the magnetosphere is a fraction of the corotation electric field determined by the efficiency of reconnection. The model allows us to estimate the convective flow speeds of plasma over the polar cap if the reconnection efficiency is known, and conversely allows us to estimate the reconnection efficiency if the flow speed over the polar cap is known.

Let me think in this way: the convective electric field in the upstream, $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, describes how many magnetic field lines are moving into the reconnection site at a given point. The integral of electric field along the X-line describes how many magnetic field lines are getting

reconnected into the magnetosphere. This is not the same as electric potential drop, where in the simplest cases requires a static EM field. Because of the difficulty of finding the separatrices, we use an indirect approach assuming equi-potential field lines to calculate the electric field along the magnetosphere boundary curve away from the reconnection site.

Be careful for the word “cross polar cap potential” here. This may be a misleading term, since how can one define an electric potential in a EM field?

13.8.7 Dissipation Mechanisms

Perhaps the most import problem associated with reconnection is the understanding of the mechanisms by which the magnetic field can dissipate its energy, and subsequently produce particle heating and acceleration [Huba 1979]. Since many field reversed plasmas are essentially collisionless, instabilities are likely to play an important role in the dissipation process.

A macroinstability (MHD type), like the tearing mode can dissipate magnetic energy even in a collisionless plasma via electron inertia, electron or ion Landau damping. However, the collisionless tearing mode saturates at a very small amplitude. Microinstabilities can also dissipate magnetic energy by producing an anomalous resistivity which can either dissipate the magnetic energy directly or enhance the growth of the tearing mode.

In any case, the kinetic physics corresponding to the fast rate of collisionless reconnection is a long-standing question. *Non-gyrotropy particle distributions* and the associated *off-diagonal term of the pressure tensor* have been considered to provide the reconnection electric field at electron-scale and ion-scale.

The *lower-hybrid-drift instability* (LHDI, Section 13.9) is often considered to be an important microinstability for dissipation near the diffusion region.

13.8.8 Sweet-Parker Solution

From Yi-Hsin Liu’s presentation of space weather.

mass conservation:

$$\nabla \cdot (n\mathbf{V}) = 0 \Rightarrow V_{in}L \simeq V_{out}\delta$$

momentum conservation:

$$\frac{\mathbf{B} \cdot \nabla \mathbf{B}}{4\pi} = nm_i \mathbf{V} \cdot \nabla \mathbf{V} \Rightarrow V_{out} \simeq \frac{B}{\sqrt{4\pi nm_i}} = V_A$$

normalized reconnection rate

$$R \equiv \frac{V_{in}}{V_A} \sim \frac{\delta}{L}$$

However, this model has a small δ/L , the rate is too small to explain the time-scales in solar flare. To explain the flares, it requires $R \sim 0.1$.

13.8.9 Petschek Solution

Reconnection in this model is much larger because $R \sim \delta/L$ goes up. However, this is not a self-consistent solution.

In PIC simulations, the diffusion region is localized like the Petschek solution. The reason we use PIC instead of MHD is that PIC captures the key physics that breaks the frozen-in condition in nature. In the GEM challenge study, PIC, hybrid, Hall-MHD, MHD with a localized artificial resistivity all give similar $R \sim 0.1$ fast reconnection rate in disparate systems. In terms of the pressure descriptions from the moment equations, it has been found that 6-moment cannot trigger fast reconnection while 10-moment can, which essentially indicates that the off-diagonal terms in the pressure tensor is critical in getting the right physics.

One possible explanation for this 0.1 rate comes from geometrical consideration. In the small δ/L limit, $R \sim \delta/L \sim 0$. In the large δ/L limit, $\delta/L \rightarrow 1$, $R \rightarrow 0$. It turns out that there should be an optimized R_{\max} in between. In the large δ/L limit,

$$\frac{\mathbf{B} \cdot \nabla \mathbf{B}}{4\pi} \simeq \frac{\nabla(B^2)}{8\pi} + nm_i \mathbf{V} \cdot \nabla \mathbf{V}$$

At the inflow region, the large angle decrease the reconnecting magnetic field. At the outflow region, it decrease the outflow speed. Both of them cause R to decrease. Constraints imposed at the inflow and outflow region (upper) bound the rate! In [Liu+, PRL 2017], an analytical expression of R is given by

$$R \equiv \frac{V_{in}}{V_A} \simeq \frac{\delta}{L} \cdot \underbrace{\left[\frac{1 - (\delta/L)^2}{1 + (\delta/L)^2} \right]}_{\text{reduction of reconnection } B} \cdot \underbrace{\sqrt{1 - \left(\frac{\delta}{L} \right)^2}}_{\text{reduction of } V_{out}}$$

Reconnection tends to proceed near the most efficient state with $R \sim \mathcal{O}(0.1)$. Nicely, the rate is insensitive to δ/L near this state (BECAUSE THE SLOPE IS SMALL??).

Amitava Bhattacharjee argued that scaling is a controversial subject:

1. How does it scale with ion/electron skin depth, resistivity, plasma beta, and system size?
 2. Is the reconnection rate insensitive to the details of the electron layer (current sheet layer), and controlled by ions?
- The whistler waves generate an out-of-plane quadrupolar magnetic field.

- The ratio of the horizontal electron outflow to the horizontal magnetic perturbation scales as k for the dispersive whistler (or kinetic Alfvén) wave. (WHICH WAVE IS IT? I have been confused by the relation between whistler waves and KAWs. Both contain a region with super-Alfvénic phase speed for large k and especially with frequency between Ω_i and Ω_e ; whistler wave involves electrons whereas KAW can only rely on ions.)

From the GEM Challenge perspective:

- Reconnection is insensitive to the mechanism that breaks field lines (electron inertia or resistivity). The length of the reconnection layer $\Delta_i \sim 10d_i$. Reconnection rate is a “universal constant”, $V_i n \approx 0.1V_A$.
- In the presence of Hall currents, whistler waves mediate reconnection. The characteristic outflow speed is the whistler phase speed (based on the upstream magnetic field).
- The inflow velocity $v_{\text{in}} = \Omega_e \delta_e^2 / \Delta_i$ where $\Delta_i \sim k^{-1} \ll L$ (system size). This rate is independent of m_e .

The GEM perspective is not universally accepted. An alternate point of view provides evidence that:

- Reconnection is not a universal constant, and depends on system parameters (such as ion/electron skin depth, plasma beta, boundary conditions).
- Reconnection rate is not independent of the system size, and in fact, often decreases as the system size increases.

He provided three examples:

1. Forced reconnection without guide field
2. Undriven reconnection with guide field
3. Undriven reconnection with open boundaries

13.8.10 3D Nature of Reconnection

How about the freedom coming from the extra dimension?

Distinct 3D features, including

- flux ropes
- kink instability
- turbulence

3D diffusion region can be fundamentally different. For example, you may find bifurcation of electron diffusion region in 3D reconnection simulations. Clue: bifurcated layer is located in between these intertwined flux ropes, and tearing modes give rise to these flux ropes! 2D only allows the parallel tearing mode, i.e., no bifurcation. 3D allows a spectrum of oblique tearing modes, unlike 2D.

3D reconnection has always been a mystery to me. People have the concepts of 2D reconnection for so long ([guide field on v.s. guide field off](#)) which makes it hard for us to understand the 3D nature of reconnection. A fair question to ask is: what differentiate a guide field from the reconnecting field?

13.9 Lower Hybrid Drift Instability

The *lower-hybrid-drift instability* (LHDI) is a microinstability which has thought to be an anomalous transport mechanism in both laboratory and space plasmas. The local linear theory of this instability is well understood. The mode is driven by the diamagnetic current produced by pressure inhomogeneities and is characterized at maximum growth by

$$\begin{aligned}\omega &\sim kV_d \lesssim \omega_{lh} \\ \gamma &\lesssim \omega \\ k &\sim \omega_{lh}/v_i \\ \mathbf{k} \cdot \mathbf{B} &= 0\end{aligned}$$

where V_d is the diamagnetic drift velocity, ω_{lh} is the lower hybrid frequency, and v_i is the ion thermal velocity.

$$\omega_{lh} = [(\Omega_i \Omega_e)^{-1} + \omega_{pi}^{-2}]^{-1/2}$$

where Ω_i is the ion cyclotron frequency, Ω_e is the electron cyclotron frequency, and ω_{pi} is the ion plasma frequency. It is so called “hybrid” because it is a mixture of two frequencies. It is unusual in that the ion and electron masses play an equally important role.

The drift velocity is

$$V_d = \frac{\nabla p \times \mathbf{B}}{neB^2}$$

Physically, the instability is reactive (fluid-like) in the strong drift velocity regime ($V_d > v_i$) and dissipative (kinetic) in the weak drift velocity regime ($V_d < v_i$). The pressure inhomogeneities can be caused by either temperature or density inhomogeneities, or both.

The corresponding lower hybrid (LH) waves are electrostatic emissions near the LH resonant frequency. They propagate perpendicularly with a small wavelength comparable to Larmor radius of thermal particles and can be capable of heating both ions and electrons. In the cold plasma regime (Section 10.6), the LH waves (with the frequency above LH resonance frequency) and whistler-mode waves belong to the same wave mode branch.

13.10 Kinetic Mode

13.11 Pressure Anisotropy Instabilities

When we shift to the MHD with anisotropic pressure tensor

$$P_{ij} = p_{\perp} \delta_{ij} + (p_{\parallel} - p_{\perp}) B_i B_j / B^2$$

where p_{\perp} and p_{\parallel} are the pressures perpendicular and parallel w.r.t. the magnetic field, respectively. For the strong magnetic field approximation, the two pressures are related to the plasma density and the magnetic field strength by two adiabatic equations,

$$\begin{aligned}\frac{d}{dt} \left(\frac{p_{\parallel} B^2}{\rho^3} \right) &= 0 \\ \frac{d}{dt} \left(\frac{p_{\perp}}{\rho B} \right) &= 0\end{aligned}$$

This is also known as the *double adiabatic theory*, which is also what many people remember to be the key conclusion from the CGL theory ((Chew, Goldberger, and Low 1956), Equation 8.43 and Equation 8.45). Here I want to emphasize the meaning of *adiabatic* again: this assumes zero heat flux. If the system is not adiabatic, the conservation of these two quantities related to the parallel and perpendicular pressure is no longer valid, and additional terms may come into play such as the stochastic heating.

Now imagine an increasing magnetic field B . For an initially Maxwellian distributed plasma like a circle in the phase space, it will get stretched in the perpendicular direction and become an oval; conversely, if the magnetic field is decreasing, it will get stretched in the parallel direction.

13.11.1 Firehose Instability

The firehose mode is the result of a pressure (or temperature) anisotropy in plasma with the parallel pressure P_{\parallel} exceeding the perpendicular P_{\perp} and magnetic $B^2/2\mu_0$ pressures. Sagdeev [1966] gave a simple intuitive explanation of this instability based on the insight that the parallel thermal motion of the adiabatic magnetised ions along the magnetic field exerts a centrifugal force on the field lines. When this force exceeds the restoring forces of the magnetic pressure and perpendicular plasma pressure, the centrifugal force wins and a small excursion of the magnetic field starts growing and propagates as a wave along the magnetic flux tube like on a string. The condition for instability is

$$P_{\parallel} - P_{\perp} > B^2/\mu_0$$

Since the pressure anisotropy on the left means that there is an excess in parallel energy in the plasma, the plasma possesses free energy which by the instability is fed into the excitation of Alfvén waves with frequencies $\omega_A \ll \omega_{ci}$, transported away with Alfvén speed and ultimately dissipated in some way. The waves excited are ordinary Alfvén waves.

If a wave growth is associated with small groups of resonant particles, then the corresponding instability is known a resonant mode. The firehose mode is non-resonant since all particles contribute to it. Generally most non-resonant instabilities can only be found by solving the full dispersion relation numerically.

See more in Section 14.1.

13.11.2 Mirror Instability

[Southwood & Kivelson](#) had a nice paper explaining the physics of mirror instability. See also Section 14.2 for a more thorough mathematical description of the instability.

The mirror instability is prevalent in planetary and cometary magnetosheaths and other high β environment. It is recognized as one of the two magnetohydrodynamic instabilities that occur in the presence of extreme velocity space (pitch angle) anisotropy in a uniform plasma, the other instability being the firehose. The mirror mode has gained increasing interest following its identification in spacecraft data from the magnetosheath and solar wind. Although the instability was originally derived from magnetohydrodynamic fluid theory, later work showed that there were significant differences between the fluid theory and a more rigorous kinetic approach, as we will see in Section 14.2.

The instability occurs when β , the ratio of plasma to magnetic pressure, satisfies the following criterion:

$$\beta_{\perp} \left(\frac{P_{\perp}}{P_{\parallel}} - 1 \right) < 1$$

The anisotropy $(p_{\perp} - p_{\parallel})/p_{\parallel}$ required for instability $\propto \beta^{-1}$. In planetary magnetosheaths the source of anisotropy is likely to be the planetary bow shock and in cometary environments the ion pickup process is a natural source of anisotropy. Both types of environment tend to have relatively large values of β .

Any nonlinear saturation mechanism of the mirror instability is likely to leave the plasma spatially structured, as is also strongly suggested by the many observations. In practice, in almost any experimental detection of a plasma instability, the wave fields and the plasma population will have evolved to some quasi-steady condition that represents a nonlinear saturated state of the instability. In the spatially structured magnetic field associated with the minor instability in both its linear and nonlinear phase, different parts of the ion distribution will resonate with an ion cyclotron wave as the wave propagates along the inhomogeneous field. This effect is likely to inhibit the growth of the ion cyclotron mode, which theoretically also exists in the same frequency band.

The mirror instability is referred to as a “fluid” instability, alluding to the fact that the phase space (pitch angle) anisotropy of the bulk of the hot plasma distribution serves as the source of energy. The instability grows because of a subtle coupling between a group of particles with small velocity parallel to the field and the rest of the population. The mirror instability has zero parallel phase velocity in the plasma frame of reference. It follows, by analogy with other uses of the term that we can call particles with near zero parallel velocity *resonant*.

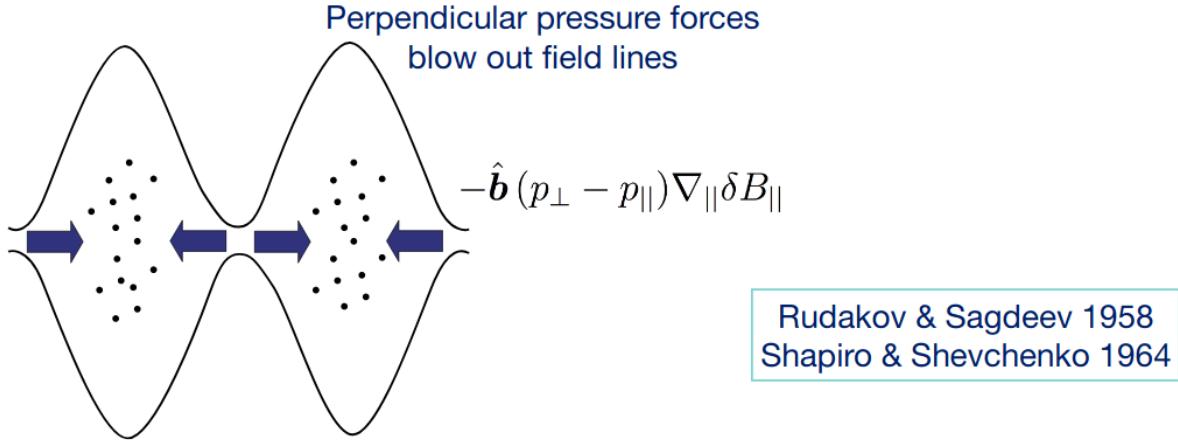


Figure 13.8: Illustration of mirror instability.

(Song+ 1994) presented a set of identification criteria for distinguishing among various forms of high β MHD waves, with mirror modes having the properties:

- Compressional fluctuations:

$$(\delta \mathbf{B}^2 - \delta \mathbf{B}_{\parallel}^2)/\delta \mathbf{B}_{\parallel}^2 < 1$$

- Anti-correlated δP_i and δP_B :

$$\delta P_i/\delta P_B < 0$$

- Linear magnetic field polarization

- Stationary in plasma rest frame (i.e. zero phase speed):

$$(\delta \mathbf{v}^2 / \mathbf{v}_0^2) / (\delta \mathbf{B}^2 / \mathbf{B}_0^2) \approx 0$$

The last criterion is referred to as the Doppler ratio and is used to distinguish between slow mode waves (value ≥ 1) and mirror modes (value near zero).

Note that mirror mode is not really a MHD mode, but from observation in the Earth’s magnetosheath we often see mirror modes on large MHD scales. However, in the solar wind, mirror mode structures seem to be much smaller than those observed in the magnetosheath. From Equation 14.20, electron physics may be involved on the electron scales.

It is still an unknown mystery from MMS observation that downstream of an interplanetary shock, very few mirror modes were observed. This is surprising, because usually the shock is a source of pressure anisotropy. Another explained fact is that typically mirror modes are not observed during fast solar wind (as of 2022).

13.11.3 Weibel Instability

The Weibel instability is a plasma instability present in homogeneous or nearly homogeneous electromagnetic plasmas which possess an anisotropy in velocity space. This anisotropy is most generally understood as two temperatures in different directions, and the generalization extends to superposition of many counter-streaming beams. In this sense, it is like the two-stream instability except that the perturbations are electromagnetic and result in filamentation as opposed to electrostatic perturbations which would result in charge bunching. In the linear limit the instability causes exponential growth of electromagnetic fields in the plasma which help restore momentum space isotropy.

Check [Wiki](#) for an illustration of Weibel instability in a simple configuration. F. F. Chen (2016) also has a section on this.

Weibel instability is common in astrophysical plasmas, such as collisionless shock formation in supernova remnants and γ -ray bursts.

13.12 Cyclotron Instability

I still don't see the dispersion relations for these types of instabilities...

13.12.1 Ion Cyclotron Instability

There is another instability (besides mirror instability) that occurs at frequencies below the ion gyrofrequency in the presence of ion pitch angle anisotropy, the *electromagnetic ion cyclotron instability* (EMIC). This is a resonant instability in which the energy for the instability is fed from a subset of the particle population that are in gyroresonance with the unstable wave.

From Earth's magnetosheath observations, mirror modes dominates even though the linear dispersion relation predicts a smaller growth rate compared with EMIC waves. Gary argued in the 1990s that this is because of a small portion of heavier species (e.g. helium) that modifies the growth rate. Local hybrid simulations (e.g. the example from hybrid VPIC in ISSS14) supports this. However, researchers have not reached a consensus on whether ion cyclotron instability suppress mirror instability or vice versa.

14 Kinetic MHD

This is taken from the hand-written lecture notes from Prof. Alexander Schekochihin. Theoretical physicists love CGS units, but I tend to use SI units here. In later part of the note, there may be mixed units, so be careful.

We start by showing how for magnetized, weakly collisional plasmas ($\nu_{\text{colli}} \ll \Omega_s$, $r_L \ll \lambda_{\text{mfp}}$, where λ_{mfp} is the mean free path), low-frequency ($\omega \ll \Omega_s$), long-wavelength ($kr_L \ll 1$) dynamics can be decided by a set of equations that look almost like the familiar MHD. We will see later on that the ways in which they are not MHD will profoundly affect the dynamics — indeed we do not fully understand the full implication of this in high- β plasmas. This is consequently one of the frontier topics in theoretical plasma astrophysics.

Let us start from first principles. Any plasma that is going to be of interest to us is described by the Vlasov-Maxwell-Landau system of equations:

$$\frac{\partial f_s}{\partial t} + \mathbf{v} \cdot \nabla f_s + \frac{q_s}{m_s} [\mathbf{E} + \mathbf{v} \times \mathbf{B}] \frac{\partial f_s}{\partial \mathbf{v}} = C(f_s) \quad (14.1)$$

The Maxwell's equation can be simplified based on our assumptions.

$$\nabla \cdot \mathbf{E} = \epsilon_0 \sum_s q_s n_s, \quad n_s = \int d\mathbf{v} f_s$$

$\nabla \cdot \mathbf{E}$ is small when $k^2 \lambda_{De} \ll 1$, and this simply gives the quasi-neutrality condition.

$$\nabla \cdot \mathbf{B} = 0$$

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j} + \cancel{\epsilon_0 \mu_0 \frac{\partial \mathbf{E}}{\partial t}}, \quad \mathbf{j} = \sum_s n_s q_s \mathbf{u}_s, \quad \mathbf{u}_s = \frac{1}{n_s} \int d\mathbf{v} \mathbf{v} f_s$$

The displacement current can be neglected since $\omega \ll kc$ for low frequency waves and non-relativistic motions.

Intuitively, we tend to think of the plasma as a fluid (or a multi-fluid of several species) with some density n_s , velocity \mathbf{u}_s and perhaps pressure, temperature, etc. This is rooted in our experience with collisional gases ($\nu \gg \omega$), which are in local Maxwellian equilibrium:

$$f_s = \frac{n_s}{(\pi v_{\text{th},s}^2)^{3/2}} e^{-\frac{(\mathbf{v}-\mathbf{v}_s)^2}{v_{\text{th},s}^2}}, \quad v_{\text{th},s} = \sqrt{\frac{2k_B T_s}{m_s}}$$

where n_s , \mathbf{u}_s and T_s are governed by fluid equations.

With this desire to think of plasmas as fluid, let us break the motion of the particles into two parts:

$$\mathbf{v} = \mathbf{u}_s(t, \mathbf{r}) + \mathbf{w}$$

where \mathbf{u}_s represents the mean velocity of species s (fluid-like description) and \mathbf{w} represents the “peculiar” velocity or internal motion (kinetic description). This amounts to a transformation of variables

$$(t, \mathbf{r}, \mathbf{v}) \rightarrow (t, \mathbf{r}, \mathbf{w}), \quad \mathbf{w} = \mathbf{v} - \mathbf{u}_s(t, \mathbf{r})$$

under which the derivatives in the new basis shall be written as

$$\begin{aligned} \frac{\partial}{\partial t} &\rightarrow \frac{\partial}{\partial t} - \frac{\partial \mathbf{u}_s}{\partial t} \cdot \frac{\partial}{\partial \mathbf{w}} \\ \nabla &\rightarrow \nabla - (\nabla \mathbf{u}_s) \cdot \frac{\partial}{\partial \mathbf{w}} \\ \frac{\partial}{\partial \mathbf{v}} &\rightarrow \frac{\partial}{\partial \mathbf{w}} \end{aligned} \tag{14.2}$$

This can be derived from the chain rule with

$$\begin{aligned} t' &= t \\ \mathbf{r}' &= \mathbf{r} \\ \mathbf{w} &= \mathbf{v} - \mathbf{u}_s(t, \mathbf{r}) \end{aligned}$$

Note that the three variables are independent, for any quantity f , we have

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{\partial f}{\partial t'} \cdot \frac{\partial t'}{\partial t} + \frac{\partial f}{\partial \mathbf{r}'} \cdot \frac{\partial \mathbf{r}'}{\partial t} + \frac{\partial f}{\partial \mathbf{w}} \cdot \frac{\partial \mathbf{w}}{\partial t} \\ &= \frac{\partial f}{\partial t} - \frac{\partial \mathbf{u}_s}{\partial t} \cdot \frac{\partial}{\partial \mathbf{w}} \end{aligned}$$

Similarly we can derive the other two relations in Equation 14.2. Then the Boltzmann equation becomes

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_s \cdot \nabla \right) f_s + (\mathbf{w} \cdot \nabla) f_s + \left(\frac{q_s}{m_s} \mathbf{w} \times \mathbf{B} + \mathbf{a}_s - \mathbf{w} \cdot \nabla \mathbf{u}_s \right) \cdot \frac{\partial f_s}{\partial \mathbf{w}} = C(f_s) \tag{14.3}$$

where

$$\mathbf{a}_s = \frac{q_s}{m_s} (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) - \frac{d\mathbf{u}_s}{dt}$$

and now we always have $\int d\mathbf{w} \mathbf{w} f_s = 0$ by definition. The strategy now is to take moments of Equation 14.3. The zeroth-order moment ($\int d\mathbf{w}$) gives

$$\begin{aligned} & \int \frac{df_s}{dt} d\mathbf{w} + \int (\mathbf{w} \cdot \nabla) f_s d\mathbf{w} + \int [\dots] \cdot \frac{\partial f_s}{\partial \mathbf{w}} d\mathbf{w} = 0 \\ & \frac{d}{dt} \int f_s d\mathbf{w} + \nabla \cdot \cancel{\int \mathbf{w} f_s d\mathbf{w}} - \int (\mathbf{w} \cdot \nabla \mathbf{u}_s) \cdot \frac{\partial f_s}{\partial \mathbf{w}} d\mathbf{w} = 0 \\ & \frac{dn_s}{dt} + \int f_s \frac{\partial}{\partial \mathbf{w}} (\mathbf{w} \cdot \nabla \mathbf{u}_s) d\mathbf{w} = 0 \\ & \frac{dn_s}{dt} + \int f_s \frac{\partial}{\partial w_j} (w_i \frac{\partial}{\partial x_i} u_{sj}) dw_j = 0 \\ & \frac{dn_s}{dt} + \int f_s \delta_{ij} \frac{\partial}{\partial x_i} u_{sj} dw_j + \int f_s w_i \cancel{\frac{\partial^2 u_{sj}}{\partial x_i \partial w_j}} dw_j = 0 \\ & \frac{dn_s}{dt} + \int f_s \frac{\partial}{\partial x_i} u_{si} dw_i = 0 \\ & \frac{dn_s}{dt} + \nabla \cdot \mathbf{u}_s \int f_s d\mathbf{w} = 0 \\ & \frac{dn_s}{dt} + (\nabla \cdot \mathbf{u}_s) n_s = 0 \end{aligned}$$

or

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{u}_s) = 0 \quad (14.4)$$

The first-order moment ($\int d\mathbf{w} m_s \mathbf{w}$) gives

$$\nabla \cdot \int d\mathbf{w} m_s \mathbf{w} \mathbf{w} f_s - m_s n_s \mathbf{a}_s = \int d\mathbf{w} m_s \mathbf{w} C(f_s) \equiv \mathbf{R}_s$$

where $\int d\mathbf{w} m_s \mathbf{w} \mathbf{w} f_s = \mathbf{P}_s$ is the pressure tensor and \mathbf{R}_s is the collisional friction. Unpacking \mathbf{a}_s , we have the momentum equation for each species s

$$m_s n_s \frac{d\mathbf{u}_s}{dt} = -\nabla \cdot \mathbf{P}_s + q_s n_s (\mathbf{E} + \mathbf{u}_s \times \mathbf{B}) + \mathbf{R}_s \quad (14.5)$$

Summing over all the species,

$$\begin{aligned} \sum_s m_s n_s \frac{d\mathbf{u}_s}{dt} &= -\nabla \cdot \sum_s \mathbf{P}_s + \cancel{\sum_s q_s n_s \mathbf{E}} + \sum_s q_s n_s \mathbf{u}_s \times \mathbf{B} + \cancel{\sum_s \mathbf{R}_s} \\ \rho \frac{d\mathbf{u}}{dt} &= -\nabla \cdot \mathbf{P} + \mathbf{j} \times \mathbf{B} \\ \rho \frac{d\mathbf{u}}{dt} &= -\nabla \cdot \mathbf{P} + \mu_0^{-1} (\nabla \times \mathbf{B}) \times \mathbf{B} \\ \rho \frac{d\mathbf{u}}{dt} &= -\nabla \cdot \left[\mathbf{P} + \frac{B^2}{2\mu_0} \mathbf{I} - \mathbf{B}\mathbf{B} \right] \end{aligned}$$

It is useful to emphasize that $d/dt = \partial/\partial t + \mathbf{u} \cdot \nabla$. Later we will see the notation of D/Dt , which is used to remind us of the fact that \mathbf{w} is involved.

We also need an equation for the magnetic field. It is Faraday's law:

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$$

From Equation 14.5,

$$\mathbf{E} = -\mathbf{u}_s \times \mathbf{B} + \frac{\nabla \cdot \mathbf{P}_s}{q_s n_s} - \frac{\mathbf{R}_s}{q_s n_s} + \frac{m_s}{q_s} \frac{d\mathbf{u}_s}{dt}$$

Based on the following arguments:

- $\nabla \cdot \mathbf{P}_s / (q_s n_s)$ is small since $k r_s / M_A \ll 1$ (long wave + incompressible plasma???)
- $\mathbf{R}_s / (q_s n_s)$ is small since $\nu_s / \Omega_s \ll 1$,
- $(m_s / q_s) d\mathbf{u}_s / dt$ is small since $\omega / \Omega_s \ll 1$

we have the simplest Ohm's law and in turn $\mathbf{u}_s = \mathbf{E} \times \mathbf{B} / B^2 = \mathbf{u}_\perp$, the perpendicular component of the velocity is the same for all species. Then we get the induction equation from Faraday's law:

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) \quad (14.6)$$

or

$$\frac{d\mathbf{B}}{dt} = \mathbf{B} \cdot \nabla \mathbf{u} - \mathbf{B} \nabla \cdot \mathbf{u} \quad (14.7)$$

The three equations we have so far are very similar to MHD, except for the pressure tensor. Obviously, all the kinetic magic is hidden in \mathbf{P} .

Going back to Equation 14.3, it is key to notice that

$$\frac{q_s}{m_s} \mathbf{w} \times \mathbf{B} \cdot \frac{\partial f_s}{\partial \mathbf{w}} = -\Omega_s \left(\frac{\partial f_s}{\partial} \right)_{w_\perp, w_\parallel}$$

where θ is the gyroangle in the perpendicular plane. This can be proved by changing to cylindrical coordinates

$$\mathbf{w} = (w_{\perp} \cos \theta, w_{\perp} \sin \theta, w_{\parallel})$$

with changing of variables:

$$\begin{aligned} \frac{\partial f_s}{\partial \theta} &= \frac{\partial f_s}{\partial w_{\perp 1}} \frac{\partial w_{\perp 1}}{\partial \theta} + \frac{\partial f_s}{\partial w_{\perp 1}} \frac{\partial w_{\perp 1}}{\partial \theta} + \frac{\partial f_s}{\partial w_{\parallel}} \frac{\partial w_{\parallel}}{\partial \theta} \\ &= \frac{\partial f_s}{\partial w_{\perp 1}} \frac{\partial w_{\perp} \cos \theta}{\partial \theta} + \frac{\partial f_s}{\partial w_{\perp 2}} \frac{\partial w_{\perp} \sin \theta}{\partial \theta} \\ &= -\frac{\partial f_s}{\partial w_{\perp 1}} w_{\perp} \sin \theta + \frac{\partial f_s}{\partial w_{\perp 2}} w_{\perp} \cos \theta \\ &= -w_{\perp 2} \frac{\partial f_s}{\partial w_{\perp 1}} + w_{\perp 1} \frac{\partial f_s}{\partial w_{\perp 2}} \\ &= \mathbf{w} \times \hat{\mathbf{b}} \cdot \frac{\partial f_s}{\partial \mathbf{w}} \end{aligned} \quad (14.8)$$

This is why we say the third term in Equation 14.1 represents a rotation in the velocity space, or more exactly, in the perpendicular velocity plane.

From Equation 14.3, if we apply the lowest order of approximation,

$$\Omega_s \left(\frac{\partial f_s}{\partial} \right)_{w_{\perp}, w_{\parallel}} = \underbrace{\frac{df_s}{dt}}_{\omega/\Omega_s \ll 1} + \underbrace{\mathbf{w} \cdot \nabla f_s}_{kr_s \ll 1} + \underbrace{(\mathbf{a}_s - \mathbf{w} \cdot \nabla \mathbf{u}_s)}_{kr_s M_A \ll 1} \cdot \frac{\partial f_s}{\partial \mathbf{w}} - \underbrace{C(f_s)}_{\nu_s \ll \Omega_s} = 0 \quad (14.9)$$

which essentially tells us that $f_s = f_s(w_{\perp}, w_{\parallel}, \theta) = f_s(w_{\perp}, w_{\parallel})$ is gyrotropic. Let us use $\langle \rangle$ to denote averaging over a gyroperiod:

$$\langle A \rangle = \int_0^{2\pi} A d\theta$$

We can use gyrotropy to simplify the pressure tensor:

$$\begin{aligned} \mathbf{P}_s &= \int d\mathbf{w} m_s \langle \mathbf{w} \mathbf{w} \rangle f_s(\mathbf{r}, w_{\perp}, w_{\parallel}, t) \\ &= \int d\mathbf{w} m_s \left[\frac{w_{\perp}^2}{2} (\mathbf{I} - \hat{\mathbf{b}}\hat{\mathbf{b}}) + w_{\parallel}^2 \hat{\mathbf{b}}\hat{\mathbf{b}} \right] f_s(\mathbf{r}, w_{\perp}, w_{\parallel}, t) \\ &= (\mathbf{I} - \hat{\mathbf{b}}\hat{\mathbf{b}}) \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} f_s + \hat{\mathbf{b}}\hat{\mathbf{b}} \int d\mathbf{w} m_s w_{\parallel}^2 f_s \\ &= \begin{pmatrix} p_{\perp s} & 0 & 0 \\ 0 & p_{\perp s} & 0 \\ 0 & 0 & p_{\parallel s} \end{pmatrix} \end{aligned}$$

where

$$p_{\perp} = \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} f_s$$

$$p_{\parallel} = \int d\mathbf{w} m_s w_{\parallel}^2 f_s$$

Equation 14.5 becomes

$$\rho \frac{d\mathbf{u}}{dt} = -\nabla \left(\underbrace{p_{\perp} + \frac{B^2}{2\mu_0}}_{\text{total scalar pressure}} \right) + \nabla \cdot \left[\hat{b}\hat{b} \left(\underbrace{p_{\perp} - p_{\parallel}}_{\text{pressure anisotropy stress}} + \underbrace{\frac{B^2}{\mu_0}}_{\text{Maxwell stress}} \right) \right] \quad (14.10)$$

The pressure anisotropy stress is the key new feature compared to usual MHD. It should be important provided $p_{\perp} - p_{\parallel} \gtrsim B^2/\mu_0$, or $(p_{\perp} - p_{\parallel})/p \gtrsim 2/\beta$. Therefore this is more likely to matter in high- β plasmas.

To summarize what we have gotten so far: to work out motions and magnetic fields in a plasma, solve Equation 14.10 for \mathbf{u} and Equation 14.6 for \mathbf{B} , where

$$\rho = \sum_s m_s \int d\mathbf{w} f_s$$

$$p_{\perp} = \sum_s \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} f_s$$

$$p_{\parallel} = \sum_s \int d\mathbf{w} m_s w_{\parallel}^2 f_s$$

We still need the kinetic equation to calculate f_s — this kinetic equation will need to be somewhat reduced to solve for the lowest-order, gyrotropic $f_s(w_{\perp}, w_{\parallel})$. In pursuit of instant justification, we can postpone doing this and first derive some results that do not need the f_s equation (i.e. the Firehose instability) as in Section 14.1. For mirror modes, let us continue from the kinetic Equation 14.9 for higher orders. We have already known that the lowest order approximation gives gyrotropic distributions.

To the first order,

$$\Omega_s \left(\frac{\partial f_s^1}{\partial \theta} \right)_{w_{\perp}, w_{\parallel}} = \frac{df_s^0}{dt} + \mathbf{w} \cdot \nabla f_s^0 + (\mathbf{a}_s - \mathbf{w} \cdot \nabla \mathbf{u}_s) \cdot \frac{\partial f_s^0}{\partial \mathbf{w}} - C(f_s^0)$$

The left-hand side can be eliminated by integrating over θ , so we have

$$\left\langle \frac{df_s}{dt} + \mathbf{w} \cdot \nabla f_s + (\mathbf{a}_s - \mathbf{w} \cdot \nabla \mathbf{u}_s) \cdot \frac{\partial f_s}{\partial \mathbf{w}} - C(f_s) \right\rangle = 0$$

where $f_s = f_s(w_\perp, w_\parallel)$. To do this averaging, we transform variables from $(t, \mathbf{r}, \mathbf{w}) \rightarrow (t, \mathbf{r}, w_\perp, w_\parallel, \theta)$. With

$$w_\parallel = \mathbf{w} \cdot \hat{\mathbf{b}}(t, \mathbf{r})$$

$$w_\perp = |\mathbf{w} - w_\parallel \hat{\mathbf{b}}|$$

and some algebras (??? Check online notes.), we have

$$\frac{Df_s}{Dt} + \frac{1}{B} \frac{DB}{Dt} \frac{w_\perp}{2} \frac{\partial f_s}{\partial w_\perp} + \left(\frac{q_s}{m_s} E_\parallel - \frac{D\mathbf{u}_s \cdot \hat{\mathbf{b}}}{Dt} - \frac{w_\perp^2}{2} \frac{\nabla_\parallel B}{B} \right) \frac{\partial f_s}{\partial w_\parallel} = C(f_s) \quad (14.11)$$

where

$$D/Dt = d/dt + w_\parallel \hat{\mathbf{b}} \cdot \nabla = \partial/\partial t + \mathbf{u}_s \cdot \nabla + w_\parallel \hat{\mathbf{b}} \cdot \nabla$$

This is not terribly transparent and it is perhaps better to write this equation in different, “more physical” variables. Let

$$f_s(w_\perp, w_\parallel) = F_s(\mu, \epsilon)$$

where $\mu = m_s w_\perp^2 / 2B$ is the magnetic moment of a gyrating particle and $\epsilon = m_s w^2 / 2 = m_s(w_\perp^2 + w_\parallel^2)/2$. Since μ is conserved when $\omega \ll \Omega_s$, F_s satisfies (???)

$$\frac{DF_s}{Dt} + \left[m_s w_\parallel \left(\frac{q_s}{m_s} E_\parallel - \frac{D\mathbf{u}_s \cdot \hat{\mathbf{b}}}{Dt} \right) + \mu \frac{dB}{dt} \right] \frac{\partial F_s}{\partial \epsilon} = C(F_s) \quad (14.12)$$

- The first term is the convective derivative in the guiding center coordinates.
- The second term is the acceleration by parallel electric field, where E_\parallel is determined by imposing $\sum_s q_s n_s = 0$.
- The third term takes account of the fact that ϵ does not include the bulk velocity.
- The fourth term is the *betatron acceleration* due to μ conservation:

$$\epsilon = \mu B + \frac{m_s w_\parallel^2}{2}$$

$$\dot{\epsilon} = \mu \dot{B} (w_\parallel \text{ constant} ???)$$

Betatron acceleration refers to situations in which the magnetic field strength increases slowly in time (compared with a gyroperiod), so that μ remains constant, but the particle kinetic energy is not constant due to the presence of electric fields (associated with the time-varying magnetic field). Then, the perpendicular energy is increased due to constancy of μ . As we will see soon in Section 14.2, this is the key for explaining mirror modes.

14.1 Firehose Instability: Linear Theory

Suppose we have some “macroscopic” solution of our (yet to be fully derived) equilibrium. We allow low-frequency, short-wavelength perturbations ($\omega \ll u/l, kl \gg 1$) of this solution, and seek solutions in the form $\mathbf{X} + \delta\mathbf{X}$ with infinitesimal perturbations $\propto e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$. Note that the velocity \mathbf{u} is treated as a perturbation term (background velocity is simply a drift).

From Equation 14.6

$$\begin{aligned} -\omega\delta\mathbf{B} &= \mathbf{B} \cdot \mathbf{k} \delta\mathbf{u} - \mathbf{B} \mathbf{k} \cdot \delta\mathbf{u} \\ &= B(k_{\parallel} \delta\mathbf{u}_{\perp} - \hat{b} \mathbf{k}_{\perp} \cdot \delta\mathbf{u}_{\perp}) \end{aligned} \quad (14.13)$$

Inserting Equation 3.11 into Equation 14.10, we have

$$\begin{aligned} -\omega\rho\delta\mathbf{u} &= -\mathbf{k} \left(\delta p_{\perp} + \frac{B\delta B}{\mu_0} \right) + \mathbf{k} \cdot \left[(\hat{b}\hat{b} + \hat{b}\delta\hat{b})(p_{\perp} - p_{\parallel} + \frac{B^2}{\mu_0}) + \hat{b}\hat{b} \left(\delta p_{\perp} - \delta p_{\parallel} + \frac{2B\delta B}{\mu_0} \right) \right] \\ &= -\mathbf{k}_{\perp} \left(\delta p_{\perp} + \frac{B\delta B}{\mu_0} \right) - \hat{b}k_{\parallel} \left[\delta p_{\parallel} + (p_{\perp} - p_{\parallel}) \frac{\delta B}{B} \right] + \delta\hat{b}k_{\parallel} \left(p_{\perp} - p_{\parallel} + \frac{B^2}{\mu_0} \right) \end{aligned} \quad (14.14)$$

\hat{b} has two parts: the Alfvénic part $\delta\mathbf{B}_{\perp}/B$ and the compressional part $\delta\mathbf{B}_{\parallel}/B$. From Equation 14.13, the Alfvénic perturbation of \hat{b} can be written as

$$\hat{b} = \frac{\delta\mathbf{B}_{\perp}}{B} = -\frac{k_{\parallel}}{\omega} \delta\mathbf{u}_{\perp}$$

Isolate the Alfvénic response in Equation 14.14 by cross-producting with \mathbf{k}_{\perp} :

$$\begin{aligned} -\omega\rho\mathbf{k}_{\perp} \times \delta\mathbf{u}_{\perp} &= k_{\parallel} \left(p_{\perp} - p_{\parallel} + \frac{B^2}{\mu_0} \right) \mathbf{k}_{\perp} \times \hat{b} \\ \omega\rho\mathbf{k}_{\perp} \times \delta\mathbf{u}_{\perp} &= k_{\parallel} \left(p_{\perp} - p_{\parallel} + \frac{B^2}{\mu_0} \right) \mathbf{k}_{\perp} \times \frac{k_{\parallel}}{\omega} \delta\mathbf{u}_{\perp} \\ \omega^2 &= k_{\parallel}^2 \left(\frac{B^2}{\mu_0\rho} + \frac{p_{\perp} - p_{\parallel}}{\rho} \right) = k_{\parallel}^2 v_{th\parallel}^2 \left(\frac{p_{\perp} - p_{\parallel}}{p_{\parallel}} + \frac{2}{\beta_{\parallel}} \right) \end{aligned}$$

Let $A = (p_{\perp} - p_{\parallel})/p_{\parallel}$. The system will be unstable if $A < -2/\beta_{\parallel}$, i.e.

$$p_{\perp} - p_{\parallel} > 2p_B \quad (14.15)$$

which leads to a growth rate

$$\gamma = k_{\parallel} v_{th\parallel} \sqrt{\left| A + \frac{2}{\beta_{\parallel}} \right|} \quad (14.16)$$

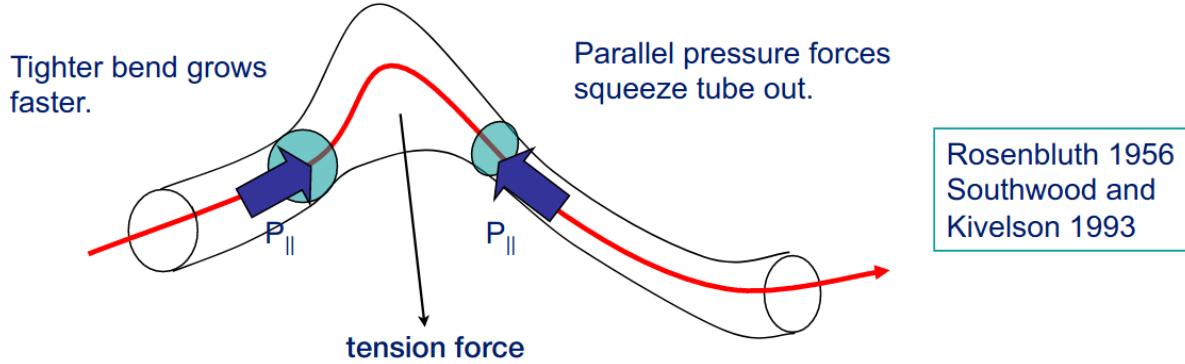


Figure 14.1: Along a flux tube, p_{\parallel} is the destabilizing force, the curvature force is the stabilizing force.

Thus, negative A ($p_{\parallel} > p_{\perp}$) locally weakens tension, i.e. slows down Alfvén waves, and makes it energetically easier to bend the field lines. When $A < -2/\beta_{\parallel}$, the elasticity of field lines is lost and we have the firehose instability.

Key points:

- Nothing surprising that $p_{\parallel} > p_{\perp}$ leads to an instability: it is a non-equilibrium situation, so a source of free energy.
- $\gamma \propto k_{\parallel}$ leads to UV catastrophe: within KMHD ($\omega \ll \Omega_i, kr_i \ll 1$), the wavenumber of peak γ is not captured. Including finite larmor radius gives (Oxford MNRAS 405, 291? ARE THE EXPRESSIONS CORRECT?)

$$\gamma_{\text{peak}} \sim \left| A + \frac{2}{\beta_{\parallel}} \right| \Omega_i$$

$$k_{\parallel \text{peak}} r_i \sim \left| A + \frac{2}{\beta_{\parallel}} \right|^{1/2}$$

so the instability is very fast ($\gamma \propto \Omega_i$ very large with strong B field) at microscale. Any high- β macroscopic solution with $p_{\parallel} > p_{\perp}$ will blow up instantly. What happens next is decided by the nonlinear saturation of the firehose. It was a transformative moment when Justin Kasper in 2002 discovered that the firehose stability boundary constrains most observed solar wind states, followed by Hellinger in 2006. Bale in 2009 showed that there is an increased fluctuation level at the boundary.

ADD FIGURE ABOUT THE FIREHOSE STABILITY REGIME FIGURE!

14.2 Mirror Instability: Linear Theory

Let us go back to Equation 14.13 and Equation 14.14 and get apart from Alfvénic what other perturbations there are and when they are stable. We have already looked at the Alfvénic perturbation $\hat{\delta b} = \delta \mathbf{B}_\perp / B$. Now consider

$$\frac{\delta B}{B} = \frac{\delta B_\parallel}{B} \quad (14.17)$$

From Equation 14.13, we have the perpendicular compression increases B:

$$\omega \frac{\delta B}{B} = \mathbf{k}_\perp \cdot \delta \mathbf{u}_\perp$$

Take $\mathbf{k}_\perp \cdot$ Equation 14.14:

$$\omega \rho \mathbf{k}_\perp \cdot \delta \mathbf{u}_\perp = \rho \omega^2 \frac{\delta B}{B} = k_\perp^2 \left(\delta p_\perp + \frac{B \delta B}{\mu_0} \right) + k_\parallel^2 \left(p_\perp - p_\parallel + \frac{B^2}{\mu_0} \right) \frac{\delta B}{B} \quad (14.18)$$

Note the p_\perp term here: we need kinetic theory to calculate this! Fortunately we have Equation 14.12 ready for calculating

$$\delta p_\perp = \int d\mathbf{w} \frac{m_s w_\perp^2}{2} \delta f_s(w_\perp, w_\parallel)$$

$\delta f_s(w_\perp, w_\parallel)$ can be obtained by calculating $F_s(\mu, \epsilon)$ and transforming back to w_\perp, w_\parallel .

Here is a cute subtlety: our macroscopic equilibrium, around which we are expanding the distribution is

$$F_{0s}(\mu, \epsilon) = f_{0s}(w_\perp, w_\parallel) = f_{0s}\left(\sqrt{\frac{2B_0\mu}{m_s}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}}\right)$$

which contains B_0 the unperturbed magnetic field. μ in F_0 contains $B_0 + \delta B$, and this has to be taken into account when transforming to w_\perp, w_\parallel . Now when we perturb everything:

$$\begin{aligned} F_s(\mu, \epsilon) &= F_{0s}(\mu, \epsilon) + \delta F_s \\ &= f_{0s}(w_\perp, w_\parallel) + \delta f_s \\ &= f_{0s}\left(\sqrt{\frac{2\mu(B_0 + \delta B)}{m_s}}, \sqrt{\frac{2[\epsilon - \mu(B_0 + \delta B)]}{m_s}}\right) + \delta f_s \\ &= f_{0s}\left(\sqrt{\frac{2\mu B_0}{m_s}} \sqrt{1 + \frac{\delta B}{B_0}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}} \sqrt{1 - \frac{m_s \mu \delta B}{(\epsilon - \mu B_0)}}\right) + \delta f_s \\ &\approx f_{0s}\left(\sqrt{\frac{2\mu B_0}{m_s}}, \sqrt{\frac{2(\epsilon - \mu B_0)}{m_s}}\right) + \frac{2\mu}{m_s} \delta B \left(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\parallel^2} \right) + \delta f_s \end{aligned}$$

Thus

$$\delta f_s = \delta F_s - w_\perp^2 \frac{\delta B}{B} \left(\frac{\partial f_{0s}}{\partial w_\perp^2} - \frac{\partial f_{0s}}{\partial w_\parallel^2} \right)$$

If f_{0s} is a bi-Maxwellian,

$$f_{0s} = \frac{n_s}{\pi^{3/2} v_{\text{th}\perp s}^2 v_{\text{th}\parallel s}^2} \exp \left(- \frac{w_\perp^2}{v_{\text{th}\perp s}^2} - \frac{w_\parallel^2}{v_{\text{th}\parallel s}^2} \right)$$

then this can be further written as

$$\delta f_s = \delta F_s + w_\perp^2 \frac{\delta B}{B} \left(\frac{1}{v_{\text{th}\perp s}^2} - \frac{1}{v_{\text{th}\parallel s}^2} \right) f_{0s} = \delta F_s + w_\perp^2 \frac{\delta B}{B} \frac{m_s n_s}{2} \left(\frac{1}{p_{\perp s}} - \frac{1}{p_{\parallel s}} \right) f_{0s}$$

We can eliminate the partial derivatives via integration by parts:

$$\begin{aligned} \int d\mathbf{w} \frac{\partial f_{0s}}{\partial w_\parallel^2} &= \int_0^{2\pi} d\theta \int dw_\perp \int \frac{1}{2w_\parallel} \frac{\partial f_{0s}}{\partial w_\parallel} dw_\parallel \\ &= \int_0^{2\pi} d\theta \int dw_\perp \int \frac{1}{2w_\parallel} df_{0s} \\ &= \int_0^{2\pi} d\theta \int dw_\perp \left[\frac{f_{0s}}{2w_\parallel} \Big|_{-\infty}^{\infty} - \int f_{0s} d\frac{1}{2w_\parallel} \right] \\ &= \int_0^{2\pi} d\theta \int dw_\perp \int \frac{1}{2w_\parallel^2} f_{0s} dw_\parallel \\ &= \int d\mathbf{w} \frac{1}{2w_\parallel^2} f_{0s} \\ \int d\mathbf{w} w_\perp^4 \frac{\partial f_{0s}}{\partial w_\perp^2} &= 2\pi \int dw_\parallel \left[\frac{1}{2} w_\perp^4 f_{0s} \Big|_{-\infty}^{+\infty} - \frac{1}{2} \int f_{0s} dw_\perp^4 \right] \\ &= -2\pi \int dw_\parallel 2w_\perp^2 f_{0s} w_\perp dw_\perp \\ &= -2 \int d\mathbf{w} w_\perp^2 f_{0s} \end{aligned}$$

This then gives us

$$\begin{aligned}
\delta p_{\perp s} &= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta f_s \\
&= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta F_s - \int d\mathbf{w} \frac{m_s w_{\perp}^4}{2} \left(\frac{\partial f_{0s}}{\partial w_{\perp}^2} - \frac{\partial f_{0s}}{\partial w_{\parallel}^2} \right) \frac{\delta B}{B} \\
&= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta F_s - \int_0^{2\pi} d\theta \int dw_{\perp} w_{\perp} \int dw_{\parallel} \frac{m_s w_{\perp}^4}{2} \left(\frac{\partial f_{0s}}{\partial w_{\perp}^2} - \frac{\partial f_{0s}}{\partial w_{\parallel}^2} \right) \frac{\delta B}{B} \\
&= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta F_s + 2 \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta f_{0s} \frac{\delta B}{B} + \int d\mathbf{w} \frac{2(\frac{1}{2} m_s w_{\perp}^2)^2}{m_s w_{\parallel}^2} f_{0s} \frac{\delta B}{B} \\
&= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} \delta F_s + \frac{\delta B}{B} \left(2p_{\perp s} - \frac{2p_{\perp s}^2}{p_{\parallel s}} \alpha_s \right)
\end{aligned} \tag{14.19}$$

where α_s is some coefficients of order 1 if f_{0s} is not bi-Maxwellian.

δF_s can be obtained by ignoring collisions and linearizing and Fourier-transforming Equation 14.12 ($\mathbf{u}_s = 0$):

$$\begin{aligned}
-i(\omega - k_{\parallel} w_{\parallel}) \delta F_s &= - \left[m_s w_{\parallel} \left(\frac{q_s}{m_s} E_{\parallel} - i(\omega - k_{\parallel} w_{\parallel}) \delta u_{\parallel s} \right) - i\omega \mu \delta B \right] \frac{\partial F_{0s}}{\partial \epsilon} \\
\delta F_s &= -i \frac{w_{\parallel} q_s E_{\parallel}}{\omega - k_{\parallel} w_{\parallel}} \frac{\partial F_{0s}}{\partial \epsilon} - \delta u_{\parallel s} m_s w_{\parallel} \frac{\partial F_{0s}}{\partial \epsilon} - \frac{\omega}{\omega - k_{\parallel} w_{\parallel}} \mu \delta B \frac{\partial F_{0s}}{\partial \epsilon}
\end{aligned}$$

The first term can be ignored if $\beta \gg 1$ (??? See the complete calculation in another note!); otherwise E_{\parallel} can be got by imposing $\sum_s q_s n_s = 0$. The second term can be shown to be equivalent to $\delta u_{\parallel s} \partial f_{0s} / \partial w_{\parallel}$:

$$\begin{aligned}
\frac{\partial f_{0s}}{\partial w_{\parallel}} &= \frac{\partial f_{0s}}{\partial \epsilon} \frac{\partial \epsilon}{\partial w_{\parallel}} + \frac{\partial f_{0s}}{\partial \mu} \frac{\partial \mu}{\partial w_{\parallel}} \\
&= \frac{\partial F_{0s}}{\partial \epsilon} m_s w_{\parallel}
\end{aligned}$$

so this will not contribute to δp_{\perp} because it integrates to 0.

The third term can be written as

$$\frac{\omega}{\omega - k_{\parallel} w_{\parallel}} \mu \delta B \frac{\partial F_{0s}}{\partial \epsilon} = \frac{\omega}{\omega - k_{\parallel} w_{\parallel}} \frac{m_s w_{\perp}^2}{2} \frac{\delta B}{B} \frac{1}{w_{\parallel}} \frac{\partial f_{0s}}{\partial w_{\parallel}} = \frac{\omega}{\omega - k_{\parallel} w_{\parallel}} m_s w_{\perp}^2 \frac{\delta B}{B} \frac{\partial f_{0s}}{\partial w_{\parallel}^2}$$

Thus, the “relevant” part of δF_s is

$$\delta F_s = -\frac{\omega}{\omega - k_{\parallel} w_{\parallel}} m_s w_{\perp}^2 \frac{\delta B}{B} \frac{\partial f_{0s}}{\partial w_{\parallel}^2}$$

and its contribution to $\delta p_{\perp s}$ is

$$\int d\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s = \frac{\delta B}{B} \frac{\omega}{|k_\parallel|} \int \frac{dw_\parallel}{w_\parallel - \frac{\omega}{|k_\parallel|}} \left[\frac{\partial}{\partial w_\parallel^2} \int d\mathbf{w}_\perp \frac{m_s^2 w_\perp^4}{2} f_{0s} \right]$$

Here we have $|k_\parallel|$ because if $k_\parallel < 0$, we can change the variable $w_\parallel \rightarrow -w_\parallel$. This involves the Landau integral, which can be evaluated with the residual theorem Equation 3.3 when integrate in the complex plane mostly along the real axis and the large semicircle in the upper half plane except for a small semicircle just below the pole (ADD FIGURE!):

$$\frac{1}{w_\parallel - \frac{\omega}{|k_\parallel|}} = P \frac{1}{w_\parallel - \frac{\omega}{|k_\parallel|}} + i\pi \delta\left(w_\parallel - \frac{\omega}{|k_\parallel|}\right)$$

so

$$\int d\mathbf{w} \frac{m_s w_\perp^2}{2} \delta F_s = \frac{\delta B}{B} \left[\frac{\omega}{|k_\parallel|} P \int \frac{dw_\parallel}{w_\parallel - \frac{\omega}{|k_\parallel|}} [\dots] + i\pi \frac{\omega}{|k_\parallel|} [\dots]_{w_\parallel} = \omega/|k_\parallel| \right]$$

The first term is small when we assume $\omega \ll k_\parallel v_{\text{ths}}$; the second term must be kept because it is the lowest-order imaginary part which will lead to instability.

For a bi-Maxwellian,

$$\left[\frac{\partial}{\partial w_\parallel^2} \int d\mathbf{w}_\perp \frac{m_s^2 w_\perp^4}{2} f_{0s} \right]_{w_\parallel = \omega/|k_\parallel|} = -\frac{2p_{\perp s}^2}{p_{\parallel s}} \frac{e^{-\frac{\omega^2}{k_\parallel^2 v_{\text{th}}^2 s}}}{\sqrt{\pi} v_{\text{th}} s}$$

The exponential term is nearly 1. If it is not a bi-Maxwellian, then we need to multiply by a coefficient $\alpha_s \sim 1$.

Equation 14.19 becomes

$$\delta p_{\perp s} = \frac{\delta B}{B} \left[2p_{\perp s} - \frac{2p_{\perp s}^2}{p_{\parallel s}} \left(\alpha_s + i\sqrt{\pi} \frac{\omega}{|k_\parallel| v_{\text{th}} s} \sigma_s \right) \right]$$

This goes into Equation 14.18:

$$\rho\omega^2 = k_\perp^2 \frac{B^2}{\mu_0} \left[\sum_s \left(1 - \frac{p_{\perp s}}{p_{\parallel s}} \alpha_s \right) \beta_{\perp s} - i \sum_s \sigma_s \frac{p_{\perp s}}{p_{\parallel s}} \beta_{\perp s} \sqrt{\pi} \frac{\omega}{|k_\parallel| v_{\text{th}} s} + 1 \right] + k_\parallel^2 \frac{B^2}{\mu_0} \left[\sum_s \frac{\beta_{\perp s}}{2} \left(1 - \frac{p_{\parallel s}}{p_{\perp s}} \right) + 1 \right]$$

The left-hand side can be neglected because $\omega \ll k_\parallel v_{\text{th}} s$. The electron thermal velocity $v_{\text{th}} e$ in the denominator can be neglected because $v_{\text{th}} e \gg v_{\text{th}} i$. The growth rate γ is the imaginary part of ω . Reorganize the last equation:

$$\sigma_i \frac{p_{\perp i}}{p_{\parallel i}} \beta_{\perp i} \sqrt{\pi} \frac{\gamma}{|k_\parallel| v_{\text{th}} i} = \sum_s \left(\frac{p_{\perp s}}{p_{\parallel s}} \alpha_s - 1 \right) \beta_{\perp s} - 1 - \frac{k_\parallel^2}{k_\perp^2} \left[\sum_s \frac{\beta_{\perp s}}{2} \left(1 - \frac{p_{\parallel s}}{p_{\perp s}} \right) + 1 \right] \quad (14.20)$$

where $\Lambda \equiv \frac{k_{\parallel}^2}{k_{\perp}^2} \sum_s \left(\frac{p_{\perp s}}{p_{\parallel s}} \alpha_s - 1 \right) \beta_{\perp s} - 1$ triggers instability if this is positive:

$$\sum_s \left(\frac{p_{\perp s}}{p_{\parallel s}} \alpha_s - 1 \right) \beta_{\perp s} > 1$$

Examining where this comes from, we see that this amounts to δp_{\perp} modifying the magnetic pressure force and turning it from positive to negative:

$$\delta p_{\perp} + \frac{B \delta B}{\mu_0} = \frac{B \delta B}{\mu_0} \left[\underbrace{\frac{1}{\text{B pressure}}}_{\text{non-resonant particle pressure}} - \underbrace{\sum_s \left(\frac{p_{\perp s}}{p_{\parallel s}} \alpha_s - 1 \right) \beta_{\perp s}}_{\text{resonant particle pressure}} + \dots \right]$$

Thus, fundamentally, pressure anisotropy makes it easier to compress or rarefy magnetic field — and things become unstable when the sign of the pressure flips and it becomes energetically profitable to create compressions and rarefactions. (ADD FIGURE!) The dispersion relation Equation 14.20 is basically a statement of pressure balance between the magnetic pressure, the non-resonant particle pressure δp_{\perp} and the resonant particle pressure $\propto \gamma$, which came from the *betatron acceleration* $\mu dB/dt$ in Equation 14.12. (See also Eq. 21 in D. J. Southwood and Kivelson (1993))

The betatron acceleration term refers to what happens in the stable case. When magnetic pressure opposes formation of δB perturbations (say, troughs), to compensate it, we must have $\gamma < 0$ and energy goes from δB to resonant particles, which are accelerated by the mirror force. The corresponding decaying of δB is the well-known *Barnes damping*.¹

More discussion on the physics is presented by D. J. Southwood and Kivelson (1993). They pointed out that although the instability is resonant, the role of resonant particle is unusual. The instability results from pressure imbalance between the bulk of the plasma and the magnetic field. For this to occur, the bulk (nonresonant) pressure response must be in antiphase with the magnetic pressure as occurs at low frequencies when the magnetic moment and the particle energy are conserved. The resonant particles produce a pressure perturbation, however, in phase with the field pressure change. A corollary is that unlike the nonresonant particles, the resonant particles experience energy changes as the instability develops. The linear growth rate of most resonant instabilities is proportional to the number of resonant particles. However, in the case of the mirror instability, the growth is inversely proportional to the number (and pressure contribution) of resonant particles. The reason for the anomalous result is that the fewer particles there are at small parallel velocity the higher the growth rate needs to be to balance the pressure imbalance generated by the nonresonant distribution.

To finish the job, note that, from Equation 14.20 (ADD FIGURE!) for a given k_{\perp}

$$\left. \frac{\partial \gamma}{\partial k_{\parallel}} \right|_{k_{\perp}} \propto \Lambda - \frac{k_{\parallel}^2}{k_{\perp}^2} \left[\sum_s \frac{\beta_{\perp s}}{2} \left(1 - \frac{p_{\parallel s}}{p_{\perp s}} \right) + 1 \right]$$

¹Landau damping of “mirror field”, Barnes 1966, also known as transit-time damping from Stix’s book.

The maximum growth rate is reached when the right-hand side goes to 0, which is equivalent to $\frac{2}{3}\Lambda$, so the maximum growth rate

$$\gamma_{\max} = \frac{|k_{\parallel}|v_{\text{th}\parallel i}}{\sqrt{\pi}} \frac{2}{3}\Lambda \frac{p_{\parallel i}}{p_{\perp i}} \frac{1}{\sigma_i \beta_{\perp i}}$$

We have assumed $\gamma \ll k_{\parallel} v_{\text{th}\parallel s}$, which is indeed true if

$$\Lambda \frac{1}{\beta_{\perp i}} = \left(\sum_s A_s \beta_{\perp s} - 1 \right) \frac{1}{\beta_{\perp i}} \ll 1$$

so our approximations are consistent.

If we are close to marginal instability,

$$\frac{k_{\parallel}}{k_{\perp}} \sim \sqrt{\Lambda} \ll 1$$

so mirror modes are highly oblique near the threshold.

Another important point is that again we encounter the UV catastrophe since $\gamma \propto k_{\parallel}$. The mirror mode is a fast, microscale instability whose peak growth rate is outside KMHD regime. Including finite larmor radius gives (Hellinger 2007 PoP 14, 082105?)

$$\gamma_{\text{peak}} \sim \left(A - \frac{1}{\beta} \right)^2 \beta \Omega_i, \quad k_{\text{peak}} r_i \sim \left(A - \frac{1}{\beta} \right) \beta$$

Thus, any high- β macroscopic solution of KMHD with $p_{\perp} > p_{\parallel}$ will blow up, just like the case for $p_{\parallel} > p_{\perp}$, and again what happens next depends on how mirror instability saturates. Note that A_e is ignored since $A_e \ll A_i$ (?). The mirror instability condition is

$$\begin{aligned} \frac{p_{\perp i}}{p_{\parallel i}} - 1 &> \frac{1}{\beta_{\perp i}} = \frac{1}{\beta_{\parallel i}} \frac{p_{\parallel i}}{p_{\perp i}} \\ \frac{p_{\perp i}}{p_{\parallel i}} \left(\frac{p_{\perp i}}{p_{\parallel i}} - 1 \right) &> \frac{1}{\beta_{\parallel i}} \end{aligned}$$

Figure 14.2 shows observation from Wind spacecraft. The solar wind indeed seems to stay within these boundaries. (ADD REFS!)

14.2.1 Comparison With Slow Mode

- Driving Mechanism
 - Mirror Mode: Driven primarily by pressure anisotropy ($T_{\perp}/T_{\parallel} > 1$) in high-beta plasmas.

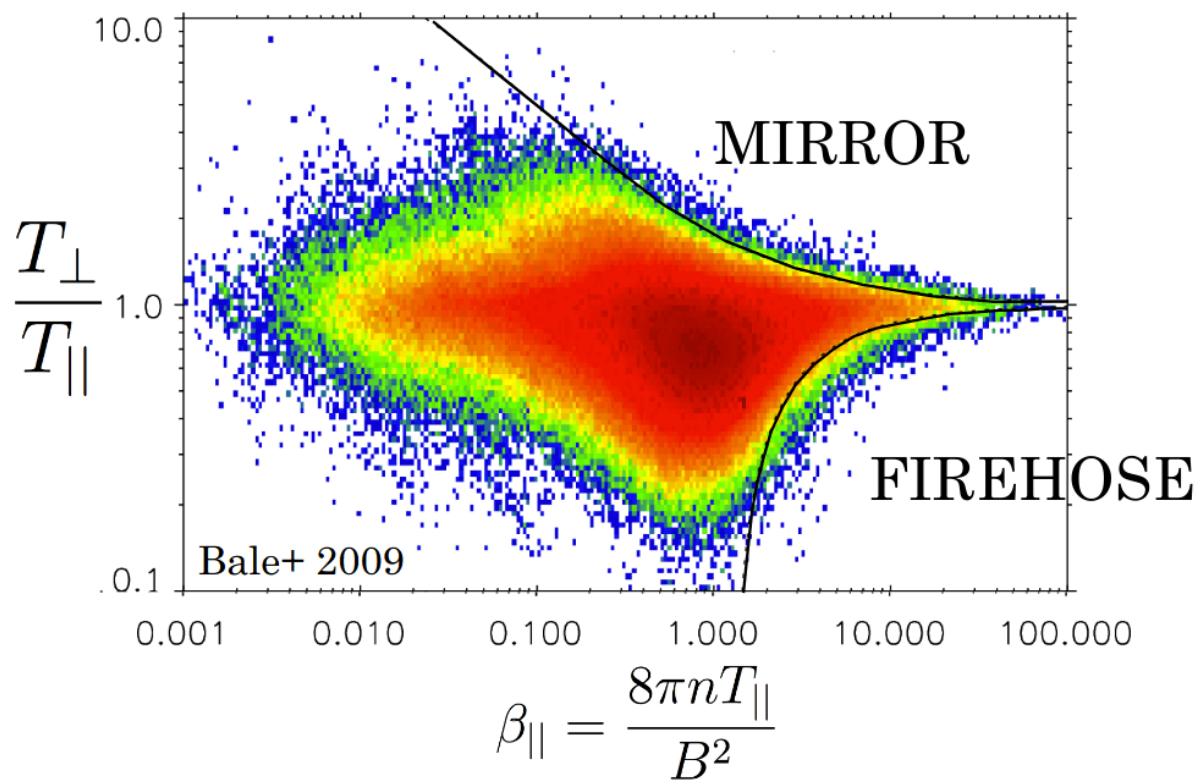


Figure 14.2: Collective solar wind observation data ($\sim 1e6$) from Wind spacecraft. The lines represent the instability thresholds for mirror and firehose instability, respectively.

- MHD Slow Mode: Driven by pressure gradients and magnetic field line curvature in low- to moderate-beta plasmas.
- Propagation Characteristics
 - Mirror Mode: Primarily propagates parallel to the background magnetic field, but can have a small perpendicular component. It is a non-propagating mode in the fluid limit (zero frequency), but it can acquire a finite frequency due to kinetic effects.
 - MHD Slow Mode: Propagates obliquely to the magnetic field, with both parallel and perpendicular components. It is a propagating mode with a finite frequency.
- Plasma Conditions
 - Mirror Mode: Typically found in high-beta ($\gtrsim 1$) plasmas, such as the Earth's magnetosheath and the solar wind.
 - MHD Slow Mode: More common in low- to moderate-beta ($\ll 1$) plasmas, such as the solar corona and the Earth's magnetosphere, where the magnetic pressure dominates.

14.3 Origin of Pressure Anisotropy

So far we have seen that the bottom line is that any macroscopic, high- β KMHD solution that has $p_{\perp} \neq p_{\parallel}$ (more precisely, $|p_{\perp} - p_{\parallel}|/p \gtrsim 1/\beta$) will be violently unstable to either firehose or mirror — both of which are fast and micro-scale modes giving rise to fluctuations outside the KMHD regime (and, by the way, also outside gyrokinetics — too close to cyclotron frequency, k_{\parallel}/k_{\perp} not small enough, $\delta\mathbf{B}/B$ also not small enough). How worried should this make us about the applicability of KMHD to high- β plasmas that are not collisional enough to be fully fluid (i.e. $\nu \ll \Omega_s$)?

The answer is, *very worried!* A key property of low-frequency, weakly collisional dynamics is that the magnetic moment $\mu = m_s w_{\perp}^2 / 2B$ is conserved by particles. The mean μ of particles of species s is

$$\langle \mu \rangle_w = \frac{1}{n_s} \int d\mathbf{w} \mu f_s = \frac{p_{\perp s}}{n_s B} = \text{const.}$$

For the purpose of a qualitative discussion, let us pretend for a moment that $n_s = \text{const}$ (incompressible plasmas, $\beta \gg 1$). Then the above conservation relation says that, locally in a fluid element (\mathbf{w} is peculiar velocity), every time you change \mathbf{B} , you must change $p_{\perp s}$ proportionally (but not $p_{\parallel s}$). Thus we expect (?)

$$\frac{1}{p_{\perp s}} \frac{dp_{\perp s}}{dt} \underset{\mu \text{ conservation}}{\sim} \underbrace{\frac{1}{B} \frac{dB}{dt}}_{\text{relaxation of pressure anisotropy by collisions}} - \underbrace{\nu_s \frac{p_{\perp s} - p_{\parallel s}}{p_{\perp s}}}_{\text{relaxation of pressure anisotropy by collisions}} \quad (14.21)$$

It is useful to remind ourselves that d/dt is in the \mathbf{u}_s frame. Balancing the two effects on the right-hand side,

$$\Delta_s = \frac{p_{\perp s} - p_{\parallel s}}{p_{\perp s}} \sim \frac{1}{\nu_s} \frac{1}{B} \frac{dB}{dt} \quad (14.22)$$

This expression is valid only if $\Delta_s \ll 1$, i.e. $\nu_s \gg \frac{1}{B} \frac{dB}{dt}$, otherwise Δ_s will grow with time as B is changed. Thus

- B increases locally $\rightarrow \Delta_s > 0 \rightarrow$ mirror
- B decreases locally $\rightarrow \Delta_s < 0 \rightarrow$ firehose

As nearly any large-scale dynamics involves local changes in \mathbf{B} , this means that nearly any macroscopic solution of KMHD in the high- β regime will be unstable. A very good example is the dynamo problem: when magnetic field is randomly stretched by turbulence, leading (in MHD) to exponential growth of magnetic energy (and, eventually, to saturated fields we observe), locally one finds structures of this sort: (ADD FIGURE!)

Generally speaking, in order to understand long-time evolution, we need some sort of mean-field theory for the large-scale effect of the microscale instabilities on the dynamics. Presumably, this is to keep preserve anisotropy, marginal out (?) the instabilities (as indeed appears to be confirmed by the solar wind measurements — see Bale + 2009).

There are two ways in which this can happen — firehose and mirror fluctuations might scatter particles, leading to higher effective collisionality and ? control the pressure anisotropy

$$-\frac{2}{\beta} \lesssim \frac{p_{\perp} - p_{\parallel}}{p} \lesssim \frac{1}{\beta}$$

They might inhibit changes of B , which is another way of keeping Δ under control. Which of these matters for dynamics, see the speculative overview of the possible consequences of either mechanism in MNRAS 440, 3226 (2014).

- nonlinear firehose: Rosin+ MNRAS 413 2011
- nonlinear mirror: Rincon+ MNRAS 447, 2015
- PIC simulations: Kunz+ PRL, 2014

14.4 Remarks

14.4.1 Remark I

If we assume incompressibility, the magnetic induction Equation 14.7 becomes

$$\begin{aligned} \frac{d\mathbf{B}}{dt} &= \mathbf{B} \cdot \nabla \mathbf{u} \\ \frac{1}{B} \frac{dB}{dt} &= \hat{\mathbf{b}} \hat{\mathbf{b}} : \nabla \mathbf{u} \end{aligned}$$

Then, from Equation 14.22,

$$p_{\perp s} - p_{\parallel s} \sim \frac{p_s}{\nu_c} \hat{b} \hat{b} : \nabla \mathbf{u}$$

where p_s/ν_c is the parallel dynamical viscosity. Putting this back into Equation 14.10, we get the lowest order ??? MHD equation. So, from the large-scale point of view, pressure anisotropy is viscous stress — but the resulting equations are ill-posed (blow up via instabilities with $\gamma \propto k_{\parallel}$).

14.4.2 Remark II

More rigorously, Equation 14.21 can be obtained via “CGL equations”, i.e. the evolution equations of $p_{\perp s}$ and $p_{\parallel s}$. Namely $\int d\mathbf{w} \frac{m_s w_{\perp}^2}{2}$ Equation 14.11:

$$p_{\perp s} \frac{d}{dt} \ln \frac{p_{\perp s}}{n_s B} = -\nabla \cdot (q_{\perp s} \hat{b}) - q_{\perp s} \nabla \cdot \hat{b} - \nu_s (p_{\perp s} - p_{\parallel s}) \quad (14.23)$$

$\int d\mathbf{w} m_s w_{\parallel}^2$ Equation 14.11:

$$p_{\parallel s} \frac{d}{dt} \ln \frac{p_{\parallel s} B^2}{n_s^3} = -\nabla \cdot (q_{\parallel s} \hat{b}) + 2q_{\perp s} \nabla \cdot \hat{b} - 2\nu_s (p_{\parallel s} - p_{\perp s}) \quad (14.24)$$

The left-hand side is the conservation of $J = \oint dl w_{\parallel}$, i.e. “bounce invariant”. The new feature here is heat fluxes:

$$\begin{aligned} q_{\perp s} &= \int d\mathbf{w} \frac{m_s w_{\perp}^2}{2} w_{\parallel} f_s \\ q_{\parallel s} &= \int d\mathbf{w} m_s w_{\parallel}^3 f_s \end{aligned}$$

They are here because particles can flow in and out of a fluid element and thus affect the conservation (or otherwise) of $\langle \mu \rangle_w$ and $\langle J^2 \rangle_w$ within it.

Finally, from Equation 14.23 and Equation 14.24,

$$\begin{aligned} \frac{d}{dt} (p_{\perp s} - p_{\parallel s}) &= (p_{\perp s} + 2p_{\parallel s}) \frac{1}{B} \frac{dB}{dt} + (p_{\perp s} - 3p_{\parallel s}) \frac{1}{n_s} \frac{dn_s}{dt} \\ &\quad - \nabla \cdot [(q_{\perp s} - q_{\parallel s}) \hat{b}] - 3q_{\perp s} \nabla \cdot \hat{b} - 3\nu_s (p_{\perp s} - p_{\parallel s}) \end{aligned}$$

???

$$\begin{aligned} \Delta_s &= \frac{p_{\perp s} - p_{\parallel s}}{p_s} \\ &\approx \frac{1}{\nu_s} \left\{ \frac{1}{B} \frac{dB}{dt} - \frac{2}{3} \frac{1}{n_s} \frac{dn_s}{dt} - \frac{\nabla \cdot [(q_{\perp s} - q_{\parallel s}) \hat{b}] + 3q_{\perp s} \nabla \cdot \hat{b}}{3p_s} \right\} \end{aligned}$$

Typically, what prominent here are electron heat fluxes. So heat fluxes can also lead to anisotropies and so macroscopic solutions of KMHD involving temperature gradients will also go unstable at microscales!

So here we are, we cannot change B at large scales, we cannot compress/rarefy the plasma and we cannot have temperature gradients without having to deal everything exploding and needing new equations. Enjoy!

15 Gyrokinetics

This is taken from the slide given by Frank Jenko from Max-Planck-Institut für Plasmaphysik.

Gyrokinetics is a theory for describing plasmas at a much finer level. If kinetic effects (e.g. finite Larmor radius, Landau damping, magnetic trapping etc.) play a role, MHD is not applicable, and one has to use a kinetic description.

Once again we start from the Vlasov-Maxwell equation. Removing the fast gyromotion under the assumption $\omega \ll \Omega$ leads to a dramatic speed-up. Thus we can think the basic element of charged rings as quasiparticles: it is described by gyrocenter coordinates and can keep the kinetic effects.

15.1 A Brief Historical Review

The word “gyrokinetics” appeared in the literature in the late 1960s, first proposed by Rutherford, Frieman, Taylor and Hastic. The goal is to provide an adequate formalism for the linear kinetic drift waves study in general magnetic configurations, including finite Larmor radius effects. The first nonlinear set of equations for the perturbed distribution function δf was given by Friemann and Liu Chen in 1982, who also introduced the gyrokinetic ordering. Gyrokinetic theory is based on Hamiltonian methods, which means that from a Lagrangian description, we remove the gyro-angle dependency by the change of coordinate systems to describe the equation of motion.

15.2 Coordinate Transformation

We perform the following coordinate transformation from the charged particle’s phase space (\mathbf{x}, \mathbf{v}) to the corresponding guiding-center phase space (\mathbf{X}, \mathbf{V}) , where:

$$\mathbf{X} = \mathbf{X}_\perp + X_\parallel \mathbf{b}_0, \quad \mathbf{X}_\perp = \mathbf{x}_\perp + \mathbf{r}_L, \quad \mathbf{r}_L = \mathbf{v} \times \mathbf{b}_0 / \Omega_c$$

$$\mathbf{V} = [\epsilon = v^2/2, \mu = v_\perp^2/2B_0, \sigma = \text{sgn}(v_\parallel)]$$

Here, $\mathbf{b}_0 = \mathbf{B}_0/B_0$, \mathbf{r}_L is the gyroradius vector, $v_{\parallel} = \mathbf{v} \cdot \mathbf{b}_0$, μ is the magnetic moment adiabatic invariant and, assuming there is no equilibrium electrostatic potential, ϵ is an equilibrium constant of motion.

In the guiding-center phase space, charged particle dynamics is naturally separated into the fast cyclotron motion and the slow guiding-center motion. One can then apply the gyrokinetic orderings and systematically average out the fast cyclotron motion (i.e., the gyrophase averaging) and obtain the asymptotically dominant (in terms of the smallness parameter ϵ) perturbed distribution function response. This perturbed distribution function in the guiding-center phase space can then be *inversely transformed* back to the charged particle phase space and applied toward the field equations (i.e., Maxwell's equations) for a self-consistent kinetic description.

- Obtain Vlasov equation in the guiding center coordinates.
- Obtain Maxwell's equations in the guiding center coordinates.

Why do we need extra steps before using Maxwell's equations? It is because in Maxwell's equations the particle information (ρ, \mathbf{j}) are described not in the guiding center coordinates. The distribution we obtain from Vlasov equation must be transformed back to ordinary coordinates and then we can do the moment integral.

In the guiding center coordinates, density and current density can be expressed as

$$n = \bar{N}_0 + \bar{\nabla} \cdot \left(\frac{c\bar{N}_0}{B\Omega} \bar{\nabla}_{\perp} \delta\phi \right) + \frac{\bar{N}_0 \delta B_{\parallel}}{B} + \int \bar{F}_1 d\mathbf{p} \quad (15.1)$$

$$\begin{aligned} \mathbf{j} = & -\frac{\bar{N}_0 q^2}{cm} \delta A_{\parallel} \hat{b} + \int \frac{q\bar{\mathbf{P}}_{\parallel}}{m} \hat{b} d\bar{\mathbf{P}} + \frac{cq\bar{N}_0}{B} \hat{b} \times \bar{\nabla} \delta\phi + \frac{3c^2 \bar{N}_0 \bar{T}}{2B^2 \Omega} \hat{b} \times \bar{\nabla} \bar{\nabla}_{\perp}^2 \delta\phi \\ & + \frac{2c\bar{N}_0 \bar{T}}{B^2} \hat{b} \times \bar{\nabla} \delta B_{\parallel} + \frac{\hat{b}}{B} \times \bar{\nabla} (c\bar{N}_0 \bar{T}) + \hat{b} \times \bar{\nabla} \int c\bar{\mu} \bar{F}_1 d\bar{\mathbf{P}} \end{aligned} \quad (15.2)$$

Note that

- In Vlasov equation in the guiding center coordinates, the $\mathbf{E} \times \mathbf{B}$ drift, gradient drift and curvature drift all appears but the polarization drift is missing.
- In Equation 15.1, only the polarization term correction appears. From Hamilton's mechanics, all term that has a explicit time dependency will not contribute here since it will break the energy conservation of the system.

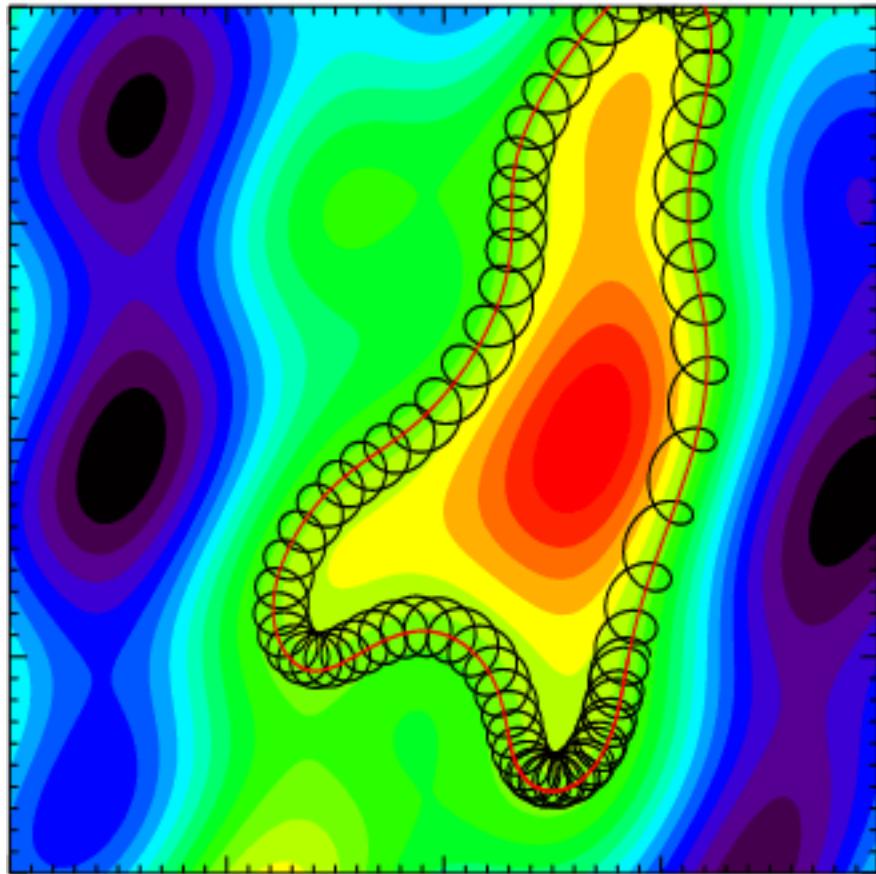


Figure 15.1: Charged particle orbit in a magnetic field pointing into the plane with electrostatic potential fluctuations.

15.3 The Gyrokinetic Ordering

15.3.1 From Kinetics to Gyrokinetics

Figure 15.1 shows the basic idea of gyrokinetic approximation. There is a strong magnetic field pointing into the plane. The electrostatic potential fluctuations are shown by the colored contours. The particle orbit is composed of two parts: a fast gyromotion and a slow $\mathbf{E} \times \mathbf{B}$ drift. In gyrokinetics we simply remove the fast gyromotion and introduce charged rings as quasiparticles, i.e. go from particle to gyrocenter coordinates.

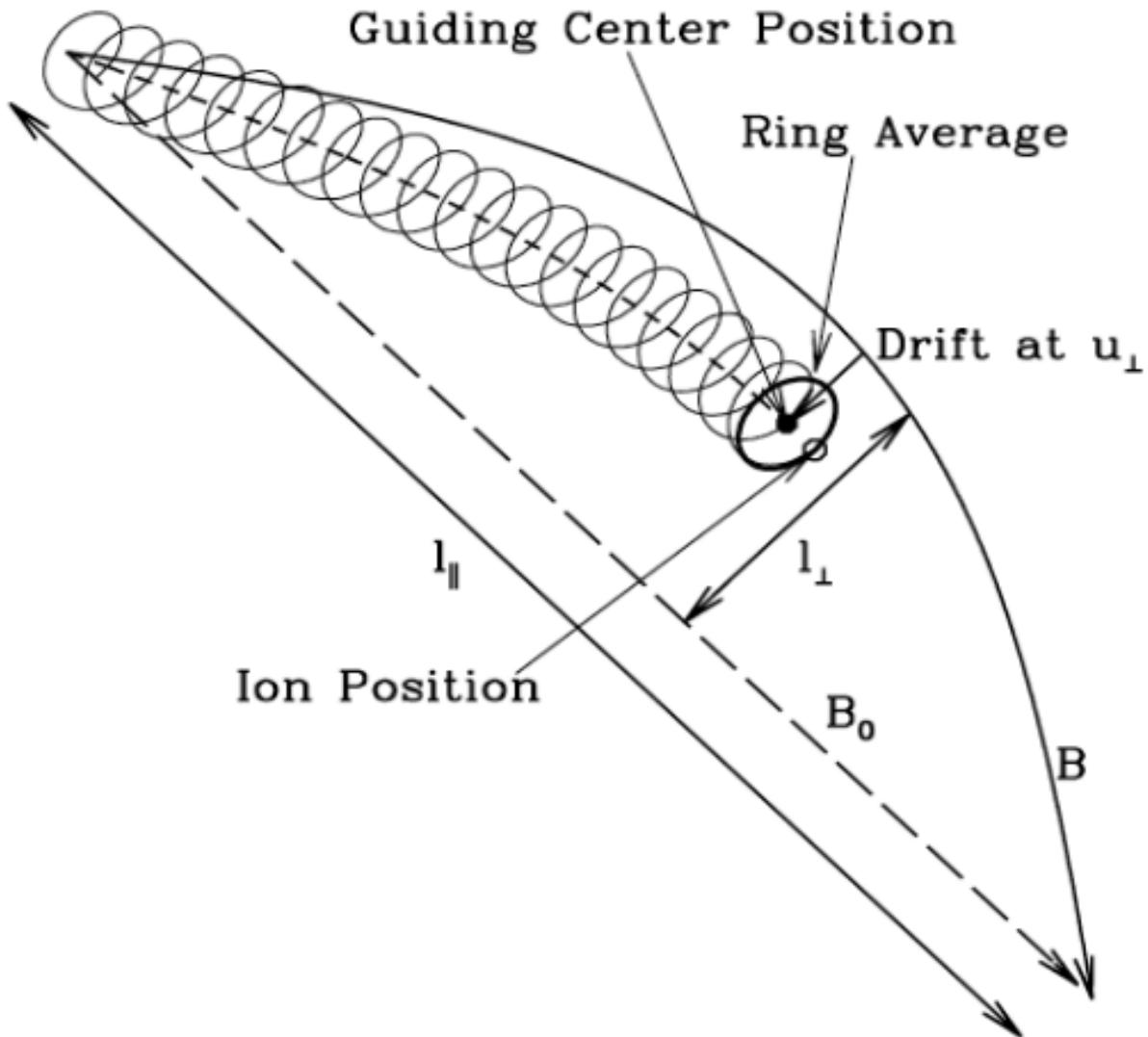


Figure 15.2: Gyrokinetic approximation.

- Slow time variation as compared to the gyromotion time scale:

$$\omega/\Omega_i \ll 1$$

- Spatial equilibrium scale much larger than the Larmor radius:

$$r_L/L \ll 1$$

- Strong anisotropy, i.e. only perpendicular components of the fluctuating quantities can be large:

$$k_{\parallel}/k_{\perp} \ll 1$$

- Small amplitude perturbations, i.e. energy of perturbation much smaller than the thermal energy:

$$e\phi/k_B T_e \ll 1$$

There exists a natural smallness parameter, $\epsilon = r_L/L$, which we are going to use in the ordering. In magnetically confined plasmas, typically we have $\epsilon \lesssim \mathcal{O}(10^{-2}) \ll 1$.

$$\frac{\omega}{\Omega} \sim \frac{r_L}{L} \sim \frac{k_{\parallel}}{k_{\perp}} \sim \frac{\delta F}{F_0} \sim \frac{\delta T}{T_0} \sim \frac{\delta n}{n_0} \sim \frac{|\delta \mathbf{B}|}{|\mathbf{B}_0|} \sim \frac{q\delta\phi}{T} \sim \epsilon$$

Usually low-frequency ($|\omega/\Omega_i| \sim \epsilon$) but short-wavelength ($k_{\perp} r_L \sim 1$) fluctuations are of interest in gyrokinetics. To include Landau resonance (??),

$$k_{\parallel} v_{\parallel} \sim \omega, \quad \text{or} \quad |k_{\perp} r_i| \sim 1$$

Noting, furthermore, for $|k_{\perp} r_i| \sim 1$ and the ratio of plasma ion pressure to the background magnetic field energy density $\beta_i \lesssim 1$:

$$\beta = \frac{P_{0i}}{P_{0B}} = \frac{m_i n_i v_{\perp i}^2 / 2}{B_0^2 / 2\mu_0} = \frac{v_{\perp i}^2}{v_A^2}$$

$$\left| \frac{\omega}{k_{\perp} v_A} \right| \sim \left| \frac{\omega}{\Omega_i} \right| \left| \frac{1}{k_{\perp} r_{iL}} \right| \beta_i^{1/2} \lesssim \epsilon$$

i.e. fast waves are systematically suppressed in the gyrokinetic orderings.

15.4 A Lagrangian Approach

If the Lagrangian of a dynamical system is known, e.g. for charged particle motion in non-canonical coordinates (\mathbf{x}, \mathbf{v}) :

$$L = \left(\frac{e}{c} \mathbf{A}(\mathbf{x}, t) + m\mathbf{v} \right) \cdot \dot{\mathbf{x}} - H(\mathbf{x}, \mathbf{v})$$

$$H = \frac{m}{2} v^2 + e\phi(\mathbf{x}, t)$$

with $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\nabla\phi - \partial_t \mathbf{A}/c$. The equation of motion are given by the Lagrange equations (I almost forget everything from theoretical mechanics...):

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad i = 1, \dots, 6$$

For charged particles,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{v}}} - \frac{\partial L}{\partial \mathbf{v}} = 0?$$

$$\dot{\mathbf{x}} = \mathbf{v}$$

$$\dot{\mathbf{v}} = \frac{e}{m} (\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

In gyrokinetics we add low-frequency, anisotropic, small-amplitude fluctuations:

$$\frac{\omega}{\Omega_i} \sim \frac{k_{\parallel}}{k_{\perp}} \sim \frac{e\phi}{T_e} \sim \epsilon$$

We need a transition from particle coordinates (\mathbf{x}, \mathbf{v}) to guiding center coordinates $(\mathbf{R}, v_{\parallel}, \mu, \varphi)$. The easy way is to construct a new Lagrangian using Lie transforms (???)

$$\Gamma = \left(mv_{\parallel} \hat{b}_0 + \frac{e}{c} \bar{\mathbf{A}}_{1\parallel} \hat{b}_0 + \frac{e}{c} \mathbf{A}_0 \right) \cdot d\mathbf{X} + \frac{mc}{e} \mu d\theta - \left(\frac{m}{2} v_{\parallel}^2 + \mu B_0 + \mu \bar{B}_{1\parallel} + e\bar{\phi}_1 \right) dt$$

where μ is the magnetic moment, θ is the gyroangle, and the overbar denotes a gyroaveraging operation. This gives the Euler-Lagrange equations

$$\begin{aligned}
\dot{\mathbf{X}} &= v_{\parallel} \hat{b} + \frac{B}{B_{\parallel}^*} \left[\frac{v_{\perp}}{B} \bar{\mathbf{B}}_{1\perp} + \frac{c}{B^2} \bar{\mathbf{E}}_1 \times \mathbf{B} + \frac{\mu}{m\Omega} \hat{b} \times \nabla(B + \bar{B}_{1\parallel}) + \frac{v_{\parallel}^2}{\Omega} (\nabla \times \hat{b})_{\perp} \right] \\
\dot{v}_{\parallel} &= \frac{\dot{\mathbf{X}}}{mv_{\parallel}} \cdot (e\bar{\mathbf{E}}_1 - \mu \nabla(B + \bar{B}_{1\parallel})) \\
\dot{\mu} &= 0
\end{aligned} \tag{15.3}$$

Equation 15.3 contains all the drifts we have seen in Chapter 7.

Applying the gyrokinetic approximation, the effective gyroaveraged potential over one gyroperiod can be written using Fourier transform (???:)

$$\begin{aligned}
\phi^{\text{eff}}(\mathbf{x}, r_L) &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \phi(\mathbf{x} + \mathbf{r}_L) \\
&= \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty} dk e^{i\mathbf{k}\cdot\mathbf{x}} \phi(\mathbf{k}) J_0(kr_L)
\end{aligned}$$

where J_0 is the zeroth order Bessel function.

15.4.1 Linear Gyrokinetics

We shall limit our considerations to that of a simple uniform plasma with an isotropic Maxwellian equilibrium distribution function. Assuming, furthermore, β (ratio between the plasma and magnetic pressures) $\ll 1$, such that there is negligible magnetic compression, the particle velocity distribution is then given by:

$$f(\mathbf{x}, \mathbf{v}, t) = f_M(\epsilon) + \delta f(\mathbf{x}, \mathbf{v}, t)$$

where $f_M(\epsilon) = n_0 / (\pi^{3/2} v_t^3) \exp(-\epsilon/v_t^2)$ is the Maxwellian distribution function, v_t is the thermal speed (HOW TO UNDERSTAND THIS???:)

$$\delta f = \frac{q}{T} f_M(\epsilon) \delta \phi + e^{-\mathbf{r}_L \cdot \nabla} \delta g$$

$T = mv_t^2/2$, δg satisfies the following linear gyrokinetic equation:

$$\begin{aligned}
\left(\frac{\partial}{\partial t} + v_{\parallel} \mathbf{b}_0 \cdot \nabla \right) \delta g &= \frac{q}{T} f_M(\epsilon) \frac{\partial}{\partial t} \langle \delta L_g \rangle_{\alpha} \\
\delta L_g &= e^{-\mathbf{r}_L \cdot \nabla} \delta L \\
\delta L &= \delta \phi - v_{\parallel} \delta A_{\parallel} / c
\end{aligned} \tag{15.4}$$

and $\langle \dots \rangle_\alpha$ denotes averaging over the gyrophase angle, α . Here, the field variables are the scalar and vector potentials, $\delta\phi$ and $\delta\mathbf{A}$, with $\delta A_{\parallel} = \delta\mathbf{A} \cdot \mathbf{b}_0$ and the $\nabla \cdot \delta\mathbf{A} = 0$ Coulomb gauge. The operator $e^{-\mathbf{r}_L \cdot \nabla}$, meanwhile, represents the transformation between the particle and guiding center positions.

The corresponding field equations are the Poisson's equation and the parallel Ampère's law, $\nabla^2 \delta A_{\parallel} = -4\pi \delta J_{\parallel}/c$. In the low-frequency and $|k\lambda_D|^2 \ll 1$ limit with λ_D being the Debye length, Poisson's equation can be approximated as the quasi-neutrality condition; $\sum_j n_0 j q_j < \delta f_j \rangle_v \simeq 0$. Here, $\langle \dots \rangle_v = \int d\mathbf{v}(\dots)$ is the velocity-space integral, and subscript j runs over the particle species. Meanwhile, substituting the parallel Ampère's law into the $\nabla \cdot \delta\mathbf{J} \simeq 0$ quasi-neutrality condition as derived by Equation 15.4 yields a generalized linear gyrokinetic vorticity equation, which is often convenient to use in studying shear/kinetic Alfvén wave dynamics.

Linear kinetic Alfvén wave properties

(WARNING: SUPER HARD TO FOLLOW!) For plane wave ω, \mathbf{k} perturbations, Equation 15.4 gives:

$$\delta g_{\mathbf{k}} = -\frac{q}{T} f_M J_0(k_{\perp} \mathbf{r}_L) \frac{\omega}{k_{\parallel} v_{\parallel} - \omega} \left(\delta\phi - \frac{v_{\parallel}}{c} \delta A_{\parallel} \right)_{\mathbf{k}}$$

J_0 is the Bessel function and $J_0(k_{\perp} \mathbf{r}_L)$ corresponds to the gyro-averaging of the coordinate transformation, that is:

$$\langle \exp(-\mathbf{r}_L \cdot \nabla) \rangle_\alpha = J_0(k_{\perp} r_L)$$

In SAW/KAW analyses, it is sometimes convenient to introduce an *effective induced parallel potential* defined by $\mathbf{b}_0 \cdot \nabla \delta\Psi = -\partial_t \delta A_{\parallel}/c$ or:

$$\delta\Psi_{\mathbf{k}} = \omega \delta A_{\parallel \mathbf{k}} / (ck_{\parallel})$$

$\delta\Psi$, thus, gives rise to the induced parallel electric field; that is, the net parallel electric field is given by:

$$\begin{aligned} \delta E_{\parallel} &= -\mathbf{b}_0 \cdot \nabla (\delta\phi - \delta\Psi) \quad \text{or} \\ \delta E_{\parallel \mathbf{k}} &= -ik_{\parallel} (\delta\phi - \delta\Psi)_{\mathbf{k}} \end{aligned}$$

The quasi-neutrality condition then straightforwardly yields: (Chen and Hasegawa 1991)

$$\sum_j \left(\frac{n_0 q^2}{T_0} \right)_j \{ \delta\phi_{\mathbf{k}} + \Gamma_{0kj} [\xi_{kj} Z_{kj} \delta\phi_{\mathbf{k}} - (1 + \xi_{kj} Z_{kj} \delta\Psi_{\mathbf{k}})] \} = 0 \quad (15.5)$$

Here, $\xi_{kj} = \omega/|k_{\parallel}|v_{tj}$, $Z_{kj} = Z(\xi_{kj})$ with Z being the well-known plasma dispersion function (Equation 12.33), and $\Gamma_{0kj} = I_0(b_{kj}) \exp(-b_{kj})$ with I_0 the modified Bessel function and $b_{kj} = k_{\perp}^2 r_{Lj}/2 = k_{\perp}^2 (T_j/m_j)/\Omega_j^2$. The linear gyrokinetic vorticity equation, meanwhile, is given by: (Chen and Hasegawa 1991)

$$i \frac{c^2}{4\pi\omega} k_{\parallel}^2 k_{\perp}^2 \delta\Psi_{\mathbf{k}} - i \sum_j \left(\frac{n_0 q^2}{T_0} \right)_j (1 - \Gamma_{0kj}) \omega \delta\phi_{\mathbf{k}} = 0 \quad (15.6)$$

Nothing that, for KAW, $|k_{\perp} r_{Li}| \sim \mathcal{O}(1)$ and $|k_{\perp} r_{Le}| \ll 1$ and, thus, $\Gamma_{0ke} \simeq 1$, Equation 15.5 and Equation 15.6 then become

$$\begin{aligned} \epsilon_{s\mathbf{k}} \delta\phi_{\mathbf{k}} &= [1 + \xi_{ke} Z_{ke} + \tau(1 + \Gamma_k \xi_{ki} Z_{ki})] \delta\phi_{\mathbf{k}} \\ &= [1 + \xi_{ke} Z_{ke} + \tau \Gamma_k (1 + \xi_{ki} Z_{ki})] \delta\Psi_{\mathbf{k}} \end{aligned} \quad (15.7)$$

and

$$\omega^2 \delta\phi_{\mathbf{k}} = k_{\parallel}^2 v_A^2 \frac{b_k}{1 - \Gamma_k} \delta\Psi_{\mathbf{k}} \quad (15.8)$$

Here, $\tau = T_{0e}/T_{0i}$, $b_k = b_{ki}$, $\Gamma_k = \Gamma_{0ki}$, and $\epsilon_{s\mathbf{k}}$ is the dielectric constant for the slow-sound (ion-acoustic) wave (SSW).

It is also instructive, as done in some literatures, to define the effective parallel potential, $\delta\phi_{\parallel\mathbf{k}} = \delta\phi_{\mathbf{k}} - \delta\Psi_{\mathbf{k}}$, and rewrite Equation 15.7 and Equation 15.8 as

$$\epsilon_{s\mathbf{k}} \delta\phi_{\parallel\mathbf{k}} = -\tau(1 - \Gamma_k) \delta\Psi_{\mathbf{k}} \quad (15.9)$$

and

$$\left[\omega^2 - k_{\parallel}^2 v_A^2 \frac{b_k}{1 - \Gamma_k} \right] \delta\Psi_{\mathbf{k}} = -\omega^2 \delta\phi_{\parallel\mathbf{k}} \quad (15.10)$$

Equation 15.9 and Equation 15.10 demonstrate the coupling between SAW and SSW via the finite $|k_{\perp} r_{Ls}|$ term. In the $|k_{\perp} r_{Li}| \sim \mathcal{O}(1)$ short-wavelength limit, SAW evolves into KAW due to both the finite $|k_{\perp} r_{Li}|$ and $|k_{\perp} r_{Ls}|$ effects. (???) More specifically, the coupled KAW-SSW dispersion relation becomes

$$\omega_{\mathbf{k}}^2 \left[1 - \frac{\tau(1 - \Gamma_k)}{\epsilon_{s\mathbf{k}}} \right] = k_{\parallel}^2 v_A^2 \frac{b_k}{1 - \Gamma_k} \quad (15.11)$$

Let us concentrate on the KAW branch and, to further simplify the analysis, assume $1 \gg \beta_i \sim \beta_e \gg m_e/m_i$. With $|\omega| \sim |k_{\parallel}v_A|$, we then have $|\xi_{ki}| = |\omega/k_{\parallel}v_{ti}| \sim \beta_i^{-1/2} \gg 1 \gg |\xi_{ke}| \sim (m_e/m_i\beta_e)^{1/2}$, and, keeping only the lowest order $\mathbf{O}(1)$ terms:

$$\epsilon_{s\mathbf{k}} \simeq 1 + \tau(1 - \Gamma_k) \equiv \sigma_k$$

From Equation 15.11, we then have

$$\omega_{\mathbf{k}}^2 \simeq k_{\parallel}^2 v_A^2 \frac{\sigma_k b_k}{1 - \Gamma_k} \quad (15.12)$$

As to wave polarizations, which are useful for wave identification in observations, we can readily derive:

$$\left| \frac{c\delta\mathbf{E}_{\perp}}{\delta\mathbf{B}_{\perp}} \right| = v_A \left[\frac{b_k}{\sigma_k(1 - \Gamma_k)} \right]^{1/2} \quad (15.13)$$

and

$$\left| \frac{c\delta E_{\parallel}}{\delta\mathbf{B}_{\perp}} \right| = v_A \left| \frac{k_{\parallel}}{k_{\perp}} \right| \tau \left[\frac{b_k(1 - \Gamma_k)}{\sigma_k} \right]^{1/2} \quad (15.14)$$

ADD PLOTS for Equation 15.12, Equation 15.13, and Equation 15.14!!!

Equation 15.14 show that, for a fixed $|k_{\parallel}/k_{\perp}|, |\delta E_{\parallel}/\delta\mathbf{B}_{\perp}|$ increases with b_k . Since wave-particle energy and momentum exchanges are proportional to $|\delta E_{\parallel}|$, short-wavelength KAW are, thus, expected to play crucial roles in the heating, acceleration, and transport of charged particles.

In addition to having a significant δE_{\parallel} , another important property of KAW, in contrast to SAW, is that KAW has a finite perpendicular (to \mathbf{B}_0) group velocity, $\mathbf{v}_{g\perp}$. Assuming $|k_{\perp}r_{Li}|^2 \ll 1$, we have, letting $\omega_A^2 \equiv k_{\parallel}^2 v_A^2$:

$$\omega_{\mathbf{k}}^2 \simeq \omega_A^2 (1 + k_{\perp}^2 \hat{r}^2) \quad (15.15)$$

where

$$\hat{r}^2 = (3/4 + \tau)r_{Li}^2 \quad (15.16)$$

Thus

$$\mathbf{v}_{g\perp} = \frac{\partial \omega_{\mathbf{k}}}{\partial \mathbf{k}_{\perp}} \simeq \frac{\omega_A^2}{\omega_{\mathbf{k}}} \hat{r}^2 \mathbf{k}_{\perp} \quad (15.17)$$

Linear mode conversion of KAW

Equation 15.15 has a significant implication in non-uniform plasmas. Consider, again, a slab plasma with a non-uniform $\omega_A^2(x)$ and $k_\perp^2 = k_x^2(x)$ being the WKB wavenumber in the non-uniformity x -direction. Equation 15.15 then indicates that KAW is propagating ($k_x^2 > 0$) in the $\omega_{\mathbf{k}}^2 > \omega_A^2(x)$ region, and it is cutoff ($k_x^2 < 0$) in the $\omega_{\mathbf{k}}^2 < \omega_A^2(x)$ region. That $\mathbf{v}_{g\perp}$ is finite also suggests that, in contrast to SAW, an initial smooth perturbation will not only evolve into short wavelengths, but also propagate toward the lower $\omega_A^2(x)$ region. These features are illustrated in ?@fig-gk-KAW-evolution(b); where the spatial-temporal evolution of KAW is solved explicitly according to the following wave equation: (DO IT MYSELF!!!)

$$\left[\hat{r}^2 \frac{\partial^2}{\partial x^2} - 1 - \frac{1}{\omega_A^2(x)} \frac{\partial^2}{\partial t^2} \right] \delta B_y(x, t) = 0 \quad (15.18)$$

Equation 15.18 can be readily derived from Equation 15.15 by letting $\omega_{\mathbf{k}} = i\partial/\partial t$ and $k_\perp = -i\partial/\partial x$. The spatial profile of $\omega_A^2(x)/\omega^2 = 1/(1 + x^2/L^2)$ is shown in ?@fig-gk-KAW-evolution(a), with L indicating the profile length-scale, so that the KAW wave-packet frequency is assumed to be consistent with the SAW frequency at $x = 0$. ?@fig-gk-KAW-evolution(b) shows the propagation of the KAW wave-packet in the direction of radial non-uniformity, consistent with Equation 15.17.

That there exists a finite perpendicular group velocity also implies, in the steady state, the removal of “singular” resonance and linear mode conversion process (Hasegawa and Chen 1976, see also Chapter 16). More specifically, the corresponding wave equation is given by:

$$\left[\hat{r}^2 \frac{\partial^2}{\partial x^2} + \left(\frac{\omega_0^2}{\omega_A^2(x)} - 1 \right) \right] \delta \hat{B}_y(x) = \delta \hat{B}_{y0} \quad (15.19)$$

Here, ω_0 is the external driving frequency. In the ideal SAW ($\hat{r} \rightarrow 0$) limit, there is the resonance singularity at x_0 , where $\omega_0^2 = \omega_A^2(x_0)$. Noting that, near $x = x_0$, $\omega_A^2(x) \simeq \omega_0^2 + (\omega_A^2)'(x_0)(x - x_0) \equiv \omega_0^2 - (\omega_0^2/L_A)(x - x_0)$, Equation 15.19 can be approximated as an inhomogeneous Airy equation and solved analytically. Equation 15.19 can then be solved, with appropriate boundary conditions, by connecting the solutions valid away from the $x = x_0$ resonance layer via the analytic solution of the inhomogeneous Airy equation valid near $x = x_0$ (Hasegawa and Chen 1975, 1976). The solutions away from the singular layer are given by:

$$\delta \hat{B}_y(x) = \begin{cases} \frac{\delta \hat{B}_{y0}}{\epsilon_A(x)} & \text{for } \omega_0^2 < \omega_A^2(x) \\ \frac{\delta \hat{B}_{y0}}{\epsilon_A(x)} - \frac{\sqrt{\pi} \delta \hat{B}_{y0}}{(\hat{r}/L_A)^{1/2}} \left(\frac{\hat{r}^2}{\epsilon_A(x)} \right)^{1/4} \exp \left[i \int_{x_0}^x \left(\frac{\epsilon_A(x')}{\hat{r}^2} \right)^{1/2} dx' + i \frac{\pi}{4} \right] & \text{for } \omega_0^2 > \omega_A^2(x) \end{cases} \quad (15.20)$$

where

$$\epsilon_A(x) = \frac{\omega_0^2}{\omega_A^2(x)} - 1$$

The corresponding numerical solutions are plotted in ?@fig-gk-KAW-MHD. (I DON'T UNDERSTAND!!!)

Both the analytical results and mode conversion process exhibit two important features. One is, instead of being singular, the amplitude at $x = x_0$ (where $\omega_A(x_0) = \omega_0$) is amplified by the Airy swelling factor $(L_A/\hat{r})^{2/3}$. Here, we recall L_A is the scale length of ω_A and \hat{r} , from Equation 15.16, is of $\mathcal{O}(r_{Li})$, and, hence, $|L_A/\hat{r}| \gg 1$. The other is the singularity at $x = x_0$ is being replaced by the Airy scale length $\Delta_0 = (\hat{r}^2 L_A)^{1/3}$. Recalling, from Equation 10.37, $|k_x| \simeq |\omega'_A|t \simeq (\omega/L_A)t$, there then exists a KAW formation time scale given by $(\omega_0/L_A)t_0 \simeq 1/\Delta_0$; i.e., $\omega_0 t_0 \simeq (L_A/\hat{r})^{2/3}$. Taking, for an example, a typical laboratory plasma, $L_A/\hat{r} \simeq \mathcal{O}(10^3)$, we have $\omega_0 t_0 \simeq \mathcal{O}(10^2)$, suggesting that it is reasonable to anticipate, in the presence of SAW continuous spectrum, the appearance of KAW in such plasmas.

The main mode identification method for KAWs is based on the measurement of the wave polarization, $|c\delta\mathbf{E}_\perp/v_A\delta\mathbf{B}_\perp|$. Observations can be found from *Van Allen Probes* in the Earth's inner magnetosphere and the *Cluster* satellites in the solar wind.

15.4.2 Nonlinear Gyrokinetics

The idea of deriving the gyrokinetic equations is very similar to the derivation of 5/10-moment equations. Extra care is needed because of the coordinate transformation to the guiding center coordinates. For species α ,

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v} \cdot \frac{\partial f_\alpha}{\partial \mathbf{x}} + \frac{q_\alpha}{m_\alpha} \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \frac{\partial f_\alpha}{\partial \mathbf{v}} = \sum_{\alpha'} C_{\alpha\alpha'}(f_\alpha, f'_\alpha)$$

We expand f in different orders:

$$f = f_0 + \delta f = f_0 + f_1 + f_2 + \dots$$

and treat the velocity as a first order quantity. For simplicity, we now ignore the species subscript α . Separating the equilibrium and perturbation terms, we have

$$\begin{aligned} & \frac{\partial f_0}{\partial t} + \frac{\partial \delta f}{\partial t} \\ & + \mathbf{v}_\perp \cdot \frac{\partial \delta f}{\partial \mathbf{x}} + \mathbf{v}_\parallel \cdot \frac{\partial \delta f}{\partial \mathbf{x}} \\ & + \frac{q}{m} \left(\delta \mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}_0}{c} + \frac{\mathbf{v} \times \delta \mathbf{B}_0}{c} \right) \cdot \frac{\partial f_0}{\partial \mathbf{v}} + \frac{q}{m} \left(\delta \mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}_0}{c} + \frac{\mathbf{v} \times \delta \mathbf{B}}{c} \right) \cdot \frac{\partial \delta f}{\partial \mathbf{v}} = \\ & C(f_0, f_0) + C(f_0, \delta f) + C(\delta f, f_0) + C(\delta f, \delta f) \end{aligned}$$

Based on the gyrokinetic ordering,

$$\begin{aligned}
\frac{\partial f_0}{\partial t} &\sim \omega_{\text{heat}} f_0 \sim \epsilon^3 \Omega f_0 \\
\frac{\partial \delta f}{\partial t} &\sim \omega \delta f \sim \epsilon^2 \Omega f_0 \\
\mathbf{v}_\perp \cdot \frac{\partial \delta f}{\partial \mathbf{x}} &\sim v_{\text{th}} k_\perp \delta f \sim \epsilon \Omega f_0 \\
\mathbf{v}_\parallel \cdot \frac{\partial \delta f}{\partial \mathbf{x}} &\sim v_{\text{th}} k_\parallel \delta f \sim \epsilon^2 \Omega f_0 \\
\frac{q}{m} \nabla_\perp \delta \phi \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\perp \delta \phi \frac{f_0}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{m} \nabla_\parallel \delta \phi \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\parallel \delta \phi \frac{f_0}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{mc} \frac{\partial \delta \mathbf{A}}{\partial t} \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{m} \omega \delta A \frac{f_0}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{mc} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} B_0 \frac{f_0}{v_{\text{th}}} \sim \Omega f_0 \\
\frac{q}{mc} (\mathbf{v} \times \delta \mathbf{B}) \cdot \frac{\partial f_0}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} \delta B \frac{f_0}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{m} \nabla_\perp \delta \phi \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\perp \delta \phi \frac{\delta f}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0 \\
\frac{q}{m} \nabla_\parallel \delta \phi \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{m} k_\parallel \delta \phi \frac{\delta f}{v_{\text{th}}} \sim \epsilon^3 \Omega f_0 \\
\frac{q}{mc} \frac{\partial \delta \mathbf{A}}{\partial \mathbf{v}} &\sim \frac{q}{m} \omega \delta A \frac{\delta f}{v_{\text{th}}} \sim \epsilon^3 \Omega f_0 \\
\frac{q}{mc} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} B_0 \frac{\delta f}{v_{\text{th}}} \sim \epsilon \Omega f_0 \\
\frac{q}{mc} (\mathbf{v} \times \delta \mathbf{B}) \cdot \frac{\partial \delta f}{\partial \mathbf{v}} &\sim \frac{q}{mc} v_{\text{th}} \delta B \frac{\delta f}{v_{\text{th}}} \sim \epsilon^2 \Omega f_0
\end{aligned}$$

ω_{heat} is ???

The zeroth order equation is

$$\frac{q}{mc} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial f_0}{\partial \mathbf{v}} = \Omega (\mathbf{v}_\perp \times \hat{b}_0) \cdot \frac{\partial f_0}{\partial \mathbf{v}} = 0$$

If we transform to a cylindrical coordinates in velocity space (Equation 14.8), this can be written as

$$(\mathbf{v}_\perp \times \hat{b}_0) \cdot \frac{\partial f_0}{\partial \mathbf{v}} = - \frac{\partial f_0}{\partial \theta} = 0$$

This means that the equilibrium distribution function f_0 does not depend on the gyrophase, i.e. is gyroscopic.

The first order equation is

$$\mathbf{v}_\perp \cdot \frac{\partial \delta f_1}{\partial \mathbf{x}} + \frac{q}{m} \left(-\nabla_\perp \delta \phi + \frac{\mathbf{v} \times \delta \mathbf{B}}{c} \right) \cdot \frac{\partial f_0}{\partial \mathbf{v}} + \frac{q}{mc} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial \delta f_1}{\partial \mathbf{v}} = C(f_0, f_0)$$

Multiplying with $(1 + \ln f_0)$ and integrating over phase space, we get

$$\int \ln(f_0) C(f_0, f_0) d\mathbf{x} d\mathbf{v} = 0$$

From the proof of Boltzmann's H-theorem, we conclude:

$$\begin{aligned} f_0 &= f_{0,M}(v) = \frac{n_0}{\pi^{3/2} v_{\text{th}}^3} e^{-v^2/v_{\text{th}}^2} \\ C(f_{0,M}, f_{0,M}) &= 0 \\ \frac{\partial f_{0,M}}{\partial \mathbf{v}} &= -2f_{0,M} \frac{\mathbf{v}}{v_{\text{th}}^2} = -f_{0,M} m \frac{\mathbf{v}}{T} \end{aligned}$$

The equilibrium distribution is Maxwellian. The first order equation thus becomes (???)

$$\mathbf{v}_\perp \cdot \frac{\partial \delta f_1}{\partial \mathbf{x}} - \Omega \frac{\partial \delta f_1}{\partial \theta} = -f_{0,M} \mathbf{v} \cdot \nabla_\perp \left(\frac{q \delta \phi}{T} \right)$$

The solution to this is a combination of a homogeneous solution and a particular solution. For the homogeneous part,

$$\mathbf{v}_\perp \cdot \frac{\partial h}{\partial \mathbf{x}} - \Omega \frac{\partial h}{\partial \theta} \Big|_{\mathbf{x}} = 0, \quad \mathbf{x} = \mathbf{X} + \mathbf{r}_L, \quad \Omega \frac{\partial h}{\partial \theta} \Big|_{\mathbf{x}} = 0$$

For the particular part, the solution is

$$-q\delta\phi(\mathbf{x}, t)f_0(\mathbf{v})/T$$

Thus the complete solution is

$$\delta f_1(\mathbf{x}, \mathbf{v}, t) = h(\mathbf{X}, v_\parallel, v_\perp, t) - q\delta\phi(\mathbf{x}, t)f_0(\mathbf{v})/T$$

We still need an equation for h for closure. Therefore we have to go to the second order equation

$$\begin{aligned}
\frac{\partial \delta f_1}{\partial t} + v_{\parallel} \hat{b}_0 \cdot \nabla \delta f_1 + \frac{q}{m} \left(-\nabla_{\perp} \delta \phi + \frac{\mathbf{v} \times \delta \mathbf{B}}{c} \right) \cdot \frac{\partial \delta f_1}{\partial \mathbf{v}} \\
- \frac{q}{m} (\nabla_{\parallel} \delta \phi) \frac{\partial f_0}{\partial v_{\parallel}} - \frac{q}{mc} \frac{\partial \delta \mathbf{A}}{\partial t} \cdot \frac{\partial f_0}{\partial \mathbf{v}} \\
+ \mathbf{v}_{\perp} \cdot \nabla \delta f_2 + \frac{q}{mc} (\mathbf{v} \times \mathbf{B}_0) \cdot \frac{\partial \delta f_2}{\partial \mathbf{v}} \\
= C(\delta f_1, f_0) + C(f_0, \delta f_1)
\end{aligned}$$

From particle coordinates to guiding center coordinates

$$\begin{aligned}
\frac{\partial h}{\partial t} + \frac{d\mathbf{X}}{dt} \cdot \frac{\partial h}{\partial \mathbf{X}} + \frac{q}{m} \left(-\nabla_{\perp} \delta \phi + \frac{\mathbf{v} \times \delta \mathbf{B}}{c} \right) \cdot \frac{\partial h}{\partial \mathbf{v}} = \\
C(h, f_0) + C(f_0, h) + \frac{q}{T} f_0 \frac{\partial \chi}{\partial t} + \Omega \frac{\partial \delta f_2}{\partial \theta} \Big|_{\mathbf{x}}
\end{aligned}$$

where $\xi \equiv \delta \phi - \mathbf{v} \cdot \delta \mathbf{A}/c$.

For any quantity Q , the gyrophase averaging at fixed guiding center position is

$$\langle Q(\mathbf{x}, \mathbf{v}, t) \rangle_{\mathbf{x}} \equiv \frac{1}{2\pi} \int_0^{2\pi} Q \left(\mathbf{x} - \frac{\mathbf{v} \times \hat{b}_0}{\Omega}, \mathbf{v}, t \right) d\theta$$

We have

$$\begin{aligned}
\langle h(\mathbf{x}, \mathbf{v}, t) \rangle_{\mathbf{x}} &= h \\
\langle \mathbf{v}_{\perp} \cdot \nabla Q \rangle_{\mathbf{x}} &= -\Omega \left\langle \frac{\partial Q}{\partial \theta} \Big|_{\mathbf{x}} \right\rangle_{\mathbf{x}} = 0
\end{aligned}$$

The equation for the gyrocenter distribution function h is then

$$\frac{\partial h}{\partial t} + v_{\parallel} \hat{b}_0 \cdot \frac{\partial h}{\partial \mathbf{X}} + \frac{c}{B_0} \left(\hat{b}_0 \times \frac{\partial \langle \chi \rangle_{\mathbf{x}}}{\partial \mathbf{X}} \right) \cdot \frac{\partial h}{\partial \mathbf{X}} = \langle C(h) \rangle_{\mathbf{x}} + \frac{q}{T} f_0 \frac{\partial \langle \chi \rangle_{\mathbf{x}}}{\partial t}$$

The quasi-neutrality $\sum_{\alpha} q_{\alpha} \delta n_{\alpha} = 0$ now becomes

$$\begin{aligned}
\sum_{\alpha} \left[-\frac{q_{\alpha}^2 n_{0\alpha}}{T_{0\alpha}} \delta \phi(\mathbf{x}, t) + q_{\alpha} \int h_{\alpha} \left(\mathbf{x} + \frac{\mathbf{v} \times \hat{b}_0}{\Omega_{\alpha}}, \mathbf{v}, t \right) d\mathbf{v} \right] = 0 \\
\sum_{\alpha} \left(-\frac{q_{\alpha}^2 n_{0\alpha}}{T_{0\alpha}} \delta \phi(\mathbf{x}, t) + q_{\alpha} \int \langle h_{\alpha} \rangle d\mathbf{v} \right) = 0
\end{aligned}$$

Last but not least we need the modified Ampère's law:

$$\nabla \times \delta \mathbf{B}(\mathbf{x}, t) = \frac{4\pi}{c} \sum_{\alpha} q_{\alpha}$$

The distribution function now depends on 6 variables, $f = f(\mathbf{X}, v_{\parallel}, \mu; t)$. The Vlasov equation can be written as

$$\frac{\partial f}{\partial t} + \dot{\mathbf{X}} \cdot \frac{\partial f}{\partial \mathbf{X}} + v_{\parallel} \frac{\partial f}{\partial v_{\parallel}} = 0$$

where \mathbf{X} is the gyrocenter position, v_{\parallel} is the parallel velocity, and μ is the magnetic moment. The appropriate field equations are now

$$\begin{aligned} \frac{n_1}{n_0} &= \frac{\bar{n}_1}{n_0} - (1 - |I_0^2|) \frac{e\phi_1}{T} + |eI_0I_1| \frac{B_{1\parallel}}{B} \\ \nabla_{\perp}^2 A_{1\parallel} &= -\frac{4\pi}{c} \sum \bar{J}_{1\parallel} \\ \frac{B_{1\parallel}}{B} &= -\sum \epsilon_{\beta} \left(\frac{\bar{p}_{1\perp}}{n_0 T} + |xI_0I_1| \frac{e\phi_1}{T} + |x^2 I_1^2| \frac{B_{1\parallel}}{B} \right) \end{aligned}$$

These equations remove the irrelevant space-time scales and become nonlinear 5D equations instead 6D.

15.5 Applications

Gyrokinetics has emerged as the standard approach to plasma turbulence. It unveils several new understandings:

- Importance of Marginal Stability and Critical Gradients
- Importance of Zonal Flows (proper Boltzmann electron response)
- Tiny electron-scale ETG turbulence can sometimes be important

Some issues under investigation where gyrokinetics may shed light on:

- Role of microturbulence in space physics?
- transport across boundary layers
- fast magnetic reconnection
- coronal and solar wind heating
- How does MHD turbulence dissipate energy?
- How do fast particles interact with turbulence?

15.6 Limitations

What gyrokinetics cannot resolve:

- cyclotron resonance
- gyro-related phenomena

At least the orginal form of gyrokinetic cannot include EMIC waves and whistler waves.

16 Field Line Resonance

The resonant mode coupling is one of major physical processes in the space plasmas including the magnetic reconnection and collisionless shock wave. The field line resonance describes *the resonant interaction between compressional fast waves and shear Alfvén waves in a non-uniform plasma* such as the Earth magnetosphere. The concept of this resonant mode coupling was first outlined by Tsutomu Tamao in 1961 and later discussed in his seminal paper on hydromagnetic coupling resonances (Tamao, 1965). Experimental work by John Samson+ (1971) made resonant mode coupling the current paradigm of planetary magnetospheric ULF wave research. Thinking of magnetosphere as a microwave oven: instead of heating food with microwaves, we heat Earth with EM waves.

16.1 Historial Review

Before 1950s, no people thought about the idea of standing waves in space simply because a vacuum had been assumed. Right after the space age began, Dungey proposed that standing Alfvén waves could be excited on geomagnetic field lines.

- Ground observations have shown discrete frequencies for EM wave power. In the magnetosphere the Alfvén speed is typically ~ 1000 km/s, while typical periods of geomagnetic pulsations are 10-600 s. Thus typical wavelengths are $10^4 - 10^6$ km, or $1 - 100 R_E$, comparable with the size of the magnetosphere itself. So uniform plasma theory is clearly inadequate.
- In the 1950s people realized that MHD waves of poloidal and toroidal modes can be coupled, and if we thought the modes as standing waves, we might explain the discrete frequencies. However, good agreements were not found because there were poor estimates of the magnetospheric plasma density.
- In the 1960s, [Sugiura 1961] showed that waves are observed simultaneously at both ends of the same field line, which indicated that the waves were guided. [Nagata+ 1963] showed pulsations at conjugate points could be matched cycle for cycle. The phase comparison could indicate the possible harmonics of standing waves. [Patel 1965] reported discovery of both transverse and compressional waves in space from Explorer 12 magnetometer data that are correlated with ground measurements, confirming that they were the same. [Cummings+ 1969] showed long-lived and frequent waves with nice numerical comparison of their periods from the poloidal and toroidal mode equations.

- In the 1970s a simpler box model was proposed to explain field line resonance. The idea of standing waves could not explain why only some field lines are preferentially excited (with the foot point of the field line near magnetic latitude 70°). In the box model, discrete frequencies are treated as sources that reflect the properties of the magnetopause thickness with a preferred KHI growth rate.
- In the 1980s the cavity model was proposed to link the discrete frequencies to the eigenmodes if we treated the magnetosphere as a cavity.
- In the 1980s the waveguide model was proposed based on the cavity idea to loose the constrain in the azimuthal direction. This could explain many ground observation of discrete frequency, large amplitude Pc5 waves but satellite measurements often saw small amplitude Pc3/4 waves. Unlike field line resonances, which are described by a single eigennumber (k_z), cavity modes have three eigennumbers, corresponding to the number of wavelengths along a field line (k_z), azimuthally around the Earth (k_y), and radially between the magnetopause (or bow shock) and the inner reflection point (n). It is argued in the 1990s that if the lowest order mode has a frequency of, say, 2 mHz, then above perhaps 10 mHz the higher order modes will be so numerous and close in frequency that they could probably not be resolved in the data given their inherent width and the frequency resolution of a typical spectrum. Rather they would appear as a continuum. As a continuum, they can not be responsible for exciting discrete frequencies in this higher frequency band (corresponding to Pc3/4). Maybe the waveguide model is only relevant to Pc5 pulsations?

Now, the big question is: where do the quantized numbers of the observed ULF waves come from [Kivelson & Southwood, 1986]? Researchers borrowed ideas from ionospheric radio propagation, laser fusion and plasma physics and came up with the names *cavity* and *waveguide*. For the box model that we will discuss in the next section, if the z boundaries are perfectly reflecting, wave fields must have standing structure in the z direction, and allowed parallel wave numbers are quantized ($k_z = m\pi/c$, as in Equation 16.19). If the boundaries are weakly absorptive, the parallel wave numbers are complex, but the real parts are still quantized as above. If we impose periodic boundary conditions in the y direction, the wavenumber in y, k_y , needs to be quantized as well.

Another question: if these discretized frequencies correspond to standing waves propagating along the field lines, at the footprints (i.e. ground measurements) the amplitudes shall be the smallest? Why are we still able to observe that on the ground?

16.2 Theory

The theory starts from linearized cold MHD Equation 10.30 and Equation 10.31.

16.2.1 Axisymmetric Spherical Coordinates

First, let's treat the problem in spherical coordinates (r, θ, ϕ) , as done by (Radoski and Carovillano 1966).

The background density is taken to be only a function of the radial distance, $\rho = \rho(r)$. In the analysis below the perturbed variables, \mathbf{B}_1 , \mathbf{E} , \mathbf{j} and \mathbf{u} are treated as axisymmetric, i.e. in spherical coordinates any perturbed function has the form

$$= \delta_r(r, \theta)\hat{r} + \delta_\theta(r, \theta)\hat{\theta} + \delta_\phi(r, \theta)\hat{\phi}$$

Under this symmetry, the variables can be separated into two independent sets referred to as the *toroidal* and *poloidal* variables:

- Toroidal variables

$$B_{1\phi}, E_r, E_\theta, u_\phi, j_r, j_\theta; \nabla \cdot \mathbf{u} = \frac{1}{r \sin \theta} \frac{\partial u_\phi}{\partial \phi} \equiv 0 \text{ (incompressible)}$$

- Poloidal variables

$$B_{1r}, B_{1\theta}, E_\phi, u_r, u_\theta, j_\phi; \nabla \cdot \mathbf{E} = \frac{1}{r \sin \theta} \frac{\partial E_\phi}{\partial \phi} \equiv 0$$

The name of the set were originally based on the magnetic field perturbation directions. One insight into these two sets of variables can be gained from the Poynting vector \mathbf{S} . The simplest MHD Ohm's law constrains the electric field \mathbf{E} to be perpendicular to the background magnetic field \mathbf{B}_0 . For the toroidal mode, \mathbf{E} lies in a meridional plane and $B_{1\phi}$ is out of plane, so \mathbf{S} is directed along the field lines. For the poloidal mode, \mathbf{S} is in a meridional plane and parallel to the wave vector \mathbf{k} , because

$$\mathbf{B}_1 = -ik \times \mathbf{E} \Rightarrow \mathbf{S} \parallel \mathbf{E} \times \mathbf{B}_1 \parallel \mathbf{E} \times (\mathbf{k} \times \mathbf{E}) = E^2 \mathbf{k} - (\mathbf{E} \cdot \mathbf{k}) \mathbf{E} \parallel \mathbf{k}$$

for a transverse EM wave ($\mathbf{E} \perp \mathbf{k}$). Thus the toroidal mode is an Alfvén mode with dispersion relation $\omega = V_A k \cos \theta$, $\mathbf{S} \parallel \mathbf{B}_0$, while poloidal mode is a fast mode with dispersion relation $\omega = V_A k$, $\mathbf{S} \parallel \mathbf{k}$.

To clarify my early misunderstanding: *a toroidal mode is not propagating in the toroidal direction!* Similarly, *a poloidal mode does not propagate in the poloidal direction*. In fact, it is more common to have a fast poloidal mode propagating in the azimuthal direction, then coupling to the Alfvén toroidal mode along the ambient magnetic field direction, then being observed on the ground. Table 16.1 from (Zong, Rankin, and Zhou 2017) summarizes properties of toroidal and poloidal mode standing ULF waves.

Table 16.1: Properties of toroidal and poloidal mode standing ULF wave

ULF mode	Magnetic field	Electric field	Wave number
Toroidal waves	B_azimuthal (B_ϕ)	E_radial (E_r)	Small wave number m
Poloidal waves	B_radial (B_r)	E_azimuthal (E_ϕ)	Large wave number m
Compressional poloidal waves	B_parallel (B_\parallel)	E_azimuthal (E_ϕ)	Large wave number m

For the axisymmetric poloidal electric field $E_\phi \sim e^{i\omega t}$, Equation 10.30 becomes (I HAVEN'T DERIVED THIS!)

$$B_0^2[\nabla^2 - (r^2 \sin^2 \theta)^{-1}]E_\phi = -\mu_0 \rho \omega^2 E_\phi$$

This is the decoupled poloidal wave equation. For a dipole field, $B_0^2 = \frac{M^2(1+3\mu^2)}{r^6}$, where M is the magnetic moment and $\mu = \cos \theta$ is the colatitude. A solution for E_ϕ is sought after a separation of the unknowns of the form

$$E_\phi = \sum_{l=1}^{\infty} f_l(r) P_l^m(\mu)$$

where P_l^m is the associated Legendre function with integer indices l and m . Why this form you may ask? It's a mathematical hypothesis based on experience and talents. With this representation the poloidal wave equation becomes

$$\sum_{l=1}^{\infty} [(1 + 3\mu^2)O_l + 2K^2]f_l P_l^m = 0 \quad (16.1)$$

where $K^2 = \frac{\mu_0 \rho}{2} \left(\frac{\omega^2 r^2}{M^2} \right)^2$ (maybe the coefficient is wrong) and O_l is the spherical Bessel operator

$$O_l = \frac{1}{r} \frac{\partial^2}{\partial r^2} r - \frac{l(l+1)}{r^2}$$

(This is almost out of my control now. If I want to fully understand this, I need to go back to math equations for physics!!!) The special property of Legendre polynomials allows us to write the solution to the wave equation as

$$\sum_{s=0,\pm 2} a_{l+s} O_{l+s} f_{l+s} + K^2 f_l = 0 \quad (16.2)$$

where the constants coefficients are

$$\begin{aligned} a_{l+2} &= \frac{3(l+2)(l+3)}{2(2l+3)(2l+5)} \\ a_l &= \frac{5l(l+1)-6}{(2l-1)(2l+3)} \\ a_{l-2} &= \frac{3(l-2)(l-1)}{2(2l-3)(2l-1)} \end{aligned}$$

From Equation 16.2, we know that radial amplitudes of opposite parity do not couple. Since only $l \geq 1$ occurs in Equation 16.1, there are two fundamental unknown radial amplitudes: one for odd l and one for even l .

We assume perfect reflection at the boundaries so that the transverse components of E , i.e. E_θ and E_ϕ , are zeros. The walls are assumed to be rigid so that $u_r = 0$. The normal component of \mathbf{B}_1 also vanishes from the governing equations.

The toroidal wave equation seems to be too complicated for me following Radoski's derivations...

From the boundary condition, the length l of the field line between two reflection point must be a multiple of half the wavelength λ , implying

$$n\lambda = 2l, \quad n = 1, 2, 3, \dots$$

From the dispersion relation, with the average Alfvén speed $\langle V_A \rangle$, one finds:

$$\omega_n = \langle V_A \rangle k = \langle V_A \rangle \frac{2\pi}{\lambda} = \frac{n\pi \langle V_A \rangle}{l}$$

16.2.2 Cylindrical Coordinates

In cylindrical coordinates (r, ϕ, z) , assuming perturbations of the form $e^{i(m\phi - \omega t)}$, we can separate the linearized equations ((Hughes 1994) I HAVEN'T DERIVED THIS!),

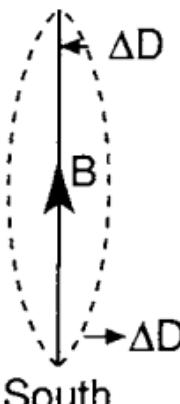
$$\left[\omega^2 \mu_0 \rho - \frac{1}{r} (\mathbf{B}_0 \cdot \nabla) r^2 (\mathbf{B}_0 \cdot \nabla) \right] \frac{u_\phi}{r} = \omega m \frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{r} \quad (16.3)$$

$$\left[\omega^2 \mu_0 \rho - r B^2 (\mathbf{B}_0 \cdot \nabla) \frac{1}{r^2 B^2} (\mathbf{B}_0 \cdot \nabla) \right] r E_\phi = i \omega B^2 (\mathbf{B}_0 \times \nabla)_\phi \left(\frac{\mathbf{B}_0 \cdot \mathbf{B}_1}{B^2} \right) \quad (16.4)$$

$$i \omega \mathbf{B}_0 \cdot \mathbf{B}_1 = \frac{1}{r} (\mathbf{B} \times \nabla)_\phi (r E_\phi) - i m B^2 \frac{u_\phi}{r}$$

TORODIAL MODE

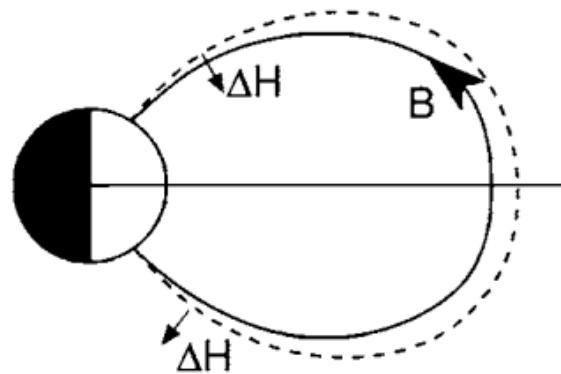
North



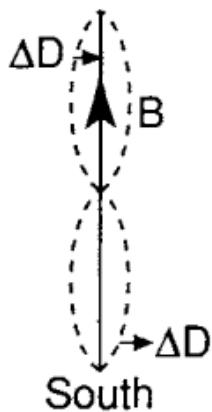
Fundamental Mode

South

POLODIAL MODE



North



Second Harmonic

South

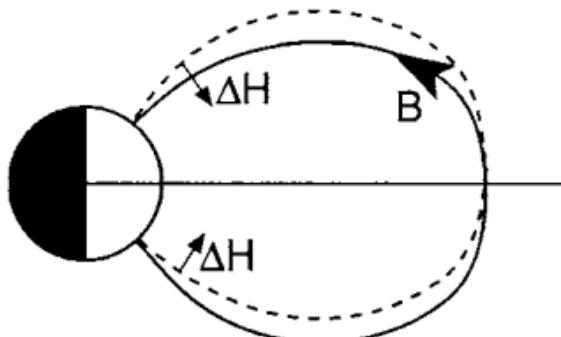


Figure 16.1: Cartoons showing the oscillation of a field line in the two lowest frequency toroidal modes (left) and poloidal modes (right). On the left the field line is drawn stretched out from north to south, and we look towards the Earth as the field line is displaced east and west. On the right, the dipolar field line is displaced within its meridian plane so all magnetic perturbations are radial. Note that for the toroidal mode, the magnetic perturbations (ΔD) have the opposite sense north and south in the fundamental and the same sense in the second harmonic. The opposite is true of the ΔH perturbation in the poloidal mode.

The first two equations are from the momentum equation. $\mathbf{B}_0 \cdot \nabla$ is the derivative along the direction of \mathbf{B} . These are wave equations coupled by the terms on the RHS which depend on $\mathbf{B}_0 \cdot \mathbf{B}_1$, the compressional part of the magnetic perturbation. The third equation is from Faraday's law and Ohm's law. It shows how $\mathbf{B}_0 \cdot \mathbf{B}_1$, E_ϕ and u_ϕ are related and closes the set. Since the transverse dispersion relation depends only on k_\parallel , it is tempting to think of the LHS of Equation 16.3 and Equation 16.4 as representing pseudo-transverse mode oscillations, and the RHS as representing the coupling due to the fast mode which has a compressional component. When $m \neq 0$, the phase variation of the toroidal mode in the ϕ direction leads to a compressional perturbation in the magnetic field. The polarizations of the toroidal and poloidal oscillations are no longer orthogonal, and thus there is coupling.

If $m = 0$ the RHS of Equation 16.3 vanishes. The LHS then describes a mode in which the electric field is purely radial and the magnetic and velocity perturbations are azimuthal. Magnetic L shells decouple and each shell oscillates azimuthally independently of each other. This is the Alfvénic *toroidal* mode. If $m \rightarrow \infty$, for the RHS to remain finite, $\mathbf{B}_0 \cdot \mathbf{B}_1 \rightarrow 0$ so the RHS of Equation 16.4 vanishes. Equation 16.4 then describes a mode in which \mathbf{E} is azimuthal and \mathbf{u} and \mathbf{b} are contained in a meridian plane. This is the compressional *poloidal* mode.

16.2.3 Cartesian Box Model

Earliest theories had cold ideal MHD equations Equation 10.30 or Equation 10.31 expressed in the spherical or cylindrical coordinates, which are not easy to solve. As we will see, the essence of FLR can be obtained in the Cartesian coordinates. We simplify the actual dayside magnetosphere, which like a compressed dipole, to something we can solve analytically. Imagine a field line with both footpoints connecting to the conducting ionosphere, we can map this curved field line into a straight line extending along z . In the x -direction, the outer boundary is the magnetopause, and the inner boundary is the reflection point (plasmapause or ionosphere). The y -direction represents the azimuthal direction, therefore at the front of the magnetosphere it is more aligned with the y -direction in GSE/GSM and at the flanks/sides it is more aligned with the x -direction. If the length in y is finite ($b < \infty$), we call it a *cavity*; if the length in y is infinite ($b \rightarrow \infty$), we call it a *waveguide*. The names cavity and waveguide are inherited from electrodynamics. This is shown in the schematic Figure 16.2 and Figure 16.3.

Along the z -direction, we impose the ionospheric boundary for closed field lines $u_x = u_y = B_{1z} = 0$. Because of the ideal MHD assumption, the electric field along the field lines $E_z = 0$. In the y -direction, we impose either a periodic boundary condition for the cavity model that leads to a quantized wavenumber k_y , or a infinite boundary. In the x direction, the boundaries are provided by large V_A gradients at both the outer boundary x_m (magnetopause) and the inner boundary x_t (plasmapause). We impose a reflective boundary $u_x = 0$ at $x = x_t, x_m$. *The inhomogeneity in x means that we can only consider Fourier components in the y and z directions.* Therefore, we seek wave modes of the form $\propto \exp[i(k_y y + k_z z - \omega t)]$ with the ansatz

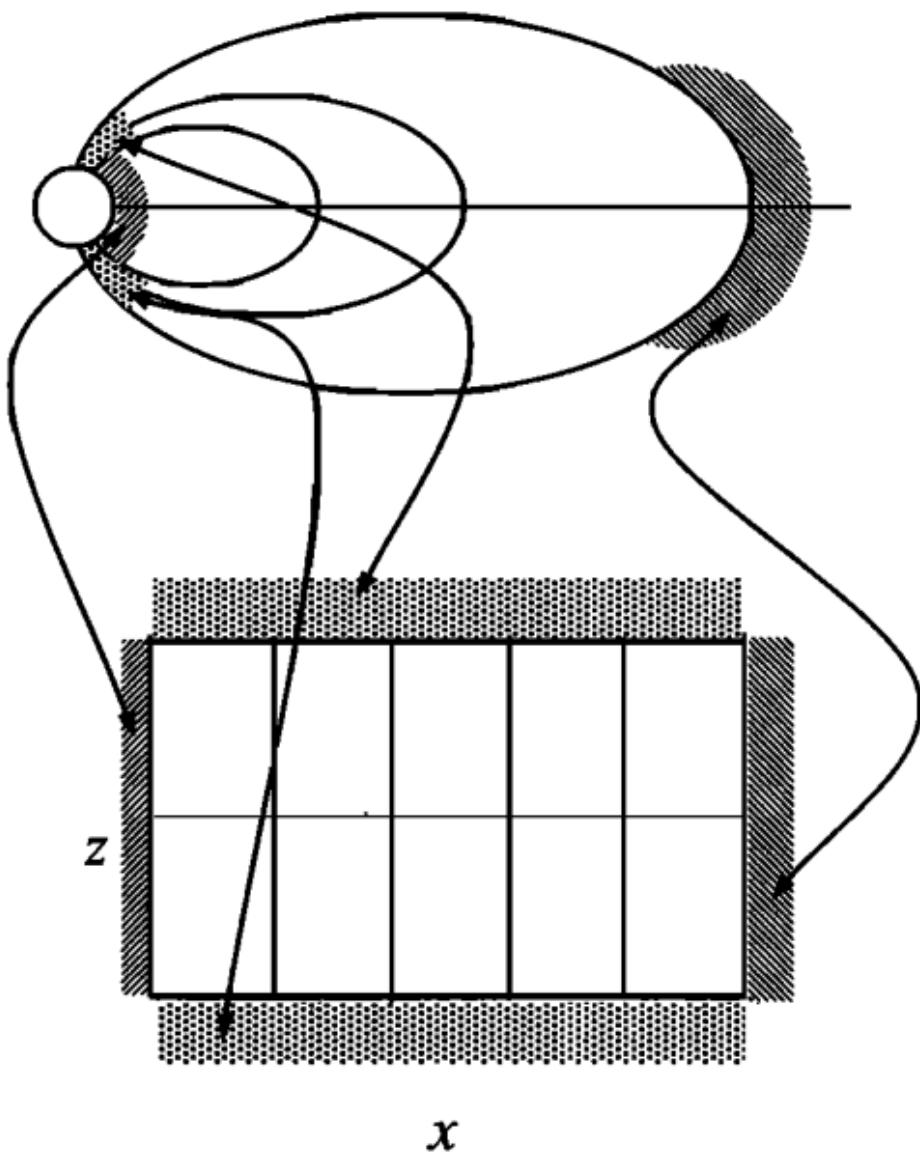


Fig. 2. Schematic of the mapping of the quasidipolar magnetosphere onto the cuboidal box used for the analysis of this paper. Boundaries corresponding to the magnetopause, the northern hemisphere and southern hemisphere ionospheres, and a low latitude near-equatorial boundary surface are illustrated.

Figure 16.2: From one of Southwood & Kivelson's paper.

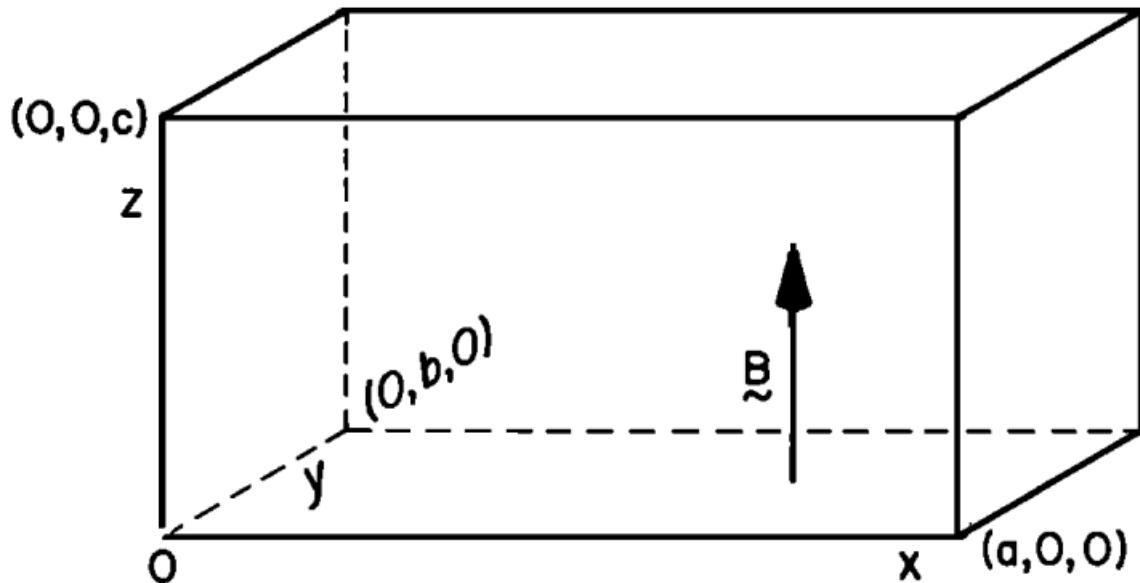


Fig. 3. Schematic representation of the magnetosphere as a rectangular box containing a uniform magnetic field parallel to the z axis. The ionospheric boundaries are planes at $z=0$ and $z=c$. The magnetopause is a plane at $x=a$ and a reflecting boundary is assumed at $x=0$. The y -direction represents the azimuthal direction in the magnetosphere. Plasma density is assumed to be either constant (the uniform model) or to vary with x (the nonuniform model).

Figure 16.3: From one of Southwood & Kivelson's paper.

$$\begin{aligned}
\mathbf{k} &= (0, k_y, k_z) \\
\mathbf{E} &= (E_x(x), E_y(x), 0) e^{i(k_y y + k_z z - \omega t)} \\
\mathbf{B}_1 &= (B_{1x}, B_{1y}, B_{1z}) e^{i(k_y y + k_z z - \omega t)} \\
\mathbf{B}_0 &= B_z \hat{z} = B \hat{z} \\
\rho &= \rho(x) \\
\mathbf{u} &= \mathbf{u}_0 + \mathbf{u}_1 = \mathbf{u}_1
\end{aligned}$$

The linearized momentum Equation 10.29 gives

$$\begin{aligned}
\rho \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) &= \mathbf{j} \times \mathbf{B} \\
\rho \frac{\partial \mathbf{u}}{\partial t} &= \frac{1}{\mu_0} (\nabla \times \mathbf{B}) \times \mathbf{B}
\end{aligned}$$

Separating the two perpendicular x and y directions and applying plane wave decomposition,

$$\begin{aligned}
-i\omega \rho \mu_0 u_x &= (\nabla \times \mathbf{B}_1)_y B_z - (\nabla \times \mathbf{B}_1)_z B_y \\
&= \left(\frac{\partial B_{1x}}{\partial z} - \frac{\partial B_{1z}}{\partial x} \right) B \\
&= -\frac{\partial}{\partial x} (BB_{1z}) + ik_z BB_{1x} \\
-i\omega \rho \mu_0 u_y &= (\nabla \times \mathbf{B}_1)_z B_x - (\nabla \times \mathbf{B}_1)_x B_z \\
&= \left(\frac{\partial B_{1y}}{\partial z} - \frac{\partial B_{1z}}{\partial y} \right) B \\
&= k_z BB_{1y} - k_y BB_{1z}
\end{aligned}$$

From the simplest form of the generalized Ohm's law $\mathbf{E} = -\mathbf{u}_1 \times \mathbf{B}_0$,

$$\begin{aligned}
E_x &= -u_y B \\
E_y &= -u_x B
\end{aligned}$$

Inserting into the linearized momentum equation, we have

$$\begin{aligned}
-i\omega \rho \mu_0 \frac{E_y}{B} &= -\frac{\partial}{\partial x} (BB_{1z}) + ik_z BB_{1x} \\
\omega \rho \mu_0 \frac{E_x}{B} &= -k_y BB_{1z} + k_z BB_{1y}
\end{aligned} \tag{16.5}$$

From linearized Faraday's law $\nabla \times \mathbf{E} = -\dot{\mathbf{B}}_1$,

$$\begin{aligned}
-\frac{\partial E_y}{\partial z} &= -\dot{B}_{1x} \\
\frac{\partial E_x}{\partial z} &= -\dot{B}_{1y} \\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\dot{B}_{1z} \\
B_{1x} &= \frac{1}{\omega}(-k_z E_y) \\
B_{1y} &= \frac{1}{\omega}(k_z E_x) \\
B_{1z} &= \frac{1}{\omega}\left(-i \frac{\partial E_y}{\partial x} - k_y E_x\right)
\end{aligned} \tag{16.6}$$

Substituting Equation 16.6 into Equation 16.5 gives

$$\begin{aligned}
-i\omega\rho\mu_0 \frac{E_y}{B} &= -\frac{\partial}{\partial x} \left[B \frac{1}{\omega} \left(-i \frac{\partial E_y}{\partial x} - k_y E_x \right) \right] + ik_z B \frac{-k_z}{\omega} E_y \\
&\quad -i(\omega^2\rho\mu_0/B^2 - k_z^2)E_y = k_y \frac{\partial E_x}{\partial x} - i \frac{\partial^2 E_y}{\partial x^2} \\
-i\omega\rho\mu_0 \frac{E_x}{B} &= k_z B \frac{k_z}{\omega} E_x - k_y B \frac{1}{\omega} \left(-i \frac{\partial E_y}{\partial x} - k_y E_x \right) \\
&\quad (\omega^2\rho\mu_0/B^2 - k_y^2 - k_z^2)E_x + ik_y \frac{\partial E_y}{\partial x} = 0
\end{aligned}$$

Let $R^2 = \mu_0\rho(x)\omega^2/B^2 = \omega^2/V_A(x)^2$, we have

$$\begin{aligned}
-i(R^2 - k_z^2)E_y &= k_y \frac{\partial E_x}{\partial x} - i \frac{\partial^2 E_y}{\partial x^2} \\
(R^2 - k_y^2 - k_z^2)E_x + ik_y \frac{\partial E_y}{\partial x} &= 0
\end{aligned} \tag{16.7}$$

Eliminating E_x from Equation 16.7 gives

$$\frac{\partial^2}{\partial x^2} E_y - k_y^2 \frac{\partial}{\partial x} R^2(x) \frac{1}{(R^2 - k_z^2)(R^2 - k_z^2 - k_y^2)} \frac{\partial E_y}{\partial x} + (R^2 - k_z^2 - k_y^2)E_y = 0 \tag{16.8}$$

There is a very important observation about polarization from Equation 16.7:

$$\frac{E_x}{E_y} = -\frac{ik_y}{R^2 - k_y^2 - k_z^2} \frac{\partial E_y}{\partial x} \frac{1}{E_y} \quad (16.9)$$

Our familiar fast MHD wave is *no longer linearly polarized in a nonuniform plasma!* A key thing on the RHS is the dependence on the sign of k_y as well as $\partial E_y / \partial x$. For fast waves propagating in opposite directions (i.e. eastward and westward in the azimuthal direction), the sense of polarization (RH, LH) will change as k_y changes sign across local noon. Right at the resonance point where $R^2 - k_z^2 - k_y^2 = 0$, we get linearly polarized Alfvén wave. $\partial E_y / \partial x$ changes sign across the resonance point, thus we also have another change in polarization detected on the ground for different L-shells (or magnetic latitudes).

Equation 16.8 can also be written as a second order differential equation of B_{1z} , the compressional component of the magnetic field (see Equation 3.9), as (Kivelson and Southwood 1985) did in proposing the idea of cavity modes:

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{\frac{\partial R^2}{\partial x}}{R^2 - k_z^2} \frac{\partial B_{1z}}{\partial x} + (R^2 - k_y^2 - k_z^2) B_{1z} = 0$$

or equivalently,

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{\omega^2 \partial V_A^{-2} / \partial x}{\omega^2 / V_A^2 - k_z^2} \frac{\partial B_{1z}}{\partial x} + \left(\frac{\omega^2}{V_A^2} - k_y^2 - k_z^2 \right) B_{1z} = 0 \quad (16.10)$$

Let us first get some intuitions about Equation 16.8 (following (Glassmeier et al. 1999), but note that there are sign errors and wrong equations in the original paper!). This equation exhibits strong singularities found in the denominator of its second term, much as first described by Tamao (1965). The following solutions are possible. If $R^2 - k_z^2 - k_y^2 > 0$, from which $R^2 - k_z^2 > 0$ follows, no singularities appear. Assuming $k_y \approx 0$, the above equation reduces to

$$\frac{\partial^2}{\partial x^2} E_y + (R^2 - k_z^2) E_y = 0$$

For a linear density profile, i.e. $R^2 = \alpha_0^2 + \alpha^2 x$, with the definition of the turning point x_t via $R^2 = k_z^2$, and the transformation $s = \alpha^{2/3}(x - x_t) > 0$,

$$\begin{aligned} \frac{\partial E_y}{\partial x} &= \frac{\partial E_y}{\partial(\alpha^{-2/3}s + x_t)} = \alpha^{2/3} \frac{\partial E_y}{\partial s} \\ \frac{\partial^2 E_y}{\partial x^2} &= \alpha^{4/3} \frac{\partial^2 E_y}{\partial s^2} \end{aligned}$$

$$\frac{\partial^2 E_y}{\partial s^2} + s E_y = 0$$

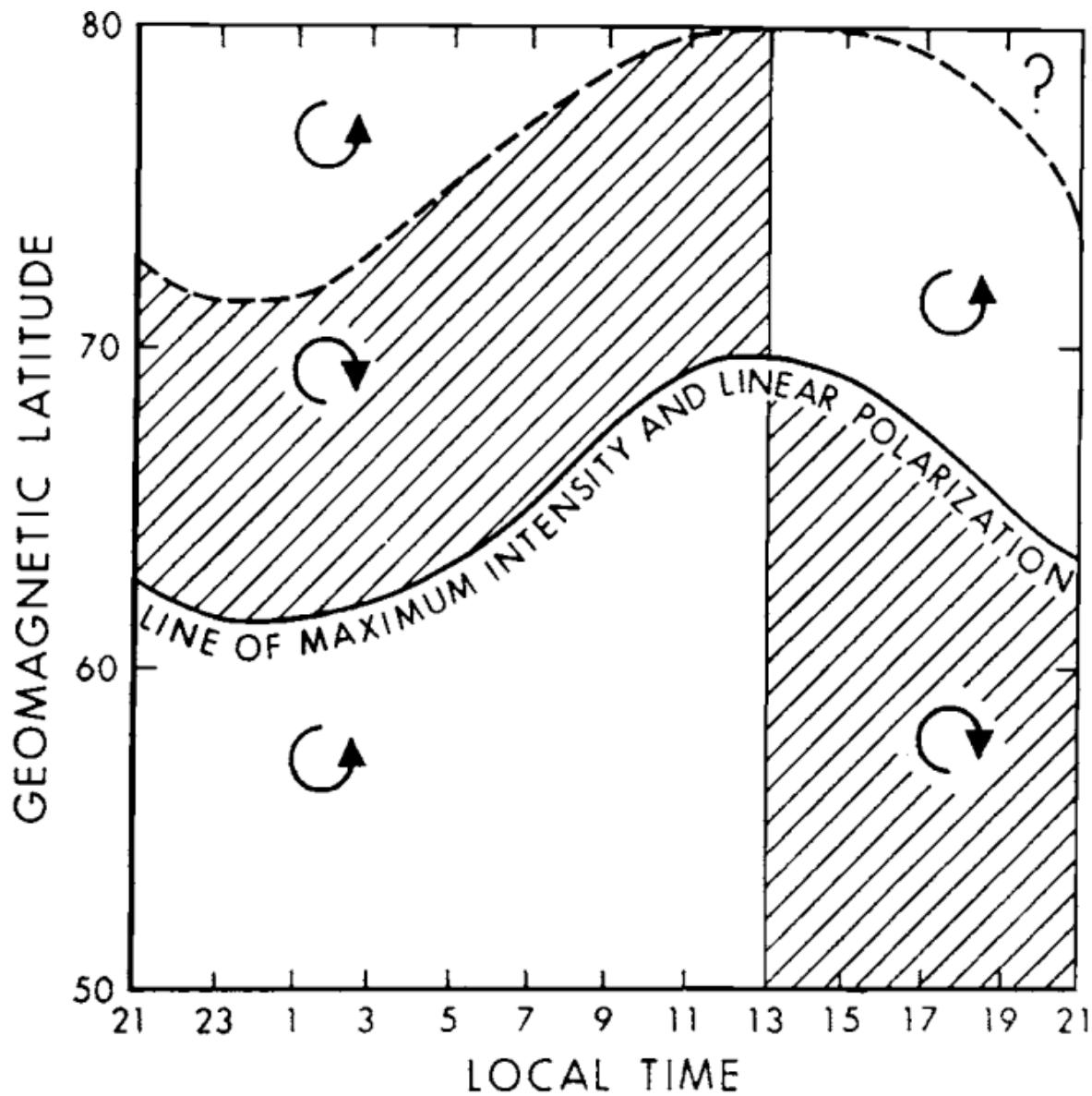


Figure 16.4: The variation of amplitude and the sense of polarization of $Pc5$'s seen at high latitudes with latitude and magnetic local time. (After [Samson +, 1974].)

This is the Airy or Stokes equation (Section 3.11.1) with the two principal solutions displayed in ?@fig-airy-flr. The solution $Bi(s)$ is unphysical as it implies unlimited growth of E_y behind the turning point at $s = 0$. Thus $E_y(s) = Ai(s)$ is the required solution. The turning point actually is the point of total reflection of the wave field. Its appearance can be understood on the following grounds. The “effective” local wavenumber in x is given by

$$k_x^2(x, \omega) = \frac{\omega^2}{V_A^2(x)} - k_y^2 - k_z^2 \quad (16.11)$$

If V_A^2 increases with x , that is with s in the Airy function plot, k_x^2 has to decrease as ω, k_y and k_z stay constant. Eventually k_x^2 may become negative, which implies an imaginary wave number k_x . At this turning point the wave will be reflected.

`KeyNotes.plot_airy_minus_x()`

Next if $R^2 - k_z^2 - k_y^2 < 0$, $R^2 - k_z^2 = 0$ may occur. Assuming again a linear density profile, defining a resonance point x_r via $R^2 = k_z^2$, and $s = x - x_r$, the electric field perturbation transforms close to the resonance point $x - x_r = 0$ into

$$\frac{\partial^2 E_y}{\partial s^2} - \frac{1}{s} \frac{s_t}{s - s_t} \frac{\partial E_y}{\partial s} + \alpha^2(s - s_t) E_y = 0 \quad (16.12)$$

where $s_t = k_y^2/\alpha^2$.

At the resonance point its solution exhibits a clear singularity with unlimited growth of E_y . In front of the turning point $s > s_t$ the solution is similar to an Airy function, while behind it singular behaviour is observed at the resonance point $s = 0$. I do not show the solution BECAUSE I DON'T KNOW HOW TO WRITE IT DOWN! But let us now follow (Kivelson and Southwood 1985) and check the equivalent perturbation Equation 16.10 for B_{1z} .

When we also assume linear mass density variation with x in Equation 16.10 such that

$$\begin{aligned} R^2 - k_z^2 &= \alpha^2(x - x_r) \\ R^2 - k_z^2 - k_y^2 &= \alpha^2(x - x_t) \end{aligned}$$

where $x_r = k_z^2/\alpha^2$, $x_t = (k_z^2 + k_y^2)/\alpha^2 = x_r + k_y^2/\alpha^2$, we have

$$\frac{\partial^2 B_{1z}}{\partial x^2} - \frac{1}{x - x_r} \frac{\partial B_{1z}}{\partial x} + \alpha^2(x - x_t) B_{1z} = 0 \quad (16.13)$$

The solution of Equation 16.13 is discussed by (1986). The singular point where $x = x_r$ and $R^2 = k_z^2$ corresponds to the *resonance point* where the wave frequency matches the Alfvén mode frequency. The point where $x = x_t$ and $R^2 = k_y^2 + k_z^2$ is the *turning point* of the equation.

When $x > x_t$, any wave described by the equation oscillates with x ; when $x < x_t$ independent solutions grow or evanesce with x on each side of the resonance singularity x_r .

Series solutions valid in the neighborhood of $x = x_r$ and $x = x_t$ have been derived by a variety of authors. The Frobenius method of solution of partial differential equations can be used to find solutions valid near x_r , where the analytic solution is of the form

$$R_t(x) = k_y^2(x - x_r)^2 + \mathcal{O}[(x - x_r)^4] \quad (16.14)$$

and the singular solution is of the form

$$s_t(x) = 1 + \frac{1}{2}R_t(x) \ln(x_r - x) + \mathcal{O}[(x - x_r)^2] \quad (16.15)$$

For $k_y^2 = \alpha^{4/3}$, the two solutions for B_{1z} is shown in Figure 16.5. On the far right side of the plot the functions are close to spatial quadrature and have zeros whose spacing decreases with increasing x . In this regime, as we will mention later, the WKB approximation is good and the amplitude and phase are proportional to $s^{1/4}$ and $s^{3/2}$, respectively. The amplitudes have been plotted just such that an incoming wave from large x would be represented by $R_t(x) + iS_t(x)$ if a time variation $e^{-i\omega t}$ is specified. The corresponding mode propagating toward large x would be $R_t(x) - iS_t(x)$. Thus $R_t(x), S_t(x)$ represent standing wave forms. However, the presence of the resonance (at the far left of the plots) precludes there being perfect reflection or perfect standing wave solutions.

`@sco KeyNotes.plot_b1z_ode()`

The following physical interpretation is tempting. The MHD wave propagating into the magnetosphere is a fast mode wave generated by, e.g. plasma instabilities at the magnetopause. Eventually the wave reaches the turning point where reflection occurs. If conditions are favourable, that is if a resonance point occurs, part of the fast mode wave energy can tunnel into the resonance, where coupling from the fast mode disturbance into an Alfvén mode perturbation takes place. The resonance point *always* lies beyond the turning point, but energy tunnels to the resonance point and the reflection is found to be less than perfect. This scenario is schematically shown in Figure 16.6, where the reflection or turning point is assumed to coincide with the magnetopause. The point of maximum coupling or “resonant mode coupling” is given by $\omega = k_{\parallel}V_A(x)$, where ω is the fast mode wave frequency, k_{\parallel} is the field-aligned component of the fast mode wave vector and $V_A(x)$ is the local Alfvén velocity. This is also the Alfvén dispersion relation.

Equation 16.10 has a long history that appeared first in the context of a radio wave obliquely incident on a region of increasing electron density, all the way back to 1951, as mentioned in (Kivelson and Southwood 1986). It also appears in calculating the absorption of laser energy by a plasma and heating a plasma by radio wave injection to excite internal Alfvén waves. Energy is absorbed at the location in the plasma corresponding to the singular point of the governing

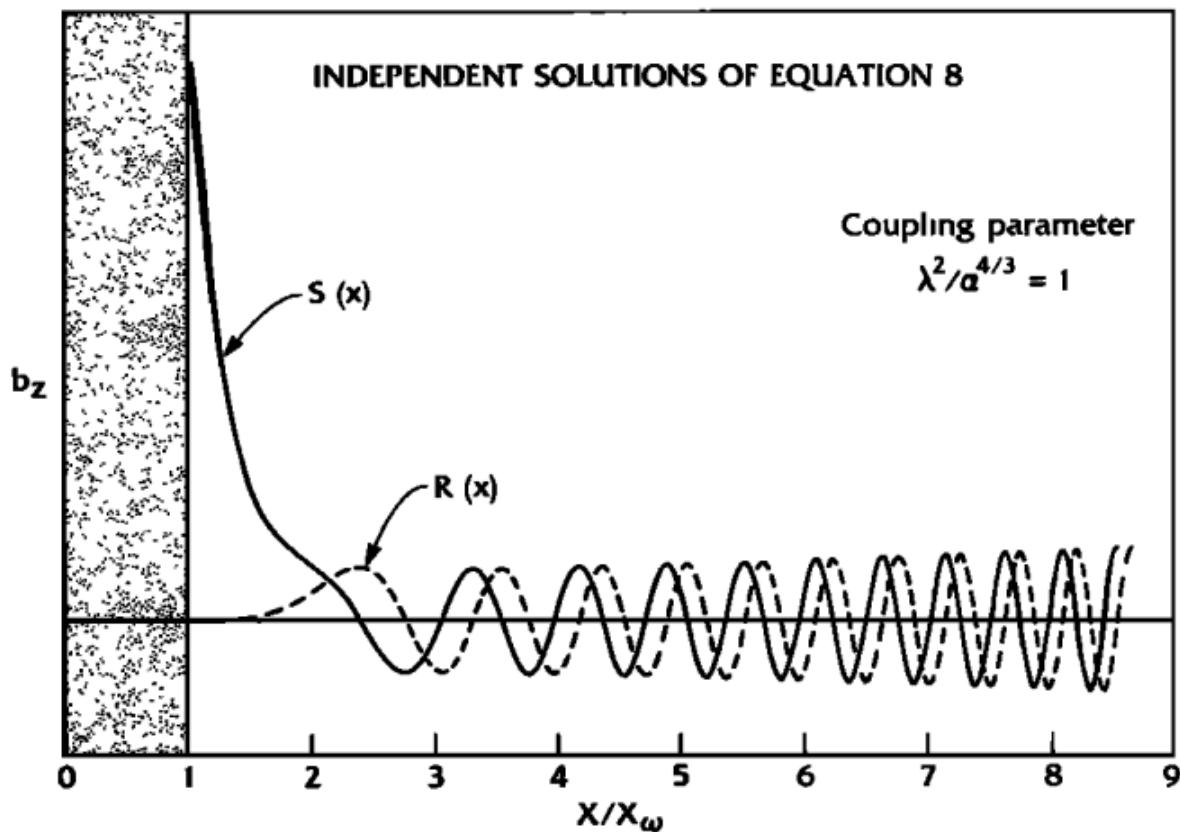


Figure 16.5: Two independent solutions of Equation 16.13 for $x > x_r$. The particular pair illustrated are solutions for the case $\lambda^2/\alpha^{4/3} = 1$ for which $x_t/x_r = 2$. The dashed curve is the solution that is analytic at the singular point (cf. Equation 16.14). The solution plotted as a solid curve is finite at the singularity with infinite derivatives as evident from Equation 16.15. Amplitudes have been selected to match at large x where the oscillating solutions are in spatial quadrature.

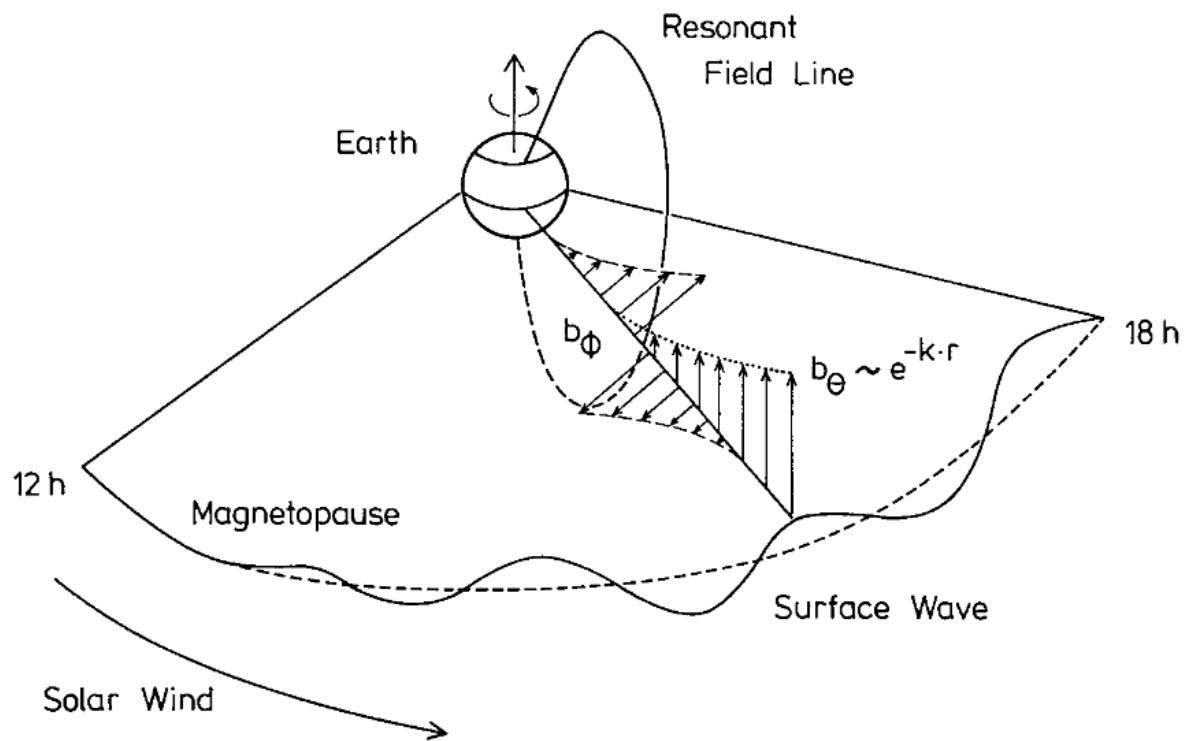


Figure 16.6: Schematic view of the field line resonance, driven by unstable surface wave.

differential ?@eq-flr-wave-bz or its equivalent. Without the introduction of additional effects such as dissipation or dispersion near resonance, the singularity in the differential equation implies that the amplitude of the Alfvén wave grows secularly.

Some studies (e.g. Wright (1994)) prefer to use plasma displacement ξ instead of \mathbf{u} , where $\mathbf{u} = \dot{\xi}$, and perturbed magnetic field component B_{1z} to describe the model:

$$\begin{aligned}\frac{1}{V_A^2} \frac{\partial^2 \xi_x}{\partial t^2} - \frac{\partial \xi_x}{\partial z^2} &= -\frac{1}{B_0} \frac{\partial B_{1z}}{\partial x} \\ \frac{1}{V_A^2} \frac{\partial^2 \xi_y}{\partial t^2} - \frac{\partial \xi_y}{\partial z^2} &= -\frac{1}{B_0} \frac{\partial B_{1z}}{\partial y} \\ B_{1z} &= -B_0 \left(\frac{\partial \xi_x}{\partial x} + \frac{\partial \xi_y}{\partial y} \right)\end{aligned}\tag{16.16}$$

This equation set is equivalent to Equation 10.31. The first two equations are the x and y component of the momentum equation, and the last equation is the magnetic field advection equation (Faraday's law + Ohm's law). Recall that B_{1z} is the compressional perturbation to the ambient magnetic field, and ξ_y corresponds to u_y which is the toroidal Alfvénic perturbation. Then on the RHS of the second equation, it is the azimuthal gradient in the fast mode $\partial B_{1z}/\partial y$ that drives the response in the azimuthal place displacement ξ_y ; the LHS is a simple harmonic oscillator and the RHS is the forcing term. Since we are seeking for solutions of the form $e^{i(k_y y + k_z z - \omega t)}$, we can see that:

- if $k_y = 0$ for the fast mode, then there are no azimuthal gradients to drive FLRs and hence you get no wave coupling;
- if $k_y \rightarrow \infty$, then $\xi_y \rightarrow 0$ and ξ_x describes a decoupled poloidal Alfvén wave.
- if $k_y \neq 0$ or ∞ , the wave modes are coupled together and energy initially in the fast mode may mode convert to Alfvén wave energy on localized field lines.

Uniform density and background magnetic field

Let $\rho = \text{const.}$, $V_A = \text{const.}$. From Equation 16.16,

$$\begin{aligned}\ddot{B}_{1z} &= -B_0 \left(\frac{\partial \ddot{\xi}_x}{\partial x} + \frac{\partial \ddot{\xi}_y}{\partial y} \right) = V_A^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) B_{1z} \\ &= V_A^2 \nabla^2 B_{1z}\end{aligned}$$

Consider normal modes of waveguides of the form $e^{\mathbf{k} \cdot \mathbf{r} - i\omega t}$, where $\mathbf{k} = (k_x, k_y, k_z)$. The dispersion relation for the fast mode is then

$$\omega^2 = V_A^2 (k_x^2 + k_y^2 + k_z^2)\tag{16.17}$$

In an infinite medium this relation gives nondispersive propagation. Since we are interested primarily in propagation along the waveguide (i.e., in the y direction) we shall define v_p and v_g to be the phase and group velocities along \hat{y} . Employing Equation 16.17 gives

$$v_p = \frac{\omega}{k_y} = \frac{k}{k_y} V_A$$

$$v_g = \frac{\partial \omega}{\partial k_y} = \frac{k_y}{k} V_A$$

which yields the familiar waveguide relation

$$v_p v_g = V_A^2 \quad (16.18)$$

The shear Alfvén wave has the same velocity parallel to the field line (because it is cold). However, in a waveguide the boundary conditions in the x and z directions restrict the choice of wavenumbers and introduce dispersion. Suppose that the boundaries in x are perfectly reflecting (e.g. $\xi_x = \frac{\partial B_{1z}}{\partial y} = 0$) as are those in z ($\xi_x = \xi_y = B_{1z} = 0$) which represent the ionospheric boundary for closed field lines. Then

$$k_x = \pm n\pi/a, \quad n = 1, 2, 3, \dots \quad (16.19)$$

$$k_z = \pm m\pi/c, \quad m = 1, 2, 3, \dots$$

where a is the box length in x and c is the box length in z (i.e. length of the field lines).

Given values of k_x and k_z , we may use Equation 16.17 to find k_y as a function of ω ,

$$k_y^2 = \frac{\omega^2}{V_A^2} - k_x^2 - k_z^2$$

If k_y is real then the mode may propagate along the guide. However, if we believe that the boundary at $x = x_m = a$ is not a reflector but is driven, it would be appropriate to impose a wavenumber k_y along the outer boundary and solve for k_x given k_z . This is the case when the magnetopause is driven by the Kelvin-Helmholtz instability (i.e. k_y given by the surface waves), which was proposed in the 1970s to be the driver of field line resonances. In this situation k_x is found to be imaginary, and the mode is evanescent (exponentially decaying) in x .

Nonuniform density and magnetic field

The simplest assumptions in this case would be $\rho = \rho(x)$, $\mathbf{B}_0 = B_0 \hat{z}$, and that the Alfvén speed V_A is monotonically decreasing with x , $\partial V_A / \partial x < 0$. Then again we get Equation 16.10. Solving this equation for given boundary conditions in x yields a set of orthogonal eigenfunctions $B_{1z}(x)$ and eigenfrequencies.

First of all, we shall notice that a singularity/resonance occurs at x_r where

$$\omega^2 = k_z^2 V_A^2(x_r) \quad (16.20)$$

Across x_r , there will be a 180° phase shift for the perturbed terms (Equation 16.9). Energy is been absorbed at the resonance point from fast mode to Alfvén mode. Dissipation is required, otherwise the amplitude of the Alfvén wave grows secularly. Ionospheric dissipation near the resonant field line is likely to be one important process limiting the growth of the resonance amplitude.

Secondly, there is a turning point at x_t defined via $k_x^2 = 0$ (fixed k_y, k_z, ω):

$$\omega^2 = (k_y^2 + k_z^2)V_A^2(x_t) \quad (16.21)$$

On the right of x_t , the positive exponent represents a wave propagating in the positive x direction towards the magnetopause, and the negative exponent represents one propagating away from the magnetopause. This is like an imperfect standing wave solution. Across x_t on the left, the oscillating solutions convert to a decaying solution and an exponentially growing solution (Section 3.11.1).

If the Alfvén speed V_A is monotonically decreasing with x , from Equation 16.20 and Equation 16.21 we always have $0 < x_r < x_t < x_m$. In the low Alfvén speed region $[x_t, x_m]$ the fast mode may propagate, while in the high Alfvén speed region $[0, x_t]$ the fast mode is evanescent. Thus the resonant singularity x_r is in the evanescent tail of the fast mode.

The first-order derivative term in Equation 16.10 dominate near $x = x_r$, but is small in the propagating interval $[x_t, x_m]$. This is particularly true in the WKB limit (Section 3.11), where we consider waves of short wavelength in x ; i.e. second-order derivative term is much larger than the first-order derivative term. In the lowest order WKB analysis we may neglect the first-order derivatives and solve

$$\frac{\partial^2 B_{1z}}{\partial x^2} + \left(\frac{\omega^2}{V_A^2} - k_y^2 - k_z^2 \right) B_{1z} = 0 \quad (16.22)$$

and the WKB solution is given in the form

$$B_{1z}(x) = Ak_x^{-1/2}e^{i\int k_x dx} + Bk_x^{-1/2}e^{-i\int k_x dx}$$

where A and B are constants, and k_x is the “effective” wavenumber defined in Equation 16.11, an explicit function of x .

The Bohr-Sommerfeld (or phase integral, Section 3.11.4) condition which the quasi-standing wave must satisfy between x_t and x_m is

$$\int_{x_t}^{x_m} k_x(x, \omega_n) dx = (n + \alpha)\pi \quad n = 1, 2, 3\dots \quad (16.23)$$

where $x_m = a$ is the location of the outer boundary in x and the phase factor α is determined by the boundary conditions in x . For a perfect reflector at a ($\xi_x = 0$) and evanescent decay at small x , we find $\alpha = -1/4$. If there is a resonance in the waveguide, it will modify α ; however, since the resonance is in the evanescent tail there will be no leading order change in α .

Equation 16.23 is an integral relation for the eigenfrequencies of the fast mode: we specify k_y , k_z , and n (the mode number in x), then find the n -th eigenfrequency ω_n as that frequency for which the criterion is met. Once we compute ω_n , we can then compute the locations of the excited field line x_r from Equation 16.20. These are the most observable discrete modes.

Driven FLR vs Cavity Modes vs Waveguide Modes

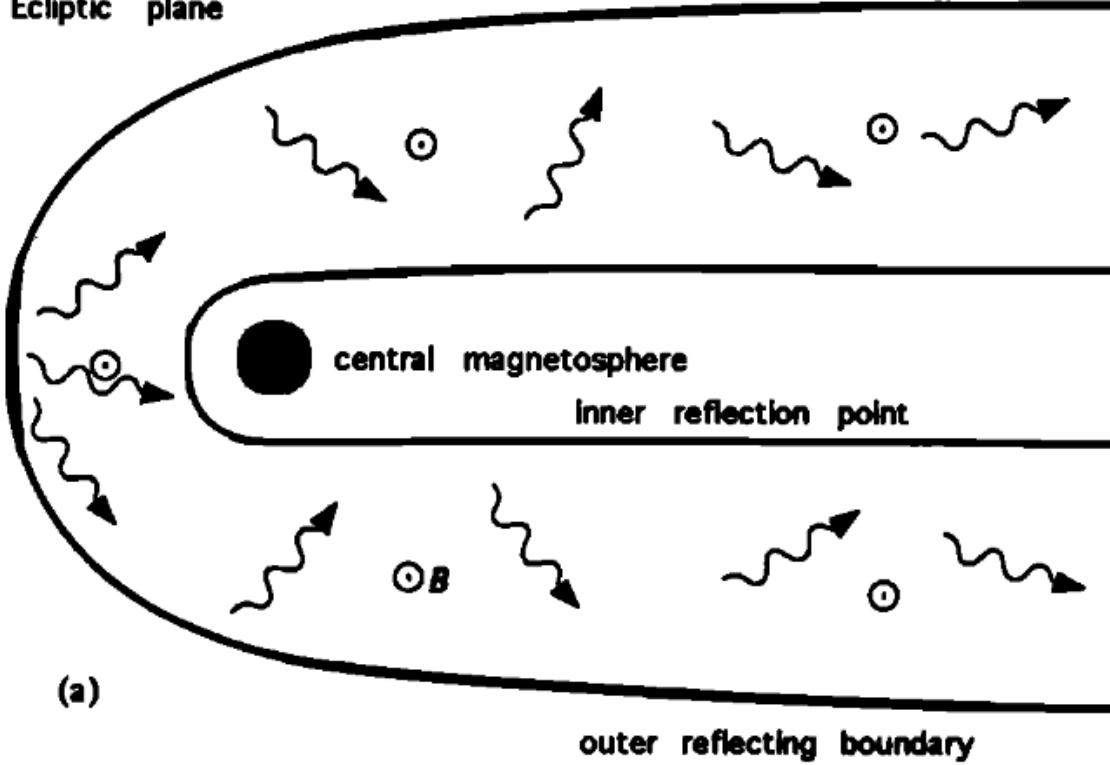
The surface wave driven FLR seems reasonable. However, the biggest problem for this hypothesis to be valid is that the mapped phase speeds to the magnetopause were too high compared to in-situ measurements. In addition, in spite of numerous magnetopause crossings by ISEE 1 and 2, regular oscillations of the magnetopause corresponding to the Kelvin-Helmholtz waves were not seen. These two facts indicated that we should seek for a new explanation for the driver.

In the 1980s, cavity modes and waveguide modes were proposed to remove the constraint on driven FLR. The analogy of a magnetosphere to a cavity came before the analogy to a waveguide. In a cavity model the azimuthal direction ϕ or the y direction in the box geometry is finite, and the wavenumber k_y is determined by the given scale size b . In the axisymmetric cylindrical coordinates (r, ϕ, z) , where ϕ is the azimuthal angle and z is the coordinate along the field lines, the field has the form $\sim e^{im\phi}$, $m = 0, 1, 2$.¹ This is a weak point of the theory because the magnetosphere is far from axisymmetric and there are no other obvious boundaries to define the cavity. If the system is similar to a waveguide, then k_y can have a continuum of values, and the waveguide allows propagation over a wide band of frequencies.

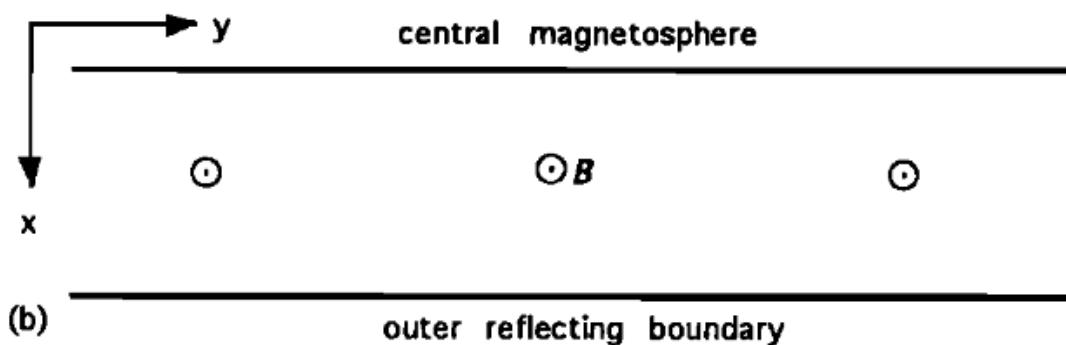
If instead we use a waveguide model, the azimuthal direction (or y -direction) will have no imposed boundary condition. Theoretically a continuum of wave modes k_y can excite FLR, and the discrete resonance frequencies are determined by the Bohr-Sommerfeld condition Equation 16.23 where harmonics of toroidal and poloidal are given by k_z . The discrete frequency spectrum of the fast waveguide modes is suitable for driving a series of Alfvén resonance; however, it is not obvious that the continuous frequency spectrum of the modes will be able to drive resonances at discrete frequencies. A key observation to make waveguide mode more valid is that *only small k_y contributes to the resonance*. ((Walker et al. 1992), [Wright 1992], (Wright 1994)). If k_y is not fixed (but k_z is fixed), then Equation 16.23 is the dispersion relation for the waveguide expressing ω as a function of k_y . The Alfvén speed is a strong function of the L shell, typically $\sim L^{-3}$. Thus, if the turning point x_t is reasonably deep within the magnetosphere, the range over which $\omega^2/V_A^2 \gg k_y^2 + k_z^2$ contributes most significantly to the integral Equation 16.23. The result is that over a wide range of wavelengths the frequency is very insensitive to k_y . Only when k_y is quite large this is not true, and then the turning point is near the magnetopause, and the wave does not penetrate very deeply into the magnetosphere. Each waveguide mode has a phase velocity ω/k_y and a group velocity $d\omega/dk_y$. Since Equation 16.23 defines ω as a function of k_y , we can substitute Equation 16.11

¹ $E(\phi = 0) = E(\phi = 2\pi)$

Ecliptic plane



(a)



(b)

Fig. 1. (a) The magnetospheric waveguide viewed in the ecliptic plane. Fast mode waves (depicted as wiggly arrows) enter the waveguide and subsequently suffer reflection from an outer boundary (perhaps the magnetopause or the bow shock) and an inner reflection point (the plasmapause or a "turning point"). (b) Half of the waveguide (from noon, around the dusk flank and into the magnetotail) is approximated as an infinitely long waveguide containing a uniform magnetic field and plasma whose density depends on x and y .

Figure 16.7: From Wright 1992.

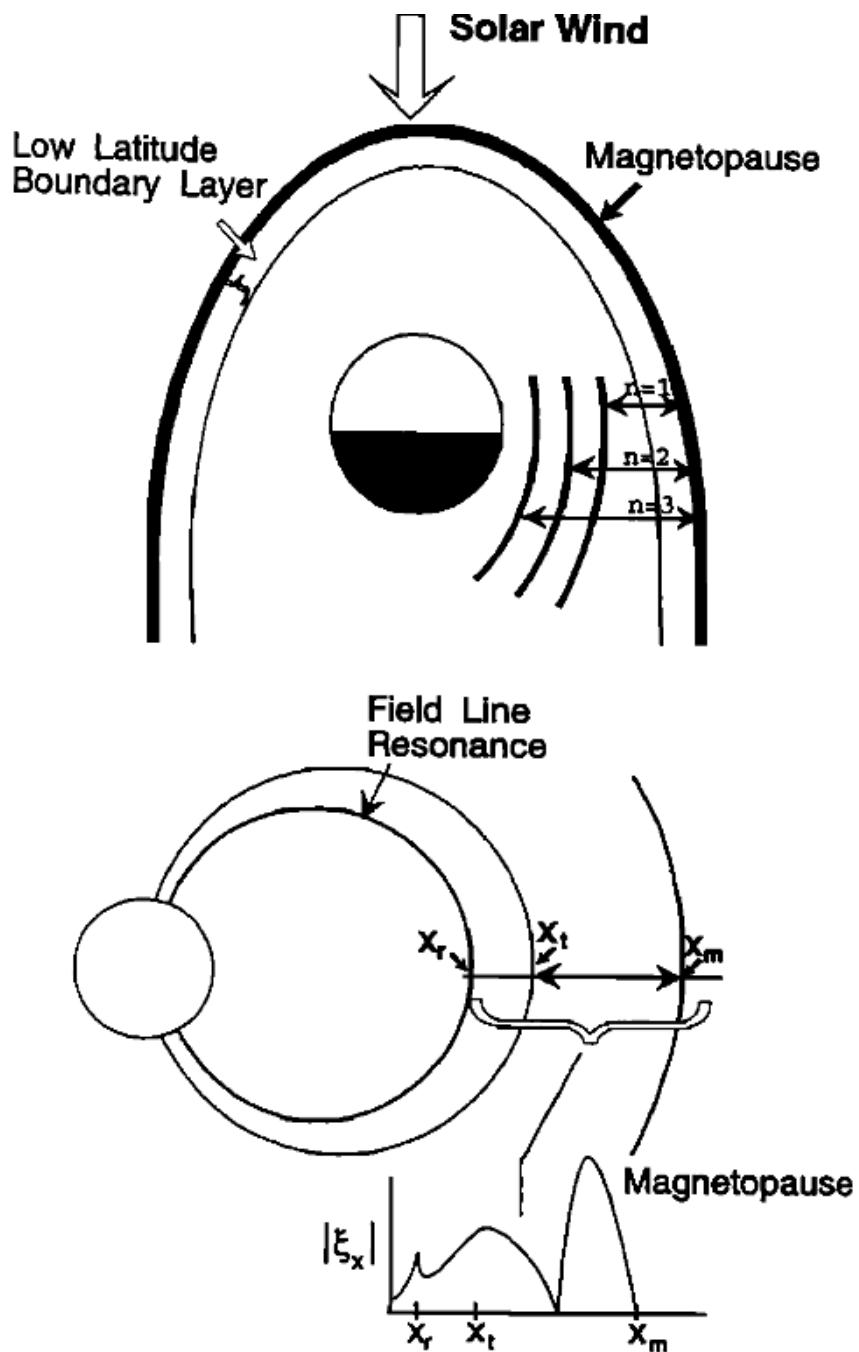


Fig. 3. The waveguide model we have used for our calculations. Top: The turning points in the equatorial plane. The first three eigenmodes are shown. Bottom: A cut in the local dawn meridian. A schematic of the radial displacement for $n=2$ is given in the insert at the bottom.

Figure 16.8: From (Samson et al. 1992).

into Equation 16.23 and differentiate with respect to k_y (assuming α is a constant):

$$\begin{aligned} \frac{\partial}{\partial k_y} \int_{x_t}^{x_m} \left(\frac{\omega^2}{V_A(x)^2} - k_y^2 - k_z^2 \right)^{1/2} dx &= 0 \\ \int_{x_t}^{x_m} \frac{1}{2} k_x^{-1} \left(\frac{1}{V_A^2} \frac{\partial \omega^2}{\partial k_y} - 2k_y \right) dx &= 0 \\ k_y \frac{\partial \omega}{\partial k_y} \int_{x_t}^{x_m} k_x^{-1} V_A^{-2} dx &= \int_{x_t}^{x_m} k_x^{-1} dx \\ v_p v_g \equiv \frac{\omega}{k_y} \frac{d\omega}{dk_y} &= \frac{\int_{x_t}^{x_m} k_x^{-1} dx}{\int_{x_t}^{x_m} V_A^{-2} k_x^{-1} dx} \end{aligned}$$

The RHS is the reciprocal of the weighted mean of V_A^{-2} with weighting function k_x^{-1} . Thus we can write

$$v_p v_g = \langle V_A^{-2} \rangle^{-1} \quad (16.24)$$

In a uniform waveguide filled with plasma for which the Alfvén speed was V_A the well-known waveguide relation Equation 16.18 would be recovered. In this more complicated case where the Alfvén speed is not constant throughout the waveguide the RHS of Equation 16.24 is an appropriately weight mean of V_A^{-2} . In general the phase velocity v_p is much larger than the mean Alfvén velocity in the guide, and the group velocity v_g is much smaller. This is also consistent with the weak dependence of ω on k_y such that v_g is small. The consequence is that even if the ends of the guide are open, the time taken for energy to be propagated along it is long compared with the period of the oscillations.

Therefore, the waveguide model proposes the following explanation. Disturbances in the solar wind move along the magnetopause, producing a disturbance on the boundary which is propagated tailward with speed v_w . Generally $v_w \gg V_A$. Such a disturbance is likely to excite waveguide modes with phase velocities equal to v_w . Because the frequency of the modes is very insensitive to k_y , the same frequencies (determined by eigenmodes in x and z) are excited no matter what the velocity. The disturbance moving with the wave can feed energy continuously into it. This then requires

$$k_y = \frac{\omega}{v_w}$$

i.e., the azimuthal wavenumber is determined by the velocity of the source rather than by a boundary condition on y . Thus we have a discrete value of wavenumber and a discrete fast mode frequency spectrum, without relying on the resonance modes from the cavity model.

Such a disturbance could arise from more than one type of source. If it arose from an impulse in the solar wind, then one would expect the waveguide response to be a ringing at the natural frequencies of its modes with growth on the time scale required to establish the mode and decay on the time scale for leakage to the resonance and loss in the ionosphere or elsewhere.

Alternatively the waveguide modes could be excited by the Kelvin-Helmholtz instability. In this case it could be expected that the waveguide modes would all be excited continuously and would be driven so long as the solar wind velocity and the characteristics of the magnetopause were maintained.

16.3 Energy Consideration

???

The energy that is carried into the magnetosphere across the background field by the non-guided fast mode is accumulated in the plane of resonant mode coupling (i.e. the $y-z$ plane through x_r in Figure 16.3) in the form of the guided Alfvén wave. It is this localized accumulation of energy due to resonant mode coupling between a non-guided mode and a guided mode that constitutes a field line resonance. This wave energy accumulation can be described by

$$\frac{\partial w}{\partial t} + \nabla \cdot \mathbf{S} = -h$$

where w is the wave energy density, \mathbf{S} is the Poynting flux, and h a dissipation term, describing energy loss due to e.g. ionospheric Joule heating. As the background parameters only vary in the radial direction, this equation reduces to

$$\frac{\partial w}{\partial t} = -\frac{d}{dr} S_x - h$$

Integration across the width of the coupling region in a radial direction leads to the following rate equation:

$$\frac{\partial W}{\partial t} = c_e S_{ng} - S_g^{\text{off}} - H$$

where W is the energy per area that is being accumulated in the coupling, c_e is a coupling efficiency, S_{ng} the incoming Poynting flux of the non-guided mode, and S_g^{off} the “off-angle” component of the Poynting flux of that mode to which the non-guided mode couples. Including this term S_g^{off} allows us to consider energy losses due to coupling to not strictly guided modes. A finite off-angle component of the coupled wave mode would render the energy accumulation less efficient or may even inhibit the build-up of a resonance. Off-angle components may arise if the transverse scale of the coupled wave become small enough for finite ion gyroradius or finite electron inertia becoming important. In this case the coupled mode is a *kinetic Alfvén mode*. The parameter c_e denotes the coupling efficiency, that is the fraction of energy of the non-guided mode that is converted into the guided mode. Finally, H gives the dissipative losses, integrated in the x direction.

It is now instructive to evaluate the rate Equation for the ideal MHD regime. As the Alfvén mode is a strictly guided mode its Poynting flux is directed exactly parallel to the background

magnetic field. In other words, $S_h^{\text{off}} = 0$, and Equation reads

$$\frac{\partial W}{\partial t} = c_e S_{\text{fast}}$$

where $S_{\text{fast}} = S_{ng}$ is the Poynting flux of the non-guided MHD mode. Thus, the energy density in the coupling region is continuously increasing as there is no outward transport of energy to balance the incoming Poynting flux of the fast mode. However, ionospheric Joule heating provides a significant dissipation mechanism with H limiting the energy density.

Resonant mode coupling is only a *necessary* condition for field line resonances to occur. A *sufficient* condition is critical coupling to a strictly guided mode and absence of dissipation to a degree that there is enough energy for resonance.

16.4 Non-MHD Effects

As been shown in previous sections, some features of the wave absorption can be simply understood using a cold fluid approach. In the cold plasma description, the Alfvén resonance is a singularity of the compressional wave Equation 16.10. The compressional wave equation may be solved by analytic continuation around the Alfvén resonance singularity in which case there is a loss of Poynting flux at the resonance and wave energy is converted to an Alfvén wave. Alternatively, a kinetic description of the mode conversion process involves two propagating modes, i.e. the compressional wave and the dispersive kinetic Alfvén wave, which couple near the resonance location with energy converted from the former to the latter. If we then consider a hot plasma, as the temperature increases the *Doppler shift* will spread the resonance leading to a more extended coupling region. Thus the inclusion of finite-temperature and E_{\parallel} (in KAWs) can remove the singularity in the cold fluid model, and the energy absorbed by the Alfvén resonance shows up as electron heating near the resonance.

As a recap, in a plasma with a 1-D inhomogeneity in the background plasma parameters along the direction perpendicular to the magnetic field, the linearized MHD equations can be expressed in terms of $\delta p_1 = B_0 \delta B_{\parallel} + \delta p$ and δB_{\perp} by a coupled system of equations

$$-ik_{\parallel} \left(1 - \frac{k_{\parallel}^2 c_s^2}{\omega^2} \right) \delta p_1 = \left(1 + \frac{c_s^2}{v_A^2} - \frac{k_{\parallel}^2 c_s^2}{\omega^2} \right) B_0^2 \nabla_{\perp} \cdot \left(\frac{\delta \mathbf{B}_{\perp}}{B_0} \right) \quad (16.25)$$

and

$$(\omega^2 - k_{\parallel}^2 V_A^2) B_0 \delta \mathbf{B}_{\perp} = ik_{\parallel} v_A^2 \nabla_{\perp} \delta p_1 \quad (16.26)$$

In higher frequency cases, the Alfvén resonance condition Equation 16.20 is modified as

$$\omega^2 = k_{\parallel}^2 V_A^2 (1 - \omega^2 / \omega_{ci}^2)^2 \quad (16.27)$$

because of the finite ion Larmor radius effects [Stix, 1992].

The Alfvén resonance singularity can be removed by including non-MHD effects such as electron inertia or ion Larmor radius corrections.²

Using the two-fluid theory, Equation 16.26 can be replaced by

$$\left[\omega^2(1 - \lambda_e^2 \nabla_{\perp}^2) - k_{\parallel}^2 v_A^2 (1 - r_{Ls}^2 \nabla_{\perp}^2) \right] B_0 \delta B_{\perp} = ik_{\parallel} v_A^2 \nabla_{\perp} \delta p_1 \quad (16.28)$$

which includes electron inertial effects with $\lambda_e = c/\omega_{pe}$ being the plasma skin depth or electron inertial length, and electron pressure effects on the *ion acoustic Larmor radius scale* $r_{Ls}^2 = k_B T_E / m_i \Omega_i^2$. The relative importance of electron pressure versus electron inertial effects is determined by r_{Ls}/λ_e : the pressure effect is dominant if $r_{Ls}/\lambda_e = nk_B T_e / \mu_0 B^2 \sqrt{m_i/m_e} > 1$, which occurs when $\beta_e/2 > \sqrt{m_e/m_i} = 0.02$. Near the magnetopause, electron inertial effects are typically not important except at locations where $k_{\parallel} \rightarrow 0$ such as in a sheared field or reconnection geometry.

Thus what's usually been considered to remove the singularity is the ion kinetic effects, where we end up coupling fast waves with kinetic Alfvén waves of the dispersion relation Equation 10.81. Using the kinetic theory Equation 16.26 can be modified as

$$\left[\omega^2 \hat{K}_1 - k_{\parallel}^2 v_A^2 \left(1 + \frac{T_e}{T_i} \hat{K}_2 \right) \right] B_0 \delta B_{\perp} = ik_{\parallel} v_A^2 \nabla_{\perp} \delta p_1 \quad (16.29)$$

where

$$\hat{K}_1 = ???$$

$$\hat{K}_2 = ???$$

It can be shown that under certain approximations (Pade approximation and $\zeta_e \ll 1$), the left hand side of Equation 16.29 becomes the kinetic Alfvén wave

$$k_{\parallel}^2 v_A^2 \left[1 + \left(1 + \frac{T_e}{T_i} \right) k_{\perp}^2 r_{iL}^2 \right] = ik_{\parallel} v_A^2 \nabla_{\perp} \delta p_1 \quad (16.30)$$

Ion Landau damping may be important when $\zeta_i = \omega/\sqrt{2}k_{\parallel}v_{i,th} \sim 1$ (I can feel this from Bellan, but how exactly???).

From the cold MHD theory the nonzero azimuthal/y wavenumber is necessary for wave coupling (Equation 16.16). However, finite-frequency effects can remove the decoupling when $k_y = 0$, as originally noted by (1979). The energy absorption rate can be expressed as a function of k_y and k_z so even if $k_y = 0$, as long as $k_z l^2(\omega/\Omega_i) \neq 0$ (l is a length scale related to ω, v_A, Ω_i , and k_z), we can still get energy absorption.

²Check the papers by Lin, Hasegawa, Chen, Johnson, and Cheng.

16.5 Nonlinear Effects

Cold ions theory and an extension to finite $k_{\perp}r_{iL}$ predict that a pump KAW can nonlinearly decay into an ion acoustic mode and a lower-sideband daughter KAW through a parametric instability (Section 17.4).

16.6 3D FLR

17 Nonlinear Effects

In many experiments waves are no longer describable by the linear theory by the time they are observed. Consider, for instance, the case of drift waves. Because they are unstable, drift waves would, according to linear theory, increase their amplitude exponentially. This period of growth is not normally observed — since one usually does not know when to start looking — but instead one observes the waves only after they have grown to a large, steady amplitude. The fact that the waves are no longer growing means that the linear theory is no longer valid, and some *nonlinear* effect is limiting the amplitude. Theoretical explanation of this elementary observation has proved to be a surprisingly difficult problem, since the observed amplitude at saturation is rather small.

A wave can undergo a number of changes when its amplitude gets large. It can change its shape — say, from a sine wave to a lopsided triangular waveform. This is the same as saying that Fourier components at other frequencies (or wave numbers) are generated. Ultimately, the wave can “break”, like ocean waves on a beach, converting the wave energy into thermal energy of the particles. A large wave can trap particles in its potential troughs, thus changing the properties of the medium in which it propagates (e.g. nonlinear Landau damping). If a plasma is so strongly excited that a continuous spectrum of frequencies is present, it is in a state of *turbulence*. This state must be described statistically, as in the case of hydrodynamics. An important consequence of plasma turbulence is *anomalous resistivity*, in which electrons are slowed down by collisions with random electric field fluctuations, rather than with ions. This effect is used for ohmic heating of a plasma to temperature so high that ordinary resistivity is insufficient.

Nonlinear phenomena can be grouped into three broad categories:

1. *Basically nonlinearizable problems.*

- Diffusion in a fully ionized gas, for instance, is intrinsically a nonlinear problem because the diffusion coefficient varies with density.
- Problems of hydromagnetic equilibrium are nonlinear.
- Plasma sheath.

2. *Wave-particle interactions.*

- Particle trapping can lead to nonlinear damping (Section 12.6).

- A classic example is the quasilinear effect, in which the equilibrium of the plasma is changed by the waves. Consider the case of a plasma with an electron beam (Fig.8-1). Since the distribution function has a region where df_0/dv is positive, the system has inverse Landau damping, and plasma oscillations with v_ϕ in the positive-slope region are unstable (Equation 12.32). The resonant electrons are the first to be affected by wave-particle interactions, and their distribution function will be changed by the wave electric field. The waves are stabilized when $f_e(v)$ is flattened by the waves, as shown by the dashed line in Fig.8-1 ?@fig-unstable-f, so that the new equilibrium distribution no longer has a positive slope.
- Plasma wave echoes (Section 17.5).

3. Wave-wave interactions.

- Waves can interact with each other even in the fluid description, in which individual particle effects are neglected. A single wave can decay by first generating harmonics of its fundamental frequency. These harmonics can then interact with each other and with the primary wave to form other waves at the beat frequencies. The beat waves in turn can grow so large that they can interact and form many more beat frequencies, until the spectrum becomes continuous. It is interesting to discuss the direction of energy flow in a turbulent spectrum. In fluid dynamics, long-wavelength modes decay into short-wavelength modes, because the large eddies contain more energy and can decay only by splitting into small eddies, which are each less energetic. The smallest eddies then convert their kinetic motion into heat by viscous damping. In a plasma, usually the opposite occurs. Short wave-length modes tend to coalesce into long-wavelength modes, which are less energetic. This is because the electric field energy $E^2/8\pi$ is of order $k^2\phi^2/8\pi$, so that if $e\phi$ is fixed (usually by $k_B T_e$), the small- k , long- λ modes have less energy. As a consequence, energy will be transferred to small k by instabilities at large k , and some mechanism must be found to dissipate the energy. No such problem exists at large k , where Landau damping can occur. For motions along \mathbf{B}_0 , nonlinear “modulational” instabilities could cause the energy at small k to be coupled to ions and to heat them. For motions perpendicular to \mathbf{B}_0 , the largest eddies will have wavelengths of the order of the plasma radius and could cause plasma loss to the walls by convection.

Although problems still remain to be solved in the linear theory of waves and instabilities, the mainstream of plasma research has turned to the much less well understood area of nonlinear phenomena. The examples in the following sections will give an idea of some of the effects that have been studied in theory and in experiment.

17.1 The Necessity for Sheaths

In all practical plasma devices, the plasma is contained in a vacuum chamber of finite size. What happens to the plasma at the wall? For simplicity, let us confine our attention to a

one-dimensional model with no magnetic field ?@fig-sheath. Suppose there is no appreciable electric field inside the plasma; we can then let the potential ϕ be zero there. When ions and electrons hit the wall, they recombine and are lost. Since electrons have much higher thermal velocities than ions, they are lost faster and leave the plasma with a net positive charge. The plasma must then have potential positive with respect to the wall; i.e., the wall potential ϕ_w is negative. This potential cannot be distributed over the entire plasma, since Debye shielding will confine the potential variation to a layer of the order of several Debye lengths in thickness. This layer, which must exist on all cold walls with which the plasma is in contact, is called a sheath. The function of a sheath is to form a potential barrier so that the more mobile species, usually electrons, is confined electrostatically. The height of the barrier adjusts itself so that the flux of electrons that wave enough energy to go over the barrier to the wall is just equal to the flux of ions reaching the wall.

17.1.1 The Planar Sheath Equation

Recall that the Debye length is derived from linearizing Poisson's equation. To examine the exact behavior of $\phi(x)$ in the sheath, we must treat the nonlinear problem: we shall find that there is not always a solution. ?@fig-sheath-potential shows the situation near one of the walls. At the plane $x = 0$, ions are imagined to enter the sheath region from the main plasma with a drift velocity u_0 . This drift is needed to account for the loss of ions to the wall from the region in which they were created by ionization. For simplicity, we assume $T_i = 0$, so that all ions have the velocity u_0 at $x = 0$. We consider the steady state problem in a collisionless sheath region. The potential ϕ is assumed to decrease monotonically with x . Actually, ϕ could have spatial oscillations, and then there would be trapped particles in the steady state. This does not happen in practice because dissipative processes tend to destroy any such highly organized state.

If $u(x)$ is the ion velocity, conservation of energy requires

$$\begin{aligned} \frac{1}{2}mu^2 &= \frac{1}{2}mu_0^2 - e\phi(x) \\ u &= \left(u_0^2 - \frac{2e\phi}{m_i} \right)^{1/2} \end{aligned} \tag{17.1}$$

The ion equation of continuity then gives the ion density n_i in terms of the density n_0 in the main plasma:

$$\begin{aligned} n_0 u_0 &= n_i(x)u(x) \\ n_i(x) &= n_0 \left(1 - \frac{2e\phi}{m_i u_0^2} \right)^{1/2} \end{aligned} \tag{17.2}$$

In steady state, the electrons will follow the Boltzmann relation closely,

$$n_e(x) = n_0 \exp(e\phi/k_B T_e)$$

Poisson's equation is then

$$\epsilon_0 \frac{d^2\phi}{dx^2} = e(n_e - n_i) = en_0 \left[\exp \left(\frac{e\phi}{k_B T_e} \right) - \left(1 - \frac{2e\phi}{m_i u_0^2} \right)^{1/2} \right] \quad (17.3)$$

The structure of this equation can be seen more clearly if we simplify it with the following changes in notation:

$$\chi \equiv -\frac{e\phi}{k_B T_e} \quad \xi \equiv \frac{x}{\lambda_D} = x \left(\frac{n_0 e^2}{\epsilon_0 k_B T_e} \right)^{1/2} \quad \mu \equiv \frac{u_0}{(k_B T_e / m_i)^{1/2}}$$

Then it can be written as

$$\chi'' = \left(1 + \frac{2\chi}{\mu^2} \right)^{-1/2} - e^{-\chi} \quad (17.4)$$

where the prime denotes $d/d\xi$. This is the nonlinear equation of a plane sheath, and it has an acceptable solution only if μ is large enough. The reason for the symbol μ will become apparent in the following section on shock waves.

17.1.2 The Bohm Sheath Criterion

Equation 17.4 can be integrated once by multiplying both sides by χ' :

$$\int_0^\xi \chi' \chi'' d\xi_1 = \int_0^\xi \left(1 + \frac{2\chi}{\mu^2} \right)^{-1/2} \chi' d\xi_1 - \int_0^\xi e^{-\chi} \chi' d\xi_1$$

where ξ_1 is a dummy variable. Since $\chi = 0$ at $\xi = 0$, the integration easily yield

$$\frac{1}{2}(\chi'^2 - \chi_0'^2) = \mu^2 \left[\left(1 + \frac{2\chi}{\mu^2} \right)^{1/2} - 1 \right] + e^{-\chi} - 1 \quad (17.5)$$

If $\mathbf{E} = 0$ in the plasma, we must set $\chi'_0 = 0$ at $\xi = 0$. A second integration to find χ would have to be done numerically; but whatever the answer is, the right-hand side of Equation 17.5 must be positive for all χ . In particular, for $\chi \ll 1$, we can expand the right-hand terms in Taylor series:

$$\begin{aligned} \mu^2 \left[1 + \frac{\chi}{\mu^2} - \frac{1}{2} \frac{\chi^2}{\mu^4} + \dots - 1 \right] + 1 - \chi + \frac{1}{2} \chi^2 + \dots - 1 &> 0 \\ \frac{1}{2} \chi^2 \left(-\frac{1}{\mu^2} + 1 \right) &> 0 \\ \mu^2 > 1 \quad \text{or} \quad u_0 > (k_B T_e / m_i)^{1/2} \end{aligned} \quad (17.6)$$

This inequality is know as the *Bohm sheath criterion*. It says that ions must enter the sheath with a velocity greater than the acoustic velocity v_s . To give the ions this directed velocity u_0 , there must be a finite electric field in the plasma. Our assumption that $\chi' = 0$ at $\xi = 0$

is therefore only an approximate one, made possible by the fact that the scale of the sheath region is usually much smaller than the scale of the main plasma region in which the ions are accelerated. The value of u_0 is somewhat arbitrary, depending on where we choose to put the boundary $x = 0$ between the plasma and the sheath. Of course, the ion flux $n_0 u_0$ is fixed by the ion production rate, so if u_0 varies, the value of n_0 at $x = 0$ will vary inversely with u_0 . If the ions have finite temperature, the critical drift velocity u_0 will be somewhat lower.

The physical reason for the Bohm criterion is easily seen from a plot of the ion and electron densities vs. χ ?@fig-sheath-n-variation. The electron density n_e falls exponentially with χ , according to the Boltzmann relation. The ion density also falls, since the ions are accelerated by the sheath potential. If the ions start with a large energy, $n_i(\chi)$ falls slowly, since the sheath field causes a relatively minor change in the ions' velocity. If the ions start with a small energy, $n_i(\chi)$ falls fast, and can go below the n_e curve. In that case, $n_e - n_i$ is positive near $\chi = 0$; and Equation 17.4 tells us that $\phi(x)$ must curve upward, in contradiction to the requirement that the sheath must repel electrons. In order for this not to happen, the slope of $n_i(\chi)$ at $\chi = 0$ must be smaller (in absolute value) than that of $n_e(\chi)$; this condition is identical with the condition $\mu^2 > 1$.

17.1.3 The Child-Langmuir Law

Since $n_e(\chi)$ falls exponentially with χ , the electron density can be neglected in the region of large χ next to the wall (or any negative electrode). Poisson's equation is then approximately

$$\chi'' \approx \left(1 + \frac{2\chi}{\mu^2}\right)^{-1/2} \approx \frac{\mu}{(2\chi)^{1/2}}$$

Multiplying by χ' and integrating from $\xi_1 = \xi_s$ to $\xi_1 = \xi$, we have

$$\frac{1}{2}(\chi'^2 - \chi_s'^2) = \sqrt{2}\mu(\chi^{1/2} - \chi_s^{1/2})$$

where ξ_s is the place where we started neglecting n_e . We can redefine the zero of χ so that $\chi_s = 1$ at $\xi = \xi_s$. We shall also neglect χ'_s , since the slope of the potential curve can be expected to be much steeper in the $n_e = 0$ region than in the finite- n_e region. Then the above equation becomes

$$\begin{aligned}\chi'^2 &= 2^{3/2}\mu\chi^{1/2} \\ \chi' &= 2^{3/4}\mu^{1/2}\chi^{1/4} \\ d\chi/\chi^{1/4} &= 2^{3/4}\mu^{1/2}d\xi\end{aligned}$$

Integrating from $\xi = \xi_s$ to $\xi = \xi_s + d/\lambda_D = \xi_{\text{wall}}$, we have

$$\frac{4}{3}\chi_w^{3/4} = 2^{3/4}\mu^{1/2}d/\lambda_D$$

or

$$\mu = \frac{4\sqrt{2}}{9} \frac{\chi_w^{3/2}}{d^2} \lambda_D^2$$

Changing back to the variables u_0 and ϕ , and noting that the ion current into the wall is $J = en_0 u_0$, we then find

$$J = \frac{4}{9} \left(\frac{2e}{m_i} \right)^{1/2} \frac{\epsilon_0 |\phi_w|^{3/2}}{d^2} \quad (17.7)$$

This is just the well-known Child-Langmuir law of space-charge-limited current in a plane diode.

The potential variation in a plasma-wall system can be divided into three parts. Nearest the wall is an electron-free region whose thickness d given by Equation 17.7. Here J is determined by the ion production rate, and ϕ_w is determined by the equality of electron and ion fluxes. Next comes a region in which n_e is appreciable; as shown in the linear theory from which the scale of Debye length is derived. Finally, there is a region with much larger scale length, the “presheath”, in which the ions are accelerated to the required velocity u_0 by a potential drop $|\phi| \geq \frac{1}{2}k_B T_E/e$. Depending on the experiment, the scale of the presheath may be set by the plasma radius, the collisional mean free path, or the ionization mechanism. The potential distribution, of course, varies smoothly; the division into three regions is made only for convenience and is made possible by the disparity in scale lengths. In the early days of gas discharges, sheaths could be observed as dark layers where no electrons were present to excite atoms to emission. Subsequently, the potential variation has been measured by the electrostatic deflection of a thin electron beam shot parallel to a wall.

17.1.4 Electrostatic Probes

The sheath criterion Equation 17.6 can be used to estimate the flux of ions to a negatively biased probe in a plasma. If the probe has a surface area A , and if the ions entering the sheath have a drift velocity $u_0 \geq (k_B T_e/m_i)^{1/2}$, then the ion current collected is

$$I = n_s e A (k_B T_e / m_i)^{1/2}$$

The electron current can be neglected if the probe is sufficiently negative (several times $k_B T_e$) relative to the plasma to repel all but the tail of the Maxwellian electron distribution. The density n_s is the plasma density at the edge of the sheath. Let us define the sheath edge to be the place where u_0 is exactly $(k_B T_e/m_i)^{1/2}$. To accelerate ions to this velocity requires a presheath potential $|\phi| \geq \frac{1}{2}k_B T_e/e$, so that the sheath edge has a potential

$$\phi_s \simeq -\frac{1}{2}k_B T_e/e$$

relative to the body of the plasma. If the electrons are Maxwellian, this determines n_s :

$$n_s = n_0 e^{e\phi_s/k_B T_e} = n_0 e^{-1/2} \approx 0.61 n_0$$

For our purposes it is accurate enough to replace 0.61 with a round number like 1/2; thus, the “saturation ion current” to a negative probe is approximately

$$I_B \simeq \frac{1}{2} n_0 e A (k_B T_e / m_i)^{1/2}$$

I_B , sometimes called the “Bohm current”, give the plasma density easily, once the temperature is known.

If the Debye length λ_D , and hence the sheath thickness, is very small compared to the probe dimensions, the area of the sheath edge is effectively the same as the area A of the probe surface, regardless of its shape. At low densities, however, λ_D can become large, so that some ions entering the sheath can orbit the probe and miss it. Calculations of orbits for various probe shapes were first made by I. Langmuir and L. Tonks — hence the name “**Langmuir probe**” ascribed to this method of measurement. Though tedious, these calculations can give accurate determinations of plasma density because an arbitrary definition of sheath edge does not have to be made. By varying the probe voltage, the Maxwellian electron distribution is sampled, and the current-voltage curve of a Langmuir probe can also yield the electron temperature. The electrostatic probe was the first plasma diagnostic and is still the simplest and the most localized measurement device. Unfortunately, material electrodes can be inserted only in low-density, cool plasma.

More detailed explanation of how to measure the plasma density using a Langmuir probe is given [here](#).

17.2 Ion Acoustic Shock Waves

When a jet travels faster than sound, it creates a shock wave. This is a basically nonlinear phenomenon, since there is no period when the wave is small and growing. The jet is faster than the speed of waves in air, so the undisturbed medium cannot be “warned” by precursor signals before the large shock wave hits it. In hydrodynamic shock waves, collisions are dominant. Shock waves also exist in plasmas, even when there are no collisions. A magnetic shock, the “bow shock”, is generated by the earth as it plows through the interplanetary plasma while dragging along a dipole magnetic field. We shall discuss a simpler example: a collisionless, one-dimensional shock wave which develops from a large-amplitude ion wave.

(Normal space plasma textbooks discuss this part starting from R-H relations. I don’t like that.)

17.2.1 The Sagdeev Potential

?@fig-shock-potential shows the idealized potential profile of an ion acoustic shock wave. The reason for this shape will be given presently. The wave is traveling to the left with a velocity u_0 . If we go to the frame moving with the wave, the function $\phi(x)$ will be constant in time, and we will see a stream of plasma impinging on the wave from the left with a velocity u_0 . For simplicity, let T_i be zero, so that all the ions are incident with the same velocity u_0 , and let the electrons be Maxwellian. Since the shock moves much more slowly than the electron thermal speed, the shift in the center velocity of the Maxwellian can be neglected. The velocity of the ions in the shock wave is, from energy conservation,

$$u = \left(u_0^2 - \frac{2e\phi}{m_i} \right)^{1/2}$$

If n_0 is the density of the undisturbed plasma, the ion density in the shock is

$$n_i = \frac{n_0 u_0}{u} = n_0 \left(1 - \frac{2e\phi}{m_i u_0^2} \right)^{-1/2}$$

The electron density is given by the Boltzmann relation. Poisson's equation then gives

$$\epsilon_0 \frac{d^2\phi}{dx^2} = e(n_e - n_i) = en_0 \left[\exp \left(\frac{e\phi}{k_B T_e} \right) - \left(1 - \frac{2e\phi}{m_i u_0^2} \right)^{1/2} \right]$$

This is, of course, the same Equation 17.3 as we had for a sheath. A shock wave is no more than a sheath moving through a plasma. We now introduce the dimensionless variables

$$\chi \equiv \frac{e\phi}{k_B T_e} \quad \xi \equiv \frac{x}{\lambda_D} = x \left(\frac{n_0 e^2}{\epsilon_0 k_B T_e} \right)^{1/2} \quad \mu \equiv \frac{u_0}{(k_B T_e / m_i)^{1/2}}$$

Note that we have changed the sign in the definition of χ so as to keep χ positive in this problem. The quantity μ is called the *Mach number* of the shock. Poisson's equation can now be written

$$\chi'' = e^\chi - \left(1 - \frac{2\chi}{\mu^2} \right)^{-1/2} = -\frac{dV(\chi)}{d\chi} \quad (17.8)$$

which differs from the sheath Equation 17.4 only because of the change in sign of χ .

The behavior of the solution of Equation 17.8 was made clear by R. Z. Sagdeev, who used an analogy to an oscillator in a potential well. The displacement x of an oscillator subjected to a force $-mdV(x)/dx$ is given by

$$\ddot{x} = -dV/dx$$

If the right-hand side of Equation 17.8 is defined as $-dV/dx$, the equation is the same as that of an oscillator, with the potential χ playing the role of χ , and $d/d\xi$ replacing d/dt . The

quasipotential $V(\chi)$ is sometimes called the Sagdeev potential. The function $V(x)$ can be found from Equation 17.8 by integration with the boundary condition $V(\chi) = 0$ at $\chi = 0$:

$$V(\chi) = 1 - e^\chi + \mu^2 \left[1 - \left(1 - \frac{2\chi}{\mu^2} \right)^{1/2} \right] \quad (17.9)$$

For μ lying in a certain range, this function has the shape shown in ?@fig-shock-quasipotential. If this were a real well, a particle entering from the left will go to the right-hand side of the well ($x > 0$), reflect, and return to $x = 0$, making a single transit. Similarly, a quasiparticle in our analogy will make a single excursion to positive χ and return to $\chi = 0$, as shown in ?@fig-shock-soliton. Such a pulse is called a *soliton*: it is a potential and density disturbance propagating to the left in ?@fig-shock-soliton with velocity u_0 . (I DON'T FULLY UNDERSTAND!!!)

Now, if a particle suffers a loss of energy while in the well, it will never return to $x = 0$ but will oscillate (in time) about some positive value of x . Similarly, a little dissipation will make the potential of a shock wave oscillate (in space) about some positive value of ϕ . This is exactly the behavoir depicted in ?@fig-shock-potential. Actually, dissipation is not needed for this: reflection of ions from the shock front has the same effect. To understand this, imagine that the ions have a small thermal spread in energy and that the height $e\phi$ of the wave front is just large enough to reflect some of the ions back to the left, while the rest go over the potential hill to the right. The reflected ions cause an increase in ion density in the upstream region to the left of the shock front (?@fig-shock-potential). This means that the quantity

$$\chi' = \frac{1}{n_0} \int_0^\xi (n_e - n_i) d\xi_1$$

is decreased. Since χ' is the analog of dx/dt in the oscillator problem, our virtual oscillator has lost velocity and is trapped in the potential well of ?@fig-shock-quasipotential.

17.2.2 The Critical Mach Numbers

Solutions of either the soliton type or the wave-train type exist only for a range of μ . A lower limit for μ is given by the condition that $V(\chi)$ be a potential well, rather than a hill. Expanding Equation 17.9 for $\chi \ll 1$ yields

$$\mu^2 > 1 \quad u_0 > (k_B T_e / m_i)^{1/2}$$

This is exactly the same, both physically and mathematically, as the Bohm criterion for the existence of a sheath (Equation 17.6).

An upper limit to μ is imposed by the condition that the function $V(\chi)$ of [?@fig-shock-quasipotential](#) must cross the χ axis for $\chi > 0$; otherwise, the virtual particle will not be reflected, and the potential will rise indefinitely. From Equation 17.9, we require

$$e^\chi - 1 < \mu^2 \left[1 - \left(1 - \frac{2\chi}{\mu^2} \right)^{1/2} \right]$$

for some $\chi > 0$. If the lower critical Mach number is surpassed ($\mu > 1$), the left-hand side, representing the integral of the electron density from zero to χ , is initially larger than the right-hand side, representing the integral of the ion density. As χ increases, the right-hand side can catch up with the left-hand side if μ^2 is not too large. However, because of the square root, the largest value χ can have is $\mu^2/2$. This is because $e\phi$ cannot exceed $\frac{1}{2}m_i u_0^2$; otherwise, ions would be excluded from the plasma in the downstream region. Inserting the largest value of χ into the above equation, we have

$$\exp(\mu^2/2) - 1 < \mu \quad \text{or} \quad \mu < 1.6$$

This is the upper critical Mach number. Shock waves in a cold-ion plasma therefore exist only for $1 < \mu < 1.6$.

As in the case of sheaths, the physical situation is best explained by a diagram of n_i and n_e vs. χ ([?@fig-shock-n](#)). This diagram differs from [?@fig-sheath-n-variation](#) because of the change of sign of ϕ . Since the ions are now decelerated rather than accelerated, n_i will approach infinity at $\chi = \mu^2/2$. The low critical Mach number ensures that the n_i curve lies below the n_e curve at small χ , so that the potential $\phi(x)$ starts off with the right sign for its curvature. When the curve n_{i1} crosses the n_e curve, the soliton ([?@fig-shock-soliton](#)) has an inflection point. Finally, when χ is large enough that the areas under the n_i and n_e curves are equal, the soliton reaches a peak, and the n_{i1} and n_e curves are retraced as χ goes back to zero. The equality of the areas ensures that the net charge in the soliton is zero; therefore, there is no electric field outside. If μ is larger than 1.6, we have the curve n_{i2} , in which the area under the curve is too small even when χ has reached its maximum value of $\mu^2/2$.

17.2.3 Wave Steepening

If one propagates an ion wave in a cold-ion plasma, it will have the phase velocity given by

$$v_p = \left(\frac{k_B T_e + \gamma_i k_B T_i}{m_i} \right)^{1/2} = \left(\frac{k_B T_e}{m_i} \right)^{1/2}$$

corresponding to $\mu = 1$. How, then, can one create shocks with $\mu > 1$? One must remember that the above phase velocity was a linear result valid only at small amplitudes. As the amplitude is increased, an ion wave speeds up and also changes from a sine wave to a sawtooth shape with a steep leading edge ([?@fig-wave-steepening](#)). The reason is that the wave electric field has accelerated the ions. In [?@fig-wave-steepening](#), ions at the peak of the

potential distribution have a large velocity in the direction of v_ϕ than those at the trough, since they have just experienced a period of acceleration as the wave passed by. In linear theory, this difference in velocity is taken into account, but not the displacement resulting from it. In nonlinear theory, it is easy to see that the ions at the peak are shifted to the right, while those at the trough are shifted to the left, thus steepening the wave shape. Since the density perturbation is in phase with the potential, more ions are accelerated to the right than to the left, and the wave causes a net mass flow in the direction of propagation. This causes the wave velocity to exceed the acoustic speed in the undisturbed plasma, so that μ is larger than unity.

17.2.4 Double Layers

A phenomenon related to sheaths and ion acoustic shocks is that of the [double layer](#). This is a localized potential jump, believed to occur naturally in the ionosphere, which neither propagates nor is attached to a boundary. The name comes from the successive layers of net positive and net negative charge that are necessary to create a step in $\phi(x)$. Such a step can remain stationary in space only if there is a plasma flow that Doppler shifts a shock front down to zero velocity in the lab frame, or if the distribution functions of the transmitted and reflected electrons and ions on each side of the discontinuity are specially tailored so as to make this possible. In natural situations double layers are likely to arise where there are gradients in the magnetic field \mathbf{B} , not where \mathbf{B} is zero or uniform, as in laboratory simulations. In that case, the $\mu\nabla B$ force can play a large role in localizing a double layer away from all boundaries. Indeed, the thermal barrier in tandem mirror reactors is an example of a double layer with strong magnetic trapping.

17.3 The Ponderomotive Force

Light waves exert a radiation pressure which is usually very weak and hard to detect. Even the esoteric example of comet tails, formed by the pressure of sunlight, is tainted by the added effect of particles streaming from the sun. When high-powered microwaves or laser beams are used to heat or confine plasmas, however, the radiation pressure can reach several hundred thousand atmospheres! When applied to a plasma, this force is coupled to the particles in a somewhat subtle way and is called the *ponderomotive force*. Many nonlinear phenomena have a simple explanation in terms of the ponderomotive force.

The easiest way to derive this nonlinear force is to consider the motion of an electron in the oscillating \mathbf{E} and \mathbf{B} fields of a wave. We neglect dc \mathbf{E}_0 and \mathbf{B}_0 fields. The electron equation of motion is

$$m\dot{\mathbf{v}} = -e[\mathbf{E}(\mathbf{r}) + \mathbf{v} \times \mathbf{B}(\mathbf{r})]$$

This equation is exact if \mathbf{E} and \mathbf{B} are evaluated at the instantaneous position of the electron. The nonlinearity comes partly from the $\mathbf{v} \times \mathbf{B}$, which is second order because both \mathbf{v} and \mathbf{B} vanish in the equilibrium, so that the term is no larger than $\mathbf{v}_1 \times \mathbf{B}_1$, where \mathbf{v}_1 and \mathbf{B}_1 are the linear-theory values. The other part of the nonlinearity, as we shall see, comes from evaluating \mathbf{E} at the actual position of the particle rather than its initial position. Assume a wave electric field of the form

$$\mathbf{E} = \mathbf{E}_s(\mathbf{r}) \cos \omega t$$

where $\mathbf{E}_s(\mathbf{r})$ contains the spatial dependence. In first order, we may neglect the $\mathbf{v} \times \mathbf{B}$ term in the equation of motion and evaluate \mathbf{E} at the initial position \mathbf{r}_0 . We have

$$\begin{aligned} m\dot{\mathbf{v}}_1 &= -e\mathbf{E}(\mathbf{r}_0) \\ \mathbf{v}_1 &= -(e/m\omega)\mathbf{E}_s \sin \omega t = \dot{\mathbf{r}}_1 \\ \delta\mathbf{r}_1 &= (e/m\omega^2)\mathbf{E}_s \cos \omega t \end{aligned}$$

It is important to note that *in a nonlinear calculation, we cannot write $e^{i\omega t}$ and takes its real part later*. Instead, we write its real part explicitly as $\cos \omega t$. This is because products of oscillating factors appear in nonlinear theory, and the operations of multiplying and taking the real part do not commute.

Going to second order, we expand $\mathbf{E}(\mathbf{r})$ about \mathbf{r}_0 :

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}(\mathbf{r}_0) + (\delta\mathbf{r}_0 \cdot \nabla)\mathbf{E}|_{r=r_0} + \dots$$

We must now add the term $\mathbf{v}_1 \times \mathbf{B}_1$, where \mathbf{B}_1 is given by Maxwell's equation:

$$\begin{aligned} \nabla \times \mathbf{E} &= -\dot{\mathbf{B}} \\ \mathbf{B}_1 &= -(1/\omega)\nabla \times \mathbf{E}_s|_{r=r_0} \sin \omega t \end{aligned}$$

The second-order part of the equation of motion is then

$$m\dot{\mathbf{v}}_2 = -e[(\delta\mathbf{r}_1 \cdot \nabla)\mathbf{E} + \mathbf{v}_1 \times \mathbf{B}_1]$$

Inserting the expressions of \mathbf{v}_1 , $\delta\mathbf{r}_1$ and \mathbf{B}_1 into the above and averaging over time, we have

$$m \left\langle \frac{d\mathbf{v}_2}{dt} \right\rangle = -\frac{e^2}{m\omega^2} \frac{1}{2} [(\mathbf{E}_s \cdot \nabla)\mathbf{E}_s + \mathbf{E}_s \times (\nabla \times \mathbf{E}_s)] = \mathbf{f}_{NL} \quad (17.10)$$

Here we have used $\langle \sin^2 \omega t \rangle = \langle \cos^2 \omega t \rangle = \frac{1}{2}$. The double cross product can be written as the sum of two terms, one of which cancels the $(\mathbf{E}_s \cdot \nabla)\mathbf{E}_s$ term.

What remains is

$$\mathbf{f}_{NL} = -\frac{1}{4} \frac{e^2}{m\omega^2} \nabla E_s^2$$

This is the effective nonlinear force on a single electron. The force per m^3 is \mathbf{f}_{NL} times the electron density n_0 , which can be written in terms of ω_p^2 . Since $E_s^2 = 2 \langle E^2 \rangle$, we finally have for the ponderomotive force the formula

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2} \nabla \frac{\langle \epsilon_0 E^2 \rangle}{2} \quad (17.11)$$

If the wave is electromagnetic, the second term in Equation 17.10 is dominant, and the physical mechanism for \mathbf{F}_{NL} is as follows. Electrons oscillate in the direction of \mathbf{E} , but the wave magnetic field distorts their orbits. That is, the Lorentz force $-ev \times \mathbf{B}$ pushes the electrons in the direction of \mathbf{k} (since \mathbf{v} is in the direction of \mathbf{E} , and $\mathbf{E} \times \mathbf{B}$ is in the direction of \mathbf{k}). The phases of \mathbf{v} and \mathbf{B} are such that the motion does not average to zero over an oscillation, but there is a secular drift along \mathbf{k} . If the wave has uniform amplitude, no force is needed to maintain this drift; but if the wave amplitude varies, the electrons will pile up in regions of small amplitude, and a force is needed to overcome the space charge. This is why the effective force \mathbf{F}_{NL} is proportional to the *gradient* of $\langle E^2 \rangle$. Since the drift for each electron is the same, \mathbf{F}_{NL} is proportional to the density — hence the factor ω_p^2/ω^2 .

If the wave is electrostatic, the first term in Equation 17.10 is dominant. Then the physical mechanism is simply that an electron oscillating along $\mathbf{k} \parallel \mathbf{E}$ moves farther in the half-cycle when it is moving from a strong-field region to a weak-field region than vice versa, so there is a net drift.

Although \mathbf{F}_{NL} acts mainly on the electrons, the force is ultimately transmitted to the ions, since it is a low-frequency or dc effect. When electrons are bunched by \mathbf{F}_{NL} , a charge-separation field \mathbf{E}_{cs} is created. The total force felt by the electrons is

$$\mathbf{F}_e = -e\mathbf{E}_{cs} + \mathbf{F}_{NL}$$

Since the ponderomotive force on the ions is smaller by $\Omega_p^2/\omega_p^2 = m_e/m_i$, the force on the ion fluid is approximately

$$\mathbf{F}_i = e\mathbf{E}_{cs}$$

Summing the last two equations, we find that the force on the plasma is \mathbf{F}_{NL} .

A direct effect of \mathbf{F}_{NL} is the [self-focusing of laser light in a plasma](#). In [?@fig-laser-self-focusing](#) we see that a laser beam of finite diameter causes a radially directed ponderomotive force in a plasma. This force moves plasma out of the beam, so that ω_p is lower and the dielectric constant ϵ is higher inside the beam than outside. The plasma then acts as a convex lens, focusing the beam to a smaller diameter.

17.4 Parametric Instabilities

The most thoroughly investigated nonlinear wave-wave interactions are the “parametric instabilities”, so called because of an analogy with parametric amplifiers, well-known devices in electrical engineering. A reason for the relatively advanced state of understanding of this subject is that the theory is basically a linear one, but linear about a non-oscillating equilibirum.

The parametric instabilities arise when a nonlinearity such as a pressure gradient couples waves. The waves must allow frequency and wavenumber matching which are consequences of energy and momentum conservation. But parametric instabilities are not limited to waves but also arise in nonlinear resonances such as a plasma sheath. A very common parametric wave instability arises from the decay of *strong electromagnetic wave*, the “pump”, into two electrostatic waves, an *electron plasma waves* and an *ion acoustic waves*. It is called a “decay” instability when the pump excites a *lower* frequency sideband, the electron plasma wave, and the difference mode, an ion acoustic wave. These three modes are the only ones in an unmagnetized plasma.

See [here](#) for an observation example from a lab experiment. The following video demonstrates some intuitive lab experiments from parametric excitation to parametric instability.

17.4.1 Coupled Oscillators

Consider the mechanical model of [?@fig-parametric-amplifier](#), in which two oscillators M1 and M2 are coupled to a bar resting on a pivot. The pivot P is made to slide back and forth at a frequency ω_0 , while the natural frequencies of the oscillators are ω_1 and ω_2 . It is clear that, in the absence of friction, the pivot encounters no resistance as long as M1 and M2 are not moving. Furthermore, if P is not moving and M2 is put into motion, M1 will move; but as long as 2 is not the natural frequency of M1, the amplitude will be small. Suppose now that both P and M2 are set into motion. The displacement of M1 is proportional to the product of the displacement of M2 and the length of the lever arm and, hence, will vary in time as

$$\cos \omega_2 t \cos \omega_0 t = \frac{1}{2} \cos[(\omega_2 + \omega_0)t] + \frac{1}{2} \cos[(\omega_2 - \omega_0)t]$$

If ω_1 is equal to either $\omega_2 + \omega_0$ or $\omega_2 - \omega_0$, M1 will be resonantly excited and will grow to large amplitude. Once M1 starts oscillating, M2 will also gain energy, because one of the beat frequencies of ω_1 with ω_0 is just ω_2 . Thus, once either oscillator is started, each will be excited by the other, and the system is unstable. The energy, of course, comes from the “pump” P, which encounters resistance once the rod is slanted. If the pump is strong enough, its oscillation amplitude is unaffected by M1 and M2; the instability can then be treated by a linear theory. In a plasma, the oscillators P, M1, and M2 may be different types of waves.

17.4.2 Frequency Matching

The equation of motion for a simple harmonic oscillator x_1 is

$$\ddot{x}_1 + \omega_1^2 x_1 = 0 \quad (17.12)$$

where ω_1 is its resonant frequency. If it is driven by a time-dependent force which is proportional to the product of the amplitude E_0 of the driver, or pump, and the amplitude x_2 of a second oscillator, the equation of motion becomes

$$\ddot{x}_1 + \omega_1^2 x_1 = c_1 x_2 E_0 \quad (17.13)$$

where c_1 is a constant indicating the strength of the coupling. A similar equation holds for x_2 :

$$\ddot{x}_2 + \omega_2^2 x_2 = c_2 x_1 E_0 \quad (17.14)$$

Let $x_1 = \bar{x}_1 \cos \omega t$, $x_2 = \bar{x}_2 \cos \omega' t$, and $E_0 = \bar{E}_0 \cos \omega_0 t$. Equation 17.14 becomes

$$\begin{aligned} (\omega_2^2 - \omega'^2) \bar{x}_2 \cos \omega' t &= c_2 \bar{E}_0 \bar{x}_1 \cos \omega_0 t \cos \omega t \\ &= c_2 \bar{E}_0 \bar{x}_1 \frac{1}{2} \{ \cos[(\omega_0 + \omega)t] + \cos[(\omega_0 - \omega)t] \} \end{aligned}$$

The driving terms on the right can excite oscillators x_2 with frequencies

$$\omega' = \omega_0 \pm \omega \quad (17.15)$$

In the absence of nonlinear interactions, x_2 can only have the frequency ω_2 , so we must have $\omega' = \omega_2$. However, the driving terms can cause a frequency shift so that ω' is only approximately equal to ω_2 . Furthermore, ω' can be complex, since there is damping (which has been neglected so far for simplicity), or there can be growth (if there is an instability). In either case, x_2 is an oscillator with finite Q and can respond to a range of frequencies about ω_2 . If ω is small, one can see from that both choices for ω' may lie within the bandwidth of x_2 , and one must allow for the existence of two oscillators, $x_2(\omega_0 + \omega)$ and $x_2(\omega_0 - \omega)$.

Now let $x_1 = \bar{x}_1 \cos \omega'' t$ and $x_2 = \bar{x}_2 \cos[(\omega_0 \pm \omega)t]$ and insert into Equation 17.13:

$$\begin{aligned} (\omega_1^2 - \omega''^2) \bar{x}_1 \cos \omega'' t &= c_1 \bar{E}_0 \bar{x}_2 \frac{1}{2} (\cos\{\omega_0 + (\omega_0 \pm \omega)t\} + \cos\{\omega_0 - (\omega_0 \pm \omega)t\}) \\ &= c_1 \bar{E}_0 \bar{x}_2 \frac{1}{2} (\cos[(2\omega_0 \pm \omega)t] + \cos \omega t) \end{aligned} \quad (17.16)$$

The driving terms can excite not only the original oscillation $x_1(\omega)$, but also new frequencies $\omega'' = 2\omega_0 \pm \omega$. We shall consider the case $|\omega_0| \gg |\omega_1|$, so that $2\omega_0 \pm \omega$ lies outside the range of

frequencies to which x_1 can respond, and $x_1(2\omega_0 \pm \omega)$ can be neglected. We therefore have three oscillators, $x_1(\omega)$, $x_2(\omega_0 - \omega)$, and $x_2(\omega_0 + \omega)$, which are coupled by Equation 17.13 and Equation 17.14:

$$\begin{aligned} (\omega_1^2 - \omega^2)x_1(\omega) - c_1 E_0(\omega_0)[x_2(\omega_0 - \omega) + x_2(\omega_0 + \omega)] &= 0 \\ [\omega_2^2 - (\omega_0 - \omega)^2]x_2(\omega_0 - \omega) - c_2 E_0(\omega_0)x_1(\omega) &= 0 \\ [\omega_2^2 - (\omega_0 + \omega)^2]x_2(\omega_0 + \omega) - c_2 E_0(\omega_0)x_1(\omega) &= 0 \end{aligned} \quad (17.17)$$

The dispersion relation is given by setting the determinant of the coefficients equal to zero:

$$\begin{vmatrix} \omega^2 - \omega_1^2 & c_1 E_0 & c_1 E_0 \\ c_2 E_0 & (\omega_0 - \omega)^2 - \omega_2^2 & 0 \\ c_2 E_0 & 0 & (\omega_0 + \omega)^2 - \omega_2^2 \end{vmatrix} = 0 \quad (17.18)$$

A solution with $\Im(\omega) > 0$ would indicate an instability.

For small frequency shifts and small damping or growth rates, we can set ω and ω' approximately equal to the undisturbed frequencies ω_1 and ω_2 . Equation 17.15 then gives a frequency matching condition:

$$\omega_0 \approx \omega_2 \pm \omega_1 \quad (17.19)$$

When the oscillators are waves in a plasma, ωt must be replaced by $\omega t - \mathbf{k} \cdot \mathbf{r}$. There is then also a wavelength matching condition

$$\mathbf{k}_0 \approx \mathbf{k}_2 \pm \mathbf{k}_1 \quad (17.20)$$

describing spatial beats; that is, the periodicity of points of constructive and destructive interference in space. The two conditions Equation 17.19 and Equation 17.20 are easily understood by analogy with quantum mechanics. Multiplying the former by Planck's constant \hbar , we have

$$\hbar\omega_0 = \hbar\omega_2 \pm \hbar\omega_1$$

E_0 and x_2 may, for instance, be electromagnetic waves, so that $\hbar\omega_0$ and $\hbar\omega_2$ are the photon energies. The oscillator x_1 may be a Langmuir wave, or plasmon, with energy $\hbar\omega_1$. Equation 17.17 simply states the conservation of energy. Similarly, Equation 17.16 states the conservation of momentum $\hbar\mathbf{k}$.

For plasma waves, the simultaneous satisfaction of Equation 17.15 and Equation 17.16 in one-dimensional problems is possible only for certain combinations of waves. The required relationships are best seen on an $\omega - k$ diagram [?@fig-parallelogram-construction](#). (A) shows the dispersion curves of an electron plasma wave ω_2 (Bohm-Gross wave) and an ion acoustic wave ω_1 . A large-amplitude electron wave (ω_0, \mathbf{k}_0) can decay into a backward moving electron wave (ω_2, \mathbf{k}_2) and an ion wave (ω_1, \mathbf{k}_1) . The parallelogram construction ensures that $\omega_0 = \omega_1 + \omega_2$ and $\mathbf{k}_0 = \mathbf{k}_1 + \mathbf{k}_2$. The positions of (ω_0, \mathbf{k}_0) and (ω_2, \mathbf{k}_2) on the electron curve

must be adjusted so that the difference vector lies on the ion curve. Note that an electron wave cannot decay into two other electron waves, because there is no way to make the difference vector lie on the electron curve.

B shows the parallelogram construction for the “parametric decay” instability. Here, (ω_0, \mathbf{k}_0) is an incident electromagnetic wave of large phase velocity ($\omega_0/k_0 \approx c$). It excites an electron wave and an ion wave moving in opposite directions. Since $|\mathbf{k}_0|$ is small, we have $|\mathbf{k}_1| \approx -|\mathbf{k}_2|$ and $\omega_0 = \omega_1 + \omega_2$ for this instability.

C shows the $\omega - k$ diagram for the “parametric backscattering” instability, in which a light wave excites an ion wave and another light wave moving in the opposite direction. This can also happen when the ion wave is replaced by a plasma wave. By analogy with similar phenomena in solid state physics, these processes are called, respectively, “stimulated Brillouin scattering” and “stimulated Raman scattering.”

D represents the two-plasmon decay instability of an electromagnetic wave. Note that the two decay waves are both electron plasma waves, so that frequency matching can occur only if $\omega_0 \simeq 2\omega_p$. Expressed in terms of density, this condition is equivalent to $n \simeq n_c/4$, when n_c is the critical density associated with ω_0 ($n_C = m\epsilon_0\omega_0^2/e^2$). This instability can therefore be expected to occur only near the “quarter-critical” layer of an inhomogeneous plasma.

17.4.3 Instability Threshold

Parametric instabilities will occur at any amplitude if there is no damping, but in practice even a small amount of either collisional or Landau damping will prevent the instability unless the pump wave is rather strong. To calculate the threshold, one must introduce the damping rates Γ_1 and Γ_2 of the oscillators x_1 and x_2 . Equation 17.12 then becomes

$$\ddot{x}_1 + \omega_1^2 x_1 + 2\Gamma_1 \dot{x}_1 = 0$$

For instance, if x_1 is the displacement of a spring damped by friction, the last term represents a force proportional to the velocity. If x_1 is the electron density in a plasma wave damped by electron-neutral collisions, Γ_1 is $v_c/2$ (???) Examination of Equation 17.13, Equation 17.14, and Equation 17.18 will show that it is all right to use exponential notation and let $d/dt \rightarrow i\omega$ for x_1 and x_2 , as long as we keep E_0 real and allow \bar{x}_1 and \bar{x}_2 to be complex. Equation 17.13 and Equation 17.14 become

$$\begin{aligned} (\omega_1^2 - \omega^2 - 2i\omega\Gamma_1)x_1(\omega) &= c_1 x_2 E_0 \\ [\omega_2^2 - (\omega - \omega_0)^2 2i(\omega - \omega_0)\Gamma_2]x_2(\omega - \omega_0) &= c_2 x_1 E_0 \end{aligned} \tag{17.21}$$

We further restrict ourselves to the simple case of two waves — that is, when $\omega \simeq \omega_1$ and $\omega_0 - \omega \simeq \omega_2$ but $\omega_0 + \omega$ is far enough from ω_2 to be nonresonant — in which case the third row and column of Equation 17.18 can be ignored. If we now express x_1 , x_2 , and E_0 in

terms of their peak values, as in Equation 17.16, a factor of 1/2 appears on the right-hand sides of Equation 17.21. Discarding the nonresonant terms and eliminating x_1 and x_2 from Equation 17.21, we obtain

$$(\omega^2 - \omega_1^2 + 2i\omega\Gamma_1)[(\omega_0 - \omega)^2 - \omega_2^2 - 2i(\omega_0 - \omega)\Gamma_2] = \frac{1}{4}c_1 c_2 \bar{E}_0^2$$

At threshold, we may set $\Im(\omega) = 0$. The lowest threshold will occur at exact frequency matching; i.e., $\omega = \omega_1, \omega_0 - \omega = \omega_2$. Then the last equation gives

$$c_1 c_2 (\bar{E}_0^2)_{\text{threshold}} = 16\omega_1 \omega_2 \Gamma_1 \Gamma_2$$

The threshold goes to zero with the damping of *either* wave.

17.4.4 Physical Mechanism

The parametric excitation of waves can be understood very simply in terms of the ponderomotive force (Section 17.3). As an illustration, consider the case of an electromagnetic wave (ω_0, \mathbf{k}_0) driving an electron plasma wave (ω_2, \mathbf{k}_2) and a low-frequency ion wave (ω_1, \mathbf{k}_1) (figure parallelogram-construction B). Since ω_1 is small, ω_0 must be close to ω_p . However, the behavior is quite different for $\omega_0 < \omega_p$ than for $\omega_0 > \omega_p$. The former case gives rise to the “oscillating two-stream” instability (which will be treated in detail), and the latter to the “parametric decay” instability.

Suppose there is a density perturbation in the plasma of the form $n_1 \cos k_1 x$; this perturbation can occur spontaneously as one component of the thermal noise. Let the pump wave have an electric field $E_0 \cos \omega_0 t$ in the x direction, as shown in ?@fig-oscillating-2stream. In the absence of a dc field \mathbf{B}_0 , the pump wave follows the relation $\omega_0^2 = \omega_p^2 + c^2 k_0^2$, so that $k_0 \approx 0$ for $\omega_0 \approx \omega_p$. We may therefore regard E_0 as spatially uniform. If ω_0 is less than ω_p (HOW IS THIS POSSIBLE???), which is the resonant frequency of the cold electron fluid, the electrons will move in the direction opposite to E_0 , while the ions do not move on the time scale of ω_0 . The density ripple then causes a charge separation, as shown in ?@fig-oscillating-2stream. The electrostatic charges create a field E_1 , which oscillates at the frequency ω_0 . The ponderomotive force due to the total field is given by Equation 17.11:

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2} \nabla \frac{\langle (E_0 + E_1)^2 \rangle}{2} \epsilon_0$$

Since E_0 is uniform and much larger than E_1 , only the cross term is important:

$$\mathbf{F}_{NL} = -\frac{\omega_p^2}{\omega^2} \frac{\partial}{\partial x} \frac{\langle 2E_0 E_1 \rangle}{2} \epsilon_0$$

This force does not average to zero, since E_1 changes sign with E_0 . As seen in [?@fig-oscillating-2stream](#), F_{NL} is zero at the peaks and troughs of n_1 but is large where ∇n_1 is large. This spatial distribution causes F_{NL} to push electrons from regions of low density to regions of high density. The resulting de electric field drags the ions along also, and the density perturbation grows. The threshold value of F_{NL} is the value just sufficient to overcome the pressure $\nabla n_{i1}(k_B T_i + k_B T_e)$, which tends to smooth the density. The density ripple does not propagate, so that $\Re(\omega_1) = 0$. This is called the *oscillating two-stream instability* because the sloshing electrons have a time-averaged distribution function which is double-peaked, as in the two-stream instability (Sect. ???).

17.4.5 The Oscillating Two-Stream Instability

I do not understand parametric instability so far. Too hard.

17.4.6 The Parametric Decay Instability

17.5 Plasma Echoes

Echoes in a plasma are the excitement of new waves due to nonlinear interaction. The excitement may happen at an arbitrarily large time, which is the main source of difficulties in understanding Landau damping. Since Landau damping *does not involve collisions or dissipation*, it is a *reversible* process. That this is true is vividly demonstrated by the remarkable phenomenon of plasma echoes.

17.6 Nonlinear Landau Damping

When the amplitude of an electron or ion wave excited, say, by a grid is followed in space, it is often found that the decay is not exponential, as predicted by linear theory, if the amplitude is large. Instead, one typically finds that the amplitude decays, grows again, and then oscillates before settling down to a steady value. Although other effects may also be operative, these oscillations in amplitude are exactly what would be expected from the nonlinear effect of particle trapping discussed in Section [12.6](#).

17.7 Equations of Nonlinear Plasma Physics

17.7.1 The Korteweg-de Vries Equation

$$\frac{\partial U}{\partial \tau} + U \frac{\partial U}{\partial \xi} + \frac{1}{2} \frac{\partial^2 U}{\partial \xi^3} = 0$$

where U is amplitude, and τ and ξ are timelike and spacelike variables, respectively.

17.7.2 The Nonlinear Schrödinger Equation

$$i \frac{\partial \psi}{\partial t} + p \frac{\partial^2 \psi}{\partial x^2} + q |\psi|^2 \psi = 0 \quad (17.22)$$

where ψ is the wave amplitude, $i = \sqrt(-1)$, and p and q are coefficients whose physical significance will be explained shortly. Equation ?? differs from the usual Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} + \frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} - V(x, t)\psi = 0$$

in that the potential $V(x, t)$ depends on ψ itself, making the last term nonlinear.

18 Turbulence

Turbulence is perhaps the most beautiful unsolved problem of classical physics. History of turbulence in interplanetary space is, perhaps, even more interesting since its knowledge proceeds together with the human conquest of space. The behavior of a flow which rebels against the deterministic rules of classical dynamics is called *turbulent*. Practically turbulence appears everywhere when the velocity of the flow is high enough. At a first sight turbulence looks strongly irregular and chaotic, both in space and time: it seems to be impossible to predict any future state of the motion. However, it is interesting to recognize the fact that, when we take a picture of a turbulent flow at a given time, we see the presence of a lot of *different turbulent structures of all sizes* which are actively present during the motion. We recognize that the gross features of the flow are reproducible but details are not predictable. We have to use a *statistical approach* to turbulence, just as it is done to describe stochastic processes, even if the problem is born within the strange dynamics of a deterministic system.

Small fluctuations in plasmas lead to turbulence, and turbulent eddies can very effectively transport mass, momentum and heat from the hot core across confining magnetic field lines out to the cooler plasma edge. Predicting this phenomenon of turbulent-transport is essential for solar wind study and the understanding and development of fusion reactors.

Turbulence became an experimental science since Osborne Reynolds who, at the end of 19th century, observed and investigated experimentally the transition from laminar to turbulent flow. He noticed that the flow inside a pipe becomes turbulent every time a single parameter, a combination of the mass density ρ , viscosity coefficient η , a characteristic velocity U , and length L , would increase. This parameter $Re = UL\rho/\eta$ is now called the *Reynolds number*. When Re increases beyond a certain threshold of the order of $Re \simeq 4000$, the flow becomes turbulent.

Predictability in turbulence can be recast at a statistical level. In other words, when we look at two different samples of turbulence, even collected within the same medium, we can see that details look very different. What is actually common is a generic stochastic behavior. This means that the global statistical behavior does not change going from one sample to the other. *Fully developed turbulent flows are extremely sensitive to small perturbations but have statistical properties that are insensitive to perturbations.* Fluctuations of a certain stochastic variable ψ are defined as the difference from the average value $\delta\psi = \psi - \langle\psi\rangle$, where brackets mean some averaging process. There are, at least, three different kinds of averaging procedures that may be used to obtain statistically-averaged properties of turbulence:

1. Space averaging over flows that are statistically homogeneous over scales larger than those of fluctuations.

2. Ensemble averaging where average is taken over an ensemble of turbulent flows prepared under nearly identical external conditions. Each member of the ensemble is called a *realization*.
3. Time averaging, which is useful only if the turbulence is statistically stationary over time scales much larger than the time scale of fluctuations. In practice, because of the convenience offered by locating a probe at a fixed point in space and integrating in time, experimental results are usually obtained as time averages. The ergodic theorem (Halmos, 1956) assures that time averages coincide with ensemble averages under some standard conditions.

A different property of turbulence is that *all dynamically interesting scales are excited*, that is, *energy is spread over all scales* and a kind of self-similarity is observed.

Since fully developed turbulence involves a hierarchy of scales, a large number of interacting degrees of freedom are involved. Then, there should be an asymptotic statistical state of turbulence that is independent on the details of the flow. Hopefully, this asymptotic state depends, perhaps in a critical way, only on simple statistical properties like energy spectra, as much as in statistical mechanics equilibrium where the statistical state is determined by the energy spectrum (Huang, 1987). Of course, we cannot expect that the statistical state would determine the details of individual realizations, because realizations need not to be given the same weight in different ensembles with the same low-order statistical properties.

It should be emphasized that there are no firm mathematical arguments for the existence of an asymptotic statistical state. Reproducible statistical results are obtained from observations, that is, it is suggested experimentally and from physical plausibility. Apart from physical plausibility, it is embarrassing that such an important feature of fully developed turbulence, as the existence of a statistical stability, should remain unsolved. However, such is the complex nature of turbulence.

Amitava Bhattacharjee gave a nice tutorial on the [MHD turbulence in space plasmas](#).

18.1 Weak and Strong Turbulence

Weak turbulence

- Nonlinearity: Interactions between waves or eddies are weak in the sense that the energy in these interactions is much smaller than the total energy of the system.
- Perturbative Approach: Weak turbulence theory is often treated via a perturbative expansion where nonlinear terms are considered small corrections to the linear behavior of waves. This allows for an analytical treatment of wave-wave interactions.
- Three-wave interactions: The fundamental interactions involve three waves satisfying resonance conditions in frequency and wavenumber (e.g., two waves combining to form a third wave).

- Predictability: Since the interactions are weak, the time evolution of the system is relatively predictable, and statistics of the turbulent state can be calculated more directly.
- Energy Cascades: Energy still gets transferred across scales from large to small, creating a characteristic energy spectrum.

Strong turbulence

- Intense Nonlinearity: Interactions between waves or eddies become comparable in strength to the total energy of the system. Linear wave descriptions break down significantly.
- Complex Interaction Web: Simple three-wave interactions no longer fully describe the behavior. More complex, multi-wave interactions and turbulent structures play a dominant role.
- Less Analytical Tractability: Strong turbulence is difficult, often impossible, to fully solve analytically, because linear approximation no longer works. Numerical simulations and phenomenological modeling become essential tools.
- Chaotic Behavior: The system can show extreme sensitivity to initial conditions, making long-term predictions about the detailed evolution very challenging.
- Energy Cascades and Dissipation: The transfer of energy across scales becomes more rapid, and may no longer follow simple power-law energy spectra like in weak turbulence. Dissipation at small scales becomes significant.

In the context of water waves, an example of weak turbulence is the small ripples on a relatively calm pond, and an example of strong turbulence is the crashing waves on a stormy ocean. Intuitively, we can tell that the weak turbulence is easier to describe mathematically than the strong turbulence.

18.2 MHD

Nunov highly recommended a review (Schekochihin 2022). Another nice reference for the solar wind turbulence is presented (Bruno and Carbone 2013), where I cite many materials in this note.

Magnetohydrodynamic turbulence concerns the chaotic regimes of magnetofluid flow at high Reynolds number. Magnetohydrodynamics (MHD) deals with what is a quasi-neutral fluid with very high conductivity. The fluid approximation implies that the focus is on macro length-and-time scales which are much larger than the collision length and collision time respectively.

The guide magnetic field in turbulence is usually considered as the steady state background field. It can either be imposed by external sources, or created by large-scale eddies.

18.2.1 Comparison with Hydrodynamics

- HD turbulence: interaction of eddies.
- MHD turbulence: interaction of wave packets moving with Alfvén velocities.

18.2.2 Incompressible MHD equations

In the context of studying turbulence, the most commonly used math tool is the incompressible MHD equations. Incompressibility means $\nabla \cdot \mathbf{u} = 0$, but not necessarily constant density. Physically it means infinite sound speed $V_s \sim \sqrt{\delta p / \delta \rho} \rightarrow \infty$, i.e. the density will not change no matter how large the pressure variation is. In many scenarios, assuming constant density is good enough and lead to further simplification of math.¹

The incompressible MHD equations for constant mass density, $\rho = 1$, are

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} &= -\nabla P + \mathbf{B} \cdot \nabla \mathbf{B} + \nu \nabla^2 \mathbf{u} \\ \frac{\partial \mathbf{B}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{B} &= \mathbf{B} \cdot \nabla \mathbf{u} + \eta \nabla^2 \mathbf{B} \\ \nabla \cdot \mathbf{u} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \quad (18.1)$$

where P is the total (thermal + magnetic) pressure, ν is the kinematic viscosity, and η is the magnetic diffusivity. In the above equation, the magnetic field is in Alfvén units (same as velocity units).

The total magnetic field can be split into a mean field \mathbf{B}_0 and a fluctuating field \mathbf{b} :

$$\mathbf{B} = \mathbf{B}_0 + \delta \mathbf{b}$$

We can then obtain the magnetic field equation for the perturbation term only:

$$\frac{\partial \mathbf{b}}{\partial t} = \mathbf{v}_A \cdot \nabla \mathbf{u} + \mathbf{b} \cdot \nabla \mathbf{u}$$

If the Hall effect is considered, then there will be additional terms in the magnetic field equation

$$\frac{\partial \mathbf{b}}{\partial t} = \mathbf{v}_A \cdot \nabla \mathbf{u} + \mathbf{b} \cdot \nabla \mathbf{u} + d_i \nabla \times [\mathbf{v}_A \times (\nabla \times \mathbf{b})] + d_i \nabla \times [\mathbf{b} \times (\nabla \times \mathbf{b})] + \eta \nabla^2 \mathbf{b} \quad (18.2)$$

The symmetric forms of \mathbf{u} and \mathbf{B} hint us that we may be able to combine them together; this is exactly what Walter Elsässer did back in 1950.

¹constant density leads to incompressibility, but not vice versa.

18.2.3 Coupling between charged fluid and magnetic field

Elsässer variables are used to extract the Alfvénic component from MHD. The perturbations in Elsässer form are written as

$$\begin{aligned}\mathbf{z}^+ &= \delta\mathbf{v} + \delta\mathbf{b} \\ \mathbf{z}^- &= \delta\mathbf{v} - \delta\mathbf{b}\end{aligned}\quad (18.3)$$

where $\mathbf{b} = \mathbf{B}/\sqrt{\mu_0\rho}$ or $\mathbf{b} = \mathbf{B}/\sqrt{4\pi\rho}$ in CGS units.

\mathbf{z}^\pm corresponds to anti-parallel/parallel propagating modes:

- Parallel wave: $\delta\mathbf{v} = -\delta\mathbf{b} \Rightarrow \mathbf{z}^+ = 0, \mathbf{z}^- = 2\delta\mathbf{v}$
- Anti-parallel wave: $\delta\mathbf{v} = \delta\mathbf{b} \Rightarrow \mathbf{z}^+ = 2\delta\mathbf{v}, \mathbf{z}^- = 0$

Note that each field can contain multiple fluctuating modes: $\mathbf{z}^- = \sum_k \mathbf{z}_k^-, \delta\mathbf{B} = \sum_k \delta\mathbf{B}_k$. \mathbf{z}^\pm is a field quantity.

The Alfvénicity condition

$$\delta\mathbf{v} = \pm\delta\mathbf{b} = \pm\delta\mathbf{B}/\sqrt{\mu_0\rho_0} \quad (18.4)$$

often appears in the context of discussing Alfvén waves.

The incompressible MHD wave equation in fluctuating Elsässer form is:

$$\frac{\partial\mathbf{z}^\pm}{\partial t} \mp \mathbf{v}_A \cdot \nabla_{\parallel} \mathbf{z}^\pm + \mathbf{z}^\mp \cdot \nabla \mathbf{z}^\pm = -\nabla p + \nu_+ \nabla^2 \mathbf{z}^\pm + \nu_- \nabla^2 \mathbf{z}^\mp \quad (18.5)$$

where $\nu_\pm = \frac{1}{2}(\nu \pm \eta)$. The incompressibility and divergence-free B condition are written as

$$\nabla \cdot \mathbf{z}^\pm = 0 \quad (18.6)$$

The zero divergence means that there are no forcing or dissipation terms. Since we have already assumed uniform density in Equation 18.1, $v_A = B_0/\rho = B_0$, we can write Equation 18.5 in an alternative form

$$\frac{\partial\mathbf{z}^\pm}{\partial t} \mp \mathbf{B}_0 \cdot \nabla_{\parallel} \mathbf{z}^\pm + \mathbf{z}^\mp \cdot \nabla \mathbf{z}^\pm = -\nabla p + \nu_+ \nabla^2 \mathbf{z}^\pm + \nu_- \nabla^2 \mathbf{z}^\mp \quad (18.7)$$

Equation 18.5 unveils the interesting phenomena in Alfvénic turbulence study. The second term on the left-hand side is a linear term that represents propagation of waves parallel to the mean field. The third term represent the non-linear interaction of counter-propagating waves, during which energy is transferred to smaller scales. This is exactly turbulence. In MHD, turbulence is manifested by the collision of Alfvén waves.

Using the linear perturbation theory, $\mp \mathbf{v}_A \cdot \nabla_{\parallel} \mathbf{z}^\pm$ scales as $(k_{\parallel} v_A) z^\pm$, and $\mathbf{z}^\mp \cdot \nabla \mathbf{z}^\pm$ scales as $(k_{\perp} z^\mp) z^\pm$. When $k_{\parallel} v_A \gg k_{\perp} z^\mp$, the turbulence is *weak*; when $k_{\parallel} v_A \simeq k_{\perp} z^\mp$, the turbulence is *strong*. Assuming an energy supply at small k , then first in the weak MHD turbulence the

energy $E(k_\perp) \propto k_\perp^{-2}$, and later in the strong MHD turbulence $E(k_\perp) \propto k_\perp^{-3/2}$. At larger k , eventually we will encounter the dissipation scale, where the slope changes sharply because of the balance/imbalance of the energy cascade of the interacting waves.

Several mechanisms may happen in the dissipation scale [Huang, Howes, McCubbin, 2024]: * Transit-Time Damping (TTD): analogy to Landau damping for magnetic field, also known as Barnes damping. Particles can be energized by mirror fields by first being accelerating by an induced electric field in the perpendicular direction (as opposed to Landau damping where the energy transfer occurs in the parallel direction), and then got transferred to the parallel direction. (Also appear in Section 14.2, since it is closely related to mirror modes.) When an electromagnetic wave propagates through a plasma with varying magnetic field strength, its electric and magnetic fields oscillate. If the frequency of these oscillations matches the natural frequency at which particles are reflected by the magnetic mirror force (2nd adiabatic motion?), a resonance occurs. During this resonance, particles effectively “surf” on the wave, gaining energy from it. This energy gain comes at the expense of the wave, causing it to damp. The energy gained by the particles is primarily in the direction perpendicular to the magnetic field. TTD is a collisionless damping mechanism, which has a unique signature in velocity space. It creates a bipolar pattern of energy transfer, where some particles gain energy and others lose energy, centered around a resonant velocity. The strength of TTD is directly related to the gradient in the magnetic field strength: a stronger gradient leads to more efficient damping.

When the nonlinear term can be neglected, we have normal mode waves. For a system with only \mathbf{z}^+ ,

$$\begin{aligned} \frac{\partial \mathbf{z}^+}{\partial t} - \mathbf{V}_A \cdot \nabla \mathbf{z}^+ &= 0 \\ \rightarrow \mathbf{z}^+ &\sim f(\mathbf{x} + \mathbf{V}_A t) \end{aligned} \tag{18.8}$$

For a system with only \mathbf{z}^- ,

$$\begin{aligned} \frac{\partial \mathbf{z}^+}{\partial t} + \mathbf{V}_A \cdot \nabla \mathbf{z}^+ &= 0 \\ \rightarrow \mathbf{z}^+ &\sim f(\mathbf{x} - \mathbf{V}_A t) \end{aligned} \tag{18.9}$$

where \mathbf{x} is the coordinate parallel to \mathbf{B} . Both Equation 18.8 and Equation 18.9 give the Alfvén normal modes

$$\omega = |k_\parallel| V_A \tag{18.10}$$

An interesting question is: Does an MHD Alfvén wave have more energy in its magnetic fluctuations or more energy in its velocity fluctuations?

Answer: Same energy in fluctuations of both types, since an MHD Alfvén wave either has only z^+ or only z^- .

Alfvénic quasimode: Nonlinearly driven, non-normal mode that need not satisfy an Alfvén wave dispersion relation but still retains some Alfvénic properties such as incompressibility and a high degree of correlation between $\delta\mathbf{b}$ and \mathbf{v} .

Consider a mode z_3^- that results from the nonlinear interaction of modes z_1^- and z_2^+ , $z_3^- = z_1^- z_2^+$. Because the nonlinear term is important,

$$\frac{\partial \mathbf{z}_3^-}{\partial t} + \mathbf{V}_A \cdot \nabla \mathbf{z}_3^- \neq 0 \rightarrow \omega_3 - k_{3\parallel} V_A \neq 0$$

Question: Consider a fluctuation in incompressible MHD that is part of a larger system of many interacting modes. If the fluctuation exists only on z^+ with no z^- component, what does it indicate?

Answer: No z^- component means $z^- = \delta \mathbf{v} - \frac{\delta \mathbf{B}}{\sqrt{4\pi\rho}} = 0$, which leads to in-phase magnetic field and velocity fluctuations. However, it does not mean that it must be an Alfvén wave traveling in one direction? (Ask Seth!)

18.2.4 Importance of Alfvén Waves

- Alfvén Waves can scatter fast ions in tokamak plasmas.
- Alfvén Waves may contribute to coronal heating.
- Alfvén Waves are the building blocks for more complicated structures and dynamics, and most notably, turbulence.

18.2.5 Wave Interaction Phenomenology

From Equation 18.5, we can study the nonlinear interaction of wave modes. Consider a mode z_3^- that results from the nonlinear interaction of modes z_1^- and z_2^+ : $z_3^- = z_1^- z_2^+$. The simplest three-wave interaction of shear-Alfvén waves is given as

$$\begin{aligned}\omega(k) &= \omega(k_1) + \omega(k_2) \\ \mathbf{k} &= \mathbf{k}_1 + \mathbf{k}_2\end{aligned}\tag{18.11}$$

where

$$\omega(k) = |k_z| v_A$$

In a simple version, we can write down as

$$\begin{aligned}\omega_3 &= \omega_1 \pm \omega_2 \\ k_{3\parallel} &= k_{1\parallel} \pm k_{2\parallel}\end{aligned}$$

The total wave number \mathbf{k} is composed of two parts: k_z is the wave number along the guide field direction, and k_\perp is the wave number perpendicular to the guide field. Since only counter-propagating waves interact, k_{1z} and k_{2z} should have opposite signs. In order to satisfy the interaction condition, we can have either $k_{1z} = 0$ or $k_{2z} = 0$. That means wave interactions

change k_\perp but not k_z . This was verified in lab experiments with two counter-propagating modes, where we can isolate single interaction for detailed study (e.g. Howes+ 2012). An interesting point is that while from incompressible ideal MHD prediction, two modes moving along the same direction cannot have nonlinear interactions, one lab experiment showed that this is not true. This is attributed to the Hall effect, which adds additional complexity to the equations.

18.2.6 Nondimensional parameters

The important nondimensional parameters for MHD are

- Reynolds number: $\text{Re} = UL/\nu$
- Magnetic Reynolds number: $\text{Re}_M = UL/\eta$
- Magnetic Prandtl number: $P_M = \nu/\eta$

The magnetic Prandtl number is an important property of the fluid. Liquid metals have small magnetic Prandtl numbers, for example, liquid sodium's P_M is around 10^{-5} . But plasmas have large P_M .

The Reynolds number is the ratio of the nonlinear term $\mathbf{u} \cdot \nabla \mathbf{u}$ of the Navier–Stokes equation to the viscous term. While the magnetic Reynolds number is the ratio of the nonlinear term and the diffusive term of the induction equation.

In many practical situations, the Reynolds number Re of the flow is quite large. For such flows typically the velocity and the magnetic fields are random. Such flows are called to exhibit MHD turbulence. Note that Re_M need not be large for MHD turbulence. Re_M plays an important role in dynamo (magnetic field generation) problem.

The mean magnetic field plays an important role in MHD turbulence, for example it can make the turbulence anisotropic; suppress the turbulence by decreasing energy cascade etc. The earlier MHD turbulence models assumed isotropy of turbulence, while the later models have studied anisotropic aspects.

18.2.7 Critical balance

Polarized Alfvén modes interact. Due to the dot product in the nonlinear term in Equation 18.5, only the *perpendicularly* polarized components of the modes interact. The strength of the interaction can be quantified as

$$\chi \sim \frac{k_\perp z^\mp}{k_\parallel V_A}$$

If we match the Alfvén (linear) timescale, $\tau_A \sim l_{\parallel}/v_A$ (related to how fast Alfvén waves propagate along magnetic field lines), with the nonlinear timescale, $\tau_{nl} \sim l_{\perp}/\delta z$ (related to how fast turbulent eddies interact and distort each other),

$$\tau_A = \tau_{nl}$$

we have a way to predict the scaling of energy distribution across different scales, and the relationship between perpendicular and parallel structures in turbulence. This is known as the *critical balance* (Goldreich and Sridhar 1995), where $\chi \sim 1$, as is assumed in solar wind turbulence.

Assume a constant energy cascade rate $\epsilon \sim \delta z^2/\tau_{nl}$, and an energy injection rate at scale L $\epsilon_L \sim v_A^2/\tau_{nl} = v_A^3/L$. When we match the two rates and ... (I have some memory from Nunov's lecture at Nordita), we have

$$k_{\parallel} \propto k_{\perp}^{2/3} \quad (18.12)$$

This means that anisotropy grows with decreasing scale. It indicates the transition from “weak” to “strong” turbulence, or more accurately a transition from “quasi-linear” to “mode-coupling” processes.² The power spectra are given as

$$\begin{aligned} E(k_{\perp}) &\propto k_{\perp}^{-5/3} \\ E(k_{\parallel}) &\propto k_{\parallel}^{-2} \end{aligned} \quad (18.13)$$

This is among the most influential discoveries related to MHD turbulence, which makes it distinct from gas dynamic Kolmogorov turbulence by taking the magnetic field into account. By separating k_{\parallel} and k_{\perp} dynamics, the critical balance theory predicts perpendicular energy spectra Equation 18.13, which is consistent with the solar wind turbulence in the inertial range [Chen 2016]. The theory invoking this concept is now known as the anisotropic MHD turbulence.

However, why this is the case in nature is still under debate.

18.2.8 Residual energy

Residual energy is the difference in energy between magnetic and velocity fluctuations. The normalized residual energy is

$$\sigma_r = \frac{|\delta\mathbf{v}|^2 - |\delta\mathbf{b}|^2}{|\delta\mathbf{v}|^2 + |\delta\mathbf{b}|^2} \quad (18.14)$$

σ_r is zero for an Alfvén wave but is generally negative ($|\delta\mathbf{b}|^2 > |\delta\mathbf{v}|^2$) at MHD scales for solar wind observations, meaning that they are mostly likely turbulence but not waves.

²Two Alfvén wave packets moving towards each other will first be in the quasi-linear stage and then in the mode-coupling stage when they meet each other and interact nonlinearly.

Weak MHD turbulence spontaneously generates a condensate of the residual energy $E_r = E_b - E_k$ at small k_{\parallel} . Condensate broadens with k_{\perp} .

For Strong MHD turbulence, the residual energy is nonzero at all \mathbf{k} .

18.2.9 Normalized cross helicity

Cross helicity is the difference in energy between \mathbf{z}^+ and \mathbf{z}^- fluctuations (Section 19.5). The normalized cross helicity is

$$\sigma_c = \frac{|\mathbf{z}^+|^2 - |\mathbf{z}^-|^2}{|\mathbf{z}^+|^2 + |\mathbf{z}^-|^2} = \frac{2\delta\mathbf{v} \cdot \delta\mathbf{b}}{|\delta\mathbf{v}|^2 + |\delta\mathbf{b}|^2} \quad (18.15)$$

- $|\sigma_c| = 1$: unidirectional Alfvén waves (no turbulence)
- $|\sigma_c| \lesssim 1$: imbalanced turbulence
- $|\sigma_c| = 0$: balanced turbulence (or non-Alfvénic fluctuations)

The solar wind is typically imbalanced towards the anti-sunward direction. From Equation 18.14 and Equation 18.15 we can easily see that

$$\sqrt{\sigma_r^2 + \sigma_c^2} \leq 1$$

The observed distribution is shown in Figure 18.1. Note, however, σ_r depends on how you evaluate δb . In the MHD case $\mathbf{b} = \mathbf{B}/\sqrt{\mu_0\rho}$ is used, but in the kinetic case there will be an extra coefficient. The results look pretty different!

Observation from 1 AU (WIND) show that the non-Alfvénic wind typically has small alpha-proton relative drift and nearly equal temperatures of both ionic components. In terms of occurrence, there is no significant difference for a typical solar wind speed of 300 km/s, but for high speed solar wind there is a higher probability for large $|\sigma_c|$ of being Alfvénic.

- Non-Alfvénic solar winds
 - slower and lower in abundance, $n_{\alpha}/n_p < 4$
 - $v_p - v_{\alpha} \sim 0$
 - $T_p \sim T_{\alpha}$
 - could be formed by the interchange-reconnection near the Sun
- Alfvénic solar winds
 - $n_{\text{He}}/n_p \sim 4$ and does not depend on the proton velocity
 - alpha particles are about 4 times hotter than protons.
 - originates from the coronal holes

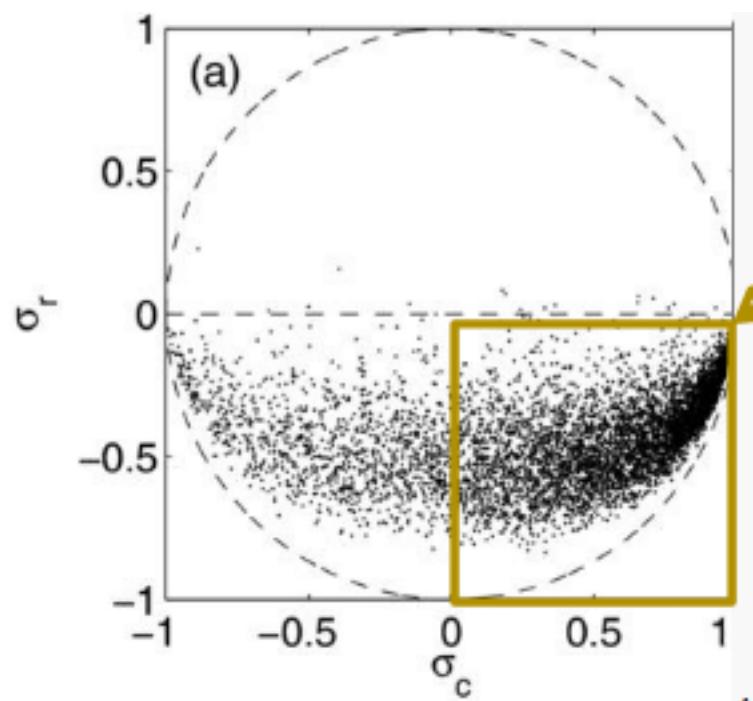


Figure 18.1: Solar wind normalized residual energy and cross helicity statistics from Chen+ 2013.

18.3 Solar Wind Turbulence

18.3.1 Taylor's hypothesis

Solar wind is a supersonic flow ($v_A/v_{\text{SW}} \ll 1$), with advection timescale much shorter than any dynamical timescales in the plasma. This means that for spacecraft observations, the time series represents an instantaneous spatial cut through the solar wind plasma. We can thus relate spacecraft frequency, f_{sc} , to wavenumber in the plasma frame, k , in a simple way via

$$k = \frac{2\pi f_{\text{sc}}}{v_{\text{SW}}}$$

This is often not appropriate in the magnetosheath, and modified Taylor's hypothesis is required close to the Sun.

The wavenumber k determined from Taylor's hypothesis is really the flow-aligned component of \mathbf{k} . For a given angle θ_{vb} , one cannot distinguish k_{\perp} and k_{\parallel} ; these can be possibly measured for small ($\sim 0^\circ$ for k_{\parallel}) and large ($\sim 90^\circ$ for k_{\perp}) at different times. In the solar wind, $P(k_{\perp}) \gg P(k_{\parallel})$, where P is the power of perturbation. For a Parker spiral-like magnetic field at 1 AU, the angle between \mathbf{B}_0 and \mathbf{v}_{sw} is rarely small ($\sim 45^\circ$), the power spectra are typically dominated by the contribution from the k_{\perp} fluctuations (Figure 18.2).

18.3.2 Solar Wind Power Spectrum

When we plot the solar wind power spectrum ($\delta|\mathbf{B}|$) as a function of time, it is usually representative of k_{\perp} ³. There are three distinct power-law ranges in the spectrum from spacecraft solar wind observations (Figure 18.4):

- Injection range:
 - Large amplitude, low frequency Alfvén waves originating from the Sun.
 - The f^{-1} spectrum is known as the “pink noise”.
 - Non-interacting
 - Inject energy to the MHD cascade at higher frequencies
- Inertial range:
 - Mostly incompressible Alfvénic turbulence.
 - $\sim f^{-5/3}$ spectrum (Kolmogorov type)
 - a cascade of energy to smaller scales
- Kinetic range:

³Of course there is an assumption here that the spatial scale and the time scale are proportional, since we cannot measure the spectrum as a function of space directly.

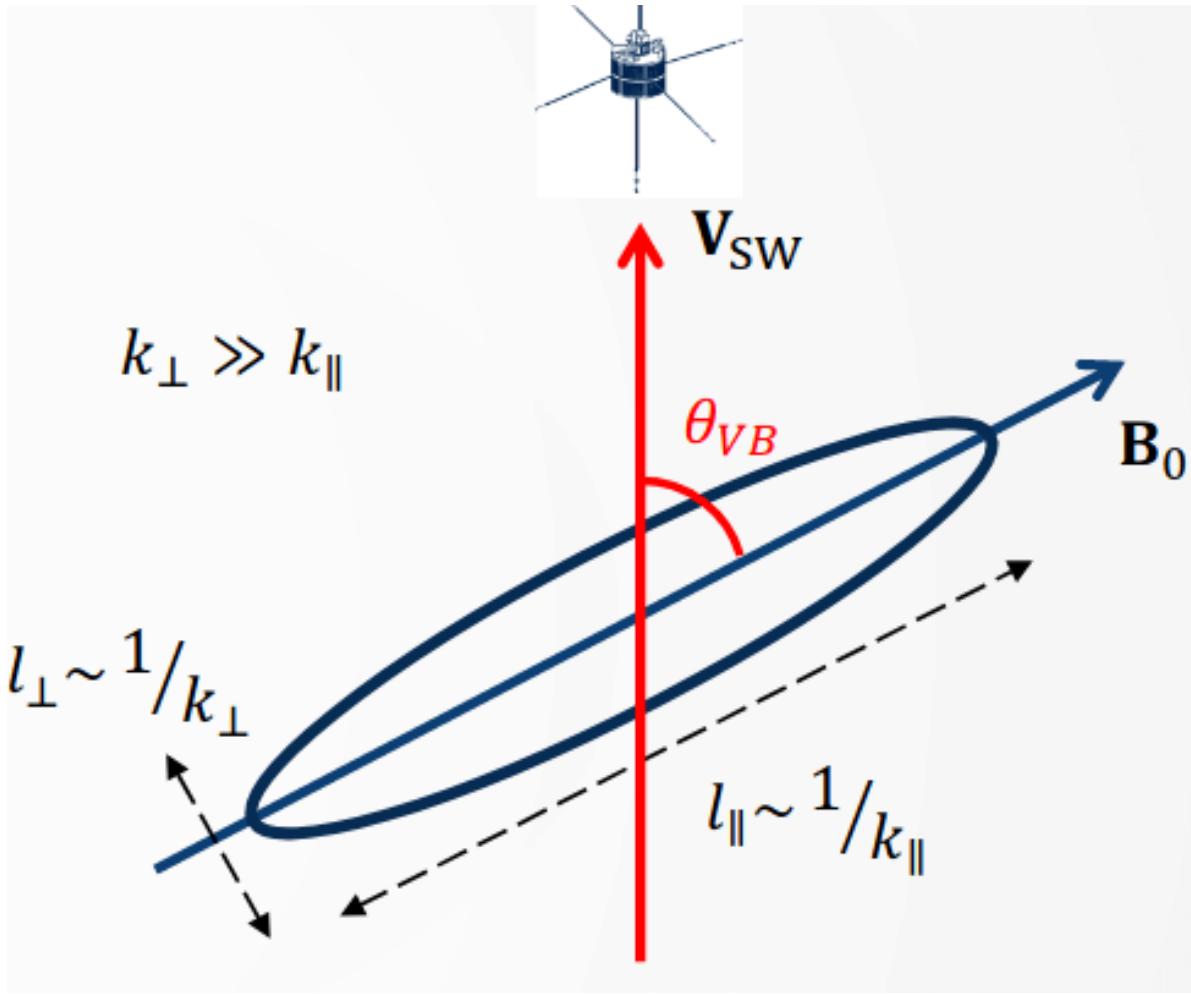


Figure 18.2: Wave scale from typical L1 satellite solar wind observation.

Mariner 10 1974:79:12 (0.12 sec)

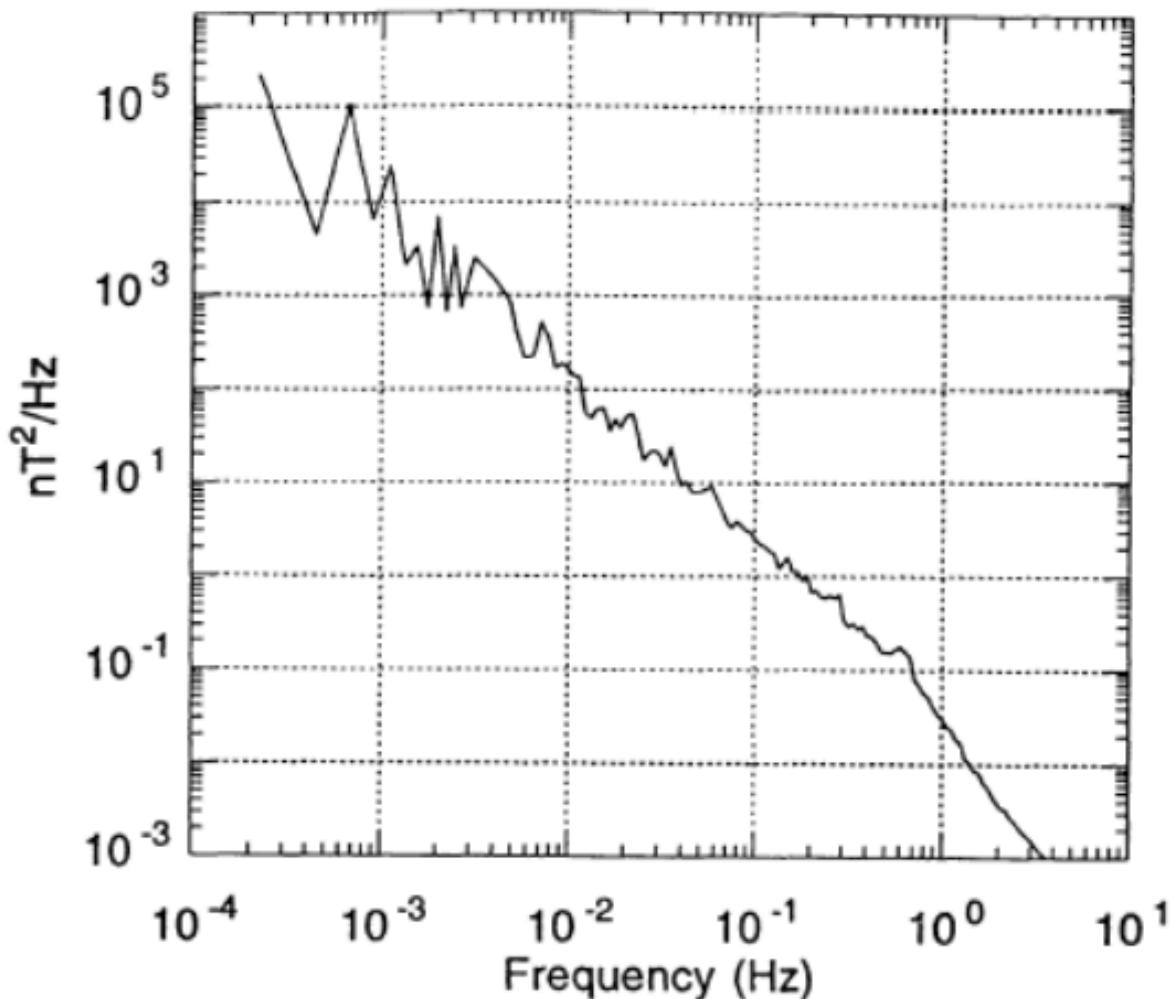


Figure 18.3: Solar wind magnetic field spectrum from Goldstein, Roberts, Matthaeus (1995).

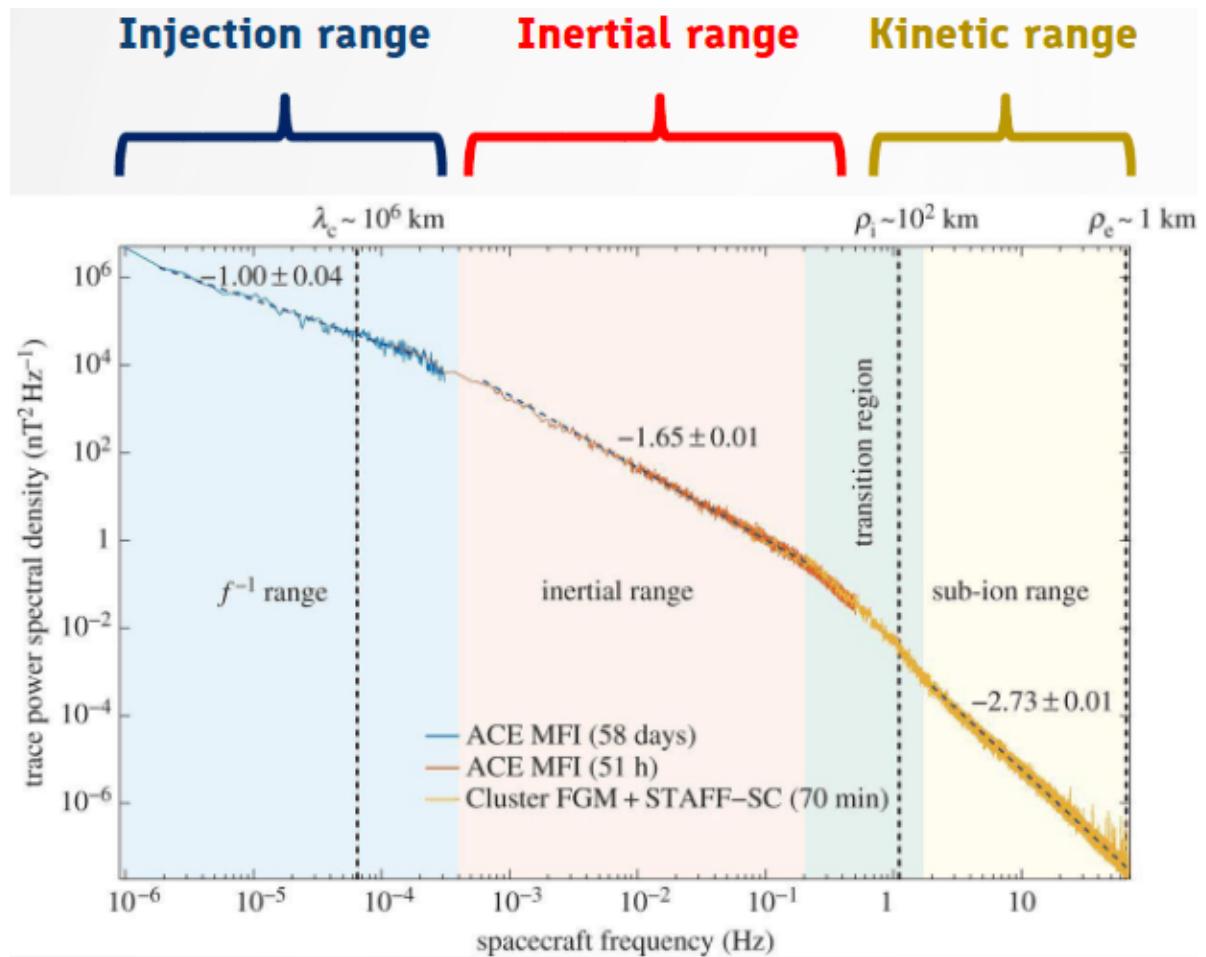


Figure 18.4: Solar wind power spectrum from Kiyoni+, 2015.

- Scales at which particles are heated
- Typically $f^{-2.8}$ spectrum
- Possibly KAW (Section 10.8.4) or whistler mode (Section 10.6) turbulence

When we move to smaller scales, anisotropy increases. This is demonstrated in Figure 18.5 as a result of $k_{\perp} \gg k_{\parallel}$, which is also why gyrokinetics is the theory for turbulence. Isotropic MHD is inadequate in turbulence study because of this major drawback.



Figure 18.5: Demonstration of parallel and perpendicular length scales of turbulence.

Fluctuation modes in the inertial range consists of 90% incompressible (Alfvén) modes and 10% compressible slow/mirror modes. Alfvénic turbulence is thought to be passive to the compressive modes: compressive modes scatter off Alfvén modes without affecting them significantly (i.e. being decoupled from each other). The compressive modes are expected to damp $\sim k_{\parallel} v_A$. Since $k_{\perp} \gg k_{\parallel}$, the damp is not significant for compressive modes, so they tend to be more anisotropic than the Alfvénic turbulence.

18.4 Turbulence and Reconnection

Turbulence and magnetic reconnection can be mutually driven, but the underlying nature of energy dissipation, intrinsic turbulence waves, and magnetic field topologies in turbulent magnetic reconnection is still poorly understood.

To study turbulence and reconnection together, one need to confirm several things beforehand:

1. The existence of reconnection, from identifying magnetic null point, plasma inflow and outflow, and the diffusion regions.
2. The existence of turbulence, by looking at the magnetic field power spectrum and confirm the cascades. From satellite data, this is done by taking a time interval, performing FFT, and check the characteristic frequency scales such as the ion cyclotron frequency f_{ci} , the lower hybrid frequency f_{ln} , etc.

If there are spectral breaks in the power spectrum diagram (PSD), especially near the characteristic frequencies, they may indicate local cyclotron resonance, which is related to wave-particle

interactions. Some researchers have found from MMS tail current sheet observations that in the reconnection outflow regions, energy is deposited in the form of kinetic Alfvén waves in low-frequency f_{ci} ion cyclotron scale and fast/slow waves in high-frequency low-hybrid scale.

19 Geometry

This chapter introduces the topology-related concepts in plasma physics, including ropes, knots, boundaries, and null points. Usually, observers and modelers have different views of topology because of the tools at hand: observers have probes which give measurements as a function of both time and space, while modelers have full spatial-temporal information under the given resolution. It always amazes me how observers can deduce the general picture of plasma structures with such limited data. Incorporating observation experience into physics as well as diagnosing numerical simulations with physics are our main goals in studying geometry.

Generally speaking, there are two ways of tackling the geometry problems: physics-based methods and statistical methods (classical and machine learning). We may find harmony when combining these two families and reach optimal results.

19.1 Local Coordinate System

Define the Jacobian $\vec{G} = \nabla \mathbf{B}$ of the magnetic field.

- Minimum Directional Derivative (MDD) Method: In the MDD method, local orthogonal coordinate directions can be obtained from eigenvectors of $\vec{G}^\top \vec{G}$.
- Minimum Gradient Analysis (MGA): In MGA, local orthogonal coordinate directions can be obtained from eigenvectors of $\vec{G} \vec{G}^\top$.

MGA produces a set of basis vectors that are analogous with the ones from Minimum Variance Analysis ([MVA](#)), with the eigenvector corresponding to the largest eigenvalue aligned with the vector that has maximal variation (which is denoted by \hat{L}_{MGA}), the second one corresponding to the intermediate (\hat{M}_{MGA}) and third the least variation (\hat{N}_{MGA}). MDD produces a set of basis vectors where the eigenvector corresponding to the largest eigenvalue shows the direction of the displacement which produces the largest variation in \mathbf{B} (the cross-sheet direction in a Harris current sheet; \hat{L}_{MDD}), idem with smaller eigenvalues (\hat{M}_{MDD} and \hat{N}_{MDD} , respectively). Both of these eigensystems have the same eigenvalues, but the eigenvectors differ and are not necessarily aligned with each others. We note here that for a one-dimensional structure, both of these eigensystem have only one well-defined eigenvector, and that for a 1D current sheet structure, the well-defined eigenvectors of different bases are orthogonal to each others, as shown in Figure~??.

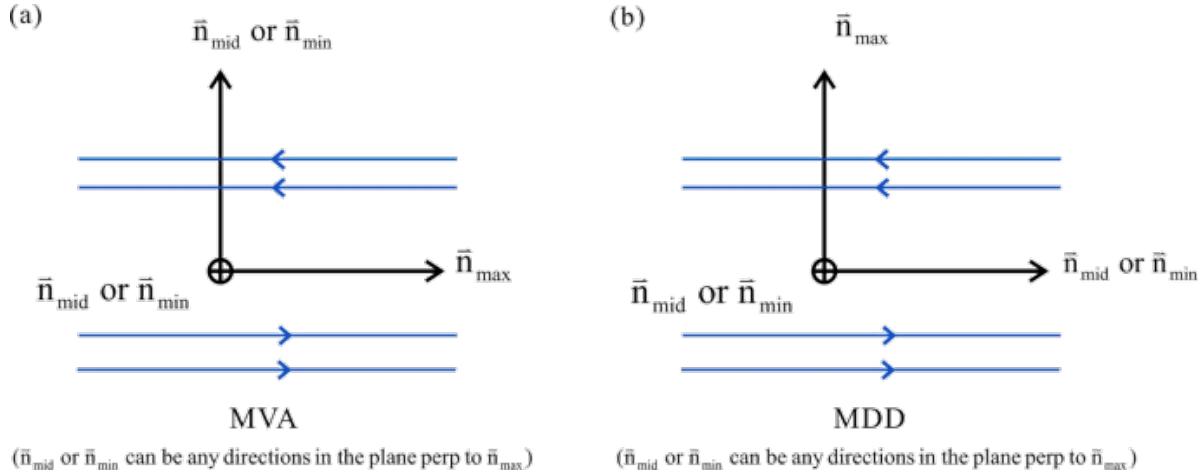


Figure 19.1: a) MVA (MGA) and b) MDD eigenvectors $\vec{n}_{\max}, \vec{n}_{\text{mid}}, \vec{n}_{\min}$, corresponding to largest, middle and minimum eigenvalues in a 1-D current sheet.

19.1.1 Dimensionality

MDD gives a way to define the local dimensionality of a structure from the eigenvalues of the matrix $\overleftrightarrow{GG}^\top$. The measures D_1, D_2, D_3 describing the one-, two- or three-dimensionality of the magnetic field and are obtained from the ratios of square roots of the eigenvalues λ_i ($i = 1, 2, 3$, in descending order) of the MDD $\overleftrightarrow{GG}^\top$.

One kind of definitions for D_1, D_2, D_3 are given below:

$$D_1 = \frac{\sqrt{\lambda_1} - \sqrt{\lambda_2}}{\sqrt{\lambda_1}}$$

$$D_2 = \frac{\sqrt{\lambda_2} - \sqrt{\lambda_3}}{\sqrt{\lambda_1}}$$

$$D_3 = \frac{\sqrt{\lambda_3}}{\sqrt{\lambda_1}}.$$

These quantities are defined to lie in the range [0,1] and their sum is one. For $D_1 \approx 1$, the magnetic field structure is primarily one-dimensional, such as a current sheet with $\mathbf{B} \approx \mathbf{B}(z)$ for a direction z normal to the current sheet. Correspondingly, for $D_2 \approx 1$, the structure is primarily a function of two coordinates, etc. These measures allow us to quantify whether or not the locally 2D treatment for neutral lines is well-founded.

If the eigenvalues are not well-separated, the directions obtained from MGA and MDD are ambiguous (Shi et al. 2019). For example, the current sheet with dimensionality 1, MGA

obtains the field-aligned direction (above and below the current sheet), while MDD obtains the normal direction to the current sheet, with two other directions being ambiguous.¹

Common representative structure:

- 1D: current sheet
- 2D: plasmoid, flux rope
- 3D: rare²

19.2 Reconnection Identification

Here we present analytical fields for an X-point configuration and an O-point configuration, respectively.

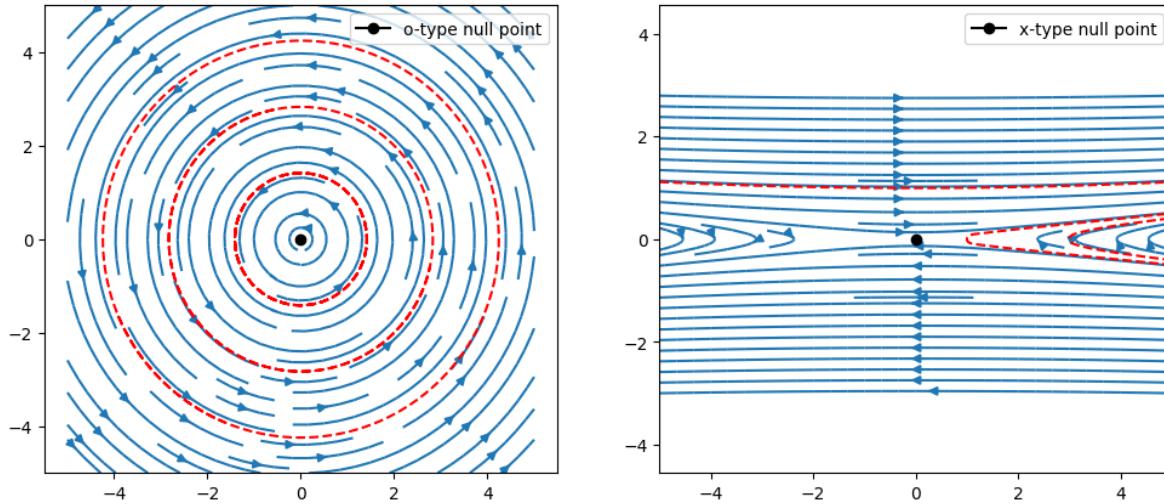


Figure 19.2: Example of (a) O-point and (b) X-point.

19.2.1 2D

Identification of 2D reconnection sites is easy. Assume the out-of-plane direction being y , with $B_y = \text{const}$. The divergence-free condition gives

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_z}{\partial z} = 0$$

¹I have not thought about Markku's statement here carefully. In his proposed method, the two L directions taken from both MDD and MGA are used to construct the local coordinate system.

²3D structures are rarely mentioned in literature. Markku also found it rare from Vlasior simulations.

In the X-Z plane, we can define a flux function

$$\psi = \int B_x dz = - \int B_z dx$$

then

$$\mathbf{B} = B_x \hat{x} + B_z \hat{z} = \hat{y} \times \nabla \psi$$

The null points are simple saddle points and extrema of the flux function.

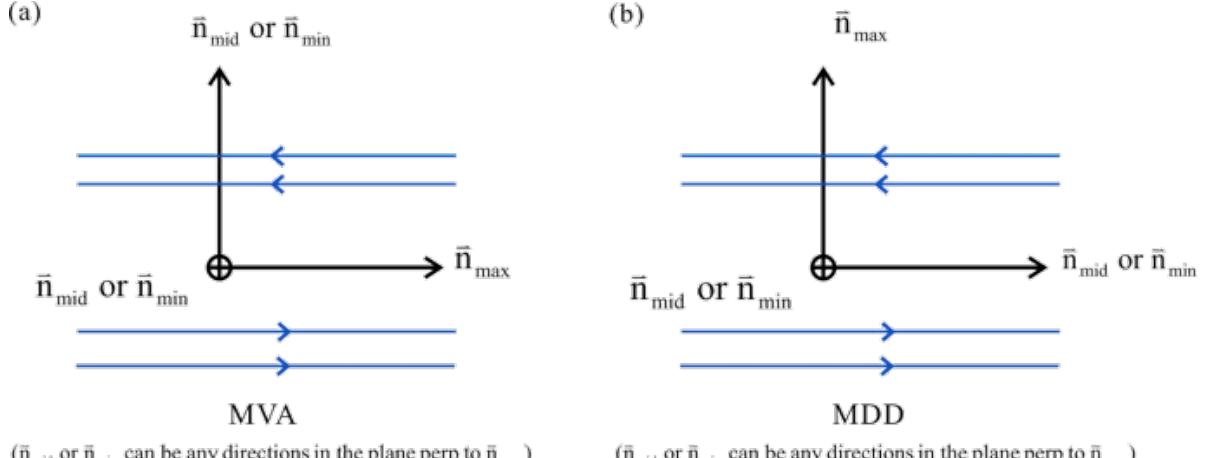


Figure 19.3: Neutral point classification in a well-defined local coordinate system. On the plane where $B_L = 0$ and on the subset of that plane where $B_N = 0$, the O-points have $\partial B_N / \partial L < 0$, and X-points $\partial B_N / \partial L > 0$. Note that this classification assumes a right-handed coordinate system.

19.2.2 3D

Identification of 3D reconnection sites is not easy.

Implement Lapenta's method.

The four-field junction (FFJ) method is proposed by Laitinen+ 2006. It works decently on steady reconnection cases, but not when things change drastically.

19.2.3 Critical Points

Let us discuss the problem with topological definitions (following Kenneth Rohde Christiansen and Aard Keimpema). Critical points are points where the vector vanishes (i.e. $\mathbf{v} = 0$). More rigidly speaking, critical points are stationary points that have a non-singular (i.e. no zero

eigenvalues) Jacobian or vector gradient $\nabla \mathbf{v}$. There are 6 different types of critical points, characterized by the behavior of nearby tangent curves. The position of these critical points can be found by searching all cells in the flow field. Critical points only occur in cells where all components of the vector pass through zero. In order to find the exact location of the critical point we will have to do interpolations.

We can then classify these by looking at the eigenvalues of the Jacobian matrix. The Jacobian matrix for a 2D vector (u, v) is given by

$$\left. \frac{\partial(u, v)}{\partial(x, y)} \right|_{x_0, y_0} = \begin{pmatrix} \partial_x u & \partial_y u \\ \partial_x v & \partial_y v \end{pmatrix} \Bigg|_{x_0, y_0} \quad (19.1)$$

As seen in Figure 19.4 these 6 different types are classified by the sign of the real and imaginary part of the eigenvalues. The real part of the eigenvalue gives rise to an attraction (if $RE < 0$) or a repulsion (if $RE > 0$). The imaginary part of the eigenvalue give rise to a rotation of vector field around the critical point. When we have an imaginary eigenvalue we are left with an entirely rotational vector field around the critical point.

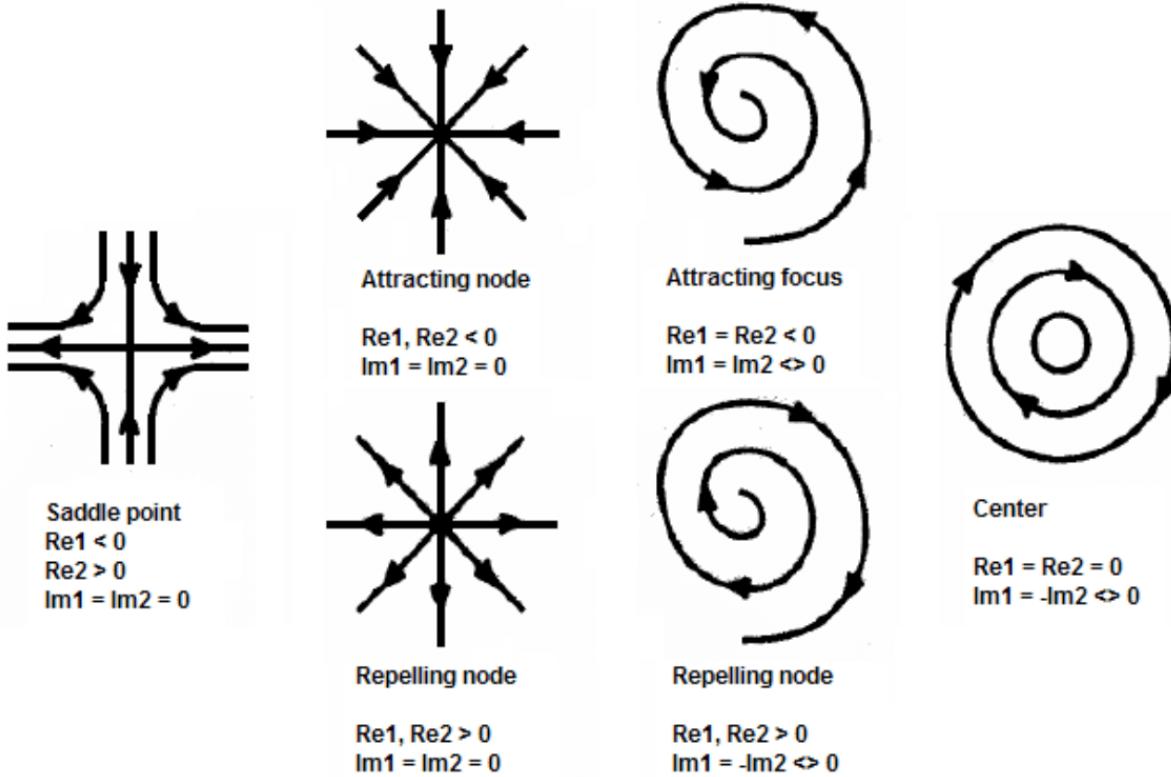


Figure 19.4: Classification of critical points, here $Re1/Re2$ is the real part of the first/second eigenvalue and $Im1/Im2$ is the imaginary part of the first/second eigenvalue.

The most important critical point is the *saddle point* here we have a combination of attraction in one direction and repulsion in the other. The importance of saddle points as we shall see in the next section is that the tangent curves near a critical point determine the global structure of the flow.

A critical point is called *hyperbolic* if it does not have any eigenvalues with zero real part. Hyperbolic critical points are structurally stable, i.e., a small perturbation does not change their topology.

Besides critical points there are also so-called attachment/detachment nodes. These are points on the wall of an object where tangent curves terminate or begin.

In 3D, the most common (i.e., nondegenerate) types of critical points are sources, sinks, repelling saddles, and attracting saddles, visualized in Figure 19.5 ((Bujack et al. 2021)). Each of these types may imply a rotating pattern in a certain plane depending on the presence of eigenvalues with nonzero imaginary part. This gives rise to the eight cases shown in the figure. A 2D separatrix and two 1D separatrices originate or terminate at a saddle (if the 2D separatrix originates at the saddle, then the 1D separatrices terminate at the saddle, and vice-versa). The directions along which the separatrices approach the saddle are given by the eigenvectors of ∇v . The 1D separatrices are computed by placing two seeds at a small user-defined offset of the saddle in the direction of the eigenvectors of ∇v corresponding to the eigenvalue whose sign appears only once. The trajectories from these seeds are integrated as explained above. The 2D separatrix is computed by placing eight seeds at a small user-defined offset of the saddle in the plane spanned by the eigenvectors of v corresponding to the eigenvalues whose signs appear twice. These seeds are the base for integrating a surface — the separatrix.

The most important input parameters are the distance in which the separatrices are seeded away from the saddles and the parameters that are handed off to the streamline and streamsurface integrators, like step sizes and maximum number of steps.

19.2.4 Separatrix

In scientific visualization, we treat vector fields defined on a discrete set of points in a grid and interpolate linearly or multilinearly within the cells. Since piecewise linear or multilinear functions are Lipschitz continuous, streamlines always exist uniquely.

A skeleton of a flow field is a figure containing the critical points of the flow field and the so-called integral curves connecting them. An integral curve can be thought of as the path a test particle takes when released infinitesimally close to a critical point. By adding e.g. an arrow the direction of the integral curves can be displayed; giving a global image of the flow. In the topology context, the skeleton is known as the **separatrix**. Separatrices are solutions which differ from their neighbors in their limit sets (?) or in their behavior in the large (?). Separatrices include critical points, periodic orbits, and invariant manifolds of saddles.

In practice, separatrices are constructed by

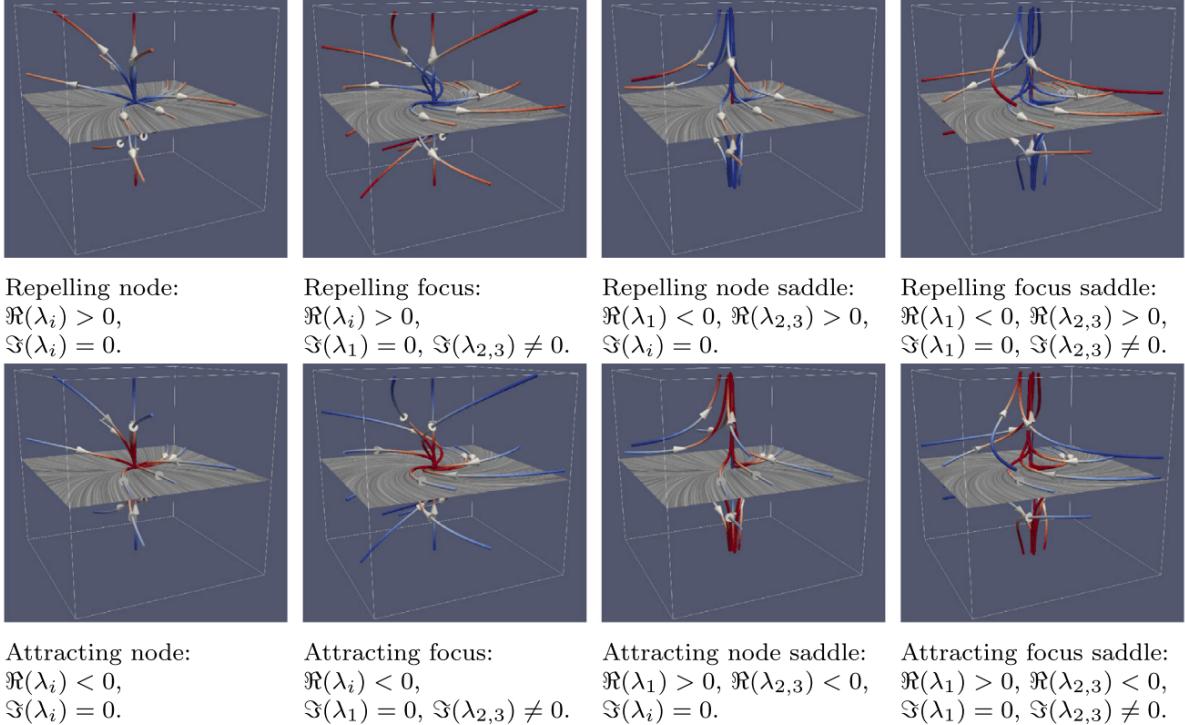


Figure 19.5: The different types of nondegenerate 3D critical points visualized with streamlines color coded in red/blue for forward/backward integration and line integral convolution in the plane of the common sign. Here the presence of the subscript i means all $i \in (1, 2, 3)$, and the order of the eigenvalues does not matter.

- Detecting critical points analytically in each cell
- Classifying them based on the eigenvalues of their Jacobian
- Seeding separatrices in a small offset of the saddles in the direction of their eigenvectors

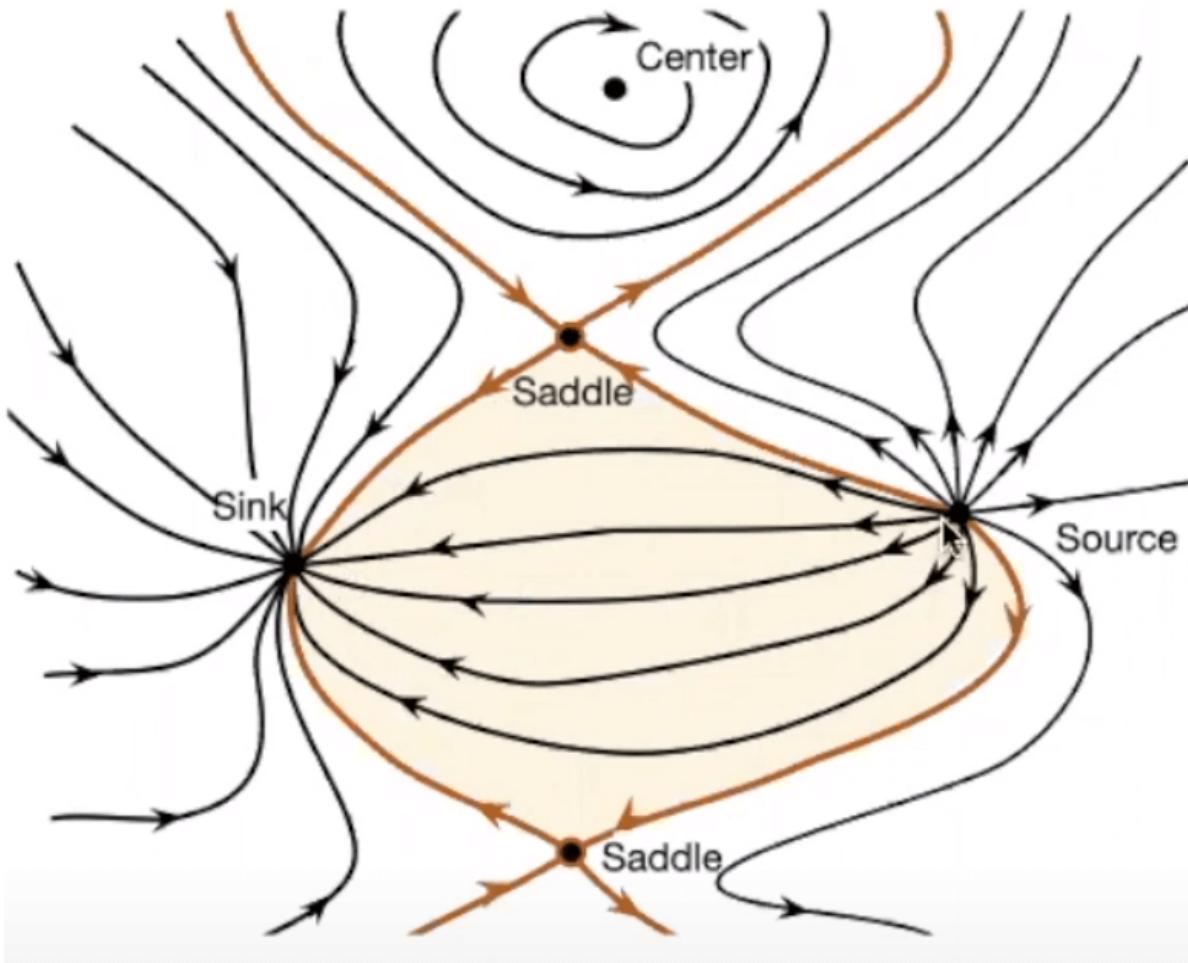


Figure 19.6: Identification of separatrices in 2D.

In 3D, instead of lines we may have separatrix surfaces.

19.3 Helicity

In fluid dynamics, helicity is, under appropriate conditions, an invariant of the Euler equations of fluid flow, having a topological interpretation as a measure of *linkage* and *knottedness* of vortex lines in the flow (Moffatt 1969).

Let $\mathbf{u}(x, t)$ be the velocity field and $\nabla \times \mathbf{u}$ the corresponding vorticity field. Under the following three conditions, the vortex lines are transported with (or “frozen in”) the flow:

1. the fluid is inviscid;
2. either the flow is incompressible ($\nabla \cdot \mathbf{u} = 0$), or it is compressible with a barotropic relation $p = p(\rho)$ between pressure p and density ρ ;
3. any body forces acting on the fluid are conservative. Under these conditions, any closed surface S on which $n \cdot (\nabla \times \mathbf{u}) = 0$ is, like vorticity, transported with the flow.

Let V be the volume inside such a surface. Then the helicity in V is defined by

$$H = \int_V \mathbf{u} \cdot (\nabla \times \mathbf{u}) dV \quad (19.2)$$

For a localised vorticity distribution in an unbounded fluid, V can be taken to be the whole space, and H is then the total helicity of the flow. H is invariant precisely because the vortex lines are frozen in the flow and their linkage and/or knottedness is therefore conserved, as recognized by Lord Kelvin (1868). Helicity is a pseudo-scalar quantity: it changes sign under change from a right-handed to a left-handed frame of reference; it can be considered as a measure of the handedness (or chirality) of the flow. Helicity is one of the four known integral invariants of the Euler equations; the other three are energy, momentum and angular momentum.

The invariance of helicity provides an essential cornerstone of the subject topological fluid dynamics and MHD, which is concerned with global properties of flows and their topological characteristics.

19.4 Magnetic Helicity

The helicity of a smooth vector field defined on a domain in 3D space is the standard measure of the extent to which the field lines wrap and coil around one another. As to magnetic helicity, this “vector field” is magnetic field. It is a generalization of the topological concept of [linking number](#) to the differential quantities required to describe the magnetic field. As with many quantities in electromagnetism, magnetic helicity for describing magnetic field lines is closely related to fluid mechanical helicity for describing fluid stream lines.

If magnetic field lines follow the strands of a twisted rope, this configuration would have nonzero magnetic helicity; left-handed ropes would have negative values and right-handed ropes would have positive values.

Formally,

$$H = \int \mathbf{A} \cdot \mathbf{B} dr^3 \quad (19.3)$$

where

- H is the helicity of the entire magnetic field
- \mathbf{B} is the magnetic field strength
- \mathbf{A} is the vector potential of \mathbf{B} and $\mathbf{B} = \nabla \times \mathbf{A}$

Magnetic helicity has units of Wb^2 in SI units and Mx^2 in Gaussian Units. Note that $\mathbf{A} \cdot \mathbf{B}$ should not be considered as “helicity density” because of gauge freedom.

It is a conserved quantity in electromagnetic fields, even when magnetic reconnection dissipates energy (Woltjer 1958). The concept is useful in solar dynamics and in dynamo theory. Helicity is approximately conserved during magnetic reconnection and topology changes. Helicity can be injected into a system such as the solar corona. When too much builds up, it ends up being expelled through coronal mass ejections. The simple proof in ideal MHD can be found on Wikipedia.

Magnetic helicity is a gauge-dependent quantity, because \mathbf{A} can be redefined by adding a gradient to it (gauge transformation):

$$\mathbf{A}' = \mathbf{A} + \nabla\phi$$

However, for perfectly conducting boundaries or periodic systems without a net magnetic flux, the magnetic helicity is gauge invariant. A gauge-invariant relative helicity has been defined for volumes with non-zero magnetic flux on their boundary surfaces. If the magnetic field is turbulent and weakly inhomogeneous a magnetic helicity density and its associated flux can be defined in terms of the density of field line linkages.

The topological properties of a magnetic field are interpreted in terms of magnetic helicity. The total helicity of a collection of flux tubes arises from the linking of flux tubes with one another (mutual helicity) and the internal magnetic structure of each flux tube (self-helicity). Reconnection changes the topology and magnetic connectivity of flux tubes. This can also be viewed as a redistribution of self- and mutual helicities. If total magnetic helicity is approximately conserved, it is possible to put quantitative limits upon the changes in self- and mutual helicities. This can be interpreted as the change in magnetic flux tube linkage (due to reconnection) and amount of twist present in the reconnected flux tubes. [Wright & Berger, 1989]

Simple examples:

1. A single untwisted closed flux loop has $H = 0$.
2. A single flux rope with a magnetic flux of ϕ that twists around itself T times has a helicity of $H = T\phi^2$.
3. Two interlinked untwisted flux loops with fluxes ϕ_1 and ϕ_2 have $H = \pm 2\phi_1\phi_2$ where the sign depends on the sense of the linkedness.

There are generalizations to allow for gauge-invariant definitions of helicity. [Berger & Field (1984)] defined the relative magnetic helicity to be

$$H = \int_V \mathbf{A} \cdot \mathbf{B} - \mathbf{A}_0 \cdot \mathbf{B}_0 dV \quad (19.4)$$

where $\mathbf{B}_0 = \nabla \times \mathbf{A}_0$ is the potential field inside V with the same field outside of V (see also Finn & Antonsen 1985).

In toroidal laboratory experiments, it is natural to consider the volume contained within conducting wall boundaries that are coincident with closed flux surfaces (i.e., the magnetic field along the wall is parallel to the boundary).

The time evolution of magnetic helicity is given by

$$\frac{dH}{dt} = -2c \int_V \mathbf{E} \cdot \mathbf{B} dV + 2c \int_S \mathbf{A}_p \times \mathbf{E} \cdot d\mathbf{S}$$

where we choose $\nabla \times \mathbf{A}_p = 0$ and $\mathbf{A}_p \cdot d\mathbf{S} = 0$ on S . The first term represents helicity dissipation when $E_{\parallel} \neq 0$, which is always zero in ideal MHD. The second term represents helicity fluxes in and out of the system, for example, flux emergence from the solar photosphere corresponds to helicity injection in the corona.

19.5 Cross Helicity

The cross helicity measures the imbalance between interacting waves, which is important in MHD turbulence (Section 18.2.9). It is given by

$$H_C = \int_V \mathbf{v} \cdot \mathbf{B} dV \quad (19.5)$$

In ideal MHD, the rate of change of H_C is

$$\frac{dH_C}{dt} = - \oint_S d\mathbf{S} \cdot \left[\left(\frac{1}{2} v^2 + \frac{\gamma}{\gamma-1} \frac{p}{\rho} \right) \mathbf{B} - \mathbf{v} \times (\mathbf{v} \times \mathbf{B}) \right]$$

This vanishes when $d\mathbf{S} \cdot \mathbf{B} = d\mathbf{S} \cdot \mathbf{V} = 0$ along the boundary S or when the boundary conditions are periodic. Cross helicity is an ideal MHD invariant when this integral vanishes.

There are discretized forms of cross helicity from the observation point of view. Check it out if you want to know more.

19.6 Flux Rope Identification

Using turbulence parameters to find flux ropes (Zhao et al. 2020):

$$\begin{aligned}\sigma_r &< -0.5 \\ |\sigma_c| &< 0.4 \\ |\sigma_m| &> 0.7\end{aligned}$$

where σ_m is the *normalized reduced magnetic helicity*, which is a measure of \mathbf{B} rotation. Strictly speaking, the magnetic helicity depends on the spatial properties of the magnetic field topology, and thus cannot be directly evaluated from single spacecraft measurements. However, Matthaeus+ (1982) described a reduced form of magnetic helicity that can be estimated with measurements from a single spacecraft based on the magnetic power spectrum. The normalized reduced magnetic helicity can be estimated by

$$\sigma_m(\nu, t) = \frac{2\Im[W_T^*(\nu, t) \cdot W_N(\nu, t)]}{|W_R(\nu, t)|^2 + |W_T(\nu, t)|^2 + |W_N(\nu, t)|^2} \quad (19.6)$$

where ν is the frequency associated with the Wavelet function and the sampling period of the measured magnetic field in the radial tangential normal (RTN) coordinate system. The spectra $W_R(\nu, t)$, $W_T(\nu, t)$, and $W_N(\nu, t)$ are the wavelet transforms of time series of \mathbf{B}_{1R} , \mathbf{B}_{1T} , and \mathbf{B}_{1N} , respectively; and $W_T^*(\nu, t)$ is the conjugate of $W_T(\nu, t)$. From the resulting spectrogram of the magnetic helicity, σ_m , one can determine both the magnitude and the handedness (chirality) of underlying fluctuations at a specific scale. A positive value of σ_m corresponds to right-handed chirality and a negative value to left-handed chirality.

19.7 Magnetopause Identification

From pressure balance argument, let $\beta^* = (p_{\text{th}} + p_{\text{dyn}})/p_B$, we can have a simple criterion:

$$\beta^* \simeq 1$$

20 Solar Wind

The *solar wind* is a high-speed particle stream continuously blown out from the Sun into interplanetary space. It extends far beyond the orbit of the Earth, and terminates in a shock front, called the heliopause, where it interfaces with the weakly ionized interstellar medium. The heliopause is predicted to lie between 110 and 160 AU (1 Astronomical Unit is 1.5×10^{11} m) from the center of the Sun. Voyager 1 & 2, which were launched in 1977, have passed through the heliopause, and are still functional!

In the vicinity of the Earth, (i.e., at about 1 AU from the Sun) the solar wind velocity typically ranges between 300 and 1400 km s⁻¹. The average value is approximately 500 km s⁻¹, which corresponds to about a 4 day time of flight from the Sun. Note that the solar wind is both *super-sonic* and *super-Alfvénic*.

The solar wind is predominately composed of protons and electrons. Generally helium (He⁺⁺) is about 4%, and oxygen (O⁺) is less than 1%.

Amazingly enough, the solar wind was predicted theoretically by Eugine Parker in 1958, a number of years before its existence was confirmed using satellite data. Parker's prediction of a super-sonic outflow of gas from the Sun is a fascinating scientific detective story, as well as a wonderful application of plasma physics.

The solar wind originates from the solar corona. The solar corona is a hot, tenuous plasma surrounding the Sun, with characteristic temperatures and particle densities of about 10⁶K and 10¹⁴ m⁻³, respectively. Note that the corona is *far hotter* than the solar atmosphere, or photosphere. In fact, the temperature of the photosphere is only about 6000 K. It is thought that the corona is heated by Alfvén waves emanating from the photosphere together with a turbulent cascading process. The solar corona is most easily observed during a total solar eclipse, when it is visible as a white filamentary region immediately surrounding the Sun.

Let us start, following Chapman, by attempting to construct a model for a static solar corona. The equation of hydrostatic equilibrium for the corona takes the form

$$\frac{dp}{dr} = -\rho \frac{GM_{\odot}}{r^2} \quad (20.1)$$

where $G = 6.67 \times 10^{-11}$ m³ s⁻² kg⁻¹ is the gravitational constant, and $M_{\odot} = 2 \times 10^{30}$ kg is the solar mass. The plasma density is written

$$\rho \simeq n m_p \quad (20.2)$$

where n is the number density of protons. If both protons and electrons are assumed to possess a common temperature, $T(r)$, then the coronal pressure is given by

$$p = 2nk_B T \quad (20.3)$$

The thermal conductivity of the corona is dominated by the electron thermal conductivity, and takes the form (???)

$$\kappa = \kappa_0 T^{5/2}$$

where κ_0 is a relatively weak function of density and temperature. For typical coronal conditions this conductivity is extremely high: i.e., it is about twenty times the thermal conductivity of copper at room temperature. The coronal heat flux density is written

$$\mathbf{q} = -\kappa \nabla T$$

For a static corona, in the absence of energy sources or sinks, we require

$$\nabla \cdot \mathbf{q} = 0$$

Assuming spherical symmetry, this expression reduces to

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \kappa_0 k_B T^{5/2} \frac{dT}{dr} \right) = 0$$

Adopting the sensible boundary condition that the coronal temperature must tend to zero at large distances from the Sun, we obtain

$$T(r) = T(a) \left(\frac{a}{r} \right)^{2/7} \quad (20.4)$$

The reference level $r = a$ is conveniently taken to be the base of the corona, where $a \sim 7 \times 10^5$ km, $n \sim 2 \times 10^{14}$ m⁻³, and $T \sim 2 \times 10^6$ K.

Equation 20.1, Equation 20.2, Equation 20.3, and Equation 20.4 can be combined and integrated to give

$$p(r) = p(a) \exp \left\{ \frac{7}{5} \frac{G M_\odot m_p}{2k_B T(a) a} \left[\left(\frac{a}{r} \right)^{5/7} - 1 \right] \right\}$$

Note that as $r \rightarrow \infty$ the coronal pressure tends towards a finite constant value:

$$p(\infty) = p(a) \exp \left\{ -\frac{7}{5} \frac{G M_\odot m_p}{2k_B T(a) a} \right\}$$

There is, of course, nothing at large distances from the Sun which could contain such a pressure (the pressure of the interstellar medium is negligibly small). Thus, we conclude, with Parker, that the static coronal model is unphysical.

Since we have just demonstrated that a static model of the solar corona is unsatisfactory, let us now attempt to construct a dynamic model in which material flows outward from the Sun.

20.1 Parker Solar Wind Model

By symmetry, we expect a purely radial coronal outflow. The radial momentum conservation equation for the corona takes the form

$$\rho u \frac{du}{dr} = -\frac{dp}{dr} - \rho \frac{GM_{\odot}}{r^2} \quad (20.5)$$

where u is the radial expansion speed. The continuity equation reduces to

$$\frac{1}{r^2} \frac{d(r^2 \rho u)}{dr} = 0 \quad (20.6)$$

In order to obtain a closed set of equations, we now need to adopt an equation of state for the corona, relating the pressure, p , and the density, ρ . For the sake of simplicity, we adopt the simplest conceivable equation of state, which corresponds to an *isothermal* corona. Thus, we have

$$p = \frac{2k_B \rho T}{m_p} \quad (20.7)$$

where T is a constant. Note that more realistic equations of state complicate the analysis, but do not significantly modify any of the physics results.

Equation 20.6 can be integrated to give

$$r^2 \rho u = I \quad (20.8)$$

where I is a constant. The above expression simply states that the mass flux per unit solid angle, which takes the value I , is independent of the radius, r . Equation 20.5, Equation 20.7, and Equation 20.8 can be combined together to give

$$\frac{1}{u} \frac{du}{dr} \left(u^2 - \frac{2k_B T}{m_p} \right) = \frac{4k_B T}{m_p r} - \frac{GM_{\odot}}{r^2} \quad (20.9)$$

Let us restrict our attention to coronal temperatures which satisfy

$$T < T_c \equiv \frac{GM_{\odot} m_p}{4k_B a}$$

where a is the radius of the base of the corona. T_c is the defined temperature where the right-hand side of Equation 20.9 is zero at $r = a$. For typical coronal parameters (see previous section), $T_c \simeq 5.8 \times 10^6 \text{ K}$, which is certainly greater than the temperature of the corona at $r = a$. For $T < T_c$, the right-hand side of Equation 20.9 is negative for $a < r < r_c$, where (???)

$$\frac{r_c}{a} = \frac{T_c}{T} \quad (20.10)$$

and positive for $r_c < r < \infty$. The right-hand side of Equation 20.9 is zero at $r = r_c$, implying that the left-hand side is also zero at this radius, which is usually termed the “critical radius”. There are two ways in which the left-hand side of Equation 20.9 can be zero at the critical radius. Either

$$u^2(r_c) = u_c^2 \equiv \frac{2k_B T}{m_p} \quad (20.11)$$

or

$$\frac{du(r_c)}{dr} = 0 \quad (20.12)$$

Note that u_c is the coronal sound speed.

As is easily demonstrated, if Equation 20.11 is satisfied then du/dr has the same sign for all r , and $u(r)$ is either a monotonically increasing, or a monotonically decreasing, function of r . On the other hand, if Equation 20.12 is satisfied then $u^2 - u_c^2$ has the same sign for all r , and $u(r)$ has an extremum close to $r = r_c$. The flow is either super-sonic for all r , or sub-sonic for all r . These possibilities lead to the existence of four classes of solutions to Equation 20.9, with the following properties:

1. $u(r)$ is sub-sonic throughout the domain $a < r < \infty$. $u(r)$ increases with r , attains a maximum value around $r = r_c$, and then decreases with r .
2. a unique solution for which $u(r)$ increases monotonically with r , and $u(r_c) = u_c$.
3. a unique solution for which $u(r)$ decreases monotonically with r , and $u(r_c) = u_c$.
4. $u(r)$ is super-sonic throughout the domain $a < r < \infty$. $u(r)$ decreases with r , attains a minimum value around $r = r_c$, and then increases with r .

These four classes of solutions are illustrated in [?@fig-parker-sol](#).

Each of the classes of solutions described above fits a different set of boundary conditions at $r = a$ and $r \rightarrow \infty$. The physical acceptability of these solutions depends on these boundary conditions. For example, both Class 3 and Class 4 solutions can be ruled out as plausible models for the solar corona since they predict super-sonic flow at the base of the corona, which is not observed, and is also not consistent with a static solar photosphere. Class 1 and Class 2 solutions remain acceptable models for the solar corona on the basis of their properties around $r = a$, since they both predict sub-sonic flow in this region. However, the Class 1 and Class 2 solutions behave quite differently as $r \rightarrow \infty$, and the physical acceptability of these two classes hinges on this difference.

Equation 20.9 can be rearranged to give

$$\frac{du^2}{dr} \left(1 - \frac{u_c^2}{u^2} \right) = \frac{4u_c^2}{r} \left(1 - \frac{r_c}{r} \right)$$

where Equation 20.10 and the definition of T_c have been used. The above expression can be integrated to give

$$\left(\frac{u}{u_c} \right)^2 - \ln \left(\frac{u}{u_c} \right)^2 = 4 \ln \frac{r}{r_c} + 4 \frac{r_c}{r} + C \quad (20.13)$$

where C is a constant of integration.

Let us consider the behaviour of Class 1 solutions in the limit $r \rightarrow \infty$. It is clear from [?@fig-parker-sol](#) that, for Class 1 solutions, u/u_c is less than unity and monotonically decreasing as $r \rightarrow \infty$. In the large- r limit, Equation 20.13 reduces to

$$\ln \frac{u}{u_c} \simeq -2 \ln \frac{r}{r_c}$$

so that

$$u \propto \frac{1}{r^2}$$

It follows from Equation 20.8 that the coronal density, ρ , approaches a finite, constant value, ρ_∞ , as $r \rightarrow \infty$. Thus, the Class 1 solutions yield a finite pressure,

$$p_\infty = \frac{2k_B \rho_\infty T}{m_p}$$

at large r , which cannot be matched to the much smaller pressure of the interstellar medium. Clearly, Class 1 solutions are unphysical.

Let us consider the behaviour of the Class 2 solution in the limit $r \rightarrow \infty$. It is clear from [?@fig-parker-sol](#) that, for the Class 2 solution, u/u_c is greater than unity and monotonically increasing as $r \rightarrow \infty$. In the large- r limit, Equation 20.13 reduces to

$$\left(\frac{u}{u_c} \right)^2 \simeq 4 \ln \frac{r}{r_c}$$

so that

$$u \simeq 2 u_c \left(\ln \frac{r}{r_c} \right)^{1/2}$$

It follows from Equation 20.8 that $\rho \rightarrow 0$ and $r \rightarrow \infty$. Thus, the Class 2 solution yields $p \rightarrow 0$ at large r , and can, therefore, be matched to the low pressure interstellar medium.

We conclude that the only solution to Equation 20.9 which is consistent with physical boundary conditions at $r = a$ and $r \rightarrow \infty$ is the Class 2 solution. This solution predicts that the solar corona expands radially outward at relatively modest, sub-sonic velocities close to the Sun, and gradually accelerates to super-sonic velocities as it moves further away from the Sun. Parker termed this continuous, super-sonic expansion of the corona the *solar wind*.

Note that while the velocity prediction accords well with satellite observations, the Parker model's prediction for the density of the solar wind at the Earth is significantly too high compared to satellite observations. Consequently, there have been many further developments of this model. In particular, the unrealistic assumption that the solar wind plasma is isothermal has been relaxed, and two-fluid effects have been incorporated into the analysis.

20.2 Parker Spiral

The *interplanetary magnetic field* (IMF), also called *Parker spiral*, is the component of the solar magnetic field that is dragged out from the solar corona by the solar wind flow to fill the Solar System. Depending on the polarity of the photospheric footpoint, the heliospheric magnetic field spirals inward or outward; the magnetic field follows the same shape of spiral in the northern and southern parts of the heliosphere, but with opposite field direction. These two magnetic domains are separated by a current sheet (an electric current that is confined to a curved plane). This heliospheric current sheet has a shape similar to a twirled ballerina skirt (Figure 20.1), and changes in shape through the solar cycle as the Sun's magnetic field reverses about every 11 years.

```
KeyNotes.plot_IMF_ecliptic()
```

```
KeyNotes.plot_IMF_3D()
```

20.3 Anisotropy in the Solar Wind

Observationally, Pioneer 6 showed that the ion temperature anisotropy in the solar wind at 1AU generally has $T_{\parallel} > T_{\perp}$, together with other two interesting discoveries:

1. high fluctuations of flow velocity outside the solar ecliptic plane;
2. anisotropic ion thermal distribution ($T_{\parallel}/T_{\perp} \sim [2, 5]$);
3. presence of a 3rd species, helium, from charge-to-mass ratio analysis of the angular and energy distributions.

It may possibly be explained by the conservation of the 1st adiabatic invariant Scarf, Wolfe, and Silva (1967).

The magnetic moment $\mu = mv_{\perp}^2/2B$ is conserved as the collisionless solar wind flows outward from the sun. Near the solar equator the mean field magnitude declines with

$$B_r(r) \simeq B_r(r_0) \left(\frac{r_0}{r} \right)^2$$

and

$$B_{\phi}(r) \simeq \frac{\Omega_r}{u(r)} B_r(r)$$

from the Parker spiral solar wind model and $\Omega_r = 2.94 \times 10^{-6}$ rad/s being the angular frequency of the rotation of the sun.

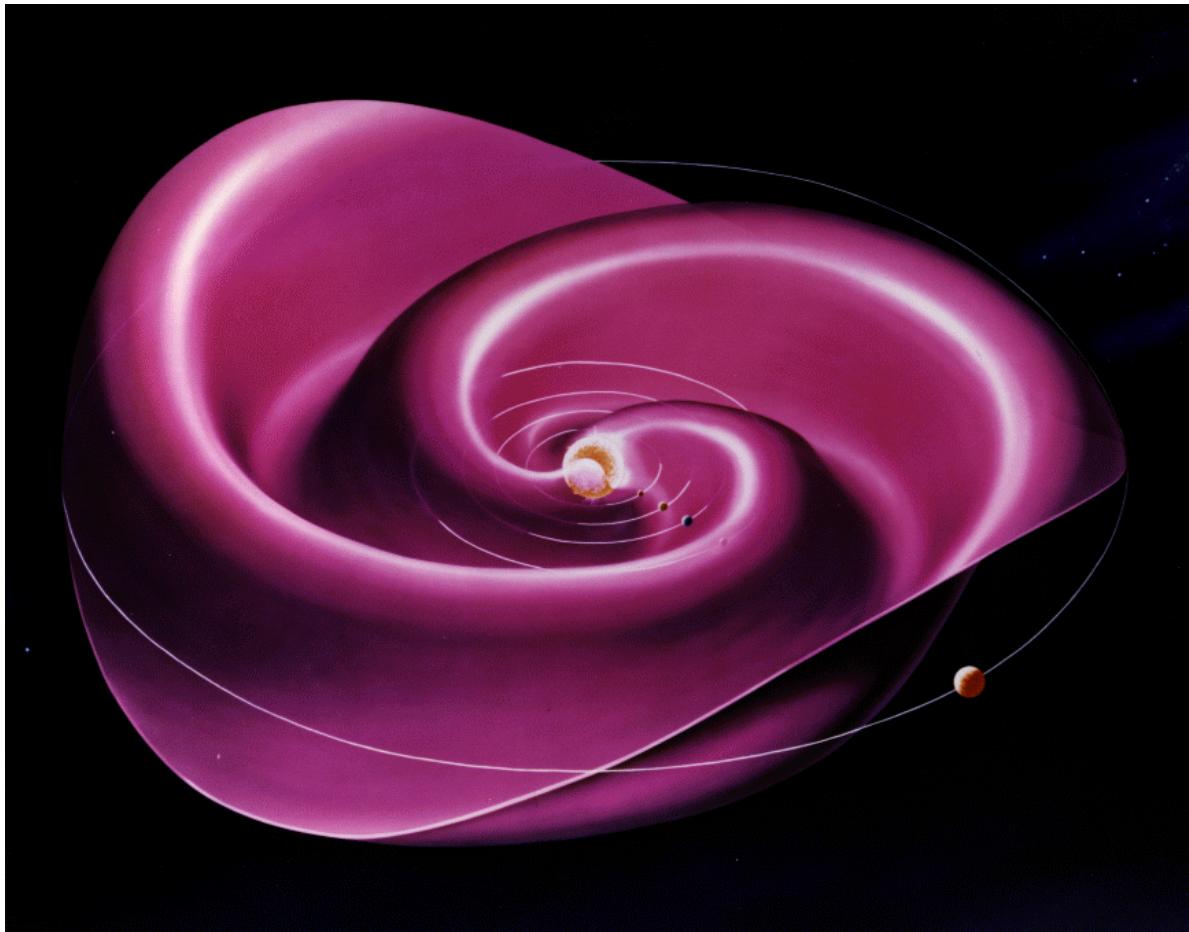


Figure 20.1: The heliospheric current sheet is a three-dimensional form of a Parker spiral that results from the influence of the Sun's rotating magnetic field on the plasma in the interplanetary medium.

The adiabatic equation in the perpendicular direction indicates that the perpendicular thermal energy $\langle mv_{\perp}^2/2 \rangle = k_B T_{\perp}$ declines with B . Assuming in the rest frame the distribution function is a bi-Maxwellian of the form

$$f(v) = \left(\frac{m}{2\pi}\right)^{3/2} \frac{1}{k_B T_{\perp} (k_B T_{\parallel})^{1/2}} \exp\left(-\frac{mv_{\perp}^2}{2k_B T_{\perp}} - \frac{mv_{\parallel}^2}{2k_B T_{\parallel}}\right)$$

The conservation of the total thermal energy

$$W = \int d^3v \frac{mv^2}{2} f(v)$$

yields

$$W = k_B T_{\perp} + k_B T_{\parallel}/2 = \text{const.}$$

These allows us to evaluate the variations in T_{\perp} and T_{\parallel} originating from isotropic distribution on the surface of the sun. Starting from $T_{\perp}(0.3AU) = T_{\parallel}(0.3AU) \simeq 1.3 \times 10^5 \text{ K}$, the predicted anisotropy T_{\parallel}/T_{\perp} at Earth can go beyond 20! Therefore, in fact, the reasonable question to ask is why the actual solar wind anisotropy factor is so small. In simulations, we always apply isotropic distribution in the upstream solar wind condition, which is primarily due to the fact that we are mostly using fluid models (i.e. MHD) for global magnetosphere-solar wind interactions. For kinetic models, *we need more realistic distribution setups.*

Well, we know now when pressure anisotropy develops, two kinds of plasma instabilities can be triggered: firehose when $T_{\perp} < T_{\parallel}$ and mirror when $T_{\perp} > T_{\parallel}$. Further studies require kinetic theory to describe their behaviors.

On the other hand, the opposite case, $T_{\perp} > T_{\parallel}$, is also observed and believed to be related to local ion heating by macroscale compressions (e.g. high/low speed streams interaction) or plasma instabilities (Bame et al. 1975).

The mirror instability criterion as an additional relation to determine the pressure anisotropy downstream of the shock from the book Plasma instabilities and nonlinear effects by Hasegawa 1975,

$$1 + \sum_{\text{species}} \beta_{\perp} \left(1 - \frac{\beta_{\perp}}{\beta_{\parallel}}\right) < 0$$

20.4 Switchbacks

When Parker Solar Probe (PSP) sent back the first observations from its voyage to the Sun, scientists found signs of a wild ocean of currents and waves quite unlike the near-Earth space much closer to our planet. This ocean was spiked with what became known as *switchbacks*: rapid flips in the Sun's magnetic field that reversed direction like a zig-zagging mountain road.

Our current understanding is that switchbacks may be generated from either magnetic reconnection, turbulence, or plasma velocity shears. There are models for each, but they all require more testing from future PSP data.

See more in [Switchbacks Science: Explaining Parker Solar Probe's Magnetic Puzzle](#).

20.5 MHD Description

Because of the presence of a strong magnetic field carried by the wind, low-frequency fluctuations in the solar wind are usually described with MHD. However, due to some peculiar characteristics, the solar wind turbulence contains some features hardly classified within a general theoretical framework. (Tu and Marsch 1995) presents a thorough review of the turbulent phenomena in the solar wind from observations in the ecliptic. In the 1990s, with the launch of the Ulysses spacecraft, investigations have been extended to the high-latitude regions of the heliosphere.

Ruelle and Takens (1971) who showed that a strange attractor in the phase space of the system is the best model for the birth of turbulence. This gave a strong impulse to the investigation of the phenomenology of turbulence from the point of view of dynamical systems. Turbulence in the solar wind has been used as a big wind tunnel to investigate scaling laws of turbulent fluctuations, multifractals models, etc. Therefore, the solar wind can be seen as a very big laboratory where fully developed turbulence can be investigated not only per se, rather as far as basic theoretical aspects are concerned.

20.6 Wave Modes in the Solar Wind

Observations in the solar wind suggest that the compressive component of inertial-range solar-wind turbulence is dominated by slow modes. The low collisionality of the solar wind allows for nonthermal features to survive, which suggests the requirement of a kinetic plasma description.

The role of kinetic slow waves (KSWs) have largely been ignored as it is heavily damped for the typical plasma parameters of the solar wind (Barnes, 1966; Narita & Marsch, 2015). Recently, KSWs (or pressure balanced structures) have often been invoked to explain the compressible fluctuations and the anticorrelation of magnetic field strength B and the density n at fluid scales (Howes et al. 2012; Verscharen, Chen, and Wicks 2017) and have also been revisited at sub-ion scales to explain the observed increase in compressibility (Lacombe, Alexandrova, and Matteini 2017).

21 Shock

Shock is a phenomenon that appears when the speed of an object exceeds the characteristic wave speeds. It is characterized by Mach number, the ratio of the shock speed to the characteristic wave speed in the medium. Consider a subsonic disturbance moving through a conventional neutral fluid. As is well-known, sound waves propagating ahead of the disturbance give advance warning of its arrival, and, thereby, allow the response of the fluid to be both smooth and adiabatic. Now, consider a supersonic disturbance. In this case, sound waves are unable to propagate ahead of the disturbance, and so there is no advance warning of its arrival, and, consequently, the fluid response is *sharp* and *non-adiabatic*. This type of response is generally known as a *shock*.

Intuitively, the number of different types of shocks depends on the types of waves in the system. In plasma physics, the simplest system for investigating shocks is ideal MHD. Since information in MHD fluids is carried via three different waves – namely, *fast* or *compressional-Alfvén waves*, *intermediate* or *shear-Alfvén waves*, and *slow* or *magnetosonic waves* – we might expect MHD fluids to support three different types of shocks, corresponding to disturbances traveling faster than each of the aforementioned waves.

In general, a shock propagating through an MHD fluid produces a significant difference in plasma properties on either side of the shock front. The thickness of the front is determined by a balance between *convective* and *dissipative* effects. However, dissipative effects in high temperature plasmas are only comparable to convective effects when the spatial gradients in plasma variables become extremely large. Hence, MHD shocks in such plasmas tend to be *extremely narrow*, and are well-approximated as *discontinuities* in plasma parameters. The MHD equations, and Maxwell's equations, can be integrated across a shock to give a set of jump conditions which relate plasma properties on each side of the shock front. If the shock is sufficiently narrow then these relations become independent of its detailed structure. We will derive the jump conditions for a narrow, planar, steady-state, MHD shock in Section 21.1.

The realization of extreme sharpness of a collisionless shock like the Earth's bow shock immediately posed a serious problem for the MHD description of collisionless shocks. In collisionless MHD there is no known dissipation mechanism that could lead to the observed extremely short transition scales Δ in high Mach number flows which are comparable to the ion gyro-radius r_{Li} . MHD neglects any differences in the properties of electrons and ions and thus barely covers the very physics of shocks on the observed scales. In the MHD frame, shocks are considered as infinitely narrow discontinuities, narrower than the MHD flow scales $L \gg \Delta \gg \lambda_d$; on the other hand, these discontinuities must physically be much wider than

the dissipation scale λ_d with all the physics going on inside the shock transition. This implies that the conditions derived from collisionless MHD just hold far upstream and far downstream of the shock transition, i.e. far outside the region where the shock interactions are going on. In describing shock waves, collisionless MHD must be used in an asymptotic sense, providing the remote boundary conditions on the shock transition. One must look for processes different from MHD in order to arrive at a description of the processes leading to shock formation and shock dynamics and the structure of the shock transition. In fact, from the MHD single-fluid viewpoint, the shock should not be restricted to the steep shock front, it rather includes the entire shock transition region from outside the foreshock across the shock front down to the boundary layer at the surface of the obstacle. And this holds as well even in two-fluid shock theory that distinguishes between the behaviour of electrons and ions in the plasma fluid. The issue of dissipation is partly indicated in Section 21.1.3.

The basic process of shock formation is the growth of a small disturbance in the plasma by the action of the intrinsic *nonlinearity* of flow, independent of the cause of the initial disturbance. The latter can be an external driver like a piston or a blast, it can also be an internal instability. Shocks form when nonlinearity causes steepening of the disturbance in space and some processes exists which prevents breaking of the steep wave. Such processes are of *dissipative* or *dispersive* nature and are discussed in ascending importance.

It is important to emphasize that the various modes of waves are responsible for the generation of anomalous dissipation, shock ramp broadening, generation of turbulence in the shock environment and shock ramp itself, as well as for particle acceleration, shock particle reflection and the successive effects. The idea is that in a plasma that consists of electrodynamically active particles the excitation of the various plasma wave modes in the EM field as collective effects is the easiest way of energy distribution and transport. There is very little momentum needed in order to accelerate a wave, even though many particles are involved in the excitation and propagation of the wave, much less momentum than accelerating a substantial number of particles to medium energy. Therefore any more profound understanding of shock processes cannot avoid bothering with waves, instabilities, wave excitation and wave-particle interaction.

As a quick summary, in some special cases, for example, shock waves in an ideal neutral gas, the global behavior does not depend on the details of the small-scale physics, because the jump conditions across a hydrodynamic shock are fully determined by the conservation of mass, momentum, and energy. For more complicated systems, such as magnetohydrodynamics with anisotropic ion pressure, the conservation laws constrain the jump conditions, but the pressure anisotropy behind the shock cannot be determined without knowledge of small-scale processes.

A general review of collisionless shocks is given by Balogh and Treumann (2013).

21.1 MHD Theory

The conserved form of MHD equations can be written as:

$$\begin{aligned}\nabla \cdot \mathbf{B} &= 0 \\ \frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{U} \times \mathbf{B}) &= 0 \\ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{U}) &= 0 \\ \frac{\partial(\rho \mathbf{U})}{\partial t} + \nabla \cdot \mathbf{S} &= 0 \\ \frac{\partial K}{\partial t} + \nabla \cdot \mathbf{w} &= 0\end{aligned}\tag{21.1}$$

where

$$\mathbf{S} = \rho \mathbf{U} \mathbf{U} + \left(p + \frac{B^2}{2\mu_0} \right) \mathbf{I} - \frac{\mathbf{B} \mathbf{B}}{\mu_0}\tag{21.2}$$

is the total (i.e., including electromagnetic, as well as plasma, contributions) stress tensor¹, \mathbf{I} the identity tensor,

$$K = \frac{1}{2} \rho U^2 + \frac{p}{\gamma - 1} + \frac{B^2}{2\mu_0}\tag{21.3}$$

the total energy density, and

$$\mathbf{w} = \left(\frac{1}{2} \rho U^2 + \frac{\gamma}{\gamma - 1} p \right) \mathbf{U} + \frac{\mathbf{B} \times (\mathbf{U} \times \mathbf{B})}{\mu_0}\tag{21.4}$$

the total energy flux density.

Let us move into the rest frame of the shock. For a 1D shock, suppose that the shock front coincides with the y - z plane. Furthermore, let the regions of the plasma upstream and downstream of the shock, which are termed regions 1 and 2, respectively, be *spatially uniform* and *time-static*, i.e. $\partial/\partial t = \partial/\partial y = \partial/\partial z = 0$. Moreover, $\partial/\partial x = 0$, except in the immediate vicinity of the shock. Finally, let the velocity and magnetic fields upstream and downstream of the shock all lie in the x-y plane. The situation under discussion is illustrated in the figure below.

Here, ρ_1 , p_1 , \mathbf{U}_1 , and \mathbf{B}_1 are the downstream mass density, pressure, velocity, and magnetic field, respectively, whereas ρ_2 , p_2 , \mathbf{U}_2 , and \mathbf{B}_2 are the corresponding upstream quantities.²

¹One point that is probably confusing for many people is the difference between the ram pressure ρU^2 and the dynamic pressure $\rho U^2/2$, which differs by a factor of 2. If the density is constant (which is a subset of incompressible fluid), the Bernoulli equation $\rho V^2/2 + P_t + P_B = \text{const.}$ is satisfied. However, for compressible fluid the pressure balance equation is $\rho V^2 + P_t + P_B = \text{const.}$

²Fitzpatrick's note use V for velocity. I prefer U for fluid velocities and V for particle velocities.

The basic RH relations are listed in [MHD shocks](#). In the immediate vicinity of the planar shock, Equation [21.1](#) reduce to

$$\begin{aligned}\frac{dB_x}{dx} &= 0 \\ \frac{d}{dx}(U_x B_y - U_y B_x) &= 0 \\ \frac{d(\rho U_x)}{dx} &= 0 \\ \frac{dS_{xx}}{dx} &= 0 \\ \frac{dS_{xy}}{dx} &= 0 \\ \frac{dw_x}{dx} &= 0\end{aligned}$$

Integration across the shock yields the desired jump conditions:

$$\begin{aligned}[B_x] &= 0 \\ [U_x B_y - U_y B_x] &= 0 \\ [\rho U_x] &= 0 \\ [\rho U_x^2 + p + B_y^2/2\mu_0] &= 0 \\ [\rho U_x U_y - B_x B_y/\mu_0] &= 0 \\ \left[\frac{1}{2} \rho U^2 U_x + \frac{\gamma}{\gamma-1} p U_x + \frac{B_y(U_x B_y - U_y B_x)}{\mu_0}\right] &= 0\end{aligned}\tag{21.5}$$

where $[A] = A_2 - A_1$ is the difference across the shock. These relations are often called the Rankine-Hugoniot relations for MHD. There are 6 scalar equations and 12 (6 upstream and 6 downstream) scalar variables all together. Assuming that all of the upstream plasma parameters are known, there are 6 unknown parameters in the problem—namely, B_{x2} , B_{y2} , U_{x2} , U_{y2} , ρ_2 , and p_2 . These 6 unknowns are fully determined by the six jump conditions. If we loose the planar assumption, then we typically write the x -component as the normal component (U_n, B_n) and the combined y - and z -components as the tangential component (U_t, B_t) :

$$\begin{aligned}[B_n] &= 0 \\ [U_n B_t - U_t B_n] &= 0 \\ [\rho U_n] &= 0 \\ [\rho U_n^2 + p + B_t^2/2\mu_0] &= 0 \\ [\rho U_n U_t - B_n B_t/\mu_0] &= 0 \\ \left[\frac{1}{2} \rho U^2 U_n + \frac{\gamma}{\gamma-1} p U_n + \frac{B_t(U_n B_t - U_t B_n)}{\mu_0}\right] &= 0\end{aligned}\tag{21.6}$$

Luckily this is still deterministic. However, as you can see later, the general case is very complicated.

A clear exposition of the two types of strong discontinuities, namely the shock wave and the tangential discontinuity can be found in §84, Landau & Lifshitz. By definition

- shocks are transition layers across which there is a transport of particles, whereas
- discontinuities are transition layers across which there is no particle transport.

Mathematically, let the shock plane speed be U_s , then

$$\begin{aligned} [\rho(U_x - U_s)] &\neq 0 \quad \text{for shock} \\ [\rho(U_x - U_s)] &= 0 \quad \text{for discontinuity} \end{aligned} \tag{21.7}$$

Thus in shocks $[U_n] \neq 0$, and in discontinuities $[U_n] = 0$. Take a reference frame fixed to the discontinuity with x-axis along the normal. Since mass, momentum and energy is conserved across the discontinuity, we must have from Equation 21.5 for inviscid flows (no magnetic field, y-direction represents the tangential direction),

$$\begin{aligned} [\rho U_x] &= 0 \\ [\rho U_x^2 + p] &= 0, [\rho U_x U_y] = 0 \\ [\rho U_x \left(\frac{1}{2} U^2 + H \right)] &= 0 \end{aligned}$$

where H is the enthalpy. In tangential discontinuities, no particle transport means $U_{1x} = U_{2x} = 0$. Then the x-momentum jump implies $[p] = 0$, where the y-momentum jump sets no restrictions on U_y . There is also no restriction on ρ . Energy equation is also satisfied. Thus, in tangential discontinuities, the density and tangential velocity components can be discontinuities, whereas the pressure must be continuous and the normal velocity component must be zero.

The categories of the solution of Equation 21.5 are shown in Table 21.1. The \pm signs denote the changes of the downstream compared with the upstream (+ means increase, - means decrease).

Table 21.1: Classes of MHD shocks and discontinuities

Type	Particle Transport	ρ	\mathbf{U}	p	\mathbf{B}	T
Tangential	No	\pm	$U_n = 0$	continuous	$B_n = 0$	\pm
Contact	No	\pm	continuous	continuous	continuous	\pm
Slow	Yes	+	-	+	B_t -	+

Type	Particle Transport	ρ	\mathbf{U}	p	\mathbf{B}	T
Intermediate	Yes	continuous	\pm	\pm	\pm	\pm
Rotational	No	continuous	$U_n = B_n/\sqrt{\mu_0 \rho},$ $U_t - B_t/\sqrt{\mu_0 \rho} = 0$	continuous	$B_n = 0$	\pm
Fast	Yes	+	-	+	B_t	+

1. The contact discontinuity (CD) is a special case of tangential discontinuity (TD) in which we assume $[U_t] = 0$, i.e., the tangential velocity (and so the velocity) is continuous, but not the density and other thermodynamic variables. Since in a CD the thermal pressure remains constant, any change in density must be compensated by a change in temperature. However, a temperature jump is quickly dissipated by electron heat conduction, which hints that CD do not persist long. TDs are often observed in the solar wind. The Hot Flow Anomaly (HFA) requires a TD and a normal electric field pointing towards the discontinuity.
2. The Earth's magnetopause (Section 22.7.1) is generally a tangential discontinuity. When there is no flux rope been generated, the magnetopause can be treated as the surface of pressure balance between magnetic pressure, ram pressure and thermal pressure. However, when reconnection triggers flux rope generation, it may become a rotational discontinuity (RD) (TO BE CONFIRMED!).
3. Intermediate (Alfvénic) shocks are incompressive and isentropic. The rotational discontinuity is a special case of the intermediate shock. The tangential velocity relation $[U_t - \frac{B_t}{\sqrt{\mu_0 \rho}}] = 0$ can be derived from Equation 21.6 assuming the *Walen relation* holds, $U_n = B_n/\sqrt{\mu_0 \rho}$, i.e. this is an Alfvénic shock. All thermodynamic quantities are continuous across the shock, but the tangential component of the magnetic field can rotate. RD is frequently observed in fast solar wind. Intermediate shocks in general however, unlike rotational discontinuities, can have a discontinuity in the pressure.
4. Fast- and slow-mode shocks are compressive and are associated with an increase in entropy. Fast/slow shocks have increasing/decreasing magnetic pressure from the upstream to the downstream of the shock. For example, the Earth's bow shock is a fast, supercritical shock (See criticality in Section 21.1.3).
5. RD and TD are similar. One easy way to distinguish these two are performing a MVA analysis, get the normal magnetic field component, and see if B_N (not the jump!) is zero. If $B_N = 0$, then it is a TD ("tangential" means no normal component); otherwise it is a RD.

The solutions can also be summarized in [the context of Riemann problem](#) or visually in Figure 21.1.

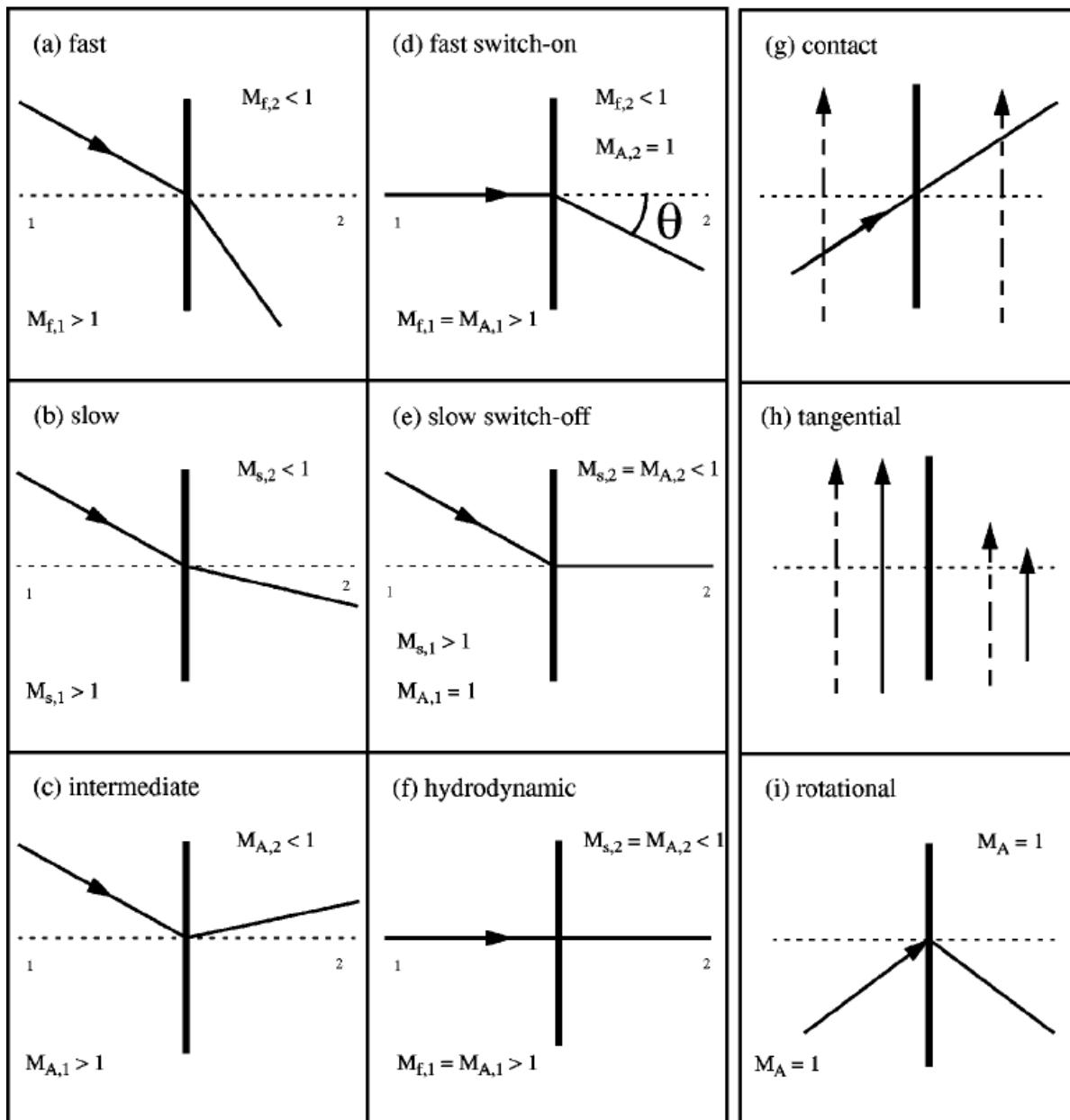


Figure 21.1: Some properties of MHD shocks and discontinuities. The thick vertical line is the shock surface. The shock normal is dotted. The full arrowed lines are magnetic field lines that are refracted through the shock surface. The dashed arrowed lines are velocity vectors. Region 1 is upstream, 2 is downstream. Reprinted from De Sterck, Low, and Poedts (1998).

21.1.1 Evolutionarity

The hyperbolic nature of the conservation laws allows wave propagation only if it is in accord with *causality*. Causality is a general requirement in nature, meaning in this case that the drop in speed across a shock must be large enough for the normal component of the downstream flow to fall below the corresponding downstream mode velocity. For a fast shock this implies the following ordering of the normal flow and magnetosonic velocities to both sides of the shock:

$$\begin{aligned} U_{1n} &> c_{1ms}^+ \\ U_{2n} &< c_{2ms}^+ \end{aligned}$$

where the numbers 1, 2 refer to upstream and downstream of the fast shock wave.

The first condition is necessary for the shock to be formed as a priori; it is the second condition which (partially) accounts for the *evolutionarity*. Otherwise the small fast-mode disturbances excited downstream and moving upward towards the shock would move faster than the flow, they would overcome the shock and steepen it without limit. Since this cannot happen for a shock to form, the downstream normal speed must be less than the downstream fast magnetosonic speed. Furthermore, for fast shocks the flow velocity must be greater than the intermediate (Alfvén) speed on both sides of the shock, while for slow shocks it must be less than the intermediate speed on both sides. These conditions hold because of the same reason as otherwise the corresponding waves would catch up with the shock front, modify and destroy it and no shock could form.

21.1.2 Coplanarity

For a stationary ideal MHD shock wave with no other wave activity or kinetic processes present outside the shock transition, such that dissipation takes place solely inside the narrow shock transition and this transition region can be considered as infinitesimally thin with respect to all other physical scales in the plasma, the electric field in the shock rest frame is strictly perpendicular to the magnetic field, $\mathbf{E} = -\mathbf{U} \times \mathbf{B}$. The stationary Faraday's law

$$\nabla \times \mathbf{E} = \dot{\mathbf{B}} = 0$$

leads to the vanishing of the difference in the tangential components of the magnetic field to both sides

$$(U_{n2} - U_{n1})(\mathbf{B}_{t2} \times \mathbf{B}_{t1}) = 0$$

For a shock $[U_n] \neq 0$, hence

$$\mathbf{B}_{t2} \times \mathbf{B}_{t1} = 0$$

i.e. the two tangential components to both sides are strictly parallel.

With more details, $[U_t]$ from the [RH relations](#) and obtain

$$[U_n \mathbf{B}_t] = \frac{B_n^2}{\mu_0 \dots} [\mathbf{B}_t]$$

Hence the cross product of the left with the right hand side must vanish:

$$\begin{aligned} [\mathbf{B}_t] \times [U_n \mathbf{B}_t] &= 0 \\ (\mathbf{B}_{t2} - \mathbf{B}_{t1}) \times (U_{n2} \mathbf{B}_{t2} - U_{n1} \mathbf{B}_{t1}) &= 0 \\ (U_{n1} - U_{n2})(\mathbf{B}_{t1} \times \mathbf{B}_{t2}) &= 0 \\ \mathbf{B}_{t1} \parallel \mathbf{B}_{t2} \end{aligned} \tag{21.8}$$

The resulting coplanarity theorem implies that the magnetic field across the shock has a reduced 2-D geometry: upstream and downstream tangential fields are parallel to each other and coplanar with the shock normal \hat{n} .

Coplanarity does not strictly hold, however. For instance, when the shock is non-stationary, i.e. when its width changes with time or in the direction tangential to the shock, then in Faraday's law $\partial \mathbf{B} / \partial t = \nabla \times \mathbf{E} \neq 0$, and coplanarity becomes violated.

Also, any upstream low frequency EM wave that propagates along the upstream magnetic field, possesses a magnetic wave field that is perpendicular to the upstream field. When it encounters the shock, this tangential component will be transformed and amplified across the shock. This naturally introduces an out-of-plane magnetic field component, thereby violating the co-planarity condition. There are also other effects which violate coplanarity at a non-MHD shock.

21.1.3 Criticality

Shock is a dissipative structure in which the kinetic and magnetic energy of a directed plasma flow is partly transferred to heating of the plasma. The dissipation does not take place, however, by means of particle collisions for a shock in space. Collisionless shocks can be divided into *super-* and *sub-critical*, according to their Mach-numbers $M < M_c$ being smaller or $M > M_c$ larger than some critical Mach-number M_c . M_c refers to a threshold value of the shock's Mach number above which certain physical processes become dominant, leading to distinct changes in the shock's behavior and structure.³

³In aerodynamics, the [critical Mach number](#) M^* of an aircraft is the *lowest* Mach number at which the airflow over some point of the aircraft reaches the speed of sound, but does not exceed it. For an aircraft in flight, the speed of the airflow around the aircraft differs considerably in places from the airspeed of the aircraft; this is due to the airflow having to speed up and slow down as it travels around the aircraft's structure. When the aircraft's airspeed reaches the critical Mach number, the speed of the airflow in some areas near the airframe reaches the speed of sound, even though the aircraft itself has an airspeed lower than Mach 1.0. This creates a weak shock wave. As the aircraft exceeds the critical Mach number, its drag coefficient increases suddenly, causing dramatically increased drag, and may lead to deterioration in control of the aircraft.

Depending on the specific physical processes, we can have different critical Mach numbers associated with collisionless shocks:

1. M_c for electron injection: This is the minimum Mach number above which thermal electrons in the upstream plasma can be accelerated to high energies and injected into the shock acceleration process. This injection is essential for producing the observed cosmic ray electron spectra.
2. M_c for ion reflection: This is the minimum Mach number above which a significant fraction of upstream ions are reflected from the shock front. This ion reflection plays a crucial role in the shock dissipation process and can lead to the formation of foreshocks.
3. M_c for whistler precursor formation: This is the minimum Mach number above which whistler waves can be excited upstream of the shock, forming a precursor wave structure that can modify the shock's properties.
4. M_c for nonlinear wave steepening: This is the minimum Mach number above which non-linear wave steepening effects become important, leading to the formation of solitons and other non-linear wave structures in the shock transition region.

For a resistive MHD shock Marshall [1955] had numerically determined the critical Mach number to $M_c \approx 2.76$ (see Figure 21.2). In various scenarios, M_c typically fall in the range of 2-10 for most astrophysical and space plasma environments.

Subcritical shocks are capable of generating sufficient dissipation to account for retardation, thermalisation and entropy in the time the flow crosses the shock from upstream to downstream. The relevant processes are based on wave-particle interaction between the shocked plasma and the shock-excited turbulent wave fields.

For supercritical shocks this is, however, not the case. Supercritical shocks must evoke mechanisms different from simple wave-particle interaction for getting rid of the excess energy in the bulk flow that cannot be dissipated by any classical anomalous dissipation. Above the critical Mach number the simplest efficient way of energy dissipation is rejection of the inflowing excess energy from the shock by reflecting a substantial part of the incoming plasma back upstream. The non-thermal processes for dissipating excess energy include

- Particle acceleration to very high energies;
- Generation of strong, complex magnetic fields;
- Significant heating of the plasma.

To show how the critical Mach number of a shock arises from the Rankine-Hugoniot relations we consider the strictly perpendicular case with vanishing upstream pressure $P_1 = 0$. The jump

conditions become very simple in this case:

$$\begin{aligned} n_1 U_1 &= n_2 U_2 \\ v_1 B_1 &= v_2 B_2 \\ n_1 U_1^2 + \frac{B_1^2}{2\mu_0 m} &= n_2 U_2^2 + \frac{P_2}{m} + \frac{B_2^2}{2\mu_0 m} \\ \frac{U_1^2}{2} + \frac{B_1^2}{\mu_0 m n_1} &= \frac{U_2^2}{2} + \frac{\gamma}{\gamma - 1} \frac{P_2}{m n_2} + \frac{B_1^2}{\mu_0 m n_1} \end{aligned}$$

where B is the only existing tangential component of the magnetic field here, and $\gamma = 5/3$ is the adiabatic index (valid for fast, adiabatic transitions across the shock). This is the simplest imaginable case of an MHD shock, and it is easy to solve these equations. Figure 21.2 shows the resulting relation between the normalised downstream flow U_2/U_1 and downstream sound speed $c_{s2}/U_1 = \sqrt{\gamma P_2/n_2}/v_1$ as function of upstream Alfvén Mach number $M_A = U_1 \sqrt{\mu_0 m n_1}/B_1$.

The two curves in the figure cross each other at the critical Mach number which in the present case is $M_c = 2.76$ and where the downstream sound speed exceeds the flow speed. Below the critical Mach number the downstream flow is still supersonic (though clearly sub-magnetosonic!). Only above the critical Mach number the downstream flow velocity falls below the downstream sound speed. There is thus a qualitative change in the shock character above it that is not contained in the Rankine-Hugoniot conditions.

So as a quick recap, the inflow of matter into a supercritical shock is so fast that the time scales on which dissipation would take place are too long for dissipating the excess energy and lowering the inflow velocity below the downstream magnetosonic velocity. Hence, the condition for criticality, is that the downstream flow velocity becomes equal to the downstream *sonic* speed, which yielded the critical Mach number, $M_c \simeq 2.76$.

The determination of the critical Mach number poses an interesting question of why sonic speed appears in the downstream. The finite magnetic field compression ratio sets an upper limit to the rate of resistive dissipation that is possible in an MHD shock. Plasmas possess several dissipative lengths, depending on which dissipative process is considered. Any nonlinear wave that propagates in the plasma should steepen as long, until its transverse scale approaches the longest of these dissipative scales. Then dissipation sets on and limits its amplitude.

Thus, when the wavelength of the fast magnetosonic wave approaches the resistive length, the magnetic field decouples from the wave by resistive dissipation, and the wave speed becomes the sound speed downstream of the shock ramp. The condition for the critical Mach number is then given by $U_{n2} = c_{2s}$. Similarly, for the slow-mode shock, because of its different dispersive properties, the resistive critical-Mach number is defined by the condition $U_{1n} = c_{1s}$ ⁴. Since these quantities depend on wave angle, they have to be solved numerically. Prior studies showed that critical fast-mode Mach number varies between 1 and 3, depending on the upstream plasma parameters and flow angle to the magnetic field. It is usually called *first* critical Mach

⁴so it depends on the upstream sound speed?

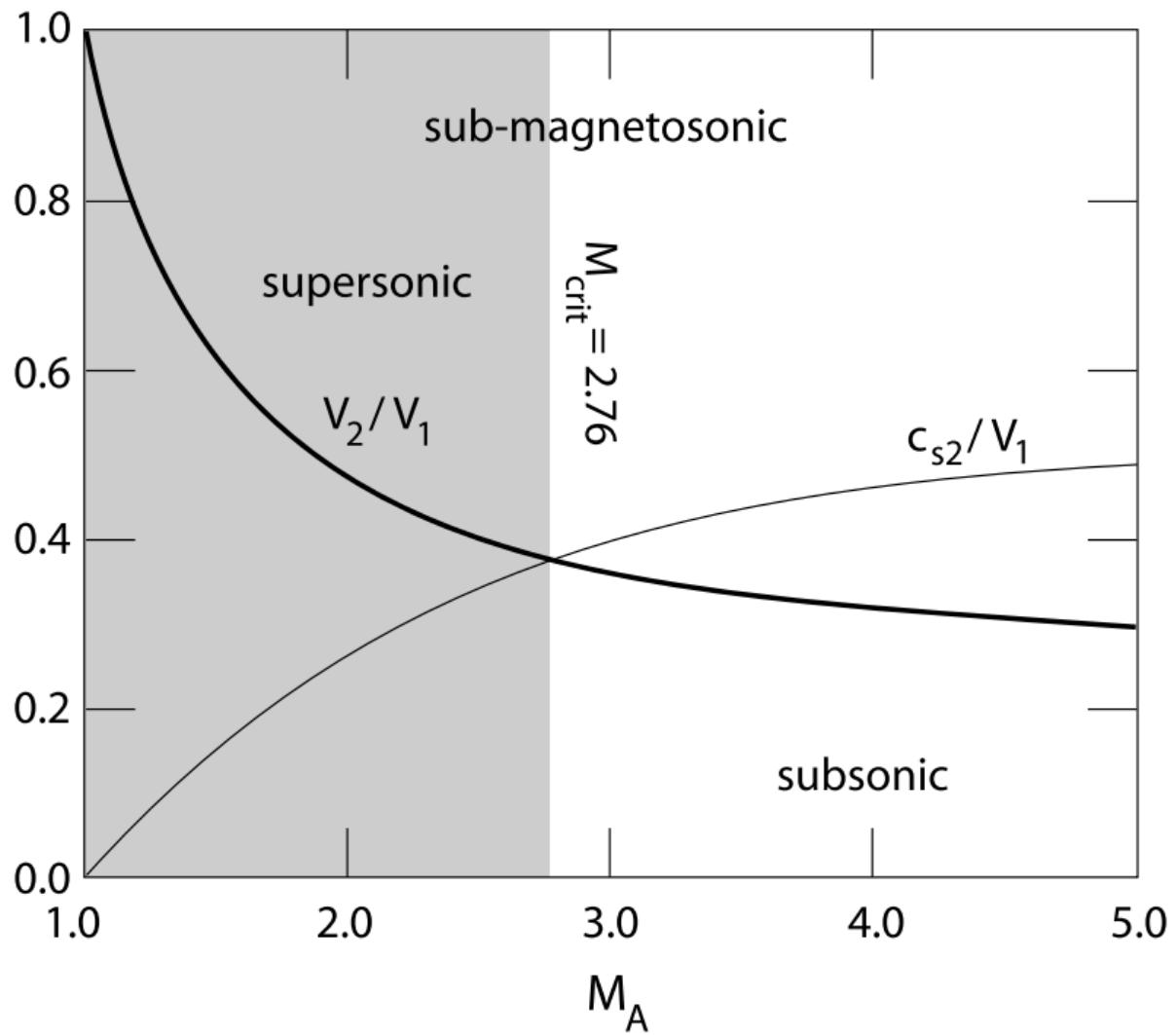


Figure 21.2: Dependence of the downstream normalised flow V_2/V_1 and sound c_{s2} velocities on the upstream Alfvén Mach number for an ideal MHD perpendicular shock with zero upstream pressure $P_2 = 0$. The crossing of the two curves defines the critical Mach number which is $M_c = 2.76$.

number, because there is theoretical evidence in simulations for a second critical Mach number, which comes into play when the shock structure becomes time dependent, whistlers accumulate at the shock front and periodically cause its reformation. The dominant dispersion is then the whistler dispersion. An approximate expression for this second or whistler critical Mach number is

$$M_{2c} \propto \left(\frac{m_i}{m_e} \right)^{1/2} \cos \theta_{Bn} \quad (21.9)$$

where the constant of proportionality depends on whether one defines the Mach number with respect to the whistler phase or group velocities. For the former it is $1/2$, and for the latter $\sqrt{27/64}$ [Oka+, 2006].

It is clear that it is the smallest critical Mach number that determines the behaviour of the shock. In simple words: $M > 1$ is responsible for the existence of the shock under the condition that an obstacle exists in the flow, which is disturbed in some way such that fast waves can grow, steepen and form shocks. When, in addition, the flow exceeds the next lowest Mach number for a given θ_{Bn} the shock at this angle will make the transition into a supercritical shock and under additional conditions, which have not yet been ultimately clarified, will start reflecting particles back upstream. If, because of some reason, this would not happen, the flow might have to exceed the next higher critical Mach number until reflection becomes possible. In such a case the shock would become metastable in the region where the Mach number becomes supercritical, will steepen and shrink in width until other effects and – ultimately – reflection of particles can set on.

21.1.4 Parallel Shock

The first special case is the so-called parallel shock in which both the upstream and downstream plasma flows are parallel to the magnetic field, as well as perpendicular to the shock front. In other words,

$$\begin{aligned} \mathbf{U}_1 &= (U_1, 0, 0), & \mathbf{U}_2 &= (U_2, 0, 0) \\ \mathbf{B}_1 &= (B_1, 0, 0), & \mathbf{B}_2 &= (B_2, 0, 0) \end{aligned} \quad (21.10)$$

Substitution into Equation 21.5 yields

$$\begin{aligned} \frac{B_2}{B_1} &= 1 \\ \frac{\rho_2}{\rho_1} &= r \\ \frac{U_2}{U_1} &= r^{-1} \\ \frac{p_2}{p_1} &= R \end{aligned} \quad (21.11)$$

with

$$r = \frac{(\gamma + 1)M_1^2}{2 + (\gamma - 1)M_1^2} \quad (21.12)$$

$$R = 1 + \gamma M_1^2(1 - r^{-1}) = \frac{(\gamma + 1)r - (\gamma - 1)}{(\gamma + 1) - (\gamma - 1)r}$$

Here, $M_1 = U_1/c_{s1}$, where $c_{s1} = \sqrt{(\gamma p_1/\rho_1)}$ is the upstream sound speed. Thus, the upstream flow is supersonic if $M_1 > 1$, and subsonic if $M_1 < 1$. Incidentally, as is clear from the above expressions, a parallel shock is unaffected by the presence of a magnetic field. In fact, this type of shock is identical to that which occurs in neutral fluids, and is, therefore, usually called a *hydrodynamic shock*.

It is easily seen from Equation 21.10 that there is no shock (i.e., no jump in plasma parameters across the shock front) when the upstream flow is exactly sonic: i.e., when $M_1 = 1$. In other words, $r = R = 1$ when $M_1 = 1$. However, if $M_1 \neq 1$ then the upstream and downstream plasma parameters become different (i.e., $r \neq 1$, $R \neq 1$) and a true shock develops. In fact, it is easily demonstrated that

$$\frac{\gamma - 1}{\gamma + 1} \leq r \leq \frac{\gamma + 1}{\gamma - 1}$$

$$0 \leq R \leq \infty \quad (21.13)$$

$$\frac{\gamma - 1}{2\gamma} \leq M_1^2 \leq \infty$$

Note that the upper and lower limits in the above inequalities are all attained simultaneously.

The previous discussion seems to imply that a parallel shock can be either compressive (i.e., $r > 1$) or expansive (i.e., $r < 1$). Is there a preferential direction across the shock? In other words, can we tell the upstream and the downstream? Yes, with the additional physics principle of the *second law of thermodynamics*. This law states that the *entropy* of a closed system can spontaneously increase, but can never spontaneously decrease. Now, in general, the entropy per particle is different on either side of a hydrodynamic shock front. Accordingly, the second law of thermodynamics mandates that the downstream entropy must *exceed* the upstream entropy, so as to ensure that the shock generates a net increase, rather than a net decrease, in the overall entropy of the system, as the plasma flows through it.

The (suitably normalized) entropy per particle of an ideal plasma takes the form

$$S = \ln \left(\frac{p}{\rho^\gamma} \right)$$

Hence, the difference between the upstream and downstream entropies is

$$[S] = \ln R - \gamma \ln r$$

Now, using Equation 21.12,

$$r \frac{d[S]}{dr} = \frac{r}{R} \frac{dR}{dr} - \gamma = \frac{\gamma(\gamma^2 - 1)(r - 1)^2}{[(\gamma + 1)r - (\gamma - 1)][(\gamma + 1) - (\gamma - 1)r]}$$

Furthermore, it is easily seen from Equation 21.13 that $d[S]/dr \geq 0$ in all situations of physical interest. However, $[S] = 0$ when $r = 1$, since, in this case, there is no discontinuity in plasma parameters across the shock front. We conclude that $[S] < 0$ for $r < 1$, and $[S] > 0$ for $r > 1$. It follows that the second law of thermodynamics requires hydrodynamic shocks to be *compressive*: i.e., $r \equiv \rho_2/\rho_1 > 1$. In other words, the plasma density must always *increase* when a shock front is crossed in the direction of the relative plasma flow. It turns out that this is a general rule which applies to all three types of MHD shock. In the shock rest frame, the shock is associated with an irreversible (since the entropy suddenly increases) transition from supersonic to subsonic flow.

The upstream Mach number, M_1 , is a good measure of shock strength: i.e., if $M_1 = 1$ then there is no shock, if $M_1 - 1 \ll 1$ then the shock is weak, and if $M_1 \gg 1$ then the shock is strong. We can define an analogous downstream Mach number, $M_2 = U_2/(\gamma p_2/\rho_2)^{1/2}$. It is easily demonstrated from the jump conditions that if $M_1 > 1$ then $M_2 < 1$. In other words, in the shock rest frame, the shock is associated with an irreversible (since the entropy suddenly increases) transition from supersonic to subsonic flow. Note that $r \equiv \rho_2/\rho_1 \rightarrow (\gamma + 1)/(\gamma - 1)$, whereas $R \equiv p_2/p_1 \rightarrow \infty$, in the limit $M_1 \rightarrow \infty$. In other words, as the shock strength increases, the compression ratio, r , asymptotes to a finite value, whereas the pressure ratio, P , increases without limit. For a conventional plasma with $\gamma = 5/3$, the limiting value of the compression ratio is 4: i.e., the downstream density can never be more than four times the upstream density. We conclude that, in the strong shock limit, $M_1 \gg 1$, the large jump in the plasma pressure across the shock front must be predominately a consequence of a large jump in the plasma *temperature*, rather than the plasma *density*. In fact, the definitions of r and R imply that

$$\frac{T_2}{T_1} \equiv \frac{R}{r} \rightarrow \frac{2\gamma(\gamma - 1)M_1^2}{(\gamma + 1)^2} \gg 1$$

as $M_1 \rightarrow \infty$. Thus, a strong parallel, or hydrodynamic, shock is associated with intense plasma heating.

As we have seen, the condition for the existence of a hydrodynamic shock is $M_1 > 1$, or $U_1 > U_{S1}$. In other words, in the shock frame, the upstream plasma velocity, U_1 , must be supersonic. However, by Galilean invariance, U_1 can also be interpreted as the propagation velocity of the shock through an initially stationary plasma. It follows that, in a stationary plasma, a parallel, or hydrodynamic, shock propagates along the magnetic field with a supersonic velocity.

21.1.5 Perpendicular Shock

The second special case is the so-called *perpendicular shock* in which both the upstream and downstream plasma flows are perpendicular to the magnetic field, as well as the shock front. In other words,

$$\begin{aligned}\mathbf{U}_1 &= (U_1, 0, 0), \quad \mathbf{U}_2 = (V_2, 0, 0) \\ \mathbf{B}_1 &= (0, B_1, 0), \quad \mathbf{B}_2 = (0, B_2, 0)\end{aligned}\tag{21.14}$$

Substitution into Equation 21.5 yields

$$\begin{aligned}\frac{B_2}{B_1} &= r \\ \frac{\rho_2}{\rho_1} &= r \\ \frac{U_2}{U_1} &= r^{-1} \\ \frac{p_2}{p_1} &= R\end{aligned}\tag{21.15}$$

where

$$R = 1 + \gamma M_1^2 (1 - r^{-1}) + \beta_1^{-1} (1 - r^2)\tag{21.16}$$

and r is a real positive root of the quadratic

$$F(r) = 2(2 - \gamma)r^2 + \gamma[2(1 + \beta_1) + (\gamma - 1)\beta_1 M_1^2]r - \gamma(\gamma + 1)\beta_1 M_1^2 = 0\tag{21.17}$$

Here, $\beta_1 = 2\mu_0 p_1 / B_1^2$.

Now, if r_1 and r_2 are the two roots of Equation 21.17 then

$$r_1 r_2 = -\frac{\gamma(\gamma + 1)\beta_1 M_1^2}{2(2 - \gamma)}$$

Assuming that $\gamma < 2$, we conclude that one of the roots is negative, and, hence, that Equation 21.17 only possesses one physical solution: i.e., there is only one type of MHD shock which is consistent with Equation 21.14. Now, it is easily demonstrated that $F(0) < 0$ and $F(\gamma + 1/\gamma - 1) > 0$. Hence, the physical root lies between $r = 0$ and $r = (\gamma + 1)/(\gamma - 1)$.

Using similar analysis to that employed in the previous subsection, it is easily demonstrated that the second law of thermodynamics requires a perpendicular shock to be compressive: i.e., $r > 1$. It follows that a physical solution is only obtained when $F(1) < 0$, which reduces to

$$M_1^2 > 1 + \frac{2}{\gamma \beta_1}$$

This condition can also be written

$$\mathbf{U}_1^2 > \mathbf{V}_{s1}^2 + \mathbf{V}_{A1}^2 = V_{+1}^2$$

where $V_{A1} = B_1/\sqrt{\mu_0\rho_1}$ is the upstream Alfvén speed. $V_{+1} = (V_{S1}^2 + V_{A1}^2)^{1/2}$ can be recognized as the velocity of a fast wave propagating perpendicular to the magnetic field (Section 10.7.4). Thus, the condition for the existence of a perpendicular shock is that the relative upstream plasma velocity must be *greater* than the upstream fast wave velocity. Incidentally, it is easily demonstrated that if this is the case then the downstream plasma velocity is less than the downstream fast wave velocity. We can also deduce that, in a stationary plasma, a perpendicular shock propagates across the magnetic field with a velocity which exceeds the fast wave velocity.

In the strong shock limit, $M_1 \gg 1$, Equation 21.16 and Equation 21.17 become identical to Equation 21.12. Hence, a strong perpendicular shock is very similar to a strong hydrodynamic shock (except that the former shock propagates perpendicular, whereas the latter shock propagates parallel, to the magnetic field). In particular, just like a hydrodynamic shock, a perpendicular shock cannot compress the density by more than a factor $(\gamma + 1)/(\gamma - 1)$. However, according to Equation 21.15, a perpendicular shock compresses the magnetic field by the same factor that it compresses the plasma density. It follows that there is also an upper limit to the factor by which a perpendicular shock can compress the magnetic field.

21.1.6 Oblique Shock

Let us now consider the general case in which the plasma velocities and the magnetic fields on each side of the shock are neither parallel nor perpendicular to the shock front. It is convenient to transform into the so-called *de Hoffmann-Teller frame* in which $|\mathbf{U}_1 \times \mathbf{B}_1| = 0$, or

$$U_{x1}B_{y1} - U_{y1}B_{x1} = 0 \quad (21.18)$$

In other words, it is convenient to transform to a frame which moves at the local $\mathbf{E} \times \mathbf{B}$ velocity of the plasma. The key idea is to extract the velocity component perpendicular to \mathbf{B}_1 from \mathbf{v}_1 . One temptive idea is to just remove the perpendicular part:

$$\begin{aligned} \mathbf{V}_{dHT} &= \mathbf{U}_1 - (\mathbf{U}_1 \cdot \mathbf{b}_1)\mathbf{b}_1 \\ \mathbf{U}'_1 &= \mathbf{U}_1 - \mathbf{V}_{dHT} = (\mathbf{U}_1 \cdot \mathbf{b}_1)\mathbf{b}_1 \end{aligned}$$

Note that the transformation is not unique, since one can always add a parallel velocity component. Although the above transformation is correct, it introduces two issues:

1. The ram pressure ρV_n^2 is changed between the two coordinates.
2. Equation 21.22 may not possess a valid solution given certain parameters (e.g. $MA_1 = 5, \theta_1 = 65^\circ$ will give $r < 1$.)

To fix this, we can use another transformation:

$$\begin{aligned}\mathbf{V}_{\text{dHT}} &= \mathbf{U}_1 - \frac{U_x}{B_x} \mathbf{B} \\ \mathbf{U}'_1 &= \mathbf{U}_1 - \mathbf{V}_{\text{dHT}} = \frac{U_x}{B_x} \mathbf{B}\end{aligned}\tag{21.19}$$

A nice property of this transformation is that the normal velocity component is kept the same, so is the ram pressure. Therefore, we can apply the same quantities in the lab frame as in the de Hoffmann-Teller frame.

Taking Equation 21.18 into the 2nd jump condition of Equation 21.5 gives

$$U_{x2} B_{y2} - U_{y2} B_{x2} = 0\tag{21.20}$$

or $|\mathbf{U}_2 \times \mathbf{B}_2| = 0$. Thus, in the de Hoffmann-Teller frame, the upstream plasma flow is parallel to the upstream magnetic field, and the downstream plasma flow is also parallel to the downstream magnetic field. Furthermore, the magnetic contribution to the jump condition Equation 21.5 (last eq.) becomes identically zero, which is a considerable simplification.

Equation 21.18 and Equation 21.20 can be combined with the general jump conditions Equation 21.5 to give⁵

$$\begin{aligned}\frac{\rho_2}{\rho_1} &= r \\ \frac{B_{x2}}{B_{x1}} &= 1 \\ \frac{B_{y2}}{B_{y1}} &= r \frac{U_1^2 - V_{A1}^2}{U_1^2 - r V_{A1}^2} \\ \frac{U_{x2}}{U_{x1}} &= \frac{1}{r} \\ \frac{U_{y2}}{U_{y1}} &= \frac{U_1^2 - V_{A1}^2}{U_1^2 - r V_{A1}^2} \\ \frac{p_2}{p_1} &= 1 + \frac{\gamma U_1^2(r-1)}{V_{s1}^2 r} \left[\cos^2 \theta_1 - \frac{r V_{A1}^2 \sin^2 \theta_1 [(r+1)U_1^2 - 2r V_{A1}^2]}{2(U_1^2 - r V_{A1}^2)^2} \right]\end{aligned}\tag{21.21}$$

where $U_{x1} = U_1 \cos \theta_1$ is the component of the upstream velocity normal to the shock front, and θ_1 is the angle subtended between the upstream plasma flow and the shock front normal.⁶⁷

⁵One way to define the tangential direction \hat{t} is to use \mathbf{U}_{t1} as a reference: $\mathbf{U}_{t1} = \mathbf{U}_1 - (\mathbf{U}_1 \cdot \hat{n})\hat{n}$. Then $\hat{t} = \mathbf{U}_{t1}/|\mathbf{U}_{t1}|$. Note that both the normal and tangential components are signed numbers.

⁶The shock normal points from downstream to upstream, and the plasma flow points from upstream to downstream. Usually we take an angle smaller than 90°, so the definition would be $\theta_1 = \cos^{-1}(-\mathbf{U}_1 \cdot \hat{n}/|\mathbf{U}_1|)$.

⁷Note that the velocity is defined in the dHT frame, not the lab frame! dHT is used here simply because the jump conditions are easier to solve. Also note that the jump conditions solved under the dHT frame is different from the lab frame!

Finally, given the compression ratio, r , the square of the normal upstream velocity, U_1^2 , is a real root of a cubic equation known as the *shock adiabatic*:⁸

$$0 = (U_1^2 - rV_{A1}^2)^2 \{ [(\gamma + 1) - (\gamma - 1)r]U_1^2 \cos^2 \theta_1 - 2rV_{s1}^2 \} \\ - r \sin^2 \theta_1 U_1^2 V_{A1}^2 \{ [\gamma + (2 - \gamma)r]U_1^2 - [(\gamma + 1) - (\gamma - 1)r]r V_{A1}^2 \} \quad (21.22)$$

As before, the second law of thermodynamics mandates that $r > 1$.

Weak shock limit

Let us first consider the weak shock limit $r \rightarrow 1$. In this case, it is easily seen that the three roots of the shock adiabatic reduce to the slow, intermediate (or Shear-Alfvén), and fast waves, respectively, propagating in the normal direction to the shock front:

$$U_1^2 = V_{-1}^2 \equiv \frac{V_{A1}^2 + V_{S1}^2 - [(V_{A1} + V_{S1})^2 - 4 \cos^2 \theta_1 V_{S1}^2 V_{A1}^2]^{1/2}}{2}$$

$$U_1^2 = \cos^2 \theta_1 V_{A1}^2$$

$$U_1^2 = V_{+1}^2 \equiv \frac{V_{A1}^2 + V_{S1}^2 + [(V_{A1} + V_{S1})^2 - 4 \cos^2 \theta_1 V_{S1}^2 V_{A1}^2]^{1/2}}{2}$$

We conclude that slow, intermediate, and fast MHD shocks degenerate into the associated MHD waves in the limit of small shock amplitude. Conversely, we can think of the various MHD shocks as *nonlinear* versions of the associated MHD waves. It is easily demonstrated that

$$V_{+1} > \cos \theta_1 V_{A1} > V_{-1}$$

In other words, a fast wave travels faster than an intermediate wave, which travels faster than a slow wave. It is reasonable to suppose that the same is true of the associated MHD shocks, at least at relatively low shock strength. It follows from Equation 21.21 that $B_{y2} > B_{y1}$ for a fast shock, whereas $B_{y2} < B_{y1}$ for a slow shock. For the case of an intermediate shock, we can show, after a little algebra, that $B_{y2} \rightarrow -B_{y1}$ in the limit $r \rightarrow 1$. We can conclude that (in the de Hoffmann-Teller frame) fast shocks refract the magnetic field and plasma flow (recall that they are parallel in our adopted frame of the reference) *away* from the normal to the shock front, whereas slow shocks refract these quantities *toward* the normal. Moreover, the tangential magnetic field and plasma flow generally *reverse* across an intermediate shock front. This is illustrated in ?@fig-mhd-shock-rest-frame.

When r is slightly larger than unity it is easily demonstrated that the conditions for the existence of a slow, intermediate, and fast shock are $U_1 > V_{-1}$, $U_1 > \cos \theta_1 V_{A1}$, and $U_1 > V_{+1}$, respectively.

⁸This equation is only valid in the dHT frame. In the lab frame, there shall be another equation for solving the compression ratio r from upstream conditions. However, the two must give identical results for the uniqueness of shock.

Strong shock limit

Let us now consider the strong shock limit, $U_1^2 \gg 1$. In this case, the shock adiabatic yields $r \rightarrow r_m = (\gamma + 1)/(\gamma - 1)$, and

$$U_1^2 \simeq \frac{r_m}{\gamma - 1} \frac{2V_{S1} \sin^2 \theta_1 [\gamma + (2 - \gamma) r_m] V_{A1}^2}{r_m - r}$$

There are no other real roots. The above root is clearly a type of fast shock. The fact that there is only one real root suggests that there exists a critical shock strength above which the slow and intermediate shock solutions cease to exist. (In fact, they merge and annihilate one another.) In other words, there is a limit to the strength of a slow or an intermediate shock. On the other hand, there is no limit to the strength of a fast shock. Note, however, that the plasma density and tangential magnetic field cannot be compressed by more than a factor $(\gamma + 1)/(\gamma - 1)$ by any type of MHD shock.

$$\theta_1 = 0$$

Consider the special case $\theta_1 = 0$ in which both the plasma flow and the magnetic field are normal to the shock front. In this case, the three roots of the shock adiabatic are

$$\begin{aligned} U_1^2 &= \frac{2r V_{S1}^2}{(\gamma + 1) - (\gamma - 1)r} \\ U_1^2 &= r V_{A1}^2 \\ U_1^2 &= r V_{A1}^2 \end{aligned}$$

We recognize the first of these roots as the hydrodynamic shock discussed in Section 21.1.4. This shock is classified as a slow shock when $V_{S1} < V_{A1}$, and as a fast shock when $V_{S1} > V_{A1}$. The other two roots are identical, and correspond to shocks which propagate at the velocity $U_1 = \sqrt{r} V_{A1}$ and “switch-on” the tangential components of the plasma flow and the magnetic field: it can be seen from Equation 21.21 that $U_{y1} = B_{y1} = 0$ whilst $U_{y2} \neq 0$ and $B_{y2} \neq 0$ for these types of shock.

There we have “switch-on” and “switch-off” shocks which refer to the generation and elimination of tangential components of the plasma flow and the magnetic field. Incidentally, it is also possible to have a “switch-off” shock which eliminates the tangential components of the plasma flow and the magnetic field. According to Equation 21.21, such a shock propagates at the velocity $U_1 = \cos \theta_1 V_{A1}$ ⁹. Switch-on and switch-off shocks are illustrated in [?@fig-shock-switch-on-off](#).

$$\theta_1 = \pi/2$$

⁹This is a very strong indication that Alfvén waves are involved in switch-on/off shocks!

Consider another special case $\theta_1 = \pi/2$. As is easily demonstrated, the three roots of the shock adiabatic are

$$U_1^2 = r \left(\frac{2V_{S1}^2 + [\gamma + (2 - \gamma)r] V_{A1}^2}{(\gamma + 1) - (\gamma - 1)r} \right)$$

$$U_1^2 = 0$$

$$U_1^2 = 0$$

The first of these roots is clearly a fast shock, and is identical to the perpendicular shock discussed in Section 21.1.5, except that there is no plasma flow across the shock front in this case. (IS IT BECAUSE OF THE HT FRAME?) The fact that the two other roots are zero indicates that, like the corresponding MHD waves, slow and intermediate MHD shocks do not propagate perpendicular to the magnetic field.

MHD shocks have been observed in a large variety of situations. For instance, shocks are known to be formed by supernova explosions, by strong stellar winds, by solar flares, and by the solar wind upstream of planetary magnetospheres.

21.1.7 Switch-On and Switch-Off Shocks

Parallel shocks in MHD should, theoretically, behave exactly like gasdynamic shocks, not having any upstream tangential magnetic field component and should also not have any downstream tangential field. This conclusion does not hold rigorously, however, since plasmas consist of charged particles which are sensitive to fluctuations in the field and can excite various waves in the plasma via electric currents which then become the sources of magnetic fields. The kinetic effects in parallel and quasi-parallel shocks play an important role in their physics and are well capable of generating tangential fields at least on scales shorter than the ion scale.

However, even in MHD as we have seen in the previous subsection, one stumbles across the interesting fact that this kind of shocks must have peculiar properties. The reason is that they are not, as in gasdynamics, the result of steepened *sound waves*, in which case they would simply be purely electrostatic shocks. At the contrary, the waves propagating parallel to the magnetic field are Alfvén and magnetosonic waves. Alfvén waves contain transverse magnetic field components. These transverse wave fields, in a parallel shock, are in fact tangential to the shock. Hence, if a purely parallel shock steepens, the transverse Alfvén waves do steepen as well, and the shock after the transition from upstream to downstream switches on a tangential magnetic component which originally was not present. Such shocks are called *switch-on* shocks. Similarly one can imagine the case that a tangential component behind the shock is by the same process switched off by an oppositely directed switch-on field, yielding a *switch-off* shock.

The problem of whether or not such shocks exist in MHD is related to the question whether or not an Alfvén wave steepens nonlinearly when propagating into a shock. To first order this steepening for an ordinary Alfvén wave is zero. However, to second order a wave trailing the leading Alfvén wave feels its weak transverse magnetic component. This trailing wave

therefore propagates slightly oblique to the main magnetic field and thus causes a second order density compression which in addition to generating a shock-like plasma compression changes the Alfvén velocity locally. In the case when the trailing wave is polarised in the same direction as the leading wave it also increases the transverse magnetic field component downstream of the compression thereby to second order switching on a tangential magnetic component. A whole train of trailing waves of same polarisation will thus cause strong steepening in both the density and tangential magnetic field.

Clearly, this kind of shocks is a more or less exotic case of MHD shocks whose importance is not precisely known, with very rare cases of observation.

21.2 Double Adiabatic MHD

The classical approach by Chew, Goldberger, and Low (1956) utilizes the MHD framework by assuming isotropic distributions parallel and perpendicular to the magnetic field, which results in scalar pressures on the two sides of the shock. This is now known as the CGL or double adiabatic theory.

When we shift to the MHD with anisotropic pressure tensor

$$P_{ij} = p_\perp \delta_{ij} + (p_\parallel - p_\perp) B_i B_j / B^2$$

where p_\perp and p_\parallel are the pressures perpendicular and parallel w.r.t. the magnetic field, respectively. For the strong magnetic field approximation, the two pressures are related to the plasma density and the magnetic field strength by two adiabatic equations,

$$\begin{aligned} \frac{d}{dt} \left(\frac{p_\parallel B^2}{\rho^3} \right) &= 0 \\ \frac{d}{dt} \left(\frac{p_\perp}{\rho B} \right) &= 0 \end{aligned} \tag{21.23}$$

This is where the name double adiabatic theory originates, which is also what many people remember to be the key conclusion from the CGL theory. However, these are constants at a fixed location in time: it is *not* correct to apply these across the shock. Also note the meaning of *adiabatic*: this means zero heat flux. If the system is not adiabatic, the conservation of these two quantities related to the parallel and perpendicular pressure is no longer valid, and additional terms may come into play such as the stochastic heating.

The general jump conditions for discontinuities in a collisionless anisotropic magnetized plasma in the CGL approximation were derived by Abraham-Shrauner (1967).

The general jump conditions for an anisotropic plasma are given in CGS units by Hudson (1970):

$$\begin{aligned}
& [\rho U_n] = 0 \\
& [U_n \mathbf{B}_t - \mathbf{U}_t B_n] = 0 \\
& [p_\perp + (p_\parallel - p_\perp) \frac{B_n^2}{B^2} + \frac{B_t^2}{8\pi} + \rho U_n^2] = 0 \\
& [\frac{B_n \mathbf{B}_t}{4\pi} \left(\frac{4\pi(p_\parallel - p_\perp)}{B^2} - 1 \right) + \rho U_n \mathbf{U}_t] = 0 \\
& [\rho U_n \left(\frac{\epsilon}{\rho} + \frac{U^2}{2} + \frac{p_\perp}{\rho} + \frac{B_t^2}{4\pi\rho} \right) + \frac{B_n^2 U_n}{B^2} (p_\parallel - p_\perp) \\
& \quad - \frac{\mathbf{B}_t \cdot \mathbf{U}_t B_n}{4\pi} \left(1 - \frac{4\pi(p_\parallel - p_\perp)}{B^2} \right)] = 0 \\
& [B_n] = 0
\end{aligned}$$

where ρ is the mass density, U and B are the velocity and magnetic field strength. Subscripts t and n indicate tangential and normal components with respect to the discontinuity. Quantities p_\perp and p_\parallel are the elements of the plasma pressure tensor perpendicular and parallel with respect to the magnetic field. Quantity ϵ is the internal energy, $\epsilon = p_\perp + p_\parallel/2$, and $[Q] = Q_2 - Q_1$, where subscripts 1 and 2 signify the quantity upstream and downstream of the discontinuity. These equations refer to the conservation of physical quantities, i.e. the mass flux, the tangential component of the electric field, the normal and tangential components of the momentum flux, the energy flux, and, finally, the normal component of the magnetic field. To solve the jump equations for anisotropic plasma conditions upstream and downstream of the shock, one has to use an additional equation, since the set of equations is underdetermined. One common choice is the magnetic field/density jump ratio.

The following derivations follow (Erkaev, Vogl, and Biernat 2000). Let us introduce two dimensionless parameters, A_s and A_m , which are determined for upstream conditions as

$$\begin{aligned}
A_s &= \frac{p_{\perp 1}}{\rho_1 v_1^2} \\
A_m &= \frac{1}{M_A^2}
\end{aligned}$$

where M_A is the upstream Alfvén Mach number. For common solar wind conditions, both of these parameters are quite small (~ 0.01).

For shocks, the tangential components of the electric and magnetic fields are *coplanar* (Equation 21.8). Thus, the components of the magnetic field upstream of the shock are given as $B_{n1} = B_1 \cos \theta_1$ and $B_{t1} = B_1 \sin \theta_1$, where θ_1 is the angle between the magnetic field vector and the vector \hat{n} normal to the discontinuity. Similarly, the components of the bulk velocity upstream of the shock are chosen as $U_{n1} = U_1 \cos \alpha$ and $U_{t1} = U_1 \sin \alpha$, where α the angle between the bulk velocity and the normal component of the velocity. Furthermore, a parameter

λ is used to denote the pressure anisotropy

$$\lambda = p_{\perp}/p_{\parallel}$$

and another parameter r is used to denote the ratio of density

$$r \equiv \frac{\rho_2}{\rho_1} = \frac{U_{n1}}{U_{n2}}$$

21.2.1 Perpendicular Shock

For a perpendicular shock, $B_n = 0$, we have the conservation relations reduce to

$$\begin{aligned} [\rho U_n] &= 0 \\ [U_n \mathbf{B}_t] &= 0 \\ [p_{\perp} + \frac{B_t^2}{8\pi} + \rho U_n^2] &= 0 \\ [\rho U_n \mathbf{U}_t] &= 0 \\ [\rho U_n \left(\frac{\epsilon}{\rho} + \frac{U^2}{2} + \frac{p_{\perp}}{\rho} + \frac{B_t^2}{4\pi\rho} \right)] &= 0 \end{aligned}$$

The quantities downstream of the discontinuity are

$$\begin{aligned} B_{t2} &= r B_{t1} \\ U_{t2} &= U_{t1} \\ p_{\perp 2} &= p_{\perp 1} + \frac{B_{t1}^2}{8\pi} (1 - r^2) + \rho_1 U_{n1}^2 (1 - \frac{1}{r}) \end{aligned}$$

Substituting these into the energy equation leads to

$$\begin{aligned} 2\lambda_1(3\lambda_2 + 1)\xi^3 - \lambda_1(4\lambda_2 + 1)(2A_S + A_M + 2)\xi^2 \\ + \lambda_2[2\lambda_1(4A_S + 1 + 2A_M) + 2A_S]\xi + A_M\lambda_1 &= 0 \end{aligned}$$

where $\xi = 1/r$.

Now we can do some simple estimations. Assume we have isotropic upstream solar wind with $n = 2$ amu/cc, $\mathbf{v} = [600, 0, 0]$ km/s, $\mathbf{B} = [0, 0, -5]$ nT in GSM coordinates, and $T = 5 \times 10^5$ K. We want to estimate the downstream anisotropy given a density/tangential magnetic field jump of 3.

```
KeyNotes.shock_estimation()
```

Another thing to note is that, if you set the jump ratio to 4 (maximum value when $\gamma = 5/3$) in the above calculations, the downstream anisotropy will become 0.6. This indicates that under this set of upstream conditions, the jump ratio shall never be close to 4 if the anisotropy $T_{\perp}/T_{\parallel} > 1$!

21.3 Subcritical Shocks

Subcritical shocks, also known as laminar shocks, have Mach numbers between 1 and M_c which can be described by the combined action of dispersion and dissipation present in dispersive waves in collisionless plasmas. Subcritical shocks have been believed to be rare in space; they were mostly restrictedly associated to heavy mass loading of the solar wind as is the case in the vicinity of comets and unmagnetised planets like Venus and Mars, in particular at Venus with its dense atmosphere. However, they might be much more frequent simply due to the properties of nonlinear dispersive waves which are capable of steeping and evolving into shocks.

Evolution of subcritical shocks in the latter case is now quite well understood, even though the generation of anomalous resistance and anomalous dissipation below the critical Mach number still poses many unresolved problems. It is well established that the subcritical shock evolves through the various phases of steeping of a low frequency magnetosonic wave the character of which has been identified of being on the whistler mode branch. This steeping process is completely collisionless. The modes propagate against the upstream flow, forming a train of localised wave modes where the steeping is produced by sideband generation of higher spatial harmonics all propagating (approximately) at the same phase (group) velocity such that their amplitudes are in phase and superimpose on the mother wave. When the gradient length of the leading wave packet becomes comparable to the dissipation scale L_d , dissipation sets on. At this time the smaller scale higher harmonic sidebands either outrun the leading wave packet ending up as standing, spatially damped precursor wave modes in front of the shock, or forming a spatially damped trailing wake of the packet. This depends on whether the dispersion is convex or concave (sign of $\partial^2\omega/\partial k^2$). This dispersive effect limits the amplitude of the shock. At the same time the ramp is formed out of the wave packet by the dissipation generated inside the shock.

Generation of dissipation is most likely due to electron current instabilities of the shock ramp current on a scale that is shorter than the ion inertial scale. So far the instability has not yet been identified, but we have given strong arguments that it is the modified two stream instability which signs responsible. The anomalous collision rate is at the lower hybrid frequency in the shock ramp, quite high in this case and sufficient for providing the necessary dissipation for entropy generation, shock heating and compression. In addition, other small scale effects might occur which we have only given a hint on but not discussed in depth.

As long as the shocks are subcritical with $M < M_c$ the distinction between quasi-perpendicular and quasi-parallel shocks is not overwhelmingly important, at least as long as the shock normal angle is far from zero. However, as we will see in the next sections, when the Mach number increases and finally exceeds the critical Mach number, $M > M_c$, the distinction becomes very important.

21.4 Supercritical Perpendicular Shock

Quasi-perpendicular shocks are the first and important family of collisionless magnetised shocks which reflect particles back upstream in order to satisfy the shock conditions. Discussion of the particle dynamics gives clear definition for distinguishing them from quasi-parallel shocks by defining a shock normal angle with respect to the upstream magnetic field. They exist for shock normal angles $< 45^\circ$. Reflected particles at quasi-perpendicular shocks cannot escape far upstream along the magnetic field. After having performed half a gyro-circle back upstream they return to the shock ramp and ultimately traverse it to become members of the downstream plasma population; they also form a foot in front of the shock ramp. We discuss the reflecting shock potential and the explicit shock structure. Most theoretical insight is provided by numerical simulations which confirm reflection, foot formation and reformation of the shock. The latter being caused by steeping of the foot disturbance until the foot itself becomes the shock transition, reflecting particles upstream. Reformation modulates the shock temporarily but on the long terms guarantees its stationarity. Ion and electron dynamics are explicitly discussed in view of the various instabilities involved as well as particle acceleration and shock heating. Finally, a sketchy model of a typical quasi-perpendicular shock transition is provided.

In order to help maintain a shock in the supercritical case the shock must forbid an increasing number of ions to pass across its ramp, which is done by reflecting some particles back upstream. This is not a direct dissipation process, rather it is an emergency act of the shock. It throws a fraction of the incoming ions back upstream and by this reduces both the inflow momentum and energy densities. Clearly, this reflection process slows the shock down by attributing a negative momentum to the shock itself. The shock slips back and thus in the shock frame also reduces the difference velocity to the inflow, i.e. it reduces the Mach number. In addition, however, the reflected ions form an unexpected obstacle for the inflow and in this way reduce the Mach number a second time.

21.4.1 Particle Dynamics

Let's return to the orbit a particle interacting with a supercritical shock when it becomes reflected from the shock. In the simplest possible model one assumes the shock to be a plane surface, and the reflection being specular turning the component v_n of the instantaneous particle velocity \mathbf{v} normal to the shock by 180° , i.e. simply reflecting it. Here we follow the explicit calculation for these idealised conditions as given by S. J. Schwartz, Thomsen, and Gosling (1983) who treated this problem in the most general way. One should, however, keep in mind that the assumption of ideal specular reflection is the extreme limit of what happens in reality, which is no more than a convenient assumption. In fact, reflection must by no means be specular because

1. The shock ramp is not a rigid wall; the particles penetrate into it at least over a distance of a fraction of their gyroradius.
2. Particles interact with waves and even excite waves during this interaction and during their approach of the shock.

Figure 5.1 shows the coordinate frame used at the planar (stationary) shock, with shock normal \hat{n} , magnetic \hat{b} and velocity \hat{v} unit vectors, respectively. Shown are the angles θ_{Bn} , θ_{Vn} , θ_{BV} . The velocity vector \mathbf{V}_{HT} is the de Hoffmann-Teller velocity which lies in the shock plane and is defined in such a way that in the coordinate system moving along the shock plane with velocity \mathbf{V}_{HT} the plasma flow is along the magnetic field, $\mathbf{U} - \mathbf{V}_{HT} = -v_{\parallel}\hat{b}$. Because of the latter reason it is convenient to consider the motion of particles in the de Hoffmann-Teller frame. The guiding centers of the particles in this frame move all along the magnetic field. Hence, using $\mathbf{U} = -U\hat{v}$, $\hat{n} \cdot \hat{v} = \cos \theta_{Vn}$, $\hat{n} \cdot (\hat{b}, \hat{x}, \hat{y}) = (\cos \theta_{Bn}, \sin \theta_{Bn}, 0)$,

$$v_{\parallel} = U \frac{\cos \theta_{Vn}}{\cos \theta_{Bn}}, \quad \mathbf{V}_{HT} = U \left(-\hat{v} + \frac{\cos \theta_{Vn}}{\cos \theta_{Bn}} \hat{b} \right) = \frac{\hat{n} \times \mathbf{U} \times \mathbf{B}}{\hat{n} \cdot \mathbf{B}}, \quad V_{HT,n} \equiv 0 \quad (21.24)$$

The de Hoffmann-Teller velocity is the same to both sides of the shock ramp, because of the continuity of normal component B_n and tangential electric field \mathbf{E}_t . Thus, in the de Hoffmann-Teller frame there is no induction electric field $\mathbf{E} = -\hat{n} \times \mathbf{U} \times \mathbf{B}$. The remaining problem is two-dimensional (because trivially \hat{n} , \hat{b} and $-v_{\parallel}\hat{b}$ are coplanar, which is nothing else but the coplanarity theorem holding under these undisturbed idealized conditions).

In the de Hoffmann-Teller (primed) frame the particle velocity is described by the motion along the magnetic field \hat{b} plus the gyromotion of the particle in the plane perpendicular to \hat{b} :

$$\mathbf{v}'(t) = v'_{\parallel}\hat{b} + v'_{\perp}[\hat{x} \cos(\omega_{ci}t + \phi_0) \mp \hat{y} \sin(\omega_{ci}t + \phi_0)] \quad (21.25)$$

The unit vectors \hat{x} , \hat{y} are along the orthogonal coordinates in the gyration plane of the ion, the phase ϕ_0 accounts for the initial gyro-phase of the ion, and \pm accounts for the direction of the upstream magnetic field being parallel (+) or antiparallel (-) to \hat{b} .

In specular reflection (from a stationary shock) the upstream velocity component along \hat{n} is reversed, and hence (for cold ions) the velocity becomes (???)

$$\mathbf{v}' = -v'_{\parallel}\hat{b} + 2v_{\parallel} \cos \theta_{Bn} \hat{n}$$

which (with $\phi_0 = 0$) yields for the components of the velocity

$$\frac{v'_{\parallel}}{U} = \frac{\cos \theta_{Vn}}{\cos \theta_{Bn}} (2 \cos^2 \theta_{Bn} - 1) \quad \frac{v'_{\perp}}{U} = 2 \sin \theta_{Bn} \cos \theta_{Vn}$$

These expressions can be transformed back into the observer's frame by using \mathbf{V}_{HT} . It is, however, of greater interest to see under which conditions a reflected particle turns around in its upstream motion towards the shock. This happens when the upstream component of

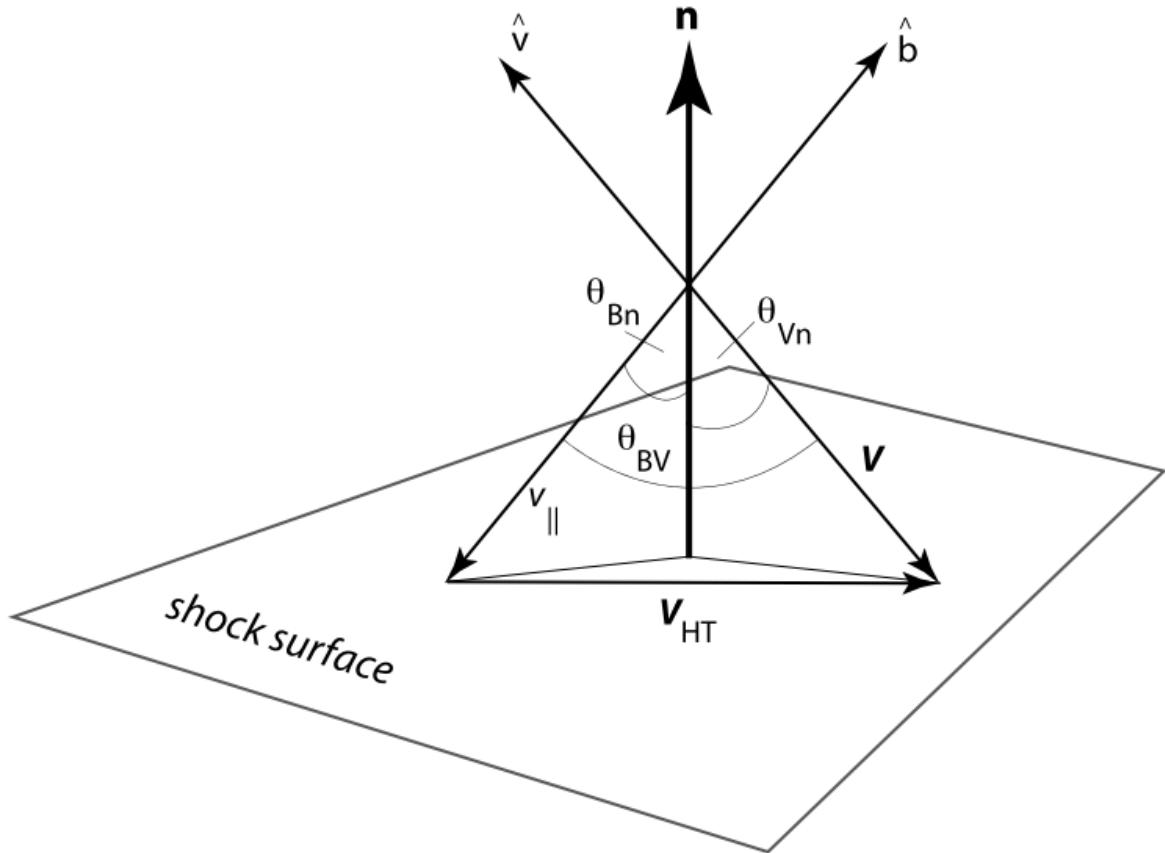


Figure 21.3: The shock coordinate system showing the shock normal \hat{n} , velocity and magnetic field directions \hat{v} , \hat{b} , the three angles θ_{Bn} , θ_{Vn} , θ_{BV} between \hat{b} and \hat{n} , velocity \mathbf{U} and \hat{n} , and velocity \mathbf{U} and \hat{b} , respectively. The velocity \mathbf{V}_{HT} in the shock plane is the deHoffmann-Teller velocity (after S. J. Schwartz, Thomsen, and Gosling 1983).

the velocity $v_x = 0$ of the reflected ion vanishes. For this we need to integrate Equation 21.25 which for $\phi_0 = 0$ yields

$$\mathbf{x}'(t) = v'_\parallel t \hat{b} + \frac{v_\perp}{\omega_{ci}} [\hat{x} \sin \omega_{ci} t \pm \hat{y} (\cos \omega_{ci} t - 1)]$$

Scalar multiplication with \hat{n} yields the ion displacement normal to the shock in upstream direction. The resulting expression

$$\mathbf{x}'_n(t^*) = v'_\parallel t \cos \theta_{Bn} + \frac{v_\perp}{\omega_{ci}} \sin \theta_{Bn} \sin \omega_{ci} t^* = 0 \quad (21.26)$$

vanishes at time t^* when the ion re-encounters the shock with normal velocity

$$v_n(t^*) = v'_\parallel \cos \theta_{Bn} + v_\perp \sin \theta_{Bn} \cos \omega_{ci} t^*$$

The maximum displacement away from the shock in normal direction is obtained when setting this velocity to zero, obtaining for the time t_m at maximum displacement (again including the initial phase here)

$$\omega_{ci} t_m + \phi_0 = \cos^{-1} \left(\frac{1 - 2 \cos^2 \theta_{Bn}}{2 \sin^2 \theta_{Bn}} \right) \quad (21.27)$$

This expression must be inserted in \mathbf{x}_n yielding for the distance a reflected ion with gyroradius $r_{ci} = V/\omega_{ci}$ can achieve in upstream direction

$$\Delta x_n = r_{ci} \cos \theta_{Vn} [(\omega_{ci} t_m + \phi_0)(2 \cos^2 \theta_{Bn} - 1) + 2 \sin^2 \theta_{Bn} \sin(\omega_{ci} t_m + \phi_0)] \quad (21.28)$$

For a perpendicular shock $\theta_{Bn} = 90^\circ$ and $\phi_0 = 0$ this distance is $\Delta x_n \simeq 0.7 r_{ci} \cos \theta_{Vn}$, less than an ion gyro radius. The distance depends on the shock normal angle, decreasing for non-planar shocks. Note that the argument of \cos^{-1} in Equation 21.27 changes sign for $\theta_{Bn} \leq 45^\circ$. Equation 21.26 has solutions for positive upstream turning distances only for shock normal angles $\theta_{Bn} > 45^\circ$, for an initial particle phase $\phi_0 = 0$.¹⁰ Reflected ions can return to the shock in one gyration time only when the magnetic field makes an angle with the shock normal that is larger than this value. For less inclined shock normal angles the reflected ions escape along the magnetic field upstream of the shock and do not return within one gyration. This sharp distinction between shock normal angles $\theta_{Bn} < 45^\circ$ and $\theta_{Bn} > 45^\circ$ thus provides the natural (kinematic specular) discrimination between quasi-perpendicular and quasi-parallel (planar) shocks we were looking for.

The theory of shock particle reflection holds, in this form, only for cold ions, which implies complete neglect of any velocity dispersion and proper gyration of the ions. The ions are considered of just moving all with one and the same oblique flow velocity \mathbf{U} . In a warm

¹⁰A finite initial phase $\phi_0 = 0$ may, however, modify this conclusion shifting the boundary between quasi-parallel and quasi-perpendicular shock to angles larger or smaller than 45° , depending on the sign of the initial phase.

plasma each particle has a different speed, and it is only the group of bulk velocity ions which are described by the above theory. Fortunately, these are the particles which experience the reflecting shock potential strongest and are most vulnerable to specular reflection. When temperature effects will be included, the theory is more involved in a number of ways:

1. The de-Hoffmann-Teller velocity must be redefined to include the microscopic particle motion.
2. The assumption of ideal specular reflection becomes questionable, as the particles themselves become involved into the generation of the shock potential. However, observations in space suggest that, for high flow velocities and supercritical Mach numbers, the simple kinematic reflection is a sufficiently well justified mechanism.

The formal treatment for warm ions are shown in P153 of Balogh and Treumann (2013). The result is that including thermal effect may

1. modify the angle of transition, and
2. will substantially affect the distance up to that a specularly reflected particle at the quasi-perpendicular shock can penetrate the upstream flow, i.e. it affects the width of the quasi-perpendicular shock-foot, even in the case when the reflection process is genuinely specular.

21.4.2 Foot Formation and Acceleration

Shock reflected ions in a quasi-perpendicular shock cannot escape far upstream (see Figure 21.4). Their penetration into the upstream plasma is severely restricted by Equation 21.28. Within this distance the ions perform a gyration orbit before returning to the shock.

Since the reflected ions are about at rest with respect to the inflowing plasma they are sensitive to the inductive convection electric field $\mathbf{E} = -\mathbf{U}_1 \times \mathbf{B}_1$ behaving very similar to pick-up ions and becoming accelerated in the direction of this field to achieve a higher energy (S. J. Schwartz, Thomsen, and Gosling 1983). When returning to the shock their maximum (minimum) achievable energy is

$$\epsilon_{\max} = \frac{m_i}{2} [(v'_\parallel + V_{\text{HT}})^2 + (V_{\text{HT}} \pm v_\perp)^2] \quad (21.29)$$

This energy is larger than their initial energy with that they have initially met the shock ramp and, under favourable conditions, they now might overcome the shock ramp potential and escape downstream. Otherwise, when becoming reflected again, they gain energy in a second round until having picked up sufficient energy for passing the shock ramp.

In addition to this energization of reflected ions which in the first place have not made it across the shock, the reflected ions when gyrating and being accelerated in the convection electric field constitute a current layer just in front of the shock ramp of current density $J_y \sim e n_{i,\text{refl}} v_{y,\text{refl}}$

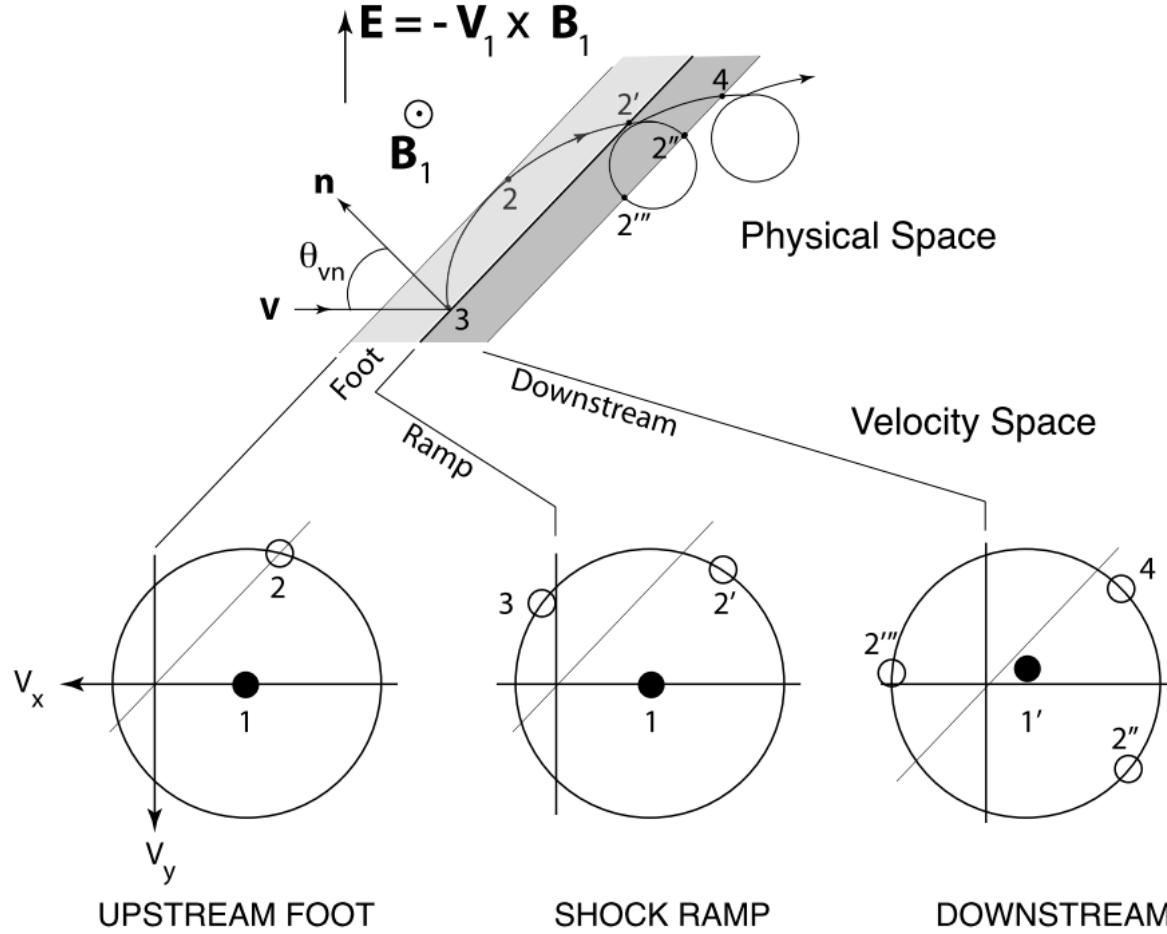


Figure 21.4: Top: Reflected ion orbits in the foot of a quasi-perpendicular shock in real space. The ion impacts under an instantaneous angle θ_{Vn} , is reflected from the infinitely thin shock, performs a further partial gyration in the upstream field \mathbf{B}_1 where it is exposed to the upstream convection electric field $\mathbf{E} = -\mathbf{U}_1 \times \mathbf{B}_1$ in which it is accelerated as is seen from the non-circular section of its orbit in the shock foot. It hits the shock ramp a second time now at energy high enough to overcome the shock potential, passing the ramp and arriving in the compressed downstream magnetic field behind the shock where it performs gyrations of reduced gyro-radius. Bottom: The ion distribution function mapped into velocity space v_x , v_y for the indicated regions in real space, upstream in the foot, at the ramp, and downstream of the shock ramp. Upstream the distribution consists of the incoming dense plasma flow (population 1, dark circle at $v_y = 0$) and the reflected distribution 2 at large negative v_y . At the ramp in addition to the incoming flow 1 and the accelerated distribution 2' there is the newly reflected distribution 3. Behind the ramp in the downstream region the inflow is decelerated 1' and slightly deflected toward non-zero v_y , and the energised passing ions exhibit gyration motions in different instantaneous phases, two of them (2'', 4) directed downstream, one of them (2''') directed upstream. (redrawn after Sckopke et al. 1983).

which gives rise to a foot magnetic field of magnitude $B_{z, \text{foot}} \sim \mu_0 j_y \Delta x_n$. It is clear that this foot ion current, which is essentially a drift current in which only the reflected newly energized ion component participates, constitutes a source of free energy as it violates the energetic minimum state of the inflowing plasma in its frame. Being the source of free energy it can serve as a source for excitation of waves via which it will contribute to filling the lack of dissipation. However, in a quasi-perpendicular shock there are other sources of free energy as well which are not restricted to the foot region.

Figure 21.5 shows a sketch of some of the different free-energy sources and processes across the quasi-perpendicular shock. In addition to the shock-foot current and the presence of the fast cross-magnetic field ion beam there, the shock ramp is of finite thickness. It contains a charge separation electric field E_x which in the supercritical shock is strong enough to reflect the lower energy ions. In addition it accelerates electrons downstream thereby deforming the electron distribution function.

The presence of this field, which has a substantial component perpendicular to the magnetic field, implies that the magnetized electrons with their gyro-radii being smaller than the shock-ramp width experience an electric drift $V_{ye} = -E_x/B_{z2}$ along the shock in the ramp which can be quite substantial giving rise to an electron drift current $J_{ye} = -en_{e,\text{ramp}} V_{ye} = en_{e,\text{ramp}} E_x/B_{z2}$ in the y-direction. This current has again its own contribution to the magnetic field, which at maximum is roughly given by $B_z \sim \mu_0 J_{ye} \Delta x_n$. Here we use the width of the shock ramp. The electron current region might be narrower, of the order of the electron skin depth $d_e = c/\omega_{pe}$. However, as long as we do not know the number of magnetized electrons which are involved into this current nor the width of the electric field region (which must be less than an ion gyro-radius because of ambipolar effects) the above estimate is good enough.

The magnetic field of the electron drift current causes an overshoot in the magnetic field in the shock ramp on the downstream side and a depletion of the field on the upstream side contributing to the steepness of the ramp. When this current becomes strong it contributes to current-driven cross-field instabilities like the modified two-stream instability.

Finally, the mutual interaction of the different particle populations present in the shock at its ramp and behind provide other sources of free energy. A wealth of instabilities and waves is thus expected to be generated inside the shock. To these micro-instabilities add the longer wavelength instabilities which are caused by the plasma and field gradients in this region. These are usually believed to be less important as the crossing time of the shock is shorter than their growth time. However, some of them propagate along the shock and have therefore substantial time to grow and modify the shock profile. In the following we will turn to the discussion of numerical investigations of some of these processes reviewing their current state and provide comparison with observations.

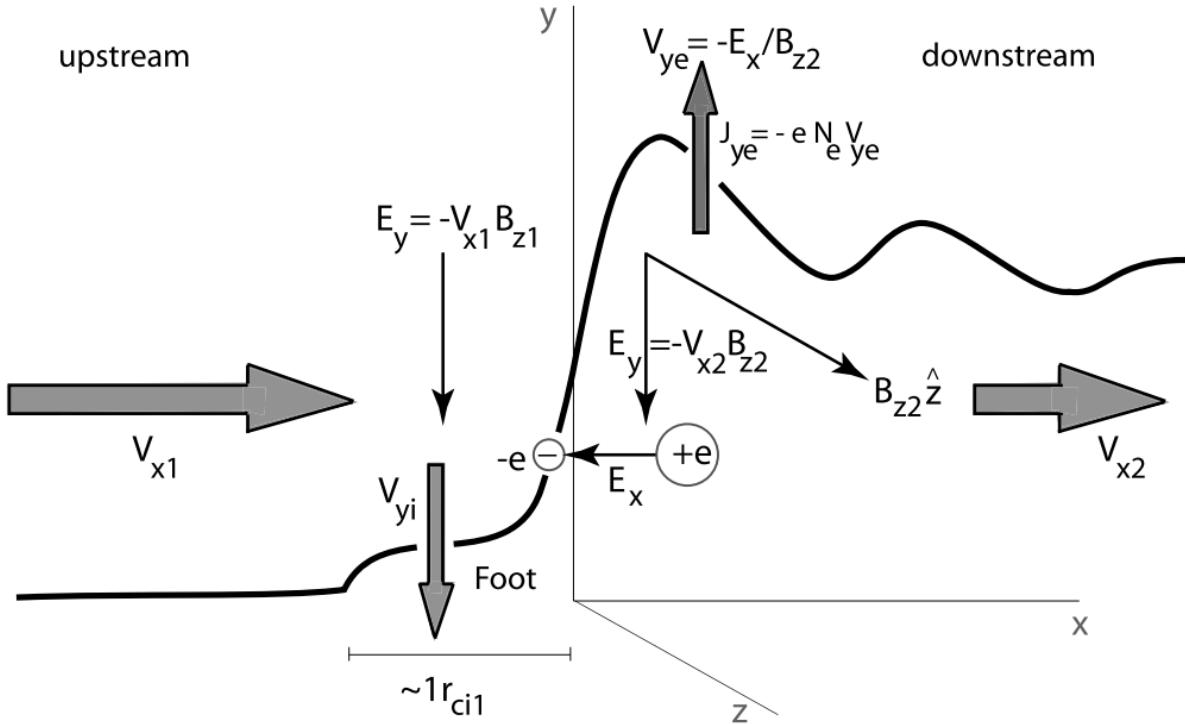


Figure 21.5: Geometry of an ideally perpendicular supercritical shock showing the field structure and sources of free energy. The shock is a compressive structure. The profile of the shock thus stands for the compressed profile of the magnetic field strength $|B|$, the density N , temperature T , and pressure $P = Nk_B T$ of the various components of the plasma. The inflow of velocity \mathbf{V}_1 and outflow of velocity \mathbf{V}_2 is in the x direction, and the magnetic field is in the z direction. Charge separation over an ion gyroradius r_{ci} in the shock ramp magnetic field generates a charge separation electric field E_x along the shock normal which reflects the low-energy ions back upstream. These ions see the convection electric field E_y of the inflow, which is along the shock front, and become accelerated. The magnetic field of the current carried by the accelerated back-streaming ions causes the magnetic foot in front of the shock ramp. The shock electrons are accelerated antiparallel to E_x perpendicular to the magnetic field. The shock electrons also perform an electric field drift in y -direction in the crossed E_x and compressed B_{z2} fields which leads to an electron current J_y along the shock. These different currents are sources of free energy which drives various instabilities in different regions of the perpendicular shock. (Balogh and Treumann 2013)

21.4.3 Shock Potential Drop

One of the important shock parameters is the electric potential drop across the shock ramp – or if it exists also across the shock foot. This potential drop is not necessarily a constant but changes with location along the shock normal. We have already noted that it is due to the different dynamical responses of the inflowing ions and electrons over the scale of the foot and ramp regions. Its theoretical determination is difficult, however when going to the de Hoffmann-Teller frame the bulk motion of the particles is only along the magnetic field, and in the stationary electron equation of motion the $\mathbf{V}_e \times \mathbf{B}$ -term drops out and, to first approximation, the cross shock potential is given by the pressure gradient (when neglecting any contributions from wave fields). The expression is then simply

$$\Delta\Phi(x) = \int_0^x \frac{1}{eN_e(n)} [\nabla \cdot \bar{P}_e(n)] d\mathbf{n} \quad (21.30)$$

Integration is over n along the shock normal \hat{n} . For a gyrotropic electron pressure, valid for length scales longer than an electron gyroradius, $\vec{P}_e = P_{e\perp} \mathbf{I} + (P_{e\parallel} - P_{e\perp}) \mathbf{B}\mathbf{B}/BB$ one obtains (Goodrich and Scudder 1984), taking into account that $\mathbf{E} \cdot \mathbf{B}$ is invariant,

$$\frac{d}{dn} \phi(n) = -\frac{E_\parallel}{\cos \theta_{Bn}} = \frac{1}{eN_e} \left[\frac{d}{dn} P_{e\parallel} - (P_{e\parallel} - P_{e\perp}) \frac{d}{dn} (\ln B) \right]$$

which, when used in the above expression, yields

$$e\Delta\Phi(x) = \int_0^x dn \left\{ \frac{dT_{e\parallel}}{dn} + T_{e\parallel} \frac{d}{dn} \ln \left[\frac{N(n)}{N_1} \frac{B_1}{B(n)} \right] + T_{e\perp} \frac{d}{dn} \ln \left[\frac{B(n)}{B_1} \right] \right\}$$

This expression can approximately be written in terms of the gradient in the electron magnetic moment $\mu_e = T_{e\perp}/B$ as follows:

$$e\Delta\Phi(x) \simeq \Delta(T_{e\parallel} + T_{e\perp}) - \int_0^x dn \frac{d\mu_e(n)}{dn} B(n)$$

with T_e in energy units. When the electron magnetic moment is conserved, the last term disappears, yielding a simple relation for the potential drop $e\Delta\Phi(x) \simeq \Delta(T_{e\parallel} + T_{e\perp})$ as the sum of the changes in electron temperature. The perpendicular temperature change can be expressed as $\Delta T_{e\perp} = T_{e\perp,1} \Delta B/B_1$ which is in terms of the compression of the magnetic field. Non-adiabatic effects contribute via the dropped integral term, which breaks the adiabatic invariant.

The parallel change in temperature is more difficult to express. One could express it in terms of the temperature anisotropy $A_e = T_{e\parallel}/T_{e\perp}$ as has been done by Kuncic et al [2002], and then vary A_e . But this depends on the particular model. It is more important to note that this adiabatic estimate of the potential drop does not account for any dynamical process which generates waves and substructures in the shock. It thus gives only a hint on the order of magnitude of the potential drop across the foot-ramp region in quasi-perpendicular shocks.

21.4.4 Shock Structure

Figure 21.6 shows observations from one of the first unambiguous satellite crossings of a quasi-perpendicular supercritical (magnetosonic Mach number $M_{ms} \sim 4.2$) shock at the Earth's bow shock.

21.4.4.1 Observational evidence

The crossing occurred on an inbound path of the two spacecraft ISEE 1 (upper block of the figure) and ISEE 2 (lower block of the figure) from upstream to downstream in short sequence only minutes apart. In spite of some differences occurring on the short time scale the two shock crossings are about identical, identifying the main shock transition as a spatial and not as a temporal structure. Temporal variations are nevertheless visible on the scale of a fraction of a minute.

In this case, θ_{Bn} is close to 90° prior to shock crossing (in the average $\theta_{Bn} \sim 85^\circ$), and fluctuates afterwards around 90° identifying the shock as quasi-perpendicular. Accordingly, the shock develops a foot in front of the shock ramp as can be seen from the slightly enhanced magnetic field after 22:51 UT in ISEE 1 and similar in ISEE 2, and most interestingly also in the electron pressure. At the same time the bulk flow velocity starts decreasing already, as the result of interaction and retardation in the shock foot region. The foot is also visible in the electron density which increases throughout the foot region, indicating the presence of electrons which, as is suggested by the increase in pressure, must have been heated or accelerated.

The best indication of the presence of the foot is, however, the measurement of energetic ions (second panel from top). These ions are observed first some distance away from the shock but increase drastically in intensity when entering the foot. These are the shock-reflected ions which have been accelerated in the convection electric field in front of the shock ramp. Their occurrence before entrance into the foot is understood when realising that the shock is not perfectly perpendicular. Rather it is quasi-perpendicular such that part of the reflected ions having sufficiently large parallel upstream velocities can escape along the magnetic field a distance larger than the average upstream extension of the foot. For nearly perpendicular shocks, this percentage is small.

The shock ramp in Figure 21.6 is a steep wall in B and P_E , respectively. The electron temperature T_E increases only moderately across the shock while the ion temperature T_P jumps up by more than one magnitude, exceeding T_E downstream behind the shock. This behaviour is due to the accelerated returning foot-ions which pass the shock. P_E , B , and N_E exhibit overshoots behind the shock ramp proper. Farther away from the shock they merge into the highly fluctuating state of lesser density, pressure, and magnetic field that can be described as some kind of turbulence. Clearly, this region is strongly affected by the presence of the shock which forms one of its boundaries, the other boundary being the obstacle (Earth) which is the main responsible for the formation of the shock.

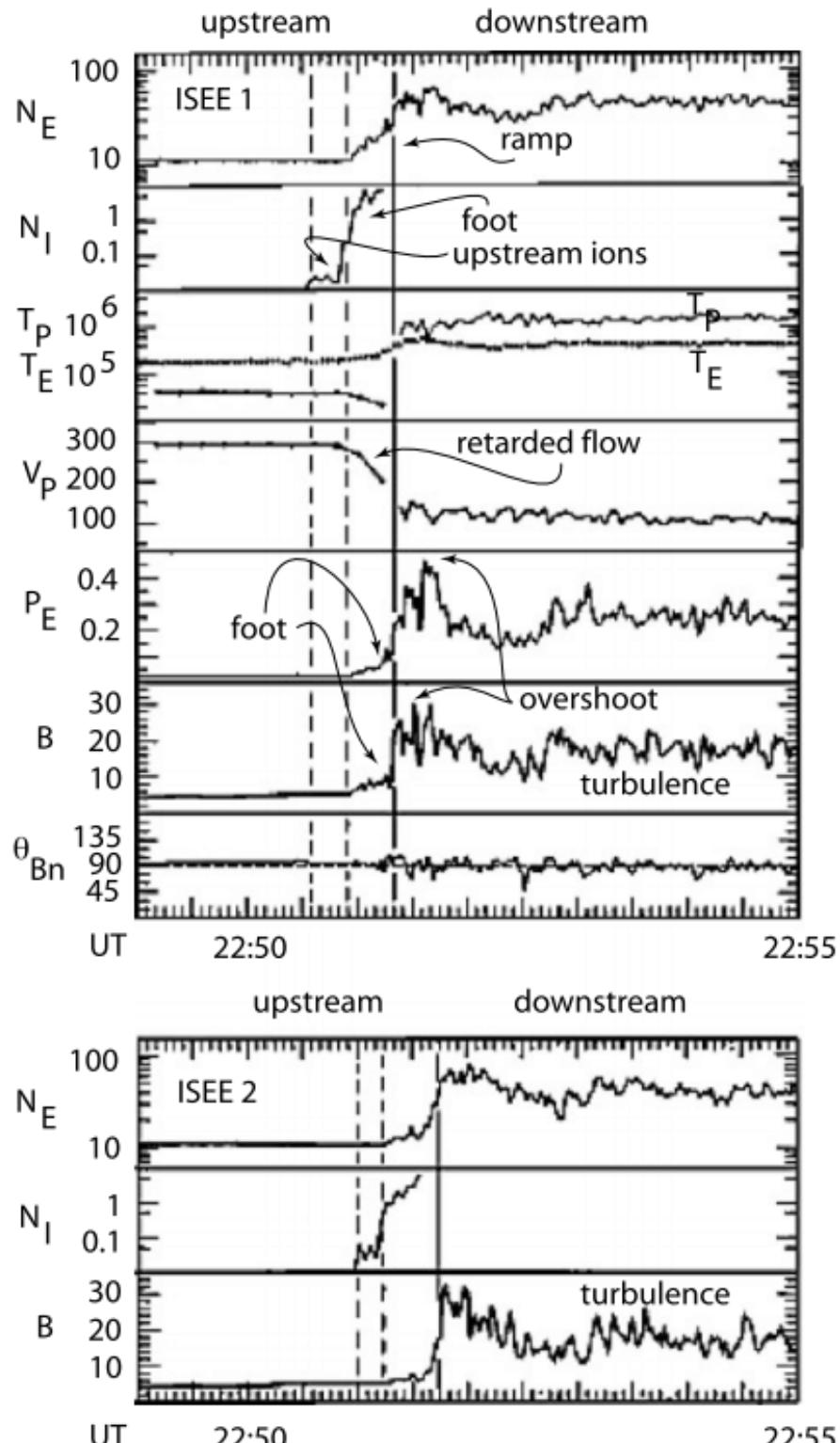


Figure 21.6: Time profiles of plasma and magnetic field parameters across a real quasi-perpendicular shock that had been crossed by the ISEE 1 and 2 spacecraft on November 7, 1977 in near-Earth space (Sckopke et al. 1983). N_E is the electron density, N_i the reflected energetic ion density, both in cm^{-3} , T_P , T_E are proton and electron in K. V_P is the proton (plasma) bulk velocity in km s^{-1} , P_e electron pressure in 10^{-9} N m^{-2} , B the magnitude of the magnetic field in nT, and θ_{Bn} . The vertical lines mark the first appearance of reflected ion, the outer edge of the foot in the magnetic profile, and the ramp in the field magnitude, respectively. The abscissa is the Universal Time UT referring to the measurements. The upper block are observations from ISEE 1, the lower block observations from ISEE 2.

The evidence provided by the described measurements suggests that the quasi-perpendicular shock is a quasi-stationary entity. This should, however, not be taken apodictic. Stationarity depends on the spatial scales as well as the time scales. A shock is a very inhomogeneous subject containing all kinds of spatial scales; being stationary on one scale does not imply that it is stationary on another scale. For a shock like the Earth's bow shock considered over times of days, weeks or years the shock is of course a stationary subject. However on shorter time scales of the order of flow transition times this may not be the case. A subcritical shock may well be stationary on long and short time scales. However, for a supercritical shock the conditions for forming a stationary state are quite subtle. From a single spacecraft passage like that described above it cannot be concluded to what extent, i.e. on which time scale and on which spatial scale and under which external conditions (Mach number, angle, shock potential, plasma- ,...) the observed shock can be considered to be stationary (a discussion of the various scales has been given, e.g, by Galeev et al, 1988). Comparison between the two ISEE spacecraft already shows that the small-scale details as have been detected by both spacecraft are very different. This suggests that – in this case – on time scales less than a minute variations in the shock structure must be expected.

Generally spoken, one must be prepared to consider the shock locally (on the ion gyroscale) and temporarily (on the ion-cyclotron frequency scale) as a non-stationary phenomenon (this has been realised first by Morse et al, 1972) which depends on many competing processes and, most important though only secondarily related to non-stationarity, a shock as a whole is *not in thermal equilibrium*. It needs to be driven by some energy source external to the shock in order to be maintained. It will thus be very sensitive to small changes in the external parameters and will permanently try to escape the non-equilibrium state and to approach equilibrium. Since its non-equilibrium is maintained by the conditions in the flow, it is these conditions which determine the time scales over which a shock evolves, re-evolves and changes its state. Real supercritical shocks, whether quasi-perpendicular or quasi-parallel, are in a permanently evolving state and thus are intrinsically nonstationary.

21.4.4.2 Simulation studies

Perpendicular or quasi-perpendicular collisionless shocks are relatively easy to treat in numerical simulations. Already from the first one-dimensional numerical experiments on collisionless shocks it became clear that such shocks have a very particular structure. This structure, which we have describe in simplified version in Figure 21.5 and which could to some extent also be inferred from the observations of Figure 21.6, becomes ever more pronounced the more refined the resolution becomes and the better the shorter scales can be resolved.

As already mentioned, collisionless shocks are in thermodynamic non-equilibrium and therefore can only evolve if a free energy source exists and if the processes are violent enough to build up and maintain a shock. Usually in a freely evolving system the free energy causes fluctuations which serve dissipating and redistributing the free energy towards thermodynamic and thermal equilibria. Thermal equilibria are characterised by equal temperatures among

the different components, e.g. $T_e = T_i$ which is clearly not given in the vicinity of a shock as seen from Figure 21.6. Thermodynamic equilibria are characterized by Gaussian distributions for all components of the plasma. To check this requires information about the phase space distribution of particles. Shocks contain many differing particle distributions, heated, top-flat, beam distributions, long energetic tails, and truncated as well as gyrating distributions which we will encounter later. Consequently, they are far from thermodynamic equilibrium.

For a shock to evolve the amount of free energy needed to dissipate is so large that fluctuations are unable to exercise their duty. This happens at large Mach numbers. The shock itself takes over the duty of providing dissipation. It does it in providing all kinds of scales such short that a multitude of dissipative processes can set on.

Scales

For a quasi-perpendicular shock propagating and evolving in a high- β plasma there is a hierarchy of such scales available. The kinetic $\beta_{\text{kin}\perp} = 2\mu_0 N m_i V_n^2 / 2B^2 \equiv M^2 > 1$ implies that the kinetic energy in the flow exceeds the magnetic energy. Hence the flow dominates the magnetic field, which is transported by the flow. In plasmas with $\beta_{\text{kin}\perp} = M^2 < 1$ the magnetic field dominates the dynamics, and shock waves perpendicular to the magnetic field cannot evolve. Parallel shocks are basically electrostatic in the $\beta_{\text{kin}\perp} \gg 1$ -case and can evolve when the flow is sufficiently fast along the field, as is observed in the auroral magnetospheres of the magnetised planets in the heliosphere. On the other hand, for large Mach numbers and $\beta \geq 1$ conditions shocks do exist, as the example of the solar wind shows.

The different scales can be organized with respect to the different regions of the shock.

1. The macroscopic scale of the foot region, which determines the width of the foot, is the ion gyroradius based on the inflow velocity $r_{ci,1} = V_1/\omega_{ci,1}$. With the slight modification of replacing the upstream magnetic field with the (inhomogeneous) ramp magnetic field $B_r(x)$ this also becomes approximately the scale of the macroscopic electric potential drop in the ramp, $\Delta_{\phi,r} \sim r_{ci,r} \sim V_1/\omega_{ci,r}$. This is an important difference which can be taken as a *golden rule* for estimates of the structure of quasi-perpendicular shocks even though, of course, these values are dynamical values which change from position to position across the foot and ramp. Observation evidence from the CLUSTER satellites are given in (Bale, Mozer, and Horbury 2003).
2. The ion inertial length $d_i = c/\omega_{pi}$, which is also a function of space inside the ramp because of the steep density increase $N(x)$. It determines the dispersive properties of the fast magnetosonic wave which is locally responsible for steeping and shock ramp formation.
3. The thermal ion gyroradius $r_{ci} = v_{th,i}/\omega_{ci}$. It determines the transition from unmagnetized to magnetized ions and from non-adiabatic to adiabatic heating of the ions.

4. The density gradient scale $L_P = (\nabla_x \ln P)^{-1}$. It determines the importance of drift waves along the shock which, when excited, structure the shock in the third dimension perpendicular to the shock normal and the magnetic field.
5. The electron inertial length $d_e = c/\omega_{pe}$. It is the scale length of whistlers which are excited in front of the shock and are generally believed to play an essential role in shock dynamics.
6. The thermal electron gyroradius $r_{ce} = v_{th,e}/\omega_{ce}$. It determines whether electrons behave magnetized or non-magnetized. In the shock they are usually magnetized under all conditions of interest. However, when non-adiabatic heating becomes important for electrons it takes place on scales comparable to r_{ce} .
7. The Debye length λ_D . It determines the dispersive properties of ion acoustic waves which are responsible for anomalous resistivity and for smaller scale density substructures in the shock like the phase space holes which evolve on scales of several Debye lengths. It also determines the scales of the Buneman two-stream (BTS) and modified two-stream (MTS) instabilities which are the two most important instabilities in the shock foot.

The importance of some of these scales has been discussed by C. Kennel, Edmiston, and Hada (1985) assuming that some mostly anomalous resistance has been generated in the plasma. In this case the speed of the fast magnetosonic wave, which is responsible for fast shock formation, is written as

$$c_{ms}^2 = c_{ia}^2 + \frac{V_A^2}{1 + k^2 R^2}, \quad R = \begin{cases} R_\eta = (\eta/\mu_0)(k/\omega) & \eta \neq 0 \\ d_e = c/\omega_{pe} & \eta \rightarrow 0 \end{cases}$$

taking explicitly care of the dispersion of the wave which leads to wave steeping. The macroscopic scale of shock formation enters here through the definition of R which in the collisionless case becomes the electron skin depth. Starting from infinity far away from the shock one seeks for growing solutions of the linear magnetic disturbance $b_z \sim \exp \lambda_x$ in the stationary point equation

$$R_e^2 b_z'' + R_\eta b_z' = D b_z, \quad D \equiv \frac{1 - M^2}{1 - c_{ia}^2/V^2}$$

where the prime $' \equiv \partial/\partial x$ indicates derivation with respect to x . With $b_z \rightarrow 0$ for $x \rightarrow -\infty$ this yields for the spatial growth rate

$$\lambda_> = -\frac{R_\eta}{2d_e^2} + \left[\frac{D}{d_e^2} + \left(\frac{R_\eta}{2d_e^2} \right)^2 \right]^{1/2} \rightarrow \frac{D^{1/2}}{d_e} \quad \text{for } R_\eta \ll d_e$$

which identifies the approximate shock transition scale as proportional to the electron skin depth, $\Delta \simeq c/\omega_{pe} D^{1/2}$, just what one intuitively would believe to happen for freely moving electrons and ions. Since the upstream sound speed $c_{ia} \ll V$ is small compared with the fast flow V , we have $D \approx 1 - M^{-2}$, and the shock ramp width becomes slightly larger than the electron skin depth $d_e = c/\omega_{pe}$, viz.

$$\Delta \simeq M(M^2 - 1)^{-1/2} d_e$$

For large Mach numbers this width approaches d_e . However, we have already seen that at large Mach numbers the competition between dispersion and dissipation does not hold anymore in this simple way.

With increasing wave number k the fast magnetosonic mode merges into the whistler branch with its convex dispersion curve. This implies that dispersive whistler waves will outrun the shock becoming precursors of the shock. Whistlers propagate only outside their resonance cone. The limiting angle between k and the magnetic field B for which the whistler outruns the shock is given by $\theta_{wh,lim} \leq \cos^{-1}[M_A(m_e/m_i)^{1/2}]$, artificially limiting the Alfvén Mach number $M_A = V/V_A < 43$. A linear whistler wave precursor can stand in front of the quasi-perpendicular shock as long as the Mach number $M < M_{wh}$ is smaller than the critical whistler Mach number

$$M_{wh} = \frac{1}{2} \left(\frac{m_i}{m_e} \right)^{1/2} |\cos \theta_{Bn}| \quad \text{linear case}$$

$$M_{wh,nl} = M_{wh} \sqrt{2} \left[1 - \frac{27\beta}{128M_{wh}} \right]^{1/3} \quad \text{nonlinear case}$$

The second expression results when including the nonlinear growth of the whistler amplitude during the steeping process [for the derivation of this expression see, e.g., Kazantsev, 1961; Krasnoselskikh et al, 2002].

In 1D simulations with all quantities changing only along the shock normal \hat{n} and the \mathbf{k} -vectors of waves along \hat{n} as well, one chooses angles between (\mathbf{k}, \hat{n}) and \mathbf{B} larger than this in order to have clean effects which are not polluted by those whistlers. However, the maximum phase speed of whistlers does not exceed the Alfvén speed by much. Hence, as long as the upstream velocity is less than this maximum whistler speed, a standing whistler precursor will be attached to the shock in front. When the upstream velocity exceeds this velocity, phase standing whistlers become impossible. This happens at the critical whistler Mach number. The shock structure becomes more complicated then by forming shock substructures [Galeev+, 1988] on scales of $d_e = c/\omega_{ce}$, and the shock might become non-stationary [Krasnoselskikh+, 2002].

1D Structure

One-dimensional observations as those presented in Figure 21.6 confirm the theoretical prediction of the gross structure of a quasi-perpendicular shock. Observations are possibly limited by not resolving the spatial structure of the shock on smaller scales and the lack of inference about the evolution of the shock. To achieve a clearer picture of both the structure and the evolution, the observations must be supported by numerical simulations.

The Shock Transition Scale

The width of the shock transition is particularly important in its relation to the width of the electrostatic potential drop across the shock. There are essentially three transition scales: the magnetic scale ΔB , the density scale ΔN , and the electric potential scale ΔE . Since the

shock is not in pressure equilibrium, the first two scales are not necessarily be proportional to each other. However, the electric field and density gradient might be related, so one expects that $\Delta N \sim \Delta E$ even though this is not necessarily so, in particular not when instabilities arise which cause very small scale electric field gradients. In principle one can distinguish three different cases [Lembege et al, 1999] which describe different physics:

1. $\Delta E \gg \Delta B$. This is a case that has been reported to have been observed in Bow shock crossings. The magnetic ramp is much steeper in this case than the structure of the electric field. The latter smears out over the foot and ramp regions. In this case the electrons will behave adiabatically, while the ions may be only partially or even non-magnetized.
2. $\Delta E \sim \Delta B$. In this case there will be a significant deviation from adiabatic behavior of the electrons in the shock transition. Electron heating and motion will not be adiabatic anymore, and the electron distribution will significantly be disturbed.
3. $\Delta E \ll \Delta B$. This case which is also called the ‘isomagnetic’ transition [Eselevich, 1982; Kennel et al, 1985] corresponds to shock transitions with electrostatic substructuring which are sometimes also called *subshocks*.

The generation of these quasi-perpendicular shock structures by a variety of instabilities might even turn the shock foot and ramp regions into regions where large anomalous collisions and thus resistances are generated as the result of wave-particle interactions. In this case the shock returns to become resistive again due to preventing large numbers of reflected ions from passing across the steep shock ramp and large shock potential, using the kinetic energy of the reflected particle population for the generation of a broad wave spectrum which acts to scatter the particles around in the foot and ramp regions and, possibly, also up to some distance in the transition region behind the ramp. This kind of confinement of reflected particles over long times will then be sufficiently long for providing the heating and dissipation which is required for sustaining a resistive shock which, then, is the result of the combined action of ion viscosity and anomalous resistivity, i.e. *anomalous collisions*. In addition, the scattering of the trapped reflected particle population necessarily results in plasma heating, and some particles will become accelerated to high velocities in these interactions as well. It is then possible that these particles provide the seed population for energetic particles which have been accelerated to high energies in the well-known shock-Fermi-one and shock-Fermi-two acceleration mechanisms.

21.4.5 Shock Reformation

Supercritical shocks do under certain conditions reform themselves periodically – or quasi-periodically –, which is kind of a non-stationarity of the shock that does not destroy the shock but, at the contrary, keeps it intact in a temporarily changing way. We will come later to the problem of real non-stationarity.

For illustration, Figure 21.7, on its left, shows a low mass-ratio example of the temporary evolution of a shock during shock reformation in a magnetic field stack plot. On its right the

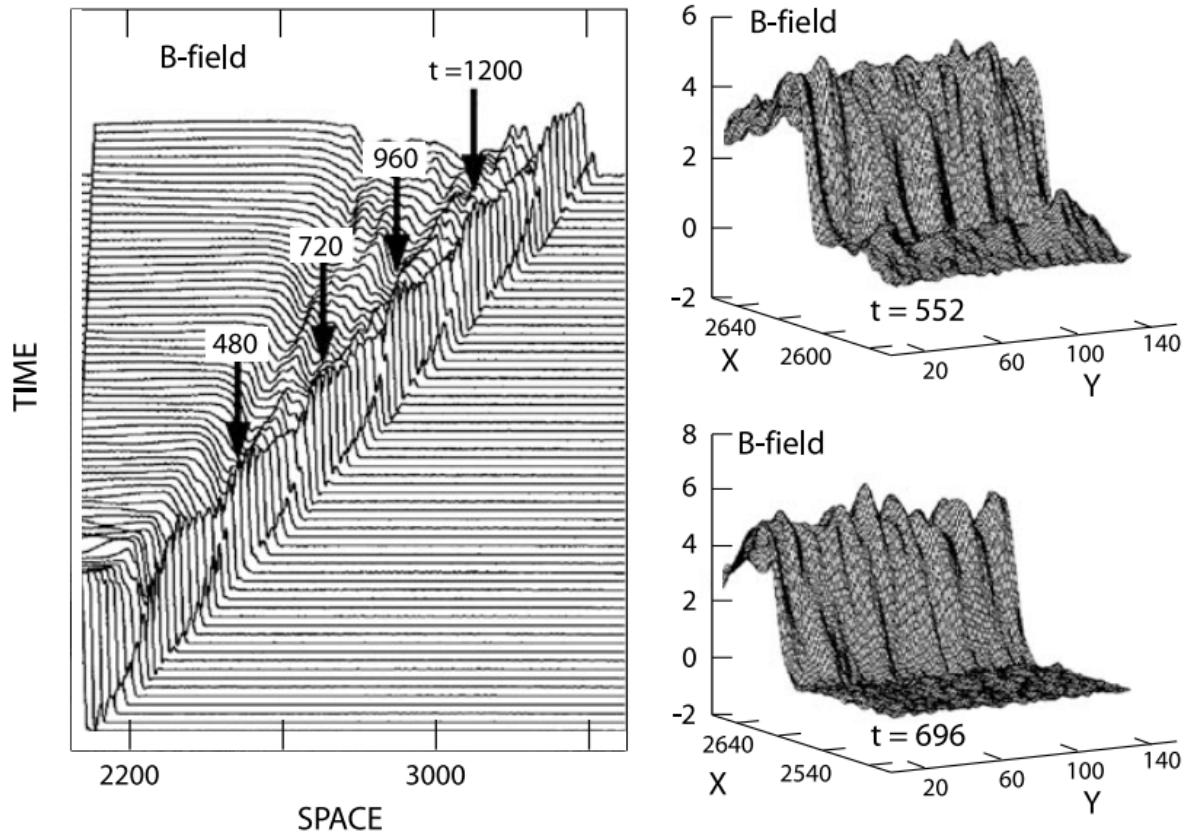


Figure 21.7: Magnetic field from full particle PIC simulations of shock reformation [Lembege & Savoini, 2002]. Left: Reformation cycles of the magnetic field in the shock. Time is measured in inverse electron plasma frequencies ω_{pe}^{-1} . The reformation times are indicated by the arrows in the plot with time given when the cycle is complete. Right: Two snapshots in time of the view of the shock front in the magnetic field at reformation. The interesting finding is that the front in this two-dimensional view is not a smooth plane but is quite distinctly structured in space and at the same time evolving.

structure of the shock ramp at two different two reformation times is plotted along the shock ramp. There is a distinct reformation cycle in this simulation and also a distinct structure of the ramp/shock front in the tangential direction which is far from being smooth, a fact to which we will return during discussion of non-stationarity of shocks. The shock not only reforms cyclically in time, it also develops ripples along its surface which travel like waves along the shock.

Full PIC simulations show that reformation at least occurs at small ion- $\beta_i \sim 0.2$. In these simulations the shock is produced by injecting a uniform plasma from -x and letting it reflect from a stationary wall at the right end of the simulation box. The plasma carries a uniform magnetic field in the (x,z)-plane, and the plasma is continuously injected in the +x-direction.¹¹ Since the right-hand reflecting boundary is stationary the shock, which is generated via the ion-ion beam instability in the interaction of the incoming and reflected ion beams, moves to the left at velocity given by the supercritical shock Mach number $M_A \sim 4.5$. The upstream plasma has $\beta_i = \beta_e = 0.05$, and the shock normal angle is $\theta_{Bn} = 87^\circ$.

Two runs of these simulations are shown in Figure 21.8, one is for a mass ratio of 400, the other for a mass ratio of 1840. The left-hand side of the figure shows stack plots of time profiles of the nearly perpendicular magnetic field B_z with time running in equidistant units upward on the ordinate. Since the plasma is injected from the left and reflected at the right boundary the shock is seen to move from the right to the left. The magnetic profiles are strikingly similar for both mass ratios. In both cases a relatively flat foot develops in front of the steeper shock ramp caused by the shock reflected ions. The magnetic field of this foot starts itself increasing with time with growth being strongest close to the upstream edge of the foot until the foot field becomes so strong that it replaces the former shock ramp and itself becomes the new and displaced shock ramp.

This is seen most clearly in the upper low mass-ratio part of the figure. The foot takes over, steepens and becomes itself the shock. One can recognise in addition that, even earlier, the intense foot already had started reflecting ions by himself and developing its own flat pre-foot region. This pre-foot evolves readily to become the next foot, while the old ramps become part of the downstream turbulence.

The one-dimensional full particle PIC simulations by Umeda and Yamazaki (2006) at Mach number $M_A = 10$ and medium mass ratio $m_i/m_e = 100$ throw additional light on the reformation process when keeping in mind that reformation is not as strongly dependent on the mass ratio as originally believed. Figure 21.9 shows a collection of their results which this time are represented in the shock frame of reference.

The simulations have been performed by assuming an initial Rankine-Hugoniot equilibrium in the PIC code. The non-physicality of this initialisation is manifested in the initial evolution over the first few ion cyclotron periods. During this time the simulation adjusts itself to the correct physics, and the non-physical disturbance decays. The shock frame has shifted by this

¹¹This shock setup is common in local simulations, but is quite different from the global shock simulations. We are looking at new methods for setting up the local shock runs.

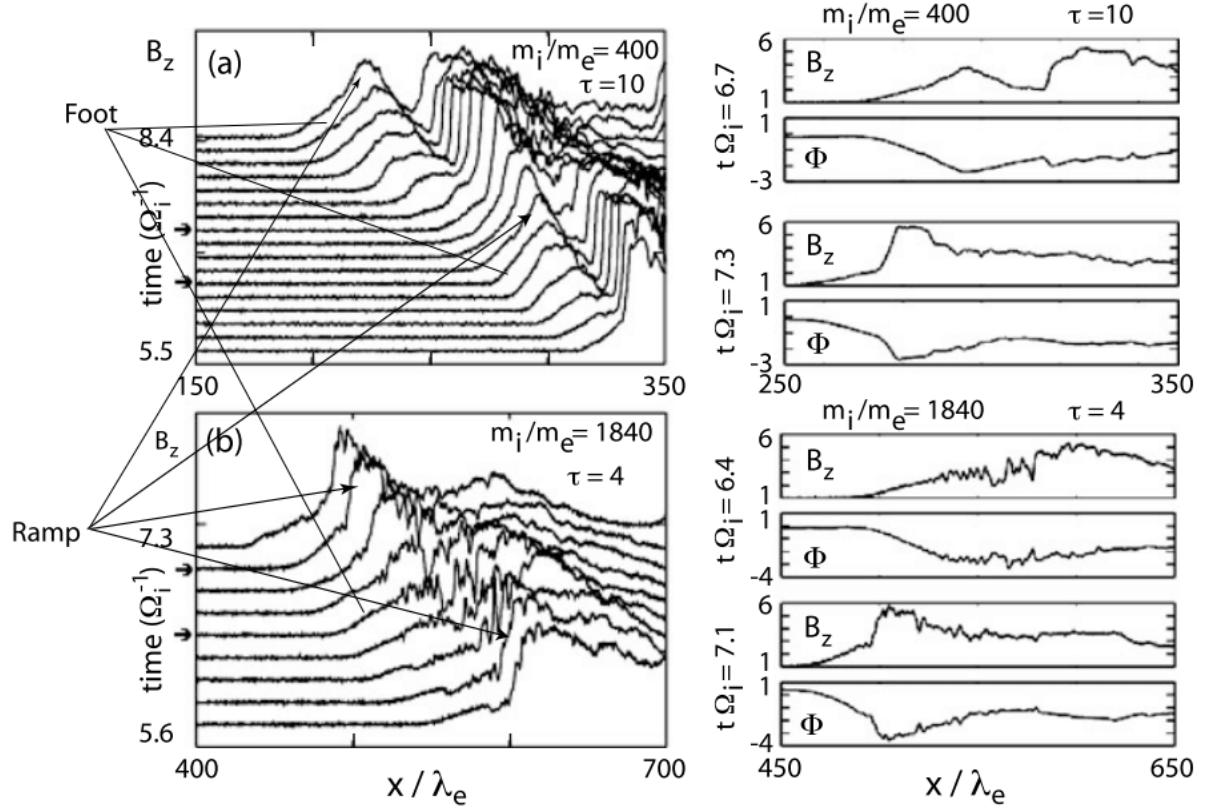


Figure 21.8: Left: One-dimensional PIC simulations (after Scholer, Shinohara, and Matsukiyo 2003) of quasi-perpendicular $\theta_{Bn} = 87^\circ$ shock reformation for mass ratios $m_i/m_e = 400$ and 1840 . Time is in $\Omega_i^{-1} = \omega_{ci}^{-1}$, space in d_e . The parameter $\tau = \omega_{pe}^2/\omega_{ce}^2$ is taken small in both cases. The higher mass ratio shows a violent time evolution because of the high electron mobility. Reformation of the shock is due to evolution of the shock feet. The original foot region builds up until becoming itself the shock assuming the role of the ramp. Afterwards a secondary foot evolves in front of this new ramp. Right: Spatial shock profiles at two time sections (see arrows on the left). The higher mass ratio run shows a more subtle structure in B_z and shock potential ϕ , but the gross features are similar. The potential drop exists already in the foot but the main drop occurs in the ramp. The lower mass ratio has a more concentrated foot region.

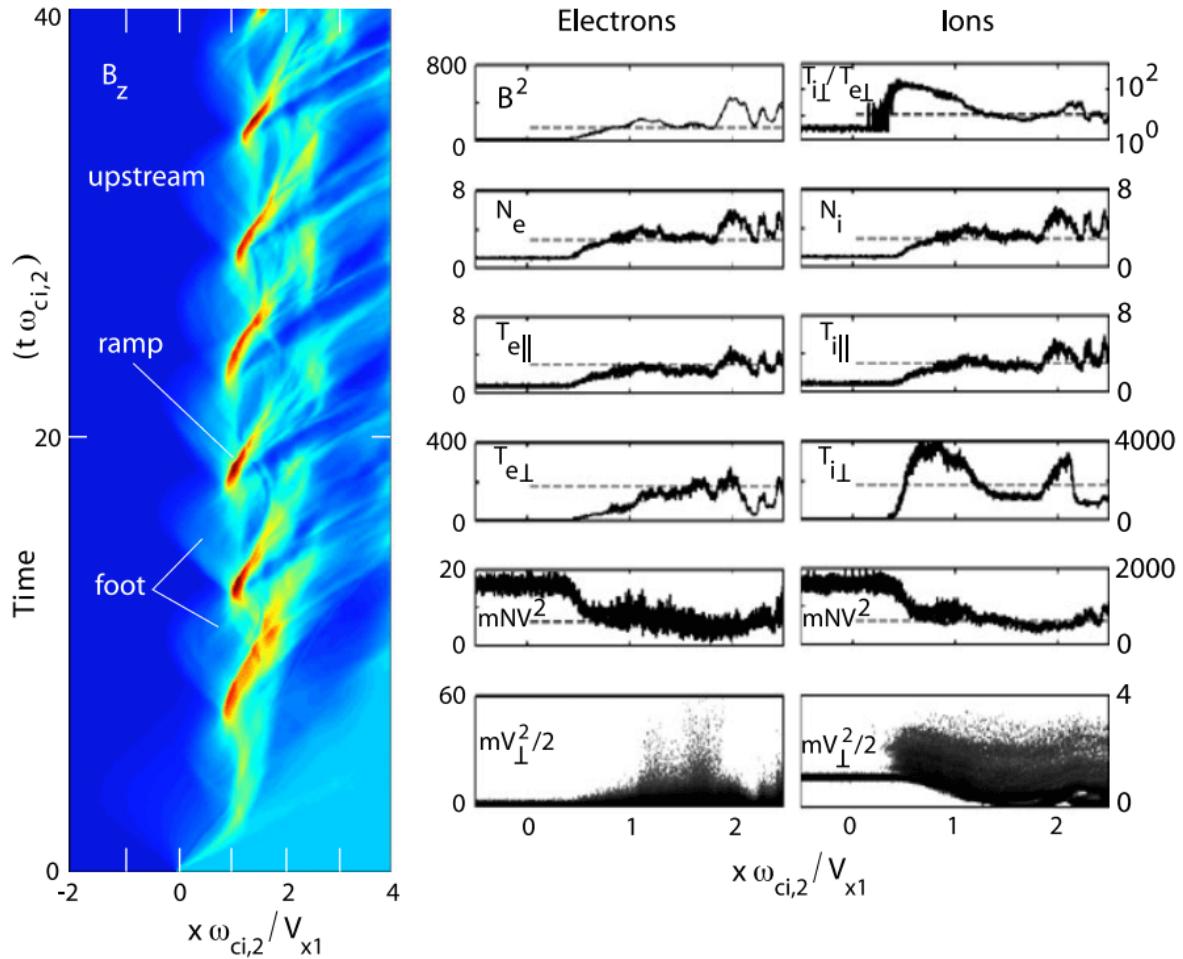


Figure 21.9: Left: Evolution of the magnetic field in a quasi-perpendicular high Mach number $M_A = 10$ PIC simulation (after Umeda and Yamazaki 2006). Here the presentation is in the shock frame of reference, and the shock has been initialised by assuming Rankine-Hugoniot initial jump conditions. The non-physical nature of this assumption is visible in the initial evolution and fast displacement of the shock to the right. After the initial unphysical disturbance has disappeared a self-consistent physical state is reached in which the shock quasi-periodically reforms itself. The competition between the shock foot and ramp formation is nicely seen in the colour plot of the magnetic field B_z . Right: Electron and ion plasma parameters in computational units. Of interest is only their relative behaviour, not the absolute values. The profiles are taken at time $t\omega_{ci,2} = 38.1$ (No 2π factor!). They show the compression of the plasma and heating of electrons and ions. Parallel electron and ion heating is comparable, but ions are heating much stronger than electrons in perpendicular direction causing a large perpendicular temperature anisotropy downstream of the shock.

to a new position, which in the shock frame is located farther downstream (which takes into account of the moment transferred to the shock by the reflection of the upstream ions who lower the shock speed).

The further evolution of the shock shows the quasi-periodic reformation and the play between the foot and the ramp formation. The periodicity is roughly $\sim 6\omega_{ci,\text{down}}^{-1} \sim T_{i,\text{down}}$. When the foot takes over to become the ramp, the ramp jumps ahead in a fraction of this time. Afterwards the formation of the foot retards the ramp motion, and the ramp softens and displaces itself downstream to become a downstream moving spectrum of magnetic oscillations which is injected into the downstream region in the form of wave packets. The various plasma parameters in the left part of the figure show in addition the compression of plasma and field, and the dominance of perpendicular ion heating which is, of course, due to the accelerated foot ions which pass into downstream.

21.4.6 Ion Dynamics

21.4.6.1 Ion Dynamics in Shock Reformation

Reflected particles are forming a foot on the perpendicular shock and may contribute to the reformation of the shock.

Figure 21.10 shows the ion phase space plots for the two mass ratios $m_i/m_e = 400$ (top) and 1840 (bottom) respectively. Both plots show just the enlarged shock foot transition region over the same scale of $100d_e$. The electron $\beta_e = 0.2$ has been kept constant in both simulations, while the ion β_i has been changed. Only the normal component of the ion velocity is shown for the nearly perpendicular supercritical shock. In both plots the magnetic field B_z has been drawn as a thin continuous line showing the magnetic shock profile over the spatial distance Δx .

The upper (low-mass-ratio) low- β_i panel shows the cold dense ion inflow at velocity $v_{ix} \sim 5$ (in units of the upstream Alfvén velocity) being retarded to nearly Mach number 1 already when entering the foot. This retardation is due to its interaction with the intense but cold (narrow in velocity space) reflected ion beam which is seen as the narrow negative v_{xi} -velocity beam originating from the shock ramp. This reflected ion beam needs a certain distance to interact with the upstream plasma inflow. This distance is the length the beam-beam excited waves need to grow. But once the interaction becomes strong enough, the reflected ions are scattered by the waves into a hot ion clump in addition to being turned around by gyration. Both effects cause a reduction in velocity vx of the reflected ions which, being accelerated by the convection electric field, turn to flow in y direction and cause the magnetic bump that develops in this region of the foot. It is interesting to remark that in the (v_x, x) -plane the reflected ions close with the upstream flow into an hot ion ring distribution (vortex) just in front of the ramp of which the hot ion clump that brakes the inflow is the upstream boundary. Behind the ramp, which is the point of bifurcation of the ion distribution, i.e. the location where the reflection is

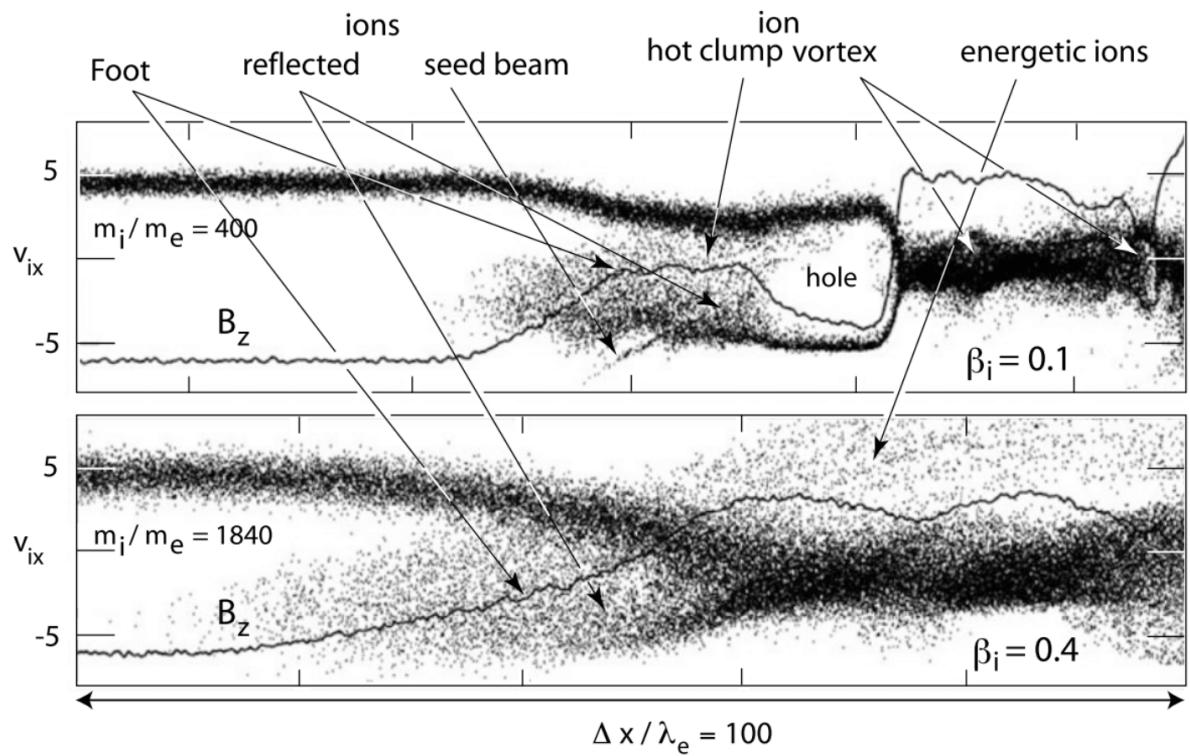


Figure 21.10: Ion phase space plot from 1D full PIC simulations Scholer, Shinohara, and Matsukiyo (2003) in the low- β regime with $M_A = 4.2$.

at work, a broad hot ion distribution arises which at some locations shows rudimentary remains of ion vortices from former reformation cycles. Their magnetic signatures are the dips seen in the magnetic field. The next reformation cycle can be expected to completely close the ion vortex in the foot and to transform the ramp from its current position to the position of the foot. The first sign of this process is already seen in the foot ion distribution, which shows the birth of a faint new reflected ion beam at high negative speeds. This beam is not participating in the formation of the ring but serves as the seed of the newly reflected population.

The same behaviour is found in large mass ratio simulations as long as β_i is small. This is obvious from the large mass-ratio magnetic field shown in Figure 21.8. As long as β_i remains to be small, the shock undergoes reformations also for realistic mass ratios. In other words, as long as the plasma is relatively cool the real shocks found in nature should develop feet which at a later time quasi-periodically become the shock ramp.

This changes completely, when β_i increases as is suggested from consideration of the lower panel in Figure 21.10. There a realistic mass ratio has been assumed, but $\beta_i = 0.4$, and no reformation is observed, at least not during the realisable simulation times. Instead, the shock develops a very long foot region that is extended twice as far into the upstream region as in the low- β_i case. Clearly, the high ion temperature smears out the reflected ion population over the entire gap region between the upstream and reflected beam regions, and no vortex can develop. This implies that the foot remains smooth and does not evolve into a ramp.

However, inspecting the panels of Figure 21.10 it becomes immediately clear that suppression of reformation is a relative process. Reformation will be suppressed only when the thermal speed $v_{\text{th},i}$ of the ions is large enough to bridge the gap between the reflected and incoming ion beams, i.e. large enough to fill the hole. Semi-empirically one can establish a condition for shock reformation as $v_i < \alpha V_{n1}$ when taking into account that the normal speed of incoming ions is simply specularly turned negative. Since this is never exactly the case, the coefficient will roughly be in the interval $1.5 < \alpha < 2$. This condition for reformation to occur can be written as

$$M_A > \frac{\beta_i^{1/2}}{\alpha} \quad (21.31)$$

where the Alfvén Mach number is defined on V_{1n} . The larger the Mach number becomes the less suppression of reformation will play a role, and at high Mach numbers one expects that either reformation becomes a normal process or that other time-dependent processes set in which lead to a non-stationary state where the shock reformation becomes a chaotic and unpredictable process. As we have argued earlier this is quite normal as the shock is thermodynamically and thermally in a non-equilibrium state: it is a region where electrons and ions have violently different temperatures; it is not in pressure equilibrium; upstream and downstream temperatures are different; and it hosts a number of non-Maxwellian phase space distributions all concentrated in a small volume of real space.

21.4.6.2 Ion instabilities and ion waves in the foot and ramp region

From the field and particle properties demonstrated in simulations of quasi-perpendicular supercritical shocks, it has become obvious that in the different regions of the shock transition the particle distributions carry free energy. This is true for the foot region, the shock transition and overshoot as well as the downstream region. And it is true for both species, electrons and ions. This free energy is the source of a number of instabilities which excite waves of different kinds in the various shock transition regions which can be measured.

The instabilities are not primarily driven by temperature anisotropy, but rather by differences in bulk flow properties of the different species. We therefore ignore the temperature anisotropy differences even though such instabilities may arise, in particular when ion-whistlers can be excited by temperature anisotropy. To some extent the occurrence of two (counter-streaming in the x-direction of shock normal) ion beams already fakes a bulk temperature on the ion component thus generating some relationship between a two-beam situation and a temperature anisotropy. Similarly transversely heated reflected ions superposed on a low perpendicular temperature inflowing ion background fakes a perpendicular temperature anisotropy. For our purposes, however, the bulk flow differences are more interesting and have, in fact, been more closely investigated right from the beginning.

Let us first consider the foot region. The free energy sources here are the relative drifts between the incoming electrons and reflected ions and the incoming electrons and incoming ions. The presence of the reflected ions causes a decrease of the ion bulk velocity in the foot region. This implies that the incoming electrons are decelerated so that the current in shock normal direction is zero, i.e. the flow is current-free in normal direction. However, this has the consequence that a relative bulk velocity between electrons and reflected ions or electrons and incoming ions arises. These differences will contribute to the excitation of instabilities. In this section we will restrict to ion instabilities leaving the essentially more interesting ion-electron instabilities for the next section. A list of the most important instabilities in the foot region is given in Table 21.2. It is interesting to note that only a few of these instabilities have ever been identified in actual observations and in the simulations even though theoretically they should be present. This can have several reasons, too small growth rates, too strong Landau damping, for instance, in the presence of hot ions, convective losses or very quick saturation due to heating effects, competition with other waves or wave-wave interaction and so on.

The most relevant instabilities generated in the shock foot have been identified as the whistler instability for nearly perpendicular supercritical but low-Mach number shocks, leading to foot formation but not being decisive for feet, as it has turned out that feet in this Mach number and shock-normal angle ranges are produced by accumulation of gyrating ions at the upward edge of the foot.

Table 21.2: Instabilities in the foot region (Wu et al. 1984)

Instability	Excitation by	Source of free energy	Direction of propagation
Ion-ion streaming instability	Reflected and transmitted ions	Relative streaming between the ion species	$\theta_{kB} = 90^\circ$
Kinetic cross-field streaming instability	Reflected ion	Relative streaming between the reflected ions and the solar wind electrons	$\theta_{kB} \in (0^\circ, 90^\circ)$
Kinetic cross-field streaming instability	Transmitted ion	Relative streaming between the transmitted ions and the electrons	$\theta_{kB} \leq 90^\circ$
Lower-hybrid drift instability	Reflected ions	Relative cross-field drift between the reflected ions and the electrons; density gradient	$\theta_{kB} < 90^\circ$
Lower-hybrid drift instability	Drifting electrons	Relative cross-field drift between the reflected ions and the electrons; density gradient	$\theta_{kB} < 90^\circ$
Ion-acoustic instability	Transmitted ion	Relative streaming between the ion species and the electrons	$\theta_{kB} < 90^\circ$
Electron-cyclotron drift instability	Drifting electrons	Electron drift relative to the solar wind ions	$\theta_{kB} \simeq 90^\circ$
Whistler instability	Electrons	Electron thermal anisotropy $T_{e\perp} > T_{e\parallel}$	$\theta_{kB} \simeq 0^\circ$

It is similar in the ramp region, with the fact that not all of these instabilities and waves are confirmed from neither observation nor simulation.

21.4.6.3 Downstream region

Detached shocks that evolve in front of blunt obstacles like a magnetosphere, in particular, possess extended downstream regions which separate the obstacle from the undisturbed upstream region. These regions are the domains of compressional waves – of the family of magnetosonic waves – with upstream directed phase and group velocities larger than the downstream flow velocity that in a characteristic flow time across the downstream distance from the shock to the

obstacle can make it upstream to reach the shock, which forms the spatial envelope of these waves thereby confining them to downstream of the shock transition.

This is the fluid picture of the evolution of the region downstream of the shock. For a curved collisionless supercritical shock it is very difficult to distinguish between the processes triggered by the quasi-perpendicular and quasi-parallel parts separately. It is not until we discuss more about the quasi-parallel shock are we able to have a comprehensive description of the downstream region.

Collisionless shocks cannot be described solely in the fluid picture, however. Therefore the downstream region has quite a different behaviour from that inferred from the hydrodynamic description. This is particularly true for quasi-parallel shocks, but also for quasi-perpendicular shocks. Downstream of a quasi-perpendicular super-critical shock the conditions are as well very different from those which the fluid picture prescribes which of the plasma properties basically predicts an increase in the fluid pressure anisotropy due to compression of the magnetic field and plasma under conservation of the fluid magnetic moment (the ratio T_{\perp}/B), the deviation of the flow around the obstacle, draping of the magnetic field, and a pile-up region of the magnetic field close to the obstacle where pressure balance requires dilution of the plasma.

The main reasons for a deviation from this average and laminar behaviour of the downstream region are that in the supercritical case the fluid picture does not contain the ion (and electron) reflection processes and their consequences for shock foot formation, ramp physics and formation of the shock overshoot. Since collisional dissipation is out of question, the reflection mechanisms replace the necessary dissipation. It is easy to estimate, for instance for the quasi-perpendicular Earth's bow shock, how much energy must be dissipated in order to adjust for the differences in the flow properties between upstream and downstream. For an upstream solar wind of nominal density $n = 5 \times 10^6 \text{ m}^{-3}$ and velocity $V_1 = 500 \text{ km/s}$ at 1 AU, Mach number $M_A = 5$ and moderate shock strength (compression ratio) of $B_2/B_1 = 3$, the energy density in the flow behind the shock ($M_A = 1$) is just half the energy density in the solar wind. Thus 50% of inflow energy must be dissipated by the shock. Since at most 10% of the ions are reflected (corresponding to 10% of incoming energy only), the remaining 40% are absorbed and dissipated in other ways.

A first simple effect that affects the region behind the shock ramp is that the reflected ions are accelerated by the motional electric field in the shock foot along the shock surface in the direction perpendicular to the upstream magnetic field to energies that are much higher than the kinetic energy of the incoming ions, thereby absorbing a substantial part of inflow energy. Acceleration of the 10% reflected ions by a factor 2 gives already 40% of energy dissipation. Hence this is the main dissipation mechanism. This acceleration affects the nearly periodic reformation of a quasi-perpendicular shock when these ions carry a drift current and gyrate at high speeds and large ion gyroradii in the upstream magnetic field. However, after a few such gyrations they have gained sufficient energy to break through the reflecting shock potential and overcome the shock barrier to enter the downstream flow region. Downstream of the quasi-perpendicular shock they appear as an energetic gyrating ion component with much larger perpendicular than parallel energy.

In fact these gyrating energetic ions have been observed (Sckopke et al. 1983). They are a source of free energy and excite a number of instabilities like electromagnetic ion-cyclotron waves which can grow to large amplitudes downstream of the shock and contribute to the magnetic fluctuation that are observed downstream. Even mirror modes just behind the shock have been reported [Czaykowska et al, 1998]. The rms amplitudes $b_{\text{rms}} = \sqrt{|\mathbf{b}|^2} \sim B_2$ of these waves downstream of the shock are comparable to the average downstream field B_2 . Therefore these waves contribute to the energy balance. Moreover, since they are convected downstream and damp away, they contribute to heating of the downstream plasma and shock energy dissipation.

The larger amplitude waves do not damp but decay into other shorter amplitude waves or by ponderomotive force interaction may stabilise and form large amplitude localised magnetic structures, density pulses or also narrow current sheets. The physics of these structures is still barely understood. Sometimes one speaks about downstream turbulence which, however, is not qualified and also unjustified as turbulence can hardly develop in the downstream region because it has not sufficient time for reaching a well mixed state. This would require a very large extended region and relatively quiet boundary conditions with continuous inflow at an approximately constant energy level. It would also require stationarity of the shock which is definitely not given, as we will see later when discussing the problem of stationarity. Shock reformation is already kind of a rather non-stationary process that makes the development of downstream turbulence rather doubtful.

21.4.7 Electron Dynamics

Too complicated. Takes time to digest!

More important than whistlers have turned out the Buneman and modified two-stream modes, the former dominating shocks at perpendicular angles, the latter growing slowly but dominating at more oblique angles and at later simulation times with the effect of completely restructuring both the shock feet and ramps. Both instabilities generate phase space holes which during reformation survive and are added to the downstream plasma and, in addition, being responsible for low magnetic field values.

21.4.8 Stationarity

21.4.8.1 Formation of ripples

1D theory and 1D simulations implicitly treat the shock as an infinitely extended plane surface. In addition they allow only for instabilities to evolve in the direction of the shock normal at an angle relatively close to 90° such that any waves along the shock surface are completely excluded and waves parallel to the magnetic field have very small wave numbers $k = k_x \cos \theta_{Bn} \ll k_x$ corresponding to very long parallel wavelengths. To be more realistic, 2D PIC simulations have

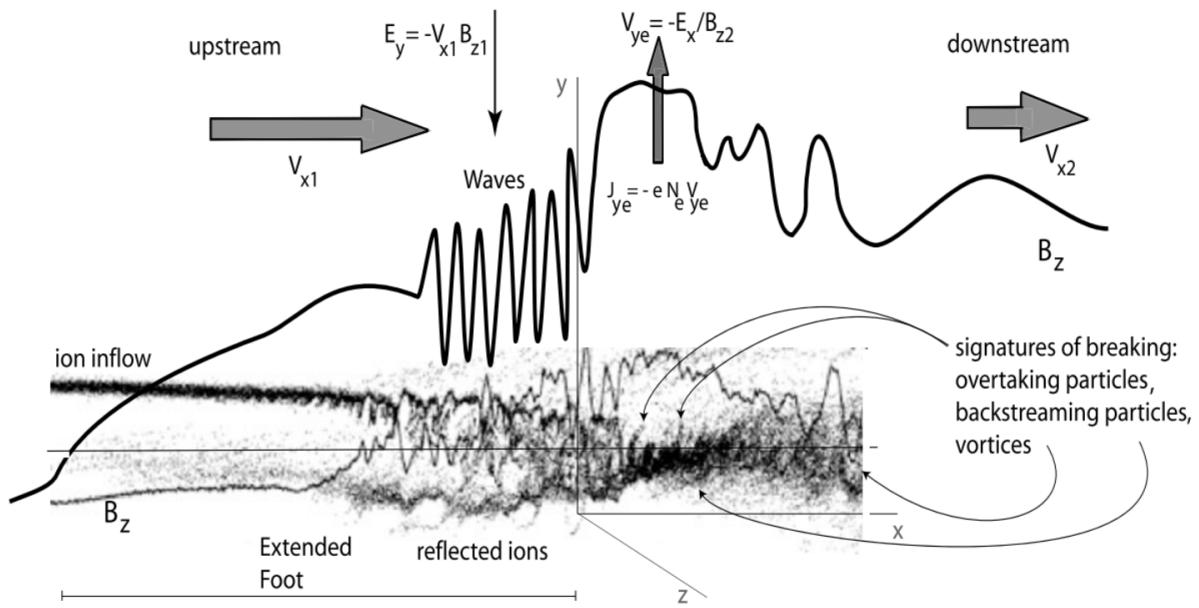


Figure 21.11: Schematic of the profile of a highly supercritical shock with waves just before shock reformation and signatures of beginning wave breaking (Balogh and Treumann 2013). It shows the structure of the ions in the ramp with the signatures of overtaking ions and backstreaming ions as well as ion vortices, all an indication of onset of breaking.

been performed to investigate the effect of the additional freedom given by the second spatial dimension which allows instabilities to evolve in other than the shock normal direction.

3D PIC simulations of quasi-perpendicular shocks at a realistic mass ratio by Shinohara & Fujimoto [2010] found very strong wave activity in the shock foot region which is permitted by the inclusion of the third dimension. These waves cause a stronger than known upstream electron acceleration and the generation of non-thermal electrons.

The lesson learned is, however, quite simple: the shock exhibits structure along its surface which can presumably be attributed to waves running along the shock front and modulating it temporarily and spatially. There are at least three types of ripples: (1) short scale (below ion inertial length) lower-hybrid waves (typically about perpendicular to the magnetic field) supported by cross field currents instabilities [Lembègue and Savoïni, 1992], (2) longer scale (of the order of one or a few ion inertial lengths) oblique whistler waves (Hellinger et al. 2007; Lembègue et al. 2009), and (3) even much longer “Alfvén wave” ripples (several ion inertial lengths) [D. Winske and Quest (1988); Saito and Umeda, 2011].

21.5 Supercritical Parallel Shock

From MHD or double adiabatic theory, parallel shocks are more special in that the magnetic field strength remains unchanged so the equations effectively describe pure gasdynamic solutions. Kuznetsov and Osin (2018) presents a simplified solution in a 1D parallel shock case with parallel and perpendicular thermal energy heat fluxes S_{\parallel} and S_{\perp} included. Note again the original CGL theory assumes 0 heat fluxes.

However, as have been indicated in Section 21.1.7, this does not cover the real physics involved into parallel shocks which must be treated on the basis of kinetic theory. These shocks possess an extended foreshock region with its own extremely interesting dynamics for both types of particles, electrons and ions, reaching from the foreshock boundaries to the deep interior of the foreshock. Based mostly on kinetic simulations, the foreshock is the region where dissipation of flow energy starts well before the flow arrives at the shock. This dissipation is caused by various instabilities excited by the interaction between the flow and the reflected particles that have escaped to upstream from the shock. Interaction between these waves and the reflected and accumulated particle component in the foreshock causes wave growth and steepening, formation of shocklets and pulsations and causes continuous reformation of the quasi-parallel shock that differs completely from quasi-perpendicular shock reformation. It is the main process of maintaining the quasi-parallel shock which by its nature principally turns out to be locally nonstationary and, in addition, on the small scale making the quasi-parallel shock close to becoming quasi-perpendicular for the electrons. This process can be defined as turbulent reformation, with transient phenomena like hot flow anomalies, foreshock bubbles, and the generation of electromagnetic radiation.

The turbulent nature implies that the quasi-parallel shock transition is less sharp than the quasi-perpendicular shock transition and thus less well defined; there exists an extended turbulent foreshock instead of a shock foot. This foreshock consists of an electron and an ion foreshock. The main population is a diffuse ion component. The turbulence in the foreshock is generated by the reflected and accelerated foreshock particle populations. An important point in quasi-parallel shock physics is the *reformation* of the shock which works completely differently from quasi-perpendicular shocks; here it is provided by upstream low-frequency electromagnetic waves excited by the diffuse ion component. Steepening of these waves during shockward propagation and addition of the large amplitude waves at the shock transition reforms the shock front. The old shock front is expelled downstream where it causes downstream turbulence. During the reformation process the shock becomes locally quasi-perpendicular for the ions supporting particle reflection.

21.5.1 Basic Properties

In a quasi-parallel shock the combined geometries of the upstream magnetic field and generally curved shock surface prevent the shock-reflected particles from immediate return to the shock. The reason is that their gyro-orbits, after having suffered reflection from the shock ramp, lie completely upstream, outside the shock ramp, such that they do not touch the shock ramp again after reflection. Since, in addition, their upstream velocities have a large component parallel to the upstream magnetic field, which increases the more the shock-normal turns parallel to the upstream magnetic field, the reflected particles are enabled to escape upstream from the shock along the magnetic field thereby escaping backward convection and forming fast upstream particle beams.

A quasi-parallel supercritical collisionless shock thus populates the upstream space with a reflected particle component. This population moves a long distance away from the shock along the magnetic field. Was the upstream flow, in the case of the quasi-perpendicular shock, completely uninformed about the presence of the shock up to a distance of the mere width of the shock foot so, at the quasi-parallel shock, it receives a first signal of the presence of a shock already at quite a large upstream distance when the first and fastest reflected particles arrive on the magnetic field lines that connect the flow to the shock. It is, in fact, only these particles that can inform the flow about the presence of a supercritical shock, because any low-frequency plasma wave cannot propagate far upstream for supercritical Mach numbers $M > M_c$, while any electromagnetic radiation that is generated at the shock has frequency $\omega_{\text{rad}} > \omega_{\text{pe}}$. The upstream flow recognizes the reflected particles in its own frame of reference as a high-speed magnetic-field aligned beam. Thereby a beam-beam configuration is created which leads to a number of beam-driven instabilities. These excite various plasma waves that fill the space in front of the shock and modify its properties.

Figure 21.12 shows a sketch of the magnetic profile of a supercritical quasi-parallel shock which contrasts the profile of a quasi-perpendicular shock that has been given in Figure 21.5.

The quasi-parallel magnetic shock profile is much stronger distorted than that of a quasi-perpendicular shock, such that it becomes difficult to identify the location of the genuine shock ramp on the profile.

The main structural difference between quasi-perpendicular and quasi-parallel shocks is that quasi-perpendicular shocks possess a narrow $\sim 1r_{ci}$ wide foot region that is tangential to the shock surface, while quasi-parallel shocks possess an extended foreshock region. Interestingly, in curved shocks which arise, for instance, in front of spatially confined obstacles, both phases of a supercritical shock can coexist at the same time, being spatially adjacent to each other. An example is shown in Figure 22.14 in the sketch of the curved Earth's bow shock. Dealing with quasi-parallel shocks means to a large extent dealing with the processes that are going on in the foreshock. It will thus be quite natural to start with a discussion of the properties of the foreshock.

However, before continuing we point out that in spite of the strict distinction between the quasi-perpendicular and quasi-parallel shocks there is also a close relation between the two. Both, being supercritical, can exist only because they reflect ions; and both possess an upstream region in front of the shock transition that is populated by the reflected ions. That this region is narrow in the case of quasi-perpendicular shocks is a question of the ions being tied to the magnetic field, which also holds in the case of the quasi-parallel shock. At quasi-perpendicular shocks the reflected ions do readily return to the shock. At the quasi-parallel shock they ultimately do also return to the shock, but only after having been processed far away from the shock in the foreshock, having coupled to the flow, having passed several stages in this processing, and having become the energetic component of the main flow. As such they finally arrive at the shock together with the stream. In between their main duty was to dissipate the excess energy, which they possessed when arriving for the first time at the shock and which could not be dissipated in the narrow shock-ramp transition region. This could have been achieved only in the broad extended foreshock which, seen from this point of view, is already the shock. Therefore, the main physical difference between the quasi-perpendicular and quasi-parallel shocks is the dissipation processes.

21.5.2 Foreshock

Figure 21.13 shows at one glance the main differences in the (average) bulk plasma parameters between quasi-perpendicular and quasi-parallel shocks as measured with the AMPTE IRM spacecraft at many crossings of the bow shock. The data used in this figure have been stapled, averaged and plotted with respect to the time normalised to the shock ramp crossing. For such a normalisation one uses the shock-normal upstream velocity to recalculate the time. This procedure is not very certain for quasi-parallel shocks since – as we will see later – the shock ramp is ill defined in a quasi-parallel shock. However, for a simple comparison of the main differences this uncertainty is less severe.

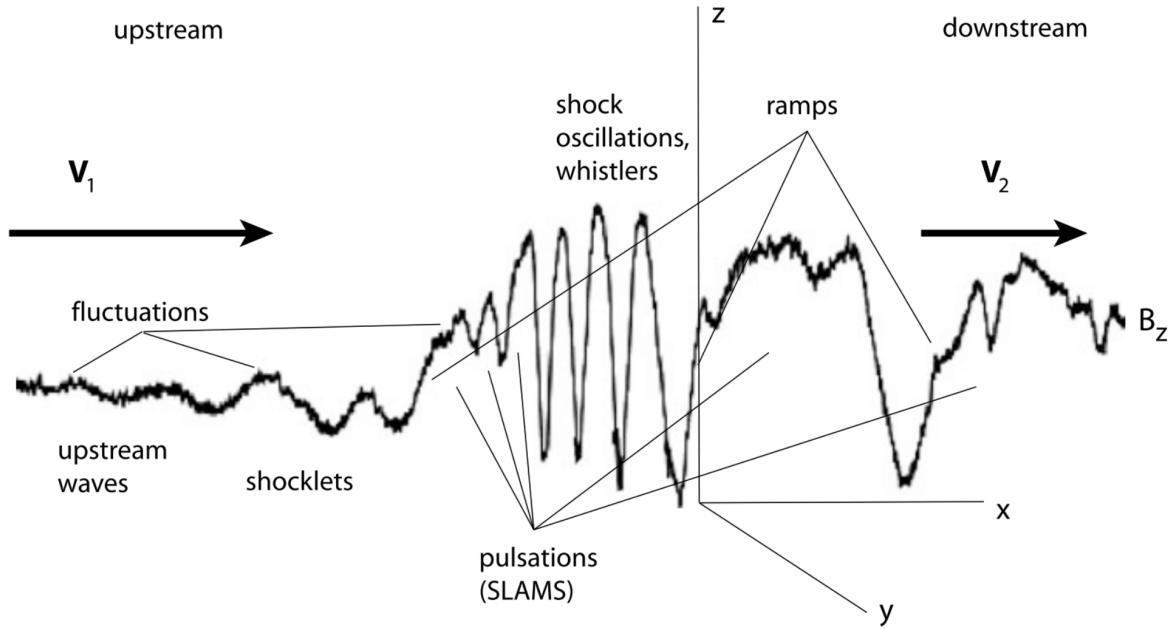


Figure 21.12: Schematic 1D profile taken along the nominal instantaneous shock normal of a supercritical quasi-parallel shock as seen in the magnetic field component B_z . This is the analogue to the quasi-perpendicular shock profile. It shows the main features in the vicinity of the quasi-parallel shock transition: the large amplitude upstream waves with the turbulent fluctuations on top of the waves, the formation of shocklets, i.e. steep flank formation on the waves exhibiting small-scale fluctuations on top of the wave, which act already like small shocks, very-large amplitude pulsations (magnetic pulsations or SLAMS) which turn out to be the building blocks of the shock, multiple shock-ramps at the leading edges of the pulsations belonging to diverse ramp-like steep transitions from upstream to downstream lacking a clear localization of the shock transition and any attached phase-locked whistlers. (Note that the entire figure is, in fact, the shock transition, as on this scale no clear decision can be made where the shock ramp is located.) Not shown here are the out of plane oscillations of the magnetic field that accompany the waves and the particle phase space.

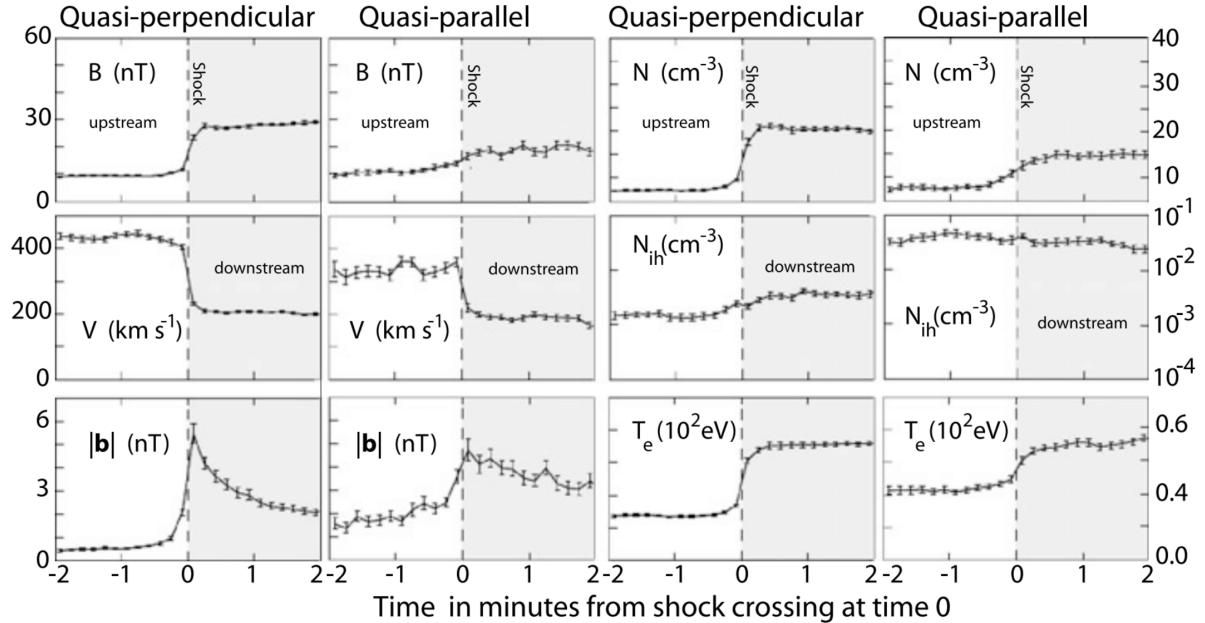


Figure 21.13: Comparison of average bulk plasma parameters in quasi-perpendicular and quasi-parallel shocks [data from Czaykowska et al, 2000]. The figure shows (on the left) the mean magnetic field B , bulk flow velocity V , average magnetic fluctuation amplitude at ultra-low frequencies, and (on the right) plasma density N , density of high energy ions N_{ih} , and electron temperature T_e . The shaded regions are the downstream parts of the AMPTE IRM crossings of the bow shock. The data have been obtained by normalizing the time with respect to crossing the nominal shock ramp by using the measured normal component of the bulk flow velocity, which is not so well-defined in the case of quasi-parallel shocks.

The shaded area in the figure corresponds to the downstream region. Shown are – in pairs of quasi-perpendicular/quasi-parallel values – the magnetic field B , bulk velocity V , average fluctuation amplitude in the ultra-low frequency waves $|\mathbf{b}|$, plasma density N , high-energy ion density N_{ih} of energy >15 keV, and electron temperature T_e . The general conclusion from this figure is that all quantities in the quasi-perpendicular case exhibit a much sharper transition than in the quasi-parallel case. Moreover, the quasi-perpendicular averages are quieter than those of the quasi-parallel case. Also, in general, the quasi-parallel levels are higher than the quasi-perpendicular. In almost all cases the pre-shock levels are enhanced in the quasi-parallel shock case with over the pre-shock levels of quasi-perpendicular shocks. This is seen most impressively in the energetic ion density, which is nearly constant over this distance/time scale at quasi-parallel shocks and much higher than that in quasi-perpendicular shocks, signifying on the one hand the importance of energetic particles in quasi-parallel shock dynamics, on the other hand the capability of quasi-parallel shocks to accelerate particles to substantial energies.

The physics of quasi-parallel shocks cannot be understood without reference to the foreshock. The foreshock is that part of the upstream shock region that is occupied with reflected particles. At a curved shock, like the Earth's bow shock, the foreshock starts on the shock surface at the location where the upstream magnetic field shock-normal angle $\theta_{Bn} \leq 45^\circ$. From that point on electrons and ions escape along the magnetic field in upstream direction. Since electrons generally move at a larger parallel velocity than ions they are less vulnerable to the convective motion of the upstream magnetic field line to which they are tied, and so there is generally a region closer to the foreshock boundary magnetic field line where only upstream electrons are found. This region is confined approximately between the line that marks the electron foreshock boundary and the more inclined line that marks the ion foreshock boundary. An example of this geometry was depicted in Figure 22.14 for the bow shock of the Earth.

In discussing the properties of the foreshock one thus has to distinguish of which foreshock is the topic. However, the properties of the electron foreshock are not as decisive for the formation of a quasi-parallel shock as are the properties of the ion foreshock. Most of the time, we refer to the ion foreshock as the foreshock.

The ion foreshock is not a homogeneous and uniform region. The reflected ion component evolves across the ion foreshock from the ion foreshock boundary to the center of the ion foreshock and from there towards the shock. Speaking of a reflected ion component that can unambiguously be identified as being reflected, i.e. streaming into the upward direction, makes sense only in the immediate vicinity of the ion foreshock boundary. Here the reflected ions appear as a fast ion beam the source of which can be traced back to the shock. Deeper in the foreshock the beam component cannot be identified anymore.

21.5.2.1 The Ion Foreshock Boundary Region

First identifications of reflected beam protons in space in the magnetic flux tube connected to the Earth's bow shock wave were reported by Gosling et al [1978] and Paschmann et al [1981] who distinguished those beams by their poorly resolved distribution functions from more diffuse protons deeper in the foreshock. Interestingly, observations in the foreshock of interplanetary travelling shocks did not show any indication of such beams but only the diffuse ion component. Figure 21.14 gives an observational example of such a reflected ion beam that propagates very close to the foreshock boundary upstream away from the shock. The bulk flow is the narrow cold beam in the left part of the figure which is displaced in negative v_x -direction. The reflected beam is less dense (lower count rates) but much more energetic. It is displaced in $+v_x$ -direction, i.e. streaming away from the shock, and has also a $-v_y$ -component, i.e. it constitutes a gyrating bunch of ions moving away from the shock. In the right part of the figure it is seen that the beam is moving away along the magnetic field line that is connected to the shock, while the bulk of the plasma flows in positive direction.

These beams along the foreshock boundary play some role in the foreshock dynamics as they seem to represent a source population for the entire ion foreshock. Whether and why this is really so is not yet fully understood as the shock should reflect ions at almost every place in its quasi-parallel state. However, it seems that only the group of ions that escape from the shock along the foreshock boundary can form such beams. This points on a further interesting relation between quasi-perpendicular and quasi-parallel shocks at a curved shock surface with a smooth transition from quasi-perpendicular to quasi-parallel as sketched in Figure 22.14 and realised in space, for instance, at planetary bow-shocks. It seems as so these beams escape from the quasi-perpendicular region of the shock along the nearly tangential field lines. This would also be in agreement with the observation [Gosling et al, 1984] that the foreshocks of extended interplanetary shocks do not show any signs of reflected ion beams. They are only very weakly curved being nearly planar, and do not possess a recognisable quasi-perpendicular area on the surface.

Kucharek et al. (2004) analysed ion distributions along magnetic field lines that were connected to the quasi-perpendicular area of Earth's bow shock. Investigating the origin of those beams they found that, indeed, the observed ion beams at the foreshock boundary result from reflection at the quasi-perpendicular shock and that, without the presence of a quasi-perpendicular region, there would presumably be no distinct foreshock boundary and no ion beams escaping into the foreshock. If this is the case, then the population in the quasi-parallel ion foreshock is indeed provided by two different ion sources, the beams from the quasi-perpendicular shock region and the genuine foreshock ion population. The latter has no beam character but is rather a diffuse ion component [see also Meziane et al, 2004]. Interplanetary shocks possess only the latter one.

Roughly 2% of the ion inflow leaves the shock ramp upstream in the form of a beam along the

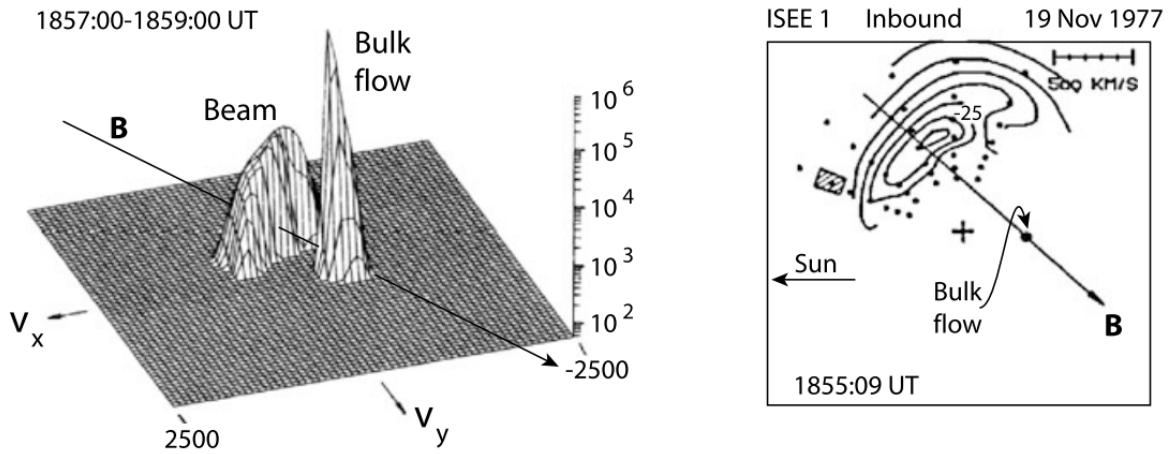


Figure 21.14: ISEE 1 observation of a reflected ion beam on November 19, 1977 propagating along a magnetic field line that was connected to the quasi-parallel Earth's bow shock. Left: A two-minute average pseudo-3D ion velocity-space profile in the (v_x, v_y) -plane showing the undisturbed and cold (narrow) plasma inflow in negative v_x -direction, and the fast and warm (broad) beam of reflected ions propagating in positive v_x -direction and spreading in v_y . This beam is quite anisotropic in temperature. Velocities are in km s^{-1} . The scale on the right is count rates, and background count rates were suppressed by choosing only values above 50 s^{-1} . Right: Contour plot of a similar beam a little earlier showing that the beam is centred on the magnetic field that connects to the shock, is quite narrow along the field and about 2–3 times as broad perpendicular to the field [data from Paschmann et al, 1981]. The cross indicates the origin (zero velocity), the dot the bulk flow centre. The $10^{-25} \text{ s}^3 \text{ cm}^{-6}$ level flux contour has been marked.

magnetic field.¹² Kucharek et al. (2004) argue that the ions which escape along the magnetic field, are reflected from the ramp/overshoot region where they have been in resonance with low-frequency plasma waves, which they assume to be large amplitude Alfvén-whistler waves. These ions experience pitch-angle scattering and pitch-angle diffusion towards small pitch angles, and subsequently can escape along the magnetic field in the upstream direction. Since the conditions for escape depend in the first place on the pitch-angle scattering process, the beams should be highly variable in time and location. The mechanism might still sound a bit speculative as long as no simulation proves its reality, but any mechanism which is able to pitch-angle scatter ions along the magnetic field in a quasi-perpendicular shock will naturally cause ion beams to escape from the ramp both in the upstream and in the downstream directions. Such simulations require a three-dimensional treatment which is not in reach yet.¹³

A measured example of the ion foreshock-boundary field-aligned ion beam distribution (Kis et al. 2004) is shown in Figure 21.15 for an upstream flow velocity $V_1 \simeq 660 \text{ km s}^{-1}$, Mach number $M_A = 8$, and average shock-normal angle $\theta_{Bn} \simeq 15^\circ$, at an upstream distance from the Earth's bow shock. This distribution is a so-called reduced distribution; it is the integrated over pitch-angle ϕ and perpendicular velocity v_\perp magnetic field-aligned phase space distribution function $f(v_\parallel) \sim 2\pi \int v_\perp^2 dv_\perp f(v_\perp, v_\parallel)$ that has been appropriately binned and smoothed. The information to be taken out of this figure is that the reduced ion-beam distribution (solid line) is narrow in velocity, maximising at a speed (in absolute terms $|V_{b\parallel}| \simeq 800 \text{ km/s}$) that is only slightly larger than the parallel flow velocity ($V_{1\parallel} \simeq 640 \text{ km/s}$, when taking into account the shock-normal angle), i.e. $|V_{b\parallel}| \simeq 1.25 V_{1\parallel}$. It is directed opposite to the flow. For comparison, a reduced parallel diffuse ion distribution is shown in the same plot taken deeper in the foreshock. This distribution is about symmetric to the magnetic field direction, indicating the about circular phase-space distribution of the diffuse ion component which appears as an energetic tail on the main ion distribution (note that the small gap on the left of the dashed diffuse-ion distribution curve is questionable as it dips below the 1-count level).

Meziane et al. (2004) reported CLUSTER observations of foreshock-boundary ion beams simultaneously with diffuse ions. They found that the nominal ion-beam velocity $V_b \simeq 1.7 V_1$ had no relation to any known shock-reflection mechanism like specular reflection, a conclusion which supports pitch-angle scattering as the beam-injection process as this is independent of the speed of the inflow.¹⁴

Kis et al. (2004) have carefully analysed the relation between the observed ion distribution at the upstream spacecraft position and the calculated shock-normal angle θ_{Bn} , determined from the local upstream magnetic field direction and the predicted shape of the bow shock. The result is shown in Figure 21.16 for three successive times on February 18, 2003 when the CLUSTER spacecraft was outside the bow shock. At 1130 UT and 1355 UT the shock

¹²From our 1D simulations, the reflecting ratio is about 1%. I think it would be interesting to investigate why. The energy perspective is a good starting point.

¹³Our chance may have arrived!

¹⁴I have strong doubts here. According to my understanding, the authors support specular reflection in this case study!

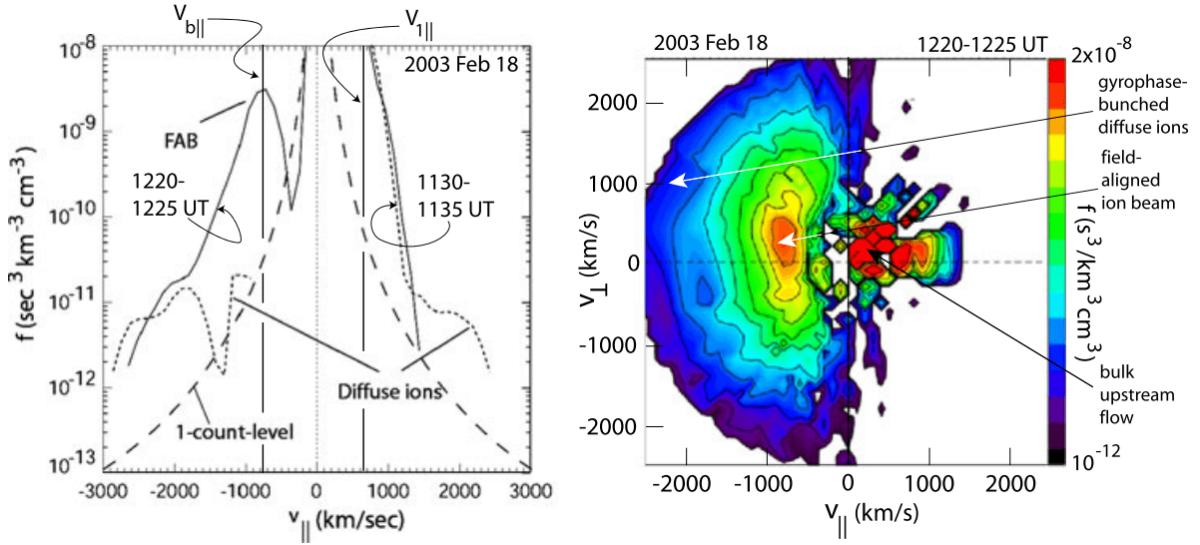


Figure 21.15: Left: The reduced field-aligned ion distribution functions in the foreshock region showing the field-aligned reflected ion beam (FAB) at 1220–1225 UT on February 18, 2003 which arrives at the spacecraft location along the magnetic field line that is connected to the quasi-perpendicular region of the Mach number $M_A \sim 8$ supercritical shock (smooth solid line). The upstream bulk velocity was $V_1 \approx 660 \text{ km s}^{-1}$. Also shown is the reduced parallel distribution function (dashed) for the diffuse ion distribution observed at 1130–1135 UT the same day (for the geometry see Figure 6.7). It is clearly seen that the latter is about symmetric to the magnetic field direction indicating the symmetry of the ring of diffuse ions, while the foreshock-boundary field-aligned beam is flowing in the direction away from the shock into the upstream medium (negative velocities). Note also that the diffuse distribution appears as a smooth tail on the full ion distribution. The small gap on the left is uncertain as it dips into the 1-count level. Right: The two-dimensional phase space plot for the time interval 1220–1225 UT when the field-aligned beam was observed. Indicated are the bulk upstream flow, field-aligned ion beam, and the gyro-phase bunched residue of the diffuse upstream ions (data from Kis et al. 2004).

normal angles $\theta_{Bn} \sim 15^\circ$ and shock-spacecraft distances ~ 7.5 respectively $\sim 6.7 R_E$ were similar. Despite this similarity the observed ion phase-space distributions were completely different. At 1130 UT no low-energy gyrating ions were observed, while they were present at the later time 1355 UT. Hence, at 1130 UT the spacecraft must have been closer to the ion foreshock boundary, i.e. the foreshock boundary was more inclined than at the later time such that beam particles scattered from the foreshock boundary have not arrived at the location of the spacecraft. Due to the velocity filter effect they have been separated out at spacecraft distance. The reconstruction of the position of the ion-foreshock boundary using the measured upstream conditions and shape of the shock for this period is shown in Figure 21.16. Indeed, at 1220 UT the spacecraft was close to the foreshock boundary and, as expected, detected the ion beam (as seen in Figure 21.15).

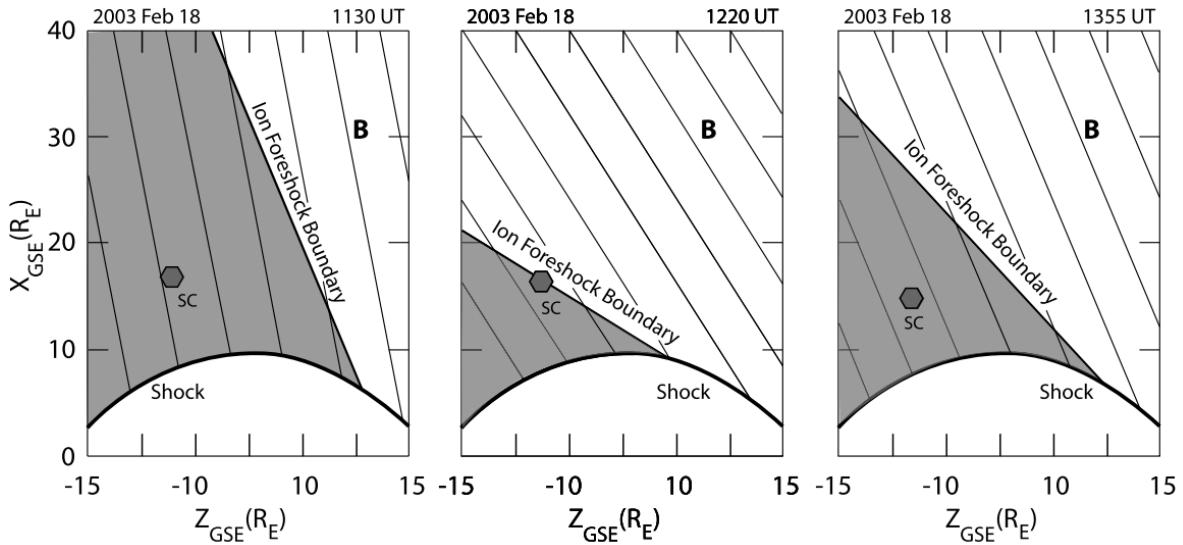


Figure 21.16: Reconstruction of the shape and location of the ion-foreshock boundary from the measured upstream plasma properties (direction of magnetic field, speed and density) for three successive times in order to explain the ion-phase space observations in Figure 21.15. The cold magnetic field-aligned ion beam is observed when the spacecraft is located in the vicinity of the ion-foreshock boundary at 1220 UT, while at 1130 UT and 1355 UT no beam was detected. It is assumed that the beam is generated along the foreshock boundary at the position where at the quasi-perpendicular shock surface the shock-normal angle is roughly about $\theta_{Bn} \sim 60^\circ$.

This observation supports the above advocated view that the ion-foreshock beam is generated in the transition region from quasi-perpendicular to quasi-parallel shock. The remaining questions to be answered are:

- What pitch-angle scattering mechanism is responsible for the generation of such a beam?

- What is the fate of the foreshock-boundary beam-ions during their propagation along the foreshock boundary?
- Do they contribute to the foreshock ion population and if, in what way?

Currently we are not able to answer either of these questions definitely. In particular, the pitch-angle scattering mechanism is unknown or at least uncertain. When discussing wave generation, we will touch on the problem of the fate of the beam. Below we present evidence for the scattering of the beam ions and merging into the upstream foreshock diffuse-ion population.

21.5.2.2 Diffuse Ions

Reflected upstream magnetic field-aligned ion beams are observed at the foreshock boundary only. The second (and main) ion component encountered in the foreshock is the diffuse ion population which is detected there as the energetic extension of the inflow plasma. It is widely assumed that the origin of this component is also at the shock as there is no other energetic particle source available. However, close investigation of the diffuse ion component in the foreshock of the Earth's bow shock wave has demonstrated that these ions are not produced in a specular reflection process at the shock. Rather their origin is of diffusive nature.

So far it has not been possible to identify the source of these ions though some models have been proposed which we will briefly discuss below in relation to the appearance of foreshock waves. These particles are also linked to the shock particle accelerations. Here, we merely discuss some of their properties and provide evidence that they are indeed coming from the shock, proving that the shock is an energetic ion source.

The foreshock-boundary ion beams also merge into the foreshock particle distribution by scattering from the foreshock boundary off their self-generated wave spectrum and subsequently being convected downstream by the bulk flow. Paschmann et al [1981] have mapped this merging process by following the evolution of the foreshock-boundary beam distribution shown in Figure 21.14 during the convection. In that figure the beam was detected close to the foreshock boundary. Figure 21.17 show its form deeper in the foreshock when it has spread substantially in angle evolving into half of a ring distribution already. Even deeper inside the foreshock the reflected ion distribution assumes the shape of a full ring around the bulk distribution as is shown in Figure 21.18.

The three observations depicted in Figures 6.4, 6.8 and 6.9 are from different times; it has, however, been checked that they are at distances corresponding to increasing distance from the foreshock boundary such that the assumption of the convectively processed beam evolution is well founded (or at least reasonable) even though it has not been directly proven.¹⁵ It is interesting to note that a gap remains between the original ion-phase space beam distribution and the bulk-flow distribution, which is another indication that the evolving distribution is part of the evolution of the ion-foreshock beam. Of a distribution that does not evolve out

¹⁵It has been demonstrated by Yann using the hybrid Vlasov model Vlasiator.

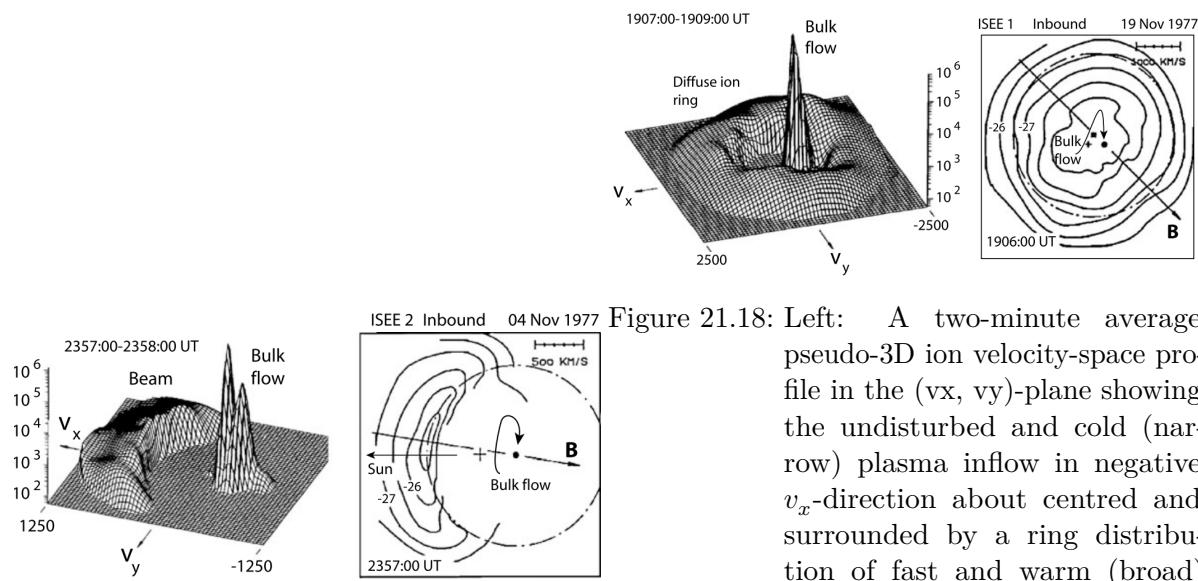


Figure 21.17: Left: A one-minute average pseudo-3D ion velocity-space profile in the (v_x, v_y) -plane showing the spreading of the beam in angle around the bulk flow without merging into the bulk flow. Right: Contour plot of the partial ring distribution.

The evolution of the ion foreshock-boundary ion beam during its convection downstream into the foreshock and propagating deep inside the foreshock away from the ion foreshock boundary as has been seen by ISEE 2 and 1 on 04 November 1977. Velocities are in km s^{-1} . The scale on the right is count rates, and background count rates were suppressed by choosing only values above 50 s^{-1} . The direction of the magnetic field is also shown. The 10^{-26} and $10^{-27} \text{ s}^3 \text{ cm}^{-6}$ level flux contours have been marked. [data from Paschmann et al, 1981]

Figure 21.18: Left: A two-minute average pseudo-3D ion velocity-space profile in the (v_x, v_y) -plane showing the undisturbed and cold (narrow) plasma inflow in negative v_x -direction about centred and surrounded by a ring distribution of fast and warm (broad) diffuse ions that have been reflected from the shock but have been processed in the foreshock region when propagating from the foreshock boundary into the foreshock. These ions are hollow in the sense that they separate from the bulk distribution but have a nearly isotropic distribution function. Right: Contour plot of a similar ring a little earlier showing that the ring centre (star) is slightly displaced from the bulk flow (dot) on the magnetic field that connects to the shock and from the centre (cross) of the phase space frame.

of a beam one expects a less regular behaviour and, generally, no such well expressed gap between bulk flow and beam in velocity space. In fact, the main energetic ion component in the foreshock is irregular and lacks a well expressd gap.

The discrepancy between the smooth no-gap foreshock distributions [Sentman et al, 1981a] and the gap-observations [Paschmann et al 1981] has, in fact, been noted much earlier [Sanderson et al, 1981; Scholer, 1985] without giving an explanation but suggesting a continuous ion source at the parallel shock. Below we provide further arguments for the two sources, foreshock-boundary beam and continuous extended shock-surface source hypothesis. Nonetheless, this conclusion must be taken with care because Figure 21.18 shows the gap progressively closing. Being sufficiently far, i.e. even farther away from the ion foreshock boundary then in this figure, it will not anymore be possible to distinguish between beam-evolved and genuine diffusive-ion distributions. Ultimately, both distributions will have merged indistinguishably. Still there is no agreement whether and where, i.e. at what distance from the ion foreshock boundary this merging of the two populations occurs.¹⁶

Trattner et al [1994], using AMPTE IRM measurements of diffuse ion densities upstream of the quasi-parallel bow shock found that the diffuse ion density *decreases exponentially* with shock distance. This investigation was substantially improved by Kis et al. (2004) in an attempt to infer about the source of the upstream diffuse energetic ions using the four CLUSTER spacecrafs. These authors determined the parallel-density gradient of diffuse ions in the energy range from 10 to 32 keV as a function of distance from the bow shock, as shown in Figure 21.19. We learned from this figure that in the deep foreshock, i.e. at distances far away from the ion-foreshock boundary, the diffuse ion component is densest close to the shock with density decaying exponentially with increasing upstream distance from the shock along the magnetic field. This spatial decay of the diffuse partial ion density $N_i(\mathcal{E}, z) \sim \exp[-z/L(\mathcal{E})]$ is different for particles of different energy \mathcal{E} . The e-folding distance $L(\mathcal{E}) \sim \mathcal{E}$ turns out to increase linearly with energy, i.e. low energy particles are confined to the shock. The higher the ion energy the deeper can the ions penetrate into the upstream plasma. The proportionality constant determined from these data under the special conditions of the Earth's bow shock is $\sim 0.14R_E/\text{keV}$. This behavior of the energetic foreshock ions provides indisputable evidence for the extended parallel shock-surface origin of the diffuse ion component. The source of the diffuse ions lies at the quasi-parallel shock. In order to be found at a distance upstream of the shock the ions undergo a diffusion process along the magnetic field. These ions are thus completely different from the beam ions found at the ion-foreshock boundary.

The e-folding distance for the diffuse ions is given by $L(\mathcal{E}) = \kappa_{\parallel}(\mathcal{E})/V_1$, with spatial diffusion coefficient $\kappa(\mathcal{E}) = \frac{1}{3}v l_{\parallel}(\mathcal{E})$, where l_{\parallel} is the diffusion length (parallel ion mean free path) and v the particle velocity¹⁷. From balance between convective inflow and diffusion into upstream direction, one can write $l_{\parallel}(\mathcal{E}) = 3L(\mathcal{E})\sqrt{\mathcal{E}_1/\mathcal{E}} \sim \sqrt{\mathcal{E}/\mathcal{E}_1}$, where \mathcal{E} is the particle energy and \mathcal{E}_1 is the upstream flow energy. In the solar wind the flow energy is a few keV, and a 20-keV diffuse ion will have a typical parallel diffusion length (or mean free path) of $l_{\parallel} \sim (1-2)R_E$.

¹⁶This is something we can check with new simulations!

¹⁷The diffusion coefficient has the correct dimension m^2s^{-1} .

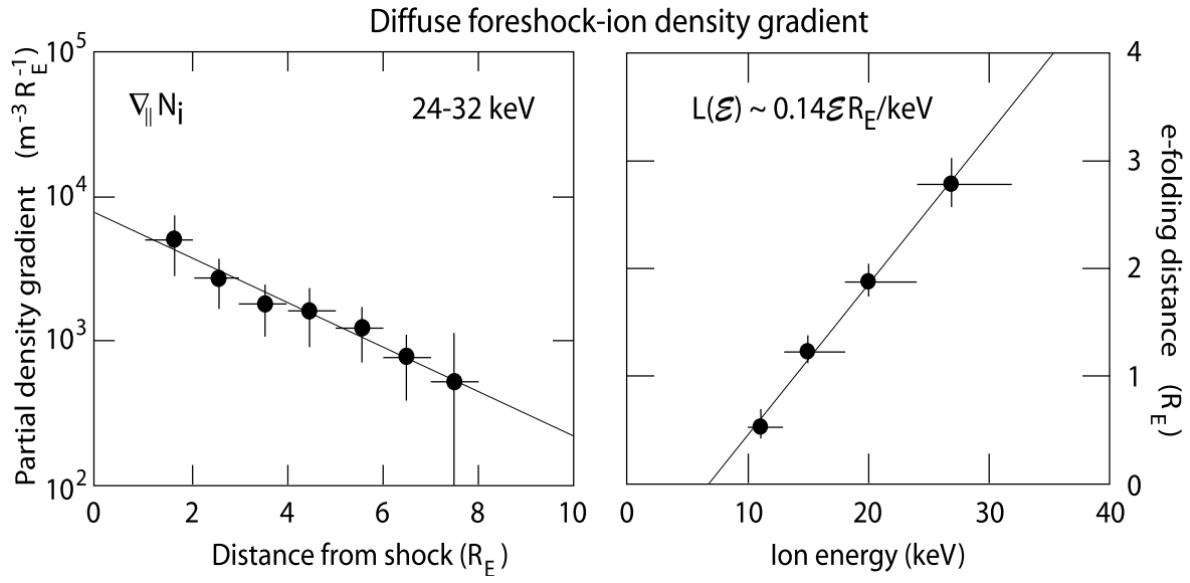


Figure 21.19: The patchwork model of J. Schwartz Steven and Burgess (1991) of a quasi-parallel supercritical shock reformation. Left: Magnetic pulsations (SLAMS) grow in the ion foreshock and are convected toward the shock where they accumulate, thereby causing formation of an irregular shock structure. Note also the slight turning of the magnetic field into a direction that is more perpendicular to the shock surface with the shock surface itself becoming very irregular. Right: The same model with the pulsations being generated in the relatively broad ULF-wave-unstable region in greater proximity to the ion-foreshock boundary. When the ULF waves evolve to large amplitude and form localised structures these are convected toward the shock, grow, steepen, overlap, accumulate and lead to the build up of the irregular quasi-parallel shock structure which overlaps into the downstream direction.

This is a rather short distance, orders of magnitude shorter than the collisional mean free path of an ion. Hence, the diffusion estimate suggests that strong wave-particle interactions can be held responsible for the scattering and acceleration of the diffuse particle component, which enables it to diffuse and escape upstream from the shock and populate the foreshock. The diffusion process is energy dependent with the most energetic ions diffusing fastest.

These interactions should take place in the quasi-parallel shock transition because, as we have shown above, the diffuse upstream-ion density maximises closest to the shock. It is interesting to estimate the corresponding upstream-ion collision frequency $\nu_{c,ui} \simeq \nu/l_{\parallel}$. For the 20 keV-upstream ions this yields $\nu_{c,ui} \sim 0.2\text{Hz}$. This value is comparable to the ion cyclotron frequency $\omega_{ci}/2\pi = (0.1 - 0.3)\text{ Hz}$ in the $B \simeq 8\text{nT}$ ramp magnetic field (Kis et al. 2004) during the time of observation.

It seems that waves, electromagnetic and/or electrostatic, related to the ion-cyclotron frequency are involved into the process of upstream ion diffusion. Since this diffusion is energy dependent, this process is not a simple pitch-angle diffusion as in the case of the generation of the ion beam that propagates along the foreshock boundary. The diffuse ion component must have experienced a substantial acceleration in this process, and this acceleration is located at or around the shock transition and contrasts with the ion-beam acceleration which is a scattering process followed up by pick-up acceleration when the upstream-propagating beam ions are subject to the effect of the main-bulk-stream convection-electric field in which they become accelerated in the direction perpendicular to the magnetic field to roughly four times the energy of the bulk flow, thereby evolving into the ring distribution that characterises their phase space distribution.

21.5.3 Low-Frequency Upstream Waves

In the frame of the upstream bulk flow the two ion components that populate the ion foreshock carry a substantial amount of free energy which is subject to dissipation. Since this dissipation is collisionless it can proceed only through the excitation of waves and wave turbulence through instability upstream of the quasi-parallel supercritical shock. On the other hand it is obvious that the presence of neither of the components can be understood without complete knowledge of the waves in the foreshock and their interaction with the particles.

Since their first detection by Olson et al [1969], Russell et al [1971], Fairfield [1974] and Fairfield & Feldman [1975], observation of shock-upstream waves has been a long-standing issue. Their existence was predicted by Tidman & Northrop [1968], followed by MHD [Barnes, 1970] and kinetic [Hasegawa, 1972] theories of electromagnetic wave excitation and propagation upstream of a collisionless quasi-parallel supercritical shock. Wu [1972] suggested that they might develop into discrete wave-packets as had been inferred from observation by Russell et al [1971]. The ISEE 1-3 spacecraft allowed for a more elaborate investigation of the properties of upstream waves. More recently, CLUSTER measurements have been used to investigate the temporal and spatial structure of upstream waves and wave turbulence. Wave generation is coupled

to particle-energy loss and to particle scattering both playing a substantial role in particle acceleration. We start with a review of the properties of upstream waves in the ion-beam and diffuse ion region from observations and discuss mechanisms of their generation.

[Burgess 1997] reviewed the observed wave organizing by frequencies. He distinguished between Low Frequency Waves (5 mHz- few 100 Hz) and High Frequency Waves (>1 kHz), the latter covering the electrostatic waves from ion-sound to electron plasma waves, as well as radiation. Radiation generation will be discussed in detail later. Here, we only note the almost continuous presence of waves in the ion-acoustic band which have been known since Rodriguez & Gurnett [1975] to populate the complete foreshock region. These spectra might be composed of several different modes, ion-sound, electron-acoustic, Buneman modes, electron-cyclotron harmonics, and others. Their generation mechanism is not clear yet. They might, via a number of different instabilities, be the result of the presence of the hot foreshock-electron component, which also invades the ion foreshock, or they are excited by unresolved narrow electron beamlets that emanate from the quasi-parallel shock. They might also be excited by plasma inhomogeneities, spatial inhomogeneities in the electron distribution, or they are the result of nonlinear wave-wave interaction which is expected to take place in the foreshock. Little we have known about these high frequency waves since Burgess' remarks.¹⁸

However, there has been substantial progress in the understanding of the low frequency waves and their role in quasi-parallel shock dynamics. The waves that are most important in shock formation propagate in the ultra-low frequency range <0.1 Hz. Usually they have large (magnetic) wave amplitudes, around $|\mathbf{b}|/B \sim 0.2 - 1.0$, which identifies them as highly nonlinear. These large amplitude waves had already been observed by Russell et al [1971] to have wave forms from monochromatic to solitary waves, frequently with steep edges resembling shocklets and suggesting that the waves have experienced nonlinear steeping during their evolution and propagation. They sometimes show the typical signs of fluctuations that are connected to these edges and obviously propagate in the whistler mode. Thus these forms are indeed little shock-like structures.

In the same frequency window, large-amplitude pulsations have been identified. These are very typical for quasi-parallel shocks. Schwartz & Burgess [1991] and Schwartz et al [1992], identifying them in the AMPTE magnetic field measurements, coined the name SLAMS for them, which stands for *Short, Large Amplitude Magnetic Structures*. They are also known as upstream pulsations. Their duration is 10–20 s; they have very large amplitudes $|\mathbf{b}|/B \sim 5$, indeed, but appear as a more coherent structure that is embedded in the ultra-low frequency wave turbulence. Like the ultra-low frequency waves, they propagate in upstream direction in the plasma rest frame while being swept toward the shock by the convective flow. Their polarization is mixed with – possibly – left-hand polarisation (in the plasma rest frame) slightly dominating, suggesting their ultra-low-frequency (ULF) wave origin. Sometimes the polarisation is different on both sides of the upstream pulsation, indicating that they have evolved by some process which produces both kinds of polarisation, which is similar to a solitary wave.

¹⁸From the simulation perspective, it is also difficult to capture these high frequency waves. The best we can get for now is about up to 10 Hz.

The propagation velocity of SLAMS exhibits an interesting amplitude dependence. The upstream directed pulsation speed increases with amplitude, which also is a solitary wave-like property. Moreover, they grow when approaching the shock and entering the increasing density gradient of diffuse ions, and they play an important role in shock reformation. Schwartz et al [1992] suspected that these pulsations are the building blocks of quasi-parallel shocks.

Another interesting property is that SLAMS contain thermal plasma with properties of the upstream flow, while being surrounded by the hot foreshock plasma. One would therefore believe that their source region is located at the ion-foreshock boundary. Being created there by an ion-ion beam plasma interaction they might grow nonlinearly until reaching quasi-equilibrium like solitary structures, having captured the upstream plasma, and afterwards being convected toward the shock into the heart of the ion foreshock.

21.5.3.1 Ion-Beam Waves

Each of the two different upstream-ion populations is responsible for the excitation of its own instabilities. In this section we deal only with those waves which are excited by the foreshock-boundary ion beam. Figure 21.15 suggests that the situation is that of an ion-ion beam interaction (for the moment neglecting the gyrophase-bunched diffuse-ion component). In the frame of the upstream flow the reflected foreshock-boundary ion beam propagates upstream along the magnetic field at parallel speed $v_{b\parallel} \simeq -(2-3)V_{1\parallel}$, where $V_{1\parallel} \simeq 600 \text{ km s}^{-1}$. The upstream ion temperature (in energy units) was $T_i \simeq (1-2) \text{ keV}$, yielding a thermal ion velocity of $V_{i,\text{th}} \simeq 200 \text{ km s}^{-1}$. The beam can be taken as warm with thermal speed $v_{b,\text{th}} \leq v_{b\parallel}$. Hence, $v_{b\parallel} > V_{i,\text{th}}$ with magnetized background and beam ions. On the other hand, the electrons are hot with $T_e \simeq 100 \text{ eV}$.¹⁹

Expected Wave Modes Theoretically, we expect that the beam can excite long-wavelength negative-helicity Alfvén waves via the firehose instability when $p_\parallel > p_\perp$. In addition, because the ion beam is mildly warm, it can excite the resonant left-hand ion-ion beam instability, which is possible for $v_{b\parallel} > V_A \sim 100 \text{ km s}^{-1}$. When the beam thermal speed can be considered to be small, it may excite right-hand resonant ion-ion beam modes. Both modes propagate upstream along the magnetic field with the beam on the background of the upstream flow. They are not as fast as the beam, however, and are thus subject to downstream convection with the flow towards the shock. The firehose mode, at the contrary, moves against the beam and thus by itself moves downstream in the direction of the shock position when excited. A link to the plasma dispersion relations is discussed in Section 10.7.5.

In all three cases the foreshock-boundary beam will lead to the excitation of low-frequency Alfvén and ion cyclotron waves which in the shock frame approach the shock while having their source on the foreshock-boundary field line. During this shock-directed convection and/or propagation they populate the foreshock region with low frequency electromagnetic fluctuations, which might further interact with the diffuse foreshock-ion component. On the other hand,

¹⁹Why are the electrons hot when T_e is much smaller than ion? Is the comparison based on speed?

since the phase and group velocities $\omega/k_{\parallel} \sim (\partial\omega/\partial k_{\parallel}) \sim V_A \ll v_{b\parallel}$ of these waves are of the order of the Alfvén velocity and are, thus, much less than the beam and flow velocities, downstream convection will quickly remove them from the foreshock-boundary source region. Hence, their further evolution in the foreshock is determined by the competition between nonlinear wave steeping and – most important as will become clear below – interaction with the diffuse foreshock-ion component.

Including the electrons (while so far neglecting the electron foreshock component) leaves us with an ion-acoustic unstable phase-space configuration. In the upstream ion-plasma frame a relatively dense foreshock-boundary ion-beam is propagating upstream on a cold ion-hot electron plasma. In order to keep the plasma current-free the electron component is slightly retarded, creating conditions under which ion-acoustic waves can be excited. On the other hand, the configuration is not able to excite neither the Buneman-two stream instability the modified-two stream instability, because the beam propagates solely parallel to the magnetic field. This is shown in Figure 21.20. The instability is excited by the velocity difference between the ion foreshock-boundary ion beam $F_{ib}(v_{\parallel})$ and the slightly shifted to the left hot ($T_e \sim 100$ eV) electron distribution because the velocity difference ΔV between ion beam and electron component $v_{e,\text{th}} > \Delta V \sim 10^3 \text{ km s}^{-1} > c_{ia} \sim 100 \text{ kms}^{-1}$ exceeds the ion acoustic speed while being less than the electron thermal velocity $v_{e,\text{th}}$. These ion acoustic waves occupy a relatively broad spectrum with downstream parallel phase velocities $< v_{ib}$ and become swept towards the shock. Since they quickly leave the foreshock boundary and since the beam is not hot, ion-Landau damping in the source region plays no role. However, when entering the diffuse-ion foreshock region these waves encounter the hot diffuse ion component, will interact with it, and will thereby experience (positive and negative) Landau damping.

Observations Observationally, it is difficult to distinguish between foreshock-boundary waves and diffuse-ion generated waves. Two types of waves that can be related to the presence of the upstream beams have been reported in the vicinity of the bow shock. Fairfield [1974] and Fairfield & Feldman [1975] found monochromatic large-amplitude discrete wave packets in the frequency range $\omega/2\pi \sim 0.4$ Hz which Hoppe & Russell [1980] showed to be right-hand polarized whistlers at $\omega \simeq 10\omega_{ci}$ propagating on the plasma rest frame and being unable to escape far upstream. The other class of waves is of smaller amplitude $|b|/B \sim 0.1$ and frequency of the order of $\omega/2\pi \sim 1$ Hz. These waves form trains which are directly tied to the upstream foreshock-boundary ion beams [Hoppe et al, 1982]. In fact, as Hoppe et al [1982] demonstrated they occur only in the presence of the observed reflected 2–5 keV foreshock-boundary ion beams [Eastman et al, 1981] which Kis et al. (2007) have shown to evolve from a beam into a gyrating particle component that in the deep foreshock is superimposed on the diffuse ion component before it merges into it. However, as Feldman et al [1983] have shown from measuring the electron distribution function, it is not the ion beam who excites these whistlers. The probability that they are driven by the particular electron distribution in the foreshock is rather higher.

The waves observed propagated obliquely ($< 60^\circ$) with respect to the magnetic field, at an average angle of $\sim 45^\circ$. The observed spectrum and the dispersion relation determined from

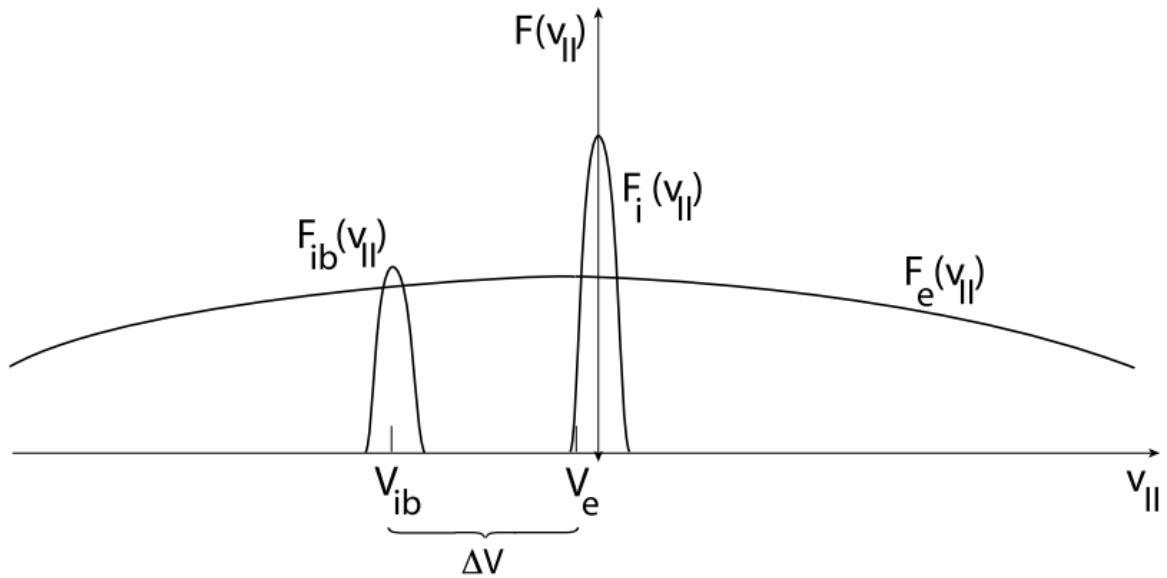


Figure 21.20: Reduced parallel distribution functions along the ion-foreshock boundary showing the cold main flow ion distribution $F_i(v_{\parallel})$, the hot main flow electron distribution $F_e(v_{\parallel})$ shifted to the left into the direction of the foreshock ion beam for keeping zero-current conditions, and the ion-foreshock ion beam distribution $F_{ib}(v_{\parallel})$. This configuration is unstable with respect to ion-ion beam instabilities and the ion-beam driven ion-acoustic instability.

the measurements are shown in the high-frequency part of Figure 6.12. This figure has been combined from the high-frequency observations of Hoppe et al [1982] and the low-frequency observations of Le & Russell [1992], both obtained from the ISEE spacecraft measurements. Tentative wave numbers have been determined for the high-frequency waves from measuring the time delay of the wave front arrivals at the two spacecraft ISEE 1 and ISEE 2 spacecraft yielding (surprisingly) short wavelengths < 100 km. Knowing the wave number, the frequency has been back-Doppler shifted $\omega = \omega_{\text{ISEE}} - \mathbf{k} \cdot \mathbf{V}_1$ into the plasma frame yielding unusually high frequencies $\omega \sim (20 - 100)\omega_{ci}$. This procedure determines the dispersion relation (naturally with large errors as indicated by the bars), which is found to be about linear in the narrow high frequency range of the waves in the rest frame of the plasma. The waves that have been left-hand polarised in the spacecraft frame turn out to become right-hand polarised in the plasma rest frame.

From these properties it was initially concluded that the waves propagate in the whistler branch and have been excited by the cool ion-ion beam instability, which would be consistent with our initial discussion. This conclusion is, however, questionable. The high rest-frame frequencies are not in agreement with model calculation [Sentman et al, 1981a; Gary, 1993] using the realistic observed beam properties. These yield wave frequencies of the order of $\omega \sim 0.1\omega_{ci}$ for both the ion-ion beam and firehose modes. In fact, Eastwood et al. (2005) report CLUSTER observations of similar waves but with much lower frequency $\omega/\omega_{ci} \sim 0.1$ and wavelength the order of $\sim 1 R_E \sim 6000$ km, which is in excellent agreement with the theoretical predictions. These waves are cold ion beam-excited fast-kinetic (magnetosonic) whistlers, in the terminology of Gary (1993) and Krauss-Varban et al [1994]. The high-frequency waves 1 Hz whistlers are instead most probably driven by the particular electron distribution in the foreshock [Feldman et al, 1983] rather than by the cold-ion beam instability. This claim is also supported by the fact that deeper in the ion foreshock, where the electron distribution becomes more isotropic, these waves do not occur separately from smaller magnetic field structures (SLAMS or shocklets).

In quite good agreement with the CLUSTER measurements are the ISEE foreshock waves analysed by Le & Russell [1992] who found the spectral peak at frequencies $\omega \sim 10^{-2}$ Hz with the spectrum broadening with increasing distance from the ion-foreshock boundary. This is shown in the upper left part of the spectrum in Figure 21.21 where the measurements of Hoppe et al [1982] and Le & Russell [1992] have been combined. Even though the conditions on the two observation times were not identical, one finds that the high-frequency whistlers identified by Hoppe et al [1982] occur on the approximate high-frequency extension of the Le & Russell [1992] 1930–2030 UT spectrum which is closest to the ion foreshock boundary. Note, however, that the low-frequency part of the Hoppe et al [1982]-spectrum (not shown here other but indicated by a thin straight line in the figure) was flatter and in the (relatively short) overlap region with the Le & Russell [1992]-spectrum did not show any indication of the magnetosonic foreshock-boundary waves, which must be due to the particular conditions prevailing during this observation period. Thus the Hoppe et al [1982] waves are a (high frequency) wave species that is different from these more common low-frequency/long-wavelength fast-magnetosonic waves. These latter waves are also of much higher (5 orders of magnitude) spectral density as seen from Figure 21.21 and has been confirmed by the later CLUSTER measurements. Moreover,

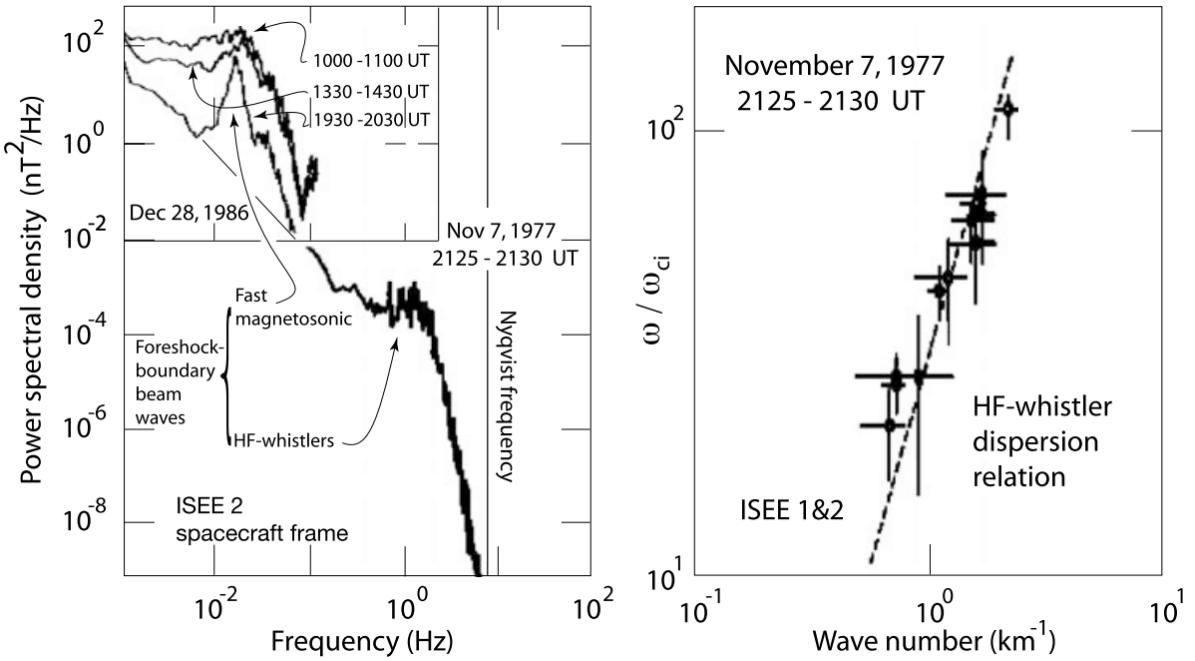


Figure 21.21: Left: The combined power spectral density of magnetic fluctuations excited by the ion foreshock boundary ion beam with central frequency around 1 Hz [lower part, data from Hoppe et al, 1982] and at <0.1 Hz [left upper part, data from Le & Russell, 1992] as measured by the ISEE spacecraft. This spectrum is obtained in the ISEE spacecraft frame. The low frequency spectrum is shown at three different times corresponding to (from below upward) increasing distance from inside the foreshock to the foreshock boundary. It is about five orders of magnitude more intense than the high-frequency waves. It is worth noting that the high frequency spectrum of Hoppe et al [1982] toward lower frequencies in the overlap with the Le & Russell [1992] spectrum does not show any indication of the lower frequency peak. As indicated by the thin straight line, it would smoothly continue into the 1930 UT branch at 0.01 Hz. This lack of the low frequency peak is probably accidental and due to the particular conditions at the time of measurement. Right: The wave dispersion relation for the high frequency 1 Hz whistler-mode waves determined from a tentative estimate of the relevant wave numbers and transformed into the plasma rest frame. It is seen that these waves in the observation range have about linear dispersion and frequencies $\omega \sim (20 - 100)\omega_{ci}$, far above the ion-cyclotron frequency ω_{ci} [data from Hoppe et al, 1982]

the magnetosonic waves are left-hand polarized in the plasma frame, and their compressive fast-magnetosonic character is proved from the in-phase variation of the density and magnetic field fluctuations shown in Figure 21.22.

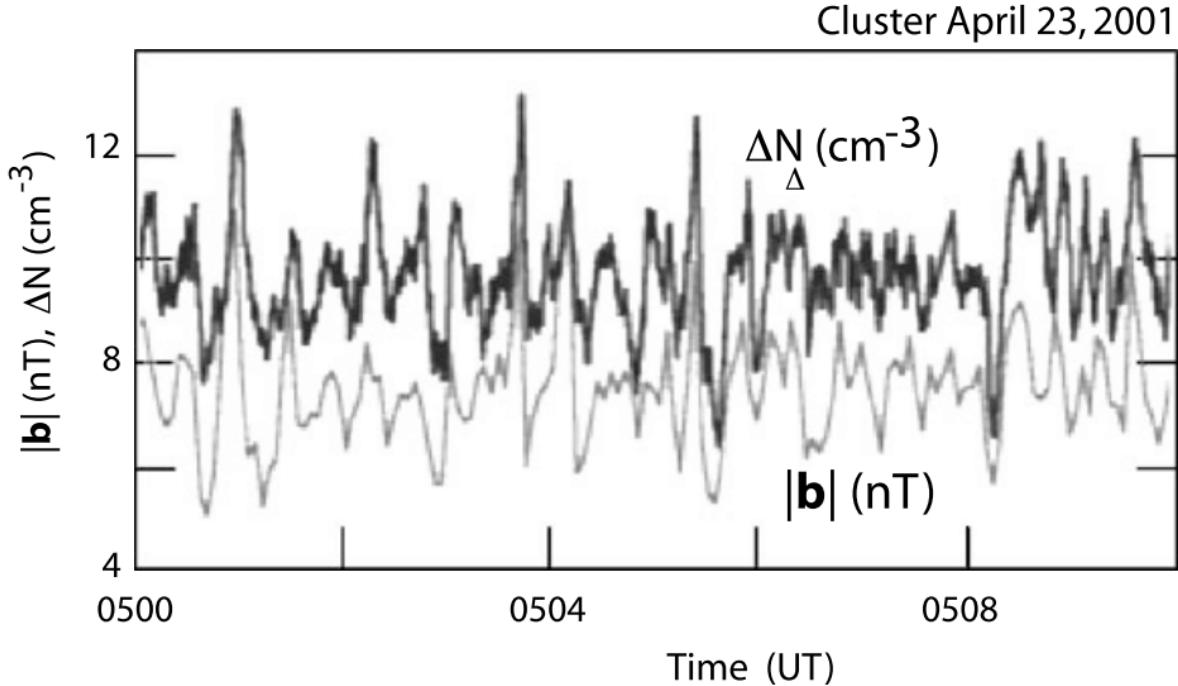


Figure 21.22: Left: The excellent correlation between the density and magnetic field variations in the ion-beam excited foreshock-boundary low-frequency waves as measured by CLUSTER [data from Eastwood et al, 2002]. The variations are practically in phase thus identifying the fluctuations as fast magnetosonic.

The presence of these low-frequency ion-beam generated waves implies that the ion beams interact with the fluctuating electromagnetic field. In this process they become scattered and diffuse in phase-space, which is the reason for the spreading of the ion beam in phase space and the final merging into the diffuse background distribution. Kis et al. (2007) have followed this evolution of the beam as we have described above. It is responsible also for the gradual spreading of the spectrum in the upper left part of Figure 21.21. A spectrum similar to that given there can be found in the paper of Kis et al. (2007). Archer et al [2005], from CLUSTER observations, inferred the nature of these ultra-low frequency (ULF) waves and showed that their correlation lengths along the wave vector direction \mathbf{k} is of the order of $1\text{--}3R_E$, while it can be a factor of three larger in the direction perpendicular to \mathbf{k} , rendering these waves oblate though nearly planar.

During times of hot beams Eastwood et al [2003] observed upstream propagating left-hand low-frequency waves which have been excited on the Alfvén-ion-cyclotron branch of the kinetic dispersion relation. These waves are kinetic Alfvén waves which have been excited by the

hot upstream propagating beam similar to the one shown on the right in Figure 21.15. In the spacecraft frame these waves because they are swept downstream by the flow, appear as right-handed waves. Eastwood et al [2003] report that with onset of the waves the beam gets more diffuse. This can either be interpreted as the reaction of the waves on the beam or that the spacecraft enters the region where the initially cold beam enters the foreshock, spreads in velocity and after becoming hot enough generates the observed waves.

Simulation studies available of the evolution of the reflected ion beam and the generation of upstream waves is still of great interest as of 2024!

21.5.3.2 Diffuse Ion Waves

Long-period 30 s waves are the rule in the foreshock. They occur together with the diffuse foreshock-ion component [cf., e.g., Sanderson et al, 1983; Mellott, 1986] which is located deeper inside the foreshock. Because of this reason, any waves that are excited by the diffuse ion component are restricted to the interior of the foreshock. One even has defined some fuzzy boundary of these waves called ULF-wave boundary [Russell & Hoppe, 1983; Greenstadt & Baum, 1986], which is even more inclined against the upstream magnetic field than the ion-foreshock boundary and outside of which the ULF wave activity should be weak. The reason is that ULF waves, if propagating upstream, can move at most with fast-magnetosonic velocity which for a supercritical shock is smaller than the stream and also less than the reflected ion beam velocities. The advection by the flow will blow them downstream towards the shock and confine them to a region relatively close to the shock bounded to upstream by the ULF-boundary.

The dispersion relation of the low-frequency waves in the foreshock from the three-dimensional observations of CLUSTER is given in Figure 21.23. The left part of the figure shows the dispersion relation in the plasma rest frame and the angle of wave propagation with respect to the local average upstream magnetic field. The scatter is quite large. Nevertheless it is surprising to find a well expressed nearly linear part on the dispersion relation even though the dispersion relation is composed of the contributions of several distinct modes with different polarizations. The representation is linear, which, on the k -axis, emphasises the short wavelengths (larger k values). From linear theory one expects that the long wavelengths (small k) should be more pronounced. Unfortunately, this region could not be resolved sufficiently at the available CLUSTER separation distances at that time. These low frequency waves, in the plasma frame, seem to propagate close to perpendicular to the magnetic field and, therefore, are probably in the fast magnetosonic mode. Moreover, it seems that a straight beam dispersion relation contributes to the perpendicular propagating waves. It also seems that in the low-frequency (magnetosonic) waves have negative frequency (i.e. propagating in opposite direction). Linear dispersion theory for the parameters during the observation time yields the curves on the right in the figure showing the coupling between the presumable beam mode and the four plasma modes at wave numbers $kV_A/\omega_{ci} < 0.3$. The experimentally determined dispersion relation resembles this clustering of couplings at small k and very low frequencies . However, application of

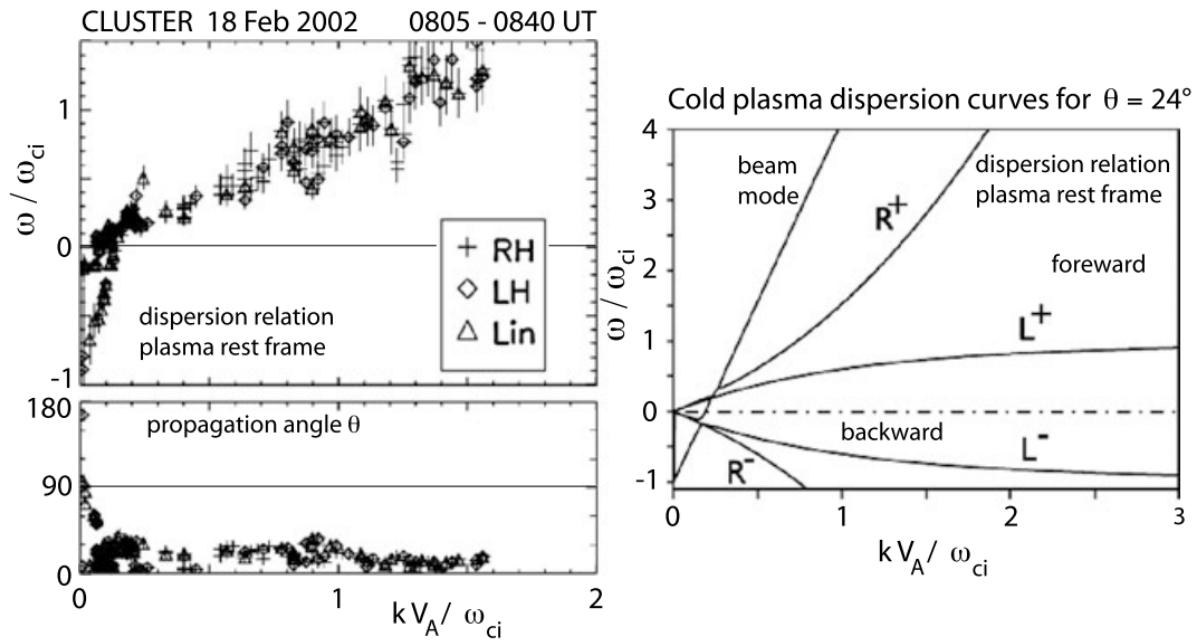


Figure 21.23: Left: The dispersion relation of low-frequency waves as measured by CLUSTER in the foreshock away from the foreshock boundary [data taken from Narita & Glassmeier, 2005], determined from the “wave-telescope” analysis method. The upper panel shows the dispersion relation $\omega(k)$ consisting of several parts and distinguished by the sense of polarization as right-hand, left-hand and linear. The different branches cluster. There are forward (positive frequency parallel to the magnetic field) and backward (negative frequency antiparallel) branches. The lower panel shows the wave propagation angle for the shorter wavelengths being mostly oblique near $\theta \approx (30 - 25)^\circ$, for longer wavelengths being more perpendicular to the magnetic field. Right: The five low-frequency modes in a cold beam plasma system of similar conditions as the foreshock plasma and for the propagation angle $\theta = 24^\circ$ showing similarity to the observed dispersion relation [data from Narita et al, 2003].

linear theory is dangerous if not questionable because the foreshock plasma is highly disturbed, the low frequency waves have rather large amplitudes and can barely be considered as linear waves. In addition the plasma is highly inhomogeneous exhibiting steep gradients in density and magnetic field. For an Alfvén speed $V_A \simeq 30 \text{ km s}^{-1}$ and an ion-cyclotron frequency $\omega/2\pi \simeq 1 \text{ Hz}$, wavelengths around the linear wave coupling are of the order of $k/2\pi \simeq 100 \text{ km}$ and should thus be affected by the plasma inhomogeneities (note that decreasing the reference cyclotron frequency decreases the wavelength even further). Nevertheless, at the very low frequencies and very long waves the waves are probably in the fast magnetosonic wave band.

More mysterious are the higher frequency-short wavelength waves. From comparison with the linear dispersion relation they seem to fit on the right-handed whistler branch (high frequency magnetosonic whistler R^+). However, inspection of the dispersion relation indicates that all kinds of polarization are scattered along the dispersion curve. The curve itself is very irregular even though the propagation angle of the waves seems to be about constant at weakly oblique angles. This wave composition suggests that we are not dealing here with one single wave mode but rather with the short wavelength part of a turbulent spectrum which has generated all kinds of short wavelength fluctuations with different almost randomly distributed polarization in a forward cascading process from long to short wavelengths. This idea has been elaborated in more detail in Narita et al [2006] who found that the shorter wavelength spectrum is indeed about featureless and power-law, close to a $-5/3$ -Kolmogorov-spectrum of stationary turbulence with most of the power in the perpendicular (non-compressive) magnetic component providing another argument for Alfvénic and magnetosonic turbulence. This is shown on the left in Figure 21.24. The figure also indicates the respective gyro- and inertial wave numbers, $k_{ci} = v_i/\omega_{ci}$ and $k_{in} = c/\omega_{pi}$. There is indication that the spectra change slope at around these numbers due to onset of anomalous ion viscosity and ion inertia. The large maximum on the perpendicular power curve near $k \sim 10^{-3} \text{ km}^{-1}$ corresponds to a wavelength of $\lambda \sim 6000 \text{ km}$ reported earlier [e.g., Eastwood et al, 2003] and is at the right position for energy injection by low frequency instability.

However, the conclusion that the deep foreshock is subject to fully developed fast magnetosonic/Alfvénic turbulence is by no means fully justified. This becomes obvious from considering the probability distribution of the field fluctuations as given on the right in Figure 21.24. Would the fluctuations be normally distributed then their distribution functions would have the shape of the dashed curves. Instead the distributions are skewed into tails. In the case of the parallel component the distribution lacks symmetry. Such tails may indicate that the turbulence is non-stationary, intermittent or inhomogeneous. All three cases might hold in the foreshock, in particular as the foreshock is a rather limited spatial region which is bounded from two sides and is subject to plasma injections and plasma losses. Nevertheless, the spectra determined contain signs of strongly nonlinear and turbulent interactions, and the waves, particularly the low-frequency waves, are of large amplitude and interact with the plasma ion component as also with other waves. These waves affect the upstream backstreaming ion component scattering, heating and accelerating it. On the other hand the upstream ion component is responsible for the existence of the waves as it is the ultimate energy source of the waves. And, to close the cycle, the waves cannot escape upstream very far from the foreshock since in a supercritical

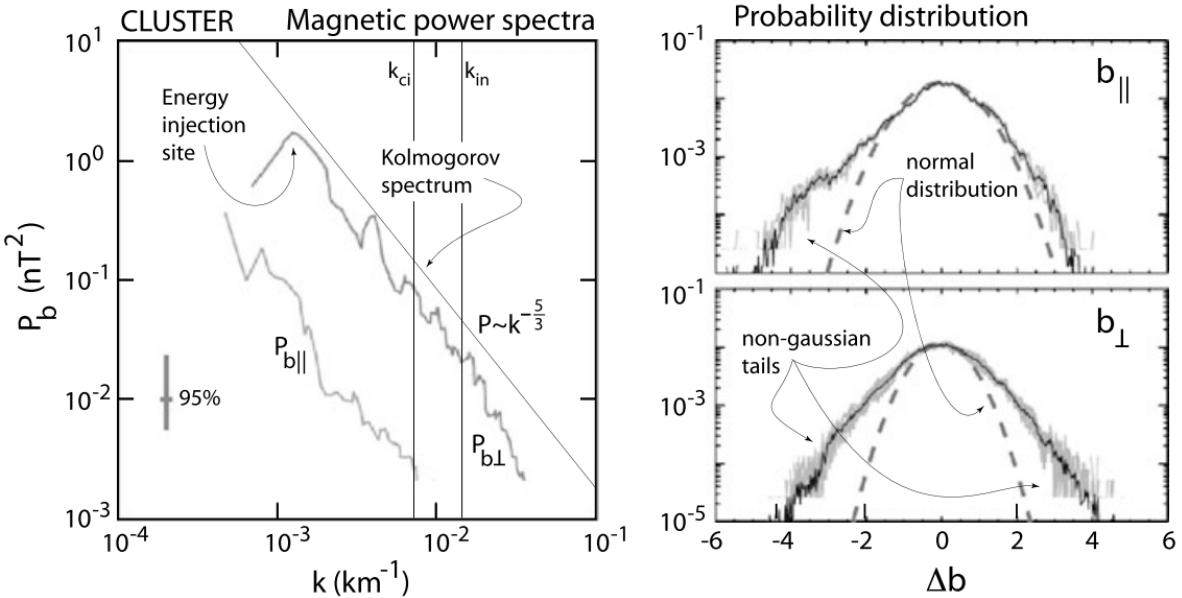


Figure 21.24: Left: Short wavelength spectra for the magnetic fluctuations measured by CLUSTER in the foreshock away from the foreshock boundary. These spectra are obtained by application of the “wave-telescope” analysis method. The perpendicular fluctuations are more than one magnitude more intense than the parallel fluctuations while the spectra decay about according to the Kolmogorov $-5/3$ -law of stationary inertial turbulence. Indication of a cut-off is seen at wave numbers larger than the inertial wave number k_{in} where strong dissipation sets on. From the maximum at $k \sim 10^{-3}$ one concludes that in this wavelength range the spectral energy is injected by instability of fast and Alfvénic magnetosonic waves which cascade nonlinearly forward towards shorter wavelengths. Right: The probability distributions of the parallel and perpendicular magnetic fluctuations. The dashed distributions are log-normal of same maximum. These distributions exhibit extended tails and thus indicate non-stationary and probably intermittent not fully developed magnetic turbulence [after Narita et al, 2006].

shock the Mach number of the flow is higher than any Mach number based on the wave speed and is thus high enough for the flow to advect the wave spectrum towards the shock. As long as the waves do not completely dissipate their energy during this convection, the electromagnetic wave energy accumulates at the location where they arrive at the shock.

The consequence of this accumulation is that the waves affect the shock, cause instability of the shock surface, and reorganise the shock front. This makes quasi-parallel shocks principally non-stationary (or as sometimes said, turbulent) and subject to some kind of irregular reformation. One may thus expect that a supercritical quasi-parallel shock does not represent a solid shock surface over a large area. It consists of a more or less dense “patchwork” of areas which together form a shock but which also appear and disappear in an irregular manner, come and go, and organise the shock in a certain volume where the entropy increases but where a multitude of very difficult to handle processes take place that can be investigated only experimentally or with the help of properly designed numerical simulations.

An important question that has not been addressed anywhere in the investigation of the ion-foreshock dynamics concerns the role of the foreshock electron component which, in the large ion-foreshock domain, consists of two populations, the 100 eV upstream electron population which belongs to the quasi-neutral upstream flow, and the hot several kev-electron population which is the product of the shock reflected electrons we are going to discuss in the next section. This component is isotropic but irregular and might contribute to ion-foreshock instabilities thereby affecting the foreshock turbulence and shock formation.

21.5.4 Electron Foreshock

Even though the ion-foreshock occupies the larger part of the foreshock and, which is more important, plays the decisive role in the dynamics of a quasi-parallel supercritical shock, it would be an ignorant attitude not to mention the part that aside of the main flow is populated solely by the reflected higher energy electron component, i.e. the electron-foreshock region. Schematically its extent and relation to the ion-foreshock is shown in Figure 22.14. The electron foreshock has been identified already by Scarf et al [1971] in the OGO 5 satellite observations. Scarf et al [1971] noted the occurrence of enhanced electron fluxes near the bow shock, which were related to observations of high-frequency electric-field spikes at 30 kHz. They concluded that these spikes resulted from Langmuir waves that had been excited by electron beams arriving from the bow shock along shock-connected magnetic field lines.²⁰ It could be confirmed that these electron fluxes were magnetic-field aligned and of higher energy than the bulk electrons in the main flow. This was interpreted as electron beams emitted into upstream direction from the shock, even though no mechanism was known that could provide the required shock-electron acceleration/reflection – in fact, a de Hoffman-Teller-frame shock-reflection mechanism had already been proposed by Sonnerup [1969] and was reinvented and elaborated on for electrons much later by Wu [1984].

²⁰A detailed overview of the early observations is given by Klimas [1985].

21.5.4.1 Electron Beams

In analogy to the ion-foreshock boundary the electron-foreshock boundary (which is the ultimate upstream boundary of the foreshock) carries narrow bursts of electron beams, which escape into upstream direction from the shock along the magnetic field and which are slightly displaced by the convective flow into downstream direction.

An example of a narrow energy spectrum of such a beam measured by the IMP-8 satellite is shown on the left in Figure 21.25. The spacecraft was in the electron foreshock boundary for a relatively short time only. Being near apogee, the spacecraft was about standing. Thus the short contact time with the beam most probably indicates that the electron-foreshock boundary is fairly narrow or that the electron beams are bunched into electron bursts. The electron beam also occupies only a small volume in velocity space corresponding to a narrow bump on the distribution [Anderson et al, 1981]. Only one energetic electron-gyroradius deeper in the electron foreshock the electron beam becomes depleted and just forms an energetic tail of hot halo electrons on the electron distribution function of the flow.

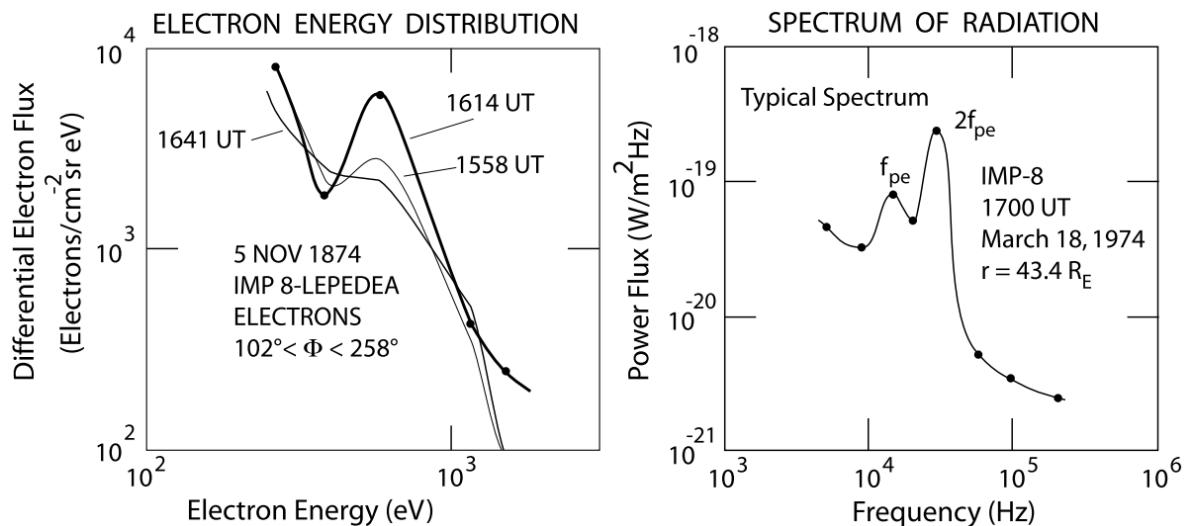


Figure 21.25: Left: The electron energy distribution as seen by IMP-8 at three times when crossing the foreshock boundary. Right: The high-frequency wave spectrum of the electromagnetic radiation as measured some distance away from the shock near the foreshock boundary. Two types of emissions are visible, the so-called ‘harmonic’ emission $f = 2f_{pe}$ at twice the electron plasma frequency $f_{pe} = \omega_{pe}/2\pi$, and the weaker so-called ‘fundamental’ emission at f_{pe} [after Gurnett & Frank, 1975].

Figure 21.26 shows electron phase-space observations from the electron instruments on ISEE 1 & 2 during crossings of the electron-foreshock boundary. The left part of the figure [taken from Fitzenreiter et al, 1984] is a much higher time and velocity (respectively energy) resolution plot

than that in the former figure. It is nicely seen how the electron distribution evolved from a field-aligned (nearly) Maxwellian distribution prior to contact with the foreshock boundary, through a field-aligned-beam-like distribution at contact, into a distribution with an energetic tail along the magnetic field, where the electron beam has been completely washed out. This transition takes place within a time interval of 15 s, just allowing to obtain the three measurements. The pseudo-three-dimensional plot of the electron distribution function on the right was obtained on another day when the spacecraft passed the electron-foreshock boundary and a series of distributions were recorded from upstream, across the foreshock boundary, and into the electron foreshock. Here only the foreshock-boundary distribution is shown when the narrow electron beam has evolved flowing into upstream anti-shockward direction along the magnetic field. Note the narrow angular extension of the beam, its weakness, and its comparably large velocity which allows to identify it just outside the bulk of the hot 100 eV background electrons which are much hotter than the beam (indicated by the small beam width in v_{\parallel}).

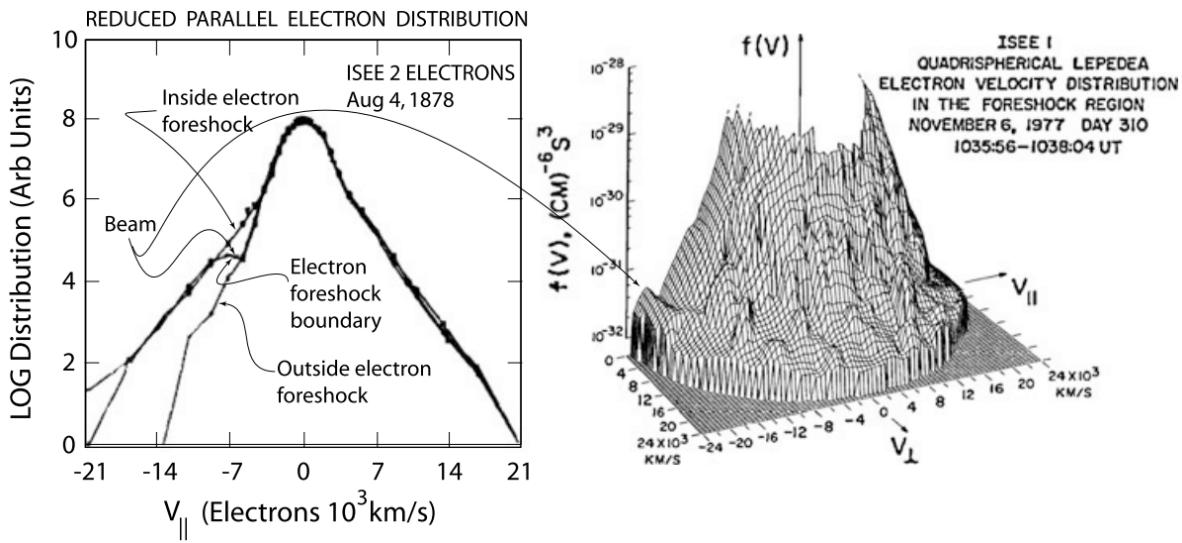


Figure 21.26: Left: Three successive reconstructions of the reduced parallel electron distribution function during crossing of the electron foreshock boundary. The distributions are only seconds apart, at 1650:37 UT before touching the foreshock boundary, at 1650:43 UT just crossing it, and at 1650:52 UT being behind it in the electron foreshock. The beam is visible only during the short crossing time. Afterwards the electron distribution shown a heated non-symmetric tail, indicating that the spacecraft sees the depleted beam in the hot electron-foreshock plasma [data from Fitzenreiter et al, 1984]. Right: The full two-dimensional electron distribution function during another crossing of the electron-foreshock boundary at the moment when the beam is visible as a narrow enhancement of the distribution in the negative parallel velocities (upstream along the magnetic field). An instant later, when the spacecraft entered into the electron foreshock the beam had disappeared [after Anderson et al, 1981].

The conclusion that can be drawn from these observations, which in the follow-up time had also been confirmed by measurements of other spacecraft like AMPTE IRM and AMPTE UKS, is that, similar to the ion foreshock, the electron foreshock consists of a very narrow region along about the shock-tangential upstream-magnetic field line (or flux tube) which is populated by a cold, weak, and fast electron beam, and a broad electron region adjacent downstream of this flux tube where the electron distribution exhibits tails and in addition contains a hot, fairly isotropic electron component. The electron beam (like the ion beam at the ion-foreshock boundary) has its origin near the shock-tangential magnetic field line which is connected to the quasi-perpendicular part of the shock. The mechanism of its generation has not yet been unambiguously clarified. Presumably it consists of a combination of the de Hoffman-Teller frame mechanism of Sonnerup [1969] and the bending of the shock surface plus some kind of stochastic acceleration in the electric fields which evolve in the shock transition, ramp and overshoot. Such fields have indeed been reported recently on scales below the electron skin depth $d_e = c/\omega_{pe}$ to be very large, of the order of $\leq 100 \text{ mV m}^{-1}$ parallel and perpendicular to the magnetic field [Bale & Mozer, 2007]. The presence of the hot electron component in the foreshock, on the other hand, is not due to this reflection mechanism but requires another mechanism that heats and accelerates electrons at a quasi-parallel shock over a large area of its surface.

Langmuir Waves

So-called ‘gentle electron beams’, which are just those beams one observes at the electron-foreshock boundary – fast, but not too fast (V_b few times V_1), parallel ($V_b \equiv V_b, V_{b\perp} = 0$), weak ($N_b \ll N_e$), cool ($v_{eb} < v_e \ll V_b$), are known to be the drivers of Langmuir waves with dispersion relation

$$\omega^2(\mathbf{k}) = \omega_{pe}^2 + 3v_e^2k^2, \quad v_{\text{res}} > \sqrt{3}v_e$$

via the resonant kinetic gentle-electron beam instability. The Langmuir resonance conditions is $\omega - \mathbf{k} \cdot \mathbf{v}_{\text{res}} = 0$, where $\mathbf{v}_{\text{res}} \sim \mathbf{v}_b$ is the resonant electron velocity. Combining it with the dispersion relation, it is easy to see that in order for the wave number k to be real the resonant velocity must satisfy the condition on the right in the above equation.

Electron Foreshock-Boundary waves

Figure 21.27, taken from Décréau et al [2001], shows a beautiful double passage of the CLUSTER-spacecraft quartet through the Earth’s bow shock on December 22, 2000 in the light of the CLUSTER plasma wave spectrum recorded by the WHISPER plasma wave instrument aboard the four CLUSTER spacecraft. Figure 21.28 shows an example of these high-resolution measurements.

CLUSTER came from the magnetosphere, entered the downstream region of the shock, passed the shock to upstream, re-entered the shock, and escaped downstream again. The relevant signatures of each region are seen in all four spacecraft. The uppermost light blue trace is the electron plasma frequency $f_{pe} = \omega_{pe}/2\pi$ which maps the local plasma density. The shock appears as the intense broadband emission with maximum intensity in the low-frequency waves. In the density it is mapped as a steep drop in f_{pe} to the low upstream density values. Close to

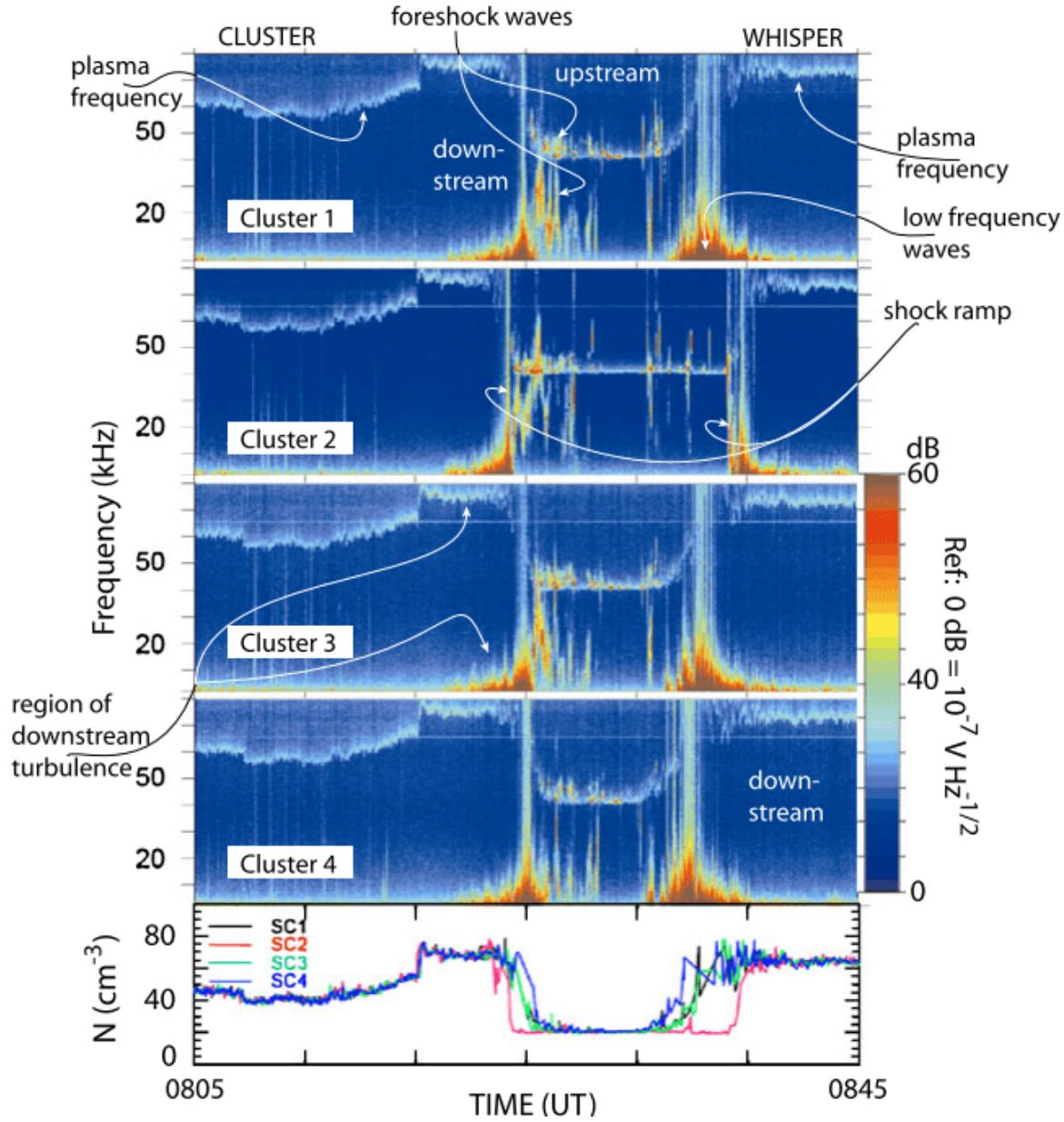


Figure 21.27: A CLUSTER spacecraft passage across the Earth's bow shock region on December 22, 2000. The spacecraft cross the downstream magnetosheath region (left) and the shock, enter the upstream solar wind region and back downstream. The figure shows broadband (0 kHz – 80 kHz) plasma wave electric field spectra from all four CLUSTER pacecraft (Cluster 1–Cluster 4). The uppermost light blue emissions in each of the panels are at the electron plasma frequency f_{pe} which maps the local plasma density N estimates of which are shown in the lowest panel. Passages of the shock are signalled by intense broadband waves starting at low frequencies and being correlated with steep drops in N . In the low density upstream solar wind the spots of strong intensification of f_{pe} indicate contact with the electron foreshock boundary beam and excitation of Langmuir waves. These are highly structured, sometimes stretching above, sometimes hanging down below f_{pe} . Also seen are intermediate frequency emissions near the shock [data taken from Décréau et al, 2001].

the shock a number of intensifications in f_{pe} can be recognised. These occur at the time of contact with the electron foreshock boundary when intense plasma waves are excited in the Langmuir mode. It is interesting to note that the intensifications occur in spots and are not necessarily narrow-band. They may exceed f_{pe} , or they may also drop below it: the plasma frequency may have ‘hair’ or ‘beards’. Moreover, sometimes intense lower-frequency emissions are detected half way between the plasma frequency and the low frequency emissions in the foreshock. The most intense of these emissions belong to the region closer to the shock where the density gradient has not yet settled to the upstream values, but weaker emissions of the same kind occur farther away in the upstream low density domain.

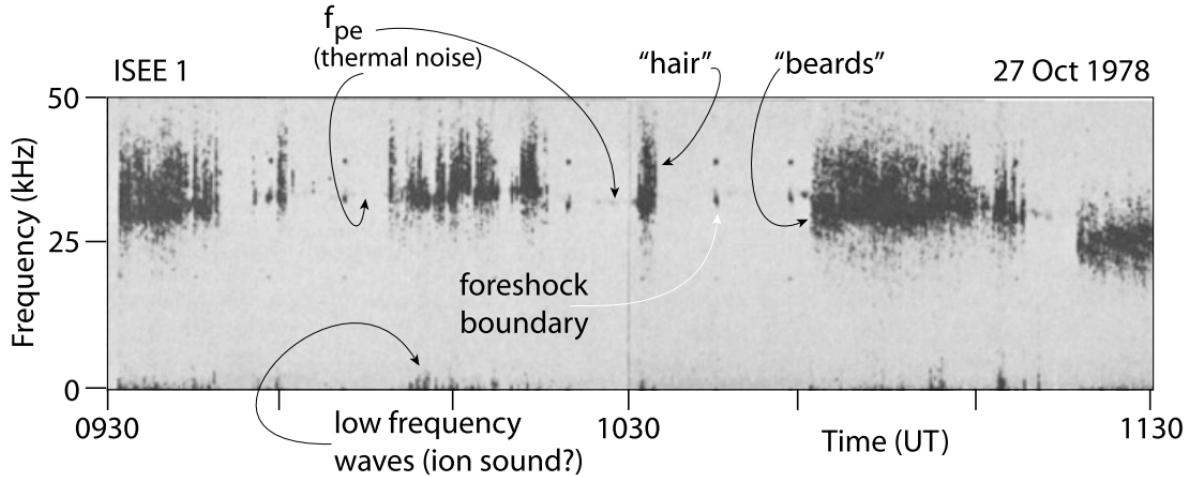


Figure 21.28: An example of ISEE 1 high frequency wave observations during contact and passage of the electron foreshock boundary. The thin faint spotty line around 30–40 kHz is the local plasma frequency $f_{pe} \sim 30$ kHz, seen here only as thermal noise with lesser instrumental sensitivity than in the CLUSTER observations. The occasional intense dark spots as the one marked by the white arrow are brief contacts with the foreshock boundary field line when the beam occurs. deeper in the foreshock the spectrum broadens nonsymmetrically, evolving ‘hair’ and ‘beards’. Some moving structures can be identified in the hair. The emissions deeper inside the foreshock are correlated with the occurrence of low frequency emissions below a few kHz which are probably ion-sound waves [after Etcheto & Fauchez, 1984].

It is highly suggestive that the observed intense spots in the plasma frequency, with the ‘hair’ parts exceeding and the ‘beard’ parts hanging down from the local f_{pe} , are related to magnetic field-aligned electron beams emanating from the shock and propagating upstream along the electron foreshock-boundary magnetic field as these are the same signatures as those observed with the ISEE and AMPTE spacecraft, though with much better frequency resolution and higher sensitivity here. Taking into account that the upstream plasma frequency away from the shock is fairly constant this implies that the shock-normal angle θ_{Bn} changes on an irregular

and fast time scale, i.e. that the direction of the upstream magnetic field is highly variable even under conditions of apparent quiescence in the upstream medium. The conclusion drawn from this would *not* be that the upstream magnetic field is subject to violent variations as in the electron foreshock there is no reason for the magnetic field to be strongly affected. Rather this variability points on the temporal variation of the shock area near the tangential field line. Even a small change in the shock normal there will cause a large variation in the location of the tangential magnetic flux tube at a distance upstream of the shock. The main reason is that the super-critical shock is neither in thermal nor in thermodynamic equilibrium and is thus by its very nature subject to changes in all its physical parameters.

21.5.4.2 Radiation

Shocks are frequently referred to as sources of radiation. Famous examples are supernova shocks, which are visible in almost all wavelengths, from radio through visible light up to x-rays [e.g., Dickel & Wang, 2004], and solar type II shocks with their main radiation signatures seen in the radio waves.

Supernova shocks are relativistic shocks which are not treated here. Their Mach numbers range from weakly relativistic to highly relativistic, but the energy per particle in them remains to be less than the rest energy of an electron, $m_e c^2 = 0.511 \text{ MeV}$, which allows to treat them classically. This does not hold anymore for the central supernova engine which drives the flow and which in some cases results in the generation of ultra-relativistic jets. There the shocks become non-classical, and not only radiation losses but also particle generation must be taken into account in their description.

Theory of shock-emitted radiation is based on plasma processes which under the prevailing collisionless conditions in the shock and foreshock plasmas refer to wave-wave coupling as the main generation mechanism. Direct emission from particles is unimportant, because the energy losses a particle experiences when becoming retarded or reflected at a shock, are not transformed into radiation. In all non-relativistic cases radiation losses can completely be neglected compared with all other energy losses. Nevertheless, the observed radiation is of interest because in many cases, where no measurements are possible to be performed in situ, radiation is the only direct and presumably identifiable signature a shock leaves, when seen from remote. The other signature is the generation of energetic particles, which will be treated next, but energetic particles are a more diffuse indicator of a shock, because their propagation is vulnerable to scattering from other particles, obstacles and, in the first place, scattering by magnetic fields.

21.5.5 Quasi-Parallel Shock Reformation

In quasi-parallel supercritical shocks there is not such a stringent distinction between the region upstream of the shock and the shock itself like in quasi-perpendicular shocks. The foreshock

and the shock itself cannot be considered separately due to the presence of the reflected and diffuse particle components in the foreshock. These are the source of a large number of waves.

The interaction of these waves with the shock is one of the main issues in quasi-parallel shock physics. In the present section this will become clear when we will be dealing with the formation, behaviour and structure of quasi-parallel shocks as it has been inferred less from observation than from numerical simulations. The reason is that the real observations in space do not allow to separate the particles and waves from the shock. They all occur simultaneously and are interrelated and can never be observed in their initial state. The observations to which we will nevertheless occasionally refer will leave the impression of large-amplitude noisy fluctuations.

21.5.5.1 Low-Mach Number Quasi-parallel Shocks

It is usually assumed that low-Mach number shocks are stable, i.e. show no substantial time variation or reformation. It is not completely transparent why this should generally be so.

1. The critical Mach number has been shown by Kennel et al [1985] to become small at narrow shock normal angles $\theta_{Bn} \rightarrow 0$, in which case even low-Mach number quasi-parallel shocks should become supercritical and reflect ions.
2. Any fast ions of parallel velocity $v_{\parallel} > V_1$ that have been heated in the shock can in principle escape from the quasi-parallel shock upstream along the magnetic field and should appear in the foreshock where they contribute to wave generation.

Therefore, it makes sense to investigate the state of quasi-parallel shocks in view of their stability and wave generation even for low Mach numbers. In addition, any waves that are generated in the shock ramp or transition with upstream directed \mathbf{k} and fast enough parallel phase or group velocities could also escape from the shock in upstream direction. This could, in particular, be possible just for low Mach number quasi-parallel shocks.

To check this possibility Omidi et al [1990] have performed one-dimensional hybrid simulations finding that initially the quasi-parallel shock consisted of phase-standing dispersive (magnetosonic) whistler waves with the last whistler wave cycle constituting the shock ramp. As expected, the wave vectors of these phase-locked magnetosonic whistlers are aligned with the shock normal. At later times, backstreaming ions along the upstream magnetic field excite a long-wavelength whistler wave packet upstream of the shock. In the one-dimensional simulation the wave vector is restricted to the shock normal while it is known from theory that the growth rate is largest along the magnetic field. These oblique whistler waves should thus show up in two-dimensional simulation and may be visible at larger distance for sufficiently large upstream phase velocities.

Such two-dimensional (hybrid) simulations with a non-inertial electron fluid have been performed by Scholer & Fujimoto [1993] for a Mach number of $MA = 2.2$ and angles $\theta_{Bn} = 20^\circ, 30^\circ, 45^\circ$ and by Dubouloz & Scholer [1995] for an angle $\theta_{Bn} = 30^\circ$ and high Mach number $MA = 5$ (Figure 21.29). The lower Mach number simulations show the presence of a substantial number

of backstreaming ions which cause an ion-ion instability in the upstream region. However, the excitation and properties of the waves depend strongly on the shock normal angle. Initially, as in the case of Omidi et al [1990] phase-locked short-wavelength whistlers appear which are replaced at later times by upstream long-wavelength whistlers with phase velocity directed and amplitude growing towards the shock ramp but upstream directed group velocity, i.e. the shock radiates energy away towards upstream, as one would naively expect, because the shock being supercritical must reject the excess inflow of energy which it can do by both, reflecting particles and emitting waves into upstream direction. These waves are excited by the backstreaming ion component in a strongly nonlinear interaction process because of the evolving steep ion-density gradient, which is of the same scale as the whistler wavelength. The \mathbf{k} -vector turns away from the magnetic field having comparable components parallel to \mathbf{B} and parallel to the shock normal \hat{n} . For small θ_{Bn} a remnant of the initial phase-locked whistlers survives but disappears at $\theta_{Bn} = 30^\circ$. Close to the shock, where the backstreaming ion density is high, the waves have short wavelengths, and \mathbf{k} is almost parallel to \hat{n} . In the high Mach number simulations no shock is produced but instead reflected ions were artificially injected with same Mach number as the incoming flow but with much higher temperature $v_i = 14.1V_A$, forming a spatially uniform ion beam. The intention was to investigate the effect of the hot reflected ions. This is shown on the right in the above figure. The result resembles the former one where a shock was generated by reflection at a wall, but the effect in the injected beam case is stronger because of the higher Mach number. Hence it is the hot reflected ion component that is responsible for the wave dynamics and the shock dynamics.²¹

All this can be seen from the two-dimensional intensity contours of these waves in the foreshock, which are plotted at a relatively early time in the shock evolution in the simulations in Figure 6.26. On the right in this figure the geometry is given, with the magnetic field fluctuation vector \mathbf{b} in the (y,z) -plane. The bottom panel on the left shows contours of the b_z fluctuations in the (x,z) -plane. The nominal shock ramp is at $x \approx 145d_i$ at this time. The upper panel shows the non-coplanar component b_y -contours in the same representation. Behind the shock the fluctuations are irregular and disorganised. However, in front of the shock a clear wave structure is visible with strongly inclined wave fronts and of roughly $\sim 10d_i$ wavelengths in z parallel to the nominal shock surface. The wavelength in x is about three times as short.

These waves are seen in both components, b_z , b_y , are low amplitude at large distance from the shock but reach very large amplitudes simultaneously in both components during shock approach while, at the same time, bending and assuming structure in z -direction that is different from the regular elongated shape at large distance. This deformation of wave front may be due to the residual whistlers near to the shock, but it implies that the shock has structure on the surface in both directions x and z and is not anymore as planar as was initially assumed. The shock becomes locally curved on the scale of the shock-tangential wavelength. The waves deform the shock and, in addition, being themselves of same amplitude as the shock ramp, become increasingly indistinguishable from the shock itself. *The shock is, so to say, the last of the large-amplitude magnetic wave pulsations in downstream direction, and the shock-magnetic*

²¹I think these reflecting wall setups are unrealistic!

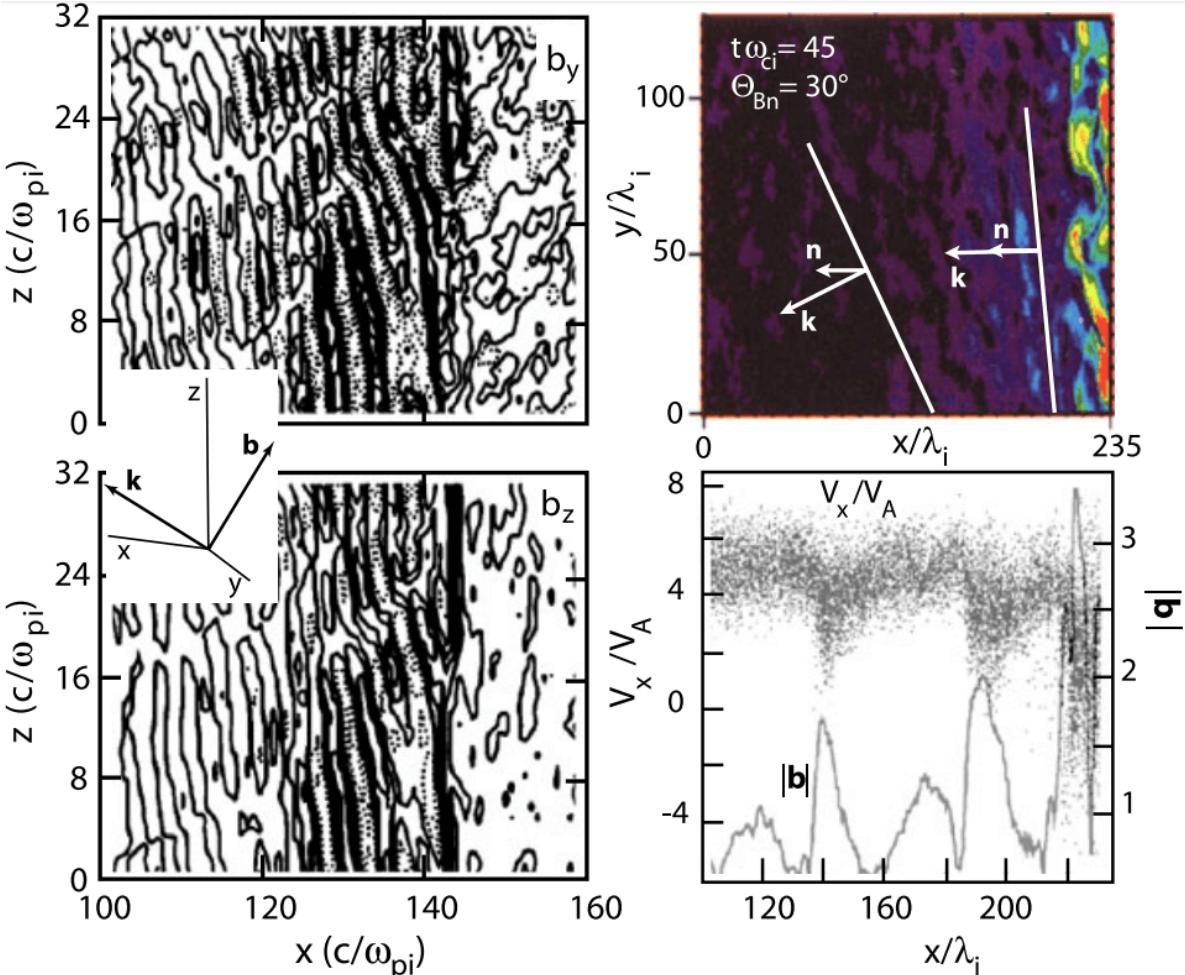


Figure 21.29: Two-dimensional hybrid simulations of the evolution of upstream waves. Left column: Upstream wave in a low-Mach number $MA = 2.2$ quasi-parallel shock [after Scholer & Fujimoto, 1993]. The nominal shock is in the (y,z) -plane. Ions escaping to upstream generate the oblique upstream magnetosonic waves. The contour plot of the two normalized to the upstream magnetic field components of the magnetic fluctuations shown is taken at relatively early times $t\omega_{ci} = 68$. It shows the nearly plane magnetic wave fronts inclined against the shock in direction x and having wavelengths of $\sim 10c/\omega_{pi}$ in z while being much shorter in x . In the vicinity of the shock the wave fronts turn more parallel to the shock and produce a non-coplanar magnetic component $|b_y|$ which is of same order as the $|b_z|$. Moreover, even though the shock has relatively low Mach number, it is not completely stable but shows structure in z direction which is produced by the presence of the reflected upstream particles and the upstream waves. At places it is impossible to identify one single shock ramp. Right column: Two-dimensional hybrid simulations of the evolution of giant magnetic pulsations (SLAMS) in front of a quasi-parallel supercritical shock [after Dubouloz & Scholer]. Top: The simulation plane showing the structure of the (normalized) magnetic fluctuation field $|b|$ at time $t\omega_{ci} = 45$ and shock normal angle $\theta_{Bn} = 30^\circ$. The accumulation of the growing wave fronts at the shock transition, their increasing amplitudes, and their turning towards becoming parallel to the shock is clearly visible from the rotation of the two wave fronts and their \mathbf{k} vectors shown in white. Away from the shock transition the angle between \mathbf{k} and the shock normal \hat{n} is large. Close to the shock the two vectors are about parallel. The magnetic field is in the wave front, so θ_{Bn} is close to 90° here. Bottom: Pulsation amplitude and ion phase space. The fluctuations evolve into large amplitude pulsations when approaching (and making up) the shock. The strong extensation of the upstream

field is not anymore coplanar, because the waves have contributed a substantial component by that points in the direction out of the coplanarity plane.

We have emphasised this phrase, because it expresses the importance of the low-frequency upstream magnetic waves in quasi-parallel shock physics. Contrary to quasi-perpendicular shocks where the reflected gyrating ions in combination with the reflected-ion excited modified-two stream instability were responsible for the shock dynamics and different kinds and phases of shock reformation, quasi-parallel shock reformation and much of its physics is predominantly due to the presence of large-amplitude and spatially distinct upstream waves. These are the generators of the shock and, due to their presence, the shock changes its character. It is highly variable in time and position along the shock surface and is – close to the shock transition on a smaller scale – ‘less quasi-parallel’ (or more perpendicular, i.e. the shock-normal angle θ_{Bn} has increased on the scale of the upstream waves). The latter is due to the out-of coplanarity-plane component of the upstream waves. This conclusion remains basically valid also in full particle simulations. It had been suggested already earlier on the basis of ISEE 1 & 2 observation of the evolution of the upstream ultra-low frequency wave component [Greenstadt et al, 1993].

The gradual evolution of the shock normal angle Θ_{Bn} has been demonstrated in other hybrid simulations by Scholer & Fujimoto [1993] and Dubouloz & Scholer [1995] who investigated the evolution of the shock normal angle in dependence on distance from the shock. This is shown in Figure 21.30 for two-dimensional hybrid simulations with initial shock-normal angles $\theta_{Bn0} = 2^\circ$ and $\theta_{Bn0} = 20^\circ$, respectively, which we anticipate here. In both cases Θ_{Bn} evolves from quasi-parallel to quasi-perpendicular angles. Qualitatively there is little difference between the two cases. At the shock ramp θ_{Bn} is deep in the domain of quasi-perpendicular shocks. The only difference is that for the nearly parallel case the angle jumps to quasi-perpendicular quite suddenly, just before approaching the shock ramp, while the evolution is more gradual for the larger initial θ_{Bn} . In both cases the evolution is not smooth, however, which is due to the presence of large-amplitude foreshock waves. Transition to quasi-perpendicular occurs for the initially nearly parallel case at the nominal shock ramp while for the initially quasi-parallel case it occurs at an upstream distance of about $100d_i$ from the shock. This is sufficiently far away for the upstream flow ions to feel the change in the shock normal.

One notices that this transition is on the ion scale, implying that in the region close to the shock the ions experience the shock occasionally (because of the large fluctuations in θ_{Bn}) – and when ultimately arriving at the shock – as quasi-perpendicular. It is thus not clear, whether the electrons do also see a quasi-perpendicular shock, here (which cannot be concluded from these hybrid simulations, however). The ISEE measurements of the electron distribution function by Feldman et al [1983] at the shock do not show a difference between quasi-perpendicular and quasi-parallel shocks. This fact suggests, in addition that, close to the shock transition, quasi-parallel shocks behave like quasi-perpendicular shocks as well on the electron scale, which is just what we have claimed.

We finally note that the behaviour of the shock normal angle gives a rather clear identification of the location of the shock transition in the quasi-parallel case, as indicated in Figure 21.30 by shading. Three distinctions can be noticed:

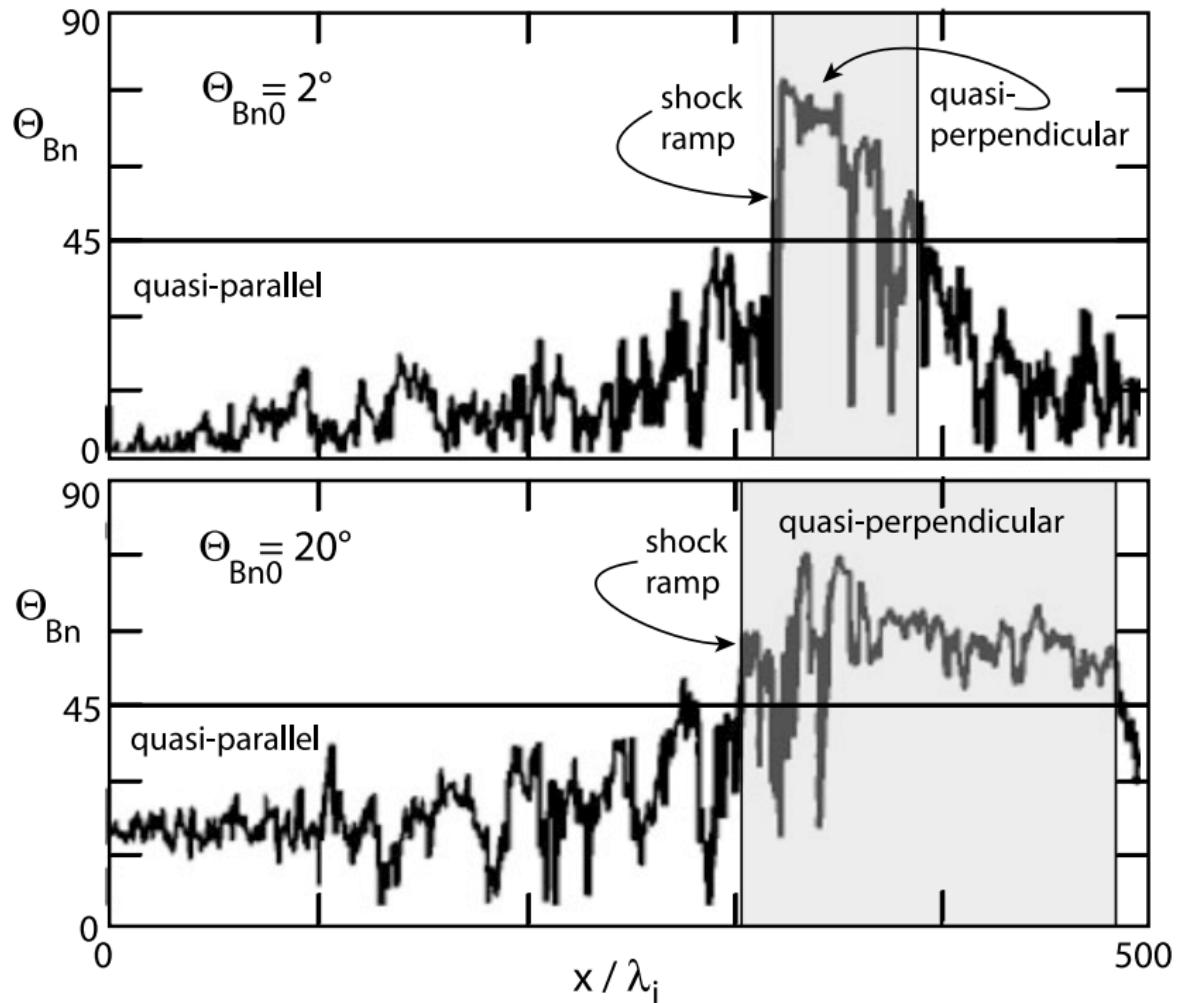


Figure 21.30: The evolution of the shock normal angle θ_{Bn} on distance from the shock in two-dimensional hybrid simulations for two initial quasi-parallel shock-normal angles $\theta_{Bn0} = 2^\circ$ and $\theta_{Bn0} = 20^\circ$, respectively [Scholer & Fujimoto, 1993]. The horizontal line at 45° is the division between quasi-perpendicular and quasi-parallel shock normal angles. In both cases θ_{Bn} evolves from quasi-parallel direction into quasi-perpendicular direction. The shaded areas identify the quasi-perpendicular domains.

1. At larger initial shock-normal angles the transition to quasi-perpendicular angles occurs earlier, i.e. farther upstream than for nearly parallel shocks. This is due to the stronger effect of the large amplitude upstream waves in this case;
2. At larger initial shock normal angles the quasi-perpendicular shock transition is considerably broader than for nearly parallel shocks, i.e. it extends farther downstream before the main quasi-parallel direction of the magnetic field in the downstream region takes over again and dominates the direction of the magnetic field
3. At an initial shock-normal angle of 20° , this region is $\sim 150d_i$ wide, implying that the magnetic field direction behind a quasi-parallel shock remains to be quasi-perpendicular over quite a long downstream distance measured from the shock ramp. For the nearly parallel shock this volume is only about $\sim 50d_i$ wide.

This observation must have interesting implications for the physics downstream of quasi-parallel shocks. For instance, applied to the Earth's bow shock, where $d_i \sim 10^3$ km. both distances correspond to regions wider than the order of $> 5R_E$ which is larger than the nominal width of the magnetosheath! Thus, behind the bow shock a substantial part of the magnetosheath plasma should behave as if the bow shock would have been a completely quasi-perpendicular shock.

21.5.5.2 Turbulent Reformation

When speaking about turbulent reformation we have in mind that a quasi-parallel super-critical shock is basically a transition from one lower entropy plasma state to another higher entropy plasma state that is mediated by a substantially broad wave spectrum. Such a transition has been proposed by J. Schwartz Steven and Burgess (1991) based on the detection of the large amplitude magnetic pulsations (SLAMS) in the foreshock.

Figure 21.31 on the left shows this model assuming that somewhere upstream in the foreshock magnetic pulsations have been excited which become convected downstream toward the shock by the convective flow, grow in amplitude and number and accumulate at the shock transition to give rise to a spatially and temporally highly variable transition from upstream of the shock to downstream of the shock. An important clue in this argument was the observation that, first, the pulsations grow in amplitude when approaching the shock and that, second, they slow down. This slowing down is effectively an increase in their upstream directed velocity on the plasma frame with growing amplitude such that their speed nearly compensates for the downstream convection of the flow.

On the right of the figure, which is suggested by the observations of Kis et al. (2007), a larger region is seen. Here the pulsations are the result of growing ULF waves which are generated in a volume inside the foreshock but relatively close to the ion-foreshock boundary. These waves grow to large amplitudes until evolving into pulsations which the flow carries toward the shock. Growth, slowing down, and accumulation then lead to the pile up of the pulsations at

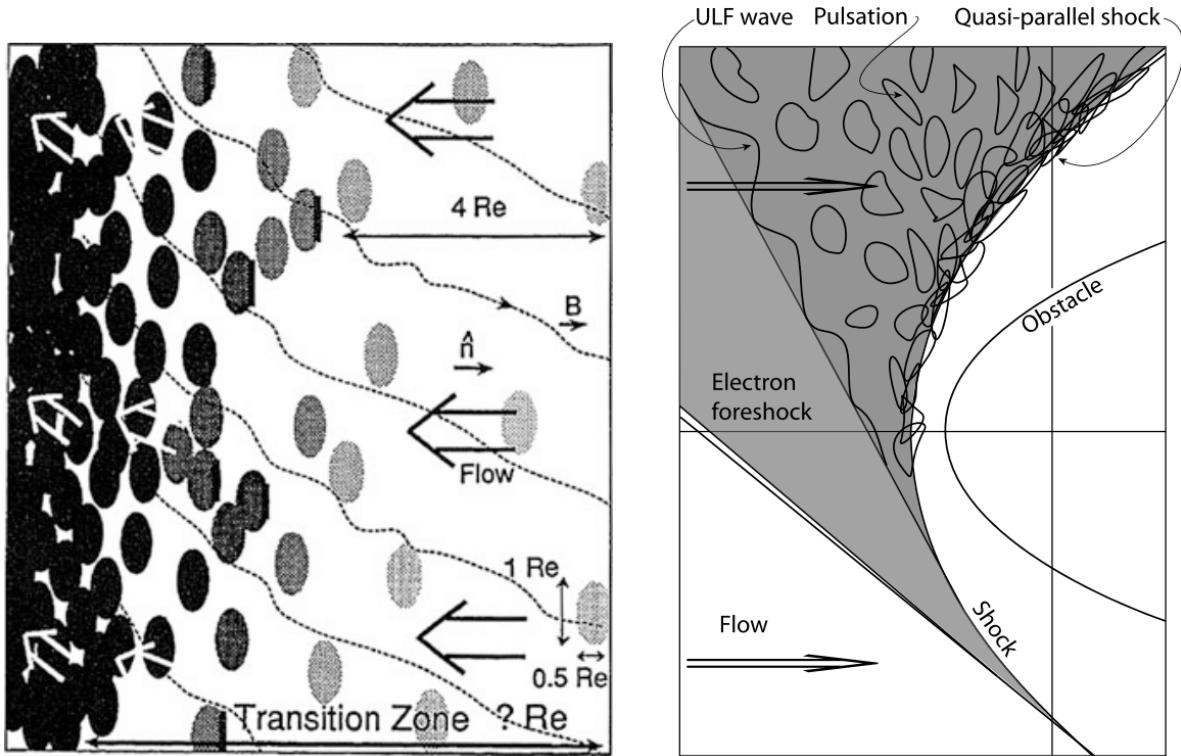


Figure 21.31: The patchwork model of J. Schwartz Steven and Burgess (1991) of a quasi-parallel supercritical shock reformation. Left: Magnetic pulsations (SLAMS) grow in the ion foreshock and are convected toward the shock where they accumulate, thereby causing formation of an irregular shock structure. Note also the slight turning of the magnetic field into a direction that is more perpendicular to the shock surface with the shock surface itself becoming very irregular. Right: The same model with the pulsations being generated in the relatively broad ULF-wave-unstable region in greater proximity to the ion-foreshock boundary. When the ULF waves evolve to large amplitude and form localised structures these are convected toward the shock, grow, steepen, overlap, accumulate and lead to the build up of the irregular quasi-parallel shock structure which overlaps into the downstream direction.

the shock location and formation of the turbulent shock structure. Only the shock-tangential upstream magnetic field lines are plotted in the right figure. In the left part, several field lines are schematically shown exhibiting the fluctuations imposed by the background level of ULF fluctuations. Moreover, a certain bending of the field lines is included here in approaching the shock transition with the field lines turning more perpendicular the closer they come to the shock. This bending is what we claim to be a parallel shock turning quasi-perpendicular at a scale very close to the shock. In this schematic drawing, however, there would be no reason for the magnetic field to turn this way. What closer observations and simulations show is that the turning of the field is the result of the presence of the large amplitude magnetic pulsations. This will become clearer below.

Observations

Of course, the model shown in Figure 21.31 is a schematic model only which, however, has some merits in explaining the observations. The signature of a quasi-parallel shock in the magnetic field is indeed quite different from that of a quasi-perpendicular shock. The magnetic signature of a quasi-parallel shock is quite broadband in that it lacks a clear location of the shock front, which contains a broad region of very large amplitude compressive oscillations in magnetic magnitude and in the direction of the magnetic field that subsequently is recognized as a passage across the quasi-parallel shock.

An example is shown in Figure 21.32 as measured by the CLUSTER spacecraft. This figure shows eight hours of observation by CLUSTER in the immediate vicinity of the quasi-parallel shock. It is difficult to say where in the figure the shock transition is located as the large fluctuations in the magnetic field magnitude and directions mask the various back and forth passages across the shock that are contained in the data. Clearly, at the beginning near 1400 UT the spacecraft was in the downstream region. The fluctuations show that during almost the entire sequence the magnetic field exhibits compressive fluctuations. These belong to the shock transition. At the same time large fluctuations in the direction of the magnetic field are also observed. In the compressions of the magnetic field buried are also upstream pulsations (SLAMS), and many of the changes in direction belong to the ULF waves present at and near the shock. The changes in direction indicate that the shock does not behave like a stationary flat surface. Instead, it shows structure with highly fluctuating local shock normal directions.

Lucek et al. (2002) checked this expectation by determining the local shock-normal angle θ_{Bn} and comparing it with the prediction estimated from magnetic field measurements by the ACE spacecraft which was located farther out in the upstream flow. The interesting result is that during the checked time-interval of passage of the quasi-parallel shock the prediction for the shock normal was around $20 - 30^\circ$, as expected for quasi-parallel shocks. However, this value just set a lower bound on the actually measured shock normal angle. The measured θ_{Bn} was highly fluctuating around much larger values and, in addition, showed a tendency to be close to 90° . It strengthens the claim that quasi-parallel shocks are locally, on the small scale, very close to perpendicular shocks, a property that they borrow from the large magnetic waves by which they are surrounded. In fact, we may even claim that locally, on the small scale ($\sim \mathcal{O}(d_i)$), quasi-parallel shocks are quasi-perpendicular.

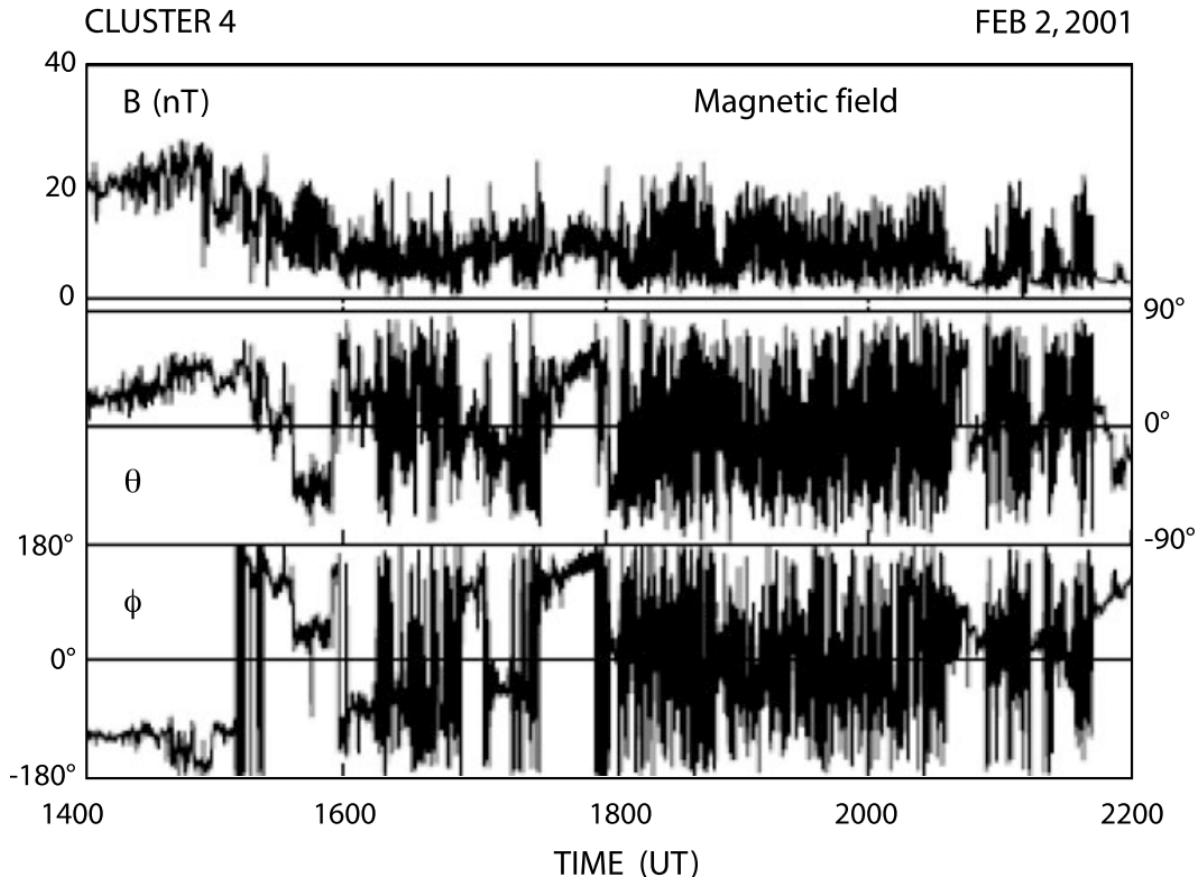


Figure 21.32: Eight hours of CLUSTER magnetic field data during a long passage near and across the quasi-parallel supercritical (Alfvénic Mach number $M_A \sim 12 - 13$, ion inertial length $d_i \sim 140$ km) bow shock. The time resolution was 4 s. The top panel shows the variation in the magnitude of the magnetic field. The two lower panels are the respective elevation and azimuthal angles θ and ϕ , in a GSE coordinate frame (Lucek et al. 2002). Large variation in the magnetic compression and direction can be seen to be associated with this quasi-parallel shock crossing. Buried in these large variations on this highly time-compressed scale are many magnetic pulsations (SLAMS). The compressive large amplitude fluctuations in the upper panel are typical for a quasi-parallel shock transition.

The data suggest that, indeed, the quasi-parallel shock is the result of a build-up from upstream waves which continuously reorganise and reform the shock. Figure 21.33 shows three representative examples of such upstream waves which are far from being continuous wave trains. The upper panel is taken from the large density fluctuation region in the shock transition. This region turns out to consist of many embedded magnetic pulsations (SLAMS) of very large amplitudes. In the present case amplitudes reach $|\mathbf{b}| \sim 25$ nT. These pulsations have steep flanks and quite irregular shape, exhibit higher frequency oscillations probably propagating in the whistler mode while sitting on the feet or shoulders of the pulsations. It is most interesting that the different CLUSTER spacecraft – at spacecraft separation < 1000 km – do not observe one coherent picture of a particular pulsation. This implies that the pulsations in the shock are of shorter scale than spacecraft separation: the different spacecraft observe different structures respectively different pulsations. In addition, the magnetic field directions (not shown in the figure) are very different from spacecraft to spacecraft and thus from pulsation to pulsation, and even for one pulsation at its front edge and trailing edge different magnetic field directions are observed. The directions of the magnetic normal across a pulsation change on very short spatial scales. The quasi-parallel shock front has thus a rather irregular shape, which will be bent locally with changing direction of its normal.

The second panel shows an isolated pulsation farther away from the shock transition. This pulsation exhibits a much more coherent way on the four different CLUSTER spacecraft. It seems as if the pulsation is still in evolution as three of the spacecraft see a nearly coherent structure while the fourth which is farther away sees it in a different state. Concluding from this event, isolated pulsations seem to have larger dimensions and lower amplitudes, which would be consistent with the assumed solitary properties of pulsations. The third panel shows a shocklet, i.e. a structure which presumably has little in common with pulsations. It is embedded into long wavelength ULF wave trains, evolves into steep front and drives whistlers attached to this front across the flow. These waves were already observed by Russell et al [1971]. Their properties indeed resemble those of sub-critical little shocks which propagate against the flow, though with slower speed such that they effectively are slowly convected towards the shock. Observations of plasma wave spectra by AMPTE IRM are shown in Figure 6.31. There the large amplitude SLAMS pulsations appear as broadband wave signatures resembling narrow shock structures.

Another interesting property of magnetic pulsations in the shock transition region where they overlap to form the quasi-parallel shock is presented by Behlke et al [2003]. Measurement of the electric cross-SLAMS potential identify a substantial unipolar drop in the electric potential of several 100 V, corresponding to a potential ramp, when passing from upstream to downstream across the pulsation. Such a drop signifies the presence of an electric field in one direction across the pulsation. Taking the mean size of a pulsation to be roughly 1000 km, the mean electric field is $\langle E \rangle \simeq 400$ mV m⁻¹. However, this field drops mainly at the leading edge of the pulsation.²² Such a field presumably corresponds to a steep pressure gradient in the pulsation.

²²It could (but is unlikely) also be generated by an anomalous collision frequency. This remains to be tested by further observations and data analysis.

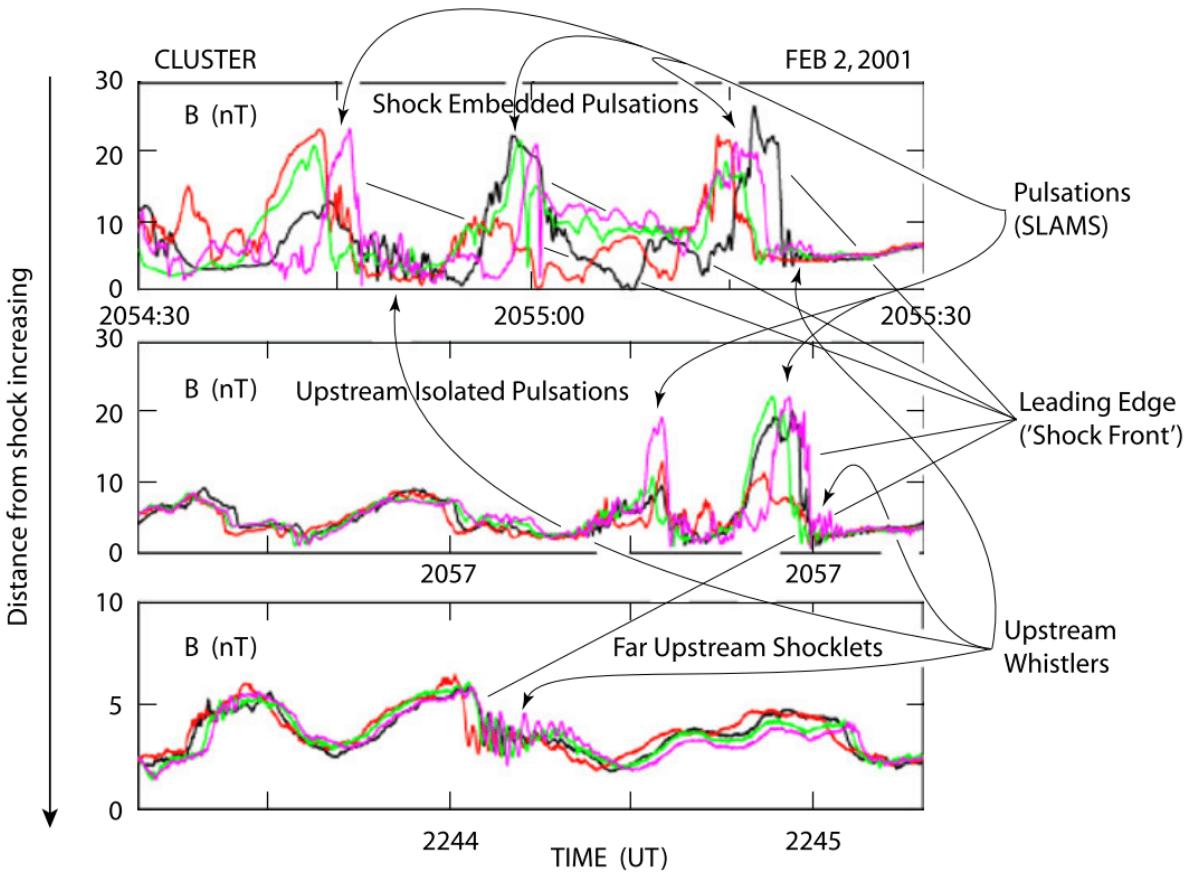


Figure 21.33: CLUSTER magnetic field measurements of magnetic pulsations (SLAMS) near and remote from a quasi-parallel shock on February 2, 2001 (Lucek et al. 2002). Only the magnetic field magnitude is shown for all four CLUSTER (colour coded) spacecraft. The spacecraft separation was between a few 100 km and 1000 km. Top: Clustered pulsations in the shock transition. Three events of large amplitudes are shown. These structures are very irregular with steep fronts. Note that in spite of the small spacecraft separation the shapes of the structures differ strongly from spacecraft to spacecraft. Moreover, determination of the pulsation fronts and normals (not shown) indicates high variability over the spacecraft separation distance. Thus the structures are of relatively small scale and large amplitude. Middle: Isolated pulsation at greater distance from shock outside compression region. The structures are seen almost simultaneously at the spacecraft and thus must be of larger size. Compared to the embedded pulsations the amplitudes are lower, and the structures are more regular. Bottom: A shocklet observed outside the pulsation region in the domain of ULF waves. The steep shock-like front is well expressed with the attached whistler waves it carries with it. Note the much lower amplitude than the pulsations.

The measurements of Behlke et al [2003] anticipated the later report of strong electric fields in the shock by Bale & Mozer [2007]. In addition to this observation it was found [Behlke et al, 2004] that the single pulsations were subject to a fairly large number of high frequency/Debye scale structures in the electric field seen in the WHISPER recordings (bottom panel), which belong to electron holes or solitons which form in the pulsation gradient regions as shown in Figure 6.32. The bipolar electric field and unipolar potential across one – indeed very symmetric – soliton is seen in this figure. These observations suggest that the pulsations are indeed the main constituents of a quasi-parallel shock with the dynamics on the micro-scale of a quasi-parallel shock going on mainly in the single pulsations of which the shock transition is built.

The occurrence of these intense nonlinear electrostatic electron plasma waves at the quasi-parallel shock transition is intriguing. It forces one to draw another very important main conclusion from these observations: that quasi-parallel shocks are sources of electron acceleration into electron beams, which are capable to move upstream along the magnetic field over a certain distance and excite electron plasma waves at intensity high enough to enter into the nonlinear regime, forming solitons and electron holes (BGK modes). This is possible only if quasi-parallel shocks are quasi-perpendicular as well on the electron scale.

Indeed, electron reflection into beams which generate Langmuir waves is presumably possible only when in the supercritical quasi-parallel shock transition region the magnetic field changes from quasi-parallel to quasi-perpendicular on the electron scale $\sim d_e$. Indications of such a change on the ion scale d_i have been noted above at a number of occasions, but the detection of solitary structures in the electron plasma waves in relation to quasi-parallel shock transitions provides a very strong argument for this to be true on a scale which is well below the ion scale. Only if this is the case, there will be ample reason for electrons to become reflected and accelerated into beams from the transition region in a quasi-parallel shock. As we already noted, we may, therefore, expect that quasi-parallel supercritical shocks on the electron scale are not anymore quasi-parallel but change to become locally quasi-perpendicular, while on the larger ion scale they still remain quasi-parallel. Because of this reason high Mach number quasi-parallel shocks will be embedded into a diffuse high energy electron component that emits radiation.

If this conclusion will turn out to be true, it will have important consequences for collisionless shock physics. Supercritical, collisionless, nonrelativistic shocks will, in fact on the electron scale, always behave quasi-perpendicularly, and it may be suspected that this conjecture will also hold for relativistic shocks though probably for other reasons (like the generation of transverse magnetic fields by the Weibel instability, which becomes dominant in relativistic shocks [see, e.g., Jaroschek et al, 2004, 2005]). This implies also that the true quasi-parallel shock physics cannot be properly elucidated when ignoring electron effects as is, for instance, done in hybrid simulations.

1D Hybrid Simulations

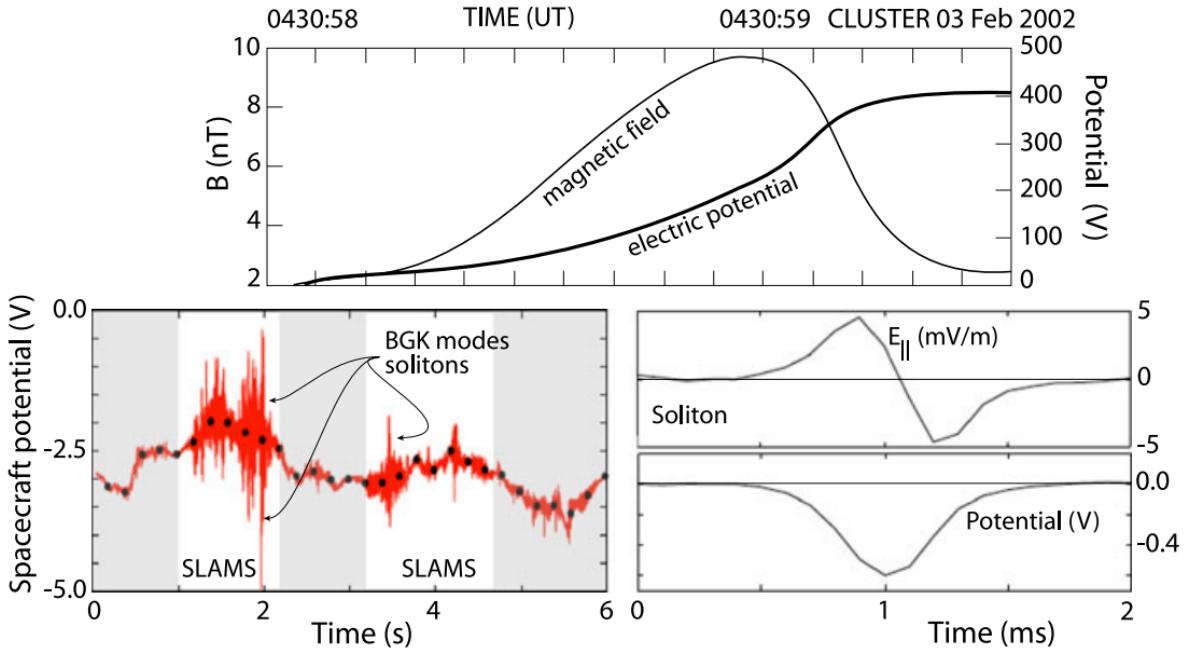


Figure 21.34: Observation of electric field structures in large magnetic pulsations (SLAMS) in the quasi-parallel shock transition region [Behlke et al, 2003, 2004]. Structures on three different time scales are shown, corresponding also to three different spatial scales. Top: CLUSTER passage across a (moderately large amplitude) magnetic pulsation in the shock transition. The (smoothed) magnetic field structure is a slightly steepened magnetic bump. The stationary parallel electric potential field across this structure shows a potential ramp with steep gradient at the leading edge of the pulsation. The potential drop of 400 V corresponds to an electric field of $\sim 0.47 \text{ V m}^{-1}$. Note that the time scale in this panel is 90 s. Bottom left: Six seconds of a CLUSTER passage through the shock transition. The black dots show the spacecraft potential variation which maps the local density variation. Overlaid is the high frequency WHISPER trace of the plasma frequency. In the magnetic pulsation regions (white) the plasma frequency exhibits huge excursions to both sides similar to those on wave observations. These excursions trace the BGK (non-symmetric) modes and (symmetric) solitons. Bottom right: One example of one of the solitons on a 2 ms time scale. It is nicely seen how symmetric the parallel potential trough and the corresponding bipolar parallel electric field shape look like in the solitary wave structure.

We do already know from the observations that quasi-parallel shocks exist at small θ_{Bn} with their foreshocks being populated by a diffuse ion component that excites upstream waves and mediates the beam-generated upstream ion-foreshock boundary waves. The impossibility for this diffuse component of being entirely due to shock reflection in the quasi-perpendicular part of the shock, immediately proves that the quasi-parallel (or even the nearly parallel shock) must be able to reflect particles upstream. Hence, either a quasi-parallel shock is capable of generating a large cross-shock potential, or it is capable of stochastically – or nearly stochastically – scattering ions in the shock transition region in pitch angle and energy in such a way that part of the incoming ion distribution can escape upstream, or – on a scale that affects the ion motion – a quasi-parallel shock close to the shock transition becomes sufficiently quasi-perpendicular that ions are reflected in the same way as if they encountered a quasi-perpendicular shock.

Observations suggest that the latter is the case, while observations also suggest that large potential drops occur in the large-amplitude magnetic pulsations (SLAMS) where they accumulate in the shock ramp [Behlke et al, 2004]. Hence, reflection of ions will be due to the combination of both effects, the electric potential drop and the magnetic deflection. In fact, this can be a quite complicated process for an ion passing across a number of magnetic pulsations, in each of which it is being retarded and at the same time deflected by a small angle until its normal velocity component is decreased sufficiently that a further deflection in pitch angle suffices to let it return into the upstream region.

Early hybrid 1D simulations have already confirmed some characteristics of quasi-parallel shocks. First 1D hybrid simulations in an extended simulation box [Burgess, 1989] suggested that the reformation of quasi-parallel shocks is about cyclic and is caused by the impact of large-amplitude upstream waves. Scholer & Terasawa [1990], using $\theta_{Bn} = 20^\circ$ and $MA = 3.5$ with an upstream ion thermal velocity $v_{th,i} = V_A \$$ (corresponding to $\beta_i = 1$) in 1D hybrid simulations (with small numerical resistivity) showed that the reflected ions are not coming from the core of the incident upstream ion distribution but originate in the shell of this distribution, having initial velocities $v \simeq 1.7v_{th,i}$. These ions escape from the shock quite far upstream and excite ULF waves with upstream directed velocity of $\sim 1.3V_A$ at distances up to $> 300d_i$, which are convected downstream to reach the shock. In this 1D hybrid simulation the only mode in which they can propagate is the compressive fast magnetosonic mode.

These waves are in fact what in observations has been identified as pulsations (SLAMS) but is not yet recognized as such, here. During downstream convection the waves grow and slow down in the interaction with the foreshock ion component. When approaching the shock they generate a large amount of new reflected ions. These slow the incident ion population down and steepen the wave crest, which becomes the new shock front. In the time between the arrival of the compressive waves the shock is about stationary and develops phase-locked upstream whistlers which the arriving next wave crest destroys. From these simulations it could not be concluded what process produced the reflected ions, however, as 1D simulations among suffering from other efficiencies select only one particular direction of wave numbers and are thus not general enough for drawing final conclusions.

The nature, generation and effects of the large-amplitude upstream waves have been further investigated in more detail in 1D [Krauss-Varban & Omidi, 1991; Scholer & Burgess, 1992; Scholer, 1993, among others], and in 2D hybrid simulations [Krauss-Varban & Omidi, 1993; Scholer, 1993; Dubouloz & Scholer, 1995]. Since shocks are three-dimensional, it is clear that two-dimensional numerical simulations at same resolution come closer to reality. However, they suffer from restrictions in size of the simulation box and simulation time. Since reality does not confront us with an initial state, large boxes and long times are needed.

In order to identify a particular wave mode, the dispersion of the wave must be investigated. This dispersion relation depends on the frame in which it is taken, because the energy/frequency of a wave is not invariant with respect to coordinate transformations; in a medium moving with velocity \mathbf{V} it is Doppler shifted according to $\omega' = \omega(\mathbf{k}) - \mathbf{k} \cdot \mathbf{V}$, where $\omega(\mathbf{k})$ is the dispersion relation in the rest frame of the flow. While the Doppler shift at high frequency is negligible, it completely changes the dispersion of ultra-low frequency waves at large Mach numbers. Figure 21.35 on its left shows the deformation of the fast mode dispersion relation in the shock frame at large Mach number $M > 1$ and for waves propagating upstream in the plasma frame. The deformation causes negative frequencies of the waves which imply downstream directed phase velocities, which is nothing else but the intuitive downstream convection of the waves by the flow. However, the minimum in the dispersion relation implies that waves of a particular frequency and wave number have zero group velocities. In the shock frame the energy of these waves is stationary. Smaller wave numbers have energy moving downstream, larger wave number have energy moving upstream away from the shock. The right part of the figure shows simulations of upstream waves according to 1D hybrid simulations by Krauss-Varban & Omidi [1991] for $\theta_{Bn} = 30^\circ$ and a Mach number $M_A = 2.8$. The entire dispersion of the simulated waves is negative. The waves are all convected toward the shock as their Mach number is less than the streaming Mach number. Near the shock most of the wave energy moves downstream and will cross the shock. Still some shorter wavelength waves (large k) move in energy upstream in the shock frame. Farther away from the shock most of the wave energy encountered is seen to move upstream.

Scholer [1993] investigated these waves further in one-dimensional hybrid simulations performing several numerical experiments on them, taking away the shock and instead injecting a diffuse ion component from downstream. The main finding is that the large amplitude upstream magnetic pulsations (SLAMS) evolve out of the ULF wave spectrum in the interaction with the diffuse ion component. In accord with observation the pulsations move upstream in the plasma frame. Thereby their upstream leading edge steepens and is right-hand circularly polarised like required for whistlers. However, dispersion is unimportant: the main cause of the evolution of large pulsations is nonlinearity when the wave interacts with the diffuse ion distribution. This distribution has a steeper shock directed density gradient than the pulsation wavelength. Moreover, the flow becomes decelerated at the leading edge of the pulsation (as is seen in Figure 21.29), and here the velocity difference between the flow and the diffuse ion component drastically decreases, which shifts the k vector of the resonant wave to larger values, and the

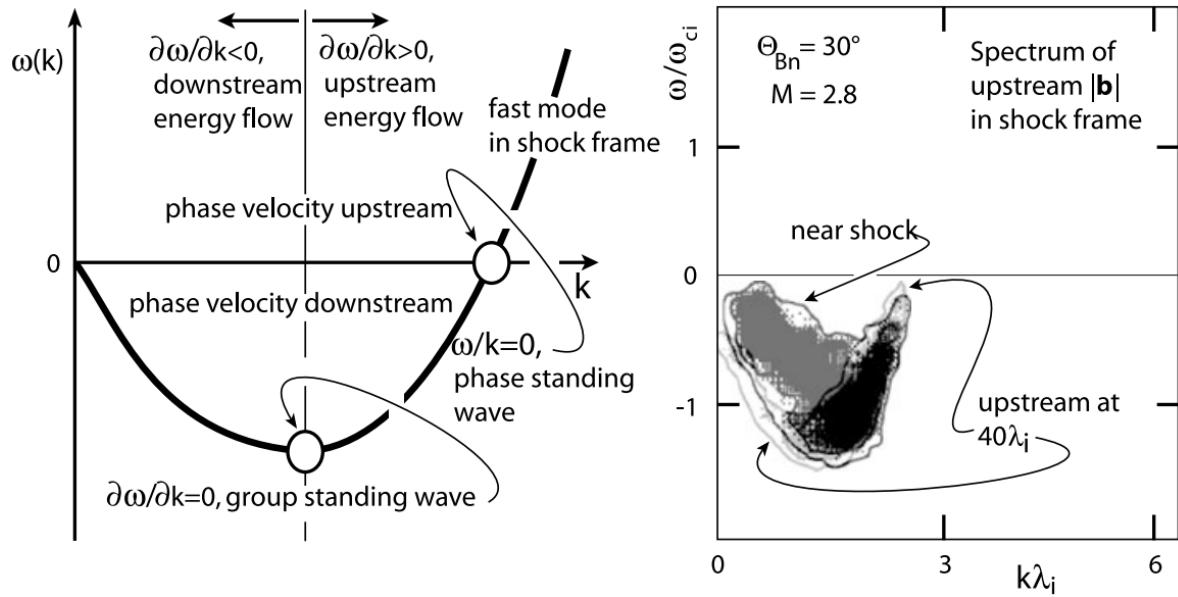


Figure 21.35: Fast mode dispersion relation in simulations in the shock frame. Left: The Doppler shifted fast mode dispersion relation in a supercritical flow in the shock frame. The Dispersion relation assumes negative frequencies corresponding to the downstream convection by the flow of Mach number $M_A = 2.8$ and at $\theta_{Bn} = 30^\circ$. Waves at zero group velocity have energy at rest in the shock frame. Negative group velocities imply downstream transport of energy, positive group velocities imply upstream transport. Right: Simulated upstream wave dispersion spectra near the shock and upstream of the shock. Near the shock wave energy accumulated around standing and downstream transport. Away from the shock the wave energy still moves upstream [data from Krauss-Varban & Omidi, 1991].

wavelength decreases during convection of the pulsation toward the shock.²³ The standing whistlers in the leading edge are simply generated by the current flowing in the edge. It is thus concluded that it is the gradient in the hot diffuse ion component over a length of the same order as the wavelength which produces the pulsations. Ultimately these pulsations cause a quasi-periodic reformation of the shock, as we have described earlier, which thus has been proven by 1D hybrid simulations. The same result is obtained when the simulation starts right away without a shock but with an injected hot beam (which is no surprise as the generation of the shock, before it was removed in the former simulations in order to keep with the wave field, was due to the plasma flow-reflected ion beam interaction).

2D Hybrid Simulations

The 2D evolution of the pulsation (SLAMS) was studied later by Scholer & Fujimoto [1993] and Dubouloz & Scholer [1995] (see Figure 21.29). It basically confirmed the conclusions drawn from 1D simulations with the following three important modifications:

1. The wave fronts of the pulsations (SLAMS) rotate into a direction that is more parallel to the shock thereby increasing the shock-normal angle locally to become quasi-perpendicular.
2. The pulsations have short wavelength in shock normal direction, but are of substantially longer but finite lengths in the direction tangential to the shock, which provides structure to the shock in tangential direction.
3. Shock reformation is a result due to the steeping and accumulation of the pulsations and is a quasi-periodic process, but the downstream structure of the shock over some distance is caused by the downstream convection of the old shock front, i.e. the bulk of the pulsations that had accumulated at the location of the former shock transition.

It is interesting to note that from one reformation cycle to the next rather large fluctuations in the magnetic field and density exist in the transition from upstream to downstream which are capable of trapping particles. The 2D simulations do also confirm the conclusion that the diffuse upstream ion component is responsible for the growth of the pulsations (SLAMS). The 2D simulations by Dubouloz & Scholer [1995], in particular, followed the same scheme as the 1D simulations, injecting hot diffuse ions into upstream in order to control the interaction between the diffuse ion component and large amplitude pulsations (SLAMS).

Full Particle PIC Simulations

In hybrid simulations, the ions are macro-particles while the electrons represent a charge-neutralising background. Clearly, such simulations are unrealistic if whistlers become involved.

21.5.6 Parallel Shock Particle Reflection

There are two possible mechanisms:

²³Note that no resonant ion beam-whistler interaction exists as the beam is hot and diffuse.

- A quasi-parallel shock is capable of generating a large cross-shock potential, or it is capable of stochastically – or nearly stochastically – scattering ions in the shock transition region in pitch angle and energy in such a way that part of the incoming ion distribution can escape upstream.
- On a scale that affects the ion motion, a quasi-parallel shock close to the shock transition becomes sufficiently quasi-perpendicular that ions are reflected in the same way as if they encountered a quasi-perpendicular shock.

21.6 Instabilities and Waves

In the context of collisionless shocks the instabilities of interest can be divided into two classes. The first class contains those waves which can grow themselves to become a shock. It is clear that these waves will be of low frequency and comparably large scale because otherwise they would not evolve into a large macroscopic shock. The primary candidates are magnetosonic, Alfvén and whistler modes. A number of waves can form secondarily after an initial seed shock ramp and grow in some way out of one of these wave modes: these are ion modes which have now been identified to be responsible for structuring, shaping and reforming the shock. In fact real oblique shocks — which are the main class of shocks in interplanetary space and probably in all space and astrophysical objects — cannot survive without the presence of these ion waves which can therefore be considered the wave modes that really produce shocks in a process of taking and giving between shock and waves.

The second class includes waves that accompany the shock and provide anomalous transport coefficients like anomalous collision frequencies, friction coefficients, heat conductivity and viscosity. These waves are also important for the shock as they contribute to entropy generation and dissipation. However, they are not primary in the sense that they are not shock-forming waves.

Among them there is another group that only carries away energy and information from the shock. These are high-frequency waves, mostly electrostatic in nature, produced by electrons or when electromagnetic they are in the free-space radiation modes. In the latter case they carry the information from remote objects as radiation in various modes, radio or x-ray to Earth, informing of the existence of a shock. In interplanetary space it is only radio waves which fall into this group as the radiation measure of the heliospheric shocks is too small to map them into x-rays.

Here we restrict mostly to low frequency EM waves in warm plasma, $\omega \leq \omega_{ci}$, while only mention the high frequency EM waves in the end. Such waves are excited by plasma streams or kinetic anisotropies in one or the other way. A simple summary is given in Table 21.3.

Table 21.3: Types of instabilities and waves related to shocks

Mode	Wave Type	Handedness	Other Properties
Firehose	Alfvén	left	Parallel prefered anisotropy
Ion-Ion Beam	Fast	right	Cool beam
Ion cyclotron	Alfvén	left	Warm beam
KAW	Alfvén	right	Electron, parallel electric field
Whistler	Alfvén	right	Hall term, electron

21.6.1 Ion Instabilities $\omega \leq \omega_{ci}$

21.6.1.1 Firehose mode

The simplest instability known which distorts the magnetic field by exciting Alfvén waves that propagate along the magnetic field is the firehose mode. The wave excited are ordinary Alfvén waves, however, and are not suited for shock formation.

When the ion beam is fast and cold it does not go into resonance because its velocity is too high. In this case all ions participate in a nonresonant instability which in fact is a thermal firehose mode where the ion beam has sufficient energy to shake the field line. This mode propagates antiparallel to the ion beam, has small phase speed and negative helicity. This mode has large growth rate for large n_b/n_e and v_b/v_A simply because then there are many beam ions and the centrifugal force is large while the beam velocity lies outside any resonant wave speed. This instability becomes stronger when the ion beam is composed of heavier ions as the larger mass of these increases the centrifugal force effect.

21.6.1.2 Kinetic Alfvén waves

KAWs (Section 10.8.4) possess a finite E_{\parallel} which can accelerate electrons; in the other way, electron moving along the magnetic field in the opposite direction become retarded and feed their energy into KAWs.

Normally this is likely to be a minor effect, as the interaction of ions which are reflected from a solitary pulse and move back upstream ahead of the pulse will cause a stronger instability. The reflected ions will represent a beam that is moving against the initial plasma inflow which by itself is another ion beam neutralised by the comoving electrons. The free energy presented in the two counter-streaming beams leads to various instabilities as viewed by Gary (1993).

21.6.1.3 Kinetic growth rate

At low frequencies it suffices for our purposes of understanding shock physics to deal with a three-component plasma consisting of two ion species and one neutralising electron component which we assume to follow a Maxwellian velocity distribution. Moreover, we assume that the drifting ion components are Maxwellians as well. In conformity with the above remarks on a resonant instability we assume that the dominant ion component has large density $n_i \gg n_b$, and the second component represents a weak fast beam of density n_b propagating on the ion-electron background with velocity $v_b \gg v_i \approx 0$. Following Gary (1993) it is convenient to distinguish the three regimes:

1. cool beams ($0 < v < v_b$)
2. warm beams ($v \sim v_b$)
3. hot beams ($v \gg v_b$)

Figure 21.36 shows the beam configurations for these three cases and the location of the wave resonances respectively the position of the unstable frequencies.

21.6.1.4 Cold Ion Beam: Right-Hand Instability

Assume that the ion beam is thermally isotropic and cool, i.e. its velocity relative to the bulk plasma is faster than its thermal speed. In this case a right-handed resonant instability occurs. In the absence of a beam $v_b = 0$ the parallel mode is a right-circularly polarised magnetosonic wave propagating on the lowest frequency whistler dispersion branch with $\omega \approx k_{\parallel}v_A$. In presence of a drift this wave becomes unstable, and the fastest growing frequency is at frequency $\omega \simeq k_{\parallel}v_b - \omega_{ci}$. This mode propagates parallel to the beam, because $\omega > 0$, $k_{\parallel} > 0$, and $v_b > 0$. The numerical solution of this instability for densities $0.01 \leq n_b/n_i \leq 0.1$ at the wave-number k_{\parallel} of fastest growth rate identifies a growth rate of the order of the wave frequency $\gamma \sim \omega$ and

$$\gamma_m \simeq \omega_{ci} \left(\frac{n_b}{n_e} \right)^{1/3}$$

for the maximum growth rate γ_m , where $n_e = n_i + n_b$ is the total density from quasi-neutrality. This instability drives waves propagating together with the beam in the direction of the ion beam on the plasma background which has been assumed at rest. If applied for instance, to shock reflected ions then for 2% reflected ions the maximum growth rate is $\gamma_m \sim 0.2\omega_{ci}$, and $v_b \sim 1.2\omega_{ci}/k_{\parallel}$, $k_{\parallel} \sim 0.2\omega_{ci}/v_A$ which gives $v_b \sim 6v_A$. In the solar wind the Alfvén velocity is about $v_A \approx 30$ km/s. Hence the velocity difference between shock reflected ions and solar wind along the magnetic field should be roughly ~ 180 km/s.²⁴ The thermal velocity of the ion beam must thus be substantially less than this value, corresponding to a thermal beam energy less than $T_b \ll 100$ eV which in the solar wind, for instance, is satisfied near the tangential

²⁴If the typical Vlasior simulation values are used, $B = 5$ nT, $n = 10^6/\text{cc}$, $v_A = 109$ km/s, $\Delta v \sim 650$ km/s.

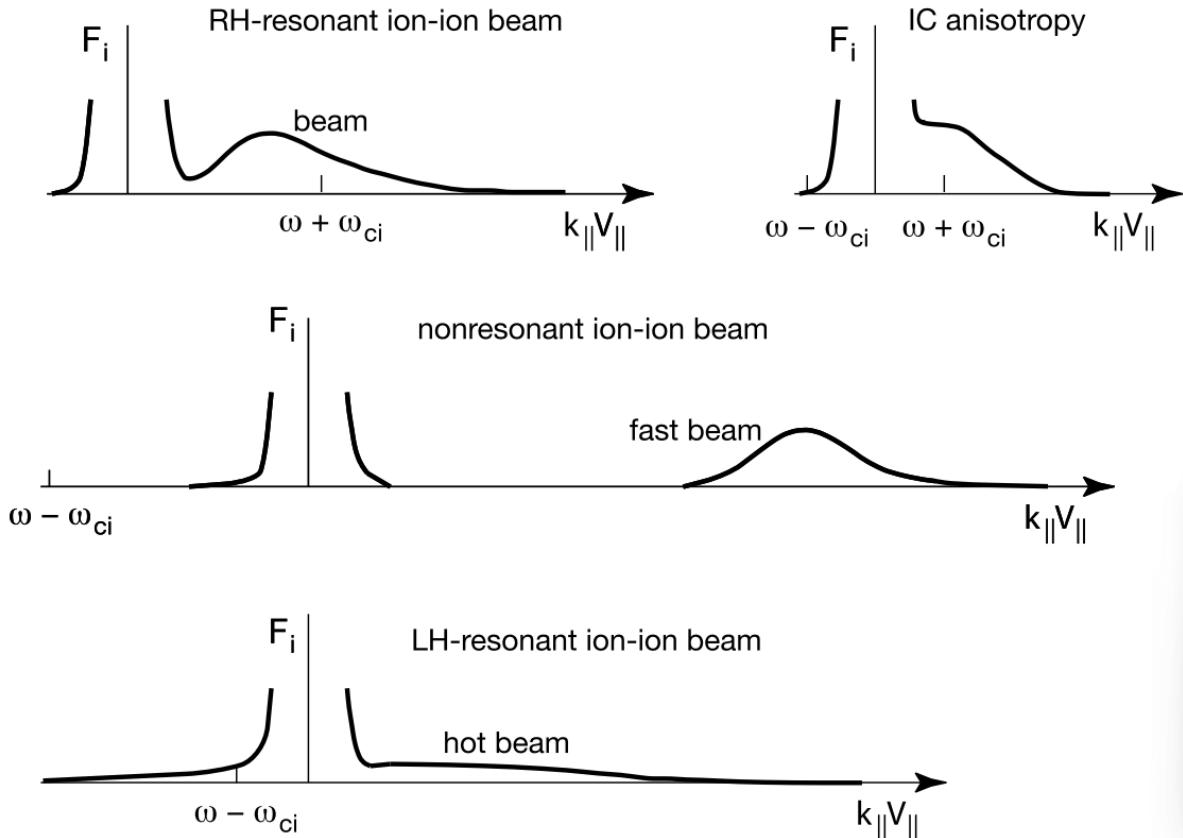


Figure 21.36: The three cases of ion beam-plasma interaction and the location of the unstable frequencies. Shown is the parallel (reduced) distribution function $F_i(k_{\parallel}v_{\parallel})$, where for simplicity the (constant) parallel wavenumber k_{\parallel} has been included into the argument. Right handed resonant modes (RH) are excited by a cool not too fast beam. When the beam is too fast the interaction becomes nonresonant. When the beam is hot, a resonant left hand mode (LH) is excited. In addition the effect of temperature anisotropy is shown when a plateau forms on the distribution function (after Gary 1993).

field line. The solar wind travels at 300–1200 km/s. Complete reflection should produce beam speeds twice these values.²⁵

The cyclotron resonance condition associated with the generated fast magnetosonic mode is

$$\omega = v_b k_{\parallel} - \omega_{ci} \quad (21.32)$$

where v_b is the beam velocity and ω_{ci} the ion gyrofrequency. It can be approximated as $\omega = v_A k_{\parallel}$.

21.6.1.5 Warm Ion Beam: Left-Hand Instability

When the temperature of the ion beam increases and the background ions remain to be cold, then beam ions appear on the negative velocity side of the bulk ion distribution and go into resonance there with the left-hand polarised ion-Alfvén wave. The maximum growth rate is a fraction of the growth rate of the right-hand low frequency whistler mode.²⁶ Nevertheless it can excite the Alfvén-ion cyclotron wave which also propagates parallel to the beam. For this instability the beam velocity must exceed the Alfvén speed $v_b > v_A$.

At oblique propagation both the right and left hand instabilities have smaller growth rates. But interestingly, it has been shown by Goldstein et al. (1985) that the fastest growing modes then appear for oblique \mathbf{k} and harmonics of the ion cyclotron frequency $\omega \sim n\omega_{ci}$, $n = 1, 2, \dots$ ²⁷

21.6.2 Electron Instabilities and Radiation $\omega \sim \omega_{pe}$

Other than ion beam excited instabilities electron-beam instabilities are not involved in direct shock formation (unless the electron beams are highly relativistic which in the entire heliosphere is not the case). The reason is that the frequencies of electron instabilities are high. However, just because of this reason they are crucial in anomalous transport being responsible for anomalous collision frequencies and high frequency field fluctuations. The reason is that the high frequency waves lead to energy loss of the electrons retarding them while for the heavier ions they represent a fluctuating background scattering them. In this way high frequency waves may contribute to the basic dissipation in shocks even though this dissipation for supercritical shocks will not be sufficient to maintain a collisionless shock or even to create a shock under collisionless conditions. This is also easy to understand intuitively, because the waves need time to be created and to reach a substantial amplitude. This time in a fast stream is longer than the time the stream needs to cross the shock. So waves will not accumulate there; rather the

²⁵The statement in (Balogh and Treumann 2013) is hard to follow. Maybe what they tried to argue is that it is the perpendicular portion of the shock that creates these reflected ions and leads to the instability. Later in the book they argued that on a small scale quasi-parallel shocks become perpendicular.

²⁶so literally the Alfvénic branch?

²⁷In the early Vlasiator paper 2014, the simulated RH growth rate is lower than theory, even though in a 1D3V almost parallel configuration.

fast stream will have convected them downstream long before they have reached substantial amplitudes for becoming important in scattering.

When we are going to discuss electromagnetic waves which can be excited by electrons we also must keep in mind that such waves can propagate only when there is an electromagnetic dispersion branch in the plasma under consideration. These electromagnetic branches in (ω, \mathbf{k}) -space are located at frequencies below the electron cyclotron frequency ω_{ce} . The corresponding branch is the whistler mode branch. Electrons will (under conditions prevailing at shocks) in general not be able to excite electromagnetic modes at higher frequencies than ω_{ce} . We have seen before that ion beams have been able to excite whistlers at low frequencies but above the ion-cyclotron frequency. This was possible only because of the presence of the high frequency electron whistler branch as a channel for wave propagation. EM waves excited by electrons propagate on the whistler branch or its low frequency Alfvénic extension, both of which are right-handed. They also excite a variety of electrostatic emissions.

21.6.3 Whistlers

Gary (1993) has investigated the case of whistler excitation by an electron beam. He finds from numerical solution of the full dispersion relation including an electron beam in parallel motion that with increasing beam velocity v_b the real frequency of the unstable whistler decreases, i.e. the unstably excited whistler shifts to lower frequencies on the whistler branch while remaining in the whistler range $\omega_{ci} < \omega < \omega_{ce}$. Both the background electrons and beam electrons contribute resonantly. The most important finding is that the whistler mode for sufficiently large $\beta_i \sim 1$ (which means low magnetic field), n_b/n_e and T_b/T_e has the lowest beam velocity threshold when compared with the electrostatic electron beam instabilities as shown in Figure 21.37. This finding implies that in a relatively high- β plasma a moderately dense electron beam will first excite whistler waves. In the shock environment the conditions for excitation of whistlers should thus be favourable whenever an electron beam propagates across the plasma along the relatively weak magnetic field. The electrons in resonance satisfy $v_{\parallel} = (\omega - \omega_{ce})/k_{\parallel}$ and, because $\omega \ll \omega_{ce}$ the resonant electrons move in the direction opposite to the beam. Enhancing the beam temperature increases the number of resonant electrons thus feeding the instability.

On the other hand, increasing the beam speed shifts the particles out of resonance and decreases the instability. Hence for a given beam temperature the whistler instability has a maximum growth rate a few times the ion cyclotron frequency.

21.7 Shock Particle Reflection

The process of particle reflection from a shock wave is one of the most important processes in the entire physics of collisionless shocks. However, the mechanism of particle reflection has not yet been fully illuminated.

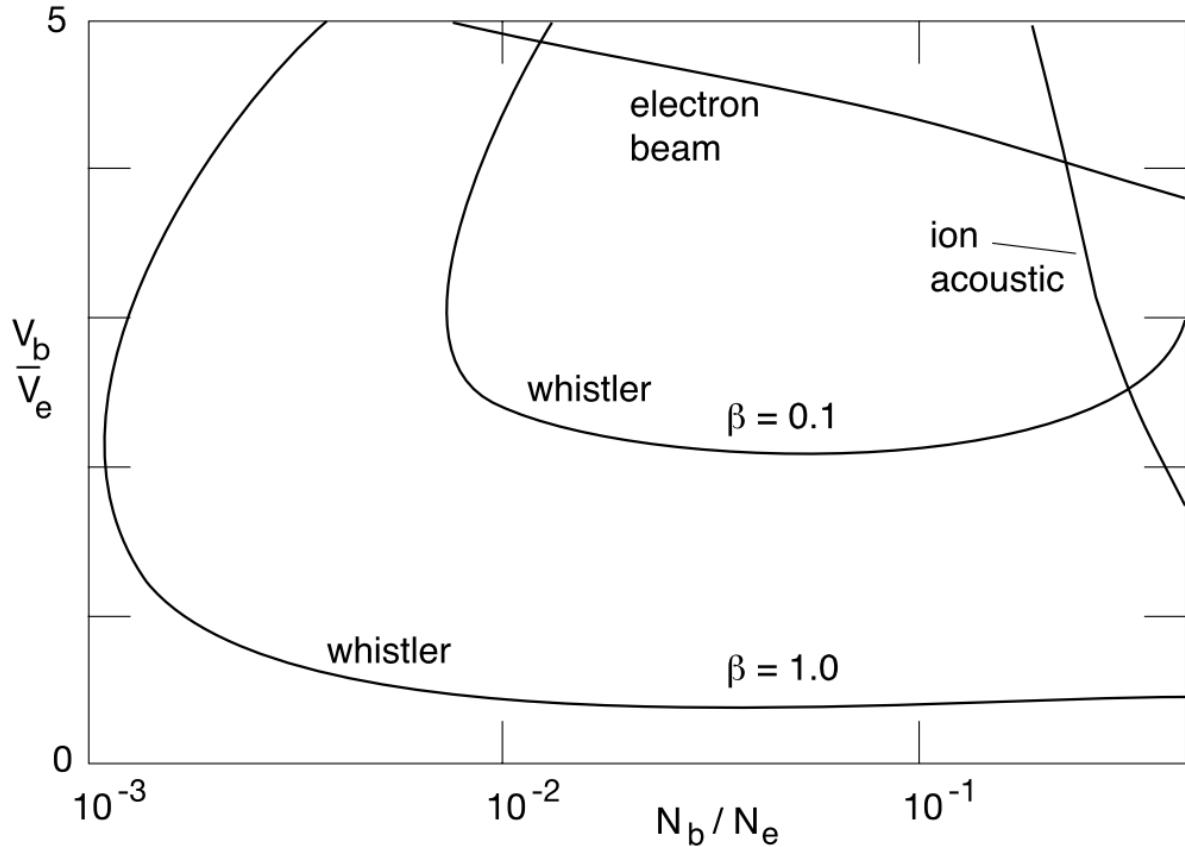


Figure 21.37: The regions of instability of the electron beam excited whistler mode in density and beam velocity space for two different β compared to the ion acoustic and electron beam modes. Instability is above the curves. The whistler instability has the lowest threshold in this parameter range (after Gary 1993).

Particle reflection is caused by a combination of electrostatic and magnetic forces and is nearly specular in nature. It is required in supercritical shocks as it is the only process that can compensate for the incapability of dissipative processes inside the shock ramp to digest the fast inflow of momentum and energy into the shock. Shock particle reflection is not dissipative by itself even though in a fluid picture which deals with moments of the distribution function it can be interpreted as kind of an ion viscosity, i.e. it generates an anomalous viscosity coefficient which appears as a factor in front of the second derivative of the ion velocity in the ionic equation of motion. As such it also appears in the ion heat-transport equation. The kinematic ion viscosity can be expressed as

$$\mu_{\text{vis}} = m_i n \nu_i \lambda_{\text{mfp}} \simeq P_i / 2\omega_{ci}$$

through the ion pressure P_i and the ion-cyclotron frequency ω_{ci} when replacing the mean free path through the ion gyro-radius. In this sense shock particle reflection constitutes by itself a very efficient non-dissipative dissipation mechanism. However, its direct dissipative action is to produce real dissipation as far as possible upstream of the shock in order to dissipate as much energy of motion as remains to be in excess after formation of a shock ramp, dissipation inside the ramp, and reflection of ion back upstream. The shock does this by inhibiting a substantial fraction of inflow ions to pass across the shock from upstream into the downstream region. It is sending these ions back into the upstream region where they cause a violently unstable upstream ion beam-plasma configuration which excites a large amplitude turbulent wave spectrum that scatters the uninformed plasma inflow, heats it and retards it down to the Mach number range that can be digested by the shock. In this way the collisionless shock generates a shock transition region that extends far upstream with the shock ramp degrading to the role of playing a subshock at the location where the ultimate decrease of the Mach number from upstream to downstream takes place.

Shock reflection has another important effect on the shock as the momentum transfer from the reflected particle component to the shock retards the shock in the region of reflection thereby decreasing the effective Mach number of the shock.

21.7.1 Specular Reflection

Specular reflection of ions from a shock front, i.e. reversal of velocity normal component at the shock, is the simplest case to be imagined. It requires that the ions experience the shock ramp as an impenetrable wall. This can be the case when the shock itself contains a positive reflecting electric potential which builds up in front of the approaching ion. Generation of this electric potential is not clarified yet. In a very naive approach one assumes that in flowing magnetised plasma a potential wall is created as the consequence of charge separation between electrons and ions in penetrating the shock ramp. It occurs over a scale typically of the spatial difference between an ion and an electron gyro-radius, because in the ideal case the electrons, when running into the shock ramp, are held temporarily back in the steep magnetic field gradient over this distance while the ions feel the magnetic gradient only over a scale longer than their gyro-radius and thus penetrate deeper into the shock transition.

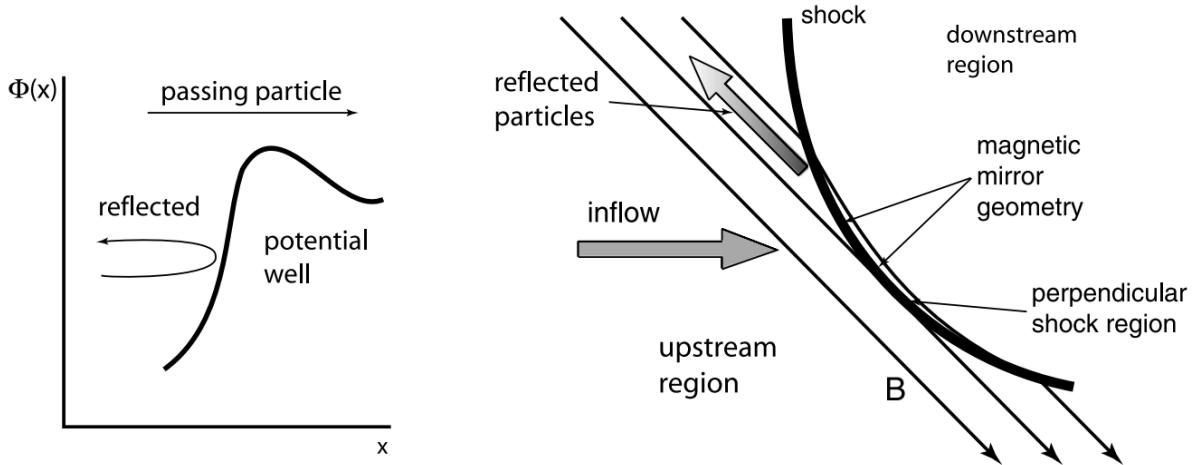


Figure 21.38: The two cases of shock reflection. Left: Reflection from a potential well $\Phi(x)$. Particles of energy higher than the potential energy $e\Phi$ can pass while lower energy particles become reflected. Right: Reflection from the perpendicular shock region at a curved shock wave as the result of magnetic field compression. Particles move toward the shock like in a magnetic mirror bottle, experience the repelling mirror force and for large initial pitch angles are reflected back upstream.

- Reflection from Shock Potential

Due to this simplistic picture the shock ramp should contain a steep increase in the electric potential $\Delta\Phi$ which will reflect any ion which has less kinetic energy $m_i V_N^2 / 2 < e\Delta\Phi$ (Figure 21.38).

- Mirror Reflection

Another simple possibility for particle reflection from a shock ramp in magnetised plasma is mirror reflection. An ion approaching the shock has components $v_{i\parallel}$. Assume a curved shock like Earth's bow shock. Close to its perpendicular part where the upstream magnetic field becomes tangential to the shock the particles approaching the shock with the stream and moving along the magnetic field with their parallel velocities experience a mirror magnetic field configuration that results from the converging magnetic field lines near the perpendicular point (Figure 21.38). Conservation of the magnetic moment $\mu = T_{i\perp}/B$ implies that the particles become heated adiabatically in the increasing field; they also experience a reflecting mirror force $-\mu\nabla_{\parallel}B$ which tries to keep ions away from entering the shock along the magnetic field. Particles will mirror at the perpendicular shock point and return upstream when their pitch angle becomes 90° at this location. (??? Leroy & Mangeney, 1984; Wu, 1984)

Specular reflection from shocks is the extreme case of shock particle reflection. Other mechanisms like turbulent reflection are, however, not well elaborated and must in any case be investigated

with the help of numerical simulations.

21.7.2 Consequences of Shock Reflection

How far the reflected ions return upstream depends on the direction of the magnetic field with respect to the shock, i.e. on the shock normal angle θ_{Bn} . For perpendicular shocks the reflected ions only pass just one gyro-radius back upstream. Seeing the convection electric field $\mathbf{E} = -\mathbf{v}_{\text{flow}} \times \mathbf{B}$ they become accelerated along the shock forming a current, the velocity of which in any case exceeds the inflow velocity (which is zero in the perpendicular direction) and for sufficiently cold ions also the ion acoustic velocity c_{ia} in which case the ion-beam plasma instability will be excited in the shock foot region where the ion current flows. This may generate anomalous collision in the shock foot region. Moreover, since the excited waves accelerate electrons along the magnetic field other secondary instabilities can arise as well.

In quasi-perpendicular and oblique shocks the ions can escape along the magnetic field. In this case an ion two-stream situation arises between the upstream beam and the plasma inflow with the consequence of excitation of a variety of electromagnetic and electrostatic instabilities. In addition, however, an ion-electron two-stream situation is caused between the upstream ions and the inflow electrons which because of the large upstream electron temperatures probably excites mainly ion-acoustic modes but can also lead to Buneman two-stream mode excitation. These modes contribute to turbulence in the upstream foreshock region creating a weakly dissipative state in the foreshock where the plasma inflow becomes informed about the presence of the shock. The electromagnetic low frequency instabilities on the other hand, which are excited in this region, will grow to large amplitude, form localised structures and after being convected by the main flow towards the shock ramp interact with the shock and modify the shock profile or even contribute to shock formation and shock regeneration.

21.8 Shock Particle Acceleration

In the context of cosmic rays that have been observed in the interstellar space, medium energy particles refer to \sim few GeV ions and \sim few MeV electrons. Above these ranges relativistic shocks must be considered. Near the Earth's bow shock the solar wind hydrogen kinetic energy is ~ 1 keV; ~ 10 keV is about the low threshold for energetic ions. Here we limit our discussions first to the non-relativistic case.

Figure 21.39 shows schematically the process of particle acceleration. Based on early estimations by Fermi (1949), a large number of shock crossings and reflections back and forth is required for the particles to reach energetic cosmic ray level. The scattering process is a stochastic process that is assumed to conserve energy; in particular they should not become involved into excitation of instabilities which consume part of their motional energy. The only actual dissipation that is allowed in this process is dissipation of bulk motional energy from where the few accelerated particles extract their energy gain. This dissipation is also attributed to direct

particle loss by either convective transport or the limited size of the acceleration region. Thus this mechanism works until the gyro-radius of the accelerated particle becomes so large that it exceeds the size of the system.

The stochastic process implies that the basic equation that governs the process is a phase space diffusion equation in the form of a Fokker-Planck equation

$$\frac{\partial F(\mathbf{p}, \mathbf{x}, t)}{\partial t} + \mathbf{v} \cdot \nabla F(\mathbf{p}, \mathbf{x}, t) = \frac{\partial}{\partial \mathbf{p}} \cdot \mathbf{D}_{pp} \cdot \frac{\partial F(\mathbf{p}, \mathbf{x}, t)}{\partial \mathbf{p}}, \quad \mathbf{D}_{pp} = \frac{1}{2} \left\langle \frac{\Delta \mathbf{p} \Delta \mathbf{p}}{\Delta t} \right\rangle$$

where $\Delta \mathbf{p}$ is the variation of the particle momentum in the scattering process which happens in the time interval Δt , and the angular brackets indicate ensemble averaging. \mathbf{D}_{pp} is the momentum space diffusion tensor. It is customary to define $\mu = \cos \alpha$ as the cosine of the particle pitch angle α and to understand among $F(p, \mu)$ the gyro-phase averaged distribution function, which depends only on $p = |\mathbf{p}|$ and μ .

The dependence on the gyro-radius imposes a severe limitation on the acceleration mechanism, i.e. the *injection problem*. In order to experience a first scattering, i.e. in order to be admitted to the acceleration process, the particle must initially already possess a gyro-radius much larger than the entire width of the shock transition region. Only when this condition is given, the shock will behave like an infinitesimally thin discontinuity separating two regions of vastly different velocities such that the particle when crossing back and forth over the shock can become aware of the bulk difference in speed and take an energetic advantage of it. This restriction rules out any particles in the core of the upstream inflow distribution from participation in the acceleration process: in order to enter the Fermi shock-acceleration mechanism a particle must be pre-accelerated or pre-heated until its gyro-radius becomes sufficiently large. This condition poses the *injection problem*, where an unresolved seed population of energetic particles are needed for further acceleration, that has not yet been resolved.

Foreshock transients (Section 21.9), especially HFAs and FBs, can accelerate particles and contribute to the primary shock acceleration. These can form secondary shocks which leads to several possible acceleration mechanisms; they can also cause local magnetic reconnection that accelerates particles. The interaction with foreshock transients provides a possible solution to Fermi's *injection problem* and increase the acceleration efficiency of primary shocks.

1. As foreshock transients convect with the upstream flow, particles enclosed within their boundary and the primary shock can experience Fermi acceleration.
2. Secondary shocks have also been observed to accelerate upstream particles on their own through the **shock drift acceleration (SDA)**²⁸ and even to form a secondary foreshock.
3. Secondary shocks can also capture and further energize primary shock-accelerated electrons through betatron acceleration.

²⁸The essence of SDA is that the electric field increases in the middle of the shock! The linked animation shows a perpendicular shock scenario where SDA is not present if there is no increase of the electric field in the shock. However, the question remains: why does the electric field increase?

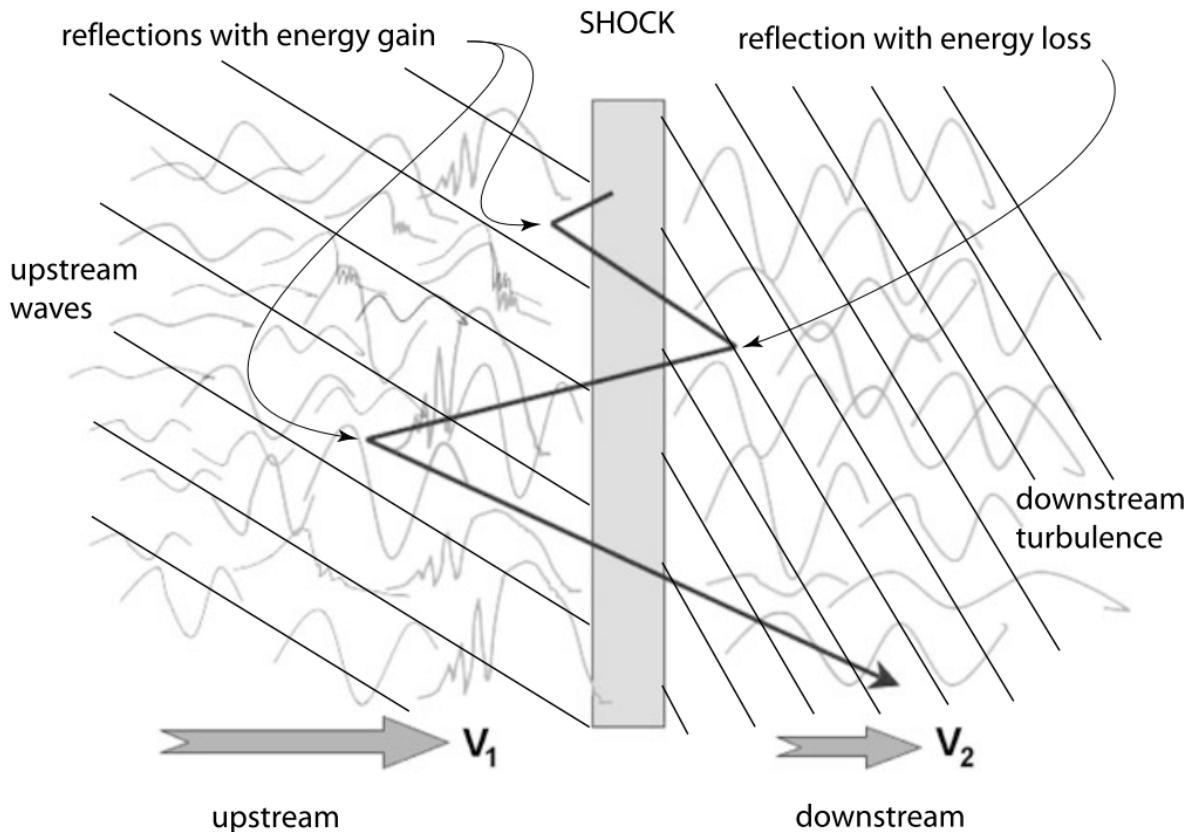


Figure 21.39: Schematic of the acceleration mechanism of a charged particle in reflection at a quasi-parallel ($\theta_{Bn} < 45^\circ$) supercritical shock. The upstream plasma flow (left, $\mathbf{V}_1 \gg \mathbf{V}_2$) contains the various upstream plasma modes: upstream waves, shocklets, whistlers, pulsations. The downstream (right) is turbulent. The energetic particle that is injected at the shock to upstream is reflected in an energy gaining collision with upstream waves, moves downstream where it is reflected in an energy losing collision back upstream. It loses energy because it overtakes the slow waves, but the energy loss is small. Returning to upstream it is scattered a second time again gaining energy. Its initially high energy is successively increased until it escapes from the shock and ends up in free space as an energetic Cosmic Ray particle. The energy gain is on the expense of the upstream flow which is gradually retarded in this interaction. However, the number of energetic particles is small and the energy gain per collision is also small. So the retardation of the upstream flow is much less than the retardation it experiences in the interaction with the shock-reflected low energy particles and the excitation of upstream turbulence.

4. Magnetic reconnection inside foreshock transients contributes to the electron and ion acceleration/heating.

Another problem awakens attention is that how the shocks are modulated by the presence of energetic particles.

In terms of particle acceleration the shock appears as a boundary between two independent regions of different bulk flow parameters which are filled with scattering centres for the particles as sketched in Figure 21.40. Theoretically ((Balogh and Treumann 2013)) any particle which returns from downstream to upstream is accelerated in the upstream flow, even in the absence of any upstream turbulence and scatterings. If the upstream medium is magnetised and is sufficiently extended to host the upstream gyration orbit, pick-up ion energization can happen via the convection electric field $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$ all along their upstream half-gyrocircles. Alternatively, the upstream turbulence can also cause ion energization.

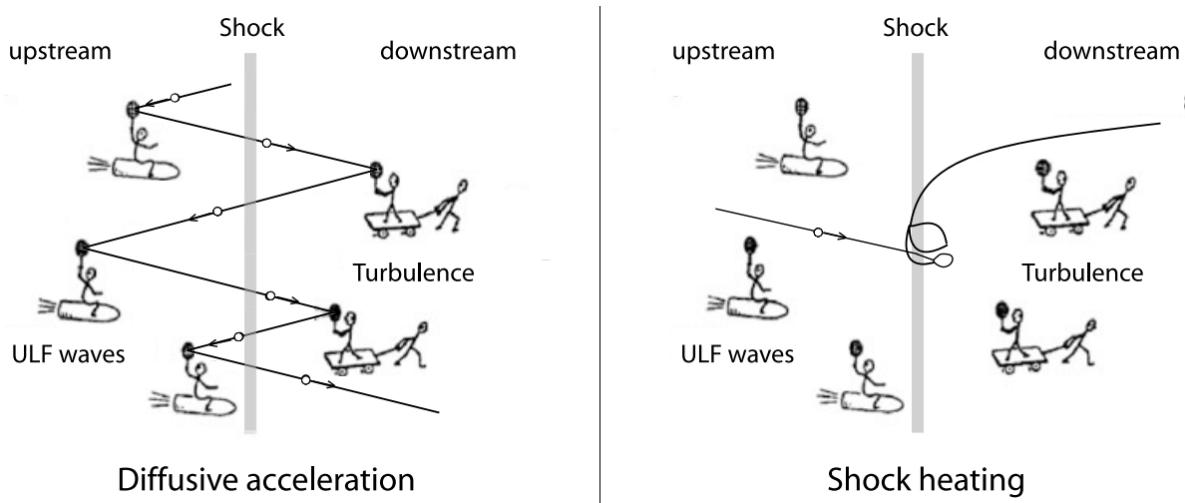


Figure 21.40: Cartoon of the diffusive shock acceleration (left) and shock heating mechanisms [after an sketch by M. Scholer and Hoshino]. In diffusive shock acceleration the particle is scattered around the shock being much faster than the shock. The requirement is the presence of upstream waves and downstream turbulence or waves. In shock heating the particle is a member of the main particle distribution, is trapped for a while at the shock and thereby thermalised and accelerated until leaving the shock.

As a summary, *diffusive shock acceleration (DSA)* (1st-order Fermi acceleration) is the most widely accepted theory for particle acceleration for quasi-parallel shocks. However, a similar theory does not work efficiently at low energies at quasi-perpendicular shocks, where the reflected ions return to the shocks almost immediately due to their gyromotion in the upstream magnetic field. Therefore, *shock drift acceleration (SDA)* and *shock surfing acceleration (SSA)* are

considered to play important roles in ion acceleration or pre-acceleration at quasi-perpendicular shocks.

21.8.1 Electron Acceleration and Heating at Shocks

For strong shocks with a quasiperpendicular geometry, while ion heating is commonly regarded as a result of particle reflection by electric and magnetic forces, the electron heating problem remains controversial. The key controversy concerning electron heating lies in how and when irreversible dissipation occurs. A widely accepted picture of electron heating at the Earth's bow shock describes that the electron phase space is inflated by a quasistatic cross-shock potential accelerating (decelerating) incoming (escaping) electrons, conserving the first adiabatic invariant, and creates an inaccessible region at energies lower than the cross-shock potential energy of electrons [Schwartz 2012]. The inaccessible region is postulated to be filled by wave-particle scattering, leading to thermalization. The picture is based on past observations of heated “flat-topped” distributions and approximately isotropic temperature increase through the shock ramp.

Recent studies concern the breaking of adiabatic conditions (e.g. Li-jen Chen 2018 PRL). MMS observations show cases of bulk acceleration of the solar wind electrons and the nonlinear relaxation resulting in heating at a quasi-perpendicular shock front. The acceleration is achieved by an electric field embedded in a low-frequency whistler-mode wave, and leads to a high parallel current density. It is hypothesized that a collection of localized strong electric fields (E_n spikes, where n is the normal direction to the shock front) may constitute part of the “cross-shock potential”.

21.9 Foreshock Transients

This section provides a list and a very short description of foreshock transients based on observations including hot flow anomalies (HFAs), spontaneous hot flow anomalies (SHFAs), foreshock bubbles (FBs), foreshock cavities, foreshock cavitons, foreshock compressional boundaries, density holes, and Short Large-Amplitude Magnetic structures (SLAMs). Table 21.4 shows a comparison of their characteristics after (Zhang et al. 2022).

21.9.1 Hot Flow Anomaly

HFAs are characterized by a low field strength and low density core with heated plasma and substantial flow deflection with sizes of several R_E . HFAs are typically driven by a solar wind tangential discontinuity (TD) that intersects the bow shock with solar wind convection electric field pointing inward on at least one side of the TD. Such a TD can locally trap foreshock ions leading to the HFA formation while propagating along the bow shock surface. HFAs may

Table 21.4: Comparison of foreshock transient phenomena at the bow shocks

Table 21.5: Basic foreshock transient properties

		Deple-	Com-	Pres-		Associ-		
		tion in	pres-	ence of		ated		
		n and	n and	suprather-	Flow	with		
		B	B	mal	deflec-	an		
				ions	tion	IMF		
						discon-		
						tinuity		
						Dura-		
						tion		
						Scale		
						size		
HFA	Yes	At edges	Yes	Strong	Strong	Yes	$10 - 10^2$ s	$\mathcal{O}(R_E)$
SHFA	Yes	At edges	Yes	Strong	Strong	No	$10 - 10^2$ s	$\mathcal{O}(R_E)$
FB	Yes	Only on the upstream edge	Yes	Strong	Strong	Yes	$10 - 10^2$ s	$\mathcal{O}(R_E)$
Cavity	Yes	At edges	Yes	Weak	Mod- est	Sometimes	$10 - 10^2$ s	$\mathcal{O}(R_E)$
Caviton	Yes	At edges	Yes	Weak	Weak	No	$10 - 60$ s	$\mathcal{O}(R_E)$
FCB	Yes on the turbulent side	At edges	Maybe	Weak	Weak	No	$10 - 10^2$ s	$\mathcal{O}(R_E)$
Density Hole	Yes	At edges	Yes	Sometimes	Sometimes	Yes	$1 - 60$ s	$\mathcal{O}(R_E)$
SLAMS	No	Yes	Maybe	Sometimes	Sometimes	No	$1 - 60$ s	$< 0.5 R_E$

Table 21.6: Foreshock transients generation mechanisms

	Generation mechanism
HFA	Interaction of IMF discontinuities with the bow shock
SHFA	Interaction of foreshock cavitons with the bow shock
FB	Kinetic interactions between suprathermal, backstreaming ions and incident SW plasma with embedded IMF discontinuities that move through and alter the ion foreshock
Cavity	Antisunward moving slabs of magnetic field lines that connect to the bow shock that are sandwiched between broader regions of magnetic field lines that remain unconnected to the bow shock
Caviton	Nonlinear interaction of ULF waves
FCB	Backstreaming ions result in increased pressure within the foreshock region leading to its expansion against the pristine SW and the generation of FCBs
Density Hole	Possibly due to backstreaming particles interacting with the original SW
SLAMS	Nonlinear wave steepening

accelerate particles efficiently through Fermi acceleration, i.e., bouncing between the converging HFA boundary and the bow shock. The observed energetic ions may also have escaped from the outer magnetosphere.

One of the most remarkable properties of HFAs is the strong deflection of the solar wind bulk flow which can be large enough that inside an HFA the flow can actually show a sunward component. A transient region of lower density in the solar wind interacting with the fast shock can cause the disruption of the fast shock and leads to a new shock that actually travels into the upstream direction with plasma behind this new shock having a much smaller momentum density and velocity than the original solar wind.

21.9.2 Spontaneous Hot Flow Anomaly

SHFAs have the same characteristics as HFAs except that they are not associated with any solar wind discontinuities. They form intrinsically in the quasi-parallel regime, likely due to the interaction between foreshock cavitons and the bow shock.

In order to distinguish SHFAs from cavitons in simulations, an additional empirical criterion of SHFA having $\beta > 10$ in at least 60% of the transient region has been used. It is chosen in order to not make assumptions on the level of heating and flow deflection inside the transients. A value of 10 indicates that the transients are dominated by the plasma instead of the magnetic field, and it is significantly above the typical β in the surrounding foreshock ($\beta \sim 1 - 4$).

21.9.3 Foreshock Bubble

When backstreaming foreshock ions interact with a solar wind rotational discontinuity (RD) that does not necessarily intersect the bow shock, FBs form upstream of the RD and convecting anti-sunward with it. Later observations and simulations found that TDs can also drive FBs. FBs are also characterized by a heated, tenuous core with significant flow deflection. Different from HFAs and SHFAs, the expansion of FBs is super-fast-magnetosonic and dominantly in the sunward direction. Because of the sunward super-fast-magnetosonic expansion, a shock forms upstream of the core, and the FB size in the expansion direction can reach 5-10 R_E , larger than typical HFAs and SHFAs. In addition to their significant dynamic pressure perturbations, FBs are also efficient particle accelerators due to the presence of the shock (e.g., shock drift acceleration and Fermi acceleration as the shock converges towards the bow shock).

21.9.4 Foreshock Cavity

Foreshock cavities are characterized by low density, low field strength core regions with high density, high field strength compressional boundaries on two sides. But different from HFAs, the flow deflection inside foreshock cavities is rather weak and plasma heating is not significant. When slabs of magnetic field lines connected to the bow shock are bounded by broader regions

of magnetic field lines that remain unconnected to the bow shock, only the slabs are filled with energized particles reflected from the bow shock. The presence of foreshock particles enhanced the thermal pressure, causing an expansion on two sides. Such an expansion decreases the plasma density and magnetic field strength inside the slabs and increases the density and field strength at two boundaries, i.e., a foreshock cavity forms.

Note that this is somehow different from another cavity known as the magnetic cavity, which is often interpreted as a magnetic mirror structure where a decrease in magnetic field at the center corresponds to an increase of particles at the center [J. Liu+ 2020].

21.9.5 Foreshock Caviton

Foreshock cavitons are also characterized by a core region with low density and field strength bounded by two boundaries with high density and field strength, without clear heating and flow deflection. Their sizes are about one R_E . They form due to the nonlinear evolution of two types of waves: the parallel propagating right- or left-hand polarized waves and the obliquely propagating linearly polarized fast magnetosonic waves. Thus, foreshock cavitons are embedded in foreshock ULF waves, whereas foreshock cavities are isolated due to their different formation mechanisms.

21.9.6 Foreshock Compressional Boundary

FCBs have enhanced density and field strength. They occur at the boundary between the foreshock and the pristine solar wind. Because of the high thermal pressure due to the presence of foreshock ions, the foreshock region expands into the ambient pristine solar wind, leading to the formation of an FCB. FCBs are sometimes associated with local density and field strength depletion on their foreshock side. FCBs can form under either steady or nonsteady IMF conditions.

21.9.7 SLAMS and Shocklets

Short Large-Amplitude Magnetic Structures (SLAMS) are magnetic pulsations with amplitudes at least two times the ambient magnetic field strength. SLAMS have typical spatial scales up to many ion gyroradii (where the thermal ion gyroradius is typically 160 km in the solar wind) and grow rapidly with time scales of seconds. Shocklets are also magnetic structures (nonlinearly steepened magnetosonic waves), but differ from SLAMS in terms of amplitude, spatial scale, growth rate, and propagation angle.

21.9.8 Density Hole

Density holes are characterized by similarly shaped magnetic holes with enhanced density and field strength at one or both edges. The definition of density holes overlaps with HFAs, SHFAs, FBs, foreshock cavities, and foreshock cavitons, but in a broader sense. Lu et al. (2022) showed statistically that 66% of 411 density holes cannot be categorized by any of these foreshock transient types. Therefore, it is necessary to make density holes a separate category. A better definition of density holes is needed to definitely distinguish them from other foreshock transient types, which requires further studies. The formation could be due to the interaction between backstreaming particles and the original solar wind (Parks et al. 2006).

21.10 Location of Shocks

In the observation comparison paper (Slavin and Holzer 1981) for quasi-perpendicular shocks, they concluded that the variations in shock stand-off distance and shape are ordered by the sonic Mach number M_s and not other Mach numbers involve magnetic field. In other words, they think the bow shock is a gasdynamic structure.

However, even in neutral fluid theory, the determination of shock location as well as shape is still an ongoing research. Imagine the simplest scenario where there is a static ball in the air with infinite mass. Assuming purely homogenous air with known density, velocity and pressure in the upstream, can you tell me the exact location of shock stand-off distance with pen and paper?

On top of that, the introduction of EM field complicates the story. Especially in the case of a parallel shock, the plasmas get “shocked” both upstream and downstream, and the stand-off distance of the shock may not be a single point theoretically. In some sense, normal magnetic field to the boundary “thickens” the shock front.

21.11 Earth Bow Shock

Using data from the AMPTE/IRM spacecraft, (Hill et al. 1995) have shown that the double adiabatic equations do not hold in the magnetosheath. Moreover, the thermal behaviour of the magnetosheath is studied by (Phan et al. 1996) using WIND spacecraft data. They report that most parts of the magnetosheath are marginally mirror unstable: electron observations showed $T_{e\perp}/T_{e\parallel} \sim 1.3$ in the magnetosheath.

22 Magnetosphere

The big picture when considering the interaction of the solar wind and the magnetosphere is as follows:

1. Is there something that can penetrate from the solar wind into the magnetosphere?
2. Is there something that can be triggered from the interaction?
3. Is a physical process internally or externally driven?

22.1 Earth's Magnetic Environment

In the first approximation the Earth's magnetic field is that of a magnetic dipole (Section 3.8). The dipole axis is tilted 11° from the direction of the Earth's rotation axis. The current circuit giving rise to the magnetic field is located in the liquid core about 1200–3400 km from the center of the planet. The current system is asymmetric displacing the dipole moment from the center, which together with inhomogeneous distribution of magnetic matter above the core gives rise to large deviations from the dipole field on the surface. The pure dipole field on the surface would be 30 μT at the dipole equator and 60 μT at the poles. However, the actual surface field exceeds 66 μT in the region between Australia and Antarctica and is weakest, about 22 μT, in a region called *South Atlantic Anomaly* (SAA).

22.1.1 Basic Structures

In the frame of reference of the Earth the solar wind is supermagnetosonic, exceeding the local magnetosonic speed $v_{ms} = \sqrt{v_s^2 + v_A^2}$, where v_s is the sound speed, v_A is the Alfvén speed. Because fluid-scale perturbations cannot propagate faster than v_{ms} , this leads to a formation of a collisionless shock front, called the *bow shock*, upstream of the magnetosphere. Under typical solar wind conditions the apex of the shock in the solar direction is about $3 R_E$ upstream of the magnetopause. The shock converts a considerable fraction of solar wind kinetic energy to heat and electromagnetic energy. The irregular shocked flow region between the bow shock and the magnetopause is called the *magnetosheath*.

22.1.2 The Dipole Field

The dipole field is an idealization where the source current is assumed to be confined into a point at the origin. The source of planetary and stellar dipoles is a finite, actually a large, current system within the celestial body. Such fields, including the Terrestrial magnetic field, are customarily represented as a multipole expansion: dipole, quadrupole, octupole, etc. When moving away from the source, the higher multipoles vanish faster than the dipole making the dipole field a good starting point to consider the motion of charged particles. In the dipole field charged particles behave adiabatically as long as their gyro radii are smaller than the gradient scale length of the field (Section 7.8) and their orbits are not disturbed by collisions or time-varying electromagnetic field.

For the geomagnetic field it is customary to define the spherical coordinates in a special way. The dipole moment (\mathbf{m}_E) is in the origin and points approximately toward geographic south, tilted 11° as mentioned above. Similar to the geographic coordinates the latitude (λ) is zero at the dipole equator and increases toward the north, whereas the latitudes in the southern hemisphere are negative. The longitude (ϕ) increases toward the east from a given reference longitude. In magnetospheric physics the longitude is often given as the *magnetic local time* (MLT). In the dipole approximation MLT is determined by the flare angle between two planes: the dipole meridional plane containing the subsolar point on the Earth's surface, and the dipole meridional plane which contains a given point on the surface, i.e., the local dipole meridian. Magnetic noon (MLT = 12 h) points toward the Sun, midnight (MLT = 24 h) anti-sunward. Magnetic dawn (MLT = 6 h) is approximately in the direction of the Earth's orbit around the Sun.

The SI-unit of m_E is A m^2 . Sometimes it is convenient to replace m_E by $k_0 = \mu_0 m_E / 4\pi$, which is also customarily called dipole moment. The strength of the terrestrial dipole moment varies slowly. A sufficiently accurate approximation is

$$\begin{aligned} m_E &= 8 \times 10^{22} \text{ A m}^2 \\ k_0 &= 8 \times 10^{15} \text{ Wb m} (\text{SI : Wb} = \text{T m}^2) \\ &= 8 \times 10^{25} \text{ G cm}^3 \text{ (Gaussian units, } 1 \text{ G} = 10^{-4} \text{ T)} \\ &= 0.3 \text{ G } R_E^3 \text{ (} R_E \simeq 6378 \text{ km)} \end{aligned}$$

The last expression is convenient in practice because the dipole field on the surface of the Earth (at $1 R_E$) varies in the range 0.3–0.6 G.

Outside its source, the dipole field is a curl-free potential field $\mathbf{B} = -\nabla\psi$, where the scalar potential is given by

$$\psi = -\mathbf{k}_0 \cdot \nabla \frac{1}{r} = -k_0 \frac{\sin \lambda}{r^2}$$

yielding

$$\mathbf{B} = \frac{1}{r^3} [3(\mathbf{k}_0 \cdot \hat{r})\hat{r} - \mathbf{k}_0]$$

The components of the magnetic field are

$$\begin{aligned}B_r &= -\frac{2k_0}{r^3} \sin \lambda \\B_\lambda &= \frac{k_0}{r^3} \cos \lambda \\B_\phi &= 0\end{aligned}$$

and its magnitude is

$$B = \frac{k_0}{r^3} (1 + 3 \sin^2 \lambda)^{1/2}$$

The equation of a magnetic field line is

$$r = r_0 \cos^2 \lambda$$

where r_0 is the distance where the field line crosses the equator. The length element of the magnetic field line element is

$$ds = (dr^2 + r^2 d\lambda^2)^{1/2} = r_0 \cos \lambda (1 + 3 \sin^2 \lambda)^{1/2} d\lambda$$

This can be integrated in a closed form, yielding the length of the dipole field line S_d as a function of r_0

$$S_d \approx 2.7603r_0 \quad (22.1)$$

The curvature radius $RC = |d^2\mathbf{r}/ds^2|^{-1}$ of the magnetic field is an important parameter for the motion of charged particles. For the dipole field the *radius of curvature* is

$$R_c(\lambda) = \frac{r_0}{3} \cos \lambda \frac{(1 + 3 \sin^2 \lambda)^{3/2}}{2 - \cos^2 \lambda}$$

Any dipole field line is determined by its (constant) longitude ϕ_0 and the distance where the field line crosses the dipole equator. This distance is often given in terms of the *L-parameter*

$$L = r_0/R_E \quad (22.2)$$

For a given L the corresponding field line reaches the surface of the Earth at the (dipole) latitude

$$\lambda_e = \arccos \frac{1}{\sqrt{L}}$$

For example, $L = 2$ (the inner belt) intersects the surface at $\lambda_e = 45^\circ$, $L = 4$ (the heart of the outer belt) at $\lambda_e = 60^\circ$ and $L = 6.6$ (the geostationary orbit) at $\lambda_e = 67.1^\circ$.

The dipole field line length in Equation 22.1 was calculated from the dipole itself. Now we can calculate also the dipole field line length from a point on the surface to the surface on the opposite hemisphere to be

$$S_e \approx (2.7755L - 2.1747)R_E$$

which is a good approximation when $L \gtrsim 2$.

The field magnitude along a given field line as a function of latitude is

$$B(\lambda) = [B_r(\lambda)^2 + B_\lambda(\lambda)^2]^{1/2} = \frac{k_0}{r_0^3} \frac{(1 + 3 \sin^2 \lambda)^{1/2}}{\cos^6 \lambda}$$

For the Earth

$$\frac{k_0}{r_0^3} = \frac{0.3}{L} \text{ G} = \frac{3 \times 10^{-5}}{L^3} \text{ T}$$

The actual geomagnetic field has considerable deviations from the dipolar field because the dipole is not quite in the center of the Earth, the source is not a point, and the electric conductivity of the Earth is not uniform. The geomagnetic field is described by the *International Geomagnetic Reference Field* (IGRF) model, which is regularly updated to reflect the slow secular variations of the field, i.e., changes in timescales of years or longer.

22.1.3 Current Systems

The Earth's magnetosphere is the region where the near-Earth magnetic field controls the motion of charged particles. It is formed by the interaction between the geodipole and the solar wind. The deformation of the field, caused by the variable solar wind pressure, sets up time-dependent magnetospheric current systems that dominate deviations from the dipole field in the outer radiation belt and beyond.

The solar wind plasma cannot easily penetrate to the Earth's magnetic field and the outer magnetosphere is essentially a cavity around which the solar wind flows. The cavity is bounded by a flow discontinuity called the *magnetopause*. The shape and location of the magnetopause is determined by the balance between the solar wind plasma pressure and the magnetospheric magnetic field pressure. The nose, or apex, of the magnetopause is, under average solar wind conditions, at the distance of about $10 R_E$ from the center of the Earth but can be pushed to the vicinity of the geostationary distance ($6.6 R_E$) during periods of large solar wind pressure. In the dayside the dipole field is compressed toward the Earth, whereas in the nightside the field is stretched to form a long *magnetotail*. The deviations from the curl-free dipole field correspond to electric current systems according to Ampère's law $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$.

Figure 22.1 is a sketch of the magnetosphere with the main large-scale magnetospheric current systems. Individual looks at different currents in detail are presented in Figure 22.4, Figure 22.5, and Figure 22.6, with more explanations in this [review](#).

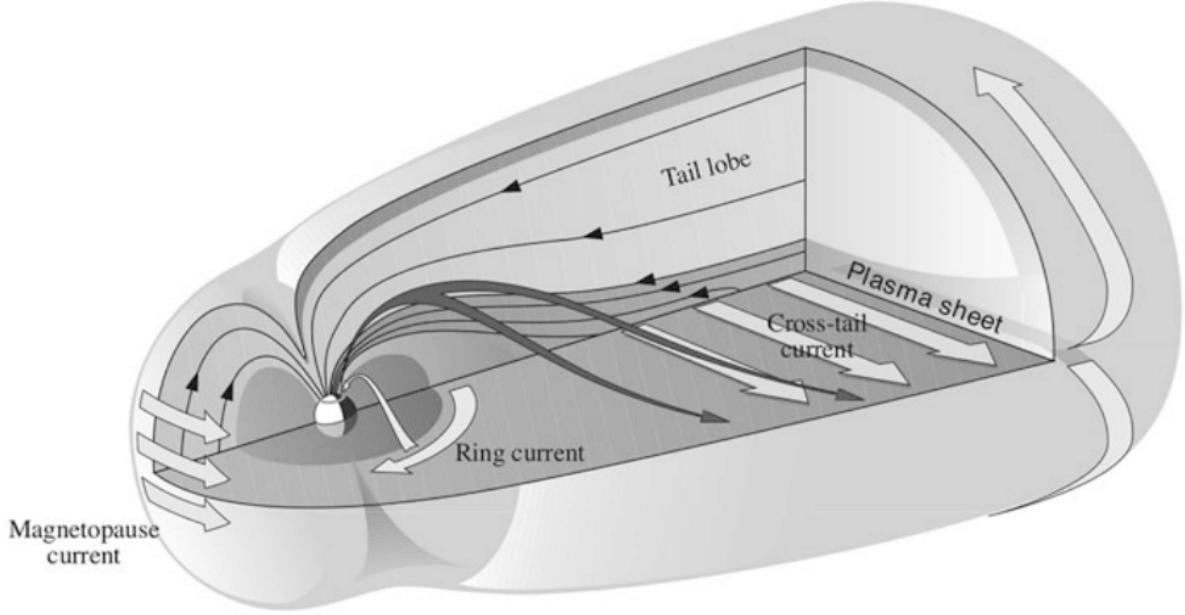


Figure 22.1: The magnetosphere and the large scale magnetospheric current systems.

The current system on the dayside magnetopause shielding the Earth's magnetic field from the solar wind is known as the *Chapman–Ferraro current* (Figure 22.2). In the first approximation the Chapman–Ferraro current density \mathbf{J}_{CF} can be expressed as

$$\mathbf{J}_{\text{CF}} = \frac{\mathbf{B}_{\text{MS}}}{B_{\text{MS}}^2} \times \nabla P_{\text{dyn}} \quad (22.3)$$

where \mathbf{B}_{MS} is the magnetospheric magnetic field and P_{dyn} the dynamic pressure of the solar wind. Because the *interplanetary magnetic field* (IMF) at the Earth's orbit is only a few nanoteslas, the magnetopause current must shield the magnetospheric field to almost zero just outside the current layer. Consequently, the magnetic field immediately inside the magnetopause doubles: about one half comes from the Earth's dipole and the second half from the magnetopause current. In plasma physics this is known as the diamagnetic current caused by the pressure gradient (Equation 8.63, Equation 8.65).

If the plasma pressure is anisotropic, the current density perpendicular to the magnetic field \mathbf{J}_{\perp} in a static system can be written

$$\mathbf{J}_{\perp} = \frac{\mathbf{B}}{B^2} \times \left[\nabla p_{\perp} + (p_{\parallel} - p_{\perp}) \frac{(\mathbf{B} \cdot \nabla) \mathbf{B}}{B^2} \right] \quad (22.4)$$

where p_{\parallel} and p_{\perp} are plasma pressure, parallel and perpendicular to the magnetic field, respectively. Equation 22.4 can be simplified to Equation 22.3 in the isotropic case.

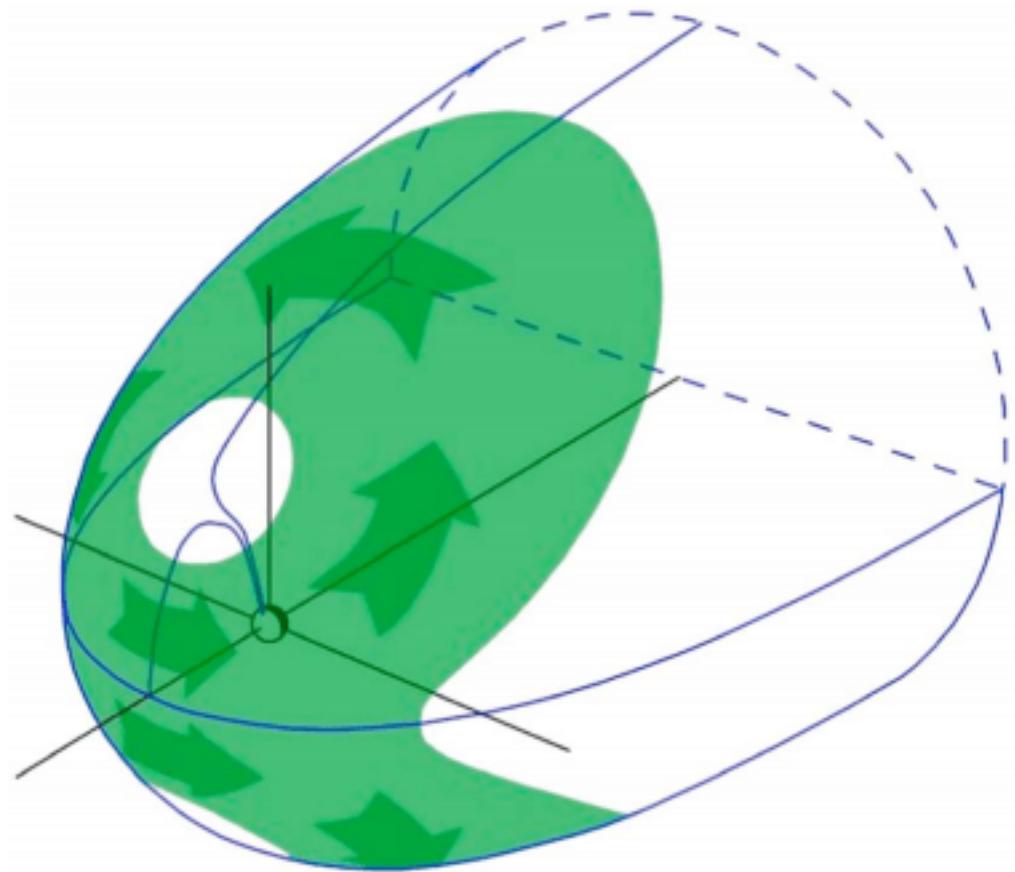


Figure 22.2: The Chapman-Ferraro dayside magnetopause currents are shown as a green-shaded surface on a simple wire diagram of the magnetopause. The Earth is the small sphere at the axis origin, and the Sun is to the lower left.

The Chapman–Ferraro model describes a teardrop-like closed magnetosphere that is compressed in the dayside and stretched in the nightside, but not very far. Since the 1960s spacecraft observations have shown that the nightside magnetosphere, the *magnetotail*, is very long, extending far beyond the orbit of the Moon. This requires a mechanism to transfer energy from the solar wind into the magnetosphere to keep up the current system that sustains the tail-like configuration.

In Figure 22.1, the overwhelming fraction of the magnetospheric volume consists of *tail lobes*, connected magnetically to the polar caps in the ionized upper atmosphere, known as the *ionosphere*. The polar caps are bounded by *auroral ovals*. Consequently, in the northern lobe the magnetic field points toward the Earth, in the southern away from the Earth. To maintain the lobe structure, there must be a current sheet between the lobes where the current points from dawn to dusk. This *cross-tail current* is embedded within the plasma sheet and closes around the tail lobes forming the nightside part of the the *magnetopause current*.

The cusp-like configurations of weak magnetic field above the polar regions known as *polar cusps* do not connect magnetically to magnetic poles, but instead to the southern and northern auroral ovals at noon, because the entire magnetic flux enclosed by the ovals is connected to the tail lobes. Tailward of the cusps the Chapman–Ferraro current and the tail magnetopause current smoothly merge with each other. Figure 22.1 also illustrates the westward flowing *ring current* (RC) and the magnetic *field-aligned currents* (FAC) connecting the magnetospheric currents to the horizontal ionospheric currents in auroral regions at an altitude of about 100 km. FACs are mainly carried by electrons. They were first suggested to explain the variations of magnetic field measured on the ground in the polar regions.

Based on the magnetosphere-ionosphere current continuity, the field-aligned current J_{\parallel} (positive if flowing into the ionosphere) is related to the magnetic field and plasma pressure in the magnetosphere (Grad, 1964; Tverskoy, 1982; Vasyliunas, 1970) as (???)

$$J_{\parallel} = \frac{B_i}{B_e} \hat{b} \cdot (\nabla W \times \nabla P) \quad (22.5)$$

where $W = \int \frac{ds}{B}$ is the magnetic flux tube volume, ds is the element of magnetic field line length, B is the magnetic field along the field line and the integration is taken between the two conjugate points, P is the plasma pressure, B_i and B_e are the magnetic fields in the ionosphere and equatorial plane, respectively. The gradients are evaluated in the equatorial plane. The formation of a FAC requires the existence of a hot plasma pressure gradient along the isosurfaces of the magnetic flux tube volume W , azimuthal plasma pressure gradient. If the azimuthal gradient is directed outward indicating that the pressure peaks are not around midnight but close to dawn and dusk, Region 1 field-aligned current can be generated in the plasma sheet, as shown in Figure 22.5.

The magnetospheric current systems can have significant temporal variations, which makes the mathematical description of the magnetic field complicated. A common approach is to apply

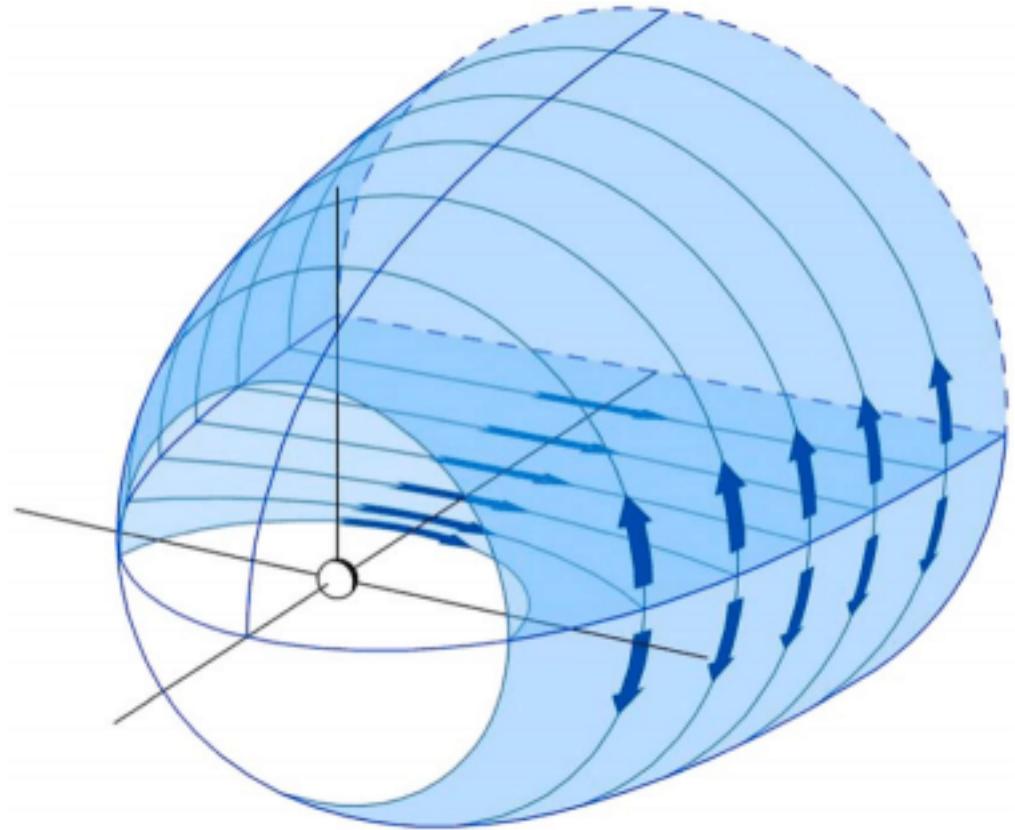


Figure 22.3: The tail current with closure via return current on magnetopause shown in light blue on the wire diagram of the magnetopause, as in Figure 22.2.

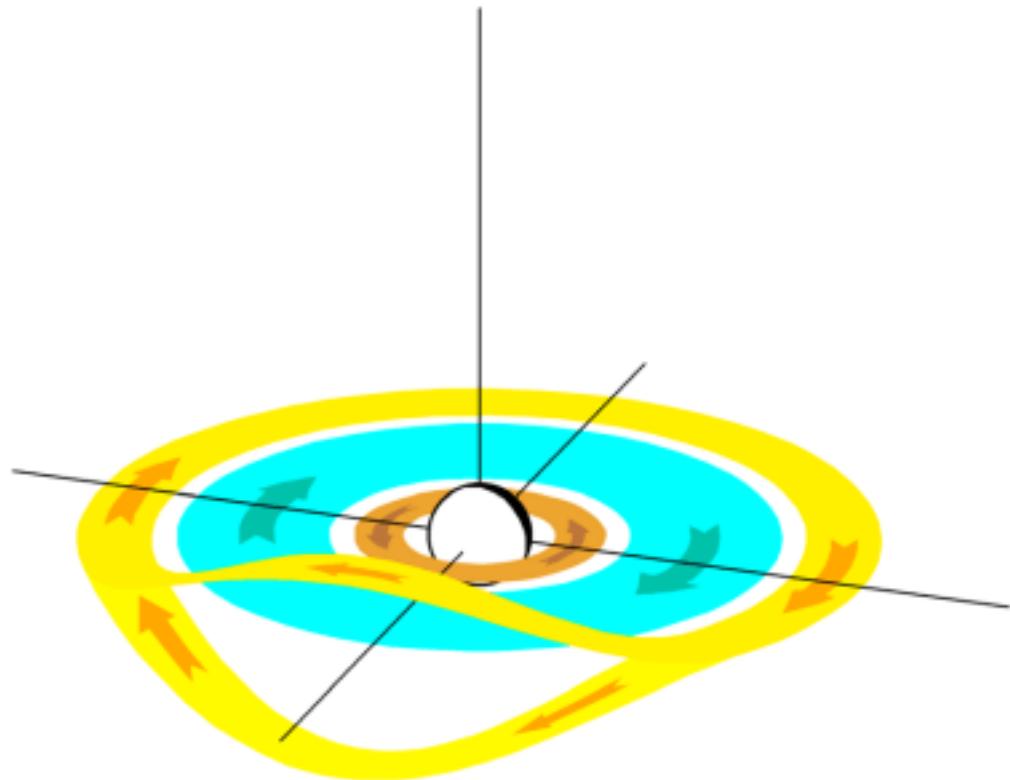


Figure 22.4: The symmetric ring current (eastward and westward, in brown and blue, respectively) including the cut ring currents on the dayside (in yellow). The viewing perspective is the same as in Figure 22.2 but now zoomed in closer to the Earth.

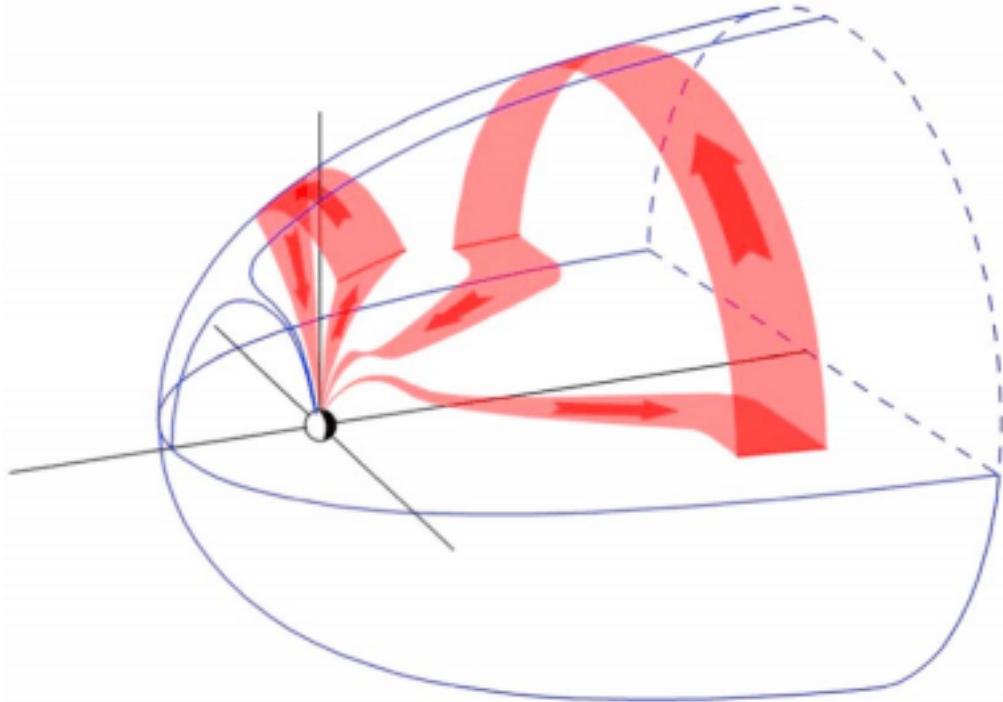


Figure 22.5: Region 1 field-aligned currents, shown as the red bands, including the two possible closure paths: directly to the magnetopause and via the far tail plasma sheet.

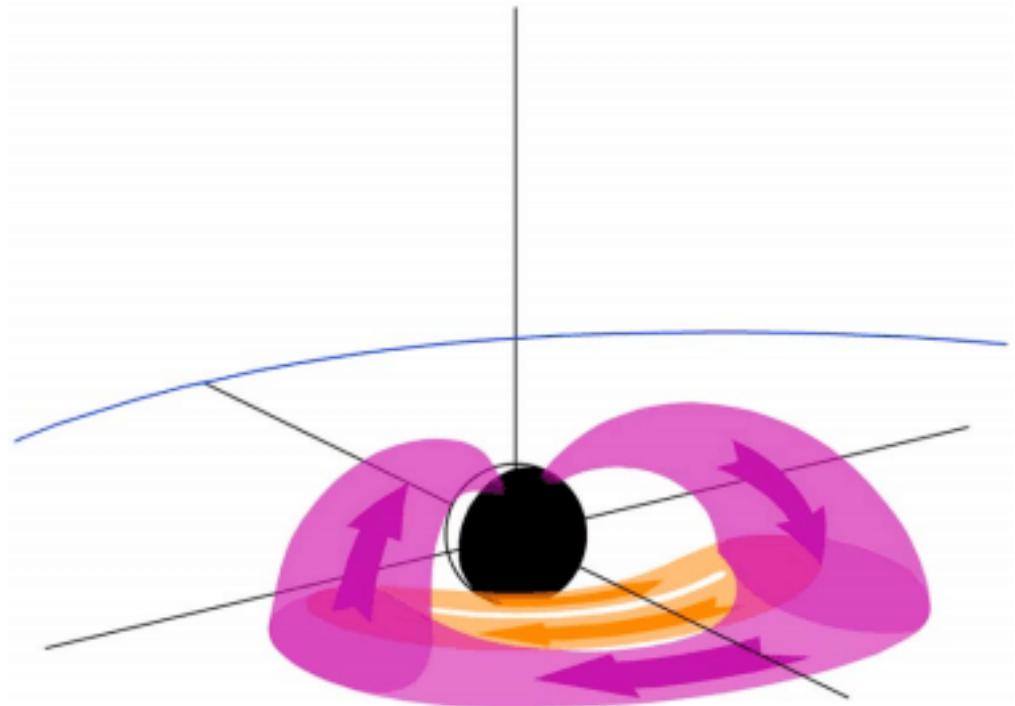


Figure 22.6: Region 2 field-aligned currents and partial ring current, shown in purple and the banana current, shown in orange. The view here is shifted to be from the evening sector.

some of the various models developed by Nikolai Tsyganenko (Tsyganenko 2013).¹

For illustrative purposes simpler models are sometimes useful. For example, the early time-independent model of Mead (1964) reduces in the magnetic equatorial (r, ϕ) plane to

$$B(r, \phi) = B_E \left(\frac{R_E}{r} \right)^3 \left[1 + \frac{b_1}{B_E} \left(\frac{r}{R_E} \right)^3 - \frac{b_2}{B_E} \left(\frac{r}{R_E} \right)^4 \cos \phi \right]$$

where B_E is the equatorial dipole field on the surface of the Earth (approximately $30.4 \mu\text{T} = 30,400 \text{nT}$) and ϕ is the longitude east of midnight. The $\cos \phi$ term describes the azimuthal asymmetry due to the dayside compression and nightside stretching of the field. The coefficients b_1 and b_2 depend on the distance of the subsolar point of the magnetopause R_s (in units of R_E), which, in turn, depends on the upstream solar wind pressure

$$\begin{aligned} b_1 &= 25 \left(\frac{10}{R_s} \right)^3 \text{nT} \\ b_2 &= 2.1 \left(\frac{10}{R_s} \right)^4 \text{nT} \end{aligned}$$

This model is fairly accurate during quiet and moderately disturbed times at geocentric distances $1.5\text{--}7 R_E$.

22.1.4 Geomagnetic Activity Indices

The intensity and variations of magnetospheric and ionospheric current systems are traditionally described in terms of *geomagnetic activity indices*. The indices are calculated from ground-based magnetometer measurements. As different indices describe different features of magnetospheric currents, there is no one-to-one correspondence between them. The choice of a particular index depends on physical processes being investigated.

The most widely used indices for global storm levels are Dst , Kp , and AE .

22.1.4.1 Dst

The Dst index aims at measuring the intensity of the ring current. It is calculated once an hour as a weighted average of the deviation from the quiet level of the horizontal magnetic field component (H) measured at four low-latitude stations distributed around the globe. *Geomagnetic storms* are defined as periods of strongly negative Dst index, signalling enhanced westward the ring current. The more negative the Dst index is, the stronger is the storm. There is no canonical lower threshold for the magnetic perturbation beyond which the state of

¹Tsyganenko models are available at [Community Coordinated Modeling Center](#).

the magnetosphere is to be called a storm and identification of weak storms is often ambiguous. As an empirical classification, we call storms with Dst from -50 to -100 nT moderate, from -100 to -200 nT intense, and those with $\text{Dst} < -200$ nT big. A similar 1-min index derived from a partly different set of six low-latitude stations (*SYM-H*) is also in use.

Dst has contributions from all currents in addition to the ring current. These include the magnetopause and cross-tail currents, as well as induced currents in the ground due to rapid temporal changes of ionospheric currents. Large solar wind pressure pushes the magnetopause closer to the Earth forcing the magnetopause current to increase to be able to shield a locally stronger geomagnetic field from the solar wind. The effect is strongest on the dayside where the magnetopause current flows in the direction opposite to the ring current. The *pressure corrected Dst* index can be defined as

$$\text{Dst}^* = \text{Dst} - b\sqrt{P_{\text{dyn}}} + c \quad (22.6)$$

where P_{dyn} is the solar wind dynamic pressure and b and c are empirical parameters, whose exact values depend on the used statistical analysis methods, e.g., $b = 7.26 \text{ nT nPa}^{-1/2}$ and $c = 11 \text{ nT}$ as determined by O'Brien and McPherron (2000).

The contribution from the dawn-to-dusk directed tail current to the Dst index is more difficult to estimate. During strong activity the cross-tail current intensifies and moves closer to the Earth, enhancing the nightside contribution to Dst. The estimates of this effect on Dst vary in the range 25–50%. Furthermore, fast temporal changes in the ionospheric currents induce strong localized currents in the ground, which may contribute up to 25% to the Dst index.

22.1.4.2 K_p

Another widely used index is the planetary K index, K_p . Each magnetic observatory has its own K index and K_p is an average of K indices from 13 mid-latitude stations. It is a quasi-logarithmic range index expressed in a scale of one-thirds: 0, 0+, 1-, 1, 1+,..., 8+, 9-, 9. K_p is based on mid-latitude observations and thus more sensitive to high-latitude auroral current systems and to substorm activity than the Dst index. K_p is a 3-h index and does not reflect rapid changes in the magnetospheric currents.

22.1.4.3 AE

The fastest variations in the current systems take place at auroral latitudes. To describe the strength of the auroral currents the *auroral electrojet indices* (AE) are commonly used. The standard AE index is calculated from 11 or 12 magnetometer stations located under the average auroral oval in the northern hemisphere. It is derived from the magnetic north component at each station by determining the envelope of the largest negative deviation from the quiet time background, called the AL index, and the largest positive deviation, called the AU index. The AE index itself is $\text{AE} = \text{AU} - \text{AL}$ (all in nT). Thus AL is the measure of the strongest

westward current in the auroral oval, AU is the measure of the strongest eastward current, and AE characterizes the total electrojet activity. AE, AU, AL are typically given with 1-min time resolution.

As the auroral electrojets flow at the altitude of about 100 km, their magnetic deviations on the ground are much larger than those caused by the ring current. For example, during typical substorm activations AE is in the range 200–400 nT and can during strong storms exceed 2000 nT, whereas the equatorial Dst perturbations exceed -200 nT only during the strongest storms.

22.2 Particles

Typical densities of the unperturbed solar wind at 1 AU extend from about 3 cm^{-3} in the fast ($\sim 750\text{ km s}^{-1}$) to about 10 cm^{-3} in the slow ($\sim 350\text{ km s}^{-1}$) solar wind, again with large deviations.

22.2.1 Outer Magnetosphere

The outer magnetosphere can be considered to begin at distances of about $7\text{--}8 R_E$ where the nightside magnetic field becomes increasingly stretched. Table 22.1 summarizes typical plasma parameters in the mid-tail region, at about $X = -20 R_E$ from the Earth. Here X is the Earth-centered coordinate along the Earth–Sun line, positive toward the Sun. The tail lobes are almost empty, particle number densities being of the order of 0.01 cm^{-3} . The central plasma sheet where the cross-tail current is embedded (Figure 22.1) is, in turn, a region of hot high-density plasma. It is surrounded by the plasma sheet boundary layer with density and temperature intermediate to values in the central plasma sheet and tail lobes. The field lines of the boundary layer connect to the poleward edge of the auroral oval. The actual numbers differ considerably from the typical values under changing solar wind conditions and, in particular, during strong magnetospheric disturbances.

Table 22.1: Typical values of plasma parameters in the mid-tail

	Magnetosheath	Tail lobe	Plasma sheet boundary	Central plasma sheet
$n(\text{cm}^{-3})$	8	0.01	0.1	0.3
$T_i(\text{eV})$	150	300	1000	4200
$T_e(\text{eV})$	25	50	150	600
$B(\text{nT})$	15	20	20	10
β	2.5	0.0003	0.1	6

Table 22.1 also includes typical parameters in the magnetosheath at the same X-coordinate. The magnetosheath consists of solar wind plasma that has been compressed and heated by the Earth's bow shock. It has higher density and lower temperature than observed in the outer magnetosphere. Table 22.1 shows that, while the magnetic field magnitude is rather similar in all regions shown, plasma beta (the ratio between the kinetic and magnetic field pressures), is a useful parameter to distinguish between different regions.

22.2.2 Inner Magnetosphere

The inner magnetosphere is the region where the magnetic field is quasi-dipolar. It is populated by different spatially overlapping particle species with different origins and widely different energies: the ring current, the radiation belts and the plasmasphere (Chapter 24). The ring current and radiation belts consist mainly of trapped particles in the quasi-dipolar field drifting due to magnetic field gradient and curvature effects around the Earth, whereas the motion and spatial extent of plasmaspheric plasma is mostly influenced by the corotation and convection electric fields.

The *ring current* arises from the azimuthal drift of energetic charged particles around the Earth; positively charged particles drifting toward the west and electrons toward the east. Basically all drifting particles contribute to the ring current. The drift currents are proportional to the energy density of the particles and the main ring current carriers are positive ions in the energy range 10–200 keV, whose fluxes are much larger than those of the higher-energy radiation belt particles. The ring current flows at geocentric distances $3\text{--}8 R_E$, and peaks at about $3\text{--}4 R_E$. At the earthward edge of the ring current the negative pressure gradient introduces a local eastward diamagnetic current, but the net current remains westward.

During magnetospheric activity the role of the ionosphere as the plasma source of ring current enhances, increasing the relative abundance of oxygen (O^+) and helium (He^+) ions in the magnetosphere. As a result a significant fraction of ring current can at times be carried by oxygen ions of atmospheric origin. The heavy-ion content furthermore modifies the properties of plasma waves in the inner magnetosphere, which has consequences on the wave–particle interactions with the radiation belt electrons.

The plasmasphere is the innermost part of the magnetosphere. It consists of cold ($\sim 1 \text{ eV}$) and dense ($\gtrsim 10^3 \text{ cm}^{-3}$) plasma of ionospheric origin. The existence of the plasmasphere was already known before the spaceflight era based on the propagation characteristics of lightning-generated and man-made very low-frequency (VLF) waves. The plasmasphere has a relatively clear outer edge, the *plasmapause*, where the proton density drops several orders of magnitude. The location and structure of the plasmapause vary considerably as a function of magnetic activity (Figure 22.7). During magnetospheric quiescence the density decreases smoothly at distances from $4\text{--}6 R_E$, whereas during strong activity the plasmapause is steeper and pushed closer to the Earth.

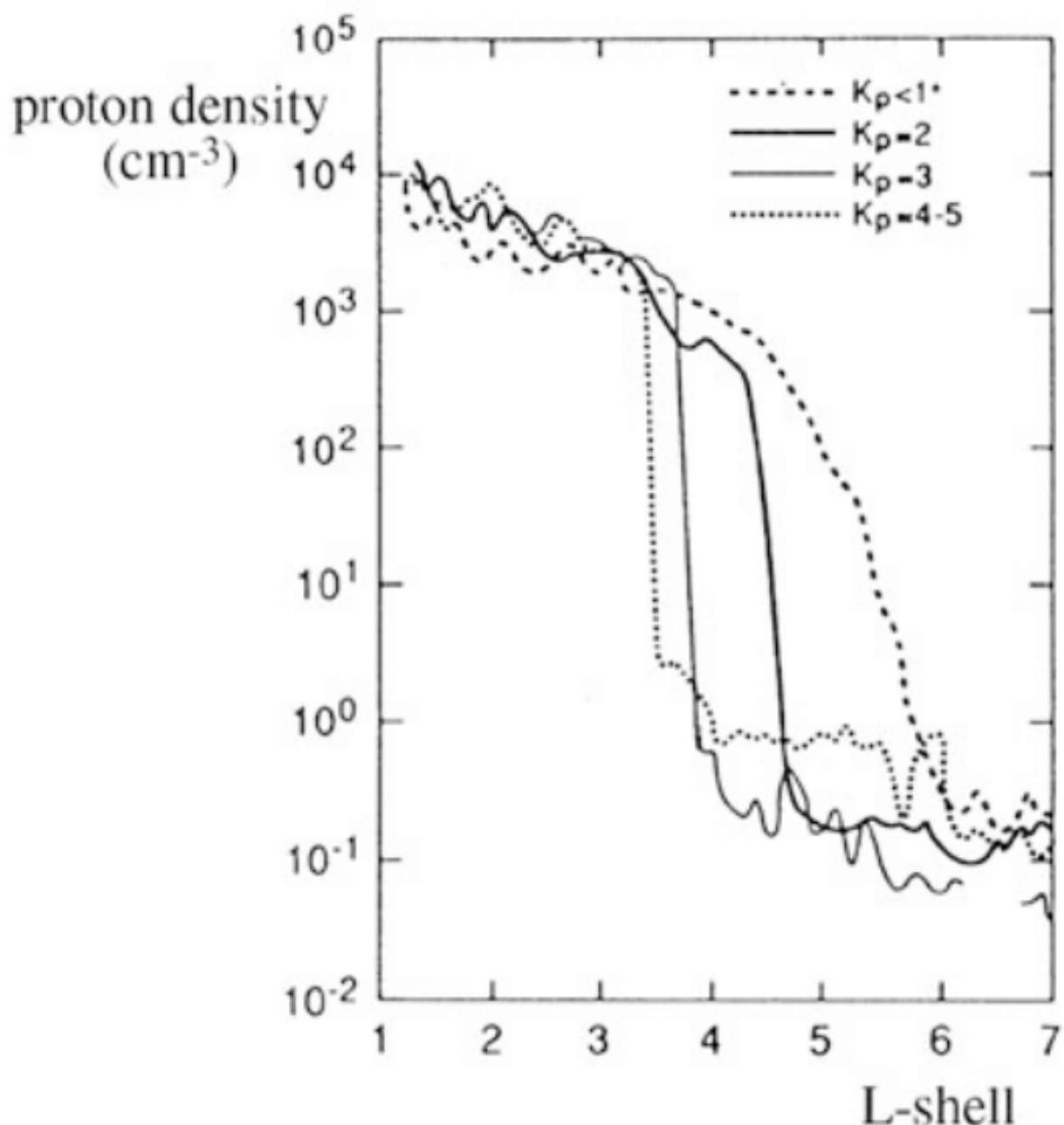


Figure 22.7: Plasma density in the night sector organized by the activity index K_p . $K_p < 1+$ corresponds to a very quiet magnetosphere, whereas $K_p = 4-5$ indicates a significant activity level, although not yet a big magnetic storm. The L-shell is defined in Equation 22.2. [Chappell (1972)]

The location of the plasmapause is determined by the interplay between the sunward convection of plasma sheet particles and the plasmaspheric plasma corotating with the Earth. By adding the convective and corotational electric fields to the guiding center motion of charged particles we find that an outward bulge called *plasmaspheric plume* develops on the duskside around 18 h magnetic local time (MLT). Plasmaspheric plumes are most common and pronounced during geomagnetic storms and substorms, but they can exist also during quiet conditions (e.g., Moldwin+ 2016). During geomagnetic storms the plume can expand out to geostationary orbit and bend toward earlier MLT.

Figure 22.8 shows global observations of the plasmasphere taken by the EUV instrument onboard the IMAGE satellite before and after a moderate geomagnetic storm in June 2000. Before the storm the plasmasphere was more or less symmetric. After the storm the plasmasphere was significantly eroded leaving a plume extending from the dusk toward the dayside magnetopause. When traversing the plume, the trapped radiation belt electrons, otherwise outside the plasmapause, encounter a colder and higher-density plasma with plasma wave environment similar to the plasmasphere proper. Consequently, the influence of the plasmasphere on radiation belt particles extends beyond its nominal boundary depicted in Figure 22.7.

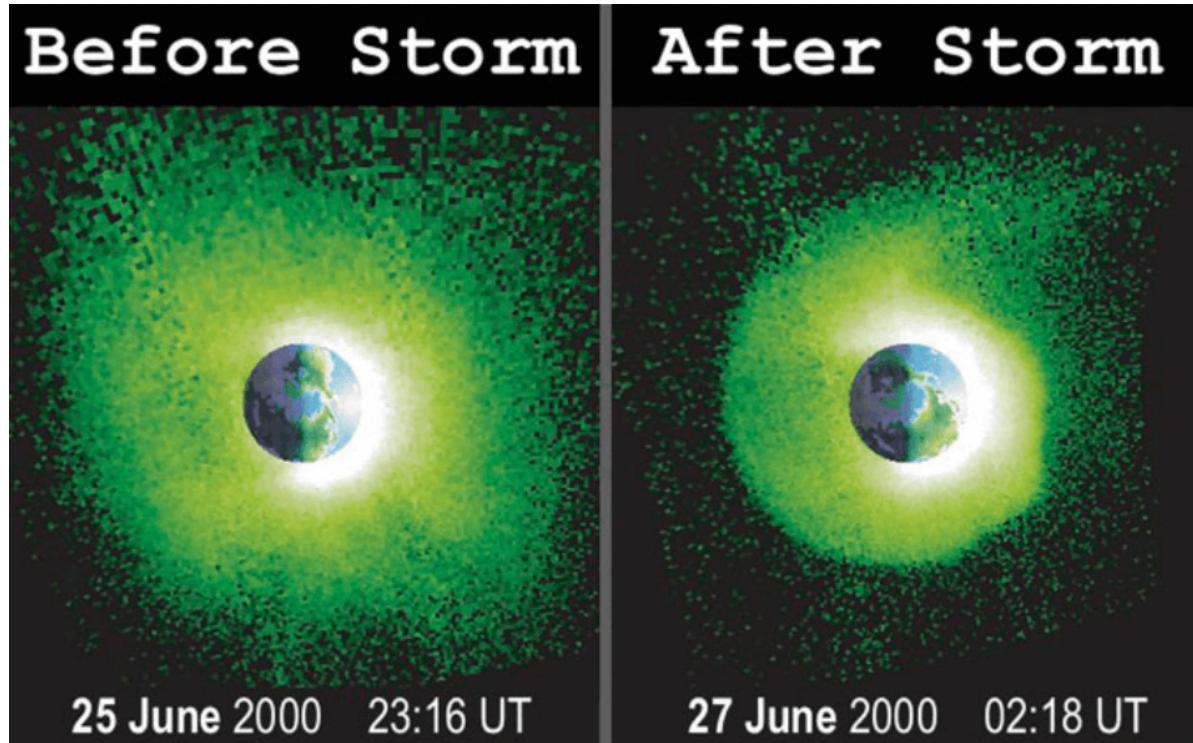


Figure 22.8: Plasmaspheric plume and plasmaspheric erosion as observed by the IMAGE EUV instrument. The picture is taken from above the northern hemisphere and the Sun is to the right. (Goldstein+2004)

The plasma parameters in the plasmasphere, in the plume and at the plasmapause are critical to the generation and propagation of plasma waves that, in turn, interact with the energetic particles in the ring current and radiation belts. Thus, the coldest and the hottest components of the inner magnetosphere are intimately coupled to each other through wave-particle interactions.

It is important to recognize that the ring current is the main driver of the field perturbations, whereas the radiation belt has negligible contributions due to its low number density.

22.2.3 Cosmic Rays

In addition to ion and electron radiation belts another important component of corpuscular radiation in the near-Earth space consists of cosmic rays. The kinetic energies of a large fraction of cosmic ray particles are so large that the geomagnetic field cannot trap them. Instead, the particles traverse through the Earth's magnetosphere without much deflection of their trajectories. Some of them hit the atmosphere interacting with nuclei of atmospheric atoms and molecules causing showers of elementary particles being possible to detect on ground. Those with highest energies can penetrate all the way to the ground.

The spectrum of cosmic ray ions at energies below about 10^{15} eV per nucleon in the near-Earth space has three main components:

- Galactic cosmic rays (GCR), whose spectrum peaks at energies above 100 MeV per nucleon, are most likely accelerated by supernova remnant shock waves in our galaxy.
- Solar cosmic rays (SCR) are accelerated by coronal and interplanetary shocks related to solar eruptions. Their energies are mostly below 100 MeV per nucleon and a fraction of them can become trapped in the inner radiation belt.
- Anomalous cosmic rays (ACR) are ions of solar origin captured and accelerated by the heliospheric termination shock, where the supersonic solar wind becomes subsonic before encountering the interstellar plasma, or in the heliosheath outside the heliopause. Some of the ions are injected back toward the Sun. Near the Earth the ACR spectrum peaks at about 10 MeV per nucleon and thus the particles can become trapped in the geomagnetic field.

Although the galactic cosmic rays cannot directly be trapped into the radiation belts, they contribute indirectly to the inner belt composition through the *Cosmic Ray Albedo Neutron Decay* (GRAND) mechanism. The cosmic ray bombardment of the atmosphere produces neutrons that move in all directions. Although the average neutron lifetime is 14 min 38 s, during which a multi-MeV neutron either hits the Earth or escapes far away from the magnetosphere, a small fraction of them decay to protons while still in the inner magnetosphere and may become trapped in the inner radiation belt.

Below about 10 GeV GCR and ACR fluxes are modulated by the 11- and 22-year solar cycles, so they provide quasi-stationary background radiation in the timescales of radiation belt

observations. The arrivals of SCRs are, in turn, transient phenomena related to solar flares and coronal mass ejections.

The cosmic ray electrons also have galactic and solar components. Furthermore, the magnetosphere of Jupiter accelerates high-energy electrons escaping to the interplanetary space. These Jovian electrons can be observed near the Earth at intervals of about 13 months when the Earth and Jupiter are connected by the IMF.

Supernova shock waves are the most likely sources of the accelerated GCR electrons, whereas in the acceleration of SCR and Jovian electrons also other mechanisms besides shock acceleration are important, in particular inductive electric fields associated with magnetic reconnection in solar flares and the Jovian magnetosphere.

The acceleration and identity of the observed very highest-energy cosmic rays up to about 3×10^{20} eV remain enigmatic. It should not be possible to observe protons with energies higher than 6×10^{19} eV, known as the *Greisen–Zatsepin–Kuzmin cutoff*, unless they are accelerated not too far from the observing site. Above the cutoff the interaction of protons with the blue-shifted cosmic microwave background produces pions that carry away the excessive energy. It is possible that the highest-energy particles are nuclei of heavier elements. This is, for the time being, an open question.

22.3 Magnetospheric Dynamics

Strong solar wind forcing drives *storms* and more intermittent *substorms* in the magnetosphere. They are primarily caused by various large-scale heliospheric structures such as interplanetary counterparts of *coronal mass ejections* (CMEs/ICMEs)², stream interaction regions (SIRs) of slow and fast solar wind flows, and fast solar wind supporting Alfvénic fluctuations. ICMEs are often preceded by interplanetary fast forward shocks and turbulent *sheath regions* between the shock and the ejecta, which all create their distinct responses in the magnetosphere. Because fast solar wind streams originate from coronal holes, which can persist over several solar rotations, the slow and fast stream pattern repeats in 27-day intervals and SIRs are often called *co-rotating interaction regions* (CIRs). However, stream interaction region is a physically more descriptive term. SIRs may gradually evolve to become bounded by shocks, but fully developed SIR shocks are only seldom observed sunward of the Earth's orbit. The duration of these large-scale heliospheric structures near the orbit of the Earth varies from a few hours to days. On average, the passage of a sheath region past the Earth takes 8–9 h and the passage of an ICME or SIR about 1 day. The fast streams typically influence the Earth's environment for several days.

²Both acronyms are commonly used. We call the ejection CME when it is observed in the Solar corona and ICME further away in the interplanetary space.

22.3.1 Magnetospheric Convection

Magnetospheric plasma is in a continuous large-scale advective motion, which in this context is, somewhat inaccurately, called *magnetospheric convection*. The convection is most directly observable in the polar ionosphere, where the plasma flows from the dayside across the polar cap to the nightside and turns back to the dayside through the morning and evening sector auroral region. The non-resistive *ideal magnetohydrodynamics* (MHD) is a fairly accurate description of the large-scale plasma motion above the resistive ionosphere. In ideal MHD the magnetic field lines are electric equipotentials and the electric field \mathbf{E} and plasma velocity \mathbf{V} are related to each other through the simple relation

$$\mathbf{E} = -\mathbf{V} \times \mathbf{B}$$

Consequently, the observable convective motion, or alternatively the electric potential, in the ionosphere can be mapped along the magnetic field lines to plasma motion in the tail lobes and the plasma sheet. As the electric field in the tail plasma sheet points from dawn to dusk and the magnetic field to the north, the convection brings plasma particles from the nightside plasma sheet toward the Earth where a fraction of them become carriers of the ring current and form the source population for the radiation belts.

In ideal MHD the plasma and the magnetic field lines are said to be frozen-in to each other. This means that two plasma elements that are connected by a magnetic field line remain so when plasma flows from one place to another. It is convenient to illustrate the motion with moving field lines, although the magnetic field lines are not physical entities and their motion is just a convenient metaphor. A more physical description is that the magnetic field evolves in space and time such that the plasma elements maintain their magnetic connection.

The convection is sustained by solar wind energy input into the magnetosphere. The input is weakest, but yet finite, when the interplanetary magnetic field (IMF) points toward the north, and is enhanced during southward pointing IMF. If the magnetopause were fully closed, plasma would circulate inside the magnetosphere so that the magnetic flux tubes crossing the polar cap from dayside to nightside would reach to the outer boundary of the magnetosphere where some type of *viscous interaction* with the anti-sunward solar wind flow would be needed to maintain the circulation. The classical (collisional) viscosity on the magnetopause is vanishingly small, but finite gyro radius effects and wave-particle interactions give rise to some level of *anomalous viscosity*³. It is estimated to provide about 10% of the momentum transfer from the solar wind to the magnetosphere.

The magnetosphere is, however, not fully closed. In the same year, when Axford and Hines presented their viscous interaction model, Dungey (1961) explained the convection in terms of magnetic reconnection. The *Dungey cycle* begins with a violation of the frozen-in condition at the dayside magnetopause current sheet. A magnetic field line in the solar wind is cut

³This is one of many examples of the questionable use of word “anomalous”. There is nothing anomalous in wave-particle interactions or processes beyond fluid description.

and reconnected with a terrestrial field line. Reconnection is most efficient for oppositely directed magnetic fields, as is the case in the dayside equatorial plane when the IMF points southward, but remains finite under other orientations. Subsequent to the dayside reconnection the solar wind flow drags the newly-connected field line to the nightside and the part of the field line that is inside the magnetosphere becomes a tail lobe field line. Consequently, an increasing amount of magnetic flux is piling up in the lobes. At some distance far in the tail the oppositely directed field lines in the northern and southern lobes reconnect again across the cross-tail current layer. At this point the ionospheric end of the field line has reached the auroral oval near local midnight. Now the earthward outflow from the reconnection site in the tail drags the newly-closed field line toward the Earth. The return flow cannot penetrate to the plasmasphere corotating with the Earth and the convective flow must proceed via the dawn and dusk sectors around the Earth to the dayside. In the ionosphere the flow returns toward the dayside along the dawnside and duskside auroral oval. Once approaching the dayside magnetopause, the magnetospheric plasma provides the inflow to the dayside reconnection inside of the magnetopause. Note that the resistive ionosphere breaks the frozen-in condition of ideal MHD and it is not reasonable to use the picture of moving field lines in the atmosphere.

The increase in the tail lobe magnetic flux and strengthening of plasma convection inside the magnetosphere during southward IMF have a strong observational basis. Calculating the east-west component of the motion-induced solar wind electric field ($E = VB_{\text{south}}$) incident on the magnetopause and estimating the corresponding potential drop over the magnetosphere, some 10% of the solar wind electric field is estimated to “penetrate” into the magnetosphere as the dawn-to-dusk directed convection electric field. Note that $\mathbf{E} = -\mathbf{V} \times \mathbf{B}$ is not a causal relationship indicating whether it is the electric field that drives the magnetospheric convection, or convection that gives rise to the motion-induced electric field. The ultimate driver of the circulation is the solar wind forcing on the magnetosphere.

The plasma circulation is not as smooth as the above discussion may suggest. If the reconnection rates at the dayside magnetopause and nightside current sheet balance each other, a steady-state convection can, indeed, arise. This is, however, seldom the case since the changes in the driving solar wind and in the magnetospheric response are faster than the magnetospheric circulation timescale of a few hours. Reconnection may cause significant erosion of the dayside magnetospheric magnetic field placing the magnetopause closer to the Earth than a simple pressure balance consideration would indicate. The changing magnetic flux in the tail lobes causes expansion and contraction of the polar caps affecting the size and shape of the auroral ovals.

Furthermore, the convection in the plasma sheet has been found to consist of intermittent high-speed *bursty bulk flows* (BBF) with almost stagnant plasma in between (Angelopoulos et al. 1992). It is noteworthy that while BBFs are more frequent during high auroral activity, they also appear during auroral quiescence. BBFs have been estimated to be the primary mechanism of earthward mass and energy transport in regions where they have been observed. Thus the high-latitude convection observed in the ionosphere corresponds to an average of the BBFs and slower background flows in the outer magnetosphere.

22.3.2 Geomagnetic Storms

As illustrated in Figure 22.9, the storms are periods of most dynamic evolution of radiation belts. They often, but not always, commence with a significant positive deviation in the horizontal component of the magnetic field (H) measured on the ground (Figure 22.10), called *storm sudden commencement* (SSC). An SSC is a signature of an ICME-driven shock and the associated pressure pulse arriving at the Earth's magnetopause. SSCs are also observed during pressure pulses related to SIRs or to ICMEs that are not sufficiently fast to drive a shock in the solar wind but still disturb and pile-up the solar wind ahead of them. If the solar wind parameters are known, the pressure effect can be removed from the Dst index as given by Equation 22.6.

Storms in the magnetosphere can also be driven by low-speed ICMEs and SIRs without a significant pressure pulse. SIR-driven storms occur if the field fluctuations have sufficiently long periods of strong enough southward magnetic field to sustain global convention electric field to enhance the ring current. Thus there are storms without a clear SSC signature in the Dst index. On the other hand, a shock wave hitting the magnetopause is not always followed by a geomagnetic storm, in particular, if the IMF points dominantly toward the north during the following solar wind structure. In such cases the positive deviation in the magnetograms is called a *sudden impulse* (SI), after which the Dst index returns close to its background level with small temporal variations only. If the dynamic pressure remains at enhanced level, Dst can maintain positive deviation for some period.

After the SSC an initial phase of the storm begins. It is characterized by a positive deviation of Dst, typically a few tens of nT. The initial phase is caused by a combination of predominantly northward IMF and high dynamic pressure. The phase can have very different durations depending on the type and structure of the solar wind driver. It can be very brief if the storm is driven by an ICME with a southward magnetic field following immediately a sheath with predominantly southward magnetic field. In such a case the storm main phase, which is a period characterized by a rapid decrease of the H component of the equatorial magnetic field, starts as soon as the energy transfer into the magnetosphere has become strong enough. If the sheath has a predominantly northward IMF, the main phase will not begin until a southward field of the ejecta enhances reconnection on the dayside magnetopause.

If there is no southward IMF either in the sheath or in the ICME, no regular global storm is expected to take place. However, pressure pulses/shocks followed by northward IMF can cause significant consequences to the radiation belt environment, as they can shake and compress the magnetosphere strongly and trigger a sequence of substorms (Section 22.3.3).

During the storm main phase, the enhanced energy input from the solar wind leads to energization and increase of the number of ring current carriers in the inner magnetosphere, as the enhanced magnetospheric convection transports an increasing amount of charged particles from the tail to the ring current region. Here substorms, discussed below, have important

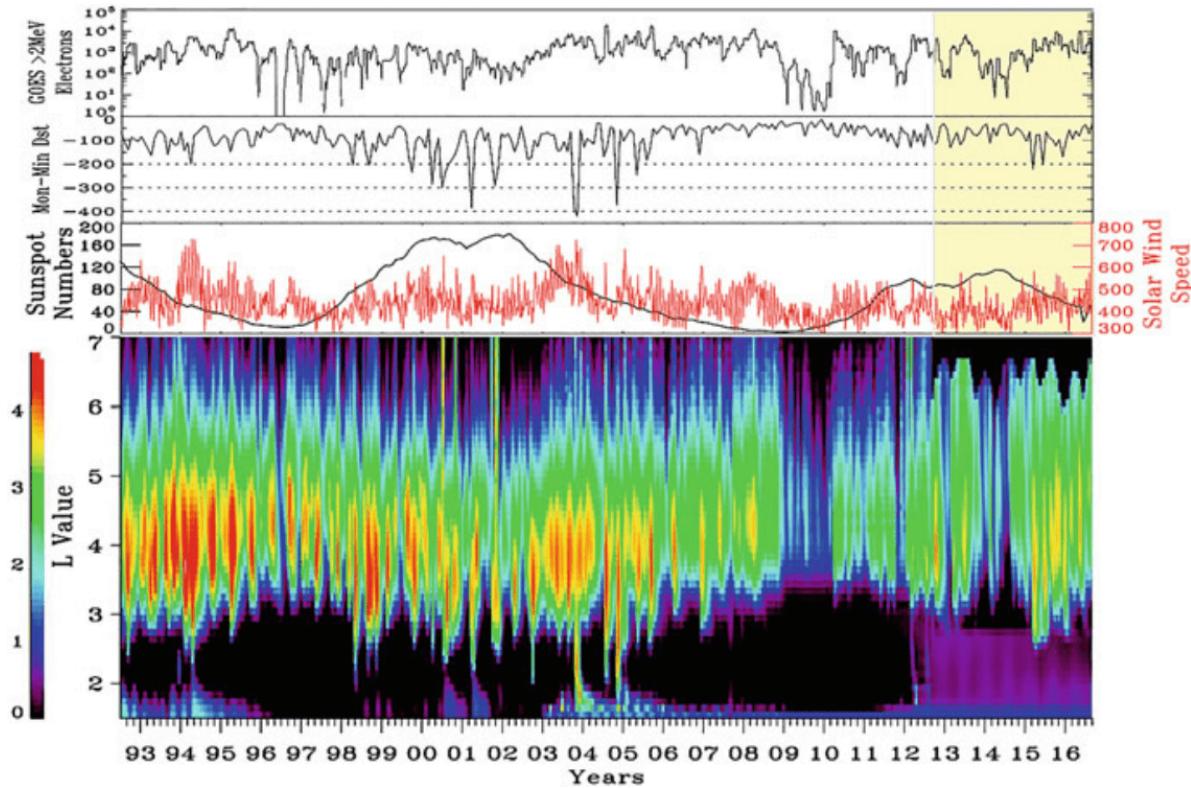


Figure 22.9: Outer radiation belt response to solar and magnetospheric activity from the SAMPEX satellite and Van Allen Probes observations over a period of more than two solar cycles. The uppermost panel shows 27-day window-averaged relativistic (>2 MeV) electron fluxes at geostationary orbit, the second panel the monthly minimum of the Dst index, and the third panel the yearly window-averaged sunspot number (black) and weekly window-averaged solar wind speed (red). The spectrogram in the lowest panel is a composite of 27-day window-averaged SAMPEX observations of relativistic (2 MeV) electron fluxes until September 2012 and Van Allen Probes REPT observations of (2.1 MeV) electron fluxes after 5 September 2012. The shift from SAMPEX to Van Allen Probes is visible in the change of sensitivity to particle flux in the slot region (Li et al. 2017)

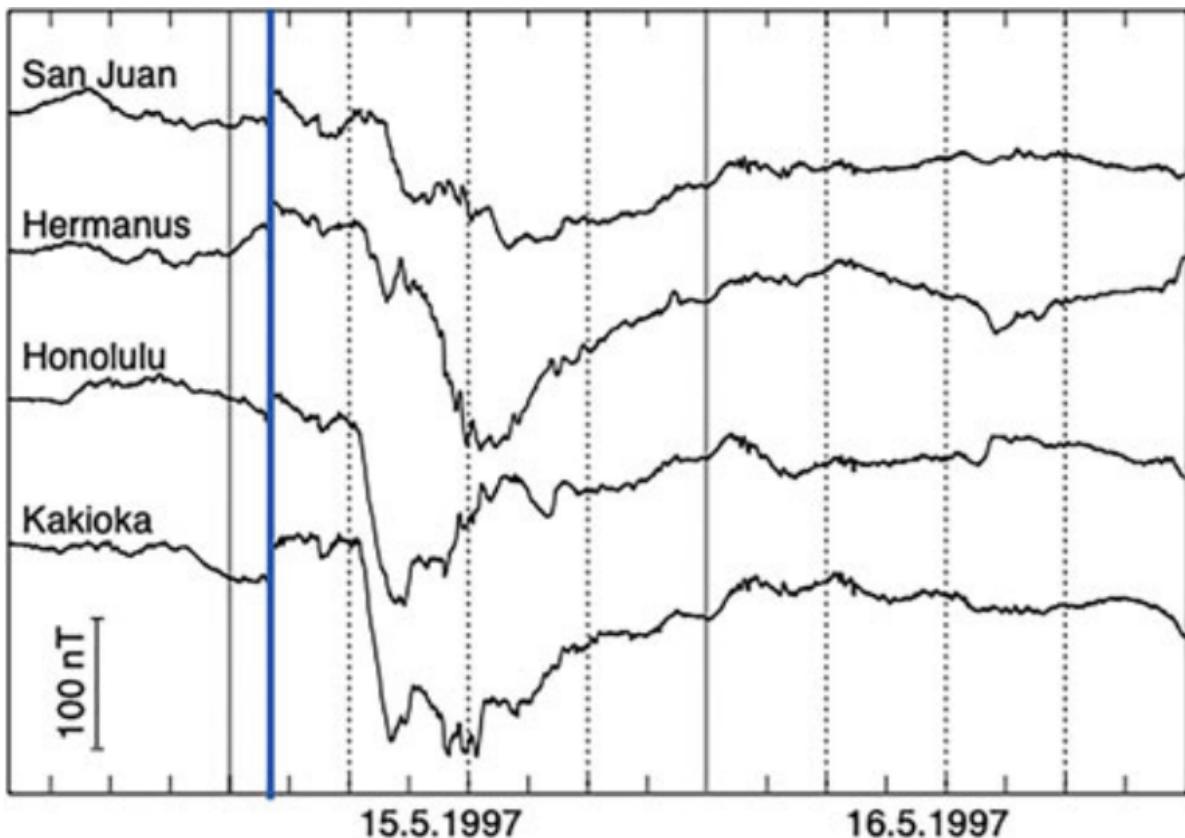


Figure 22.10: The horizontal component (H) of the magnetic field measured at four low-latitude stations during a magnetic storm on 15 May 1997. An ICME-driven solar wind shock hit the magnetosphere on 15 May at about 02 UT causing the storm sudden commencement which is indicated by a sudden positive jump of the H component at all stations (thick blue line). The main phase of the storm started after 06 UT as indicated by the strong negative deviation in the H component. The solid vertical lines give the UT midnight and the tick-marks on the horizontal axis are given for each 3 h.

contribution, as they inject fresh particles from the near-Earth tail. The ring current enhancement is typically asymmetric because not all current carrying ions are on closed drift paths but a significant fraction of them passes the Earth on the evening side and continue toward the dayside magnetopause. This is illustrated in Figure 22.10 where the Honolulu and Kakioka magnetometers show the steepest main phase development when these stations were in the dusk side of the globe.

When energy input from the solar wind ceases, the energetic ring current ions are lost faster than fresh ones are supplemented from the tail. The Dst index starts to return toward the background level. This phase is called the *recovery phase*. It is usually much longer than the main phase, because the dominating loss processes of the ring current carriers: charge exchange with the low-energy neutral atoms of the Earth's exosphere, wave-particle interactions, and Coulomb collisions, are slower than the rapid increase of the current during the main phase. As ICMEs last typically 1 day, storms driven by ICMEs trailed by a slow wind tend to have relatively short recovery phases, whereas storms driven by SIRs and ICMEs followed by a fast stream can have much longer recovery phases. This is because Alfvénic fluctuations, i.e., large-amplitude MHD Alfvén waves, in fast streams interacting with the magnetospheric boundary lead to triggering substorms, which inject particles to the inner magnetosphere. This can keep the ring current populated with fresh particles up to or longer than a week. The ring current development can also be more complex, often resulting in multi-step enhancement of Dst or events where Dst does not recover to quiet-time level between relatively closely-spaced intensifications. This typically occurs when both sheath and ICME ejecta carry southward field or when the Earth is impacted by multiple interacting ICMEs.

22.3.3 Substorms

Magnetospheric substorm is a transient solar-wind energy storage and release process within the magnetotail. It can inject fresh particles in the energy range from tens to a few hundred keV from the tail plasma sheet into the inner magnetosphere. After being injected to the quasi-dipolar magnetosphere, charged particles start to drift around the Earth, contributing to the ring current and radiation belt populations. The injections have a twofold role: They provide particles to be accelerated to high energies. Simultaneously the injected electrons and protons drive waves that can lead to both acceleration and loss of radiation belt electrons and ring current carriers.

Magnetospheric substorms result from piling of tail lobe magnetic flux in the near-to-mid-tail region during enhanced convection. The details of the substorm cycle are still debated after more than half a century of research. Observationally it is clear that substorms encompass global configurational changes in the magnetosphere, namely the stretching of the near-Earth nightside magnetic field and related thinning of the plasma sheet during the flux pile-up (*substorm growth phase*), followed by a relatively rapid return of the near-Earth field toward a dipolar shape (*expansion phase*), and a slower return to a quiet-time stretched configuration associated with thickening of the plasma sheet (*recovery phase*). A substorm cycle typically lasts

2–3 h. The strongest activity occurs following the onset of the expansion phase: The cross-tail current in the near-Earth tail disrupts and couples to the polar region ionospheric currents through magnetic field-aligned currents forming the so-called *substorm current wedge*. This leads to intense precipitation of magnetospheric particles causing the most fascinating auroral displays. During geomagnetic storms the substorm cycle may not be equally well-defined. For example, a new growth phase may begin and the onset of the next expansion may follow soon after the previous expansion phase.

A widely used, though not the only, description of the substorm cycle is the so-called *near-Earth neutral line* model (NENL model, for a review, see Baker et al. (1996)). In the model the current sheet is pinched off by a new magnetic reconnection neutral line once enough flux has piled up in the tail. The new neutral line forms somewhere at distances of 8–30 R_E from the Earth, which is much closer to the Earth than the far-tail neutral line of the Dungey cycle. Earthward of the neutral line plasma is pushed rapidly toward the Earth. Tailward of the neutral line plasma flows tailward, and together with the far-tail neutral line, a tailward moving structure called *plasmoid* forms. Sometimes recurrent substorm onsets can create a chain of plasmoids. While it is common to illustrate the plasmoid formation using two-dimensional cartoons in the noon–midnight meridional plane, the three-dimensional evolution of the substorm process in the magnetotail is far more complex. In reality a plasmoid is a magnetic flux rope whose two-dimensional cut looks like a closed loop of magnetic field around a magnetic null point.

As pointed out in Section 22.3.2, the plasma flow in the central plasma sheet is not quite smooth and a significant fraction of energy and mass transport takes place as bursty bulk flows (BBFs). The BBFs are thought to be associated with localized reconnection events in the plasma sheet roughly at the same distances from the Earth as the reconnection line of the NENL model. They create small flux tubes called *dipolarizing flux bundles* (DFBs). The name derives from their enhanced northward magnetic field component B_z corresponding to a more dipole-like state of the geomagnetic field compared to a more stretched configuration. Once created, DFBs surge toward the Earth due to the force caused by magnetic curvature tension in the fluid picture. They are preceded by sharp increases of B_z called *dipolarization fronts*. DFBs are also associated with large azimuthal electric fields, up to several mV m^{-1} , which are capable of accelerating charged particles to high energies. Whether the braking of the bursty bulk flows and coalescence of dipolarization fronts closer to the Earth cause the formation of the substorm current wedge, or not, is a controversial issue.

The NENL model has been challenged by the common observation that the auroral substorm activation starts at the most equatorward arc and expands thereafter poleward. Another model proposed is the *current disruption* (CD) model, which is built upon plasma instabilities. In the CD model, a three-dimensional plasma instability grows first near the Earth in the transition region between the stretched tail and the dipolar inner magnetosphere. This instability drives steepening waves, leading to current disruption as the tail current cannot be sustained within a strongly oscillating geometry. The current disruption then launches tailward-propagating waves, which later trigger reconnection, plasmoid and fast flows. An easy way to think about these two models is that NENL is “outside-in”, while CD is “inside-out”. As both the current

disruption and the plasma release from reconnection occur in only a few minutes, albeit roughly 100,000 km apart, it is challenging to uncover why and how the current disruption and plasmoid ejection take place. Setting aside the debate between competing explanations, what is essential is that the substorm expansions dipolarize the tail magnetic field configuration having been stretched during the growth phase and inject fresh particles into the inner magnetosphere. The particle injections can be observed as *dispersionless*, meaning that injected particles arrive to the observing spacecraft simultaneously at all energies, or *dispersive* when particles of higher energies arrive before those of lower energies. Because the dispersion arises from energy-dependent gradient and curvature drifts of the particles, a dispersionless injection suggests that the acceleration occurs relatively close to the observing spacecraft, whereas dispersive arrival indicates acceleration further away from the observation when the particle distribution has had time to develop dispersion due to energy-dependent drift motion.

Dispersionless substorm injections are typically observed close to the midnight sector at geostationary orbit ($6.6 R_E$) and beyond, but have been found all the way down to about $4 R_E$ (Friedel, Korth, and Kremser 1996). The injection sites move earthward as the substorm progresses and are also controlled by geomagnetic activity, although the extent of the dispersionless region is unclear, both in local time and radial directions. Neither have the details of acceleration of the injected particles been fully resolved. It has been suggested to be related both to betatron and Fermi acceleration associated with earthward moving dipolarization fronts. Another important aspect of dipolarization fronts for radiation belts is their braking close to Earth, which can launch magnetosonic waves that can effectively interact with radiation belt electrons.

22.4 ULF Waves

Ultra-low frequency (ULF) waves refer to waves within frequency range [0.001, 10] Hz. The name does not tell us anything about their physical origin, but simply observational fact. At Earth's magnetosphere, this frequency range overlaps largely with the MHD waves. This is the reason why early pioneers in space physics relied on MHD theory for large spatial and temporal scales to explain the physics behind these waves, albeit some deviations and deficiencies which require more refined models such as the Vlasov description. ULF waves permeate the near-Earth plasma environment and play an important role in its dynamics, for example in transferring energy from the solar wind to the magnetosphere or accelerating electrons in the Earth's radiation belts.

ULF waves were originally called micropulsations or magnetic pulsations since they were first observed by ground magnetometers. ULF pulsations are classified into two types: *pulsations continuous (Pc)* and *pulsations irregular (Pi)* with several subclasses (Pc1–5 and Pi1–2) according to their frequencies and durations. The division is based on their physical and morphological properties, and the boundaries are not strict.

Notation	Period Range [s]	Property
Pc1	0.2 - 5	EMIC
Pc2	5 - 10	EMIC, Mirror
Pc3	10 - 45	Foreshock, FLR, Mirror
Pc4	45 - 150	FLR
Pc5	150 - 600	SW, FLR
Pi1	1 - 40	
Pi2	40 - 150	

With respect to polarization, field line resonant ULF waves can be categorized into three modes: *compressional* (ΔB_{\parallel} , ΔE_{ϕ})⁴, *poloidal* (ΔB_r , ΔE_{ϕ}), and *toroidal* (ΔB_{ϕ} , ΔE_r). Here, B_r (E_r), B_{\parallel} , and B_{ϕ} (E_{ϕ}) are the radial, parallel (or compressional), and azimuthal components in the local magnetic field system, respectively. Referring to the basic MHD theory, the compressional modes are fast modes, whereas the poloidal and toroidal modes are Alfvén modes. The perturbed EM fields are related by $\mathbf{B}_1 = \frac{\mathbf{k}}{\omega} \times \mathbf{E}_1$. Think of a closed field line near the equatorial plane inside the magnetosphere: if the wave vector \mathbf{k} is along the field line, i.e. $\mathbf{k} = (0, 0, k_z)$, then there will be two cases for the EM field: poloidal where \mathbf{E}_1 in $\hat{\phi}$, \mathbf{B}_1 in \hat{r} and toroidal where \mathbf{E}_1 in \hat{r} , \mathbf{B}_1 in $\hat{\phi}$. If the wave vector \mathbf{k} is perpendicular to the field line, i.e. $\mathbf{k} = (k_x, 0, 0)$, since there is no E_{\parallel} in MHD, we only have one option \mathbf{E}_1 in $\hat{\phi}$ and \mathbf{B}_1 in \hat{z} . A phase shift is allowed, and actually in real observations (e.g. THEMIS) it is rare that you can find B and E changing in-phase. These classifications are summarised in Table 16.1.

22.4.1 Pc1 & Pc2

- Usually observed in the noon-afternoon MLT sector, easily detectable when following sudden impulses (SI) produced by sudden changes in the pressure of the solar wind plasma.
- A sudden compression of the magnetosphere by increased solar wind pressure causes maximum distortion of the quiet magnetospheric plasma near noon at high latitudes. It is on the fieldlines which thread this disturbed plasma that one is most likely to witness ULF emissions.
- Conversely, as suggested by Hirasawa (1981) sudden rarefactions of the magnetosphere would be expected to quench ULF wave growth by reducing the anisotropy and β of the plasma. (INTERESTING ONE, SHOULD CHECK AT SOMETIME!)
- Delay of 1-3 mins between the occurrence of SI and the onset of the ULF emission (ground-based magnetometers)[^growth_rate]

⁴In the literature sometimes the compressional mode is called the *compressional poloidal mode*, and the poloidal mode refers to both the compressional and Alfvénic poloidal mode.

- Drive the trapped proton radiation, greatly enhanced eV energy range protons along the B field, and energization of keV range protons caused by betatron acceleration (Arnoldy et al. 2005).

At Earth, Electromagnetic ion cyclotron (EMIC) waves are typically observed in Pc1 and Pc2 range. In the outer radiation belt, the frequency typically ranges between 0.1 to 5 Hz. The preferential region of occurrence of EMIC waves is known to be the afternoon magnetic local time (MLT) sector from 12:00 to 18:00 MLT in the region near the plasmapause and the plasmaspheric plume.

EMIC waves are normally excited by a temperature anisotropic ($T_{\perp} > T_{\parallel}$) distribution of hot ($\sim 1 - 100$ keV) ions. They are preferentially generated in regions where hot anisotropic ions and cold dense plasma populations spatially overlap (Figure 22.11).

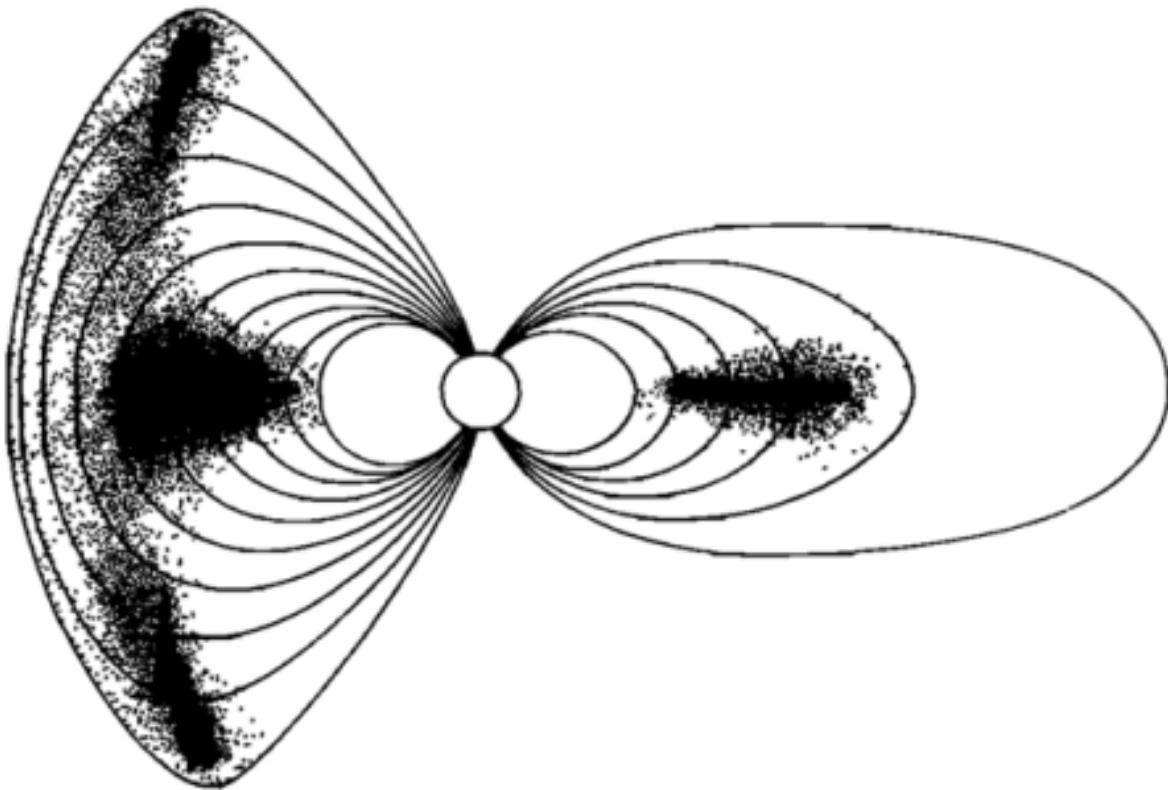


Figure 22.11: Schematic sketch of chorus/EMIC wave excitation regions, denoted by dots, on the noon-midnight meridian plane.

The excited EMIC waves can result in the energization and loss of magnetospheric particles. Through resonant wave-particle interactions, EMIC waves are able to accelerate cold ions into the thermal (~ 1 eV – 1 keV) energy range in the direction perpendicular to the ambient

magnetic field, and cause the pitch angle scattering loss of hot ions in the ring current. Particularly, EMIC waves can also resonantly interact with relativistic electrons and result in pitch angle scattering of the electrons.

Newly excited EMIC waves are often transverse and left-hand polarized, consistent with the direction of ion gyration in the magnetic field (well, no surprise). After being generated, EMIC waves can be guided along the magnetic field lines and propagate from the source region to other magnetic latitudes. Spacecraft measurements have shown that EMIC wave propagation is almost exclusively away from the equator at latitudes greater than about 11° , with minimal reflection at the ionosphere.

Some waves may even experience a polarization reversal where the wave frequency f is equal to the crossover frequency f_{co} during their higher-latitude propagation and then be reflected where f equals the bi-ion hybrid frequency f_{bi} at an even higher latitude. As a result, their polarization is crossed over from a left-hand to a right-hand or linear mode. These waves could undergo multiple equatorial crossings along magnetic flux tubes without a large radial or azimuthal drift. Because of their successive passes through the equatorial wave growth region, the waves are expected to be drastically amplified by continuing to obtain energy from the energetic protons. Nevertheless, Horne and Thorne argued that in the absence of density gradients, significant wave amplifications can only occur on the first equatorial pass because wave normal angles become large after the initial pass; wave damping by thermal heavy ions also makes it impossible for the same EMIC wave packet to bounce through its source region multiple times.

Mirror Instability & Ion Cyclotron Instability

Already, early observations in the 1970s have shown that the magnetosheath is populated by intense magnetic field fluctuations at time scales from 1 s to 10 s of seconds. Later research based primarily on data from ISEE and AMPTE satellites has shown that the mirror mode waves (Section 14.2) and kinetic Alfvén ion cyclotron (AIC) waves (i.e. EMIC waves) constitute a large majority of magnetosheath ULF waves:

1. AIC/EMIC are found predominantly near the bow shock and in the plasma depletion layer⁵ (Song, Russell, and Gary 1994; Hubert et al. 1998).
2. Mirror mode waves dominate in the central and downstream magnetosheath but can occur immediately downstream of quasi-perpendicular shocks too. (Hubert et al. 1989)
3. ULF waves are generally stronger in the dayside magnetosheath.
4. More frequency EMIC wave occurrence during quasi-parallel shocks.

The ion cyclotron instability responsible for the generation of AIC waves often grows under the same conditions as the mirror instability and in the linear approximation should dominate in lower β plasmas. The mirror instability, on the other hand, should dominate in high ion β plasmas (Lacombe and Belmont 1995). Since the initial confirmation of the existence of mirror

⁵The plasma depletion layer (PDL) is a layer on the sunward side of the magnetopause with lower plasma density and higher magnetic field compared to the corresponding upstream magnetosheath value.

modes in the Earth's magnetosheath, they have been observed throughout the heliosphere. A long-standing puzzle in space plasmas is the fact that *mirror modes are often the dominant coherent magnetic structures even for low β plasmas.*

People tried to find an answer to this puzzle. A bunch of studies in late 1980s and early 1990s (e.g. [Gary+]) argued that the presence of He^{++} tends to increase the EMIC threshold while the mirror mode growth is less affected by the presence of He^{++} ions. Yoshiharu Omura and his student Shoji presented another possibility in 2009 with hybrid PIC simulations that even though EMIC modes have higher linear growth rate, they saturates at an earlier stage than the mirror modes, especially in higher dimensions (by comparing 2D and 3D results), so that mirror mode waves can gain more free energy from temperature anisotropy.

22.4.2 Pc3 & Pc4

As already noted above, in the beginning when people proposed the ULF wave Pc divisions, many underlying physics are still unclear. The boundaries are chosen based on the observation data back then and does not necessarily contain any physical meaning.

ULF waves in the Pc3 range, with periods between 10-45 s, are a common feature of the dayside magnetosphere, where they are frequently observed both by spacecraft and ground-based observatories. They are thought to originate from the ion foreshock, extending upstream of the Earth's quasi-parallel bow shock (the angle between IMF and shock normal $\theta_{Bn} \leq 45^\circ$). There, ULF waves in the Pc3 frequency range are produced by *ion beam instabilities*, due to the interaction of shock-reflected suprathermal ions with the incoming solar wind.

For the foreshock-related Pc3/4 waves, we have the following picture. After foreshock waves are generated, they propagate through the magnetosheath (with very few observations) and reach the magnetopause. They enter the dayside magnetopause and travel antisunward into the magnetosphere as compressional Pc3 fluctuations, transporting the wave energy towards the nightside. In the inner magnetosphere, they may couple to Alfvénic field line resonances (FLRs), where their frequency matches the eigenmodes of the Earth's magnetic field lines. Pc3 FLRs was observed at low latitudes and Pc4 at midlatitudes [Yumoto+, 1985]. The amplitude of the compressional mode decays when moving further into the magnetosphere, yet they can sometimes be observed all the way to the midnight sector. Compressional Pc3 wave power associated with transmitted foreshock waves is confined near the equator. Statistical study also shows that equatorial Pc3 wave power is stronger in the prenoon or noon sector (under various geomagnetic activity levels), consistent with the foreshock extending upstream of the dawn flank bow shock for a Parker-spiral IMF orientation. However, contrary to Pc5 pulsations (150-600 s), Pc3 wave activity does *not* show a clear correlation with the level of geomagnetic disturbances.

Also, note that not all Pc3 waves are related to foreshock waves, thus we may have different survey results about the distribution of Pc3 waves. This hints the fact that we are far from understanding the whole physical mechanism of wave generation.

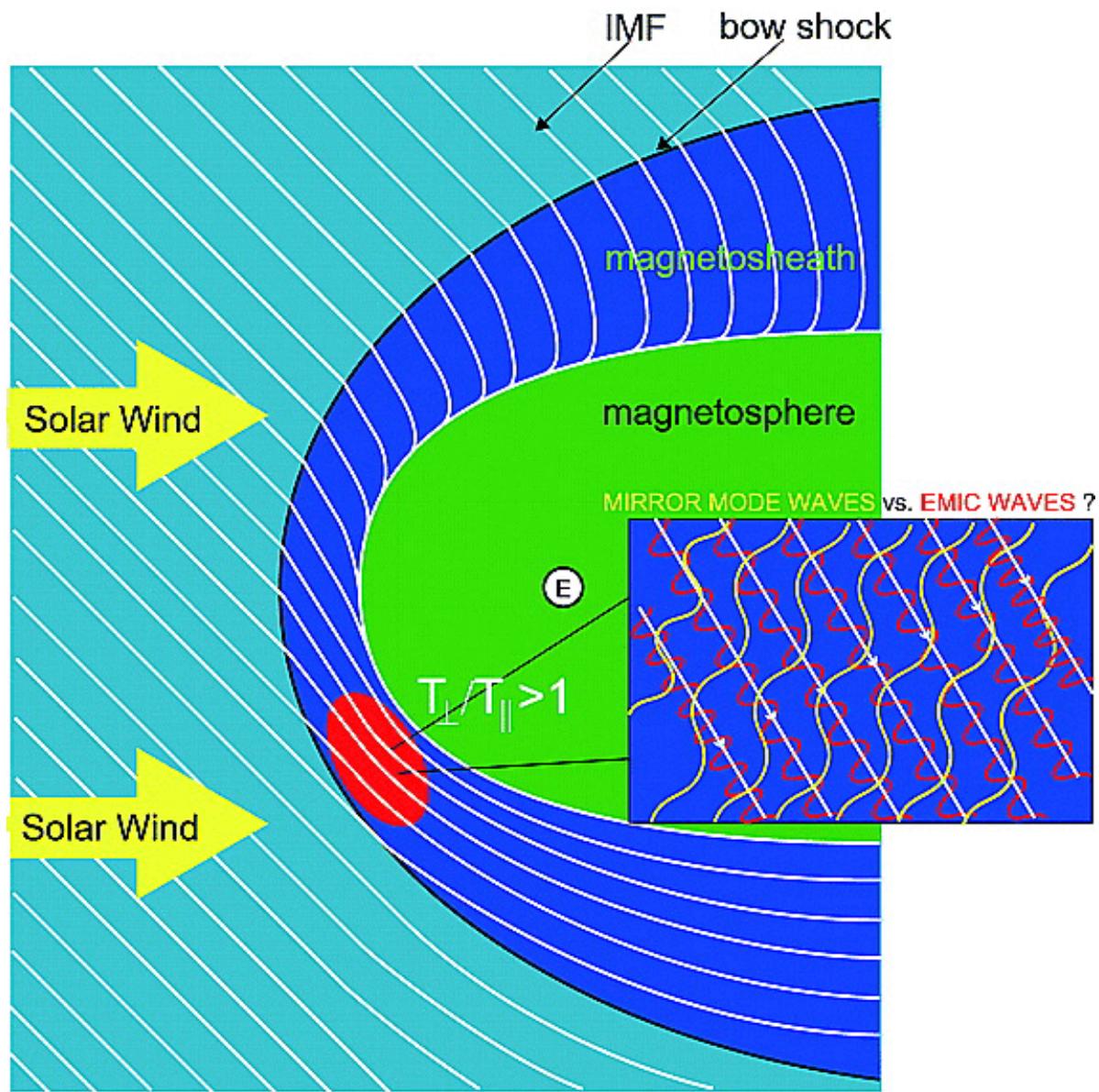


Figure 22.12: Mirror mode waves vs. EMIC waves in the magnetosphere.

There are several mechanisms by which Pc3-4 ULF waves may propagate to high latitudes:

- Harmonics of fundamental mode Pc5 resonances. Such harmonics would be expected to exhibit the same form of amplitude and phase properties that characterize FLRs and should occur at the same time as the fundamental.
- Cavity modes (Chapter 16).
- Fast mode waves propagate without mode conversion through the magnetosphere directly to the ionosphere. Such waves are subject to refraction and diffraction on their passage through the magnetosphere and may be directed to high latitudes via Fermat's Principle (???).
- Transistor model that invokes beams of precipitating *electrons* [Engebretson+ 1991] (???). The transistor model requires no wave mode coupling or wave propagation across field lines, rather the modulated precipitation of electrons in response to pressure fluctuations in the magnetosheath. The latter are attributed to the upstream ion-cyclotron resonance mechanism. The modulated electron beams convey wave information from the outer magnetosphere region containing the parent population of trapped electrons, to the near-cusp ionosphere. The resultant periodic precipitation would modulate the ionospheric conductivity and hence ionospheric currents equatorward of the cusp. Overhead field lines could then be excited by these modulated currents equatorward of the cusp, with the same frequency as the modulated electrons. Engebretson likened this behavior to that of a [transistor](#), where a small base current modulates a larger flow from collector to emitter. These ULF waves are characterized by noise-like appearance and low coherence lengths.

22.4.3 Pc5

We learned from ground, ionospheric, and space observations about the existence of *only one or at least a few* resonant field line vibrations (i.e. eigenoscillations) in the Pc5 range in the magnetosphere. As the Alfvén velocity is varying in the radial direction and as most sources of magnetospheric hydromagnetic waves are broadband sources, the resonance condition can be matched at an infinite number of geomagnetic field lines. Thus, every field line should be in resonance for a broad enough energy source. Therefore the observational fact of “magic” frequencies requires a magnetospheric frequency selection rule. (Kivelson and Southwood 1985) suggested that perturbations due to a broadband source at, e.g., the magnetopause first couple to the discrete eigenoscillations of global compressional eigenmodes. These narrow band compressional modes then couple to Alfvénic perturbations due to the field line resonance mechanism. The global modes thus select the frequency components participating in the resonant coupling. An alternative way of selection has been proposed by [Fujita+, 1996], who demonstrated that the K-H instability at the surface of a non-uniform magnetospheric plasma introduces dispersive properties of the unstable waves, which then gives rise to a narrower source spectrum. Thus, the field line resonance concept as outlined in Chapter 16 is able to explain the major features of observed resonant ULF pulsations in the terrestrial magnetosphere.

22.4.4 Sinks

According to (D. Southwood and Hughes 1983), possible sinks of ULF wave energy include at least three mechanisms:

- damping through ionospheric Joule heating,
- generalized Landau damping, and
- mode coupling.

By comparing the effects of Landau damping, Joule heating, and waveguide propagation, later researchers found from case studies that Joule heating and magnetospheric waveguide propagation are insufficient to account for the observed decay rate of ULF wave energy; Landau damping of the wave due to drift-bounce resonance with energetic ions was probably stronger, and more efficient when heavy ions such as O^+ are present.

22.4.5 Interaction with Energetic Particles

22.4.5.1 Resonance

A widely used theory of energetic particle modulation by ULF transverse waves was developed by (D. J. Southwood and Kivelson 1981, 1981). In their theory, particles experience the wave-carried electric field during their drift-bounce motion and their energy is accordingly changed. The drift-bounce resonance of energetic particles is determined from

$$\Omega - m\omega_d = N\omega_b \quad (22.7)$$

where N is an integer (normally $\pm 1, \pm 2$ or 0), m represents the azimuthal mode number of the ULF wave, and Ω, ω_d and ω_b are the wave frequency and particle drift and bounce frequencies, respectively. $N = 0$ is known as the drift resonance whereas $N \neq 1$ is the drift-bounce resonance. Because of the known energy dependence of ω_d and ω_b , the resonance energy can be calculated if the wave properties (Ω and m) are known. Observations from MMS and Cluster can be found in the review paper by Zong, Rankin, and Zhou (2017).

For particles of charge q experiencing drift motion in a ULF wave, the kinetic energy W changes at the following rate:

$$\frac{dW_A}{dt} = q\mathbf{E} \cdot \mathbf{v}_d \quad (22.8)$$

where subscript A signifies an average over many gyration periods, \mathbf{E} is the wave associated electric field, and \mathbf{v}_d is the magnetic gradient and curvature drift velocity. For an equatorially mirroring particle (vanishing curvature drift) in Earth's magnetic dipole field, the drift velocity \mathbf{v}_d can be approximated in the nonrelativistic limit by

$$\mathbf{v}_d = -\frac{3L^2 W}{qB_E R_E} \hat{e}_\phi \quad (22.9)$$

where \hat{e}_ϕ is defined eastward, R_E is Earth's radius, L is the L-shell parameter (radial distance in R_E of the equatorial crossing point a field line), and B_E is the equatorial magnetic field on Earth's surface. ULF waves propagate in the azimuthal direction with the wave-associated electric field given by

$$\mathbf{E} = E_\phi \exp i(m\phi - \omega t) \hat{e}_\phi \quad (22.10)$$

where ϕ is the magnetic longitude (increasing eastward), m is the azimuthal wave number, and ω is the wave angular frequency. From (Equation 22.8, Equation 22.9, Equation 22.10) the average rate of change of the particle energy is thus given by the following equation:

$$\frac{dW_A}{dt} = -\frac{3L^2W}{B_E R_E} \cdot E_\phi \exp i(m\phi - \omega t) \quad (22.11)$$

which should be integrated along the particle drift orbit to $t = -\infty$ to achieve an energy gain δW_A from waves. Here, the particle's drift orbit is assumed to be unperturbed (despite the energy change of the particle from the waves) with the angular drift frequency ω_d given by

$$\omega_d = \frac{d\phi}{dt} = -\frac{3L^2W}{B_E R_E^2} \quad (22.12)$$

An integration of Equation 22.11 backward in time along the particle's drift orbit Equation 22.12 leaves a result that depends on the initial condition assuming the amplitude of the sinusoidal waves remains constant. To circumvent this problem, Southwood and Kivelson (1981) assumed the wave angular frequency ω is complex with a small, positive imaginary part $\Im(\omega)$ that represents a gradually growing wave signal. This assumption enables the particle to see a sinusoid for a finite interaction time, which on integration of Equation 22.11 yields an averaged particle energy gain (???)

$$\delta W_A = -i \cdot \frac{3L^2W}{B_E R_E} \frac{E_\phi \exp i(m\phi - \omega t)}{\omega - m\omega_d} \quad (22.13)$$

Drift resonance happens when the particle drifts at the same azimuthal speed as the wave phase velocity, which indicates that $m\omega_d = \Re(\omega)$ so that the denominator of Equation 22.13 becomes $\Im(\omega) \times i$. This small imaginary term suggests that for resonant particles, δW_A must oscillate at a large amplitude in antiphase with the wave electric field. For particles with lower or higher energies (smaller or larger ω_d) the denominator is dominated by its real part and the corresponding δW_A oscillations have much smaller amplitudes and are $\pm 90^\circ$ out-of-phase with the wave electric field. In other words the amplitude δW_A of energy gain must shift in phase by 180° as the resonance is crossed from low energy to high energy.

An actual particle detector cannot measure δW_A directly, thus Southwood and Kivelson (1981) discussed the associated variations of particle fluxes and phase space densities (PSDs) that theoretical results can be compared directly with observational data. By assuming a negligible

gradient of pre-existing particle PSDs in the azimuthal direction, the wave-produced PSD variations, δf_A , can be written as

$$\delta f_A = \delta W_A \left[\frac{L}{3W} \frac{\partial f(W, L)}{\partial L} - \frac{\partial f(W, L)}{\partial W} \right]$$

which shows that δf_A is proportional to δW_A provided there is a pre-existing finite PSD gradient in energy and/or space. Here, the importance of spatial gradients in producing PSD oscillations is emphasized, which is caused by back-and-forth convection of particles as a response to the wave electric field. The PSD variations can be alternatively written as

$$\delta f_A = -\delta W_A \frac{\partial f(W, \mu)}{\partial W} = \delta W_A \frac{L}{3W} \frac{\partial f(L, \mu)}{\partial L}$$

where μ is assumed to be constant for adiabatic behavior of particles interacting with ULF waves. The linear dependence of δf_A on δW_A suggests that the phase shift of particle PSDs across the resonant energy should also be 180° . Such a phase shift is thus treated as a characteristic signature of ULF wave-particle drift resonance.

22.4.5.2 Particle trapping by compressional waves

The mirror force produced by the magnetic field-aligned component of compressional mode ULF waves can also modulate energetic particle fluxes. Observational evidence from Van Allen Probes can be found in the citations of Zong, Rankin, and Zhou (2017).

22.4.5.3 Fast acceleration of charged particles by ULF waves

The theory of drift-bounce resonance developed by Southwood and Kivelson (1981, 1981) is valid for axisymmetric magnetic fields. In this case, particles experience the wave electric field along their drift and bounce motion and suffer net energy gain or loss when the resonance condition Equation 22.7 is satisfied, usually for $N = 0, \pm 1, \pm 2$. Parallel electric fields are neglected as they are usually small in collisionless plasma⁶. The rate of energy change of a charged particle interacting with a ULF wave is expressed by

$$\frac{dW}{dt} = \mu \frac{\partial B_{\parallel}}{\partial t} + q \mathbf{E} \cdot \mathbf{v}_d$$

where \mathbf{E} , \mathbf{v}_d and μ denote the wave electric field, particle drift velocity, and magnetic moment, respectively. The first and second terms on the right-hand-side of the equation represent acceleration due to the wave magnetic and electric fields, respectively.

In an asymmetric compressed dipole field, a particle can resonate with low-m global toroidal waves when the wave frequency satisfies $\omega = (m \pm 1)\omega_d$. For example, Elkington, Hudson, and

⁶On the kinetic scale, this is not true!

Chan (1999) showed that in a magnetic topology with noon-midnight asymmetry electrons can gain energy from toroidal ULF waves; in other words, toroidal ULF waves can accelerate energetic particles in the radiation belt region under strong solar wind pressure. However, in the inner magnetosphere noon-midnight asymmetry becomes insignificant on low L-shells. Acceleration of energetic electrons by toroidal mode ULF waves may, therefore, be limited to the outer magnetosphere, while in the inner magnetosphere the poloidal mode is more likely responsible. However, in the dayside outer magnetosphere the poloidal electric field of fast mode wave may also accelerate energetic electrons.

22.4.6 EMIC Wave Events

(Usanova and Mann 2016)

EMIC waves are believed to be important for influencing the dynamics of energetic particles in the inner magnetosphere, especially in relation to ring current and radiation belt dynamics. In relation to the ring current, both the dynamical evolution of ion distributions along their drift paths and effects from the solar wind can result in EMIC wave generation from unstable ion distributions. EMIC waves often grow as a result of perpendicular temperature anisotropy, with the ion dynamics and energy and pitch angle distributions in the ring current being modified significantly by the growth of the EMIC instability. EMIC waves are also hypothesized to influence higher energy electrons in the Van Allen belts, through a Doppler-shifted cyclotron resonance, including stimulating potential scattering loss into the atmosphere.

The prevalence of dayside EMIC activity during intervals of enhanced solar wind dynamic pressure can be explained by three possible mechanisms, which can lead to the generation of anisotropic proton distributions.

22.4.6.1 Bursty EMIC wave events

Short-term magnetospheric compressions can increase the $A = T_{\perp}/T_{\parallel}$ ratio through *adiabatic heating* and can generate EMIC wave bursts during periods within which the magnetic field strength is increasing (i.e. $dB/dt > 0$). In a dipole field, the resulting anisotropy is generated if the magnetic field magnitude changes from the initial value B_0 to the final value B ,

$$A = \left(\frac{B}{B_0} \right)^{1/3} (A_0 + 1) - 1$$

where A_0 is the initial anisotropy (Olson and Lee 1983). I did not read carefully about the original paper, but I feel like this is related to Equation 14.22 and betatron acceleration.

22.4.6.2 Continuous EMIC waves events

Continuous EMIC waves events may be seen to last for several hours in ground-based magnetometer data, often occurring during intervals while the magnetosphere remains compressed in response to a continuous period of enhanced solar wind dynamic pressure. There are two proposed physical mechanisms which may result in the generation of this type of long-lasting and continuous EMIC wave event.

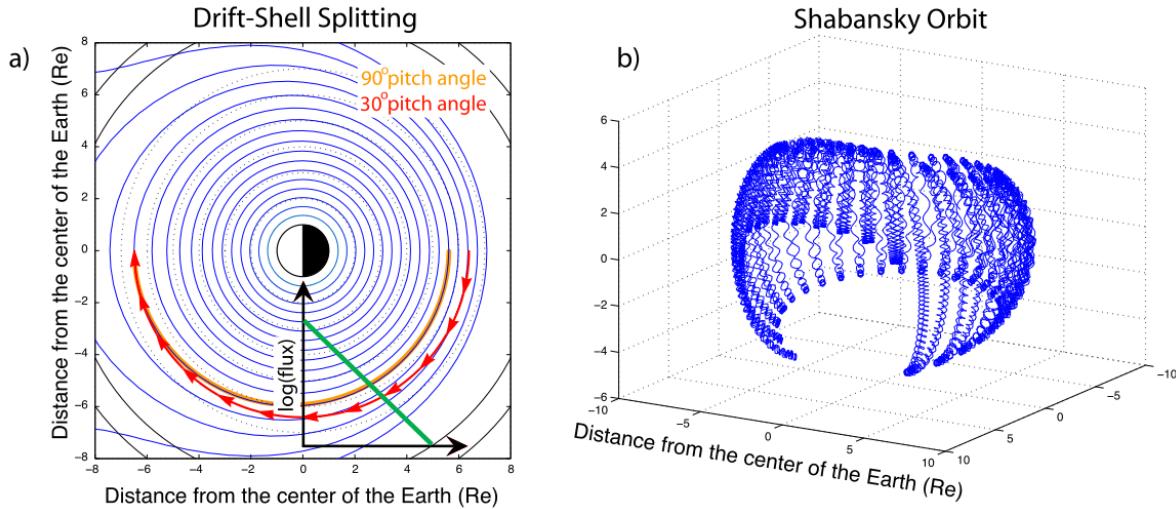


Figure 22.13: Schematic illustrating mechanisms for continuous EMIC wave event: (a) drift shell splitting and (b) Shabansky orbit [adapted from Usanova 2010b]. (a) Drift path of 30° (orange) and 90° (red) pitch angle protons from the nightside to the dayside magnetosphere; a schematic of the decreasing logarithm of proton flux (green) as a function of radial distance from the Earth is also shown. (b) Trajectory of a proton whose bounce trajectory changes from trapping between mirror points spanning the equatorial plane to trapping between two off-equatorial minimum B regions, resulting in trajectories which do not cross the equatorial plane in the dayside magnetosphere (also known as Shabansky orbits).

- Drift-Shell Splitting

Particles with different pitch-angles follow different drift paths, such that particles from different spatial locations can be brought together to create an unstable distribution. While in a symmetric dipolar magnetic field all particles drift around the Earth at the same (initial) equatorial distance, in a compressed magnetic field particles with near 90° pitch angles follow lines of constant magnetic field (conserving magnetic moment), and particles with smaller pitch-angles move along more circular orbits. This is known as *drift-shell splitting*, which is illustrated by the schematic in Figure 22.13 a.

Figure 22.13 a shows example drift paths of particles with pitch angles of 30° and 90° as they drift from the nightside to the dayside in the magnetosphere. The drift paths of these two particles intersect at the dayside subsolar magnetosphere, while the initial starting point of their trajectories was located in the nightside magnetosphere in two different locations: for a 30° pitch angle particle further away from the Earth and for a 90° pitch angle particle closer to the Earth. In a situation where the flux of particles decreases with radial distance from the Earth as shown in Figure 22.13 a the 90° pitch-angle particles drift closer to the Earth, as compared to smaller pitch-angle particles, will have a larger flux. This leads to an enhancement of the perpendicular temperature with respect to the parallel temperature and makes the distribution temperature anisotropic (with $T_\perp > T_\parallel$), and thereby potentially unstable to EMIC waves. The drift trajectory splitting is more prominent for particles at higher L-shells and for higher degrees of compression. Therefore, for the inner magnetosphere this effect may be important only during intervals of highly increased solar wind dynamic pressure.

- Shabansky Orbit

The outer L-shells of the compressed dayside magnetosphere can have two off-equatorial minima in magnetic field strength, which are more pronounced during strong compressions, and can be regarded as a compression-related feature of the reconfigured dayside magnetospheric magnetic field. In such a field topology, particles with 90° pitch angles continue to drift through the equator, but particles with lower pitch angles can mirror at high latitudes without passing through the equator, executing so-called “Shabansky” orbits [Shabansky, 1971]. As particles reach the $\partial B/\partial s < 0$ region near the dayside magnetopause, where S is distance along the magnetic field, they escape off-equator and execute repeated bounces and are trapped at mid-latitudes. This is demonstrated in Figure 22.13 b which shows the full trajectory of a proton drifting from the nightside to the dayside magnetosphere. These so-called Shabansky regions can lead to the formation of enhanced temperature anisotropy in two different locations: in the equatorial plane, where there is the absence of the lower pitch angle particles; and second, in the trapped off-equatorial regions where particles’ pitch angle undergoes a net shift towards a higher pitch angle. (???) As a result, the ion perpendicular temperature increases relative to the parallel temperature, with the resulting temperature anisotropy potentially giving rise to the growth of EMIC waves, especially in the off-equatorial regions. Similar to drift shell splitting, when magnetosphere is strongly compressed by the solar wind the regions supporting Shabansky orbits may move further inward, reaching locations inside geosynchronous orbit, to become an important source of dayside temperature anisotropy in the inner magnetosphere.

22.5 Bow Shock

Bow shock is the shock (Chapter 21) resulted from the interaction of supersonic, super-Alfvénic solar wind with the magnetosphere of an astrophysical object.

22.5.1 Foreshock

The word “foreshock” is borrowed from geophysics for an earthquake that occurs before a larger seismic event (the “mainshock”) and is related to it in both time and space. In space physics, *foreshock* is the region upstream from the bow shock that is magnetically connected to the bow shock and contains both solar wind plasma and also charged particles backstreaming from the bow shock. It is typically associated with quasi-parallel shocks, as it is much easier for charged particles to move freely along the normal direction of the shock if it is aligned with \mathbf{B} . If their speed relative to the ambient plasma is fast enough, plasma shocks without Coulomb collisions cannot dissipate all the incoming plasmas; some of them have to “return”. The counter-streaming between the foreshock plasma and the incoming plasma is very unstable and thus excites many types of waves.

Figure 22.14 demonstrate the basic geometry of Earth’s foreshock. The slanted straight lines are the magnetic field embedded in the solar wind that is flowing downward in the figure and encounters the curved bow shock (green curve). The angle between the magnetic field and the normal direction at the shock is denoted as θ_{Bn} . The portion of the bow shock where $\theta_{Bn} > 45^\circ$ (i.e., the region to the left of the Earth in the figure) is termed the “quasi-perpendicular shock”, while the region to the right, $\theta_{Bn} < 45^\circ$, is the “quasi-parallel shock” that extends all the way down to $\theta_{Bn} \sim 0^\circ$ - the “parallel shock”. At the shock near $\theta_{Bn} \sim 90^\circ$, some of the incoming solar wind electrons, which have small gyroradii, are reflected and propagate back upstream creating the region known as the “electron foreshock”. Some of the solar wind ions impinging on the bow shock near $\theta_{Bn} \sim 90^\circ$ are also reflected, but because they have larger gyroradii, quickly gyrate into the downstream. At somewhat smaller θ_{Bn} , but still associated with the quasi-perpendicular bow shock, a fraction of the solar wind ions are also reflected and travel back upstream along the magnetic field as well as being convected in the solar wind ($\mathbf{V}_{E \times B}$ in the figure) to form the quasi-perpendicular portion of the ion foreshock. Farther to the right of the ion foreshock boundary shown in Figure 22.14, ions originating at the shock at even smaller θ_{Bn} , e.g., on the quasi-parallel portion of the shock, also propagate upstream (dots in the figure), generating larger amplitude waves in the magnetic field (blue wavy lines), to form the quasi-parallel portion of the ion foreshock. These backstreaming ions may have been from the solar wind and reflected at the shock or could have leaked out of the magnetosheath. This process extends deep in the foreshock to $\theta_{Bn} \sim 0^\circ$, where the shock normal is essentially parallel to the solar wind magnetic field.

The dominant wave mode in the ion foreshock is produced by the *ion-ion beam right-hand instability* and has a typical period around 30 s in Earth’s foreshock. Their exact period however varies significantly depending on the solar wind conditions, in particular the IMF strength, between 10 and 80 s. One easy way to think about this is by remembering the gyrofrequency $\omega_c = qB/m$: the beam instabilities are associated with the cyclotron resonance, and the gyrofrequency is directly related to B . Their wavelength is of the order of $1 R_E$ and they are left-hand polarized in the spacecraft frame. Their intrinsic polarization is however right-handed, indicative of a fast mode. The polarization reversal is due to the waves attempting

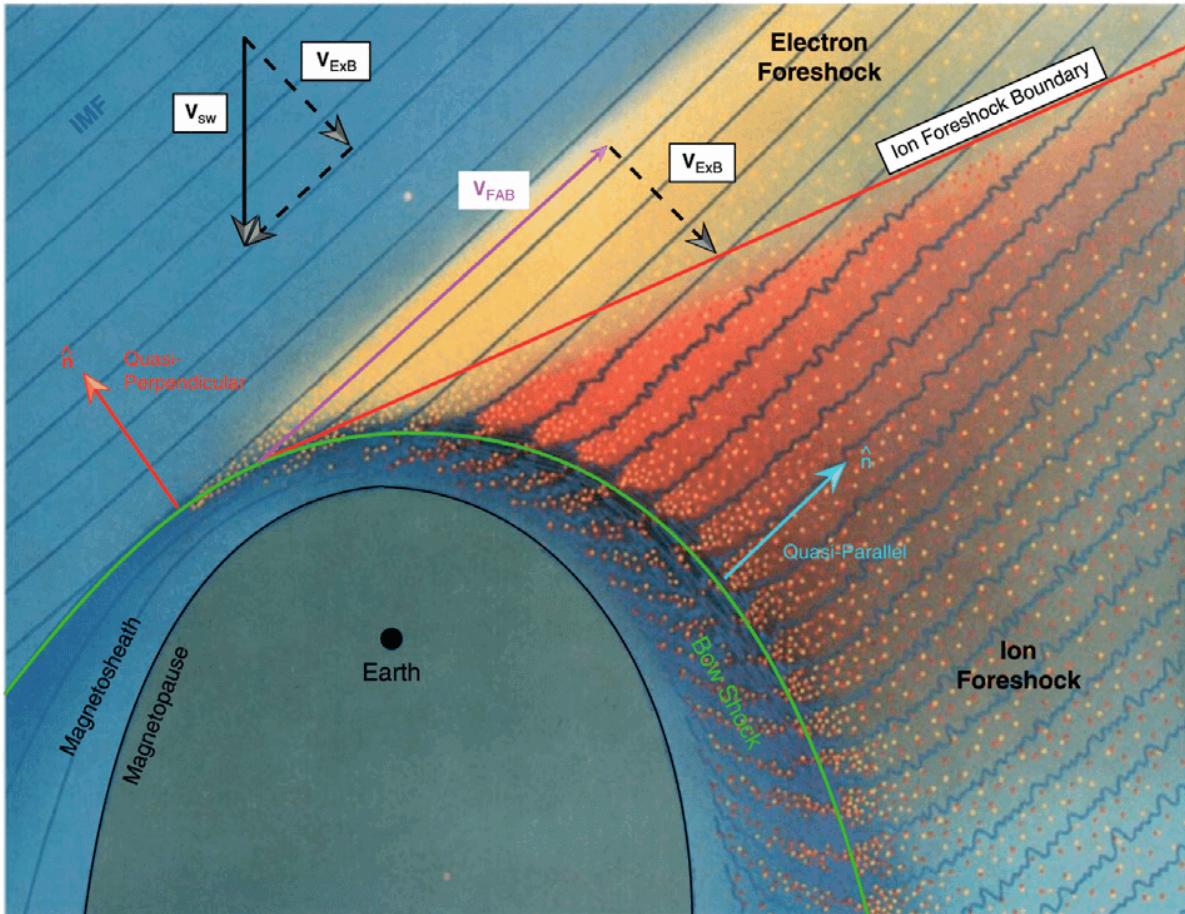


Figure 22.14: Foreshock configuration showing incident solar wind magnetic field (blue lines), impinging on the bow shock (green curved line), producing backstreaming ions (yellow dots), leaked magnetosheath ions (red dots), and upstream waves (wavy blue lines). The ion foreshock boundary is shown as the red line (Wilson III 2016). The foreshock boundary is the reflected ion trajectory tangent to the shock surface, taking into account the combination of a field-aligned motion and the solar wind convection electric field.

to propagate sunward, while they are effectively carried earthward by the faster solar wind flow. The fast mode is also consistent with the observed positively correlated magnetic field strength and plasma density fluctuations.

The ion-ion right-hand resonant instability is driven by ion beams that are cold and fast, streaming back from the bow shock along the magnetic field, characteristically produced in the quasi-perpendicular portion of the foreshock. During the initial growth of the instability, the excited electromagnetic waves have group velocities close to the beam speed, so that the field energy propagates with the beam. The waves pitch-angle scatter the beam reducing the free energy.

Electromagnetic ion beam instabilities with somewhat different properties are also found deeper in the foreshock on magnetic field lines that connect to the nearly parallel portion of the bow shock. Here the ion beams are slower and hotter and instead of being generated by reflection at the shock, are more likely to result from heating and scattering of the cold beam ions, processes related to the unsteady nature of the parallel shock, or from magnetosheath ions that leaked out into the foreshock. Other types of ion beam instabilities arise elsewhere in space, e.g., in the plasma sheet boundary layer upstream of slow-mode shocks in the magnetotail and in the vicinity of comets.

From spacecraft missions ISEE, Cluster and MMS, we have found two main populations of backstreaming ions, termed *reflected* and *diffuse*. The reflected ions have a sharply peaked energy spectrum and relatively collimated flow coming from the bow shock along the interplanetary magnetic field with number densities $\sim 1.5\%$ of the solar wind. In contrast, the diffuse ions have a much flatter energy spectrum and broad angular distributions, with lower density $\sim 0.7\%$ of the solar wind. These two populations occur in separate regions of the foreshock, with reflected ions in the quasi-perpendicular portion and the diffuse ions in the more nearly parallel portion of the foreshock, suggesting different mechanisms of origin and acceleration. In between the reflected and diffuse ion populations, were *intermediate* ions that had a crescent shape in velocity space, which corresponds to the gyro-phase bunched ions and pitch-angle scattering by the generated waves. The reflected ions were so named, because backstreaming beams of ions were observed traveling upstream in a direction determined by the interplanetary magnetic field and the convection velocity ($\mathbf{V}_{E \times B}$), consistent with simple models of specular reflection at the bow shock.

Observations from ISEE also showed that there were usually weak waves occurring with reflected beam ions, while with diffuse ions large amplitude, compressive waves often occurred. Obliquely propagating waves were also observed, often in the form of steepened waves, termed *shocklets*. In addition, deeper in the foreshock Alfvén waves were also generated in association with the diffuse ions. The location in the foreshock where the low frequency waves begin to appear, is called *ULF foreshock boundary*, near the ion foreshock boundary. Another wave produced deep in the foreshock near the quasi-parallel bow shock is termed *short large-amplitude magnetic structures* (SLAMS).

The linear theory of beam instabilities in magnetized plasmas is developed by Peter Gary. He numerically solved the complete dispersion relation for both electrostatic and electromagnetic waves excited by various sorts of free energy: i.e., beams of electrons or ions, temperature anisotropies in either the background or beam species, etc.⁷ In (Gary, Gosling, and Forslund 1981), he showed how a weak ($\sim 1\%$ density) beam initially produces a low level of waves with growth times > 25 s. Because the growth rate of the waves is small, the ion beam propagates a long distance upstream from the bow shock before the waves grow to large amplitude. But in the process of propagating upstream, the beam ions and the waves are convected deep into the quasi-parallel portion of the foreshock, where they are slowed. The beam ions loses parallel momentum as they are pitch-angle scattered by the waves, producing intermediate velocity distributions (kidney shaped). This loss of energy and momentum by the beam, and the waves, causes the solar wind to be deflected and slowed, as observed by Bame et al. (1980). Since the compressibility (i.e., the density fluctuations computed from linear theory) of oblique waves⁸ is significantly larger than for parallel propagation, oblique waves can grow and produce density fluctuations in agreement with observations.

In Gary et al. (1984), four linear electromagnetic ion instabilities are compared (Section 21.6.1). Numerical solutions of the linear dispersion relation are presented to show that for parallel propagation the righthand resonant instability has a lower instability threshold than the non-resonant instability unless the beam speed or beam density is sufficiently large. For hot ion beams a left-hand resonant instability is also excited, and if there is a large temperature anisotropy in the beam ($T_{b,\perp} \gg T_{b,\parallel}$), a left-hand ion cyclotron instability is also possible. Therefore, the deterministic factors for the driven instabilities would be

- beam density n_b
- beam velocity V_b (w.r.t. V_A)
- beam temperature T_b
- beam pressure anisotropy

In Gary, Thomsen, and Fuselier (1986), linear theory of both the right-hand resonant and non-resonant instabilities driven by an energetic, cool ion beam is used to calculate the phase angle between the fluctuating velocity of the beam ions relative to the fluctuating magnetic field in order to determine whether gyrophase bunching of the beam ions is observable. Distinct gyrophase-bunched ions have been observed in the foreshock by ISEE. Peter's linear calculations showed that for the non-resonant mode, the phase angle is $\sim 0^\circ$, while for the resonant mode it is $\sim 90^\circ$ for the most unstable wavenumber. Both observations and simulations support the right-hand resonant instability for generating the observable gyrophase bunching for the ion beam.

Despite all these achievements, a major problem remains — the observed ULF waves are primarily propagating oblique to the magnetic field, contrary to linear theory that says the parallel mode has the largest growth rate and thus in time will become the dominant mode.

⁷One important assumption is that all species are Maxwellians.

⁸I think parallel waves here means Alfvén waves, whereas oblique waves means fast waves.

Measurements by Cluster (Eastwood et al. (2005)) indicate a mean angle of 21° . An even more interesting suggestion is given later by Strumik et al. (2015). In their foreshock simulations they assume that the solar wind is essentially radial. They find that while parallel unstable modes dominate, oblique waves are also generated, and when properly averaged over phase space density, the “average” propagation direction is $\sim 20^\circ$, consistent with Eastwood et al. (2005).⁹

22.5.2 Transmission of Waves

An interesting question arises about the transmission of fast waves through the bow shock into the downstream. A list of possible mechanisms are summarized in [?@fig-foreshock-wave-transmission](#). (Turc et al. 2022) argued that the earthward magnetosonic disturbances can then propagate to and disturb the magnetopause, whereas shock reformation and mode conversion play only minor roles.

As usual in plasma physics, accompanying the ion foreshock we have electron foreshock at a much smaller scale.

The formation of ion/electron foreshocks requires counter-streaming of ions and electrons, respectively. Since there is only one velocity for each species in the electron/ion two-fluid equations, it is not possible to trigger this instability in a multi-moment model. Theoretically we may have an extremely simplified case where the ion counter-streaming is represented by two ion fluids, but there are very few studies on this topic. The minimum requirement may be gyrokinetics.

22.6 Magnetosheath

22.6.1 Jets

Magnetosheath jets are regions of enhanced dynamic pressure. Think of jets as raindrops in the magnetosheath falling on the magnetopause. They are widely associated with the quasi-parallel magnetosheath, suggesting that their origin is tied to the interactions between the foreshock and the bow shock.

The formation of magnetosheath jets may be related to both external and internal drivings.

- External
 - Solar wind pressure pulses
 - Rotational discontinuities
- Internal

⁹So my feeling is that they tried to argue that the 20° deviation is purely a statistical bias, but not physics.

- Local ripples from foreshock turbulence could lead to the refraction and funneling of plasma.
- Short large-amplitude magnetic structures (SLAMS), which are steepened foreshock fluctuations in a short time-scale but large spatial scale.

22.7 Magnetosphere

22.7.1 Magnetopause

The magnetopause is the abrupt boundary between a magnetosphere and the surrounding plasma. The collision frequency of particles in the plasma in the interplanetary medium is very low and the electrical conductivity is so high that it could be approximated to an infinite conductor. A magnetic field in a vacuum cannot penetrate a volume with infinite conductivity. Chapman and Bartels (1940) illustrated this concept by postulating a plate with infinite conductivity placed on the dayside of a planet's dipole as shown in [?@fig-magnetopause-schematic](#). The field lines on the dayside are bent. At low latitudes, the magnetic field lines are pushed inward. At high latitudes, the magnetic field lines are pushed backwards and over the polar regions. *The boundary between the region dominated by the planet's magnetic field (i.e., the magnetosphere) and the plasma in the interplanetary medium is the magnetopause.* Since the solar wind is continuously flowing outward, the magnetopause above, below and to the sides of the planet are swept backward into the geomagnetic tail as shown in the artist's concept. The region (shown in pink in the schematic) which separates field lines from the planet which are pushed inward from those which are pushed backward over the poles is an area of weak magnetic field or day-side cusp. Solar wind particles can enter the planet's magnetosphere through the cusp region. Because the solar wind exists at all times and not just times of solar flares, the magnetopause is a permanent feature of the space near any planet with a magnetic field.

If one assumed that magnetopause was just a boundary between a magnetic field in a vacuum and a plasma with a weak magnetic field embedded in it, then the magnetopause would be defined by *electrons and ions penetrating one gyroradius into the magnetic field domain*. Since the gyro-motion of electrons and ions is in opposite directions, an electric current flows along the boundary.

22.7.1.1 Estimation of the standoff distance

The location of the magnetopause is determined by the pressure balance between the internal magnetic field and external solar wind. As a first order estimation of the standoff distance to the magnetopause, we neglect the internal magnetospheric thermal pressure and let the

dynamic ram pressure from the solar wind being equal to the magnetic pressure from the Earth's magnetic field:¹⁰

$$(\rho v^2)_{\text{sw}} \approx \left(\frac{4B(r)^2}{2\mu_0} \right)_{\text{m}}$$

where ρ and v are the density and velocity of the solar wind, and $B(r)$ is the magnetic field strength of the planet in SI units.

Since the dipole magnetic field strength varies with distance as $1/r^3$ the magnetic field strength can be written as $B(r) = B_0/r^3$, where B_0 is the planet's magnetic moment, expressed in $[\text{T} \cdot \text{m}^3]$.

$$\begin{aligned} \rho v^2 &\approx \frac{2B_0^2}{r^6 \mu_0} \\ r &\approx \left(\frac{2B_0^2}{\mu_0 \rho v^2} \right)^{1/6} \end{aligned}$$

The distance from Earth to the subsolar magnetopause varies over time due to solar activity, but typical distances range from $6 - 15R_{\oplus}$. Empirical models (e.g. (Shue et al. 1997)) using real-time solar wind data can provide a real-time estimate of the magnetopause location.

Table 22.3 summarizes the basic sizes of each planet in the solar system. Venus and Mars do not have a planetary magnetic field and do not have a magnetopause. The solar wind interacts with the planet's atmosphere and a void is created behind the planet. In the case of the Earth's moon and other bodies without a magnetic field or atmosphere, the body's surface interacts with the solar wind and a void is created behind the body, often called wake. All planets may show the effect of inductance, depending on the conductivity of the ionosphere or ionosphere-like sphere. The interaction of IMF with a conductor induces current on the surface of the conductor and create an induced magnetic field which opposes the external field from penetrating into the conductor. This is a major contribution at Mars, a minor contribution at Mercury and Ganymede, and somehow being negligible at other large planets.

Table 22.3: Solar system magnetopauses

Planet	Magnetic Moment ¹¹	Standoff Distance ¹²	Variance ¹³
Mercury	0.0004	1.5	0
Venus	0	0	0
Earth	1	10	2
Mars	0	0	0
Jupiter	20000	42	25
Saturn	600	19	3
Uranus	50	25	0

¹⁰The reason for the factor of 4 is because the magnetic field strength just inside the magnetopause is twice the dipole value for a planar magnetopause.

Planet	Magnetic Moment	Standoff Distance	Variance
Neptune	25	24	1.5

22.7.2 Transport of mass, momentum, and energy

Transport of mass, momentum, and energy at the magnetopause boundary drives much of the magnetospheric dynamics and is thus an important topic in magnetospheric physics. The amount of solar wind plasma and energy that enters the magnetosphere is regulated by the orientation of the interplanetary magnetic field (IMF): during southward IMFs, the major process is considered to be magnetic reconnections (Section 13.8), while during northward IMFs, the processes include cusp reconnection, plasma mixing/reconnection in Kelvin-Helmholtz vortices, and wave-particle diffusive processes in large-amplitude ULF waves at the magnetopause boundary ([?@sec-klr](#)).

One of the more common spectral features at the magnetopause boundary is the sharp transition in wave polarization from compressional, e.g., because of foreshock waves of the quasi-parallel bow shock, to transverse (Alfvén) waves from the magnetosheath to the boundary layers. Mode conversion from compressional to Alfvén modes through the magnetopause can efficiently transport and heat plasma across the boundary. Multipoint measurements have verified that the dispersion of the broadband waves is consistent with the kinetic Alfvén waves (KAWs) [Chaston+, 2007, 2008]. KAWs also provide a natural explanation for the observed dawn-dusk asymmetry in plasma entry during northward IMF because they result from mode conversion of compressional foreshock waves, which typically bathe the dawn flank for the typical Parker spiral configuration.

According to linear theory, when an MHD fast mode compressional wave propagates across in an inhomogeneous plasma, it is coupled with the shear Alfvén wave at the local Alfvén resonance $\omega = k_{\parallel} V_A$, with ω being the wave frequency. This process can be very efficient when the fast wave propagates into a region where there is a sharp increase in the Alfvén velocity such as at the magnetopause boundary where the magnetic field increases and density decreases. Typically, the Alfvén velocity across the magnetopause increases by a factor of 10 such that an entire decade (in frequency) of wave power can be captured and localized in the boundary layer leading to massive particle transport. In the MHD description, the coupling occurs where the frequency matches the continuous spectrum and the wave becomes singular corresponding to a pileup of compressional wave energy.

Singular behavior occurs at either the Alfvén resonance, $\omega^2 = k_{\parallel}^2 v_A^2$, or at the sound resonance, where the compressional wave couples with the Alfvén or sound wave that propagates only

¹¹Normalized by Earth's magnetic moment.

¹²Normalized by planet's radii.

¹³in planet radii. The magnetosphere varies mainly in response to solar wind dynamic pressure and interplanetary magnetic field.

along the magnetic field. In higher frequency cases, the Alfvén resonance condition is modified because of the finite ion Larmor radius effects (Equation 16.27), where the Alfvén resonance singularity can be removed by including non-MHD effects such as electron inertia or ion Larmor radius corrections in Equation 16.26. It can be shown that kinetic effects naturally arises from the two-fluid/kinetic theory, including the coupling to KAWs. For more equations, see Section 16.4.

At the resonance point, the fast wave solution is coupled to the KAW solution, and transversely polarized Alfvén waves are expected to be generated and radiate away from the mode conversion location. Because the group velocity of these waves is much smaller than the compressional wave, the amplitude of the transverse fluctuations is typically much larger than the amplitude of the compressional driver, consistent with magnetopause observations.

As been discussed in Chapter 16, when the azimuthal wave number m , or k_y in the box model is zero, there should be no mode coupling. However, this is not true if the *finite-frequency effect* is considered (Section 16.4), as first noted by (1979).

22.7.3 Dungey Cycle

Dungey (1961) was the first to propose a cycle of magnetospheric convection driven by magnetic reconnection at the dayside magnetosphere. Magnetic field connected on the dayside of Earth is transported by the solar wind to the night side where it forms a long tail behind the Earth. This transfer of flux to the nightside forces the magnetosphere to undergo systematic changes in configuration that eventually lead to nightside magnetic reconnection, which returns flux to the dayside along the flanks of the magnetosphere via the different response modes.

22.7.4 Vasyliunas Cycle

The Jovian equivalent to the Dungey cycle of the Earth's magnetosphere, where the centrifugal force plays a critical role in affecting the plasma convection.

22.7.5 Low-Latitude Boundary Layer

The Earth's *low-latitude boundary layer* (LLBL) is the region where magnetosheath and magnetospheric plasmas are mixed along the magnetospheric side of the low-latitude magnetopause. It has been suggested that three processes — magnetic reconnection at high-latitudes, the Kelvin-Helmholtz instability, and Kinetic Alfvén waves — primarily contribute to the formation, all of which have been observed by in-situ satellites inside the LLBL. Recent studies further suggested that the couplings of these processes effectively enhance the formation rate.

23 Ionosphere

Ionosphere is like a transition region from neutral gas to plasma. Therefore, collisions as well as kinetic effects gradually become important as go down from space. Section 9.8 presents some basic physics related to the collisional effect in the ionosphere.

23.1 Current Systems

Electric fields associated with the viscous interaction and reconnection-driven plasma flow are transmitted along field lines to the ionosphere where they drive currents through the ionosphere. The high latitude FAC are known as *region-1 currents*, while the lower latitude currents are called *region-2 currents*. The ionospheric current connecting these two FAC systems flows parallel to the projected electric field and is called the *Pedersen current*. Since the Pedersen current flows through a resistive ionosphere in the direction of the electric field it causes Ohmic heating $\mathbf{j} \cdot \mathbf{E} > 0$. The two shells of FAC form a solenoid so that the magnetic perturbations they create are confined to the region between them and cannot be detected on the ground. Because of the interaction of the motion of ionized gases with the neutral atmosphere, the ions and electrons undergo different drift motion. Ions, due to higher collision rates with the neutral atmosphere, have a component of motion perpendicular to the $\mathbf{E} \times \mathbf{B}$ drift direction, while the electrons generally follow $\mathbf{E} \times \mathbf{B}$. The non-dissipative ($\mathbf{j} \cdot \mathbf{E} = 0$) Hall current flows at right angles to both the electric and magnetic fields. The magnetic effects of the Hall current driven by the region-1 and 2 currents can be observed on the ground and are known as the DP-2 current system.¹

In the context of ionospheric energy dissipation, the different current systems can be classified into two types based on how energy is processed.

1. If the ionospheric motion and dissipation is coupled directly to the solar wind, the system is said to be “directly-driven”. The directly-driven process manifests itself as the DP-2 (two cell pattern) ionospheric current system.
2. If magnetic energy is first stored in the tail lobes and then released some time later, driving additional convection and field-aligned and ionospheric currents, it is called “unloading”, and is associated with the DP-1 (substorm current wedge) current system.

¹Region 1 & 2 currents are currents in the magnetosphere, not ionosphere.

Both processes cause precipitation of charged particles that also deposit energy in the atmosphere.

The characteristic feature of the driven DP-2 current system is the existence of the eastward and westward electrojets flowing toward midnight along the auroral oval. Rough measures of the strength of these currents are the *auroral upper* (AU) and *auroral lower* (AL) indices. These are respectively the largest positive northward magnetic perturbation (H) measured on the ground under the eastward electrojet by any magnetic observatory in the afternoon to dusk sector, and the largest negative southward perturbation measured in the late evening to morning sector. Both AU and AL begin to grow in intensity soon after the IMF turns southward and dayside reconnection begins. The characteristic feature of the unloading DP-1 current system is the sudden development of an additional westward current that flows across the bright region of the expanding auroral bulge. This is the ionospheric segment of the substorm current wedge. The onset of this current is recorded in the AL index as a sudden decrease, corresponding to an increase in intensity of the westward current.

23.2 M-I Coupling

Ionospheric properties, principally conductivity, provide boundary conditions for magnetospheric convection, and the ionosphere is often treated as a passive part of the system. Especially during substorms, however, the boundary conditions change in a time-dependent and spatially localized fashion, allowing ionospheric feedback that can alter the magnetospheric dynamics. The coupling from the ionospheric perspective differs primarily in that the ionospheric conductance is anisotropic due to the influence of the neutral atmosphere, involving Hall as well as Pedersen conductivity. These conductivities are altered both by the connecting currents and the precipitating electrons associated with upward field-aligned currents, which increase Pedersen conductivity and field line tying. An important role is also played by field-aligned electric fields, set up locally, primarily in upward field-aligned current regions, which are the cause of auroral intensifications and, specifically auroral arcs.

For magnetosphere simulations, the simplest inner boundary is treated as a conducting sphere ($\mathbf{B}_{\text{normal}} = 0$, $\mathbf{E} = \mathbf{v} = 0$). However, this is often not realistic enough to reveal the nature. In magnetosphere simulations, the location of tail main reconnection site will be closer to the Earth by simply applying a conducting boundary. The next level extension is to employ a magnetospheric-ionospheric electrostatic coupling model. This means that we seek nonzero \mathbf{E} and \mathbf{v} at the inner boundary. The inner boundary, where the MHD quantities are connected to the ionosphere, is taken to be a shell of radius r_{in} (e.g. $r_{\text{in}} \sim 3 R_E$). The ionosphere locates at $r_{\text{ion}} \sim 1000 \text{ km} \sim 0.15 R_E$. Ideally r_{in} shall be as close to r_{ion} as possible, but typically it is restricted by computational limitations, such as extraneously high Alfvén speeds and very large \mathbf{B} field gradients closer to the Earth. Inside this shell we do not solve the governing equations (MHD/PIC/Vlasov), but assume a static dipole field. The important physical processes within

the shell are the flow of *field-aligned currents* (FACs) and the closure of these currents in the ionosphere. At each time step,

1. The magnetospheric FACs are mapped along the field lines from the inner boundary to the ionosphere assuming $j_{\parallel}/B = \text{const.}$, which are the input to the ionospheric potential equation (Raeder, Walker, and Ashour-Abdalla 1995)

$$\nabla \cdot (\boldsymbol{\Sigma} \cdot \nabla \Phi) = -j_{\parallel} \sin I \quad (23.1)$$

where $\boldsymbol{\Sigma}$ is the conductance tensor, Φ is the electric potential, j_{\parallel} is the mapped FAC density with the downward considered positive and corrected for flux tube convergence, and I is the inclination of the dipole field at the ionosphere, $\sin I = \cos(\frac{\pi}{2} - I) = \hat{b} \cdot \hat{r}$. The derivation from the charge conservation can be found in Section 9.8.7. There is another form derived by [Wolf 1983]:

$$\nabla_{\perp} \cdot \begin{pmatrix} \sigma_P / \cos^2 \delta & -\sigma_H \cos \delta \\ \sigma_H / \cos \delta & \sigma_P \end{pmatrix} \cdot \nabla_{\perp} \Phi = j_{\parallel} \cos \delta \quad (23.2)$$

where δ is the magnetic field dip angle:

$$\cos \delta = -2 \frac{\cos \theta}{\sqrt{1 + 3 \cos^2 \theta}}$$

for the northern hemisphere, where θ is the polar angle (magnetic colatitude). I DON'T KNOW THE RELATION BETWEEN THESE TWO!

2. Equation 23.1 is solved on the surface of a sphere $r = r_{\text{ion}}$. Commonly there are two types of boundary conditions: (1) $\Phi = 0$ at the equator (Raeder, Walker, and Ashour-Abdalla 1995), or (2) constant potential at or near the low-latitude boundary (e.g. LFM, BATSRUS). From here, one can either choose a static analytic model of Hall and Pederson conductance that accounts for multiple physics, or simply adopt a uniform Pederson conductance, or the height-integrated conductivity, $\Sigma_p = 5$ Siemens, while the Hall conductance Σ_H is assumed to be zero. The latter one is simplified to solve

$$\nabla^2 \Phi = -j_{\parallel} \sin I / \Sigma_p$$

A more realistic conductance requires considering EUV and diffuse auroral contributions (???) as well. The solar EUV contribution to $\boldsymbol{\Sigma}$ is considered constant in time, but naturally it varies with the solar zenith angle. For example, the empirical formulas by [Moen and Brekke 1993] can be used. The solar EUV radiation is approximated by the 10.7 cm radio flux (commonly known as *F10.7*), a widely used proxy solar UV activity, whose standard values is taken to be $100 \times 10^{-22} \text{ W/m}^2$.

The total conductance can be then expressed as

$$\Sigma_{P,H} = \sqrt{(\Sigma_{P,H}^{e-})^2 + (\Sigma_{P,H}^{\text{UV}})^2}$$

This is because $\sigma_{P,H} \propto n_e$, which in a stationary state is proportional to the square root of the production rate, and it is the production rates that can be summed linearly. (???)

Does this look well? Not yet. We know that while the high-latitude ionospheric convection is driven by the solar wind and magnetosphere interaction, at lower latitudes atmospheric neutral winds start to dominate. The next level approximation needs to take this into account. Because there is a gap between r_{ion} and r_{in} , the ionospheric footprint of their grid has a low-latitude boundary somewhere in the midlatitudes, e.g., 45° when $r_{\text{in}} \sim 2 R_E$ (Figure 23.1). Global magnetospheric models, unless they are fully coupled to models of the inner magnetosphere and the ionosphere, lack details of the ionospheric convection at latitudes equatorward of their low-latitude ionospheric boundary. To some extent, such details can be translated to the global model via the low-latitude boundary condition used to solve Equation 23.2. The easy way is to set the ionospheric potential to zero everywhere on the boundary. This corresponds to no flow across the boundary in the ionosphere or the inner boundary of the magnetosphere simulation in the equatorial plane. The choice of this boundary condition is usually justified by the argument that it helps to shield the inner magnetosphere from the cross-tail electric field.

(Merkin and Lyon 2010) tested three different boundary conditions for the potential equation:

- STANDARD: Dirichlet, the potential at the low-latitude boundary is set to zero.
 - NEUMANN: the electric field component normal to the low-latitude boundary was set to zero. This condition requires all ionospheric plasma to move normal to the boundary.
 - LOWERBC: Dirichlet, but the location of the low-latitude boundary was moved to 2° above the equator, thus allowing the plasma to move across the magnetosphere inner boundary. (???) Equation 23.3 is singular at $\theta = \pi/2$, which is why the calculation has to stop just short of the equator. (???)
3. Sparse linear algebra, GMRES together with an incomplete LU preconditioner (default for many modern solvers) are usually applied to solve the potential equation. This is generally an easy equation to solve mathematically.
 4. Once the potential equation is solved the ionospheric potential is mapped back to the r_{in} shell and used as a boundary condition for the magnetospheric flow by taking $\mathbf{v} = (-\nabla\Phi) \times \mathbf{B}/B^2$.

23.2.1 Caveats

- The mapping assumes conservation, which is not perfect. In practice $r_{\text{in}} \sim 4 R_E$ is a minimum requirement for reasonable FACs.
- Most numerical codes couples a Cartesian grid to a spherical ionosphere grid, while some couples a spherical grid to a spherical grid. For magnetosphere simulations we need a relatively simple but super fast electric potential solver, therefore structured mesh is often adopted. If a spherical grid is used, care must be taken near the pole since it is a singularity from the grid but not physics. Equation 23.2 under spherical coordinates is written as

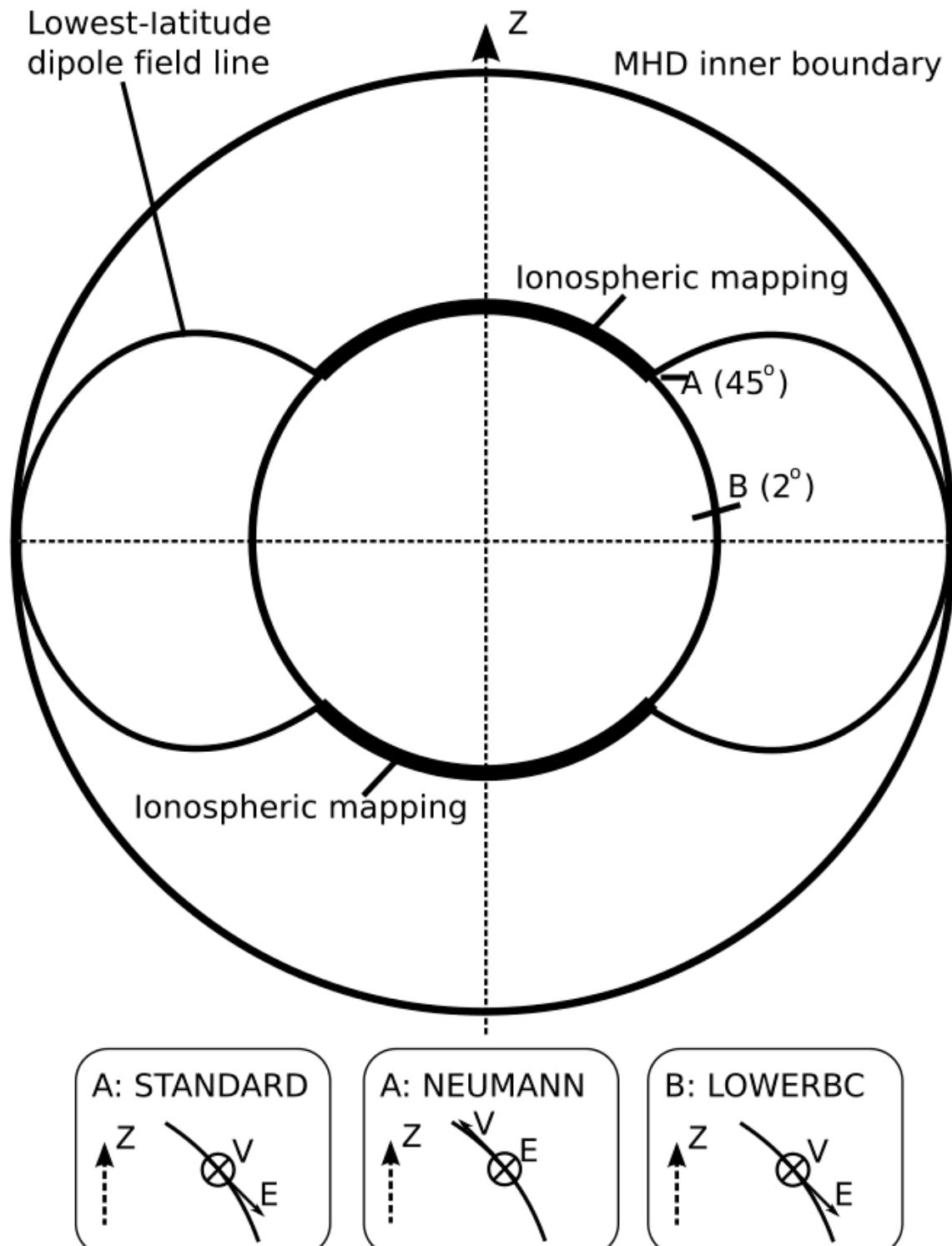


Figure 23.1: A schematic depiction of the inner boundary of the magnetosphere simulation and its ionospheric mapping. The location of the low-latitude boundary of the ionospheric grid is determined by the radius of the inner MHD boundary mapped along the dipole field from the equator. Point A (at 45°) denotes a typical location of the low-latitude boundary, where the STANDARD and NEUMANN boundary conditions are applied, while point B is where the LOWERBC boundary condition is applied. The three inset plots at the bottom depict schematically the configuration of velocity and electric field vectors with respect to the surface of the ionosphere in a meridional plane. The inset plot titles identify the type of the boundary condition and the point where it is applied. Adopted from (Merkin and

$$\begin{aligned} & \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left[\sin \theta \frac{\Sigma_P}{\cos^2 \delta} \frac{\partial \Phi}{\partial \theta} - \frac{\Sigma_H}{\cos \delta} \frac{\partial \Phi}{\partial \phi} \right] \\ & + \frac{\partial}{\partial \phi} \left[\frac{\Sigma_P}{\sin^2 \theta} \frac{\partial \Phi}{\partial \phi} + \frac{\Sigma_H}{\sin \theta \cos \theta} \frac{\partial \Phi}{\partial \theta} \right] = j_{\parallel} R^2 \cos \delta \end{aligned} \quad (23.3)$$

- Be careful about distinguishing E_{\parallel} and j_{\parallel} . $E_{\parallel} = 0$ from advection and Hall terms, but $j_{\parallel} = \nabla \times \mathbf{B} \cdot \hat{b}/\mu_0$ can be nonzero at the MHD inner boundary.
- How important it is to use a more realistic conductance model? (Merkin and Lyon 2010) shows that different BCs may give $> 10\%$ CPCP values, but I have no clue about the effect of a more complicated conductance.

23.3 Ionosphere Modeling

- What equation set is the model solving? How is the model solving them? What is being neglected?
- What species are included? How is the chemistry solved?
- Parameterizations in things such as heating, cooling, viscosity, conduction, chemistry, diffusion, collision, absorption, ionization cross sections, reaction rates...
- How are upper and lower boundaries treated? How is the pole or the open/closed field-line boundary treated?
- How are electrodynamics considered? How is the aurora specified? How is the magneto-spheric electric field imposed? Is ion precipitation considered?

For an ionosphere, a magnetic field-aligned grid is often used. This is complicated by the non-orthogonal nature of the magnetic field coordinate system. Assumptions are made.

How to solve each direction (coupled or independently)? For the ionosphere, along the field-line is treated differently than across the field-lines.

As the equations of motion are solved for, the source terms must be added. How to treat these with respect to solving in the different directions?

Build a simple 1D ionosphere model is easy (CAN I MAKE ONE MYSELF?).

Steady-state - assume $\partial \mathbf{X} / \partial t = 0$. Strangly, this is applied in situations in which the value can change on a time scale much faster than the time-step. For example, the ion velocity is often assumed to be steady-state. Ion chemistry is sometimes assumed to be steady-state.

23.3.1 Chemistry

$$\frac{\partial N}{\partial t} = S - L$$

where N is the number density of the species, S represents the sources and L represents the losses. Losses can almost always be expressed as:

$$L = RMN$$

where R is the reaction rate, M is the density of the species it is reacting with.

For a steady-state,

$$S - L = 0$$
$$N = \frac{S}{RM}$$

This is quite stable, but can be very wrong in regions of slowly changing ion densities, such as in the F-region. This is perfect for the E-region, though. It is quite easy to implement in a simple environment, but can be much more complicated as non-linear terms are included (recombination, in which M can depend on N).

An explicit time step chemistry is trivial to implement in almost all situations, but it is also the least stable, since the loss terms can become larger than the source terms and the density can quickly be driven to negative values. Subcycling can help with this, but not always.

An implicit time step chemistry is relatively stable and easier to implement than steady-state. For example in GITM, there is a blend between sub-cycling and a simplified implicit scheme that switches depending on the size of the loss term compared to the density.

Now we need to also look at the source terms. The ionization rates can be obtained from Q_{EUV} and substituting $\sigma_s^i \lambda$ instead of $\sigma_s^a \lambda$ (not in τ). [Schunk and Nagy] Chapter 8 lists a bunch of chemical equations.

After that, we write down all the sources and losses, decide a time-stepping scheme, and run the model. However, if we don't have any ion advection, the F-region will just build and build!

23.3.2 Ion Advection

In the simplest form, we assume an advection speed and let the densities advect upward and downward.

23.4 Ionosphere Waveguide

The ionosphere waveguide refers to the phenomenon in which certain radio waves can propagate in the space between the ground and the boundary of the ionosphere. Because the ionosphere contains charged particles, it can behave as a conductor. The planet operates as a ground plane, and the resulting cavity behaves as a large waveguide. This is closely related to the concept of skin depth in EM and the cutoff in wave propagation.

At Earth, extremely low frequency (ELF) (< 3 kHz) and very low frequency (VLF) (3–30 kHz) signals can propagate efficiently in this waveguide. For instance, lightning strikes launch a signal called radio atmospherics, which can travel many thousands of kilometers, because they are confined between the Earth and the ionosphere. The round-the-world nature of the waveguide produces resonances, like a cavity, which are at ~ 7 Hz.

A brief introduction can be found on [wiki](#).

23.5 Incoherent Scatter Radar

Longitudinal modes:

- Langmuir wave
- Ion acoustic wave

Cascading Langmuir turbulence

dispersive, kinetic Alfvén waves \rightarrow optical manifestation of dispersive bursts

24 Radiation Belt

A [Van Allen radiation belt](#) is a zone of energetic charged particles, most of which originate from the solar wind, that are captured by and held around a planet by that planet's magnetosphere. Earth possesses an inner belt and an outer belt.

1. the inner belt
 - MeV protons, 100 keV electrons
 - $0.2 - 2 R_E$ ($L 1 - 3$)

2. the outer belt
 - $0.1 - 10$ MeV electrons
 - $3 - 10 R_E$ (outer boundary is the magnetopause)

Our understanding of the physics mechanisms until 1990s:

1. Low energy electrons get injected into the magnetosphere from the solar wind.
2. Electrons transport towards the planet by reconnections, substorms and associated electric fields.
3. Electrons drift around the planet.
4. Magnetic fluctuations cause inward/outward diffusions.
5. Energy is gained by conservation of the 1st adiabatic invariant.
6. Loss by collisions with the atmosphere, reaching the magnetoapause, and outward radial diffusion.

During magnetic storms the magnetopause is compressed, so more electrons are lost to the magnetopause. If the mirror point is deep inside the atmosphere, charged particles will precipitate into the atmosphere and lost due to collisions. There are still many mysteries both due to lack of observations and theories. For example, the classical theory cannot explain the electron variation (intermittent injection, rapid loss) timescales!

In 1998, two new theories came out:

1. Enhanced radial diffusion
 - Solar wind flows past the magnetosphere, drives K-H instability, and in turn triggers ULF waves.
 - The ULF waves trigger FLR (Chapter [16](#)).

- The outcome is a faster radial diffusion process?
2. Wave acceleration
 - Substorms, convection, plasma instabilities.
 - Waves accelerate electrons to MeV.

The wave acceleration, or more specifically, wave-particle interaction, has been widely explored both theoretically and compared with observation data at Earth, Jupiter, and Saturn.

24.1 Waves in the Radiation Belt

- Lower hybrid waves: ???
- Whistler waves (Section 10.6)
- Chorus waves
 - 1-5 kHz ($0.1 - 1.0 f_{ce}$), well-defined narrow band
 - highly nonlinear
 - not generated by lightning, but by natural plasma instabilities (same for hiss).
 - electron anisotropy
 - loss cone scattering
- Hiss waves
 - broadband structureless signal in the plasmasphere and resembles audible hiss

Before 1990s, waves in the radiation belts are known for transferring energy from charged particles, which act as a loss process to the particles. Richard Thorne, together with his colleagues, proposed that they could also be the sources of energetic particles ([Horne & Thorne], 1998, 2003, 2005a,b). If that is the case, *the wave-particle interactions must break the adiabatic invariants*.

The theory of wave-particle interaction starts with hot plasma kinetic theory (Section 12.10.1). When the denominator goes to 0, the resonance condition is fulfilled. From observations, we learned that the wave frequencies are smaller than the gyro-frequencies. How can we have resonance then? The answer is, Doppler shift needs to be taken into account.

24.1.1 Sources of Waves

resonance can generate current => this current can generate higher frequency radiations?

Whistler waves can be explained by the linear theory. However, chorus and hiss waves are highly nonlinear and thus cannot be explained by a linear theory. We conjecture that they are caused by natural plasma instabilities, but we still have little idea what exactly these instabilities are.

As a rough physical picture, during the cyclotron resonance:

1. Waves diffuse source electrons into loss cone => electron loss and wave growth.
2. Waves diffuse trapped electrons => energy diffusion leads to electron acceleration.

24.2 Modeling

3D Fokker-Planck diffusion model¹ (e.g. (Glauert, Horne, and Meredith 2014)) has been built to model the radiation belt electrons. In the Earth's radiation belts, the evolution of the phase-averaged phase-space density $f(p, r, t)$ can be described by a diffusion equation (see also Equation 12.19):

$$\frac{\partial f}{\partial t} = \sum_{i,j} \frac{\partial}{\partial J_i} \left[D_{ij} \frac{\partial f}{\partial J_j} \right] \quad (24.1)$$

Here D_{ij} are diffusion coefficients and J_i are the action integrals, $J_1 = 2\pi m_e \mu / |q|$, $J_2 = J$, and $J_3 = q|\phi|$, where μ , J , and ϕ are the adiabatic invariants of charged particle motion, m_e is the electron mass, and q the charge. The adiabatic invariants are awkward variables to visualize and relate to data so many authors transform to other coordinates. One choice is to use pitch angle, energy, and $L^* = 2\pi M / (|\phi|R_E)$ (Equation 7.48), where $M = 8.22 \times 10^{22} \text{ A m}^2$ is the magnetic moment of the Earth's dipole field and R_E is the Earth's radius, as the three independent variables.

Assuming a dipole field, changing coordinates to equatorial pitch angle α , kinetic energy E , and L^* , and neglecting some cross derivatives, Equation 12.19 can be written as

$$\begin{aligned} \frac{\partial f}{\partial t} = & \frac{1}{g(\alpha)} \frac{\partial}{\partial \alpha} \Big|_{E,L} g(\alpha) \left(D_{\alpha\alpha} \frac{\partial f}{\partial \alpha} \Big|_{E,L} + D_{\alpha E} \frac{\partial f}{\partial E} \Big|_{\alpha,L} \right) \\ & + \frac{1}{A(E)} \frac{\partial}{\partial E} \Big|_{\alpha,L} A(E) \left(D_{EE} \frac{\partial f}{\partial E} \Big|_{\alpha,L} + D_{E\alpha} \frac{\partial f}{\partial \alpha} \Big|_{E,L} \right) \\ & + L^2 \frac{\partial}{\partial L} \Big|_{\mu,J} \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L} \Big|_{\mu,J} \right) \end{aligned} \quad (24.2)$$

¹The general form of the Fokker-Planck equation also includes an advection term, which is ignored here since the drift is negligible compared to diffusion.

where

$$g(\alpha) = T(\alpha) \sin 2\alpha$$

$$A(E) = (E + E_0)(E(E + 2E_0))^{1/2}$$

and E_0 is the electron rest mass energy. $T(\alpha)$ is related to the bounce period and in a dipole field can be approximated by

$$T(\alpha) = (1.3802 - 0.3198(\sin \alpha + \sin^{1/2} \alpha))$$

$D_{\alpha\alpha}$, $D_{\alpha E} = D_{E\alpha}$, D_{EE} , and D_{LL} are the *drift-* and *bounce-averaged* pitch angle, mixed pitch angle energy, energy and radial diffusion coefficients, respectively. When we need to clarify the fact that they are drift and bounce averaged, they will be explicitly written as $\langle D_{\alpha\alpha} \rangle^d$, $\langle D_{\alpha E} \rangle^d$, $\langle D_{EE} \rangle^d$, and $\langle D_{LL} \rangle^d$, respectively.

The bounce-averaged pitch angle and energy diffusion rates are defined by

$$\langle D_{\alpha\alpha} \rangle \equiv \left\langle \frac{(\Delta\alpha)^2}{2\Delta t} \right\rangle$$

$$\langle D_{\alpha E} \rangle \equiv \left\langle \frac{(\Delta\alpha\Delta E)^2}{2\Delta t} \right\rangle$$

$$\langle D_{EE} \rangle \equiv \left\langle \frac{(\Delta E)^2}{2\Delta t} \right\rangle$$

with units of s^{-1} , Js^{-1} , and J^2s^{-1} , respectively.

A loss term of the form f/τ_L can be added to represent losses to the atmosphere due to collisions. Here, τ_L , the loss timescale, is equal to 1/4 of the bounce time in the loss cone and infinite elsewhere. As a first-order approximation, we can neglect the cross terms and simplify Equation 24.2 as

$$\begin{aligned} \frac{\partial f}{\partial t} &= \frac{1}{g(\alpha)} \frac{\partial}{\partial \alpha} \Big|_{E,L} g(\alpha) \left(D_{\alpha\alpha} \frac{\partial f}{\partial \alpha} \Big|_{E,L} \right) \\ &\quad + \frac{1}{A(E)} \frac{\partial}{\partial E} \Big|_{\alpha,L} A(E) \left(D_{EE} \frac{\partial f}{\partial E} \Big|_{\alpha,L} \right) \\ &\quad + L^2 \frac{\partial}{\partial L} \Big|_{\mu,J} \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L} \Big|_{\mu,J} \right) \\ &\quad - \frac{f}{\tau_L} \end{aligned} \tag{24.3}$$

The first three terms on the right-hand side represent pitch angle, energy, and radial diffusion, respectively. The final term accounts for losses to the atmosphere. Pitch angle diffusion has contributions from wave-particle interactions and Coulomb collisions with the atmosphere, though the latter are only significant inside the loss cone. Energy diffusion is due to wave-particle

interactions and radial diffusion to interactions with large-scale fluctuations in the Earth's magnetic and electric fields. The diffusion coefficients are based on statistical wave models derived from data. Depending on the choice of the diffusion coefficients, the Fokker-Planck equation can be linear, quasi-linear, or nonlinear.

Let us make our lives even simpler by considering only the radial diffusion and loss term. The quasi-linear assumptions are:

- Broadband wave spectrum (with random phase), typically assumed as a Gaussian centered around a single dominant mode
- Low amplitude fluctuations (no nonlinear interactions)
- “Resonance limit” (linear growth rate of instabilities goes to 0)
- ULF wave-particle interactions lumped into the diffusion coefficient

Even when quasi-linear theory works fine, there are multiple sources of uncertainties in the radial diffusion coefficient:

- Background magnetic field model
- Bounce + drift average eliminates MLT dependence
- Azimuthal wave structure (i.e. power distribution over mode numbers)

Given the limited observation data, we want to estimate the unknown parameters D_{LL}, τ_L . This is called an *inverse problem*. In 2000, Brautigam and Albert proposed a dependence of the radial diffusion coefficient on K_p and L:

$$D_{LL} \approx 10^{-0.506K_p(t)-9.325} L^{10}$$

The standard statistical approach is to apply a Bayesian parameter estimation, where we are looking for a distribution of parameters and correcting results based on new observations. However, it is difficult to apply in high dimensions, which typically requires the Markov-chain Monte Carlo (MCMC) approach. For example, our assumption is that the parameters can be written in the following forms:

$$D_{LL}(L, t) = 10^{(aK_p(t) + b)L^c}$$

$$\tau_L(L, t) = \begin{cases} (a_0 + a_1L + a_2L^2)/K_p(t) & \text{for } L \leq L_{pp} \\ a_3/K_p(t) & \text{for } L > L_{pp} \end{cases}$$

where L_{pp} is the plasmapause location, which can be estimated with some approximations. We introduce a bunch of random variables, and the posterior distributions of these variables after MCMC give us the best fits.

A new trend is to [use physics-informed machine learning to learn the coefficients](#). First we may need to constrain the forms of the coefficients to avoid an ill-posed inverse problem. Let the physical loss term be described by a drift coefficient C :

$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L} \left|_{\mu, J} \right. \left(\frac{D_{LL}}{L^2} \frac{\partial f}{\partial L} \Big|_{\mu, J} \right) - \frac{\partial Cf}{\partial L}$$

The job of *Physical-Informed Neural Network* (PINN) is to deduce the optimal D_{LL} and C based on statistics by embedding the expected form of the equation into the loss term in the neural network. As this is statistics, we can build and test any kind of relations between variables without physical interpretation, which has been criticised a lot. For example, one may ask how you can let the phase space distribution f depend only on t and L ? The information is incomplete from a physics point of view, but datawise it may just “work”, in the sense that neural network is essentially a *universal high-dimensional function approximator*.

24.2.1 Coupling physical processes at different time scales

If we only look at low frequency regime, the high frequency wave-particle interactions behave as a diffusion process to the plasma. Thus the coupling idea is not to simulate everything on the same time scale, but simulate the radiation belt effect by macroscopic diffusion coefficients.

24.3 Loss

Quantitative understanding of the electron acceleration mechanisms falls short, mostly due to the lack of understanding of the loss mechanisms.

- Bounce loss cone
- Drift loss cone

24.4 Plasmasphere

The plasmasphere is a region within the Earth’s magnetosphere that contains *dense, cold* plasma. It is located in the innermost part of the magnetosphere, and it extends from about 1,000 to 60,000 kilometers above the Earth’s surface. The plasmasphere is formed by the Earth’s magnetic field, which traps charged particles in the region, creating a donut-shaped structure around the planet.

The plasma in the plasmasphere is composed mainly of ions and electrons, and it is much colder and denser than the plasma in other regions of the magnetosphere.

- $T \in [10^3, 10^4] \text{ K}$
- $N \in [10^1, 10^2] \text{ cm}^{-3}$

The plasmasphere plays an important role in the dynamics of the Earth’s magnetosphere and its interactions with the solar wind. It can affect the propagation of radio waves, particularly those used for communication and navigation, and it can also contribute to the formation of the aurora borealis and aurora australis.

25 Aurora

The aurora is a visible and fascinating consequence of the complex magnetospheric processes that are driven by the interaction between the solar wind and the geomagnetic field. The aurora can appear as a diffuse plae band crossing the sky from east to west, or as rapidly moving bright and colorful curtains and rays covering a large fraction of the sky. We will not attempt to describe all the various forms that auroras may appear in, but only distinguish between *diffuse* and *discrete* auroras. The particles that cause these two main types are precipitated into the atmosphere by rather different physical processes. Magnetospheric substorms and the mechanisms leading to the formation of discrete auroras are still at the frontier of magnetospheric research. This chapter is mostly based on Chapter 10 in the lecture notes of Prof. Kjell Rönnmark.

25.1 Auroral Light Emission

There are numerous ancient legends and beliefs about the aurora from various parts of the world. Many of the mythological ideas, as well as more scientific theories from the eighteenth century, explained the aurora as sunlight reflected, refracted, or scattered by various divine or natural processes. Observations reveal that it consists of discrete spectral lines, which proves that auroral light is emitted by a gas, and rules out reflected or scattered sunlight. The strongest line the auroral spectrum is found to be 557.7 nm.

Auroras are caused by charged particles with energy in the range from 10 eV to 30 keV. These particles collide in the upper atmosphere with atoms and molecules that are left in an excited state after the collision. About 90% of the aurora is caused by electrons, and the rest by protons. At altitudes about 100 km where most of the auroral light is emitted, the atmosphere consists mainly of nitrogen and oxygen as shown in Fig. ???. In these gases the quantum mechanically allowed transitions that produce strong spectral lines are all outside the visible part of the spectrum — otherwise air would not be transparent and colourless. The visible auroral emissions are due to transitions from forbidden metastable states, with low transition probabilities and long lifetimes. If the gas pressure is too high, these metastable states will relax to the ground state through collisions long before they decay by emitting light. However, at altitudes above 100 km collisions are sufficiently rare to allow the decay of an excited state of atomic oxygen. This state has an excitation energy of 4.17 eV and a lifetime of 0.8 s. It decays in a two step process. The green line is emitted in the first step, which takes the atom to a state at 1.96 eV. This intermediate state has a very long lifetime, 110 s, and the red line emission at 630.0 nm that takes it to the ground state mainly occurs at altitudes above

200 km. This red line can also be produced by direct excitation to the state at 1.96 eV by collisions with low energy electrons. Other strong lines in the auroral spectrum are emitted by nitrogen. Molecular N_2^+ ions are created in an excited state by collisions, and they decay to their ground state by emitting a bluish-violet line at 427.8 nm. Sometimes the lowest part of strong auroras are colored red by emissions near 600 nm from neutral N_2 molecules.

Figure 25.1 shows the typical altitude distribution of the auroral emissions. The cross section for excitation of the different lines depends on the energy of the incoming auroral particles. Usually, the precipitating particles have higher energy at night than during the day. Low energy electrons give up their energy at higher altitudes, and produces more red emissions. Another difference between day and night is that resonant scattering by sunlight enhances the N_2^+ line 427.8 nm relative to the green line.

25.2 Diffuse Aurora

Some diffuse aurora is nearly always found in the auroral zone. On a clear and dark night it is often seen as a diffuse band, which may appear gray if the intensity is below the color threshold of the eye. Even if no aurora can be seen from the ground, we know from satellite observations of precipitating particles and auroral spectral lines in the UV-range that the diffuse aurora forms a rather continuous band around the auroral zone.

Magnetosheath particles that enter through the magnetopause cusps are a source of diffuse aurora on the dayside. In the cusps there are field lines that connect the ionosphere to the magnetosheath, and magnetosheath particles with sufficiently small pitch angles will precipitate along these field lines. Around local noon, there will be a continuous flux of low energy (≤ 100 eV) particles, which at altitudes above 200 km cause a diffuse band of emissions at 630 nm.

On the nightside, and far into the evening and morning, diffuse auroras are caused by particles of plasmashell origin. The energy spectrum of these particles extends from below 100 eV to above 20 keV. The continuous precipitation of electrons from the plasmashell presents a problem, which has not been completely solved yet. Most of the plasmashell is on closed field lines, and the loss cone in a stationary magnetosphere should be empty (because particles will be lost). To explain the diffuse aurora, we need some mechanisms that can scatter the particles into the loss cone.

Plasma waves in the equatorial magnetosphere can cause strong pitch-angle scattering if they are resonant with the particles. Resonant in this context means that the parallel velocity of the particle and the phase velocity of the wave are related so that the particle feels an electric field that oscillates at the gyrofrequency. Significant changes in the velocities of the particles occur since they systematically gain or lose energy while they are resonant. At least for electrons with energy higher than a few keV, the required pitch-angle scattering can be provided by a type of waves in the whistler mode, known as *magnetospheric hiss*. However, the phase velocity

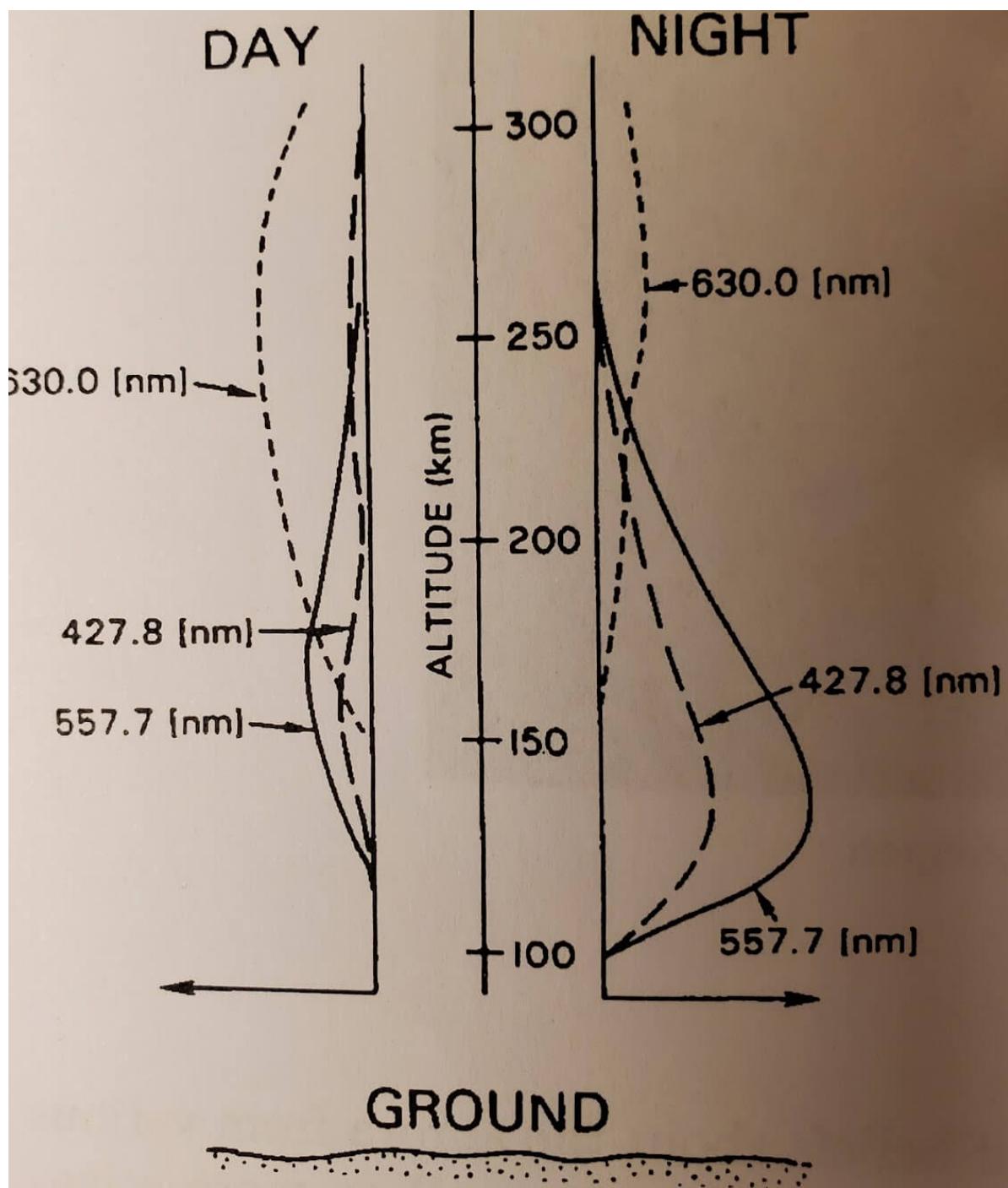


Figure 25.1: The altitude distribution of auroral emissions at day and night.

of whistler mode waves may be too high to give resonance with lower energy electrons. Low energy electrons can be scattered efficiently by waves in the upper hybrid mode. Upper hybrid waves are electrostatic waves, but it is not clear that these waves are sufficiently common to explain the diffuse precipitation of low energy electrons.

25.3 Auroral Waves and Ion Heating

Plasma waves play several roles on the auroral stage. As we have seen, waves near the equator provide the pitch angle scattering that causes diffuse aurora. Closer to Earth there are several types of plasma waves that are generated by the intense flux of particles precipitating into the aurora.

Discrete auroras are strong sources of whistler modes that appear rather different from whistlers generated by lightning. The auroral emissions are characterized by a broadband spectrum covering frequencies from a few kHz to hundreds of kHz, and they are continuously generated with only slow amplitude variations. If the signal is amplified and connected to a loudspeaker a hissing noise is heard, and these waves are called *auroral hiss*. The auroral hiss propagates from the upper ionosphere to the equatorial magnetosphere where it contributes to the magnetospheric hiss. Auroral hiss also propagates to the ground, and at auroral latitudes it will cause noise in radio receivers tuned to low frequencies.

An interesting wave emission connected to discrete auroral arcs is called Auroral Kilometric Radiation (AKR). This is an intense radio wave radiation with peak intensity at frequencies just below 300 kHz, corresponding to a kilometric wavelength. However, the frequency can vary from 50 kHz to 1 MHz. The AKR is generated by a mechanism involving subtle relativistic effects in a low density plasma at a frequency very close to the local electron gyrofrequency. The highest frequencies are generated closest to Earth, where the magnetic field is strongest. The maximum power transmitted as AKR has been estimated to 1 GW, which makes auroras the strongest sources of radio waves from Earth. EM radiation at these frequencies can propagate freely out of the magnetosphere where the density is low, but it will be reflected from the ionosphere at the level where the wave frequency equals ω_{pe} . Hence, AKR signals cannot be received on the ground, and although these emissions are very strong they were not discovered until 1965. In fact, the corresponding auroral emission from Jupiter, which is known as Jovian decametric radiation and has higher frequency, was observed by radio astronomers ten years earlier.

Electrostatic plasma oscillations and other electrostatic waves are also generated above auroral arcs. Electrostatic waves with low frequencies, in the vicinity of the ion gyrofrequencies, have important effects on the magnetospheric dynamics. These waves are generated by the auroral electrons, but they are damped by the ions. The wave energy absorbed by the ions goes mainly into their perpendicular velocities. The magnetic moment then increases, and the ions are pushed out of the ionosphere by the magnetic mirror force $\mu\nabla B$. This process can cause a flow of at least $10^{25} O^+$ per second from the auroral ionosphere during times of auroral activity.

This flow gives the magnetosphere a significant content of ionospheric particles, and reduces the plasma density in the upper ionosphere above auroral arcs. In the auroral cavity dug out by ion heating the electron density is often about $3 \times 10^5 \text{ m}^{-3}$ down to altitudes below $1 R_E$, at least when the ionosphere is in darkness. This density reduction is essential for the generation of AKR, and as we shall see also for the generation of discrete auroras.

25.4 Substorms and Discrete Aurora

Magnetospheric substorms arise due to an imbalance between the dayside and nightside reconnection rates. As a simple example, this can arise during a sudden rotation of the IMF from northward to southward. A southward IMF leads to an increased reconnection rate at the magnetopause. More plasma and more magnetic flux is then transported to the magnetotail in $\mathcal{O}(10^2)$ s, and the pressure in the tail lobes increases. The plasmashell is compressed, which on the ground can be seen as an equatorward motion and brightening of a pre-existing diffuse arc. The increased precipitation from the squeezed plasmashell will also intensify ion heating, which leads to lower densities in the upper ionosphere above the auroral arc. As the density of the cross tail current increases and the neutral sheet becomes thinner, an unstable situation builds up. At this stage, reconnection starts in some part of the tail (?@fig-substorm). This marks the onset of a magnetospheric substorm, and the process of storing energy in the magnetotail described above is often called the *substorm growth phase*.

The original usage of “substorm” comes from Akasofu and Chapman (1961), and was used to describe the short-term magnetic variations during the main phase of a magnetic storm. The current definition did not develop until a decade later, after it became clear that substorms and storms were distinctly different geomagnetic phenomena. The collection of phenomena that includes auroral breakup and expansion, the substorm current wedge, near-Earth dipolarization, and Pi2 pulsations, became collectively known as the magnetospheric substorm.

The concept of the *substorm current wedge* (SCW) has played an important role in understanding the coupling of the magnetotail to the ionosphere during substorms. It provides a simple explanation for the magnetic perturbations observed at mid and low latitudes during substorms, and is useful in understanding the magnetic variations seen in the auroral zone. In its simplest form, a model of the current wedge consists of a single loop with line currents into and out of the ionosphere on dipole field lines connected by a westward ionospheric line current and by an eastward magnetospheric line current (?@fig-substorm).

When the magnetic field lines in the tail start to reconnect, part of the plasmashell will be ejected from the tail and flow Earthward with a speed that often exceeds 400 km/s. The braking of this flow at the inner edge of the plasmashell requires a substantial $\mathbf{j} \times \mathbf{B}$ force, and hence a substantial dusk-to-dawn current, which leads to dipolarization of the inner portion of the magnetic field. At the edges of the injection region, this current is diverted into field-aligned currents that drive the auroral electrojet, a strong westward current in the ionosphere. Note that the auroral electrojet via Ohm’s law implies a westward electric field and an equatorward

flow of the ionospheric plasma. Flows from the reconnection region and dipolarization of the magnetic field are associated with field-aligned currents coupling to the ionosphere, whose net effect is then the SCW. This sequence of changes, from energy storage through explosive release, is called a *magnetospheric substorm*. The reconnection process is temporally and spatially varying, which structures the flows in scale sizes of the order of a few R_E and time scales of a few minutes.

Field-aligned auroral currents in the ionosphere are observed to have current densities of 10 A/m^2 . Considering only field-aligned currents, the current density must decrease higher up where the flux tube widens. At altitudes around $1 R_E$ the current density is still about 1 A/m^2 . In the auroral cavity where the electron density n_e is about $3 \times 10^5 \text{ m}^{-3}$, we can estimate the velocity the electrons must have to carry this current as $v_{\parallel} = j_{\parallel}/en_e \sim 2 \times 10^7 \text{ m/s}$, which corresponds to a kinetic energy slightly higher than 1 keV. When the fast flow injected from the tail by reconnection is stopped, the field-aligned current is carried by electrons that must be accelerated to keV energies. These electrons will cause aurora, and the sudden buildup of this current leads to a breakup of the quiet auroral arc that existed during the growth phase of the substorm.

The auroral breakup occurs in the region of upward field-aligned current, at the western end of the electrojet. The upward current is carried by downgoing electrons that have been accelerated by E_{\parallel} at altitudes around $1 R_E$. When these electrons start to precipitate, a very bright and dynamic auroral display begins. A suitably located observer may see a large part of the sky filled with rapidly moving auroral forms. As illustrated in the classical drawings of an auroral substorm shown by Akasofu 1968 (viewing from the north pole), the aurora spreads poleward, and after a few minutes more stable discrete auroral arcs start to form.

The inertia of the plasma flowing in from the tail with high speed will carry it into a region where the ambient, mainly magnetic, pressure is higher than the pressure in the injected plasma. As the inflow continues the pressure increases, and the injected plasma will start to expand towards the evening and morning side of the magnetosphere. This happens somewhat further from Earth where the ambient magnetic pressure is lower, on field lines that reach the ionosphere poleward of the initial breakup. The azimuthal plasma flows associated with this expansion cause the discrete arcs that form poleward of the breakup.

The simple substorm current wedge model described here is only a crude approximation to the currents that actually exist in space. It is generally believed that the upward current is localized in the premidnight sector while the downward current is more broadly distributed along the auroral oval post-midnight.¹ Upward currents are carried by downward moving electrons, while the downward current is a combination of upward flowing electrons and precipitating ions. The actual currents probably do not flow on the same L-shells. It has also been suggested that the current wedge includes currents closing in meridian planes. In this more complex model there is a second current wedge of opposite sense flowing on a lower L-shell with a different current

¹Up/down direction is determined w.r.t. the ionosphere.

strength. The effect of this loop on the ground is to reduce the apparent strength of the higher latitude current wedge.

The Dungey cycle is the source of the two-cell (DP-2) ionospheric convection pattern. Intervals of steady magnetospheric convection, where the dayside and nightside reconnection rates are roughly balanced, approach the idealized state originally envisioned by Dungey (1961). Yet reconnection is not a steady process, and even during intervals of SMC, when the solar wind driver is relatively constant, plasmashell convection can be intermittent and bursty. Detailed examination of the responses reveals that a pair of field-aligned currents in the form of a current wedge forms during each of the phenomenological stages. Thus, the SCW system is a key phenomenon across all magnetospheric dynamics related to the Dungey cycle, present from weak to extreme activity conditions, and is the link between magnetospheric dynamics and the ionosphere.

25.4.1 Knight's relation

By studying the adiabatic motion of electrons in a parallel electric field, Knight [1973] derived a formula for the auroral current-voltage (C-V) relation. Analyzing data from sounding rockets, Lyons et al. [1979] found that the energy flux of precipitating electrons was proportional to the square of the potential drop. Soon after, it was noticed that for the parameters of interest in the upward current region, Knight's relation could be approximated by [Fridman and Lemaire, 1980] (I feel like this is NOT the original Knight's relation???) Kosta mentioned the linear relation between FAC and B?

$$j_z \approx -K\Delta\phi \quad (25.1)$$

where j_z is the field-aligned current density below the potential drop and $\Delta\phi$ is the potential difference between the ionosphere and the equatorial magnetosphere. The constant K , known as the Lyons-Evans-Lundin constant, is given by

$$K = \frac{n_G e^2}{\sqrt{2\pi m T_G}}$$

where N_G and T_G are the density and temperature of electrons in the equatorial generator region. Notice that since N_G and T_G may vary in space and time, the term "constant" in this context means "independent of ϕ ". The linear C-V relation Equation 25.1 has since been adopted as a part of theoretical and numerical models. However, there are observations both support and against this relation.

Using a fluid description of the plasma, Rönnmark [1999] showed that within a purely electrostatic quasineutral model the potential drop should be proportional to the square of the upward field-aligned current

$$\Delta\phi = \frac{m_e}{2e^3} \frac{j_z^3}{n^2} \quad (25.2)$$

where n is the density in the acceleration region above the ionosphere.

25.4.2 Current circuit

The current circuit connected with a discrete auroral arc is shown in Figure 25.2. Comparing with [?@fig-substorm](#) we see that the flows have been rotated from radial to azimuthal. The length scales are also different, since the auroral electrojet is hundreds of kilometers long and the ionospheric current in Figure 25.2 which runs across an auroral arc is at most a few km. Still the physics is very similar. When pointing out some details, we will here use the geometry shown in Figure 25.2.

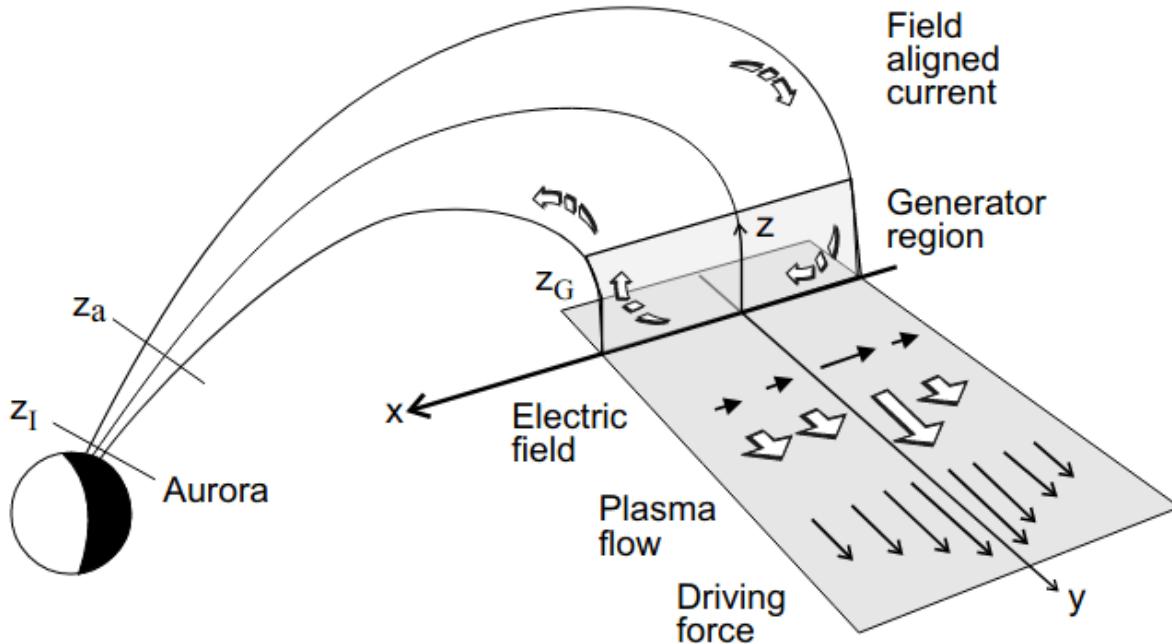


Figure 25.2: Geometry of the auroral current circuit and the generator region in the equatorial magnetosphere.

Let us assume that the auroral flux tube, extending from the ionosphere to the equatorial plane, can be separated into three parts.

1. At low altitudes we have the collision dominated ionosphere, where field-aligned currents are connected to horizontal currents.
2. The magnetospheric plasma above the ionosphere is collisionless. In a stationary state, and in the absence of collisions, the magnetic moment μ and the total energy H are conserved along the phase-space trajectory of a particle. These assumptions imply that there are no currents perpendicular to the magnetic field lines (???), and the field-aligned current in a flux tube is conserved in the second, main, part of the flux tube.
3. The third part is the equatorial generator region. Perpendicular currents are in the generator region driven by kinetic and dynamic pressure gradients, and the divergence of these

perpendicular currents is diverted to field-aligned currents. In the real magnetosphere the boundaries between these parts may be rather diffuse, and the generator region may extend far from the equatorial plane.

For simplicity we assume that the main part of an auroral flux tube is separated from the ionosphere and the generator region by well-defined boundaries. Because of quasineutrality, the density of the light and mobile electrons is determined by the ion density. The ion density will remain approximately constant during transitions between different stationary states. Such transitions, for example increases of the field-aligned current, are accomplished by shear Alfvén waves propagating up and down the field lines. The time variation of E_{\perp} associated with these Alfvén waves will cause an *ion polarization current* j_{\perp} given by

$$k_{\perp} = \frac{n_i m_i}{B^2} \frac{\partial E_{\perp}}{\partial t}$$

Combining this with the ion continuity equation

$$e \frac{\partial n_i}{\partial t} \approx -\partial_{\perp} j_{\perp} = \frac{n_i m_i}{B^2} \frac{\partial \partial_{\perp} E_{\perp}}{\partial t}$$

we can integrate to find the density change Δn_i during the growth of E_{\perp} . Let L_{\perp} be the thickness of the current sheet that j_{\perp} connects to. Choosing $E_{\perp} = 0.1 \text{ V/m}$ and $B = 10 \mu\text{T}$ as typical values for the auroral acceleration region we find

$$\left| \frac{\Delta n_i}{n_i} \right| \sim \frac{m_i}{e B^2} |\partial_{\perp} E_{\perp}| \sim \frac{10 \text{ m}}{L_{\perp}}$$

Clearly, this process can increase the plasma density significantly only within extremely thin current sheets. On the other hand, if the current sheet has a thickness of a few hundred meters or more, the density will remain almost constant. Hence, it seems reasonable to consider the plasma density in the main part of the flux tube as fixed when the current and voltage vary.

In a state of steady field-aligned current the contribution to the current by ions of mass m_i is about a factor $\sqrt{m_e/m_i}$ smaller than the contribution by electrons of mass m_e . If we as a first approximation consider this ratio as fixed, the ion and electron currents are separately conserved. In the main part of the flux tube, where the current is purely field-aligned, the plasma density will then be unaffected by the presence of a steady current. However, altitude variations in plasma properties such as ion composition and electron and ion temperatures may cause variations in the ratio between ion and electron current, and this will cause slow decreases or increases of the plasma density.

Pressure forces in the equatorial plane try to establish a strong velocity shear $\partial_x u_y$, which implies a strong $\partial_x E_x$ in Figure 25.2 ($\mathbf{E} = -\mathbf{u} \times \mathbf{B}$). Recalling eq-ionosphere_potential_derivation we find that the gradient of the ionospheric electric field is determined by the field-aligned current as

$$\frac{\partial E_x}{\partial x} = \frac{j_z}{\Sigma_P}$$

As long as there is no potential drop along the field lines, the magnetospheric and ionospheric electric fields are simply related, and the mapping of a strong velocity shera in the equatorial plane to the ionosphere demands a strong field-aligned current. However, a strong current through the auroral cavity means that the electrons must be accelerated by a potential drop (e.g. Equation 25.2).

The appearance of this potential drop on field lines that carry currents up from the ionosphere means that the ionospheric and magnetospheric electric fields become decoupled. The frozen-in condition does not hold in the acceleration region, and this breakdown of ideal MHD also means that the equatorial plasma is free to flow without dragging the ionosphere along. The potential drop $\Delta\phi$ limits the current density, and hence the braking $\mathbf{j} \times \mathbf{B}$ force. Notice that it is the low altitude acceleration in this potential drop, which in the energy spectrum of the precipitating electrons is observed as a peak at the energy $e\Delta\phi$, that distinguishes the discrete from the diffuse aurora.

The field lines carrying current up from the ionosphere connect to regions with a strong inward gradient of the driving force, which is the outer part of the volume with enhanced pressure. When the brakes in this region are released, the flow will build up a narrower region of high pressure at larger y . ??? Only the outer part of this smaller volume will continue to expand, and this process quickly creates a narrow flow channel in the y -direction. The outer edge of this channel maps to a discrete auroral arc in the ionosphere.

In the model described above, the flows are constrained to the y -direction. In the real magnetosphere, the flows can be deviated in the radial x -direction to form curls and vortices. As illustrated by Akasofu, the geometry of real auroral arcs is very complicated. The patterns are similar to the turbulence seen when water flows from a river into a lake.

I still have many questions regarding the substorms and SCW. Reading the review by (Kepko et al. 2015) makes things worse: my impression is that there has not yet existed a model for explaining the whole current system. No wonder the sawtooth study with MHD-EPIC ended up in a strange way.

25.4.3 Omega Band Auroras

Auroral luminosity undulations of the poleward boundaries of diffuse auroras were first described by Akasofu and Kimball (1964) and were named “omega bands” due to the similarity of their shapes to inverted (poleward-opening) versions of the Greek letter Ω . Omega bands are generally observed in the post-midnight and morning sectors during the recovery phases of magnetospheric substorms. They typically have sizes of 400–1000 km and usually drift eastward at speeds of 0.3–2 km/s.

25.4.4 Theta Auroras

Theta auroras (transpolar aurora arcs) are thought to be linked to high latitude reconnections, which typically happens under northward IMF conditions.

IMF B_y may also contribute to theta auroras as the twist of magnetic field lines together with high latitude reconnections can lead to the strong aurora arcs.

25.4.5 Throat Auroras

Throat Auroras, named after the finger-like structure on the dayside, are thought to be caused by transient reconnection and magnetopause depression with large B_x in the cusp region.

25.4.6 Pulsation Auroras

25.5 Region I and Region II Currents

Region II current is linked to the partial ring current around low latitude region. A simple MHD model does not have ring currents (Section 7.9), which means it cannot have Region II currents. Region I currents, however, can be generated, since it relates to the dayside particle precipitation around the polar cap cusp region and reconnection happening at the magnetopause.

Observation indicates that extreme magnetic storms can cause strongly enhanced Region II current compared with Region I current.

25.6 Aurora Oval

- The overall size is very dynamic.
- The size of the oval is more related to the nightside reconnection (as indicated by the nightside contribution of the electric potential). There is no simple relation with the dayside electric potential (CPCP?).
- Fluid model is not enough to explain observation.

25.7 Summary

- Enhanced dayside reconnection
- thin tail current sheet
- reconnection-accelerated electrons
- carry current into the auroral cavity

26 Guiding Center

The guiding center approximation (GCA) is a theoretical framework used to simplify the complex motion of charged particles in electromagnetic fields by separating the fast gyromotion from the slower drifts of the particle's guiding center (GC).

The basic equation governing the motion of a charged particle with mass m and charge e is given by

$$\frac{d\mathbf{u}}{dt} = \frac{e}{mc} (c\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

where $\mathbf{u} = \gamma\mathbf{v}$ represents the particle's four-velocity, \mathbf{v} is the particle velocity, c is the speed of light while \mathbf{E} and \mathbf{B} are the electric and magnetic field vectors, respectively.

In GCA, the orbital motion of the particle is detached from its instantaneous gyration center. This relaxes any constraints related to particle gyration, allowing systematic larger time steps to be taken and thus a considerable saving in computational time. The guiding center (GC) equations are obtained in the limit of slow-varying fields. This restricts its validity to situations in which the Larmor radius remains negligible with respect to the overall electromagnetic field scale and in which the cyclotron frequency is large enough so that adiabatic invariance holds. In this respect, the GCA is most suitable for particles with large charge-to-mass ratios such as electrons.

26.1 Summarized Version

26.1.1 Basic Concepts and Variables

State Variables: Instead of the full 6D phase space (\mathbf{x}, \mathbf{v}) , we solve for the reduced 4D guiding center state:

- \mathbf{X} : The spatial position of the guiding center.
- $p_{\parallel} = m\gamma v_{\parallel}$: The relativistic kinetic momentum parallel to the magnetic field, where γ is the Lorentz factor.

The position \mathbf{r} of a particle is decomposed into the guiding center position $\mathbf{R}(t)$ and the gyration vector ρ :

$$\mathbf{r} = \mathbf{R}(t) + \rho(\mathbf{R}, t, \psi)$$

where ψ is the gyrophase. This approximation holds when the characteristic scale of field variations L is much larger than the Larmor radius ρ ($L \gg \rho$) and time variations are slow compared to the gyroperiod.

Parameters (Constant of Motion):

- μ : The magnetic moment (adiabatic invariant).

$$\mu = \frac{p_{\perp}^2}{2mB}$$

In the GCA, μ is constant. It is calculated once from the initial conditions and treated as a parameter during the simulation.

26.1.2 The Effective Fields

The modern GCA equations can be written compactly using “effective” electromagnetic fields \mathbf{B}^* and \mathbf{E}^* . These correct the standard fields to account for drifts naturally.

Definitions: - $\mathbf{B}(\mathbf{X})$ and $\mathbf{E}(\mathbf{X})$: Standard EM fields at the guiding center location. - $\mathbf{b} = \mathbf{B}/|\mathbf{B}|$: Unit vector along the magnetic field line. - q : Particle charge.

Calculate the Effective Magnetic Field \mathbf{B}^*

This term incorporates the geometry of the field lines (curl of \mathbf{b}) which drives curvature drift.

$$\mathbf{B}^* = \mathbf{B} + \frac{cp_{\parallel}}{q} \nabla \times \mathbf{b}$$

Calculate the Effective Electric Field \mathbf{E}^*

In many implementations, it is easier to work directly with the Hamiltonian gradients, but formally:

$$\mathbf{E}^* = \mathbf{E} - \frac{1}{q} \nabla(\mu B)$$

26.1.3 Hamiltonian

For a relativistic particle, the Hamiltonian H represents the total energy.

$$H(\mathbf{X}, p_{\parallel}, \mu, t) = \sqrt{m^2 c^4 + p_{\parallel}^2 c^2 + 2m\mu B(\mathbf{X})c^2} + q\Phi(\mathbf{X})$$

- Φ : Electric scalar potential (if $\mathbf{E} = -\nabla\Phi$).
- Note: If you only have \mathbf{E} and not Φ , you will use the force term qE_{\parallel} directly in the equations of motion.

26.1.4 Equations of Motion

These are the final expressions to solve numerically. They describe the time evolution of the guiding center position \mathbf{X} and parallel momentum p_{\parallel} .

Equation for Position ($\dot{\mathbf{X}}$):

$$\frac{d\mathbf{X}}{dt} = \frac{1}{B_{\parallel}^*} \left[\frac{p_{\parallel}}{m\gamma} \mathbf{B}^* + c\mathbf{E}^* \times \mathbf{b} \right]$$

- Term 1 ($\propto \mathbf{B}^*$): Represents motion along the field line plus curvature drift (hidden inside \mathbf{B}^*).
- Term 2 ($\propto \mathbf{E}^* \times \mathbf{b}$): Represents the $\mathbf{E} \times \mathbf{B}$ drift and the ∇B drift (hidden inside \mathbf{E}^* via the $\mu \nabla B$ term).
- Denominator: $B_{\parallel}^* = \mathbf{b} \cdot \mathbf{B}^* \approx B$. This is a normalization factor essentially equal to B but corrected for phase-space volume conservation.

Equation for Parallel Momentum (\dot{p}_{\parallel}):

$$\frac{dp_{\parallel}}{dt} = \frac{q}{B_{\parallel}^*} \mathbf{B}^* \cdot \mathbf{E}^*$$

This simplifies to the standard parallel force equation:

$$\dot{p}_{\parallel} \approx qE_{\parallel} - \frac{\mu}{\gamma} \nabla_{\parallel} B$$

26.1.5 Numerical Implementation Guide

To implement this in code, follow this loop:

- A. Initialization
 1. Start with particle particle position \mathbf{x} and velocity \mathbf{v} .
 2. Compute local $\mathbf{B}(\mathbf{x})$ and \mathbf{b} .
 3. Decompose velocity: $\mathbf{v} = v_{\parallel}\mathbf{b} + \mathbf{v}_{\perp}$.
 4. Compute $\mu = \frac{m\gamma^2 v_{\perp}^2}{2B}$ (Relativistic moment).
 5. Initialize state vector $\mathbf{Y} = [\mathbf{X}, p_{\parallel}]$.
- B. Time Stepping Loop. For every time step $t \rightarrow t + \Delta t$:
 1. Interpolate Fields: Get $\mathbf{B}, \mathbf{E}, \nabla B$ at current \mathbf{X} .
 2. Compute Gradients: You need $\nabla \times \mathbf{b}$ and ∇B .
 - $\nabla \times \mathbf{b}$ can be computed numerically or analytically if the B-field model is known.
 3. Construct Effective Fields:

- $\mathbf{B}^* = \mathbf{B} + \frac{cp_{\parallel}}{q}(\nabla \times \mathbf{b})$
- $B_{\parallel}^* = \mathbf{b} \cdot \mathbf{B}^*$

4. Evaluate Derivatives (RHS):

- $\gamma = \sqrt{1 + \frac{p_{\parallel}^2 + 2m\mu B}{m^2 c^2}}$
- $\dot{\mathbf{X}} = \frac{1}{B_{\parallel}^*} \left(\frac{p_{\parallel}}{m\gamma} \mathbf{B}^* + \frac{e}{q} (q\mathbf{E} - \mu \nabla B) \times \mathbf{b} \right)$
- $\dot{p}_{\parallel} = \frac{q}{B_{\parallel}^*} \mathbf{B}^* \cdot (\mathbf{E} - \frac{\mu}{q} \nabla B)$

5. Advance: Update \mathbf{X} and p_{\parallel} .

- C. Constraints & Checks

- Time Step: The GCA allows much larger time steps than full orbit codes because you don't resolve the gyro-period. However, you must resolve the gradient length scales. A common constraint is ensuring the particle does not cross more than one cell (or characteristic length) per step.
- Validity: Check that $E_{\perp} < B$ and that the gyroradius is small compared to field gradients ($R_L \ll L_B$).

26.1.6 Conversion from Full Phase Space to Guiding Center

Transforming from the full 6D particle state (\mathbf{x}, \mathbf{v}) to the reduced Guiding Center (GC) state $(\mathbf{X}, v_{\parallel}, \mu)$ is a surjective mapping (many particle states map to the same GC state).

To perform this switch consistently—especially in hybrid codes where you might switch back and forth—it is critical to match the constants of motion (like Energy) between the two frames. Here is the step-by-step transformation procedure:

1. Calculate the Guiding Center Position (\mathbf{X})

The guiding center is located at the center of the particle's gyration. You must calculate the Larmor radius vector at the particle's current position \mathbf{x} and subtract it.

- Compute local field: Retrieve $\mathbf{B}(\mathbf{x})$ and the unit vector $\mathbf{b}(\mathbf{x}) = \mathbf{B}(\mathbf{x})/|\mathbf{B}(\mathbf{x})|$.
- Compute displacement:

$$\mathbf{X} \approx \mathbf{x} + \frac{m}{qB(\mathbf{x})} \mathbf{v} \times \mathbf{b}(\mathbf{x})$$

- This equation is valid to the first order in the Larmor radius.
- It is recommended to perform this displacement in Cartesian coordinates to avoid singularities.

2. Update Fields at the New Position

Once you have the estimated guiding center position \mathbf{X} , you must evaluate the electromagnetic fields $\mathbf{B}(\mathbf{X})$, $\mathbf{E}(\mathbf{X})$ and potentials $\Phi(\mathbf{X})$ at this new location.

3. Calculate Parallel Velocity (v_{\parallel})

The parallel velocity is defined relative to the magnetic field at the guiding center \mathbf{X} , not the particle position \mathbf{x} .

- Standard Projection:

$$v_{\parallel} = \mathbf{v} \cdot \mathbf{b}(\mathbf{X})$$

- Energy/Momentum Matching (High Precision): For higher accuracy, specifically to ensure that the particle's toroidal momentum P_{ϕ} is conserved during the switch (in axisymmetric systems), you can calculate v_{\parallel} derived from the momentum conservation equation:

$$v_{\parallel} \approx \frac{1}{B(\mathbf{X})} (\mathbf{v} \cdot \mathbf{B}(\mathbf{X}) + \dots \text{higher order terms})$$

For most standard implementations, the simple projection $\mathbf{v} \cdot \mathbf{b}(\mathbf{X})$ is sufficient.

4. Calculate Magnetic Moment (μ)

The magnetic moment is the adiabatic invariant. It is crucial to calculate μ such that the total energy \mathcal{H} of the particle is conserved during the transformation.

- Energy Conservation Condition:

$$\mathcal{H}_{\text{particle}} = \mathcal{H}_{\text{GC}}$$

$$\frac{1}{2}mv^2 + q\Phi(\mathbf{x}) = \frac{1}{2}mv_{\parallel}^2 + \mu B(\mathbf{X}) + q\Phi(\mathbf{X})$$

- Formula for μ : Rearranging the energy equation gives the most robust definition for μ during the switch:

$$\mu = \frac{1}{B(\mathbf{X})} \left[\frac{1}{2}mv^2 - \frac{1}{2}mv_{\parallel}^2 + q(\Phi(\mathbf{x}) - \Phi(\mathbf{X})) \right]$$

- If the electric potential variation $q(\Phi(\mathbf{x}) - \Phi(\mathbf{X}))$ is negligible (order of Larmor radius), this simplifies to the standard definition:

$$\mu \approx \frac{mv_{\perp}^2}{2B(\mathbf{X})}$$

where $v_{\perp}^2 = v^2 - v_{\parallel}^2$.

Summary Algorithm 1. Input: Particle \mathbf{x}, \mathbf{v} . 2. Shift: $\mathbf{X} = \mathbf{x} + \frac{m}{qB(\mathbf{x})}(\mathbf{v} \times \mathbf{b})$. 3. Project: $v_{\parallel} = \mathbf{v} \cdot \mathbf{b}(\mathbf{X})$. 4. Conserve: $\mu = \frac{\frac{1}{2}mv^2 - \frac{1}{2}mv_{\parallel}^2}{B(\mathbf{X})}$ (ignoring potential differences). 5. Output: State $(\mathbf{X}, v_{\parallel}, \mu)$.

26.1.7 Conversion from Guiding Center to Full Phase Space

Converting from the reduced Guiding Center (GC) variables $(\mathbf{X}, p_{||}, \mu)$ back to the full 6D particle phase space (\mathbf{x}, \mathbf{p}) is a “one-to-many” mapping. Because the gyrophase ζ is averaged out in the GCA, you must arbitrarily choose (or randomly assign) a gyrophase angle $\zeta \in [0, 2\pi)$ to reconstruct a specific particle position and velocity.

Here is the step-by-step reconstruction procedure.

1. Calculate Local Quantities at the Guiding Center

First, evaluate the electromagnetic fields and unit vectors at the known guiding center position \mathbf{X} . - Local Magnetic Field: $\mathbf{B} = \mathbf{B}(\mathbf{X})$ - Unit Vector: $\mathbf{b} = \mathbf{B}/|\mathbf{B}|$ - Perpendicular Momentum Magnitude: Calculate p_{\perp} from the magnetic moment μ (which is a parameter in GCA):

$$p_{\perp} = \sqrt{2mB(\mathbf{X})\mu}$$

- Gyroradius Magnitude:

$$\rho = \frac{p_{\perp}}{|q|B(\mathbf{X})}$$

2. Define a Perpendicular Basis

To define the gyration direction, you need a coordinate system perpendicular to the magnetic field. Construct two unit vectors \mathbf{e}_1 and \mathbf{e}_2 such that $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{b})$ form a right-handed orthonormal basis.

A common numerical method to find \mathbf{e}_1 is:

$$\mathbf{e}_1 = \frac{\mathbf{b} \times \hat{\mathbf{u}}}{|\mathbf{b} \times \hat{\mathbf{u}}|}$$

where $\hat{\mathbf{u}}$ is any fixed unit vector not parallel to \mathbf{b} , e.g., $\hat{\mathbf{x}}$. Then:

$$\mathbf{e}_2 = \mathbf{b} \times \mathbf{e}_1$$

3. Reconstruct Particle Position (\mathbf{x})

The particle position is the guiding center position displaced by the Larmor radius vector ρ . Select a gyrophase ζ (e.g., draw from a uniform distribution $U[0, 2\pi]$).

$$\mathbf{x} = \mathbf{X} + \rho$$

The displacement vector ρ is:

$$\rho = \rho(\cos \zeta \mathbf{e}_1 + \sin \zeta \mathbf{e}_2)$$

Note: In some high-precision hybrid schemes, specific choices for ζ (such as aligning ρ perpendicular to ∇B) are used to minimize energy discrepancies when switching models, but for general reconstruction, a random phase is standard.

4. Reconstruct Particle Momentum (\mathbf{p})

The total momentum is the sum of the parallel momentum, the perpendicular gyration momentum, and (optionally for higher accuracy) the momentum associated with the $\mathbf{E} \times \mathbf{B}$ drift.

Standard Reconstruction (0th Order): This reconstruction creates a particle purely gyrating and moving parallel to field lines.

$$\mathbf{p} = p_{\parallel} \mathbf{b} + \mathbf{p}_{gyro}$$

where the gyration momentum \mathbf{p}_{gyro} is perpendicular to ρ (tangent to the circle):

$$\mathbf{p}_{gyro} = p_{\perp} (-\sin \zeta \mathbf{e}_1 + \cos \zeta \mathbf{e}_2) \times \text{sign}(q)$$

Drift-Corrected Reconstruction (1st Order): If the GCA simulation includes significant $\mathbf{E} \times \mathbf{B}$ drifts, we should add this component to the particle velocity to ensure the full orbit matches the drift path.

$$\mathbf{p} \approx p_{\parallel} \mathbf{b} + \mathbf{p}_{gyro} + m\gamma \mathbf{v}_E$$

where

$$\mathbf{v}_E = \frac{c\mathbf{E} \times \mathbf{b}}{B}$$

26.2 Classical Guiding Center Theory

Here are the essentials points in the derivation of the classical *guiding center equations of motion (EOM)*. For a complete discussion refer to Vandervoort (1960).

26.2.1 Lorentz equation in tensor form

In order to derive the Guiding Center Approximation (GCA), we would use the tensor notations in relativity. The equation of motion for a charged particle in an EM field are written in a tensorial form as

$$\frac{dx_{\mu}}{d\tau^2} = F_{\mu\nu} \frac{dx_{\nu}}{d\tau} \quad (26.1)$$

where x_{μ} is the particle four-coordinate, $F_{\mu\nu}$ is the EM tensor and τ is the particle proper time.

The EM tensor is

$$F_{\mu\nu} = \frac{e}{mc} \begin{pmatrix} 0 & B_z & -B_y & E_x \\ -B_z & 0 & B_x & E_y \\ B_y & -B_x & 0 & E_z \\ E_x & E_y & E_z & 0 \end{pmatrix} \quad (26.2)$$

26.2.2 Fundamental assumptions

The GC formalism holds under two fundamental assumptions, namely:

1. The gyration radius must remain small compared to the scale length upon which the electromagnetic field changes significantly.
2. The particle undergoes many gyrations before the electromagnetic field changes appreciably (slow-varying fields approximation). If we let ρ be the particle gyroradius (apart from a factor $\sqrt{2}$), x_μ the particle position, X_μ the GC position, τ the proper time and $F_{\mu\nu}$ the EM field tensor, the two conditions stated above can be mathematically expressed as

$$\rho \left| \frac{\partial F_{\mu\nu}}{\partial x_\alpha} \right| \ll |F_{\mu\nu}| \quad (26.3)$$

and

$$\frac{1}{\omega} \left| \frac{\partial F_{\mu\nu}}{\partial x_\alpha} \right| \left| \frac{dX_\beta}{d\tau} \right| \ll |F_{\mu\nu}| \quad (26.4)$$

where ω is the gyrofrequency. When the previous conditions hold, we can separate the particle trajectory into a gyration and a motion of the guiding center.

26.2.3 Derivations of GC equation

The general solution of Equation 26.1 is a linear combination of the two fundamental solutions related to the four EM tensor eigenvalues $q = \pm i\omega$ and $q = \pm\lambda$, where

$$\begin{aligned} \omega &= \frac{e}{mc} \sqrt{\frac{1}{2}(B^2 - E^2) + \frac{1}{2}\sqrt{(B^2 - E^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}} \\ \lambda &= \frac{e}{mc} \sqrt{-\frac{1}{2}(B^2 - E^2) + \frac{1}{2}\sqrt{(B^2 - E^2)^2 + 4(\mathbf{E} \cdot \mathbf{B})^2}} \end{aligned} \quad (26.5)$$

ω is a generalized relativistic form of the Larmor frequency $\omega_\xi = eB/(mc)$ in the case when $E \neq 0$ (which can be verified by solving in the limit $E \rightarrow 0$). The general solution x_μ of Equation 26.1 is

$$x_\mu = \xi_\mu \rho \cos(\omega\tau) - \eta_\mu \rho \sin(\omega\tau) + \alpha_\mu \nu \cosh(\lambda\tau) + \beta_\mu \nu \sinh(\lambda\tau) \quad (26.6)$$

Note that the first two terms are related to a periodic motion, that is the gyration around magnetic field lines with gyroradius ρ . Here ρ, ν are constants defined by the initial conditions and $\xi_\mu, \eta_\mu, \alpha_\mu$ and β_μ are four-versors normalized in the manner

$$\xi_\mu^2 = \eta_\mu^2 = 1, \quad \text{and} \quad \alpha_\mu^2 = -\beta_\mu^2 = 1 \quad (26.7)$$

Solutions for $\rho = 0$ and $\nu = 0$ must be valid separately, so by replacing x_μ for these two particular cases inside Equation 26.1 we can see that $\xi_\mu, \eta_\mu, \alpha_\mu$ and β_μ form an orthogonal set of the Minkowsky space. The orthogonality relations between this set of four-vectors in particular states that the periodic motion (gyration) in the $(\xi_m u, \eta_\mu)$ -plane is perpendicular to the acceleration motion in the (α_μ, β_μ) -plane, and will be useful later. At this point, another useful relation is retrieved from the squared velocity invariance

$$\frac{dx_\mu}{d\tau} \frac{dx_\mu}{d\tau} = -c^2, \quad \text{which leads to} \quad \omega^2 \rho^2 - \lambda^2 \nu^2 = -c^2 \quad (26.8)$$

Since we are only interested in cases in which gyration appears, the only singular eigenvalue case we must consider is the one with $\lambda = 0, \omega \neq 0$, corresponding to $\mathbf{E} \cdot \mathbf{B} = 0$ and $|\mathbf{E}| \rightarrow 0$. The associated solution is

$$x_\mu = \xi_\mu \rho \cos(\omega\tau) - \eta_\mu \rho \sin(\omega\tau) + U_\mu \tau \quad (26.9)$$

Here the gyration motion with Larmor frequency $\omega_\xi = \omega$ is clearly shown, along with a uniform motion (drift) with velocity $U_\mu = (\gamma c, \mathbf{U})$ in the plane perpendicular to the (\mathbf{E}, \mathbf{B}) -plane.¹ In this particular case, the squared velocity invariance Equation 26.8 becomes

$$\omega^2 \rho^2 + U_\mu^2 = -c^2, \quad \text{or} \quad U_\mu^2 = -c^2 - \omega^2 \rho^2 < -c^2 \quad (26.10)$$

Equation 26.10 is of crucial importance in the GCA formalism, because the drift motion U_μ is the mean motion separated from the gyration and corresponds to the GC motion itself, meaning that what we just found is a relation between the energy and momentum of a particle guiding center. It is very similar to the well-known formula of the squared four-velocity $U'_\mu{}^2 = -c^2$, the only difference being an additional term $\omega^2 \rho^2$. This suggests interpreting the motion of the GC as the one of a particle located in the center of gyration, performing no gyration and possessing a squared four velocity as stated in Equation 26.10, in agreement with part of the particle kinetic energy being stored in the gyration motion through the term $\omega^2 \rho^2$. By separating the spatial and temporal parts of U_μ and rearranging some terms inside Equation 26.10, we can get a relation for the GC Lorentz factor

$$\gamma_{\text{GC}} = \sqrt{1 + \frac{|\mathbf{U}|^2}{c^2} + \frac{\omega^2 \rho^2}{c^2}} = \sqrt{1 + \frac{\gamma^2 |\mathbf{v}|^2}{c^2} + \frac{\omega^2 \rho^2}{c^2}} \quad (26.11)$$

Although we only considered a limit case, the same equation can be proven to be generally valid in the GCA formalism.

We now want to isolate the gyration motion in a general case. This is done by separating the motion as follows:

$$x_\mu = \xi_\mu x + \eta_\mu y + x'_\mu + X_\mu \quad (26.12)$$

¹For verification simply substitute x_μ and U_μ in the left and right-end side of Equation 26.1, respectively, to find that $F_{\mu\nu} U_\nu = 0$, and rewrite the indexes of such equation explicitly.

where x_μ is the space-time position of the particle and X_μ is the position of the guiding center. We associate two variables x and y to describe the periodic motion in the (ξ, η) -plane, and the term x'_μ will account for the periodic motions which do not lie in the same plane. To more easily describe the gyration, we also choose a new pair of variables

$$\zeta = \frac{1}{\sqrt{2}}(x + iy) \quad \text{and} \quad \zeta^* = \frac{1}{\sqrt{2}}(x - iy) \quad (26.13)$$

so that $\xi_\mu x + \eta_\mu y = \delta_\mu \zeta + \sigma_\mu \zeta^*$. The new variables will allow us to choose ζ_μ and δ_μ in a more convenient way and let some factor $e^{i\phi}$ account for the initial conditions of the problem. We rewrite the particle coordinate using Equation 26.13

$$x_\mu = \delta_\mu \zeta + \sigma_\mu \zeta^* + x'_\mu + X_\mu \quad (26.14)$$

The next step is substituting this expression into Equation 26.1 and expand the second derivative, which leads to a lengthy expression containing eight unknowns: ζ, ζ^* , two components of x_μ (the condition with the (σ, δ) -plane sets the other two), and four components of X_μ .

The conditions for the GCA expressed in Equation 26.4 implicitly contain a parameter of smallness $\epsilon = |u/\omega L| \ll 1$, where u is the typical particle four-velocity and L the scale length upon which the change in the EM fields $F_{\mu\nu}$ is comparable to $F_{\mu\nu}$ (slowly-varying fields approximation). This allows us to approximate $F_{\mu\nu}(x_\mu)$ (particle coordinate) by using a Taylor expansion about X_μ (the GC coordinate)

$$\begin{aligned} F_{\mu\nu}(x_\mu) &= F_{\mu\nu}(X_\mu) \\ &+ \frac{\partial F_{\mu\nu}}{\partial x_\xi}(\delta_\xi \zeta + \sigma_\xi \zeta^*) \\ &+ \frac{1}{2} \frac{\partial^2 F_{\mu\nu}}{\partial x_\xi \partial x_\pi} [\delta_\xi \delta_\pi \zeta^2 + \sigma_\xi \sigma_\pi \zeta^{*2} + (\sigma_\xi \delta_\pi + \sigma_\pi \delta_\xi) |\zeta|^2] \\ &+ \frac{\partial F_{\mu\nu}}{\partial x_\xi} x'_\xi + \dots \end{aligned}$$

We should now substitute $F_{\mu\nu}(x_\mu)$ inside Equation 26.1 using the new coordinates in order to get a quite long equation containing all quantities evaluated at the GC position X_μ up to

2nd-order, and group all 2nd-order terms on the right-end side as G_μ

$$\begin{aligned} & \delta_\mu \frac{d^2\zeta}{d\tau^2} + \sigma_\mu \frac{d^2\zeta^*}{d\tau^2} + \left(i\omega\delta_\mu + 2\frac{d\zeta}{d\tau} \right) \frac{d\zeta}{d\tau} \\ & + \left(-i\omega\sigma_\mu + 2\frac{d\sigma_\mu}{d\tau} \right) \frac{d\zeta^*}{d\tau} + \frac{d^2\delta_\mu}{d\tau^2}\zeta + \frac{d^2\sigma_\mu}{d\tau^2}\zeta^* \\ & - \frac{\partial F_{\mu\nu}}{\partial x_\xi} (\delta_\xi\zeta + \sigma_\xi\zeta^*) \frac{dX_\nu}{d\tau} - F_{\mu\nu} \left(\frac{d\delta_\nu}{d\tau}\zeta + \frac{d\sigma_\nu}{d\tau}\zeta^* \right) \\ & - \frac{\partial F_{\mu\nu}}{\partial x_\xi} (\delta_\xi\zeta + \sigma_\xi\zeta^*) \left(\delta_\nu \frac{d\zeta}{d\tau} + \sigma_\nu \frac{d\zeta^*}{d\tau} \right) \\ & + \left(\frac{d^2X_\nu}{d\tau^2} - F_{\mu\nu} \frac{dX_\mu}{d\tau} \right) + \left(\frac{d^2x'_\mu}{d\tau^2} - F_{\mu\nu} \frac{dx'_\nu}{d\tau} \right) = G_\mu \end{aligned} \quad (26.15)$$

An expression for ζ is needed to further proceed, but since the derivation is quite long we will not explicitly derive it here. The basic idea is to use the orthogonality equations for the new four-vectors σ_μ and δ_μ and the antisymmetry property of the EM tensor inside Maxwell's equations written in tensorial form, and then assume that a solution for X_μ has already been found. Considered that x_μ is a 1st-order term in the GCA, by ignoring 2nd-order quantities and writing terms independent from ζ as a numerical value P , then Equation 26.15 becomes

$$\frac{d^2\zeta}{d\tau^2} + i\Omega \frac{d\zeta}{d\tau} + \frac{i}{2} \frac{d\Omega}{d\tau} \zeta + \Omega a \zeta + P = 0 \quad (26.16)$$

where $\Omega a \zeta$ contains all the linear terms in ζ . Consequently, the corresponding solution ζ , by analogy to the *Wentzel–Kramers–Brillouin (WKB)* approximation in quantum mechanics, is

$$\zeta = \rho_0 \sqrt{\frac{\omega_0}{\Omega}} \exp \left[-i \left(\Phi + \int_{\tau_0}^{\tau} \sigma_1 d\tau \right) \right] \quad (26.17)$$

to the 0th-order, being

$$\Omega = \omega - 2i\sigma_\mu \frac{d\delta_\mu}{d\tau}, \quad \Phi = \int_{\tau_0}^{\tau} \Omega d\tau - \phi, \quad \text{and} \quad \sigma = a - \frac{a^2}{\Omega} \quad (26.18)$$

The subscripts 0, 1 indicate the associated order of approximation for quantities defined above.

Since now we know the expression for ζ , we can write the 1st-order EOM for X_μ . For this purpose, we require that in the left-end side of Equation 26.15 the terms containing the 1st- and 2nd-order derivatives of X_μ are balanced by the remaining nonoscillatory terms, that is

$$\frac{d^2X_\mu}{d\tau^2} - F_{\mu\nu} \frac{dX_\nu}{d\tau} = \frac{\partial F_{\mu\nu}}{\partial x_\xi} \left(\delta_\xi \sigma_\nu \frac{d\zeta^*}{d\tau} + \delta_\nu \sigma_\xi \frac{d\zeta}{d\tau} \right) \quad (26.19)$$

Having found the solution Equation 26.17 and making use once again of the orthogonality relations for δ_μ and σ_μ and the tensorial Maxwell's equations, we can write the equation of motion for the GC

$$\frac{d^2 X_\mu}{d\tau^2} - F_{\mu\nu} \frac{dX_\nu}{d\tau} + \rho_0^2 \omega_0 \frac{\partial \omega}{\partial x_\mu} = 0 \quad (26.20)$$

where ω (Equation 26.5) is the generalized Larmor frequency, $\rho_0^2 \omega_0 = \mu_0 = mc\gamma^2 v_\perp^2 / 2eB$ is the 0th-order approximation (apart from a constant factor) to the particle magnetic moment μ and $\omega_0 = eB/(mc)$ represents the lowest-order approximation to the Larmor frequency in the case $E \ll B$.

More precisely, the magnetic moment should be regarded as constant only in the reference frame moving at \mathbf{v}_E , because in that particular case $\mathbf{E} = 0$. Indeed, a formal analysis (Northrop and Teller (1960)) shows that the corresponding constant of motion in a general reference frame is actually an asymptotic series expansion in a smallness parameter $\epsilon = u/(\omega_0 L)$ in the form $\mu = (e/c) [\mu_0 + \epsilon \mu_1 + \epsilon^2 \mu_2 + \dots]$ so that μ_0 is not constant and can still vary in compliance with the adiabatic theory. Nevertheless, extensive numerical testing confirms that the errors due to assuming $d\mu_0/dt \approx 0$ are at most of the same order of those introduced by the GC formalism, as a consequence of the slowly varying field condition which prevents sensible changes in the magnetic moment. Hence, we safely assume that $\mu \approx (e/c)\mu_0$ is invariant.

26.2.4 First order solution for the GCA equations of motion

Although deceptively simple, the GC equation of motion Equation 26.20 is more conveniently cast in a form in which the GC velocity appears explicitly, under the nearly-crossed (EM) fields condition

$$\frac{\mathbf{E}_\parallel \cdot \mathbf{B}}{B^2 - E_\perp^2} \ll 1 \quad (26.21)$$

where the subscripts \parallel and \perp indicate the parallel and perpendicular components of the electric field with respect to the magnetic field unit vector \hat{b} . It can be proven that Equation 26.21 allows the EM tensor field $F_{\mu\nu}$ to be split into a contribution $F_{\mu\nu}^{(0)}$ constructed solely from \mathbf{E}_\perp and \mathbf{B} , plus a correction term $F_{\mu\nu}^{(1)}$ depending on \mathbf{E}_\parallel . A formal analysis leads to the conclusion that the GC four-velocity $U_\mu \equiv dX_\mu/d\tau$ can be decomposed into a 0th-order contribution $\mathbf{U}^{(0)}$ (which contains the drift velocity $\mathbf{v}_E = c\mathbf{E}_\perp \times \mathbf{B}/B^2$ as well as the velocity component parallel to the magnetic field line $v_\parallel \hat{b}$) plus 1st-order correction $\mathbf{U}^{(1)}$, leading to

$$U_\mu = (\gamma c, \mathbf{U}) \simeq (\gamma c, \gamma v_\parallel \hat{b} + \gamma \mathbf{v}_E + \mathbf{U}^{(1)}) \quad (26.22)$$

where γ is the particle Lorentz factor $\gamma = (1 - v^2/c^2)^{-1/2}$. An equation for $\mathbf{U}^{(1)}$ can be derived from Equation 26.20 using a recursive approach as shown below, while from the spatial component one obtains an equation (to the same order) for the parallel component of the GC

four-velocity γv_{\parallel} . These yield the (1st-order) GCA system of ordinary differential equations (ODEs)

$$\begin{aligned}\frac{d\mathbf{X}}{dt} &= \mathbf{v}_E + v_{\parallel} \hat{b} + \frac{\gamma_E^2}{B} \hat{b} \times \left[\frac{mc\gamma}{e} (v_{\parallel} \mathcal{L}(\hat{b}) + \mathcal{L}(\mathbf{v}_E)) + \mathcal{M} \left(\frac{B}{\gamma_E} \right) + \frac{v_{\parallel} E_{\parallel}}{c} \mathbf{v}_E \right] \\ \frac{d(\gamma v_{\parallel})}{dt} &= \frac{e}{m} E_{\parallel} - \gamma \hat{b} \cdot \mathcal{L}(\mathbf{v}_E) - \frac{\mu}{\gamma m} \hat{b} \cdot \nabla \left(\frac{B}{\gamma_E} \right)\end{aligned}\quad (26.23)$$

where $d\mathbf{X}/dt = \mathbf{U}^{(1)}/\gamma$ represents the velocity of the guiding center, e/m is the particle charge-to-mass ratio and $\mu_0 = c\mu/e$. The operators \mathcal{L} and \mathcal{M} are defined as

$$\begin{aligned}\mathcal{L}(x) &= \frac{\partial \mathbf{x}}{\partial t} + v_{\parallel} (\hat{b} \cdot \nabla) \mathbf{x} + (\mathbf{v}_E \cdot \nabla) \mathbf{x} \\ \mathcal{M}(x) &= \frac{\mu}{e\gamma} \left[\frac{\mathbf{v}_E}{c} \frac{\partial x}{\partial t} + c \nabla x \right]\end{aligned}\quad (26.24)$$

where $\gamma_E = (1 - E_{\perp}^2/B^2)^{-1/2}$ is the Lorentz factor associated to the drift velocity \mathbf{v}_E .

In the GCA Equation 26.23 information about motion perpendicular to the magnetic field is lost and the only component of the particle velocity that is actually evolved in time is u_{\parallel} , because the magnetic moment $\mu_0 \approx c\mu/e$ is now considered a constant of motion. Each term on the right hand side of Equation 26.23 corresponds to a specific drift motion. Indeed, the first term represents the $\mathbf{E} \times \mathbf{B}$ perpendicular drift, while the second one accounts for particle motion in the direction parallel to magnetic field. The first two terms in square bracket, $\mathcal{L}(\hat{b})$ and $\mathcal{L}(\mathbf{v}_E)$, describe the field curvature and polarization drifts², respectively. The next term is the $\nabla - B$ drift, while the last one represents a relativistic drift in the direction given by $\hat{b} \times \mathbf{v}_E$.

Equation 26.23 lose their validity when either $B \rightarrow 0$ (magnetic null) or $E_{\perp} \geq B$, i.e. when the nearly-crossed EM fields condition, Equation 26.21, is violated. While the first condition can easily occur inside a current sheet, the second may occur in a non-ideal MHD regime only. Violation of either condition breaks down the GCA and can lead to severe numerical errors.

26.2.5 Recursive solution

An explicit solution for Equation 26.20 can be found under Equation 26.21, which introduces a new parameter of smallness $\lambda/\omega \ll 1$.

²Notice that the $\mathcal{L}()$ operator is the Lagrangian derivative d/dt .

26.2.6 Energy conservation

The time component of Equation 26.20 leads to an evolutionary ODE for the particle energy as a function of time, namely

$$\frac{d\gamma c^2}{dt} = \frac{e}{m} \frac{d\mathbf{X}}{dt} \cdot \mathbf{E} + \frac{\mu}{m} \frac{\partial}{\partial t} \left(\frac{B}{\gamma_E} \right) \quad (26.25)$$

In the time-independent case, Equation 26.25 clearly shows that, when the GC velocity is perpendicular to the electric field, no acceleration occurs and the particle energy is conserved. For numerical purposes, however, Equation 26.25 is not solved for retrieving the Lorentz γ -factor, since large values of the right hand side could easily violate the condition $\gamma \geq 1$ at the truncation level of the scheme, unless small time steps are taken. In practice, to an 1st-order approximation, the particle velocity can be decomposed into a parallel component v_{\parallel} , a drift component v_E and a gyration component v_{\perp} . This leads to the normalization condition

$$\gamma^2(c^2 - v_{\parallel}^2 - v_{\perp}^2 - v_E^2) = c^2 \quad (26.26)$$

From the definition of the magnetic moment and the Larmor frequency $\mu_0 = mc\gamma^2v_{\perp}^2/(2eB) = \gamma^2v_{\perp}^2/(2\omega_0)$, we compute the particle Lorentz factor as

$$\gamma = \sqrt{\frac{c^2 + \gamma^2v_{\parallel}^2 + 2\mu_0\omega_0}{c^2 - v_E^2}} \quad (26.27)$$

Equation 26.27 replaces Equation 26.25 in common numerical solvers to get the Lorentz factor.

26.3 Validity and Higher-Order Corrections

The standard guiding center approximation is a first-order expansion in the Larmor radius ρ_L . Its validity depends on the variation of the magnetic field being small across the gyroradius. A rigorous criterion requires analyzing the tensor of field derivatives:

$$\frac{\sqrt{\lambda_{max}\rho_{\perp}}}{B} \ll 1$$

where λ_{max} is the maximum eigenvalue of the tensor involving gradients, curvature, and shearing of the field lines.

When field variations are strong (e.g., near sharp field reversals or current sheets), first-order approximations fail. Higher-order corrections (such as the Baños drift) may be required, which add terms proportional to the parallel curl of the magnetic field:

$$\langle \dot{z} \rangle - \langle v_{\parallel} \rangle \approx \frac{\mu B}{m\omega_0} \hat{b} \cdot (\nabla \times \hat{b})$$

This correction accounts for the discrepancy between the average particle velocity and the guiding center velocity due to field line shearing.

26.3.1 Numerical implementation

The GCA Equation 26.23 contains a set of 4 ODEs with unknowns $(\mathbf{X}, u_{\parallel})$. The terms which require evaluation at grids are

- \mathbf{B}
- $c\mathbf{E}$
- $(\hat{\mathbf{b}} \cdot \nabla)\hat{\mathbf{b}}$
- $(\hat{\mathbf{b}} \cdot \nabla)\mathbf{v}_E$
- $(\mathbf{v}_E \cdot \nabla)\mathbf{v}_E$
- $(\mathbf{v}_E \cdot \nabla)\hat{\mathbf{b}}$
- $\nabla(B/\gamma_E)$

Mignone, Haudemand, and Puzzoni (2023) reports the implementation of GCA in the PLUTO MHD model.

26.3.2 Nonrelativistic case

In the nonrelativistic case, $\gamma = 1$ so we don't need to solve Equation 26.27. Equation 26.23 can be simplified into

$$\begin{aligned} \frac{d\mathbf{X}}{dt} &= \\ \frac{d(v_{\parallel})}{dt} &= \end{aligned} \tag{26.28}$$

26.4 Hamiltonian Theory of Guiding Center Motion

Following the review by Cary and Brizard (2009).

Canonical and noncanonical Hamiltonian formulations of guiding-center motion offer improvements over non-Hamiltonian formulations: Hamiltonian formulations possess Noether's theorem (hence invariants follow from symmetries), and they preserve the Poincaré invariants (so that spurious attractors are prevented from appearing in simulations of guiding-center dynamics). Hamiltonian guiding-center theory is guaranteed to have an energy conservation law for time-independent fields—something that is not true of non-Hamiltonian guiding-center theories. The use of the phase-space Lagrangian approach facilitates this development, as there is no need to transform a priori to canonical coordinates, such as flux coordinates, which have less physical meaning.

26.4.1 Noncanonical Hamiltonian guiding center theory

The guiding-center phase space consists of

- the guiding-center position \mathbf{X} , essentially the center of the helix;
- the guiding-center parallel velocity variable $u \equiv \hat{b} \cdot \dot{\mathbf{X}}$;
- the (lowest-order) magnetic moment,

$$\mu \equiv \frac{m|\mathbf{w}|^2}{2B(\mathbf{X}, t)} \quad (26.29)$$

where $\mathbf{w} = \mathbf{v}_\perp - \mathbf{v}_E$ is the perpendicular velocity in the local frame moving with the $\mathbf{E} \times \mathbf{B}$ drift velocity $\mathbf{v}_E \equiv c\mathbf{E} \times \hat{b}/B$.

- and the ignorable gyrophase ζ , which gives the location of the particle on the circle about the guiding center.

As there are still six variables parametrizing phase space, there is no loss of information in making the guiding-center transformation $(\mathbf{x}, \mathbf{v}) \rightarrow (\mathbf{X}, u, \mu, \zeta)$. For the sake of simplicity of notation, we occasionally use the gyroaction variable $J \equiv mc/e\mu$ instead of the magnetic moment μ whenever we need to refer to the action-angle coordinates J, ζ associated with gyromotion.

The equations of motion for these variables are given by the guiding-center phase-space Lagrangian,

$$\mathcal{L}_{\text{gc}}(\mathbf{X}, u, \mu, \zeta; t) = \left[\frac{e}{c} \mathbf{A}(\mathbf{X}, t) + mu\hat{b}(\mathbf{X}, t) \right] \cdot \dot{\mathbf{X}} + J\dot{\zeta} - H_{\text{gc}} \quad (26.30)$$

in which the guiding-center Hamiltonian is given by

$$H_{\text{gc}}(\mathbf{X}, u, \mu; t) = \frac{m}{2}u^2 + \mu B(\mathbf{X}, t) + e\phi(\mathbf{X}, t) - \frac{m}{2}|\mathbf{v}_E|^2 \quad (26.31)$$

The arguments are shown here to emphasize that, for example, the magnetic-field strength $B(\mathbf{X}, t)$ is evaluated at the guiding-center position \mathbf{X} , not at the particle position \mathbf{x} . Here and throughout, the effects of a gravitational field may be found by adding the gravitational $m\Phi_G$ to the electrostatic potential energy $e\Phi$. In addition, we now drop the adjective phase space, as the guiding-center Lagrangian is always henceforth a phase-space Lagrangian.

The guiding-center Lagrangian Equation 26.30 comes not simply from gyrophase averaging, but from a transformation from the physical (particle) variables (\mathbf{x}, \mathbf{v}) to the guiding-center variables $(\mathbf{X}, u, \mu, \zeta)$. The relation between the physical location \mathbf{x} and the guiding-center position \mathbf{X} is

$$\mathbf{x} \equiv \mathbf{X} + \boldsymbol{\rho} \quad (26.32)$$

where $\boldsymbol{\rho}$ denotes the gyroradius displacement vector in the frame drifting with \mathbf{v}_E . Here we simply note that the displacement vector $\boldsymbol{\rho} \equiv \tilde{\boldsymbol{\rho}} + \bar{\boldsymbol{\rho}}$ has a part (denoted by tilde) that is

explicitly gyrophase dependent and a part (denoted by bar) that is gyrophase independent. In what follows, we show that the latter part

$$\bar{\boldsymbol{\rho}} = \frac{\hat{b}}{\Omega} \times \mathbf{v}_E = \frac{c\mathbf{E}_\perp}{\Omega B} \quad (26.33)$$

denotes the guiding-center polarization displacement (Kaufman, Phys. Fluids, 1986). ...

Our guiding-center Lagrangian Equation 26.30 is obtained from an ordering in which the scalar potential Φ is one order lower than the particle kinetic energy, unlike previous derivations of the Hamiltonian theory of guiding-center motion. In this ordering, the electric drift \mathbf{v}_E is of the same order as the perpendicular velocity \mathbf{w} , as in some non-Hamiltonian calculations (Northrop (1963)). As we will see, this ordering allows us to obtain the polarization drift in the same order as the curvature and ∇B drifts, although it appears differently in the theory. However, for consistency the parallel electric field $E_\parallel = \mathbf{E} \cdot \hat{b}$ must be smaller by one order than the perpendicular field \mathbf{E}_\perp (Equation 26.21).

The variables (J, ζ) appear in canonical form in the symplectic part of the guiding-center Lagrangian Equation 26.30 as $J\dot{\zeta}$ while the guiding-center Hamiltonian H_{gc} depends on J (or μ) alone. The Hamilton equations for (J, ζ) are

$$\begin{aligned} \dot{J} &= -\frac{\partial H_{\text{gc}}}{\partial \zeta} \equiv 0 \\ \dot{\zeta} &= \frac{\partial H_{\text{gc}}}{\partial J} \equiv \Omega \end{aligned} \quad (26.34)$$

Equation 26.34 shows that the gyroaction (or magnetic moment) is conserved by the guiding-center equations of motion. This also follows from Noether's theorem since the gyrophase ζ is an ignorable coordinate, i.e., only its time derivative appears in the guiding-center Lagrangian Equation 26.30.

If one is concerned with only the motion of the guiding center and not the evolution of the gyrophase, the term linear in $\dot{\zeta}$ can be dropped from the guiding-center Lagrangian, as it does not affect the equations of motion of the other variables, \mathbf{X} and u . In the evolution equations for \mathbf{X} and u , the adiabatic invariant μ (or J) does appear but only as a guiding-center dynamical parameter.

Variation of the guiding-center Lagrangian Equation 26.30 with respect to the variable u gives the Euler-Lagrange equation

$$0 = \frac{\partial \mathcal{L}}{\partial u} = m\hat{b} \cdot \dot{\mathbf{X}} - \frac{\partial H_{\text{gc}}}{\partial u}$$

which yields

$$u \equiv \hat{b} \cdot \dot{\mathbf{X}} \quad (26.35)$$

Thus, the guiding-center Lagrangian Equation 26.30 dictates that u is the velocity of the guiding center in the direction of the magnetic field at the guiding center.

Last, we vary the Lagrangian Equation 26.30 with respect to the guiding-center position \mathbf{X} , and obtain the Euler-Lagrange equation

$$\begin{aligned} mu\hat{b} &= e\mathbf{E} - \mu\nabla B + \frac{m}{2}\nabla|\mathbf{v}_E|^2 - mu\frac{\partial\hat{b}}{\partial t} \\ &\quad + \dot{\mathbf{X}} \times \left(\frac{e}{c}\mathbf{B} + mu\nabla \times \hat{b} \right) \\ &\equiv e \left(\mathbf{E}^* + \frac{1}{c}\dot{\mathbf{X}} \times \mathbf{B}^* \right) \end{aligned} \quad (26.36)$$

where the *effective* EM fields

$$\mathbf{E}^* \equiv -\nabla\Phi^* - \frac{1}{c}\frac{\partial\mathbf{A}^*}{\partial t} \quad \text{and} \quad \mathbf{B}^* \equiv \nabla \times \mathbf{A}^* \quad (26.37)$$

are defined in terms of the effective EM potentials

$$\begin{aligned} e\Phi^* &\equiv e\Phi + \mu B - \frac{m}{2}|\mathbf{v}_E|^2 \\ \mathbf{A}^* &\equiv \mathbf{A} + \frac{mc}{e}u\hat{b} \end{aligned} \quad (26.38)$$

The guiding-center canonical momentum is now simply expressed as $e\mathbf{A}^*/c$ and the guiding-center Hamiltonian is $e\Phi^* + mu^2/2$.

We obtain the rate of change of the variable u by taking the scalar product of Equation 26.36 with the effective magnetic field \mathbf{B}^*

$$\dot{u} = -\frac{\mathbf{B}^*}{mB_\parallel^*} \cdot \left(\nabla H_{\text{gc}} + \frac{e}{c}\frac{\partial\mathbf{A}^*}{\partial t} \right) \equiv \frac{e}{m}\frac{\mathbf{B}^*}{B_\parallel^*} \cdot \mathbf{E}^* \quad (26.39)$$

with $B_\parallel^* = \hat{b} \cdot \mathbf{B}^*$ the effective magnetic field in the parallel direction.

The time derivative Equation 26.39 contains terms that are higher order (in gyroradius) compared with the dominant terms, which are all that is usually kept. These higher-order terms, however, are needed for energy conservation.

The guiding-center velocity $\dot{\mathbf{X}}$ comes from the vector product of Equation 26.36 with \hat{b} which, using Equation 26.35, yields

$$\begin{aligned} \dot{\mathbf{X}} &= \frac{\mathbf{B}^*}{mB_\parallel^*}\frac{\partial H_{\text{gc}}}{\partial u} + \frac{c\hat{b}}{eB_\parallel^*} \times \left(\nabla H_{\text{gc}} + \frac{e}{c}\frac{\partial\mathbf{A}^*}{\partial t} \right) \\ &= u\frac{\mathbf{B}^*}{B_\parallel^*} + \mathbf{E}^* \times \frac{c\hat{b}}{B_\parallel^*} \end{aligned} \quad (26.40)$$

If the effective fields Equation 26.37 were replaced by the standard fields (\mathbf{E}, \mathbf{B}) , Equation 26.39 and Equation 26.40 would be the equations of motion for a particle in straight, constant EM fields.

The guiding-center equations of motion Equation 26.39 and Equation 26.40 can be used to derive the guiding-center energy equation

$$\frac{dH_{\text{gc}}}{dt} = \frac{\partial H_{\text{gc}}}{\partial t} + \dot{\mathbf{X}} \cdot \nabla H_{\text{gc}} + \dot{u} \frac{\partial H_{\text{gc}}}{\partial u} = e \frac{\partial \Phi^*}{\partial t} - \frac{e}{c} \frac{\partial \mathbf{A}^*}{\partial t} \cdot \dot{\mathbf{X}} \quad (26.41)$$

which implies that the guiding-center energy $E_{\text{gc}} \equiv mu^2/2 + e\Phi^*$ is a constant of the motion for time-independent fields.

An important remark must be made here concerning the polarization drift, which is absent from the guiding-center velocity Equation 26.40. This drift, however, is critical for obtaining the dielectric response of a low-frequency plasma. Instead, it appears in the transformation Equation 26.32 itself, i.e., the derivative of this relation gives

$$\dot{\mathbf{x}} = \dot{\mathbf{X}} + \dot{\tilde{\rho}} + \mathbf{v}_{\text{pol}} \quad (26.42)$$

with

$$\mathbf{v}_{\text{pol}} \equiv \frac{d\bar{\rho}}{dt} \quad (26.43)$$

representing the polarization drift, and where $\dot{\tilde{\rho}} = \dot{\zeta} \partial \tilde{\rho} / \partial \zeta + \dots$ consists of terms that oscillate on the gyroperiod time scale. The polarization drift is a pure derivative and, hence, can always be integrated. This implies, in particular, that *the polarization drift cannot lead to diffusion even in a turbulent field*. This is important, as the difference between the guiding center and the average location (found by dropping the second, oscillating term in Equation 26.32) is the polarization, the integral of the polarization drift. This difference must remain small or else the theory, which assumes that the particle remains close to \mathbf{X} , would break down.

An alternate set of guiding-center equations of motion may be derived in which the polarization drift appears explicitly in the guiding-center velocity $\dot{\mathbf{X}}$ by choosing $\bar{\rho} \equiv 0$ instead of Equation 26.33.

27 Cosmic Ray

Cosmic rays (CRs), that is the population of charged, relativistic particles with non-thermal spectra, are ubiquitous in the Universe. They pervade systems of all sizes, from stellar systems to whole galaxies, from galaxy clusters to the intercluster medium. Modelling of cosmic ray transport and interpretation of cosmic ray data ultimately rely on a solid understanding of the interactions of charged particles with turbulent magnetic fields.

Mertsch (2020) presents a thorough guide of performing test particle simulation for cosmic rays.

27.1 Basic Definitions

This part is from CERN handouts by Riaan Steenkamp, 2018.

Cosmic rays arrive at Earth isotropically, originating from all directions in outer space and spanning a vast range of energies, from about 10^9 eV up to beyond 10^{20} eV. Their sources include the Sun, supernova remnants, active galactic nuclei, and other energetic astrophysical phenomena. Number of particles in a volume element $d^3 r$ about \mathbf{r} in the momentum interval $d^3 p$ about \mathbf{p} is given by

$$dn = F(\mathbf{r}, \mathbf{p}, t) d^3 r d^3 p$$

where F is the distribution function. Within a space angle $d\Omega$ about the direction of \mathbf{p} it is

$$dn = F(\mathbf{r}, \mathbf{p}, t) d^3 r p^2 dp d\Omega$$

Typically we cannot directly measure F , but only averages over the momentum space. Averaging of F over all directions give

$$f(\mathbf{r}, p, t) = \frac{1}{4\pi} \int_{\Omega} F(\mathbf{r}, \mathbf{p}, t) d\Omega$$

where $\int_{\Omega} d\Omega = 4\pi$.

The number of particle dN in $d^3 r$ within $(p, p + dp)$ (independent of the direction of \mathbf{p}) is

$$dN = 4\pi p^2 f(\mathbf{r}, p, t) d^3 r dp = U_p(\mathbf{r}, p, t) d^3 r dp$$

where we define the differential number density U_p as the number of particles in d^3r about \mathbf{r} with momentum in the interval $(p, p + dp)$ as

$$U_p(\mathbf{r}, p, t) = 4\pi p^2 f(\mathbf{r}, p, t)$$

with units of [particles per volume].

Detectors do not measure U_p , but rather differential intensity. If particle velocity has magnitude v , the differential intensity is

$$j_p = v U_p$$

with units of [particles / unit momentum / unit area / unit time] that goes from all directions through the detector. This is usually normalized w.r.t. unit space angle

$$j_p = \frac{v U_p}{4\pi}$$

with units of [particles / unit momentum / unit area / unit time / unit radian]. In practice we measure particles w.r.t. kinetic energy intervals dT but not momentum intervals dp .

From the energy-momentum relation

$$E^2 = p^2 c^2 + E_0^2 = (pc)^2 + (m_0 c^2)^2$$

and energy conservation

$$E = E_0 + T$$

and $p = \gamma m_0 v$, we have

$$\begin{aligned} \frac{dE}{dp} &= \frac{E}{pc^2}, \quad dT = dE \\ \frac{dE}{dp} &= \frac{dT}{dp} = \frac{1}{v} \end{aligned}$$

Since we specify the same particles in two ways, we have $j_T dT = j_p dp$ and

$$j_T = j_p \frac{dp}{dT} = \frac{1}{v} j_p$$

Now we can summarize the following important relations

$$j_T = \frac{U_p}{4\pi} = p^2 f(\mathbf{r}, p, t)$$

That is, the differential intensity measured by experimentalists has simple relationship to quantities U_p and f used by theoreticians.

27.2 CR Continuity Equations

Consider the total time derivative of the distribution function $F(\mathbf{r}, \mathbf{p}, t)$.

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \mathbf{v} \cdot \nabla F + \dot{\mathbf{p}} \cdot \nabla_p F$$

with ∇_p the gradient operator w.r.t. the momentum space. Integrate this over all directions in the momentum space gives

$$\int_{\Omega} \frac{dF}{dt} d\Omega = \int_{\Omega} \left(\frac{\partial F}{\partial t} + \mathbf{v} \cdot \nabla F + \dot{\mathbf{p}} \cdot \nabla_p F \right) d\Omega$$

$$\begin{aligned} \int_{\Omega} \frac{dF}{dt} d\Omega &= \frac{d}{dt} \int F d\Omega = 4\pi \frac{df}{dt} \\ \int_{\Omega} \frac{\partial F}{\partial t} d\Omega &= \frac{d}{dt} \int F d\Omega = 4\pi \frac{\partial f}{\partial t} \end{aligned}$$

$$4\pi \frac{df}{dt} = 4\pi \frac{\partial f}{\partial t} + \int_{\Omega} \mathbf{v} \cdot \nabla F d\Omega + \int_{\Omega} \dot{\mathbf{p}} \cdot \nabla_p F d\Omega$$

We then need to evaluate the second and third terms on the RHS. From the x-component,

$$\begin{aligned} \int_{\Omega} v_x \frac{\partial F}{\partial x} d\Omega &= \int_{\Omega} \left[\frac{\partial}{\partial x} (v_x F) - \frac{\partial v_x}{\partial x} F \right] d\Omega \\ &= \frac{\partial}{\partial x} \int_{\Omega} v_x F d\Omega \end{aligned}$$

$$\begin{aligned} \langle v_x \rangle &= \frac{\int_{\Omega} v_x F d\Omega}{\int_{\Omega} F d\Omega} = \frac{1}{4\pi f} \int_{\Omega} v_x F d\Omega \\ 4\pi f \langle v_x \rangle &= \int_{\Omega} v_x F d\Omega \end{aligned}$$

Therefore

$$\int_{\Omega} v_x \frac{\partial F}{\partial x} d\Omega = 4\pi \frac{\partial}{\partial x} (\langle v_x \rangle f)$$

Generalize to the full 3D vector operator, we have an expression for the second term

$$\int_{\Omega} \mathbf{v} \cdot \nabla F d\Omega = 4\pi \nabla \cdot (\langle \mathbf{v} \rangle f)$$

The third term is more involved:

$$\begin{aligned}\int_{\Omega} \dot{\mathbf{p}} \cdot \nabla_p F d\Omega &= \int_{\Omega} [\nabla_p \cdot (\dot{\mathbf{p}}F - F \nabla_p \cdot \dot{\mathbf{p}})] d\Omega \\ &= \int_{\Omega} \nabla_p \cdot (\dot{\mathbf{p}}F) d\Omega\end{aligned}$$

since $\nabla_p \cdot \dot{\mathbf{p}} = 0$ because the EM force contains a cross product $\dot{\mathbf{p}} = q(\mathbf{E} + \mathbf{p} \times \mathbf{B}/m)$.

Using the Gauss' theorem

$$\int_{\Omega} \nabla_p \cdot (\dot{\mathbf{p}}F) d^3p = \int_{\Omega} (\dot{\mathbf{p}}F) \cdot \hat{e}_p p^2 d\Omega = \int_{\Omega} p^2 \dot{p} F d\Omega$$

If we integrate the above from p to $p + dp$,

$$\begin{aligned}\int_p^{p+dp} \nabla_p \cdot (\dot{\mathbf{p}}F) d^3p &= \int_{\Omega, p+dp} p^2 \dot{p} F d\Omega - \int_{\Omega, p} p^2 \dot{p} F d\Omega \\ &= 4\pi \left[(p^2 \dot{p} f)_{p+dp} - (p^2 \dot{p} f)_p \right] \\ &= 4\pi \frac{\partial}{\partial p} (p^2 \dot{p} f) dp\end{aligned}$$

Thus

$$\begin{aligned}\int_{\Omega} \dot{\mathbf{p}} \cdot \nabla_p F d\Omega &= \int_{\Omega} \nabla_p \cdot (\dot{\mathbf{p}}F) d\Omega \\ &= \frac{1}{p^2 dp} \int \nabla_p \cdot (\dot{\mathbf{p}}F) d^3p = \frac{4\pi}{p^2} \frac{\partial}{\partial p} (p^2 \langle \dot{p} \rangle f)\end{aligned}$$

and finally we have

$$4\pi \frac{df}{dt} = 4\pi \frac{\partial f}{\partial t} + 4\pi \nabla \cdot (\langle \mathbf{v} \rangle f) + \frac{4\pi}{p^2} \frac{\partial}{\partial p} (p^2 \langle \dot{p} \rangle f)$$

or by using $U_p = 4\pi p^2 f$ we have

$$\frac{dU_p}{dt} = \frac{\partial U_p}{\partial t} + \nabla \cdot \mathbf{S}_p + \frac{\partial}{\partial p} (\langle \dot{p} \rangle U_p)$$

where $\mathbf{S}_p = \langle \mathbf{v} \rangle U_p$ is the differential flux. Normally $\frac{dU_p}{dt} = 0$ and the above is simply a continuity equation. If there exist a source or sink Q_p , one can write

$$\frac{\partial U_p}{\partial t} + \nabla \cdot \mathbf{S}_p + \frac{\partial}{\partial p} (\langle \dot{p} \rangle U_p) = Q_p \quad (27.1)$$

The abovementioned cosmic-ray continuity equation is often called the **transport equation**.

27.3 Parker Transport Equation

Assumptions for the Parker Transport Equation:

- Guiding Center Approximation: Cosmic rays gyrate rapidly around magnetic field lines. We are interested in the motion of their guiding centers.
- Diffusion Approximation: Scattering is frequent enough that the distribution function f is nearly isotropic in the local plasma rest frame.
- Solar Wind Frame: We often work in a frame moving with the solar wind (or background plasma) velocity V_{sw} .
- Negligible Large-Scale Electric Field in Solar Wind Frame: In the frame co-moving with a perfectly conducting solar wind, the large-scale electric field is typically assumed to be zero, i.e., $\mathbf{E}' = \mathbf{E} + \mathbf{V}_{\text{sw}} \times \mathbf{B} \approx 0$. This means $\mathbf{E} \approx -\mathbf{V}_{\text{sw}} \times \mathbf{B}$.
- Omnidirectional Distribution Function: We are often interested in the omnidirectional distribution function $f_0(x, p, t)$, which is f averaged over all momentum directions in the plasma frame. Or, more commonly, the differential number density $U_p(x, p, t) = 4\pi p^2 f_0(x, p, t)$ (number of particles per unit volume per unit momentum magnitude) or $N(x, T, t)$ (differential intensity w.r.t. kinetic energy T).

In the lab frame, due to streaming and diffusion, the distribution function f is not perfectly isotropic. We can expand it as:

$$f(\mathbf{x}, p, t) = f_0(\mathbf{x}, p, t) + f_1(\mathbf{x}, p, t) + \dots$$

where f_0 is the isotropic part and f_1 is the first-order anisotropy.

$$f_1 = -\frac{\mathbf{p}}{p} \cdot \boldsymbol{\xi}(\mathbf{x}, p, t)$$

where $\boldsymbol{\xi}$ is the anisotropy vector. The particle flux (or streaming) \mathbf{S}_{cr} is related to this anisotropy.

The collision term describes the scattering of cosmic rays by magnetic field irregularities. In the diffusion approximation, this scattering tends to isotropize the distribution in the plasma rest frame. This process leads to spatial diffusion. The diffusive flux S_{diff} in the plasma frame is given by Fick's law:

$$\mathbf{S}_{\text{diff}} = -\vec{K} \cdot \nabla U_p$$

where U_p is the cosmic ray density (related to f_0) and \vec{K} is the diffusion tensor. The diffusion tensor is generally anisotropic:

$$K_{ij} = K_{\parallel} b_i b_j + K_{\perp} (\delta_{ij} - b_i b_j) + K_A \epsilon_{ijk} b_k$$

- K_{\parallel} is the diffusion coefficient parallel to the mean magnetic field direction \mathbf{b} .
- K_{\perp} is the diffusion perpendicular to \mathbf{b} .

- K_A is the antisymmetric part of the diffusion tensor describing gradient and curvature drifts.

Starting from the kinetic equation and averaging over pitch angle (assuming near isotropy in the plasma frame), one can derive an equation for f_0 . This is a lengthy process (e.g., see Schlickeiser 2002, “Cosmic Ray Astrophysics” or Parker 1965). The key steps involve:

- Transforming the kinetic equation to the plasma frame.
- Expanding f in momentum space in the plasma frame ($f = f_0 + f_1 + \dots$).
- The zeroth moment (integration over solid angle in momentum) of the kinetic equation gives the transport equation.
- The first moment gives a relation for the anisotropic part f_1 (streaming).

A common form of the Parker Transport Equation for the omnidirectional distribution function $f_0(\mathbf{x}, p, t)$ is

$$\frac{\partial f}{\partial t} + (\mathbf{V} + \mathbf{V}_d) \cdot \nabla f - \frac{1}{3} \nabla \cdot \mathbf{V} \frac{\partial f}{\partial \ln p} = \nabla \cdot (\vec{K} \cdot \nabla f) + Q$$

where $f(\mathbf{x}, p, t)$ is the particle distribution function as a function of the particle position \mathbf{x} , momentum p (isotropic momentum assumed), and time t ; \vec{K} is the spatial diffusion coefficient tensor, \mathbf{V} is the bulk plasma velocity, $\mathbf{V}_d = \frac{pv}{3q} \nabla \times (\mathbf{B}/B^2)$ is the particle drift, where $p = \gamma mv$ is particle momentum and q is particle charge.¹ The factor 1/3 comes from averaging over an isotropic pitch-angle distribution. Note that This is for an isotropic distribution; this term is sometimes incorporated into the anti-symmetric part (K_A) of the diffusion tensor \vec{K} . If so, the diffusion term becomes $\nabla \cdot (\vec{K}_s \cdot \nabla f_0 - \langle v_D \rangle f_0)$ where \vec{K}_s is the symmetric part.

- $\frac{1}{3} \nabla \cdot \mathbf{V} \frac{\partial f}{\partial \ln p}$ is the adiabatic energy change. The sign of $\nabla \cdot \mathbf{V}$ reflects whether cosmic rays lose or gain energy from the compression of the solar wind plasma.

27.4 Quasilinear Theory

More than 50 years since its inception, *quasi-linear theory (QLT)* [Jokipii (1966); Kennel and Engelmann 1966; Hall and Sturrock 1967; Hasselmann and Wibberenz 1970] is still very much the paradigm for phenomenological applications to Galactic cosmic rays. In QLT, the Fokker-Planck equation for the temporal evolution of the phase space density of CRs is derived in a perturbative approach where the force on a particle due to a turbulent magnetic field is evaluated along the unperturbed trajectory in a regular background field. The Fokker-Planck coefficients, most prominently the components of the spatial diffusion tensor, can be computed for a given model of turbulence, parametrised by the two-point function of the turbulent magnetic field. Famously, in QLT the interactions between plasma waves and particles are

¹This is the form given in the [GPAT model](#). Riaan Steenkamp shows a slightly different version for the guiding center drift in his notes which absorbs the adiabatic energy change into the diffusion tensor. Also note that $\frac{\partial f}{\partial \ln p} = p \partial f_0 / \partial p$.

found to be resonant, meaning that particles of a certain gyroradius $r_L = v/\Omega$, Ω denoting the gyrofrequency, are only affected by waves with a wavenumber k that satisfies $kr_L\mu \approx 1$ (for low-frequency waves like Alfvén waves), where μ is the cosine of the pitch-angle, that is the angle between the particle momentum \mathbf{p} and the regular magnetic field $\langle \mathbf{B} \rangle$, $\mu \equiv \mathbf{p} \cdot \langle \mathbf{B} \rangle / (|\mathbf{p}| |\langle \mathbf{B} \rangle|)$.

While some of QLT's predictions are qualitatively confirmed by data, e.g. the rigidity-dependence of the diffusion coefficients, there are a number of concerns. The most famous one is the 90° problem: Due to the resonance condition, particles with pitch-angle close to 90° ($\mu \approx 0$) can only be in resonance with very large wavenumbers k which for the usual turbulent spectra contain little energy. In the limit $\mu \rightarrow 0$, the scattering rate vanishes and particles cannot change direction along the background field resulting in *ballistic* transport. This is obviously at variance with the diffusive transport inferred from observations.

The root cause of the 90° problem is the assumption of unperturbed trajectories in QLT. This is remedied in non-linear theories, where the decay of correlations leads to a broadening of the resonance condition which allows for efficient enough scattering through 90°. However, any such extension of QLT requires additional assumptions, for instance on the form of temporal decorrelation of the particle's trajectory.

The success and popularity of QLT can be ascribed to its conceptual simplicity and validity in a number of important environments, including the solar wind, the interstellar medium and galaxy clusters. In addition, QLT is simple in principle and thus allows for a straightforward computation of the transport parameters, albeit it can become arbitrarily complex in practice. Finally, these results can be found to agree with inferences from observations, e.g. the normalisation and power law shape of the Galactic diffusion coefficient.

27.4.1 Derivation of the Fokker-Planck equation

For self-consistency, I will repeat all the equations here instead of linking to other parts of the note. Philipp Mertsch uses CGS unit in the following equations.

Charged particles in electric and magnetic fields \mathbf{E} and \mathbf{B} are subject to the Lorentz force,

$$\mathbf{F}_L = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}/c) \quad (27.2)$$

It is customary to decompose the magnetic field into a large-scale, homogeneous, regular background field, $\langle \mathbf{B} \rangle$ and a small-scale, turbulent, random field $\delta\mathbf{B}$, that is $\mathbf{B} = \langle \mathbf{B} \rangle$ with $\langle \delta\mathbf{B} \rangle = 0$. The angled brackets denote averages over an ensemble of turbulent magnetic fields.

Without loss of generality, we assume in the following that the regular field is oriented along the z -direction, $\langle \mathbf{B} \rangle = B_z \hat{z}$, unless stated otherwise. Large-scale electric fields are usually ignored, $\langle \mathbf{E} \rangle = 0$, as the large mobility of charges in astrophysical plasmas is efficiently shielding against regular electric fields (that is on scales much larger than the Debye length). Small-scale electric fields $\delta\mathbf{E}$ are necessarily present, but from Faraday's induction law, their magnitude

can be estimated to be $|\delta\mathbf{E}| \sim (v_A/c)|\delta\mathbf{B}|$ with v_A the Alfvén velocity and $v_A/c \ll 1$ in most astrophysical environments.² Thus, to lowest order in (v_A/c) , there is no electric field and as the magnetic force is not performing any work on the particle, particle energy is consequently conserved. Note that at higher orders in (v_A/c) , the particle energy can change in a second-order Fermi type process. For simplicity, we constrain ourselves here to considering the lowest order case which results in pitch-angle scattering.

A charged particle in a magnetic field forms a Hamiltonian system as long as dissipative processes (or any form of energy losses) can be ignored. A consequence of this is *Liouville's theorem*, that is the conservation of phase space volume under canonical transformations. As time evolution is a canonical transformation, phases space volume is conserved in time. Together with particle number conservation this implies the conservation of phase space density $f = f(\mathbf{r}, \mathbf{p}, t)$. This is conveniently captured by what we will call *Liouville's equation*,

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r}f + \frac{d\mathbf{p}}{dt} \cdot \mathbf{p}f = 0 \quad (27.3)$$

encoding the incompressibility of the phase space flow. Here,

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}, \quad \frac{d\mathbf{p}}{dt} = e \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \quad (27.4)$$

are the equations of motion. Note that a necessary condition for a Hamiltonian system is that the forces are conservative and differentiable (“p-divergence-free”, see more in (Bellan 2008)).

A collisionless plasma under the influence of external \mathbf{E} and \mathbf{B} fields is an example of a Hamiltonian system. Its Hamiltonian is (Jackson 1998)

$$H = \sqrt{(c\mathbf{P} - e\mathbf{A})^2 + m^2 c^4} + e\phi \quad (27.5)$$

where $\mathbf{P} = \mathbf{p} + (e/c)\mathbf{A}$ is the canonical momentum, \mathbf{A} the potential vector, m the particle mass, e its charge and ϕ the electric potential. Therefore, the phase space density of this collisionless plasma satisfies Equation 27.3 and substituting the Lorentz force, Equation 27.2, in Equation 27.3 gives the Vlasov equation,

$$\frac{\partial f}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r}f + e \left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \mathbf{p}f = 0 \quad (27.6)$$

which together with Maxwell's equations forms the basis of plasma kinetic theory. For a collisional plasma, a term needs to be added to the right-hand side, the famous collision operator. For a collisionless plasma (as appropriate for CRs) the right-hand side remains zero.

Considering turbulent fields, the phase space density also becomes a random field, $f = \langle f \rangle + \delta f$, with an expectation value, $\langle f \rangle$, and fluctuations around it, δf , that satisfy $\langle \delta f \rangle = 0$.

²In CGS units, the electric and magnetic fields have the same unit.

In any realistic astrophysical situation, it is of course impossible to know the small-scale turbulent field at all positions in order to exactly solve Equation 27.6. Instead, one can only hope to predict statistical moments of the phase space density for a statistical ensemble of turbulent magnetic fields. Traditionally, one is mostly interested in the first moment, the ensemble average, while higher order moments is also possible to obtain.

For the reason explained above, we will ignore \mathbf{E} in the following. Averaging Equation 27.6 we find (Jokipii 1972)

$$\begin{aligned}\frac{d\langle f \rangle}{dt} &= \frac{\partial \langle f \rangle}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r} \langle f \rangle + e \frac{\mathbf{v} \times \langle \mathbf{B} \rangle}{c} \cdot \mathbf{p} \langle f \rangle \\ &= - \left\langle e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \nabla_{\mathbf{p}} \delta f \right\rangle \neq 0\end{aligned}\quad (27.7)$$

Note that unlike the phase space density f , the ensemble averaged phase space density $\langle f \rangle$ is not conserved, $d\langle f \rangle/dt \neq 0$. (More on this later!)

One way to glean some physical insight from Equation 27.7 is to identify its right-hand-side with a damping term, (Earl+ 1988; Webb 1989),

$$\left\langle e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \nabla_{\mathbf{p}} \delta f \right\rangle \rightarrow \nu \left(\langle f \rangle - \frac{1}{4\pi} \int d\hat{\mathbf{p}} \langle f \rangle \right)$$

(where $\hat{\mathbf{p}} = \mathbf{p}/|\mathbf{p}|$), that is driving the phase space density towards isotropy at a rate ν , an approach that can also be motivated by gas kinetic theory (Bhatnagar+ 1954). This way, Equation 27.7 can be solved and shown to lead to a spatial diffusion equation. The parallel diffusion coefficient can be identified as $\kappa_{\parallel} = v^2/(3\nu)$ whereas the perpendicular diffusion coefficient satisfies $\kappa_{\perp}/\kappa_{\parallel} = (1 + \Omega^2/\nu^2)^{-1}$, a result referred to as the *classical scattering limit* (Gleeson 1969). Here, Ω is the particle's gyrofrequency.

In QLT, however, a more systematic solution for f is sought through an equation for the temporal evolution of the fluctuations δf . Such an equation can be obtained by subtracting the ensemble-averaged Vlasov Equation 27.7 from the original Vlasov Equation 27.6,

$$\begin{aligned}\frac{\partial \delta f}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r} \delta f + e \left(\frac{\mathbf{v} \times \langle \mathbf{B} \rangle}{c} \right) \cdot \mathbf{p} \delta f \\ \simeq -e \left(\frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \mathbf{p} \langle f \rangle\end{aligned}\quad (27.8)$$

Here, we have chosen to ignore the difference

$$e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \nabla_{\mathbf{p}} \delta f - \left\langle e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \nabla_{\mathbf{p}} \delta f \right\rangle$$

which is second order in perturbed quantities, $\delta\mathbf{B}$ and δf . This assumes, of course, that $|\delta\mathbf{B}| \ll |\langle \mathbf{B} \rangle|$ and therefore $\delta f \ll \langle f \rangle$. Equation 27.8 can now be integrated with the method of characteristics, the formal solution being

$$\delta f = \delta f_0 - \int_{t_0}^t dt' \left[e \left(\frac{\mathbf{v} \times \mathbf{B}}{c} \right) \cdot \mathbf{p} \langle f \rangle \right]_{P(t')} \quad (27.9)$$

Here, $\delta f_0 \equiv \delta f(\mathbf{r}, \mathbf{p}, t_0)$ denotes the phase space density at time t_0 and the subscript $P(t')$ indicates that positions and momenta in the square brackets are to be evaluated along the characteristics of Equation 27.8, that is the solutions of the equations of motions, Equation 27.4 with \mathbf{B} replaced by the regular field $\langle \mathbf{B} \rangle$ only (and again no electric field). These solutions P are commonly referred to as *unperturbed orbits* or *unperturbed trajectories*. For the homogeneous regular magnetic field $\langle \mathbf{B} \rangle = B_z \hat{z}$ assumed here they are of course helices along the z-direction.

We can now substitute Equation 27.9 into Equation 27.7,

$$\begin{aligned} \frac{\partial \langle f \rangle}{\partial t} + \frac{d\mathbf{r}}{dt} \cdot \mathbf{r} \langle f \rangle + e \frac{\mathbf{v} \times \langle \mathbf{B} \rangle}{c} \cdot \mathbf{p} \langle f \rangle \\ \simeq \int_{t_0}^t dt' \left\langle e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \nabla_{\mathbf{p}} \left[e \frac{\mathbf{v} \times \mathbf{B}}{c} \cdot \mathbf{p} \langle f \rangle \right]_{P(t')} \right\rangle \end{aligned} \quad (27.10)$$

where we have dropped the term $\propto \delta f_0$. At this stage, we can already see that the right-hand side will lead to diffusion terms (courtesy of the two momentum derivatives) and that it depends on the turbulent magnetic field's two-point function, integrated along the unperturbed trajectory $P(t')$. To make further progress, we consider the momentum \mathbf{p} in spherical coordinates, that is $\mathbf{p} = p (\sqrt{1 - \mu^2} \cos \phi, \sqrt{1 - \mu^2} \sin \phi, \mu)^T$ and introduce the correlation lengths l_c and correlation times τ_c of the turbulent magnetic field, defined through³

$$\langle \delta B^2 \rangle l_c \equiv \int_0^\infty d\Delta r \langle \delta B(\mathbf{r}) \delta B(\mathbf{r} + \mathbf{r}) \rangle \quad (27.11)$$

$$\langle \delta B^2 \rangle \tau_c \equiv \int_0^\infty d\Delta t \langle \delta B(t) \delta B(t + \Delta t) \rangle \quad (27.12)$$

Here, $\delta B(t)$ is short-hand for $[\delta B(\mathbf{r})]_{P(t)}$, that is $\delta B(\mathbf{r})$ evaluated along the unperturbed trajectory at time t .

The right-hand side of Equation 27.10 is still rather unwieldy and further progress requires a number of assumptions. In addition to

1. Smallness of perturbations, $|\mathbf{B}| \ll |\langle \mathbf{B} \rangle|$

these are:

2. Gyrotropy: The ensemble-averaged phase space density $\langle f \rangle$ does not depend on the azimuthal angle ϕ , so $\langle f \rangle(\mathbf{r}, p, \mu, \phi, t) \rightarrow \langle f \rangle(\mathbf{r}, p, \mu, t)$.
3. Adiabatic approximation: The phase space density only varies on time-scales much larger than the correlation time of the turbulent magnetic field, τ_c ,

$$\langle f \rangle \left/ \frac{\partial \langle f \rangle}{\partial t} \right. \gg \tau_c \quad (27.13)$$

³Strictly speaking, the correlation lengths and times are tensors because of the vector nature of the magnetic field in the two-point functions; here, however, we only require them for order of magnitude arguments, so we do not distinguish the different components.

4. Finite correlation times: The correlation times of the turbulent magnetic field are much larger than the Larmor time, $\tau_c \gg \Omega^{-1}$.
5. Homogeneous and stationary turbulence: the statistical moments are invariant under translations in space and time. In natural language, it means that the statistics of turbulence do not depend on when and where to take measurements.

Under these conditions, the ensemble averaged Vlasov equation ultimately results in a Fokker-Planck type equation (Fokker 1914; Planck 1917), also known as the Kolmogorov forward (Kolmogorov 1931) or as the Smoluchowski equation (Bogoliubov and Krylov 1939), describing diffusion in pitch-angle,

$$\frac{\partial \langle f \rangle}{\partial t} + v\mu \frac{\partial \langle f \rangle}{\partial z} = \frac{\partial}{\partial \mu} \left(D_{\mu\mu} \frac{\partial \langle f \rangle}{\partial \mu} \right). \quad (27.14)$$

Following the approach sketched above, the pitch-angle diffusion coefficient,

$$D_{\mu\mu} \equiv \frac{\langle (\Delta\mu)^2 \rangle}{2\Delta t} \quad (27.15)$$

can be expressed in terms of the correlation function of the magnetic field. (HOW???)

If we had not decided to ignore any electric field, additional terms would have appeared in the Fokker-Planck Equation 27.14, relating to changes in momentum \mathbf{p} and pitch-angle, with diffusion coefficients $D_{\mu p} = D_{p\mu}$ and D_{pp} defined analogously to Equation 27.15.

We have furthermore assumed that $v_A/v \ll 1$ in order for D_{xx}, D_{yy}, D_{xy} and D_{yx} to be negligible. Not doing so, would have resulted in the additional terms

$$\frac{\partial}{\partial x} \left(D_{xx} \frac{\partial \langle f \rangle}{\partial x} + D_{xy} \frac{\partial \langle f \rangle}{\partial y} \right) + \frac{\partial}{\partial y} \left(D_{yx} \frac{\partial \langle f \rangle}{\partial x} + D_{yy} \frac{\partial \langle f \rangle}{\partial y} \right)$$

to be added to the right-hand side of Equation 27.14.

In summary, under the influence of a turbulent magnetic field, charged particles are performing a random walk in pitch-angle which in the ensemble average results in diffusion in pitch-angle (cosine).

27.4.2 The diffusion approximation

Particle transport can be conveniently categorised if the mean-square displacement in direction i , $\langle \Delta x_i^2 \rangle$, has a power law dependence,⁴

$$\langle \Delta r_i^2 \rangle \propto (\Delta t)^\alpha \quad (27.16)$$

⁴In particle transport, ballistic refers to the movement of particles with minimal influence from their environment. Imagine a particle traveling on a frictionless, straight path, almost like a bullet. That's ballistic transport!

as

- $\alpha < 1$: sub-diffusive,
- $\alpha = 1$: diffusive,
- $\alpha > 1$: super-diffusive, in particular
- $\alpha = 2$: ballistic.

It seems clear that transport in any perturbative theory with $|\mathbf{B}| \ll |\langle \mathbf{B} \rangle|$ must be ballistic at early enough times: Particles just gyrate around $\langle \mathbf{B} \rangle$ and $\langle \Delta z^2 \rangle = (v\mu\Delta t)^2$ while $\langle \Delta x^2 \rangle = \langle \Delta y^2 \rangle = 0$ when integrated over full gyroperiods. At late times, that is for $t \gg D_{\mu\mu}^{-1}$, we would expect diffusive behaviour for the transport along the field.

In order to formalise this picture, we derive a spatial diffusion equation from the Fokker-Planck Equation 27.14. To this end, we decompose $\langle f \rangle$ into an isotropic part, g , and an anisotropic part, h ,

$$\langle f \rangle(p, \mu, t) = g(p, t) + h(p, \mu, t)$$

where

$$g(p, t) = \frac{1}{2} \int_{-1}^1 d\mu \langle f \rangle(p, \mu, t)$$

and

$$\int_{-1}^1 d\mu h(p, \mu, t) = 0$$

If g varies only slowly with time and position,

$$g \left/ \frac{\partial g}{\partial t} \right. \gg \tau_{sc} \quad \text{and} \quad g \left/ \frac{\partial g}{\partial z} \right. \gg \lambda_{sc}$$

where $\tau_{sc} \sim D_{\mu\mu}^{-1}$ and $\lambda_{sc} \sim v\tau_{sc}$ are the scattering time and mean-free path, respectively, the phase space density will be very isotropic, $h \ll g$. In this case, we can derive a spatial diffusion equation for the isotropic part g (e.g. Hasselmann and Wibberenz 1970),

$$\frac{\partial g}{\partial t} - \frac{\partial}{\partial z} \left(\kappa_{\parallel} \frac{\partial g}{\partial z} \right) = 0 \tag{27.17}$$

with the parallel diffusion coefficient

$$\kappa_{\parallel} = \frac{v^2}{8} \int_{-1}^1 d\mu \frac{(1 - \mu^2)^2}{D_{\mu\mu}} \tag{27.18}$$

Furthermore, we would expect the anisotropic part h to be dominated by the *dipole anisotropy* (??), that is $h \approx h_1 \mu$ with

$$h_1 = \frac{3}{2} \int_{-1}^1 d\mu \mu h(\mu) = -\frac{2}{v} \kappa_{\parallel} \frac{\partial g}{\partial z} \tag{27.19}$$

27.4.3 Computation of transport coefficients

So far, we have not specified the functional form of the Fokker-Planck coefficients, e.g. the pitch-angle diffusion coefficient $D_{\mu\mu}$, and its dependence on the two-point correlation function of turbulence $P_{ij}(\mathbf{k})$ that emerges in the derivation of the Fokker-Planck Equation 27.14. An alternative to the derivation of Section 27.4.1 is to directly compute the Fokker-Planck coefficients from solutions of the equations of motion. In fact, an arbitrary Fokker-Planck coefficient D_{PQ} can be defined in terms of the mean displacements of the variables in question, P and Q. For instance, the pitch-angle diffusion coefficient can be derived as the $t \rightarrow \infty$ limit of the running diffusion coefficient,

$$D_{\mu\mu}(t) = \frac{1}{2} \frac{d}{dt} \langle (\Delta\mu)^2 \rangle \quad (27.20)$$

This is a consequence of the Taylor-Green-Kubo formula (Taylor 1922; Green 1951; Kubo 1957),

$$D_{\mu\mu} = \int_0^\infty dt \langle \dot{\mu}(0) \dot{\mu}(t) \rangle \quad (27.21)$$

where the dots denote derivatives with respect to time. For diffusive transport, Equation 27.15 and Equation 27.20 coincide, of course. Moreover, this allows computing the parallel diffusion coefficient κ_\parallel without the detour of computing $D_{\mu\mu}$ first and then applying the diffusion approximation, Equation 27.18.

From the equations of motion Equation 27.2, we find (HOW???)

$$\dot{\mu} = \frac{e}{cp} (\mathbf{v} \times \mathbf{B})_z = \frac{1}{r_g B_z} (v_x \delta B_x(\mathbf{r}) - v_y \delta B_y(\mathbf{r})) \quad (27.22)$$

and thus

$$D_{\mu\mu} = \frac{1}{B_z^2 r_g^2} \int_0^\infty dt [v_x(t) v_x(0) \mathcal{P}_{yy}(t) + v_y(t) v_y(0) \mathcal{P}_{xx}(t)] \quad (27.23)$$

Here, we have defined

$$\mathcal{P}_{ij}(t) \equiv \langle \delta B_i(0) \delta B_j(t) \rangle \quad (27.24)$$

and both the velocities and the magnetic fields are to be evaluated along unperturbed trajectories. Note that the fact that the Fokker-Planck coefficients only depend on the two-point function means that we can constrain ourselves to the Gaussian part of the turbulent magnetic field.

27.4.4 Turbulence geometries and spectra

To make further progress, we need to specify the turbulence correlation tensor \mathcal{P}_{ij} . In the derivation of the Fokker-Planck equation we had to assume that turbulence is homogeneous

and stationary (assumption 5). In this case, the field can be represented very economically in Fourier space. To this end, we introduce the Fourier transform pair

$$\begin{aligned}\delta\tilde{B}_j(\mathbf{k}, t) &= (2\pi)^{-3/2} \int_{-\infty}^{\infty} d^3r \delta B_j(\mathbf{r}, t) e^{i\mathbf{k}\cdot\mathbf{r}} \\ \delta B_j(\mathbf{x}, t) &= (2\pi)^{-3/2} \int_{-\infty}^{\infty} d^3k \delta\tilde{B}_j(\mathbf{k}, t) e^{-i\mathbf{k}\cdot\mathbf{r}}\end{aligned}\quad (27.25)$$

Note that for the magnetic field to have real values, $\delta B_j(\mathbf{r}) = \delta B_j^*(\mathbf{r})$, requires a relation between the Fourier components and their complex conjugates,

$$\delta\tilde{B}_j^*(\mathbf{k}) = \delta\tilde{B}_j(-\mathbf{k})$$

The homogeneity and stationarity now guarantee that the two-point functions $\langle \delta B_i(\mathbf{r}, t) \delta B_j(\mathbf{r}', t') \rangle$ depend on the positions \mathbf{r} and \mathbf{r}' and times t and t' only through the differences $\Delta\mathbf{r} \equiv (\mathbf{r} - \mathbf{r}')$ and $(t - t')$. It is then easy to see that the two-point function in Fourier space is diagonal,

$$\begin{aligned}\langle \delta\tilde{B}_i(\mathbf{k}, t) \delta\tilde{B}_j^*(\mathbf{k}', t') \rangle &= (2\pi)^{-3} \int d^3r e^{i\mathbf{k}\cdot\mathbf{r}} \int d^3r' e^{-i\mathbf{k}'\cdot\mathbf{r}'} \langle \delta B_i(\mathbf{r}, t) \delta B_j(\mathbf{r}', t') \rangle \\ &= \delta^{(3)}(\mathbf{k} - \mathbf{k}') P_{ij}(\mathbf{k}, t - t')\end{aligned}$$

where the turbulence correlation tensor $P_{ij}(\mathbf{k}, \Delta t)$ is the Fourier transform of the coordinate space two-point function

$$\begin{aligned}P_{ij}(\mathbf{k}, \Delta t) &\equiv (2\pi)^{-3/2} \int_{-\infty}^{\infty} d^3(\Delta r) e^{i\mathbf{k}\cdot\Delta\mathbf{r}} \\ &\times \langle \delta B_i(\mathbf{r}, t) \delta B_j(\mathbf{r} - \Delta\mathbf{r}, t') \rangle\end{aligned}\quad (27.26)$$

It contains all the (statistical) information on the magnetic turbulence that enters into the computation of the Fokker-Planck coefficients:

- information on the turbulence geometry, for instance whether there is a preferred direction for the propagation of waves;
- information on the turbulence spectrum, that is the distribution of energy among different turbulent scales;
- information on the time-dependence of the correlations.

Oftentimes, it is assumed that $P_{ij}(\mathbf{k}, \Delta t)$ factorises into a magnetostatic correlation tensor $P_{ij}(\mathbf{k}) \equiv P_{ij}(\mathbf{k}, 0)$ independent of time and a time-dependent dynamical correlation function $\Gamma(\mathbf{k}, \Delta t)$,

$$P_{ij}(\mathbf{k}, \Delta t) = P_{ij}(\mathbf{k}) \Gamma(\mathbf{k}, \Delta t) \quad (27.27)$$

In the magnetostatic approximation, we ignore any time-dependence altogether, so

$$\Gamma \equiv 1$$

While in reality P_{ij} may be arbitrarily complicated, three turbulence geometries have dominated much of the literature, both in analytical studies of transport coefficients and numerical test particle simulations. These three geometries are conceptually simple and particularly amenable to analytical computations of the components of the diffusion tensor and the other Fokker-Planck coefficients:

- 3D isotropic turbulence
- slab turbulence
- a composition of slab and 2D isotropic turbulence

27.4.4.1 3D isotropic turbulence

For 3D isotropic turbulence the magnetostatic correlation tensor takes the form (Batchelor 1982)

$$P_{ij}^{3D}(\mathbf{k}) = g^{3D}(k) \left(\delta_{ij} - \frac{k_i k_j}{k^2} + \imath \sigma(k) \epsilon_{ijm} \frac{k_m}{k} \right) \quad (27.28)$$

The \mathbf{k} -dependent real functions $g^{3D}(k)$ and $\sigma(k)$ allow modelling of the overall spectrum and of a wavenumber-dependent helicity, respectively. Note that for linearly polarised waves $\sigma(k) \equiv 0$. The normalization of $g^{3D}(k)$ is fixed by requiring

$$\begin{aligned} \delta B^2 \equiv \langle \mathbf{B}^2(\mathbf{x}) \rangle &= \int d^3k (P_{xx}^{3D}(\mathbf{k}) + P_{yy}^{3D}(\mathbf{k}) + P_{zz}^{3D}(\mathbf{k})) \\ &= 2 \int d^3k g^{3D}(k) = 8\pi \int_0^\infty dk k^2 g^{3D}(k) \end{aligned} \quad (27.29)$$

27.4.4.2 Slab turbulence

In slab turbulence, it is assumed that all quantities are independent of the coordinates perpendicular to the background field (x and y) and that the turbulent field has no z -component. Consequently, the wave vectors $\mathbf{k} \parallel \hat{z}$ and if we further demand turbulence to be axisymmetric, the turbulence correlation tensor reads

$$P_{ij}^{\text{slab}}(\mathbf{k}) = g^{\text{slab}}(k_\parallel) \frac{\delta(k_\perp)}{k_\perp} (\delta_{ij} + \imath \sigma(k_\parallel) \epsilon_{ijz}) \quad (27.30)$$

for $i, j \in x, y$ and zero otherwise. In our case $k_{\parallel} = k_z$ and $k_{\perp} = \sqrt{k_x^2 + k_y^2}$. Again, $\sigma(k_{\parallel})$ allows for wavenumber-dependent helicity, but vanishes for linear polarisation. The normalisation is then

$$\begin{aligned}\delta B^2 &\equiv \int d^3k (P_{xx}^{\text{slab}}(\mathbf{k}) + P_{yy}^{\text{slab}}(\mathbf{k})) \\ &= 4\pi \int_{-\infty}^{\infty} dk_{\parallel} g^{\text{slab}}(k_{\parallel})\end{aligned}\tag{27.31}$$

While slab turbulence might seem rather restrictive a turbulence model, it is quite attractive due to its simplicity. In addition, it could be argued that it is of physical relevance in situations where the turbulence is self-generated by anisotropies in the distribution of CRs (Kulsrud and Pearce 1969; Skilling 1975): It has been shown (e.g. Tademaru 1969) that the modes with wavevectors along the background magnetic field grow fastest.

27.4.4.3 Composite (slab + 2D isotropic) turbulence

27.4.4.4 Turbulence spectra

Having reviewed three simple turbulence geometries, we need to specify the spectral shapes $g(k)$ in order to compute transport coefficients. In cascade models of turbulence, energy is injected on the largest scales in the so-called *injection (energy) range*. Non-linear interactions transfer energy to smaller scales over the so-called *inertial range*. At very small scales, the turbulent energy is dissipated in the so-called *dissipation range*. The scale separating the injection and inertial ranges is called the *outer scale of turbulence* and the scale separating the inertial and the dissipation range is called the *dissipation scale*.

Most studies deal with one of two spectra. The first one is a simple power law with spectral index q and low wavenumber cut-off k_0 , corresponding to the outer scale ($2\pi/k_0$),

$$g_{\text{PL}}(k) = \begin{cases} g_0(k/k_0)^{-q} & \text{for } k \geq k_0, \\ 0 & \text{otherwise.} \end{cases}$$

The alternative is a broken power law with a flat spectrum below the wavenumber k_0 and a power law slope q above,

$$g_{\text{BPL}}(k) = g_0 \left(1 + \left(\frac{k}{k_0} \right)^{1/s} \right)^{-qs}.$$

Here, s is parametrising the softness of the break and $s \rightarrow 0$ corresponds to a sharp break. It is assumed that the broken power law form can potentially also capture turbulence in the injection range, that is for $k < k_0$.

27.4.4.5 Slab turbulence with broken power law spectrum

The result for the pitch-angle diffusion coefficient in slab turbulence and for the broken power law spectrum (Shalchi 2009) is given as an example.

$$g^{\text{slab}}(k_{\parallel}) = \frac{C(q, \frac{1}{2})}{2\pi k_0} \delta B^2 \left(1 + \left(\frac{k}{k_0} \right)^2 \right)^{-q/2}. \quad (27.32)$$

The function $C(q, s)$ is fixed by the normalisation condition, see Equation 27.31,

$$\begin{aligned} \frac{1}{C(q, s)} &\equiv \frac{4}{k_0} \int_0^\infty dk \left(1 + \left(\frac{k}{k_0} \right)^{1/s} \right)^{-qs} \\ &= \frac{4}{k_0} \frac{\Gamma(s(q-1))\Gamma(1+s)}{\Gamma(qs)} \end{aligned} \quad (27.33)$$

where $\Gamma(\cdot)$ denotes the gamma function. We have assumed that $q > 1$ in order for the k-integral in Equation 27.33 to converge. For $q = 1$, instead, we need to assume a cut-off, that is $g^{\text{slab}} = 0$ for $k_{\parallel} > k_{\max}$.

Substituting Equation 27.32 into Equation 27.30 and Equation 27.23, one encounters the resonance function (Schlickeiser 2002)

$$R^{\text{slab}} = \pi \delta(k_{\parallel} \mu v \pm \Omega) \quad (27.34)$$

Eventually, this simplifies to

$$D_{\mu\mu} = \frac{\pi}{2} C\left(q, \frac{1}{2}\right) q k_0 \frac{\delta B^2}{B_z^2} \frac{(1 - \mu^2) \mu^{q-1} (r_L k_0)^{q-2}}{(1 + \mu^2 (r_L k_0))^{q/2}} \quad (27.35)$$

27.4.5 Field-line random walk

The computation of the pitch-angle diffusion coefficient in Equation 27.21 and Equation 27.23 is based on an evaluation of the turbulent part of the Lorentz force along trajectories around the homogeneous background field. As long as perturbations are small, this gives the dominant contribution to the parallel diffusion coefficient, Equation 27.18.

For perpendicular transport, however, there is another important contribution due to the fact that the field line is not perfectly homogeneous. Instead, the large-scale magnetic field evaluated for a particle along a field line changes direction with distance along this field line. Under certain conditions, this movement can be shown to be diffusive, see below. If the movement of the particle due to this effect is included in the computation of the mean-square displacements (or equivalently through the Taylor-Green-Kubo approach), this gives the so-called *field-line*

random walk (FLRW) contribution to perpendicular transport. The contribution without this is oftentimes called the *microscopic contribution*.

For slab turbulence, the microscopic diffusion coefficient vanishes (the transport is in fact sub-diffusive), hence FLRW gives the only contribution. For other turbulence geometries, FLRW can also contribute, but might not be dominating.

Let's again assume the regular background field $\langle \mathbf{B} \rangle = B_z \hat{z}$ to be dominating over the perturbations $\delta \mathbf{B}$. The equation determining the field line $\{x(z), y(z)\}$ is

$$\frac{dx}{dz} = \frac{\delta B_x}{B_z}, \quad \frac{dy}{dz} = \frac{\delta B_y}{B_z}$$

This can formally be integrated to obtain the mean square displacement in the perpendicular directions, e.g.

$$\langle (\Delta x)^2 \rangle = \frac{1}{B_z^2} \int_0^z dz' \int_0^z dz'' \langle \delta B_x(\mathbf{r}(z')) \delta B_x(\mathbf{r}(z'')) \rangle \quad (27.36)$$

In slab turbulence, the integrand only depends on z and it is easy to show that the perpendicular mean-square displacement $\langle (\Delta r_{\perp})^2 \rangle$ is ballistic at small z and diffusive for large z , e.g.

$$\langle (\Delta x)^2 \rangle = \begin{cases} z^2 (\delta B_x / B_z)^2 & \text{for } z \rightarrow 0, \\ 2\kappa_{\text{FLRW}} |z| & \text{for } z \rightarrow \infty, \end{cases}$$

with the FLRW diffusion coefficient

$$\kappa_{\text{FLRW}} = \frac{2\pi^2}{B_z^2} g^{\text{slab}}(0)$$

In other turbulence geometries, the integrand in Equation 27.36 also depends on x and y , such that an explicit solution is not possible without further assumptions.

If particles are assumed to diffuse along field lines, $\langle (\Delta z)^2 \rangle \propto \Delta t$, FLRW leads to subdiffusive perpendicular transport, $\langle (\Delta r_{\perp})^2 \rangle \propto \sqrt{\Delta t}$, a phenomenon known as *compound (sub)diffusion* (Jokipii 1966). Compound subdiffusion has been applied to a variety of environments like laboratory plasmas, the heliosphere, near-source transport as well as shock acceleration.

27.4.6 Short-comings of QLT

Despite its popularity, QLT exhibits a number of issues. The most well-known pathology of magnetostatic QLT is its inability to scatter particles through 90° . While present in a number of turbulence geometries, it is easiest illustrated in slab turbulence where the dependence of the pitch-angle diffusion coefficient $D_{\mu\mu}$ on the spectrum $g^{\text{slab}}(k)$ becomes very simple. In fact, inspecting Equation 27.35 we see that $D_{\mu\mu} \rightarrow 0$ for $\mu \rightarrow 0$.

The root cause for the 90° problem is the narrow resonance condition in magnetostatic QLT, $k_{\parallel}\mu r_L = \pm 1$ (Equation 27.34). Particles at finite μ are in resonance with waves of finite parallel wavenumber, $k_{\parallel} = \pm 1/(\mu r_L)$. For $\mu \rightarrow 0$, however, the resonant parallel wavenumber grows without bounds. With the turbulence spectra being falling power laws, however, there is only little energy at small scales and the pitch-angle scattering rate vanishes. In practice, there is of course no energy at all at scales below the dissipation scale.

27.5 Test Particle Simulation

In the absence of a widely accepted extension of quasi-linear theory, wave-particle interactions must also be studied in numerical simulations where the equations of motion are directly solved in a realisation of the turbulent magnetic field. The applications of such test particle simulations of cosmic rays are manifold: testing transport theories, computing parameters like diffusion coefficients or making predictions for phenomena beyond standard diffusion theories, e.g. for cosmic ray small-scale anisotropies.

Ideally, one would test these non-linear theories beyond QLT by comparing their predictions with data from observations. While this approach is being followed by the heliospheric community, a difficulty remains in that the actual turbulence (if known) turns out to be much more complex than what is routinely assumed in analytical transport theories. Alternatively, transport theories can be tested by comparing their predictions with those from numerical experiments.

Test particle simulations compute the transport of high-energy charged particles through prescribed EM fields without taking into account the effect of the high-energy particles on the EM fields. To this end, a realisation of the turbulent magnetic field is generated and the equations of motion are solved for test particles, that is the contributions of the CR particles to the EM fields are ignored. Given the trajectories of a large enough number of test particles, one can numerically compute the Fokker-Planck coefficients.

In addition to testing transport theories by comparing the analytically computed diffusion coefficients to simulated ones, there are at least two more applications of test particle simulations:

1. For sources at distances closer or similar to the scattering mean free path, the diffusive transport theory is not necessarily applicable. An often-cited pathology of computing solutions to the diffusion equation is superluminal propagation speeds. Lately, there has been increased interest in the transition between the ballistic and diffusive phases of transport and test particle simulations allow exploring this transition for given turbulent EM fields.
2. Analytical transport theories usually make predictions only for the ensemble-averaged phase-space density and it is usually assumed that the observed phase-space densities are close to the ensemble average. This has now been called into question, in particular in

view of the observation of small-scale anisotropies observed in the arrival directions of TeV-PeV CRs. Test particle simulations naturally simulate CR distributions for individual realisations of the turbulent fields and thus provide direct access to such stochasticity effects.

It needs to point that if the density in the relativistic particles is high enough, the particles can generate their own waves and induce a modification of the turbulence. This is the scenario when test particles models are no longer valid to describe the physics.

28 Boundary Conditions

Boundary condition is an important topic both in theory and simulation. This chapter aims at providing an overview of the boundary conditions we apply for electromagnetism and plasma simulations.

28.1 Conducting Boundary

The general boundary conditions on the EM field at an interface between medium 1 and medium 2 are

$$\begin{aligned}\hat{n} \cdot (\mathbf{D}_1 - \mathbf{D}_2) &= \tau \\ \hat{n} \times (\mathbf{E}_1 - \mathbf{E}_2) &= \mathbf{0} \\ \hat{n} \cdot (\mathbf{B}_1 - \mathbf{B}_2) &= 0 \\ \hat{n} \times (\mathbf{H}_1 - \mathbf{H}_2) &= \mathbf{K}\end{aligned}\tag{28.1}$$

where τ is used for the interfacial surface charge density (to avoid confusion with the conductivity), and \mathbf{K} is the surface current density. Here, \hat{n} is a unit vector normal to the interface, directed from medium 2 to medium 1. We have learned from EM that at normal incidence, the amplitude of an electromagnetic wave falls off very rapidly with distance inside the surface of a good conductor. In the limit of perfect conductivity (i.e., $\sigma \rightarrow \infty$), the wave does not penetrate into the conductor at all, in which case the internal tangential electric and magnetic fields vanish. This implies, from Equation 28.1, that the tangential component of \mathbf{E} vanishes just outside the surface of a good conductor, whereas the tangential component of \mathbf{H} may remain finite. Let us examine the behavior of the normal field components.

Let medium 1 be a conductor, of conductivity σ and dielectric constant ϵ_1 , for which $\sigma/\epsilon_1 \epsilon_0 \omega \gg 1$, and let medium 2 be a perfect insulator of dielectric constant ϵ_2 . The change density that forms at the interface between the two media is related to the currents flowing inside the conductor. In fact, the conservation of charge requires that

$$\hat{n} \cdot \mathbf{j} = \frac{\partial \tau}{\partial t} = -i \omega \tau$$

However, $\hat{n} \cdot \mathbf{j} = \hat{n} \cdot \sigma \mathbf{E}_1$, so it follows from Equation 28.1 that

$$\hat{n} \cdot (\epsilon_0 \epsilon_1 \mathbf{E}_1 - \epsilon_0 \epsilon_1 \mathbf{E}_2) = \frac{i\sigma}{\omega} \hat{n} \cdot \mathbf{E}_1$$

$$\left(1 + \frac{i\omega\epsilon_0\epsilon_1}{\sigma}\right) \hat{n} \cdot \mathbf{E}_1 = \frac{i\omega\epsilon_0\epsilon_2}{\sigma} \hat{n} \cdot \mathbf{E}_2$$

Thus, it is clear that the normal component of \mathbf{E} within the conductor also becomes vanishingly small as the conductivity approaches infinity.

If \mathbf{E} vanishes inside a perfect conductor then the curl of \mathbf{E} also vanishes, and the time rate of change of \mathbf{B} is correspondingly zero. This implies that there are no oscillatory fields whatever inside such a conductor, and that the fields just outside satisfy

$$\begin{aligned}\hat{n} \cdot \mathbf{D} &= -\tau \\ \hat{n} \times \mathbf{E} &= \mathbf{0} \\ \hat{n} \cdot \mathbf{B} &= 0 \\ \hat{n} \times \mathbf{H} &= -\mathbf{K}\end{aligned}$$

Here, \hat{n} is a unit normal at the surface of the conductor pointing into the conductor. Thus, the electric field is normal, and the magnetic field tangential, at the surface of a perfect conductor. For good conductors, these boundary conditions yield excellent representations of the geometrical configurations of the external fields, but they lead to the neglect of some important features of real fields, such as *losses in cavities* and *signal attenuation in waveguides*. [Here](#) is an estimation related to the skin depth.

28.2 Perfectly Matched Layer

In EM field solvers, often we need boundless free-space simulation to prohibit reflecting waves. Back in 1993, a technique called *perfectly matched layer* (PML) for the *absorption* of EM waves was proposed to handle this problem, so that we don't necessarily need to set boundaries sufficiently far enough from the scatterer when solving interaction problems. With the new medium the theoretical reflection factor of a plane wave striking a vacuum-layer interface is null at any frequency and at any incidence angle, contrary to the previously designed medium with which such a factor is null at normal incidence only. So, the layer surrounding the computational domain can theoretically absorb without reflection any kind of wave travelling towards boundaries, and it can be regarded as a perfectly matched layer. The new medium as the PML medium and the new technique of free-space simulation as the PML technique.

28.2.1 Derivation

The theoretical derivation starts in discussing the transverse electric wave propagation. One key concept that confused me was the magnetic conductivity denoted as σ^* . Maybe this is just a jargon: permeability is a magnetic analogy to conductivity in electric circuits. Reluctance in a magnetic circuit is inversely proportional to permeability just as electric resistance is inversely proportional to conductivity. The relationships between length and cross-sectional area are also the same. Consequently calling permeability “magnetic conductivity” is a fine way to reinforce the analogy and understand magnetic circuits using an electronic analogy.

I guess I need to review my EM courses to fully understand the derivations.

29 Particle-in-Cell

29.1 Phase Space Sampling

Let $\hat{f}(\mathbf{z})$ be a general probability density function with normalization:

$$\int_{\Omega} \hat{f}(\mathbf{z}) dV = 1$$

Now sample the phase space with N markers according to $\hat{f}(\mathbf{z})$ then the *marker distribution*, g is:

$$g(\mathbf{z}) = N\hat{f}(\mathbf{z})$$

and the phase space volume element occupied by a marker is

$$dV_i = \frac{1}{g(\mathbf{z}_i)}$$

where \mathbf{z}_i is the phase space coordinate of the marker and dV_i the volume element occupied by it.

Define the weight of a marker as the number of physical particles in the corresponding phase space volume occupied by the marker:

$$w_i = f(\mathbf{z}_i) dV_i = \frac{f(\mathbf{z}_i)}{g(\mathbf{z}_i)}$$

This marker distribution can represent the total f or the δf . (???)

We can now take moments of the distribution function in the phase space (such as density), and the Monte Carlo approximation to the integration is as follows:

$$M(A) = \int_{\Omega} A(\mathbf{z}) f(\mathbf{z}) dV = \sum_{i=1}^N A(\mathbf{z}_i) w_i + \frac{1}{\sqrt{\mathcal{N}}}$$

where N is the number of markers in the phase space region Ω . This Monte Carlo approximation makes the particle simulation feasible for a number of problems, even in 3D, at the cost of introducing an error on the order of $N^{-1/2}$.

29.2 Artificial Scalings

Common scales to characterize the plasma system:

- Global scales of the system d_g , characterized by the magnetopause standoff distance in the magnetosphere, some fraction of the solar radius in the corona, or system length for Tokamak.
- Ion kinetic scales characterized by the ion inertial length d_i

$$d_i = \sqrt{\frac{m_i}{n_i q_i^2 \mu_0}} = \frac{m_i}{q_i} \sqrt{\frac{1}{\rho_i \mu_0}}$$

and the ion gyroradius

$$r_i = \frac{m_i v_{\text{th},i}}{q_i B} = \frac{m_i}{q_i} \frac{\sqrt{p_i/\rho_i}}{B}$$

- Electron kinetic scales characterized by the electron skin depth d_e

$$d_e = \sqrt{\frac{m_e}{n_e q_e^2 \mu_0}} = \frac{m_e}{q_e} \sqrt{\frac{1}{\rho_e \mu_0}}$$

- The smallest plasma scale characterized by the Debye length λ_D

$$\lambda_D = \sqrt{\frac{\epsilon_0 v_{\text{th},e}^2}{q_e^2 n_e}} = \frac{m_e}{q_e} \frac{\epsilon_0 p_e}{\rho_e}$$

where $v_{\text{th},e} = \sqrt{p_e/\rho_e}$ is the electron thermal velocity.

Typically, d_i are orders of magnitude smaller than d_g , and d_e is even smaller. If we are interested in the interplay between the global plasma system and the kinetic processes, it is a natural question to ask how the behavior of the system depends on the ratio d_g / d_i . For example, in Ganymede's magnetosphere $d_g / d_i \sim 10$ and the kinetic effects have non-negligible contributions to the global system. On the other hand, if d_g / d_i is a very large number, then the kinetic effects will be mostly limited in a small region.¹

For quasi-neutral plasmas, $n_i = n_e \rightarrow \rho_i / \rho_e = m_i / m_e$, $q_i = q_e$, $p_i = p_e$. The ion scales are $\sqrt{m_i/m_e}$ times larger than the corresponding electron scales. For a proton-electron plasma, $\sqrt{d_i/d_e} \approx 43$. A standard trick to reduce the computational requirement is to artificially reduce the mass ratio to a smaller value ($1836 \geq d_i/d_e \geq 25$). In terms of reconnection, we

¹Reconnection can lead to wave generation, which is a process of energy and momentum conversion from the kinetic region to the global scale. Foreshock waves are also triggered by the kinetic processes and have a global impact. Therefore, this statement requires extra consideration.

have found only a relatively weak dependence of the process on the mass ratio. In practice almost all numerical studies, especially in 3-D, use a reduced ion-electron mass ratio.

Tóth et al. (2017) proposed another scaling methods to increase the ion and electron mass-to-charge ratios by a kinetic scaling factor f while keeping the MHD quantities, the mass densities ρ_e and ρ_i , the pressures p_e and p_i , the bulk velocities u_e and u_i , the magnetic field B , and the various constants ϵ_0 , μ_0 , and c unchanged. A nice property of this scaling method is that the characteristic speeds (bulk velocity, thermal velocity, Alfvén speed) are not modified; in other words, this *kinetic scaling* has no effect on ideal/resistive MHD. A direct implication of this scaling is the solution on the kinetic scales is similar for different values of f , but the spatial and temporal scales are proportional to f .

29.3 Numerical Heating and Thermalization

Both *numerical heating* and *numerical thermalization* are artifacts that can arise in PIC simulations, but they represent different phenomena. Note that “heating” is a general word: in fact, for implicit PIC schemes numerical cooling happens more often.

- Numerical Heating:
 - Mechanism: This refers to the spurious increase in the overall kinetic energy of the system due to discretization errors. It often arises from the interaction of particles with the electric and magnetic fields that are defined on a grid. The finite grid size and time step can lead to inaccuracies in the force calculation, causing particles to gain energy artificially.
 - Effect: Leads to an unphysical increase in the plasma temperature over time. This can affect the stability of the simulation and lead to inaccurate results.
 - Example: Imagine a particle accelerating in an electric field. In a real plasma, the particle would gain energy smoothly. However, in a PIC simulation, the particle might “jump” over grid points, leading to a slightly higher energy gain than expected.
- Numerical Thermalization:
 - Mechanism: This is a more subtle effect related to the artificial relaxation of the velocity distribution function (VDF) towards a Maxwellian distribution. It is caused by the “numerical collisions” between macroparticles due to their finite size and the grid. Even if the total energy is conserved, the way this energy is distributed among the particles can be artificially altered.
 - Effect: Drives the system towards an artificial thermal equilibrium, even if the real plasma would not reach such a state. This can mask kinetic effects and lead to incorrect predictions of transport properties.
 - Example: Consider a beam of electrons injected into a plasma. In reality, this beam might persist for a long time. However, in a PIC simulation, numerical collisions

can cause the beam to spread out and lose its distinct identity much faster than expected.

Key Differences:

- Energy Conservation: Numerical heating violates energy conservation, while numerical thermalization can occur even when energy is conserved.
- Timescale: Numerical heating often occurs on faster timescales compared to numerical thermalization.
- Distribution Function: Numerical heating primarily affects the average energy, while numerical thermalization modifies the shape of the distribution function.

Mitigation Strategies:

- Increasing the number of macroparticles: This reduces the graininess of the simulation and the impact of numerical collisions. Refining the grid: A smaller grid size improves the accuracy of field calculations and reduces discretization errors.
- Employing higher-order shape functions: These provide a more accurate representation of the charge density and reduce numerical heating.
- Using advanced PIC algorithms: Some algorithms are designed to conserve energy or momentum more accurately, reducing numerical artifacts.

29.4 Computational Cost

The key feature that makes PIC more feasible than the brute-force nbody method is the reduction of computational cost. Given N macro-particles in the system, a brute force method has a complexity of $\mathcal{O}(N^2)$, and a PIC method has $\mathcal{O}(N \log N)$. A rough idea how this is possible can be obtained by ignoring the pair interactions and consider collective effects especially for those particles at a large distance.

A commonly used trick is to introduce a reduced ion-to-electron mass ratio. In a 3D simulation, the cost scales with $(m_i/m_e)^3$. However, note that in low mass ratio runs (e.g. $m_i/m_e = 25$), the thermal speed of the electrons may be comparable to the Alfvén speed which causes higher wave decay through Landau damping on the electrons.

29.5 f Method

Consider the growth rate of a single unstable wave in a 1D-1V Vlasov-Poisson system of plasma. The energy conservation leads to

$$\frac{d}{dt} \left(\frac{\epsilon_0}{2} E(x)^2 \right) + \int qvE(x)f(x,v)dv = 0$$

For a single mode $E(x, t) = A(t) \sin(kx - \omega t)$, the growth rate is

$$\gamma = \frac{1}{A} \frac{dA}{dt} = - \int qvE f dv = - \sum_j q_j w_j E(x_j)$$

where w_j is the particle weight. All kinetic information lies in the marker particle position distribution.

The idea of $\delta - f$ method is to decompose the total distribution function into two parts,

$$f = f_0 + \delta f, \quad |\delta f| \ll f_0$$

and let the particle weight be proportional to the perturbation part,

$$w \propto \delta f, \quad w_j = \frac{\delta f(x_j, v_j)}{f(x_j, v_j)}$$

If the contribution from f_0 is known, for example

$$\int qvE f_0 dv = \int qv f_0(v) A \sin(kx - \omega t) dv = 0$$

then the growth rate can be written as

$$\gamma = - \int qvE \delta f dv = - \int qvE \frac{\delta f}{f} f dv = \sum_j q_j v_j w_j E(x_j)$$

Since now the particle weight is proportional to $\delta f/f$, the system noise would be significantly reduced by $(\delta f/f)^2$; in other words, for the same noise level the number of macro-particles required would be reduced by $(\delta f/f)^2$, which is typically $\mathcal{O}(10^6)$. Of course this requires the prior assumption $|\delta f| \ll f$ to be valid.

30 Hybrid Methods

The hybrid model is valid for low-frequency physics with $\omega \sim \Omega_i$ and $kr_{Li} \sim 1$ (wavelength $\lambda \sim 6r_{Li}$), where ω is the wave frequency, k is the wave number, Ω_i is the ion gyrofrequency, and r_{Li} is the ion Larmor radius. For this range of wave frequency and wavelength, the ion kinetic physics in the near-Earth instabilities are resolved with grid sizes $\sim r_{Li}$ or ion inertial length d_i . The finite ion gyroradius effects are resolved with particle time steps Δt much smaller than the gyroperiod.

A typical time step in a global hybrid magnetosphere model is $0.05\Omega_0^{-1}$, where Ω_0 is the upstream (solar wind) ion gyrofrequency. For IMF $B_0 \sim 10\text{nT}$, $\Omega_0 \sim 1\text{rad/s}$, which corresponds to $f_0 \sim 1/2\pi\text{s}^{-1}$. Based on in-situ observations, the typical ion inertial length in the tail is $\sim 0.2R_E$, but smaller near the dayside magnetopause. Therefore usually discrete grid size $\Delta x = 0.1d_i$ is barely enough to resolve tail ion kinetic dynamics but not dayside kinetic structures.

30.1 Classical Hybrid Model

Define the electric charge density ρ and current density \mathbf{J} as

$$\rho = \sum_s q_s n_s - e n_e$$

$$\mathbf{J} = \sum_s q_s n_s \mathbf{u}_s - e n_e \mathbf{u}_e$$

where q_s, n_s, \mathbf{U}_s are the charge, number density and bulk velocity of ion species s calculated by taking moments of the distribution function

$$n_s = \int d^3v f_s(\mathbf{r}_s, \mathbf{v}_s, t)$$
$$\mathbf{u}_s = \frac{1}{n_s} \int d^3v \mathbf{v}_s f_s(\mathbf{r}_s, \mathbf{v}_s, t)$$

or in the corresponding discrete forms where the distribution function is represented as a group of macro-particles with a specific shape function. In this way it behaves more like a particle cloud.

The crucial assumption in the hybrid model is the *quasi-neutrality*, that is, the electrons move fast enough to cancel any charge-density fluctuations and $\rho = 0$ is always satisfied. By assuming quasi-neutrality, we can

- avoid solving the conservation equations for electrons
- avoid solving the Maxwell's equations entirely and instead use the generalized Ohm's law instead

The electron density thus can be written by using ion densities $n_e \approx n_i \equiv \sum_s q_s n_s / e$. In addition, the electron bulk velocity may also be eliminated using Ampère's law

$$\mathbf{J} = \mu_0^{-1} \nabla \times \mathbf{B} \quad (30.1)$$

and the relation

$$\mathbf{u}_e = \mathbf{u}_i - \mathbf{J}/n_e e \quad (30.2)$$

The basic equations used in the conventional PIC hybrid model first has a particle pusher for individual ions

$$\frac{d\mathbf{x}_j}{dt} = \mathbf{v}_j$$

$$\frac{d\mathbf{v}_j}{dt} = \frac{q_j}{m_j} (\mathbf{E} + \mathbf{v}_j \times \mathbf{B})$$

where the subscript j and e indicate the indices for individual ions and the electron fluid and other notations are standard. The lowercase velocities are velocities for each macro-particle.

Alternatively, if we rely on a Vlasov system, we directly solve for the distribution function $f(\mathbf{r}_s, \mathbf{v}_s, t)$ from the Vlasov equation

$$\frac{\partial f_s}{\partial t} + \mathbf{v}_s \cdot \frac{\partial f_s}{\partial \mathbf{r}_s} + \mathbf{a}_s \cdot \frac{\partial f_s}{\partial \mathbf{v}_s} = 0$$

where

$$\mathbf{a}_s = \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v}_s \times \mathbf{B})$$

The generalized Ohm's law is used to determine the time evolution of the electric field, derived from the electron momentum equation assuming $m_e \rightarrow 0$,

$$\mathbf{E} = -\mathbf{u}_i \times \mathbf{B} + \frac{1}{\mu_0 n_e e} (\nabla \times \mathbf{B}) \times \mathbf{B} - \frac{1}{n_e e} \nabla \cdot \vec{P}_e + \mathbf{R}_{ei} \quad (30.3)$$

where the current in the Hall term has already been replaced by the curvature of \mathbf{B} . Note that the density has been written with electron number density, instead of ion density, to avoid the confusion for multiple ion species under the quasi-neutrality condition. The last term \mathbf{R}_{ei} can

either represent collision/physical resistivity, or artificial resistivity/numerical diffusion. For electromagnetic problems, it can include resistive and hyper-resistive terms of the form

$$\mathbf{R}_{ei} = \eta \mathbf{J} + \eta_H \nabla^2 \mathbf{J}$$

where the resistivity η may be thought of as accounting in a simple way for electron-ion collisions, and the hyper-resistivity η_H may represent an electron viscosity and helps with numerical stability by damping grid-scale oscillations. In addition, \mathbf{R}_{ei} can also include contributions from explicit friction or other momentum exchange terms in the particle collision models.

The magnetic fields evolve according to Faraday's law

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$$

The Hall term supports Whistler waves, which usually place the strongest limit on the time step in hybrid PIC codes. The Courant–Friedrichs–Lewy (CFL) condition for Whistler waves on the time step is typically $\Omega_{ci}\Delta t < (\Delta x/d_i)^2/\pi$. Updating the magnetic field including the Hall term turns out to be a stiff problem, thus requires smaller time step sub-cycling in many practical cases. This is most useful when the grid resolution is very fine with $\Delta x \ll d_i$.

Finally, by determining the electron pressure tensor by using an appropriate equation of state, the evolution of the system can be followed in time. For example, let $\tilde{P}_e = P_e \tilde{I}$ where P_e is the isotropic scalar electron pressure. In the simplest form

$$P_e = n_e k_B T_e$$

where $n_e \approx n_i$ and $T_e = T_i$. Note however in a plasma electron pressure is usually higher than ion temperature, so this is a very crude assumption. Another commonly used assumption is an adiabatic process

$$P_e = n_e^\gamma k_B T_e = n_0(n/n_0)^\gamma k_B T_{e0}$$

where $\gamma = 5/3$ is the adiabatic index for a monatomic ideal gas.

For more complicated systems with gradients in the initial conditions, a separate electron energy evolution equation is required. This takes the form

$$\frac{\partial P_e}{\partial t} = -\gamma \nabla \cdot (P_e \mathbf{u}_e) + (\gamma - 1) \mathbf{u}_e \cdot \nabla P_e + (\gamma - 1)(-\nabla \cdot \mathbf{Q}_e + H_{ei}) \quad (30.4)$$

where the electron bulk velocity is inferred from Equation 30.2, with the current from Equation 30.1. If used in place of a simple equation of state, Equation 30.4 is integrated in time within the same numerical loop as the magnetic field evolution. For many problems, the electron heat flux may be modeled with a heat conductivity κ as

$$\mathbf{Q}_e = -\kappa \nabla T_e$$

where $\kappa = \kappa_e + \kappa_0$, κ_e is given by a physical model and may depend on the local plasma conditions, and κ_0 is a small ($\kappa_0 \sim 0.01 - 0.1\kappa_e$ for models with a physical heat conductivity, or $\kappa_0 \sim 0.01n_0d_iV_A$ for magnetized simulations) constant numerical diffusion coefficient set separately for convenience that helps maintain numerical stability. Models for electron and ion energy exchange may be included through the term H_{ei} , which can be captured by adding energy lost in each cell by ions in collisions models back to the local electron fluid cell. (???)

It is of course possible to incorporate the full electron pressure tensor effects to handle specially reconnection physics better.

A more complete review is given by (Dan Winske et al. 2023) for the hybrid-kinetic model assuming massless electrons. The essential problem in all hybrid algorithms is how to calculate the electric field at the next time step.

30.1.1 Pros and Cons

Strengths:

- No approximations to ion physics.
- Valid for $\omega/\Omega_i \sim kr_i \sim 1$.
- No issues for high- β regimes.
- Simple implementation of particle push that can be readily optimized.
- Removes stiffest electron scales. Neglecting the electron kinetic scales typically reduces the computing cost and memory requirements by a factor of $\sqrt{m_i/m_e}$ (compared to what is used in full-PIC) for each spatial dimension and the time step may usually be increased by a similar factor.

Limitations:

- Need to resolve ion gyrofrequency.
- Stiff EMHD whistler waves $\Delta t_{CFL} \sim \Delta x^2$.
- No electron Landau damping.
- Explicit time-stepping schemes can be complex.
- No existing method conserves momentum or energy.

30.1.2 Role of Electrons

The interpretation of “massless” electrons is tricky. If we think of the full particle model, and take the asymptotic behavior when $m_e \rightarrow 0$, this does not explicitly lead to the hybrid equations. The reason is that

1. The hybrid model completely eliminates the electron inertia term from the equations of motion. *This is not just an approximation of small mass; it fundamentally changes the way electrons respond to fields.*

2. Hybrid models often (though not always) assume quasi-neutrality, where the electron and ion densities are nearly equal. This is a separate assumption from massless electrons.
3. The way the electric and magnetic fields are calculated differs between the two models. The hybrid model typically solves for the electric field directly from a combination of the generalized Ohm's law and Faraday's law, whereas the full particle model would solve Maxwell's equations fully. Again, combined with point 1, the way electrons respond to fields is completely different in the two models.

30.2 Low Density Treatment

Because the hybrid model includes terms proportional to $1/n_e$, a modification is necessary for low-density and vacuum regions. The simplest method is to apply a density lower limit n_f in the field solver, so that the density used to advance the fields is $n_e = \max(n_e, n_f)$. Typical values used for the density floor are $n_f/n_0 \sim 0.01 - 0.05$, where n_0 is a reference background density. The reference density n_0 is typically the value used to define the ion skin depth $d_i = V_A/\omega_{ci}$.

(Amano, Higashimori, and Shirakawa 2014) suggests another way to solve for the electric field

$$(\omega_{pe}^2 - c^2 \nabla^2) \mathbf{E} = \frac{e}{m_e} (\mathbf{J}_e \times \mathbf{B} - \nabla \cdot \vec{P}_e) + (\mathbf{V}_e \cdot \nabla) \mathbf{J}_e + \eta \mathbf{J}$$

which can be reduced to the Laplace equation in near-vacuum region, presenting no numerical difficulty.

Besides, the electron velocity is redefined

$$\mathbf{V}_e = \frac{\mathbf{J}_e}{\max(\rho_e, \rho_{e,min})}$$

where the minimum density $\rho_{e,min}$ is an artificially set value.

In a hybrid system, the maximum phase velocity is the electron Alfvén speed, which goes to infinity when $m_e \approx 0$. However, when doing calculations we only have ion Alfvén speed

$$v_{p,max} \simeq \frac{1}{2} \frac{B}{\sqrt{\mu_0 n_e m_e}} = \frac{1}{2} V_{A,i} \sqrt{\frac{m_i}{m_e}}$$

To keep the maximum phase velocity always below the CFL condition, one may use a modified electron mass ratio m'_e defined as

$$\frac{m'_e}{m_i} = \max\left(\frac{m_e}{m_i}, V_A^2 \left(\frac{\Delta t}{2\alpha \Delta x}\right)^2\right)$$

instead of the physical electron mass m_e . Here V_A is the Alfvén speed calculated from the local density and magnetic field, and α is the maximum allowed Courant number (≤ 0.5).

30.3 Finite Electron Inertia

The conventional hybrid simulation model dealing with kinetic ions and a massless charge-neutralizing electron fluid is known to be susceptible to numerical instability due to divergence of the whistler-mode wave dispersion, as well as division-by-density operation in regions of low density. The Alfvén wave at short wavelength comparable to ion inertia length has dispersion due to the decoupling between ion and electron dynamics. There thus appears the whistler mode whose frequency diverges as $\omega \propto k^2$. This means that the maximum phase velocity in the system increases rapidly without bound, implying numerical difficulty. The division-by-density issue originates from Equation 30.2 and appears in the Hall and electron pressure gradient terms. Consequently, a pure vacuum region is not allowed to exist in the simulation domain unless some ad hoc technique is used.

On the other hand, this can be restated as hybrid simulations with massless charge-neutralizing electrons are unrealistic if whistlers are involved. Either people use full particle models, or try to incorporate finite electron mass into the conventional hybrid model.

The finite electron inertia correction is proposed to solve the whistler-mode wave dispersion issue. The conventional way to include a finite electron inertia correction into the hybrid model is to introduce the following so-called generalized electromagnetic field $\widehat{\mathbf{E}}, \widehat{\mathbf{B}}$, defined as

$$\widehat{\mathbf{E}} = \mathbf{E} - \frac{\partial}{\partial t} \left(\frac{c}{\omega_{pe}^2} \nabla \times \mathbf{B} \right)$$

$$\widehat{\mathbf{B}} = \mathbf{B} + \nabla \times \left(\frac{c^2}{\omega_{pe}^2} \nabla \times \mathbf{B} \right)$$

in which the terms proportional to $\nabla \times \mathbf{B}$ represent electron inertia correction.

From the equation of motion for the electron fluid, it may be shown that

$$\widehat{\mathbf{E}} = -\mathbf{V}_i \times \mathbf{B} + \frac{1}{n_i e} (\nabla \times \mathbf{B}) \times \mathbf{B} - \frac{1}{n_i e} \nabla \cdot \vec{P}_e - \frac{m_e}{e} (\mathbf{V}_e \cdot \nabla) \mathbf{V}_e$$

which is similar to the generalized Ohm's law but now with the last term which also represents the correction. \mathbf{V}_e is obtain from Equation 30.2.

Given the generalized electric field $\widehat{\mathbf{E}}$, one can advance the generalized magnetic field $\widehat{\mathbf{B}}$ by using Faraday's law, which can be easily checked to satisfy

$$\frac{\partial \widehat{\mathbf{B}}}{\partial t} = -\nabla \times \widehat{\mathbf{E}}$$

Further simplifications are commonly adopted; for example, the electric field correction term and electron-scale spatial variation of density are often ignored. In this case, the magnetic field

may be recovered by solving the equation

$$\widehat{\mathbf{B}} = \left(1 - \frac{c^2}{\omega_{pe}^2} \nabla^2\right) \mathbf{B}$$

and $\widehat{\mathbf{E}} = \mathbf{E}$ is assumed. The nice feature with this approach is that the correction can be implemented as a post process to the each integration step of a standard procedure.

30.4 Comparison with Hall MHD

The zeroth and first moments of the ion Vlasov equation are

$$\begin{aligned}\frac{\partial n}{\partial t} + \nabla \cdot (n \mathbf{u}_i) &= 0 \\ \frac{\partial mn \mathbf{u}_i}{\partial t} + \nabla \cdot \left[mn \mathbf{u}_i \mathbf{u}_i - \frac{\mathbf{B} \mathbf{B}}{\mu_0} + \frac{B^2}{2\mu_0} \mathbf{I} + \vec{P} \right] &= 0\end{aligned}$$

The difference between Hall MHD and hybrid model is the treatment of the pressure tensor term. For Hall MHD with constant T_{i0}/T_{e0} ,

$$\vec{P} = p_e (1 + T_{i0}/T_{e0}) \mathbf{I}$$

For hybrid models,

$$\vec{P} = p_e \mathbf{I} + \int m_i f_i \mathbf{w} \mathbf{w} d\mathbf{w}$$

Thus Hall-MHD is a “cold-ion” model in the sense that it does not include ion finite Larmor radius (FLR) or other kinetic effects from warm distribution functions.

30.5 Normalization

There are five basic quantities in the hybrid model (length, mass, time, current density, and number density???) and three physical constants (μ_0, q, m). If we add temperature, then correspondingly k_B would appear. Usually even though the particle mass is a constant, we treat it as a parameter to represent a proton system or electron system or other particle system. Thus we need three (5 – 2) independent reference quantities for the normalized units in a hybrid model. For instance, We can take a magnetic field scale B_0 , number density scale n_0 , and mass scale m_0 . The general variable transformation from the original units to normalized units is

$$\chi = \chi_0 \tilde{\chi}$$

where $\tilde{\chi}$ denote the variable in the normalized units and the scale χ_0 shall be in the original units (e.g. SI).

Typically we use the inverse of gyrofrequency for the time scale

$$t_0 = \Omega_{c0}^{-1} = \frac{m_0}{eB_0}$$

and Alfvén speed for the velocity scale

$$v_0 = v_A = \frac{B_0}{\sqrt{\mu_0 n_0 m_0}}$$

Then the length scale is taken to be the ion skin-depth, or inertial length

$$L_0 = d_i = \frac{c}{\Omega_{c0}} = \sqrt{\frac{m_0}{\mu_0 e^2 n_0}}$$

A common trick we can use to speed up the simulation is to artificially increase the ion mass such that the length scale is increased $\propto \sqrt{m_0}$. For example, in many global hybrid Earth magnetosphere models, d_i in the upstream solar wind is artificially increased to $0.1 R_E$ (by increasing the ion mass), which is 6.8 times the realistic $d_i = 0.015 R_E$ for $n_{sw} = 6 \text{ amu/cc}$. If our grid resolution is $0.05 R_E \approx 300 \text{ km}$ (which is common as of 2020s), we will have about 10 points per ion-scale wave, which is enough to resolve the ion-scale kinetics.

However, a consequence of this scaling due to computational limitation must be emphasized. The reference Alfvén speed

$$V_{A0} = d_{i0} \Omega_{i0}$$

will also be larger than reality since we artificially increase d_{i0} but not Ω_{i0} .

Note the difference between gyrofrequency and frequency, which differs by a factor of 2π :

$$\omega = 2\pi f$$

I once made a mistake in dealing with a code that uses SI units. You may wonder how come the ion inertial length is defined by speed of light divided by the plasma ion frequency in the unit of [rad/s], and time scale in the unit of Ω_i^{-1} which is [s/rad]. In practice, we do not include 2π in neither of them!

The pressure scale can be equivalently derived from the magnetic pressure or dynamic pressure

$$P_0 = \frac{B_0^2}{\mu_0} = \rho_0 v_0^2$$

Table 30.1: hybrid model unit conversion.

Table 30.2: Example basic quantities in the hybrid model

Basic variable	Notation	Definition	Value
Magnetic field	B_0	$B_0 = B_{\text{ref}}$	$1 \times 10^{-8} \text{ T}$
Number density	n_0	$n_0 = n_{\text{ref}}$	$1 \times 10^6 \text{ m}^{-3}$
Mass	m_0	$m_0 = m_i$	$1.67 \times 10^{-27} \text{ kg}$

Table 30.3: Example derived quantities in the hybrid model

Derived variable	Notation	Definition	Value
Length	l_0	$l_0 = \sqrt{m_0/(\mu_0 e^2 n_0)}$	$2.28 \times 10^5 \text{ m}$
Velocity	v_0	$v_0 = B_0 / \sqrt{\mu_0 m_i n_0}$	$2.18 \times 10^5 \text{ m} \cdot \text{s}^{-1}$
Time	t_0	$t_0 = m_i / (e * B_0)$	1.04 s
Mass density	ρ_0	$\rho_0 = n_0 / m_i$	$1.67 \times 10^{-21} \text{ m}^{-3}$
Pressure	p_0	$p_0 = \rho_0 v_0^2$	$7.96 \times 10^{-11} \text{ N} \cdot \text{m}^{-2}$
Temperature	T_0	$T_0 = p_0 * m_i / (k_B \rho_0)$	$5.76 \times 10^6 \text{ K}$

Note the drop of the factor of 2 here: it is then kept in the dimensionless equations, e.g. $p'_B = B'^2/2$.

The temperature scale is then

$$T_0 = \frac{p_0}{n_0 k_B} = \frac{B_0^2}{2 \mu_0 k_B n_0}$$

Note the factor of 2 appeared in the pressure and temperature scales: this is to make the derivations consistent. The unit conversions are summarized in Table 30.1.

30.6 Numerical Stability

Nonlinear numerical simulations typically need some dissipation for stability. This is achieved either via:

1. Explicit terms in equations (“physical dissipation”).
2. Upwinding of advective terms (implicit dissipation via discretization)

Hybrid models usually follow 1) by adding dissipation in Ohm’s law:

$$\mathbf{E} = \underbrace{\mathbf{E}^*}_{\text{frictionless } \mathbf{E}} + \underbrace{\eta \mathbf{j}}_{\text{resistivity}} - \underbrace{\eta_H \nabla^2 \mathbf{j}}_{\text{hyper-resistivity}}$$

The reason that we need hyper-resistivity is because the Hall term is badly behaved (stiff?). Slide 33 Stainer??? The hyper-resistivity term has a similar form of an electron collisional viscosity $\eta_H \nabla^2 \mathbf{u}_e$. But the coefficient is too large for space, sometimes argued as “anomalous viscosity”.

Extra care shall be taken for conservation when including the frictional terms. For momentum conservation, we should use \mathbf{E}^* for the macro-particle pusher or Vlasov solver, and only use \mathbf{E} for updating \mathbf{B} . For energy conservation, it requires a separate electron pressure equation with frictional heating $H_e = \eta j^2 + \eta_H \nabla \mathbf{j} : \nabla \mathbf{j}$ and heat flux \mathbf{q}_e :

$$(\gamma - 1)^{-1} \left[\frac{\partial p_e}{\partial t} + \nabla \cdot (\mathbf{u}_e p_e) \right] + p_e \nabla \cdot \mathbf{u}_e = H_e - \nabla \cdot \mathbf{q}_e$$

30.6.1 Finite Grid Instability

Imagine a scenario where cold ion beams move through uniform spatial mesh. It was shown by (Rambo 1995) that non-conservative (explicit) schemes are unstable for $T_i / T_e \ll 1$ regardless of spatial resolution. The precise threshold in T_i / T_e and beam velocity depends on shape-function for macroparticles. This instability causes unstable (exponential) heating of ions until some saturation value and also violates momentum conservation. Implicit momentum and energy conserving schemes are stable w.r.t. these instabilities.

30.7 Boundary Conditions

30.7.1 Open Boundary Conditions

Open boundary conditions allow plasma and magnetic flux to flow into or out of the simulation domain. Ion particles are absorbed at the open boundaries, and new particles are injected if necessary. The injected particle flux is sampled from a drifting multi-Maxwellian velocity distribution that matches specified densities, velocities, and pressure moments at the boundary.

The magnetic field is formally split into two components $\mathbf{B} = \mathbf{B}_0 + \mathbf{B}_1$, with a fixed external field \mathbf{B}_0 and a time-varying component \mathbf{B}_1 . The external field is a vacuum field ($\nabla \times \mathbf{B}_0 = 0$) generated by a system of currents outside the plasma. Examples are the interplanetary magnetic

field and planetary dipole field in a global magnetosphere simulation, or the confinement fields generated by external coils in a magnetic mirror device. The particles are advanced in the total magnetic field \mathbf{B} . In the field solver, only \mathbf{B}_1 is advanced in time, and \mathbf{B}_0 may be dropped when computing the plasma current density in the Hall term from $\mu_0 \mathbf{J} = \nabla \times \mathbf{B} = \nabla \times \mathbf{B}_1$. The advantage of splitting \mathbf{B} in this manner is that for the open boundary conditions, \mathbf{B}_0 is left fixed in the boundary ghost cells. Otherwise, the external field may change over time by diffusing at the boundaries. For the open boundary condition on the time-varying component of \mathbf{B} , the field \mathbf{B}_1 in each ghost cell is set equal to the value of its neighbor within the simulation domain.

The electric field is handled differently. The electric field in ghost cells along an open boundary is advanced in time along with the electric field within the bulk cells. But when computing the electric field from the Ohm's law Equation 30.3 in the a ghost cell, it is assumed that there are no gradients in either \mathbf{B} or the plasma moments (n and \mathbf{u}_i) normal to the boundary. The plasma moments, like the magnetic field, within the ghost cell are set equal to their neighboring values. So, for example, an open boundary in the x direction may have pressure gradient fields E_y and E_z from y and z gradients in P_e , but there is no pressure gradient-driven component E_x . Crucially, for an open x boundary, gradients in y and z (but not x) are retained in the Hall term $\propto (\nabla \times \mathbf{B}) \times \mathbf{B}$. This treatment of the electric field at open boundaries significantly enhances numerical stability. Nevertheless, it is often necessary to also include a buffer region several cells wide with an enhanced hyper-resistivity η_H to dissipate residual oscillations.

30.7.2 Magnetosphere Inner Boundary Conditions

Inner boundary condition is often the most tricky part in magnetopshere modeling. In Angeo3D [Lin+ 2014], the inner magnetosphere ($r < 6 R_E$) is assumed to be dominated by a cold, incompressible ion fluid, which coexists with particle ions. The number density of the cold ion fluid is assumed to be

$$n_f = \frac{n_{\text{eq}}}{r^3} [1 - \tanh(r - 6.5)]$$

where r is in the unit of R_E , and $n_{\text{eq}} = 1000 \text{ cm}^{-3}$.

The inclusion of the cold ion fluid in the inner magnetosphere simplifies the conditions for the fluid-dominant low-altitude, inner boundary. Ion particles are set to be reflected at the magnetospheric inner boundary (e.g. $r = 3.5 R_E$). This simple reflection of the ion parallel velocity $\mathbf{v}_{i\parallel}$ means that loss cone effects are omitted. For a particle distribution with an isotropic pitch angle distribution in a dipole field, the particles in a full loss cone are only 0.3% of the total, which is reasonably neglected. \mathbf{B} is assumed to maintain the dipole field values at the inner boundary. The ionospheric conditions (1000 km altitude) are incorporated into the hybrid code, as in global MHD models (Raeder, Walker, and Ashour-Abdalla 1995). See Section 23.2.

30.8 Tests

30.8.1 Proton Cyclotron Anisotropy Instability

This is an electromagnetic and multi-ion verification test. We have an 1D-3V electromagnetic instability driven by $p_{i\perp}/p_{i\parallel} > 1$. Maximum growth happens at $\mathbf{k} \times \mathbf{B} = 0$ with a finite real frequency. The instability threshold is

$$\frac{P_\perp}{P_\parallel} - 1 \approx \frac{S}{\beta_\parallel^{0.4}}$$

with $S \sim 1$ (Gary 1993).

Both \mathbf{k} and \mathbf{B}_0 are parallel to the x-axis. For the initial perturbation, we choose $k_x \Delta x = 0.065$. We set an 1D simulation with 64 cells, $\Delta t \Omega_{ci} = 0.01$, dissipation=0. The nominal simulation parameters are:

$$\beta_\parallel = 1, \frac{T_\perp}{T_\parallel} = 3, \frac{L_x}{d_i} = 10.5, \frac{T_e}{T_\parallel} = 1, \gamma = \frac{5}{3}$$

PCAI results:

- Transverse velocity and magnetic field components grow from noise (left-hand Alfvén waves).
- Agree with linear theory for these parameters ($\gamma/\Omega_{ci} = 0.162$).
- Pressure anisotropy decreases via wave-particle interaction until saturation.

When we add a 20% density fraction of a minor species of alpha particles (He^{2+}), the growth rate is found to be smaller.

$$Z_\alpha = 2, M_\alpha = 4, \frac{T_{\alpha\parallel}}{T_{p\parallel}} = 2, \frac{T_{\alpha\perp}}{T_{\alpha\parallel}} = \frac{T_{p\perp}}{T_{p\parallel}} = 3, \frac{N_\alpha}{N_p} = 0.2, Z_\alpha N_\alpha + Z_p N_p = 1$$

The growth rates across a range of β and anisotropy can be computed and compared with a linear dispersion solver, e.g. [HYDROS](#).

30.8.2 Landau Damped Ion Acoustic Wave

The fundamental electrostatic mode in the hybrid-PIC model is the ion acoustic wave. This is also driven by pressure perturbations. In fluid models (e.g. Hall MHD), this wave is undamped. However, in the hybrid-PIC, Landau resonance damps the wave and locally flattens ion VDF, which is analogous to electron Landau damping of Langmuir oscillations.

The dispersion relation is

$$\frac{dZ(\zeta)}{d\zeta} = 2 \frac{T_i}{T_e}, \quad \zeta \equiv \frac{\omega - i\gamma}{kv_{\text{th},i}}$$

The nominal simulation parameters are:

$$T_i = 1/3, \gamma = 5/3, c_s = \sqrt{\gamma T_e/m_i} = 1, k_x = \pi/8, \delta n = 2 \times 10^{-2}$$

Results for nominal parameters

- Damping rate: $\gamma = -0.0932$.
- Initial perturbation damps to noise floor. Noise can be reduced by:
 1. Use more particles/cell (noise $\sim 1/\sqrt{N_p}$).
 2. Binomial smoothing/higher order shape functions.
 3. Using low-discrepancy quasi-Random numbers to seed particles (noise $\sim 1/N_p$).
 4. Most efficient: Delta-F (See Section 29.5).

30.8.3 Magnetic Reconnection Island Coalescence

Magnetic islands are 2D versions of flux-ropes. Here we set a self-driven reconnection problem of coupling of ideal island motion to micro-scale reconnection physics. Ion kinetic effects are crucial in reconnection studies.

Unstable Fadeev island equilibrium: the magnetic field \mathbf{B} is given by $\nabla \times \mathbf{A}$, where in this setup we only need

$$A_y = -\lambda B_0 \ln[\cosh(z/\lambda) + \epsilon \cos(z/\lambda)]$$

and the density is given by

$$n = n_0(1 - \epsilon^2)/[\cosh(z/\lambda) + \epsilon \cos(z/\lambda)]^2 + n_b$$

The pressure balance gives

$$\beta = \frac{2\mu_0 n_0 k_B (T_{i0} + T_{e0})}{B_0^2} = 1$$

The nominal simulation parameters are:

$$\lambda = 5d_i, \epsilon = 0.4, n_b = 0.2n_0, T_i/T_e = 1, \eta = 10^{-3}, \eta_H = 5 \times 10^{-3}, \gamma = 1$$

and for the numerical parameters, we have a 2D space with 256×128 cells, 50 particles/cell, $\Delta t \Omega_{ci} = 0.005$.

30.8.4 Collisionless Shock

This is a 2D magnetospheric shock problem, with a $M_A = 11.4$ shock injected from the right (open) boundary and a reflecting left boundary to drive the collisionless shock.

31 Ray Tracing

Wave heating is one of the most important approach to heating the plasma to high temperature. In magnetic confined plasmas, the usually used waves from high frequency (100 GHz) to low frequency (<1 MHz) include electron cyclotron wave (ECW, or whistler wave), lower hybrid wave (LHW), ion cyclotron wave (ICW) and Alfvén wave (AW). There are also terminologies such as fast wave (FW), slow wave (SW), helicon wave, etc. A simple but still accurate way to study the wave propagation and heating is using the *geometrical optics approximation*, which yields the *ray tracing equations*.

31.1 Backgrounds

In optics, the refractive index is a measure of how much a material bends or refracts light as it passes through it. It is defined as the ratio of the speed of light in a vacuum to the speed of light in the material:

$$n = \frac{c}{v}$$

where c is the speed of light in a vacuum, and v is the speed of light in the material.

In a plasma, the refractive index depends on the plasma parameters, such as the density, temperature, and magnetic field. The plasma refractive index is given by:

$$n = 1 - \frac{\omega_p^2}{\omega^2 - \omega_{ce}^2 - i\omega_{en}\omega} \quad (31.1)$$

where ω_p is the plasma frequency (Equation 10.4), $\omega_{ce} = eB/m_e$ is the electron cyclotron frequency, ω_{en} is the electron-neutral collision frequency, ω is the frequency of the incident wave, and i is the imaginary unit. The plasma refractive index can be complex, meaning that it has both real and imaginary parts:

- The real part of the refractive index determines the speed of light in the plasma, while
- The imaginary part determines the attenuation or absorption of the wave as it passes through the plasma.

The refractive index plays an important role in determining the behavior of electromagnetic waves in plasmas, and is therefore a key parameter in plasma ray tracing simulations.

31.2 How it Works

In plasma ray tracing, the refractive index of the plasma is calculated based on the plasma parameters, such as the density, temperature, and magnetic field, and the ray is then traced through the plasma using the calculated refractive index in the Snell's law:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \quad (31.2)$$

where n_1 and n_2 are the refractive indices of the media on either side of the interface, ' θ_1 ' is the angle of incidence, and ' θ_2 ' is the angle of refraction. This allows us to study how waves interact with plasma and how they can be used to probe and manipulate the properties of the plasma.

The ray tracing equations in Cartesian coordinates are

$$\begin{aligned} \frac{d\mathbf{r}}{dt} &= \frac{\partial \omega}{\partial \mathbf{k}} = -\frac{\partial D/\partial \mathbf{k}}{\partial D/\partial \omega} = \mathbf{v}_g \\ \frac{d\mathbf{k}}{dt} &= -\frac{\partial \omega}{\partial \mathbf{r}} = \frac{\partial D/\partial \mathbf{r}}{\partial D/\partial \omega} \end{aligned}$$

with the dispersion relation

$$D(\omega, \mathbf{k}, \mathbf{r}) = 0$$

where $\mathbf{r} = (x, y, z)$ is the spatial location, $\mathbf{k} = (k_x, k_y, k_z)$ is the wave vector, ω is the wave frequency, and \mathbf{v}_g is the group velocity. The geometrical optics approximation is valid in cases where *the wave length is much smaller than the system nonuniform length*, which is usually well satisfied for high frequency waves such as ECW and LHW, but should be used with caution for low frequency waves such as ICW and AW.

Written in a more explicit way, The three first-order differential equations that relate the position of the ray to its direction and the properties of the medium are:

$$\begin{aligned} \frac{dx}{dt} &= v_g \cos(\theta) \\ \frac{dy}{dt} &= v_g \sin(\theta) \\ \frac{dz}{dt} &= \frac{v_g}{k} \frac{dk}{dz} \end{aligned}$$

where θ is the angle between the ray and the z-axis.¹

For more thorough introduction, check out (Tracy et al. 2014).

¹These are given by Chat-GPT 3.5, which I don't understand. There must be some assumption for this choice of the coordinate system!

32 Tests

Verifying the validity of models is critical in performing consistent research. This chapter goes through the numerical tests we can do in plasma physics, and is heavily influenced by the [tests in the Athena MHD code](#). Detail can be found in (Stone et al. 2008).

Most test parameters are listed in dimensionless units. If SI units are required, we need unit conversions. See Section [6.2](#) for details.

32.1 Errors

1. Diffusion error: the rate at which the amplitude of the wave decreases.
2. Dispersion error: the difference between the speed at which the wave propagates in the numerical versus the analytic solution.

The error norm usually contains contributions from both. However, one could plot the error as a function of position or phase to track each individually.

32.2 Wave Tests

32.2.1 Linear wave

Linearization of Equation [8.49](#) ends up in a system of equations $\mathbf{W}_t = \mathbf{A}(\mathbf{W})\mathbf{W}_x$, where \mathbf{W} is composed of either the primitive variables or conserved variables in 1D, and the subscripts t and x are corresponding partial derivatives. The eigenvalues and eigenvectors of ideal MHD can be used to set the characteristic wave perturbations. See the details in Appendix A and B from (Stone et al. 2008).

The density, velocity, magnetic field, and total energy are all set to constant values initially. These values can be chosen so that the wave speeds are all well separated, and so that (in MHD) the wavevector is at an arbitrary angle to \mathbf{B} . The precise values chosen for the tests described here are given in the appropriate results section below.

The wave is added to as a perturbation to these constant values of the form $\delta\mathbf{U} = A\mathbf{R}\sin(2\pi x)$. Here \mathbf{U} is the vector of conserved variables, A is an amplitude, and \mathbf{R} is the right-eigenvector

corresponding to the desired wave family. The components of \mathbf{R} for each test are listed in the appropriate results section below. For all the tests shown here, $A = 10^{-6}$.

The length of the computational domain is set to be one wavelength. Periodic boundary conditions are used. After the wave has propagated one wavelength, we measure the error in the numerical solution by computing the norm of the vector resulting from summing the absolute value of errors in each variable over the grid,

$$\begin{aligned}\epsilon &= \|\Delta U\| = \sqrt{\sum_k (\Delta U_k)^2} \\ &= \sqrt{\sum_k \left(\sum_{i=1}^{N_x} |U_{k,i}^n - U_{k,i}^0| / N_x \right)^2}\end{aligned}$$

Here, $U_{k,i}$ is the numerical solution for the k -th component of the vector of conserved quantities at grid point i and time level n , $U_{k,i}^0$ is the initial numerical solution, and N_x is the number of grid points. Note the initial solution $U_{k,i}^0$ is just the analytic solution which has been discretized to the grid. Since the discretization of the initial condition also introduces error, to measure the error in the integration algorithm it is important to compute errors relative to the initial condition $U_{k,i}^0$ rather than the analytic solution. In 2D (3D), summation over $j(k)$ is required as well.

To compare to the results given here, it is important to

1. Use the same amplitude for the wave
2. Compute the errors in exactly the same way
3. Compute the errors at exactly the same time

This is an excellent quantitative test of the accuracy and convergence rate of a numerical algorithm. The only drawback is that it involves only linear amplitude oscillations. Thus, this test is not characteristic of the kind of problem codes are written to solve in the first place (after all, the dynamics of linear waves can be treated analytically). A code which does well on this test may still be very poor at shock-capturing. Still, it is nice to know an algorithm reduces to the correct answer in the linear regime. Moreover, since virtually all schemes are first-order for discontinuities such as shocks, smooth problems like this are the only way to measure the actual convergence rate of higher-order schemes.

This test has proved very useful at detecting coding bugs:

- The errors for left- and right-going waves of the same family should be identical.
- If the errors do not converge, something is wrong somewhere.
- It is necessary to use double precision and very small wave amplitudes to eliminate round-off error and nonlinear effects.

32.2.1.1 1D Adiabatic Hydrodynamics

Given Equation 8.49 (with $\mathbf{B} = 0$),

$$\begin{aligned}\rho &= 1 \\ \mathbf{u} &= \mathbf{0} \\ p &= 1/\gamma\end{aligned}$$

with $\gamma = 5/3$. The conserved variables are

$$\mathbf{U} = [\rho, \rho u_x, \rho u_y, \rho u_z, E]$$

and the right-eigenvector for a left-going wave is

$$\mathbf{R} = [1.0, -1.0, 0.0, 0.0, 1.5]$$

When switching from \mathcal{E} to p , we can see from the background and \mathbf{R} that

$$\delta p = (\gamma - 1) \left(\delta \mathcal{E} + \rho u \delta u + \frac{1}{2} u^2 \delta \rho \right) = 1.0$$

The parameters are summarized in Table 32.1.

Table 32.1: 1D adiabatic hydrodynamic linear wave parameters

Variable	Background	Perturbation \mathbf{R}
ρ	1	1.0
ρu_x	0	-1.0
ρu_y	0	0.0
ρu_z	0	0.0
p	$\frac{1}{\gamma}$	1.0
\mathcal{E}	$\frac{1}{\gamma(\gamma-1)}$	1.5

The absolute error in propagating a sound wave to the left one wavelength as a function of the number of grid points N_x

32.2.1.2 1D Adiabatic MHD

Given Equation 8.49,

$$\begin{aligned}\rho &= 1 \\ \mathbf{u} &= \mathbf{0} \\ p &= 1/\gamma \\ \mathbf{B} &= [1, \sqrt{2}, 0.5]\end{aligned}$$

with $\gamma = 5/3$. Thus, the characteristic speeds along x is

$$\begin{aligned}v_s &= \sqrt{\frac{\gamma p}{\rho}} = 1.0 \\ v_A &= \sqrt{\frac{b_x^2 + b_y^2 + b_z^2}{\rho}} = \frac{\sqrt{13}}{2} \approx 1.8 \\ v_{xA} &= \sqrt{\frac{b_x^2}{\rho}} = 1.0 \\ v_{x,\text{fast}} &= \sqrt{\frac{1}{2} \left[v_s^2 + v_A^2 + \sqrt{(v_s^2 + v_A^2)^2 - 4v_s^2 v_{xA}^2} \right]} = 2.0 \\ v_{x,\text{slow}} &= \sqrt{\frac{1}{2} \left[v_s^2 + v_A^2 - \sqrt{(v_s^2 + v_A^2)^2 - 4v_s^2 v_{xA}^2} \right]} = 0.5\end{aligned}$$

If the conserved variables are ordered as

$$\mathbf{U} = [\rho, \rho u_x, \rho u_y, \rho u_z, E, B_y, B_z]$$

then the right eigenvectors for left going waves are as follows:

- For a fast magnetosonic wave:

```
U = [
4.472135954999580e-01
-8.944271909999160e-01
4.216370213557840e-01
1.490711984999860e-01
2.012457825664615e+00
8.432740427115680e-01
2.981423969999720e-01 ]
```

- For a Alfvén wave:

```

U = [
0.0
0.0
-3.33333333333333e-01
9.428090415820634e-01
0.0
-3.33333333333333e-01
9.428090415820634e-01 ]

```

- For a slow magnetosonic wave:

```

U = [
8.944271909999159e-01
-4.472135954999579e-01
-8.432740427115680e-01
-2.981423969999720e-01
6.708136850795449e-01
-4.216370213557841e-01
-1.490711984999860e-01 ]

```

The parameters are summarized in Table 32.2.

In some cases, we need to convert from conserved variables to other set of variables, e.g. $(n, \mathbf{u}, p, \mathbf{B})$. The number density perturbation is

$$\delta n = \frac{\delta\rho}{m} = \frac{\rho_0}{m} A \sin 2\pi x \quad (32.1)$$

Assuming background $u_0 = 0$,

$$\delta u = \frac{1}{\rho} [\delta(\rho u) - (\delta\rho)u] = \frac{1}{\rho_0 + \delta\rho} \delta(\rho u) \quad (32.2)$$

Given the perturbation of energy density

$$\begin{aligned} \mathcal{E} &= \frac{p}{\gamma - 1} + \frac{\rho u^2}{2} + \frac{B^2}{2\mu_0} \\ p &= (\gamma - 1) \left[\mathcal{E} - \frac{\rho u^2}{2} - \frac{B^2}{2\mu_0} \right] \end{aligned}$$

the pressure perturbation is then

$$\delta p = (\gamma - 1) \left[\delta\mathcal{E} - \frac{1}{2}(\rho_0 + \delta\rho)\delta u^2 - \frac{1}{2}(2B_0\delta B + \delta B^2) \right] \quad (32.3)$$

Table 32.2: 1D adiabatic MHD linear wave parameters

Table 32.3: fast magnetosonic wave

Variable	Background	Perturbation \mathbf{R}
ρ	1.0	4.472135954999580e-1
ρu_x	0	-8.944271909999160e-1
ρu_y	0	4.216370213557840e-1
ρu_z	0	1.490711984999860e-1
\mathcal{E}	$\frac{1}{\gamma(\gamma-1)} + \frac{13}{8}$	2.012457825664615
B_x	1.0	0.0
B_y	$\sqrt{2}$	8.432740427115680e-1
B_z	0.5	2.981423969999720e-1

Table 32.4: Alfvén wave

Variable	Background	Perturbation \mathbf{R}
ρ	1.0	0.0
ρu_x	0	0.0
ρu_y	0	-3.33333333333333e-1
ρu_z	0	9.428090415820634e-1
\mathcal{E}	$\frac{1}{\gamma(\gamma-1)}$	0.0
B_x	1.0	0.0
B_y	$\sqrt{2}$	-3.33333333333333e-1
B_z	0.5	9.428090415820634e-1

Table 32.5: slow magnetosonic wave

Variable	Background	Perturbation \mathbf{R}
ρ	1.0	8.944271909999159e-1
ρu_x	0	-4.472135954999579e-1
ρu_y	0	-8.432740427115680e-1
ρu_z	0	-2.981423969999720e-1
\mathcal{E}	$\frac{1}{\gamma(\gamma-1)}$	6.708136850795449e-1
B_x	1.0	0.0
B_y	$\sqrt{2}$	-4.216370213557841e-1
B_z	0.5	-1.490711984999860e-1

If the input is T instead of p , we have

$$\delta T = \frac{p_0 + \delta p}{\rho_0 + \delta \rho} - \frac{p_0}{\rho_0} \quad (32.4)$$

Note that while in the linear perturbation theory we can ignore the high order (>2) terms, we do not need to do so here numerically.

32.2.1.3 2D Adiabatic MHD

In this case, we use the same values as for the 1D adiabatic MHD test, but we use a 2D grid of size $0 \leq x \leq 2$ and $0 \leq y \leq 1$. We use twice as many grid points in the x-direction at every resolution (e.g. our highest resolution is 512×256), thus the grid is rectangular, but each cell is square. The wave propagates along the diagonal of the grid, at an angle $\theta = \tan^{-1}(0.5) \approx 26.6$ degrees with respect to the x-axis. Since the wave does not propagate along the diagonals of the grid cells, we guarantee the x- and y-fluxes are different; that is the problem is truly multi-dimensional.

32.2.2 3D Adiabatic MHD

Now we use a 3D grid of size $0 \leq x \leq 3$ and $0 \leq y \leq 1.5$, and $0 \leq z \leq 1.5$. The grid is of size $2N \times N \times N$. The wave propagates along the grid diagonal, again guaranteeing a truly multidimensional test. The background state is identical to the 1D test values.

32.2.3 Anisotropic MHD

This test is taken from (Daldorff et al. 2014). The parallel and perpendicular pressures vary differently in the collisionless plasma even if the unperturbed plasma has isotropic pressure. The isotropic MHD solutions are not valid in models like anisotropic MHD, hybrid, or PIC.

The following equations are an exact solution of the anisotropic MHD equations (Equation 8.49

but replace the energy equation with two pressure equations as in Equation 8.14)

$$\begin{aligned}\rho &= \rho_0 [1 + \delta \sin(kx - \omega t)] \\ u_x &= V_f \delta \sin(kx - \omega t) \\ u_y &= 0 \\ u_z &= 0 \\ p &= p_0 [1 + \gamma \delta \sin(kx - \omega t)] \\ p_{\parallel} &= p_0 [1 + \delta \sin(kx - \omega t)] \\ B_x &= 0 \\ B_y &= B_0 (1 + \delta \sin(kx - \omega t)) \\ B_z &= 0\end{aligned}$$

where $p = \frac{p_{\parallel} + 2p_{\perp}}{3}$ and

$$V_f = \frac{\omega}{k} = \sqrt{\frac{B_0^2 + 2p_0}{\rho_0}}$$

is the propagation speed of the fast wave moving perpendicular relative to the magnetic field direction. This corresponds to a case where $\theta = 90^\circ$ in Section 10.7.4. Note that p and p_{\parallel} have different wave amplitudes, i.e. this solution is different from the isotropic (collisional) fast wave. Also note the $2p_0$ (actually $2p_{\perp}$) term instead of the usual γp term in the phase speed.

Initially, we set $\rho_0 = 1$, $p_0 = 4.5 \times 10^{-4}$ and $B_0 = 0.04$ which results in $V_f = 0.05$. The wavelength is set to $\lambda = 32$ so that $k = 2\pi/\lambda$. The above solution also satisfies the Vlasov-Maxwell equations solved by a kinetic solver if the ion and electron gyro-radii are small relative to the wavelength λ and the propagation speed V_{ph} is much less than the speed of light c .

32.2.4 Circularly polarized Alfvén wave

The following relations describe a right-hand polarized Alfvén wave

$$\begin{aligned}\rho &= 1.0 \\ p &= 0.1 \\ v_{\parallel} &= 0.0 \\ v_{\perp} &= 0.1 \sin(2\pi x_{\parallel}) \\ v_z &= 0.1 \cos(2\pi x_{\parallel}) \\ B_{\parallel} &= 1.0 \\ B_{\perp} &= 0.1 \sin(2\pi x_{\parallel}) \\ B_z &= 0.1 \cos(2\pi x_{\parallel})\end{aligned}$$

with $\gamma = 5/3$ and $x_{\parallel} = (x \cos \alpha + y \sin \alpha)$ where α is the angle at which the wave propagates with respect to the grid. Here v_{\perp} and B_{\perp} are the components of velocity and magnetic field (in the x-y plane) perpendicular to the wavevector. They are related to the components stored on the grid B_x and B_y via

$$B_{\perp} = -B_x \sin \alpha + B_y \cos \alpha$$

$$B_{\parallel} = B_x \cos \alpha + B_y \sin \alpha$$

By inverting the system, we have

$$B_x = B_{\parallel} \cos \alpha - B_{\perp} \sin \alpha$$

$$B_y = B_{\parallel} \sin \alpha + B_{\perp} \cos \alpha$$

Generally the velocity and magnetic perturbation amplitudes are related by

$$\frac{v_1}{B_1} = \frac{V_A}{B_0}$$

For 2D the computational domain is of size $L_x = 2L_y$, with $N_x = 2N_y$. Thus, the grid is rectangular, but each cell is square. The wave propagates along the diagonal of the grid, at an angle $\alpha = \tan^{-1}(0.5) \approx 26.6$ degrees with respect to the x-axis. Since the wave does not propagate along the diagonals of the grid cells, we guarantee the x- and y-fluxes are different; that is the problem is truly multi-dimensional.

The wave is an exact nonlinear solution to the MHD equations, allowing one to test the algorithm in the nonlinear regime.¹ Although nonlinear amplitude Alfvén waves are subject to a parametric instability (Section 17.4) which causes them to decay into magnetosonic waves, the instability should not be present for the circularly polarized Alfvén waves. Since the problem is smooth, it can be used for convergence testing. Running the test with smaller pressure (higher β) and/or larger amplitudes is a good test of how robust is the algorithm.

See more in [reference](#) and (Tóth 2000).

32.2.5 Whistler wave

This test involves the propagation of whistler waves taken from (Daldorff et al. 2014). This is an extension of the circularly polarized Alfvén wave with the Hall term. The whistler waves have variations in the transverse components of the magnetic field and the velocity. The longitudinal components, the density and the pressure are not perturbed. The following

¹As opposed to linearly polarized Alfvén wave, which is not an exact solution to the nonlinear MHD equations! Polarization affects decay. Ask Chaitanya for more about this point.

solution of the Hall MHD equations describes a right-hand polarized whistler wave propagating in the $+x$ -direction:

$$\begin{aligned} u_y &= -\delta u \cos(kx - \omega t) \\ u_z &= +\delta u \sin(kx - \omega t) \\ B_y &= +\delta B \cos(kx - \omega t) \\ B_z &= -\delta B \sin(kx - \omega t) \end{aligned} \quad (32.5)$$

where the wave speed is (Equation 10.25)

$$v_{\text{ph}} = \frac{\omega}{k} = \frac{W}{2} + \sqrt{\frac{B_x^2}{\rho} + \frac{W^2}{4}}, \quad W = \frac{m}{e} \frac{k B_x}{\rho}$$

Note that the wave speed depends on the wave number k . The magnetic and velocity perturbations are related as (This is different from the usual Alfvénicity Equation 18.4???) But it reduces to the Alfvénicity condition if $v_{\text{ph}} = v_A = B_x / \sqrt{\rho}$.)

$$\frac{\delta u}{\delta B} = \frac{B_x}{v_{\text{ph}} \rho} \quad (32.6)$$

We set $\rho = 1$, $u_x = 0$, $B_x = 0.2$ and $p = 5.12 \times 10^{-4}$, and wavelength $\lambda = 32$. This results in $v_{\text{ph}} = 0.22059$ that is about 10% faster than the Alfvén speed $V_A = 0.2$, so the Hall term is small but not negligible at this wavelength. The amplitude of the transverse magnetic field perturbation is set to $\delta B = 0.01$ which is 5% of the background field.

These parameters are chosen to satisfy

- $c \gg v_{\text{ph}}$
- $m_e \ll m_i$
- $\beta \ll 1$
- $\omega \ll \Omega_e$

such that the kinetic dispersion relation reduces to the Hall MHD dispersion relation for the right-hand polarized whistler waves.

(Tenerani et al. 2023) shows the Alfvénic wave dispersion and damping relations with respect to λ/d_i . When we get close to but not necessarily below the ion inertial length (for $\beta \sim 1$, d_i and r_{iL} are on the same order), these effects already take place from both Hall MHD and hybrid simulations.

What I observe while doing this low- β test with the hybrid-Vlasov code Vlasiator is that without the Hall term the wave profile will be quickly distorted starting from the trailing edge of the peaks and troughs of all quantities; with the Hall term the wave profile can be maintained much longer, and of course the phase speed is v_W which is faster than v_A as in the ideal MHD case. In Section 8.6.1, we learned that the Hall term in the Ohm's law can be neglected under

two cases, neither of which apply in this test as shown by the comparison of V_A and v_W above. This implies an important fact that in certain parameter spaces the classical Alfvén waves never exist; we may observe Alfvénic fluctuations (i.e. correlations between $\delta\mathbf{u}$ and $\delta\mathbf{B}$), but they may belong to whistler waves (R-waves), EMIC waves (L-waves), or KAWs.

32.2.6 Light wave

32.2.7 Firehose instability

32.2.8 Mirror mode instability

32.2.9 Langmuir wave

Langmuir wave is a rapid electrostatic oscillation in plasma driven by Coulomb force. Here we assume cold plasma and immobile ions. From the fluid perspective, the governing equations are given in Equation 10.2:

$$\begin{aligned}\frac{\partial n_1}{\partial t} &= -n_0 \frac{\partial u_1}{\partial x} \\ \frac{\partial u_1}{\partial t} &= -\frac{e}{m} E_1 \\ \frac{\partial E_1}{\partial x} &= -\frac{e}{\epsilon_0} n_1\end{aligned}$$

The third equation can be alternatively written using the electric potential ϕ which is defined via $\mathbf{E} = -\nabla\phi = -\partial\phi/\partial x$,

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{e}{\epsilon_0} n_1$$

Then the full equations are

$$\begin{aligned}\frac{\partial n_1}{\partial t} &= -n_0 \frac{\partial u_1}{\partial x} \\ \frac{\partial u_1}{\partial t} &= -\frac{e}{m} \frac{\partial \phi}{\partial x} \\ \frac{\partial^2 \phi}{\partial x^2} &= \frac{e}{\epsilon_0} n_1\end{aligned}$$

Langmuir oscillation can be driven self-consistently if there is a density perturbation in electrons, i.e. local charge imbalance. A classical test is to set the 1D periodic spatial domain between $[0, 2\pi]$ with 512 points, and time step $\Delta t = 0.01$. The initial perturbation density is

$n_1 = 0.02n_0 \cos(x)$, where n_0 is the uniform background electron density. Temperature is set to 0. Total simulation time is 100.²

The electron equation of motion is

$$\begin{aligned} m \frac{dv}{dt} &= -eE \\ \frac{dx}{dt} &= v \end{aligned}$$

32.3 Shock Tests

32.3.1 Brio-Wu shock tube

This test is an MHD shocktube, where the right and left states are initialized to different values. The initial left/right values are

$$\begin{aligned} \rho &= 1.0, 0.125 \\ u_x &= 0.0, 0.0 \\ u_y &= 0.0, 0.0 \\ u_z &= 1.0, -1.0 \\ B_x &= 0.75, 0.75 \\ B_y &= 0.0, 0.0 \\ B_z &= 0.0, 0.0 \\ p &= 1.0, 0.1 \end{aligned}$$

and $\gamma = 2$. The hydrodynamic portion of the initial conditions are the same as for the Sod shock tube problem.

This is a standard test for MHD codes for checking whether the code can accurately represent the shocks, rarefactions, contact discontinuities, and the compound structures of MHD.

32.3.2 Ryu and Jones Test 2A

This test is an MHD shocktube, where the right and left states are initialized to different values. It involves a three-dimensional field and velocity structure and rotation of the plane of the

²Using a simple finite difference solver, I found that solving E is unstable because the finite difference matrix on the left-hand-side of Poisson's equation is singular. Solving the electric potential with Laplace's equation is fine.

magnetic field. The initial left/right values are

$$\begin{aligned}\rho &= 1.08, 1.0 \\ u_x &= 1.2, 0.0 \\ u_y &= 0.01, 0.0 \\ u_z &= 0.5, 0.0 \\ B_x &= 3.6, 2.0 \\ B_y &= 2.0, 4.0 \\ B_z &= 2.0, 2.0 \\ p &= 0.95, 1.0\end{aligned}$$

This test contains fast shocks, slow shocks, and rotational discontinuities which propagate to each side of the contact discontinuity. The ability of the scheme to capture all 7 waves in MHD can be checked with this single test.

32.3.3 Spherical blast waves

We used a rectangular domain, $-0.5 \leq x \leq 0.5; -0.75 \leq y \leq 0.75$. The boundary conditions are periodic everywhere. This non-square domain and periodic boundary conditions produces complex shock-shock and shock-CD interactions at late times.

The initial conditions are

$$\begin{aligned}\rho &= 1.0 \\ \mathbf{u} &= [0.0, 0.0, 0.0] \\ p &= 0.1\end{aligned}$$

with $\gamma = 5/3$. Initial velocities are zero everywhere. Within the region $r < 0.1$, $p = 10.0$ (that is, 100 times the ambient pressure). For the MHD problem, the initial magnetic field is uniform everywhere with $B_x = B_y = 1/\sqrt{2}$.

Although this test is not very quantitative, it makes great movies!

At early times, it is important that the out-going blast wave is spherical and shows no grid alignment effects. At late times, the interaction of the blast wave with the CD at the edge of the evacuated bubble in the center produces filaments of dense gas by the Richtmyer-Meshkov instability. It is important these fingers are sharp and not diffused away. Moreover, for the hydrodynamical problem, the pattern of the fingers should be exactly symmetric top-to-bottom and left-to-right. For the MHD problem, the Richtmyer-Meshkov instability is suppressed, and no fingers are evident.

32.4 Instability Tests

32.4.1 Kelvin-Helmholtz instability

We use a square domain, $-0.5 \leq x \leq 0.5; -0.5 \leq y \leq 0.5$. The boundary conditions are periodic everywhere. For $|y| > 0.25$, we set $\rho = 1$ and $u_x = -0.5$, for $|y| \leq 0.25$, $\rho = 2$ and $u_x = 0.5$. The pressure is 2.5 everywhere, and $\gamma = 1.4$, giving a Mach number of about 0.377 in the $\rho = 2$ gas, and about 0.267 in the $\rho = 1$ gas. The interface between the two oppositely directed streams is a discontinuity, that is a “slip surface”. We use different densities in the two fluids to make visualization of the interface easier.

For the MHD problem, the initial magnetic field is uniform everywhere with $B_x = 0.5$.

To seed the instability, we add random numbers to both the x- and y-components of the velocity with peak-to-peak amplitude of 0.01.

At early times, one can check that the growth rate of the transverse component of the velocity agrees with the prediction from linear theory. This requires initializing a single-mode perturbation rather than a spectrum of perturbations as we have done here.

At late times, once the instability has gone fully nonlinear, it is difficult to make quantitative comparisons. However, the sharpness of the boundary between the two streams is an indication of the numerical diffusion of the scheme. For example, if the HLLE Riemann solver is used, diffusion at the interface is significant enough to suppress the instability.

32.4.2 Rayleigh-Taylor instability

For the single-mode test, we use a rectangular domain, $-0.25 \leq x \leq 0.25; -0.75 \leq y \leq 0.75$. The boundary conditions are periodic at $|x| = 0.25$, and reflecting walls at $|y| = 0.75$. For $y > 0$ the density is 2.0, while for $y \leq 0$ it is 1.0. A constant gravitational acceleration $g = 0.1$ must be added to the equations of motion. The pressure is given by the condition of hydrostatic equilibrium, that is $p = p_0 - 0.1\rho y$, where $p_0 = 2.5$, and $\gamma = 1.4$. This gives a sound speed of 3.5 in the low density medium at the interface.

The structures which appear in the nonlinear regime are very sensitive to the nature of the perturbations used to seed the instability. To avoid gridding errors associated with perturbing the interface, we instead perturb the velocities. For the single-mode perturbation, we set $u_y = 0.01[1 + \cos(4\pi x)][1 + \cos(3\pi y)]/4$.

For the multimode perturbation, we use a domain of size $-0.25 \leq x \leq 0.25; -0.375 \leq y \leq 0.375$, and set $u_y = A[1 + \cos(8\pi y/3)]/2$, where A is a random number at each zone with a peak-to-peak amplitude of 0.01.

The way in which source terms are included in the algorithm can have a strong effect on the outcome of this test. For example, for Godunov schemes, if the source term is added

using operator splitting, grid noise generated by the lack of an exact numerical equilibrium can perturb the interface and seed structure. If the source terms are included directly in the reconstruction and integration steps, it is able to hold hydrostatic equilibrium automatically.

At early times, one can check that the growth rate of the vertical component of the velocity agrees with the prediction from linear theory.

At late times, once the instability has gone fully nonlinear, it is difficult to make quantitative comparisons. However, the sharpness of the boundary between the two fluids is an indication of the numerical diffusion of the scheme. Also, the amount of fine scale struture induced by secondary KH instabilities is sensitive to the way the interface is perturbed, and how sharp the algorithm preserves the contact discontinuity. It is not always clear that sharper is better, however. For example the “contact steepener” in the PPM algorithm can introduce “stair stepping” in contact discontinuities in multidimensions, which in turn can cause KH rolls to be seeded by grid noise.

32.5 Turbulence Tests

32.5.1 Orszag-Tang vortex

We use a square domain, $0 \leq x \leq 1; 0 \leq y \leq 1$. The boundary conditions are periodic everywhere. The density ρ is $25/(36\pi)$ and the pressure is $5/(12\pi)$ everywhere, and $\gamma = 5/3$. Note that this choice gives $u_s^2 = \gamma p/\rho = 1$. The initial velocities are periodic with $u_x = -\sin(2\pi y)$ and $u_y = \sin(2\pi x)$. The magnetic field is initialized using a periodic vector potential defined at zone corners; $A_z = B_0(\cos(4\pi x)/(4\pi) + \cos(2\pi y)/(2\pi))$, with $B_0 = 1.0$. Face-centered magnetic fields are computed using $\mathbf{B} = \nabla \times \mathbf{A}$ to guarantee $\nabla \cdot \mathbf{B} = 0$ initially. This gives $B_x = -B_0 \sin(2\pi y)$ and $B_y = B_0 \sin(4\pi x)$.

The Orszag-Tang vertex is a well-known model problem for testing the transition to supersonic 2D MHD turbulence. Thus, the problem tests how robust the code is at handling the formation of MHD shocks, and shock-shock interactions. The problem can also provide some quanitative estimates of how significant magnetic monopoles affect the numerical solutions, testing the $\nabla \cdot \mathbf{B} = 0$ condition. Finally, the problem is a very common test of numerical MHD codes in two dimensions, and has been used in many previous studies. As such, it provides a basis for consistent comparison of codes.

32.6 Reconnection Tests

32.6.1 GEM challenge

This 2D setup is based on (Birn et al. 2001). The computation is carried out in a rectangular domain $-L_x/2 \leq x \leq L_x/2$ and $-L_z/2 \leq z \leq L_z/2$. The system is taken to be periodic in the x direction with ideal conducting boundaries at $z = \pm L_z/2$. Thus the boundary conditions on the magnetic fields at the z boundaries are $B_z = \partial B_x / \partial z = \partial B_y / \partial z = 0$ with corresponding conditions on the electric fields and particle or fluid quantities. Open boundary conditions are used for all quantities.

The equilibrium chosen for the reconnection challenge problem is a Harris equilibrium with a floor in the density outside of the current layer. The magnetic field is given by

$$B_x(z) = B_0 \tanh(z/\lambda)$$

where λ is the current sheet scale size, and the density by

$$n(z) = n_0 \operatorname{sech}^2(z/\lambda) + n_\infty$$

The electron and ion temperatures, T_e and T_i , are taken to be uniform in the initial state. We assume the plasma $\beta = (p_i + p_e)/p_B = 1$ initially.

The normalization of the space and time scales of the system is chosen to be the ion inertial length $d_i = c/\omega_{pi}$ and the ion cyclotron frequency Ω_i^{-1} , where $\omega_{pi}^2 = n_0 e^2 / \epsilon_0 m_i$ is evaluated with the density n_0 and the ion gyrofrequency $\Omega_i = eB_0/m_i$ is evaluated at the peak magnetic field. The velocities are then normalized to the Alfvén speed v_A . In the normalized units, $B_0 = 1$ and $n_0 = 1$. Specific parameters for the simulations are $L_x = 25.6$, $L_z = 12.8$, $\lambda = 0.5$, $n_\infty/n_0 = 0.2$, and $T_e/T_i = 0.2$. $m_i/m_e = 25$ is assumed if required.

The initial magnetic island is specified through the perturbation in the magnetic flux,

$$\psi(x, z) = \psi_0 \cos(2\pi x/L_x) \cos(\pi z/L_z)$$

where the magnetic perturbation is given by $\mathbf{B} = \hat{y} \times \nabla\psi$, or more specifically,

$$\begin{aligned} B_{1x} &= -\psi_0 \left(\frac{\pi}{L_z} \right) \cos(2\pi x/L_x) \sin(\pi z/L_z) \\ B_{1z} &= \psi_0 \left(\frac{2\pi}{L_x} \right) \sin(2\pi x/L_x) \cos(\pi z/L_z) \end{aligned}$$

In normalized units $\psi_0 = 0.1$, which produces an initial island width which is comparable to the initial width of the current layer. The rationale for such a large initial perturbation is to put the system in the nonlinear regime of magnetic reconnection from the beginning of the simulation.

The initial setup is shown in Figure 32.1.

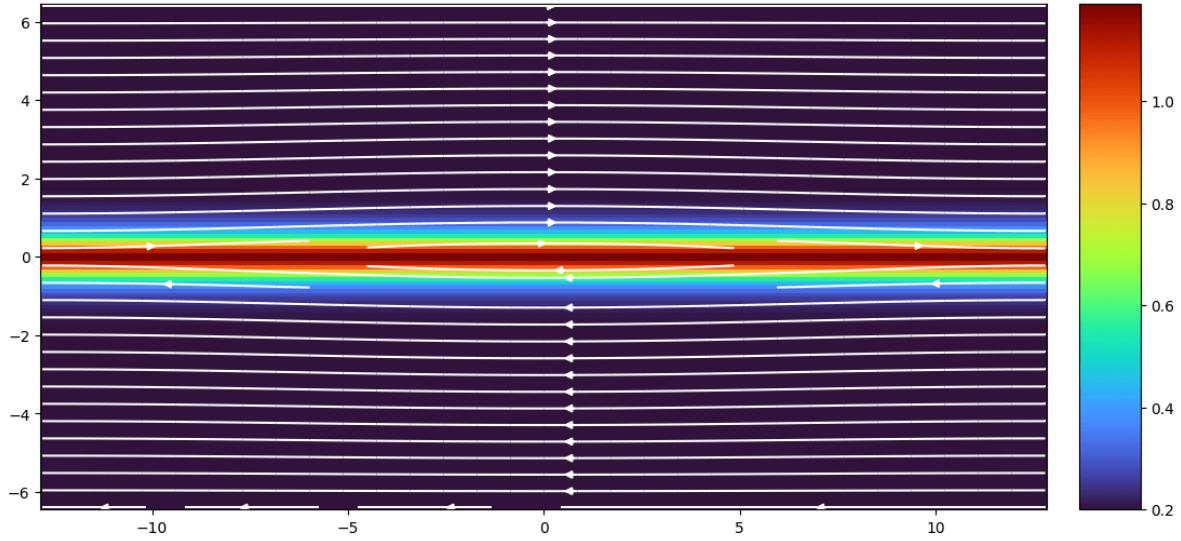


Figure 32.1: Initial density and magnetic field in the GEM current sheet setup.

32.6.1.1 Why do we need to resolve ion inertial length

Physically, electrons and ions separate at the scale of ion inertial length. Numerically, Hall term is important only when cell size is small enough to resolve the ion initial length. The reason is as follows. The Ohm's law is

$$\mathbf{E} = -(\mathbf{U} + \mathbf{U}_H) \times \mathbf{H}$$

Assume the typical flow velocity is Alfvén velocity: $\mathbf{U} = \mathbf{V}_A$. The Hall velocity is estimated as:

$$\mathbf{U}_H = -\frac{\mathbf{J}}{ne} = -\frac{\nabla \times \mathbf{B}}{\mu_0 ne} \sim -\frac{|dB|}{\mu_0 ne \Delta x} \sim -\frac{|B|}{\mu_0 ne \Delta x}$$

The approximation $dB \approx B$ is valid for relatively coarse grid size; for fine discrete cell sizes, this approximation does not hold, so magnetic field cannot cancel out in the following estimation. Let us now assume we can make this assumption. Then the ratio of Hall velocity and Alfvén velocity is:

$$\frac{|\mathbf{U}_H|}{|\mathbf{V}_A|} = \frac{c/\omega_{pi}}{\mu_0 \Delta x}$$

Ion inertial length is also important for PIC: if particle's velocity is assumed to be Alfvén velocity, then ion inertial length is the same as ion gyroradius.

32.6.2 Current sheet

The grid is a square with $-0.5 \leq x \leq 0.5$ and $-0.5 \leq y \leq 0.5$. The density and pressure are uniform everywhere, with $\rho = 1$ and $p = \beta/2$ where β is an input parameter. For $|x| > 0.25$ we set $B_y/\sqrt{4\pi} = 1$, otherwise $B_y/\sqrt{4\pi} = -1$. The velocities are $u_x = A \sin(2\pi y)$ (where A is an amplitude) and $u_y = 0$. The “standard” test uses $\beta = 0.1$ and $A = 0.1$, although part of the point of this test is to see how small (large) a value of $\beta(A)$ is required to break the code.

Although we do not know the analytic solution for this problem, it may be an excellent test of the robustness of the algorithm. For ideal MHD, initially the solution should be nonlinearly (???) polarized Alfvén waves propagating along the field in the y -direction (which quickly generate magnetosonic waves since the magnetic pressure does not remain constant). However, because of the two current sheets in the problem (at $x = \pm 0.25$), reconnection inevitably occurs. Because $\beta < 1$, this reconnection drives strong over-pressurized regions that launch magnetosonic waves transverse to the field. Moreover, as reconnection changes the topology of the field lines, magnetic islands will form, grow, and merge. The point of the test is to make sure the algorithm can follow this evolution for as long as possible without crashing. Keeping $\nabla \cdot \mathbf{B} = 0$ as the field topology undergoes complex changes could be important.

32.7 Divergence-free Field test

32.7.1 Magnetic field loop

33 Inductance

Inductance is the ability of an inductor to store energy and it does this in the magnetic field that is created by the flow of electrical current. Energy is required to set up the magnetic field and this energy is released when the field falls. As a result of the magnetic field associated with the current flow, inductors generate an opposing voltage proportional to the rate of change in current in a circuit. Inductance is caused by the magnetic field generated by electric currents flowing within an electrical circuit. Typically coils of wire are used as a coil increases the coupling of the magnetic field and increases the effect.

There are two ways in which inductance is used:

- *Self-inductance*: Self-inductance is the property of a circuit, often a coil, whereby a change in current causes a change in voltage in that circuit due to the magnetic effect caused by the current flow. It can be seen that self-inductance applies to a single circuit — in other words it is an inductance, typically within a single coil. This effect is used in single coils or chokes.
- *Mutual-inductance*: Mutual inductance is an inductive effect where a change in current in one circuit causes a change in voltage across a second circuit as a result of a magnetic field that links both circuits. This effect is used in transformers.

33.1 Unit Definition

When indicating an inductor on a circuit diagram or within an equation, generally the symbol “L” is used. On circuit diagrams, inductors are generally numbered, L1, L2, etc.

The SI unit of inductance is the henry, H which can be defined in terms of rate of change of current and voltage. The inductance of a circuit is one henry if the rate of change of current in a circuit is one ampere per second and this results in an electromotive force of one volt.
 $1H = 1Wb/A$.

33.2 Phenomena

When a current flows within a conductor, whether it be straight or in the form of a coil, a magnetic field builds up around it and this affects the way in which the current builds up after the circuit is made.

In terms of how inductance affects an electrical circuit, it helps to look at the way the circuit operates, first for a direct current (DC), and then for an alternating current (AC). Although they follow the same laws and the same effects result, it helps the explanation, the direct current example is simpler, and then this explanation can be used as the basis for the alternating current case.

33.2.1 Direct current

As the circuit is made the current starts to flow. As the current increases to its steady value the magnetic field it produces builds up to its final shape. As this occurs, the magnetic field is changing, so this induces a voltage back into the coil itself, as would be expected according to Lenz's Law.

The time constant T in seconds of the circuit which will include the inductor value L Henries and the associated circuit resistance, R Ohms can be calculated as $T = L/R$. T is the time for the current I amps to rise to 0.63 of its final steady state value of V/R . The energy stored in the magnetic field is $\frac{1}{2}LI^2$.

When the current is switched off this means that effectively the resistance of the circuit rises suddenly to infinity. This means that the ratio L/R becomes very small and the magnetic field falls very rapidly. This represents a large change in magnetic field and accordingly the inductance tries to keep the current flowing and a back *electromagnetic force* (EMF) is set up to oppose this arising from the energy stored in the magnetic field.

When the back EMF is set up, the very high voltages generated mean that sparks can appear across the switch contact, especially just as the contact is broken. This leads to pitted contacts and wear on any mechanical switches. In electronic circuits this back EMF can destroy semiconductor devices and therefore ways of reducing this back EMF are often employed.

33.2.2 Alternating current

For the case of the AC passing through an inductor, the same basic principles are used, but as the waveform is repetitive, we tend to look at the way the inductor responds in a slightly different way as it is more convenient.

By its very nature, an alternating waveform is changing all of the time. This means that the resulting magnetic field will always be changing, and there will always be an induced back

EMF produced. The result of this is that the inductor impedes the flow of the alternating current through it as a result of the inductance. This is in addition to the resistance caused by the Ohmic resistance of the wire.

It means that if the Ohmic resistance of the inductor is low, it will pass DC with little loss, but it can present a high impedance to any high frequency signal. This characteristic of an inductor can be used in ensuring that any high frequency signals do not pass through the inductor.

A further aspect of inductance is that the reactance of an inductor and that of a capacitor can act together in a circuit to cancel each other out. This is known as *resonance*, and it is widely used in bandpass filters.

33.3 Inductors

Inductors are electronic components that use inductance in an electronic circuit. These inductors are normally wound components having many turns of wire to increase the level of inductance. They may also be wound on ferromagnetic cores to further increase the level of inductance.

33.3.1 Inductance of Wires and Coils

Straight wires and coils have an inductance. Normally coils are used for inductors because the linking of the magnetic field between the different turns of the coil increases the inductance and enables the wire to be contained within a smaller volume.

If the wire was not coiled, then very long lengths of wire would often be needed making electronic components of this nature not viable. By coiling the wire the inductance is maximised enabling inductors to be incorporated into many electronic circuits.

However, even the inductance of a straight wire can affect some electronic circuits. For most low frequency applications, the inductance of a straight wire can be ignored, but as the frequency increases into the VHF region and beyond, the inductance of the wire itself can become significant, and interconnections need to be kept short to minimise the effects.

Calculations are available to enable the inductance of wires to be calculated quite accurately, but the inductance of coils is a little more complicated and depends upon a variety of factors including the shape of the coil and the constant of the material in and around the coil.

33.3.2 Applications

Inductance is a very important aspect of electronic circuit design. Although inductors are not so widely used in low frequency electronic circuit designs because the size of the electronic components required to give the levels of inductance needed is large, they are widely used for much higher frequencies in radio frequency designs, as well as within EMC — where filtering is used, often using inductors to ensure that any interference is not able to pass along wires and cables. For example a simple form of inductor is often seen on computer cables where a ferrite is added around a cable to add inductance and prevent the signals from travelling along the cable and being transmitted, thereby giving rise to the possibility of interference to other systems.

34 Capacitance

A capacitor is a little like a battery but works completely differently. A battery is an electronic device that converts chemical energy into electrical energy, whereas a capacitor is a passive electronic component that stores electrostatic energy in an electric field.

34.1 What Is a Capacitor?

A capacitor is a two-terminal electrical device that can store energy in the form of an electric charge. It consists of two electrical conductors that are separated by a distance. The space between the conductors may be filled by vacuum or with an insulating material known as a dielectric. The ability of the capacitor to store charges is known as capacitance.

Capacitors store energy by holding apart pairs of opposite charges. The simplest design for a capacitor is a parallel plate, which consists of two metal plates with a gap between them. But, different types of capacitors are manufactured in many forms, styles, lengths, girths, and materials.

Appendix

This is the appendix.

```
example_dataframe()
```

References

- Abraham-Shrauner, Barbara. 1967. “Shock Jump Conditions for an Anisotropic Plasma.” *Journal of Plasma Physics* 1 (3): 379–81. <https://doi.org/10.1017/S0022377800003366>.
- Amano, Takanobu, Katsuaki Higashimori, and Keisuke Shirakawa. 2014. “A Robust Method for Handling Low Density Regions in Hybrid Simulations for Collisionless Plasmas.” *Journal of Computational Physics* 275: 197–212. <https://doi.org/10.1016/j.jcp.2014.06.048>.
- Angelopoulos, Vassilis, Wo Baumjohann, CF Kennel, F Vo Coroniti, MG Kivelson, R Pellat, RJ Walker, H Lühr, and G Paschmann. 1992. “Bursty Bulk Flows in the Inner Central Plasma Sheet.” *Journal of Geophysical Research: Space Physics* 97 (A4): 4027–39. <https://doi.org/10.1029/91JA02701>.
- Arnoldy, Roger L, MJ Engebretson, RE Denton, JL Posch, Marc R Lessard, NC Maynard, DM Ober, et al. 2005. “Pc 1 Waves and Associated Unstable Distributions of Magnetospheric Protons Observed During a Solar Wind Pressure Pulse.” *Journal of Geophysical Research: Space Physics* 110 (A7). <https://doi.org/10.1029/2005JA011041>.
- Axford, WI, E Leer, and G Skadron. 1977. “The Acceleration of Cosmic Rays by Shock Waves.” In *International Cosmic Ray Conference*. Vol. 11. Springer. <https://doi.org/10.1007/978-3-662-25523-0>.
- Baker, Daniel N, TI Pulkkinen, V Angelopoulos, W Baumjohann, and RL McPherron. 1996. “Neutral Line Model of Substorms: Past Results and Present View.” *Journal of Geophysical Research: Space Physics* 101 (A6): 12975–3010. <https://doi.org/10.1029/95JA03753>.
- Bale, SD, FS Mozer, and TS Horbury. 2003. “Density-Transition Scale at Quasiperpendicular Collisionless Shocks.” *Physical Review Letters* 91 (26): 265004. <https://doi.org/10.1103/PhysRevLett.91.265004>.
- Balogh, André, and Rudolf A Treumann. 2013. *Physics of Collisionless Shocks: Space Plasma Shock Waves*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4614-6099-2>.
- Bame, SJ, JR Asbridge, WC Feldman, S Peter Gary, and Michael D Montgomery. 1975. “Evidence for Local Ion Heating in Solar Wind High Speed Streams.” *Geophysical Research Letters* 2 (9): 373–75. <https://doi.org/10.1029/GL002i009p00373>.
- Bame, SJ, JR Asbridge, WC Feldman, JT Gosling, G Paschmann, and N Sckopke. 1980. “Deceleration of the Solar Wind Upstream from the Earth’s Bow Shock and the Origin of Diffuse Upstream Ions.” *Journal of Geophysical Research: Space Physics* 85 (A6): 2981–90. <https://doi.org/10.1029/ja085ia06p02981>.
- Bellan, Paul M. 2008. *Fundamentals of Plasma Physics*. Cambridge University Press.
- Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah. 2017. “Julia: A Fresh Approach to Numerical Computing.” *SIAM Review* 59 (1): 65–98.

- Birn, J, JF Drake, MA Shay, BN Rogers, RE Denton, M Hesse, M Kuznetsova, et al. 2001. "Geospace Environmental Modeling (GEM) Magnetic Reconnection Challenge." *Journal of Geophysical Research: Space Physics* 106 (A3): 3715–19. <https://doi.org/10.1029/1999JA900449>.
- Bruno, Roberto, and Vincenzo Carbone. 2013. "The Solar Wind as a Turbulence Laboratory." *Living Reviews in Solar Physics* 10 (1): 1–208. <https://doi.org/10.12942/lrsp-2013-2>.
- Budden, Kenneth George. 1961. *Radio Waves in the Ionosphere. Radio Waves in the Ionosphere.*
- Bujack, Roxana, Karen Tsai, Steven K Morley, and Etienne Bresciani. 2021. "Open Source Vector Field Topology." *SoftwareX* 15: 100787. <https://doi.org/10.1016/j.softx.2021.100787>.
- Cary, John R, and Alain J Brizard. 2009. "Hamiltonian Theory of Guiding-Center Motion." *Reviews of Modern Physics* 81 (2): 693. <https://doi.org/10.1103/RevModPhys.81.693>.
- Chen, Francis F. 2016. *Introduction to Plasma Physics and Controlled Fusion*. Vol. 1. Springer.
- Chen, Liu, Fulvio Zonca, and Yu Lin. 2021. "Physics of Kinetic Alfvén Waves: A Gyrokinetic Theory Approach." *Reviews of Modern Plasma Physics* 5 (1): 1–37. <https://doi.org/10.1007/s41614-020-00049-3>.
- Chew, GF, ML Goldberger, and FE Low. 1956. "The Boltzmann Equation and the One-Fluid Hydromagnetic Equations in the Absence of Particle Collisions." *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 236 (1204): 112–18. <https://doi.org/10.1098/rspa.1956.0116>.
- Dahlin, JT, JF Drake, and M Swisdak. 2014. "The Mechanisms of Electron Heating and Acceleration During Magnetic Reconnection." *Physics of Plasmas* 21 (9). <https://doi.org/10.1063/1.4894484>.
- Dai, Lei, and Chi Wang. 2023. "Kinetic Alfvén Wave Eigenmode in Magnetosphere Magnetic Reconnection." *Reviews of Modern Plasma Physics* 7 (1): 1–38. <https://doi.org/10.1007/s41614-022-00107-y>.
- Daldorff, Lars KS, Gábor Tóth, Tamas I Gombosi, Giovanni Lapenta, Jorge Amaya, Stefano Markidis, and Jeremiah U Brackbill. 2014. "Two-Way Coupling of a Global Hall Magnetohydrodynamics Model with a Local Implicit Particle-in-Cell Model." *Journal of Computational Physics* 268: 236–54. <https://doi.org/10.1016/j.jcp.2014.03.009>.
- De Sterck, Hans, BC Low, and Stefaan Poedts. 1998. "Complex Magnetohydrodynamic Bow Shock Topology in Field-Aligned Low- β Flow Around a Perfectly Conducting Cylinder." *Physics of Plasmas* 5 (11): 4015–27. <https://doi.org/10.1063/1.873124>.
- Eastwood, JP, A Balogh, EA Lucek, C Mazelle, and I Dandouras. 2005. "Quasi-Monochromatic ULF Foreshock Waves as Observed by the Four-Spacecraft Cluster Mission: 2. Oblique Propagation." *Journal of Geophysical Research: Space Physics* 110 (A11). <https://doi.org/10.1029/2004ja010618>.
- Elkington, Scot R, Mary K Hudson, and Anthony A Chan. 1999. "Acceleration of Relativistic Electrons via Drift-Resonant Interaction with Toroidal-Mode Pc-5 ULF Oscillations." *Geophysical Research Letters* 26 (21): 3273–76. <https://doi.org/10.1029/1999gl003659>.
- Erkaev, NV, DF Vogl, and HK Biernat. 2000. "Solution for Jump Conditions at Fast Shocks in an Anisotropic Magnetized Plasma." *Journal of Plasma Physics* 64 (5): 561–78. <https://doi.org/10.1017/S002237780000893X>.

- Fermi, Enrico. 1949. "On the Origin of the Cosmic Radiation." *Physical Review* 75 (8): 1169. <https://doi.org/10.1103/PhysRev.75.1169>.
- Friedel, RHW, A Korth, and G Kremser. 1996. "Substorm Onsets Observed by CRRES: Determination of Energetic Particle Source Regions." *Journal of Geophysical Research: Space Physics* 101 (A6): 13137–54. <https://doi.org/10.1029/96JA00399>.
- Furth, Harold P, John Killeen, and Marshall N Rosenbluth. 1963. "Finite-Resistivity Instabilities of a Sheet Pinch." *The Physics of Fluids* 6 (4): 459–84. <https://doi.org/10.1063/1.1706761>.
- Gary, S Peter. 1993. *Theory of Space Plasma Microinstabilities*. 7. Cambridge university press.
- Gary, S Peter, JT Gosling, and DW Forslund. 1981. "The Electromagnetic Ion Beam Instability Upstream of the Earth's Bow Shock." *Journal of Geophysical Research: Space Physics* 86 (A8): 6691–96. <https://doi.org/10.1029/ja086ia08p06691>.
- Gary, S Peter, Charles W Smith, Martin A Lee, Melvyn L Goldstein, and David W Forslund. 1984. "Electromagnetic Ion Beam Instabilities." *The Physics of Fluids* 27 (7): 1852–62. <https://doi.org/10.1063/1.864797>.
- Gary, S Peter, Michelle F Thomsen, and Stephen A Fuselier. 1986. "Electromagnetic Instabilities and Gyrophase-Bunched Particles." *The Physics of Fluids* 29 (2): 531–35. <https://doi.org/10.1063/1.865441>.
- Glassmeier, Karl-Heinz, Carsten Othmer, Rüdiger Cramm, Martin Stellmacher, and Mark Engebretson. 1999. "Magnetospheric Field Line Resonances: A Comparative Planetology Approach." *Surveys in Geophysics* 20 (1): 61–109. <https://doi.org/10.1023/A:1006659717963>.
- Glauert, Sarah A, Richard B Horne, and Nigel P Meredith. 2014. "Three-Dimensional Electron Radiation Belt Simulations Using the BAS Radiation Belt Model with New Diffusion Models for Chorus, Plasmaspheric Hiss, and Lightning-Generated Whistlers." *Journal of Geophysical Research: Space Physics* 119 (1): 268–89. <https://doi.org/10.1002/2013JA019281>.
- Goldstein, Melvyn L, Hung K Wong, Adolfo F Viñas, and Charles W Smith. 1985. "Large-Amplitude MHD Waves Upstream of the Jovian Bow Shock: Reinterpretation." *Journal of Geophysical Research: Space Physics* 90 (A1): 302–10. <https://doi.org/10.1029/JA090iA01p00302>.
- Goodrich, CC, and JD Scudder. 1984. "The Adiabatic Energy Change of Plasma Electrons and the Frame Dependence of the Cross-Shock Potential at Collisionless Magnetosonic Shock Waves." *Journal of Geophysical Research: Space Physics* 89 (A8): 6654–62. <https://doi.org/10.1029/JA089iA08p06654>.
- Gurram, Harsha, Jan Egedal, and W Daughton. 2021. "Shear Alfvén Waves Driven by Magnetic Reconnection as an Energy Source for the Aurora Borealis." *Geophysical Research Letters* 48 (14): e2021GL094201. <https://doi.org/10.1029/2021GL094201>.
- Hameiri, E, P Laurence, and M Mond. 1991. "The Ballooning Instability in Space Plasmas." *Journal of Geophysical Research: Space Physics* 96 (A2): 1513–26. <https://doi.org/10.1029/90JA02100>.
- Hellinger, Petr, Pavel Trávníček, Bertrand Lembege, and Philippe Savoини. 2007. "Emission of Nonlinear Whistler Waves at the Front of Perpendicular Supercritical Shocks: Hybrid Versus Full Particle Simulations." *Geophysical Research Letters* 34 (14). <https://doi.org/10.1029/2007GL030239>.

- Hill, P, G Paschmann, RA Treumann, W Baumjohann, N Sckopke, and H Lühr. 1995. "Plasma and Magnetic Field Behavior Across the Magnetosheath Near Local Noon." *Journal of Geophysical Research: Space Physics* 100 (A6): 9575–83. <https://doi.org/10.1029/94JA03194>.
- Howes, GG, SD Bale, KG Klein, CHK Chen, CS Salem, and JM TenBarge. 2012. "The Slow-Mode Nature of Compressible Wave Power in Solar Wind Turbulence." *The Astrophysical Journal Letters* 753 (1): L19. <https://doi.org/10.1088/2041-8205/753/1/L19>.
- Hubert, D, C Lacombe, CC Harvey, M Moncuquet, CT Russell, and MF Thomsen. 1998. "Nature, Properties, and Origin of Low-Frequency Waves from an Oblique Shock to the Inner Magnetosheath." *Journal of Geophysical Research: Space Physics* 103 (A11): 26783–98. <https://doi.org/10.1029/GL016i002p00159>.
- Hubert, D, C Perche, CC Harvey, C Lacombe, and CT Russell. 1989. "Observation of Mirror Waves Downstream of a Quasi-Perpendicular Shock." *Geophysical Research Letters* 16 (2): 159–62. <https://doi.org/10.1029/GL016i002p00159>.
- Hudson, PD. 1970. "Discontinuities in an Anisotropic Plasma and Their Identification in the Solar Wind." *Planetary and Space Science* 18 (11): 1611–22. [https://doi.org/10.1016/0032-0633\(70\)90036-X](https://doi.org/10.1016/0032-0633(70)90036-X).
- Hughes, W Jeffrey. 1994. "Magnetospheric ULF Waves: A Tutorial with a Historical Perspective." *Solar Wind Sources of Magnetospheric Ultra-Low-Frequency Waves* 81: 1–11. <https://doi.org/10.1029/GM081p0001>.
- Ji, Hantao, William Daughton, Jonathan Jara-Almonte, Ari Le, Adam Stanier, and Jongsoo Yoo. 2022. "Magnetic Reconnection in the Era of Exascale Computing and Multiscale Experiments." *Nature Reviews Physics* 4 (4): 263–82. <https://doi.org/10.1038/s42254-021-00419-x>.
- Jokipii, Jo R. 1966. "Cosmic-Ray Propagation. I. Charged Particles in a Random Magnetic Field." *Astrophysical Journal, Vol. 146, p. 480* 146: 480. <https://doi.org/10.1086/148912>.
- Karney, CFF, FW Perkins, and Y-C Sun. 1979. "Alfven Resonance Effects on Magnetosonic Modes in Large Tokamaks." *Physical Review Letters* 42 (24): 1621. <https://doi.org/10.1103/PhysRevLett.42.1621>.
- Kennel, CF, JP Edmiston, and Tohru Hada. 1985. "A Quarter Century of Collisionless Shock Research." *Geophysical Monograph Series* 34: 1–36. <https://doi.org/10.1029/GM034p0001>.
- Kennel, Charles F, and HE Petschek. 1966. "Limit on Stably Trapped Particle Fluxes." *Journal of Geophysical Research* 71 (1): 1–28. <https://doi.org/10.1029/JZ071i001p00001>.
- Kepko, Larry, RL McPherron, O Amm, S Apatenkov, Wolfgang Baumjohann, Joachim Birn, M Lester, R Nakamura, Tuija I Pulkkinen, and V Sergeev. 2015. "Substorm Current Wedge Revisited." *Space Science Reviews* 190 (1): 1–46. <https://doi.org/10.1007/s11214-014-0124-9>.
- Kilpua, Emilia, and Hannu Koskinen. 2017. *Introduction to Plasma Physics*. Limes ry.
- Kis, A, M Scholer, B Klecker, H Kucharek, EA Lucek, and H Rème. 2007. "Scattering of Field-Aligned Beam Ions Upstream of Earth's Bow Shock." In *Annales Geophysicae*, 25:785–99. 3. Copernicus GmbH. <https://doi.org/10.5194/angeo-25-785-2007>.
- Kis, A, M Scholer, B Klecker, E Möbius, EA Lucek, H Rème, JM Bosqued, LM Kistler, and H Kucharek. 2004. "Multi-Spacecraft Observations of Diffuse Ions Upstream of Earth's Bow

- Shock.” *Geophysical Research Letters* 31 (20). <https://doi.org/10.1029/2004GL020759>.
- Kivelson, Margaret G, and David J Southwood. 1985. “Resonant ULF Waves: A New Interpretation.” *Geophysical Research Letters* 12 (1): 49–52. <https://doi.org/10.1029/GL012i001p00049>.
- . 1986. “Coupling of Global Magnetospheric MHD Eigenmodes to Field Line Resonances.” *Journal of Geophysical Research: Space Physics* 91 (A4): 4345–51. <https://doi.org/10.1029/JA091iA04p04345>.
- Kucharek, H, E Möbius, M Scholer, C Mouikis, LM Kistler, T Horbury, A Balogh, H Réme, and JM Bosqued. 2004. “On the Origin of Field-Aligned Beams at the Quasi-Perpendicular Bow Shock: Multi-Spacecraft Observations by Cluster.” In *Annales Geophysicae*, 22:2301–8. 7. Copernicus Publications Göttingen, Germany. <https://doi.org/10.5194/angeo-22-2301-2004>.
- Kuznetsov, VD, and AI Osin. 2018. “On the Parallel Shock Waves in Collisionless Plasma with Heat Fluxes.” *Physics Letters A* 382 (31): 2052–54. <https://doi.org/10.1016/j.physleta.2018.05.029>.
- Lacombe, Catherine, Olga Alexandrova, and Lorenzo Matteini. 2017. “Anisotropies of the Magnetic Field Fluctuations at Kinetic Scales in the Solar Wind: Cluster Observations.” *The Astrophysical Journal* 848 (1): 45. <https://doi.org/10.3847/1538-4357/aa8c06>.
- Lacombe, Catherine, and Gérard Belmont. 1995. “Waves in the Earth’s Magnetosheath: Observations and Interpretations.” *Advances in Space Research* 15 (8-9): 329–40. [https://doi.org/10.1016/0273-1177\(94\)00113-F](https://doi.org/10.1016/0273-1177(94)00113-F).
- Lembège, Bertrand, Philippe Savoimi, Petr Hellinger, and Pavel M Trávníček. 2009. “Nonstationarity of a Two-Dimensional Perpendicular Shock: Competing Mechanisms.” *Journal of Geophysical Research: Space Physics* 114 (A3). <https://doi.org/10.1029/2008JA013618>.
- Li, X, DN Baker, H Zhao, K Zhang, AN Jaynes, Q Schiller, SG Kanekal, JB Blake, and M Temerin. 2017. “Radiation Belt Electron Dynamics at Low 1 (< 4): Van Allen Probes Era Versus Previous Two Solar Cycles.” *Journal of Geophysical Research: Space Physics* 122 (5): 5224–34. <https://doi.org/10.1002/2017JA023924>.
- Livadiotis, G, and DJ McComas. 2013. “Understanding Kappa Distributions: A Toolbox for Space Science and Astrophysics.” *Space Science Reviews* 175 (1): 183–214. <https://doi.org/10.1007/s11214-013-9982-9>.
- Lu, Xi, Hui Zhang, Terry Liu, Andrew Vu, Craig Pollock, and Boyi Wang. 2022. “Statistical Study of Foreshock Density Holes.” *Journal of Geophysical Research: Space Physics* 127 (4): e2021JA029981. <https://doi.org/10.1029/2021JA029981>.
- Lucek, EA, TS Horbury, MW Dunlop, PJ Cargill, SJ Schwartz, A Balogh, P Brown, C Carr, K-H Fornacon, and E Georgescu. 2002. “Cluster Magnetic Field Observations at a Quasi-Parallel Bow Shock.” In *Annales Geophysicae*, 20:1699–1710. 11. Copernicus Publications Göttingen, Germany. <https://doi.org/10.5194/angeo-20-1699-2002>.
- Merkin, VG, and JG Lyon. 2010. “Effects of the Low-Latitude Ionospheric Boundary Condition on the Global Magnetosphere.” *Journal of Geophysical Research: Space Physics* 115 (A10). <https://doi.org/10.1029/2010JA015461>.
- Mertsch, Philipp. 2020. “Test Particle Simulations of Cosmic Rays.” *Astrophysics and Space Science* 365 (8): 135. <https://doi.org/10.1007/s10509-020-03832-3>.
- Meziane, K, C Mazelle, M Wilber, D LeQuéau, JP Eastwood, H Reme, I Dandouras, et al. 2004.

- “Bow Shock Specularly Reflected Ions in the Presence of Low-Frequency Electromagnetic Waves: A Case Study.” In *Annales Geophysicae*, 22:2325–35. 7. Copernicus GmbH. <https://doi.org/10.5194/angeo-22-2325-2004>.
- Mignone, Andrea, H Haudemand, and E Puzzoni. 2023. “A Guiding Center Implementation for Relativistic Particle Dynamics in the PLUTO Code.” *Computer Physics Communications* 285: 108625. <https://doi.org/10.1016/j.cpc.2022.108625>.
- Nakamura, Takuma KM, KJ Genestreti, Y-H Liu, R Nakamura, W-L Teh, Hiroshi Hasegawa, W Daughton, et al. 2018. “Measurement of the Magnetic Reconnection Rate in the Earth’s Magnetotail.” *Journal of Geophysical Research: Space Physics* 123 (11): 9150–68. <https://doi.org/10.1029/2018JA025713>.
- Northrop, Theodore G. 1963. “Adiabatic Charged-Particle Motion.” *Reviews of Geophysics* 1 (3): 283–304. <https://doi.org/10.1029/RG001i003p00283>.
- Northrop, Theodore G, and Edward Teller. 1960. “Stability of the Adiabatic Motion of Charged Particles in the Earth’s Field.” *Physical Review* 117 (1): 215. <https://doi.org/10.1103/PhysRev.117.215>.
- Olson, John V, and LC Lee. 1983. “Pc1 Wave Generation by Sudden Impulses.” *Planetary and Space Science* 31 (3): 295–302. [https://doi.org/10.1016/0032-0633\(83\)90079-X](https://doi.org/10.1016/0032-0633(83)90079-X).
- Parks, GK, E Lee, F Mozer, M Wilber, E Lucek, I Dandouras, H Rème, et al. 2006. “Larmor Radius Size Density Holes Discovered in the Solar Wind Upstream of Earth’s Bow Shock.” *Physics of Plasmas* 13 (5): 050701. <https://doi.org/10.1063/1.2201056>.
- Phan, TD, DE Larson, RP Lin, JP McFadden, KA Anderson, CW Carlson, RE Ergun, et al. 1996. “The Subsolar Magnetosheath and Magnetopause for High Solar Wind Ram Pressure: WIND Observations.” *Geophysical Research Letters* 23 (10): 1279–82. <https://doi.org/10.1029/96GL00845>.
- Radoski, Henry R, and Robert L Carovillano. 1966. “Axisymmetric Plasmasphere Resonances: Toroidal Mode.” *The Physics of Fluids* 9 (2): 285–91. <https://doi.org/10.1063/1.1761671>.
- Raeder, J, RJ Walker, and M Ashour-Abdalla. 1995. “The Structure of the Distant Geomagnetic Tail During Long Periods of Northward IMF.” *Geophysical Research Letters* 22 (4): 349–52. <https://doi.org/10.1029/94GL03380>.
- Rahman, Sk Mashfiqur, and Omer San. 2019. “A Relaxation Filtering Approach for Two-Dimensional Rayleigh–Taylor Instability-Induced Flows.” *Fluids* 4 (2): 78.
- Rambo, PW. 1995. “Finite-Grid Instability in Quasineutral Hybrid Simulations.” *Journal of Computational Physics* 118 (1): 152–58. <https://doi.org/10.1006/jcph.1995.1086>.
- Samson, JC, BG Harrold, JM Ruohoniemi, RA Greenwald, and ADM Walker. 1992. “Field Line Resonances Associated with MHD Waveguides in the Magnetosphere.” *Geophysical Research Letters* 19 (5): 441–44. <https://doi.org/10.1029/92GL00116>.
- Scarf, FL, JH Wolfe, and RW Silva. 1967. “A Plasma Instability Associated with Thermal Anisotropies in the Solar Wind.” *Journal of Geophysical Research* 72 (3): 993–1005. <https://doi.org/10.1029/JZ072i003p00993>.
- Schekochihin, Alexander A. 2022. “MHD Turbulence: A Biased Review.” *Journal of Plasma Physics* 88 (5): 155880501. <https://doi.org/10.1017/S0022377822000721>.
- Scholer, Manfred, Iku Shinohara, and Shuichi Matsukiyo. 2003. “Quasi-Perpendicular Shocks: Length Scale of the Cross-Shock Potential, Shock Reformation, and Implication for Shock

- Surfing.” *Journal of Geophysical Research: Space Physics* 108 (A1): SSH-4. <https://doi.org/10.1029/2002JA009515>.
- Schwartz, J., Steven, and David Burgess. 1991. “Quasi-Parallel Shocks: A Patchwork of Three-Dimensional Structures.” *Journal of Geophysical Research: Space Physics*. <https://doi.org/10.1029/91GL00138>.
- Schwartz, Steven J, Michelle F Thomsen, and John T Gosling. 1983. “Ions Upstream of the Earth’s Bow Shock: A Theoretical Comparison of Alternative Source Populations.” *Journal of Geophysical Research: Space Physics* 88 (A3): 2039–47. <https://doi.org/10.1029/JA088iA03p02039>.
- Sckopke, N, G Paschmann, SJ Bame, JT Gosling, and CT Russell. 1983. “Evolution of Ion Distributions Across the Nearly Perpendicular Bow Shock: Specularly and Non-Specularly Reflected-Gyrating Ions.” *Journal of Geophysical Research: Space Physics* 88 (A8): 6121–36. <https://doi.org/10.1029/JA088iA08p06121>.
- Shi, Q. Q., A. M. Tian, S. C. Bai, H. Hasegawa, A. W. Degeling, Z. Y. Pu, M. Dunlop, et al. 2019. “Dimensionality, Coordinate System and Reference Frame for Analysis of In-Situ Space Plasma and Field Data.” *Space Science Reviews* 215 (4): 35. <https://doi.org/10.1007/s11214-019-0601-2>.
- Shue, J-H, JK Chao, HC Fu, CT Russell, P Song, KK Khurana, and HJ Singer. 1997. “A New Functional Form to Study the Solar Wind Control of the Magnetopause Size and Shape.” *Journal of Geophysical Research: Space Physics* 102 (A5): 9497–9511. <https://doi.org/10.1029/97JA00196>.
- Slavin, James A, and Robert E Holzer. 1981. “Solar Wind Flow about the Terrestrial Planets 1. Modeling Bow Shock Position and Shape.” *Journal of Geophysical Research: Space Physics* 86 (A13): 11401–18. <https://doi.org/10.1029/JA086iA13p11401>.
- Song, P, CT Russell, and SP Gary. 1994. “Identification of Low-Frequency Fluctuations in the Terrestrial Magnetosheath.” *Journal of Geophysical Research: Space Physics* 99 (A4): 6011–25. <https://doi.org/10.1029/93JA03300>.
- Sonnerup, B. U. Ö. 1979. “Space Plasma Physics: The Study of Solar-System Plasmas.”
- Sonnerup, BU Ö. 1974. “Magnetopause Reconnection Rate.” *Journal of Geophysical Research* 79 (10): 1546–49. <https://doi.org/10.1029/JA079i010p01546>.
- Southwood, David J, and Margaret G Kivelson. 1981. “Charged Particle Behavior in Low-Frequency Geomagnetic Pulsations 1. Transverse Waves.” *Journal of Geophysical Research: Space Physics* 86 (A7): 5643–55. <https://doi.org/10.1029/JA086iA07p05643>.
- . 1993. “Mirror Instability: 1. Physical Mechanism of Linear Instability.” *Journal of Geophysical Research: Space Physics* 98 (A6): 9181–87. <https://doi.org/10.1029/92JA02837>.
- Southwood, DJ, and WJ Hughes. 1983. “Theory of Hydromagnetic Waves in the Magnetosphere.” *Space Science Reviews* 35 (4): 301–66. <https://doi.org/10.1007/BF00169231>.
- Stenzel, RL. 1999. “Whistler Waves in Space and Laboratory Plasmas.” *Journal of Geophysical Research: Space Physics* 104 (A7): 14379–95. <https://doi.org/10.1029/1998JA900120>.
- Stone, James M, Thomas A Gardiner, Peter Teuben, John F Hawley, and Jacob B Simon. 2008. “Athena: A New Code for Astrophysical MHD.” *The Astrophysical Journal Supplement Series* 178 (1): 137. <https://doi.org/10.1086/588755>.
- Strumik, M, V Roytershteyn, H Karimabadi, K Stasiewicz, M Grzesiak, and D Przepiórka.

2015. "Identification of the Dominant ULF Wave Mode and Generation Mechanism for Obliquely Propagating Waves in the Earth's Foreshock." *Geophysical Research Letters* 42 (13): 5109–16. <https://doi.org/10.1002/2015gl064915>.
- Swanson, Donald Gary. 2012. *Plasma Waves*. Elsevier.
- Tautz, RC, A Dosch, F Effenberger, H Fichtner, and A Kopp. 2013. "Pitch-Angle Scattering in Magnetostatic Turbulence-i. Test-Particle Simulations and the Validity of Analytical Results." *Astronomy & Astrophysics* 558: A147. <https://doi.org/10.1051/0004-6361/201322142>.
- Tenerani, Anna, Carlos González, Nikos Sioulas, Chen Shi, and Marco Velli. 2023. "Dispersive and Kinetic Effects on Kinked Alfvén Wave Packets: A Comparative Study with Fluid and Hybrid Models." *Physics of Plasmas* 30 (3): 032101. <https://doi.org/10.1063/5.0134726>.
- Tóth, Gábor. 2000. "The $\cdot b = 0$ Constraint in Shock-Capturing Magnetohydrodynamics Codes." *Journal of Computational Physics* 161 (2): 605–52. <https://doi.org/https://doi.org/10.1006/jcph.2000.6519>.
- Tóth, Gábor, Yuxi Chen, Tamas I Gombosi, Paul Cassak, Stefano Markidis, and Ivy Bo Peng. 2017. "Scaling the Ion Inertial Length and Its Implications for Modeling Reconnection in Global Simulations." *Journal of Geophysical Research: Space Physics* 122 (10): 10–336. <https://doi.org/10.1002/2017JA024189>.
- Tracy, Eugene Raymond, Alain Jean Brizard, AS Richardson, and AN Kaufman. 2014. *Ray Tracing and Beyond: Phase Space Methods in Plasma Wave Theory*. Cambridge University Press.
- Tsyganenko, NA. 2013. "Data-Based Modelling of the Earth's Dynamic Magnetosphere: A Review." In *Annales Geophysicae*, 31:1745–72. 10. Copernicus Publications Göttingen, Germany. <https://doi.org/10.5194/angeo-31-1745-2013>.
- Turc, L, OW Roberts, D Verscharen, AP Dimmock, P Kajdič, M Palmroth, Y Pfau-Kempf, et al. 2022. "Transmission of Foreshock Waves Through Earth's Bow Shock." *Nature Physics*, 1–9. <https://doi.org/10.1038/s41567-022-01837-z>.
- Umeda, Takayuki, and Ryo Yamazaki. 2006. "Full Particle Simulation of a Perpendicular Collisionless Shock: A Shock-Rest-Frame Model." *Earth, Planets and Space* 58 (10): e41–44. <https://doi.org/10.1186/BF03352617>.
- Usanova, Maria E, and Ian R Mann. 2016. "Understanding the Role of EMIC Waves in Radiation Belt and Ring Current Dynamics: Recent Advances." *Waves, Particles, and Storms in Geospace: A Complex Interplay*, 244–76. <https://doi.org/10.1093/acprof:oso/9780198705246.003.0011>.
- Vandervoort, Peter O. 1960. "The Relativistic Motion of a Charged Particle in an Inhomogeneous Electromagnetic Field." *Annals of Physics* 10 (3): 401–53. [https://doi.org/10.1016/0003-4916\(60\)90004-X](https://doi.org/10.1016/0003-4916(60)90004-X).
- Verscharen, Daniel, Christopher HK Chen, and Robert T Wicks. 2017. "On Kinetic Slow Modes, Fluid Slow Modes, and Pressure-Balanced Structures in the Solar Wind." *The Astrophysical Journal* 840 (2): 106. <https://doi.org/10.3847/1538-4357/aa6a56>.
- Walker, ADM, Jo M Ruohoniemi, KB Baker, RA Greenwald, and JC Samson. 1992. "Spatial and Temporal Behavior of ULF Pulsations Observed by the Goose Bay HF Radar." *Journal of Geophysical Research: Space Physics* 97 (A8): 12187–202. <https://doi.org/10.1029/1992JA00297>

92JA00329.

- Wang, Jing, Jiang Yu, Aojun Ren, Zuzheng Chen, Xiaojun Xu, Jun Cui, and Jinbin Cao. 2023. “In Situ Observations of Whistler-Mode Waves in Magnetic Reconnection at Mars.” *The Astrophysical Journal* 944 (1): 85. <https://doi.org/10.3847/1538-4357/acb152>.
- Wang, Liang, Ammar H Hakim, Jonathan Ng, Chuanfei Dong, and Kai Germaschewski. 2020. “Exact and Locally Implicit Source Term Solvers for Multifluid-Maxwell Systems.” *Journal of Computational Physics* 415: 109510. <https://doi.org/10.1016/j.jcp.2020.109510>.
- Wilson III, Lynn B. 2016. “Low Frequency Waves at and Upstream of Collisionless Shocks.” *Low-Frequency Waves in Space Plasmas*, 269–91. <https://doi.org/10.1002/9781119055006.ch16>.
- Winske, Dan, Homa Karimabadi, Ari Yitzchak Le, Nojan Nick Omidi, Vadim Roytershteyn, and Adam John Stanier. 2023. “Hybrid-Kinetic Approach: Massless Electrons.” In *Space and Astrophysical Plasma Simulation: Methods, Algorithms, and Applications*, 63–91. Springer. https://doi.org/10.1007/978-3-031-11870-8_3.
- Winske, D, and KB Quest. 1988. “Magnetic Field and Density Fluctuations at Perpendicular Supercritical Collisionless Shocks.” *Journal of Geophysical Research: Space Physics* 93 (A9): 9681–93. <https://doi.org/10.1029/JA093iA09p09681>.
- Woltjer, Lodewijk. 1958. “A Theorem on Force-Free Magnetic Fields.” *Proceedings of the National Academy of Sciences* 44 (6): 489–91. <https://doi.org/10.1073/pnas.44.6.489>.
- Wright, Andrew N. 1994. “Dispersion and Wave Coupling in Inhomogeneous MHD Waveguides.” *Journal of Geophysical Research: Space Physics* 99 (A1): 159–67. <https://doi.org/10.1029/93JA02206>.
- Wu, CS, D Winske, YM Zhou, ST Tsai, P Rodriguez, M Tanaka, K Papadopoulos, et al. 1984. “Microinstabilities Associated with a High Mach Number, Perpendicular Bow Shock.” *Space Science Reviews* 37: 63–109. <https://doi.org/10.1007/BF00213958>.
- Zhang, Hui, Qiugang Zong, Hyunju Connor, Peter Delamere, Gábor Facskó, Desheng Han, Hiroshi Hasegawa, et al. 2022. “Dayside Transient Phenomena and Their Impact on the Magnetosphere and Ionosphere.” *Space Science Reviews* 218 (5): 40. <https://doi.org/10.1007/s11214-021-00865-0>.
- Zhao, L-L, GP Zank, L Adhikari, Q Hu, JC Kasper, SD Bale, KE Korreck, et al. 2020. “Identification of Magnetic Flux Ropes from Parker Solar Probe Observations During the First Encounter.” *The Astrophysical Journal Supplement Series* 246 (2): 26. <https://doi.org/10.3847/1538-4365/ab4ff1>.
- Zhou, Hongyang, Gabor Toth, Xianzhe Jia, and Yuxi Chen. 2020. “Reconnection-Driven Dynamics at Ganymede’s Upstream Magnetosphere: 3-d Global Hall MHD and MHD-EPIC Simulations.” *Journal of Geophysical Research: Space Physics* 125 (8): e2020JA028162. <https://doi.org/10.1029/2020JA028162>.
- Zong, Qiugang, Robert Rankin, and Xuzhi Zhou. 2017. “The Interaction of Ultra-Low-Frequency Pc3-5 Waves with Charged Particles in Earth’s Magnetosphere.” *Reviews of Modern Plasma Physics* 1: 1–90. <https://doi.org/10.1007/s41614-017-0011-4>.