

Paulo J.V. Garcia
João M. Ferreira
Editors

LECTURE NOTES IN PHYSICS 793

Jets from Young Stars IV

From Models to Observations and Experiments



Springer



Lecture Notes in Physics

Founding Editors: W. Beiglöck, J. Ehlers, K. Hepp, H. Weidenmüller

Editorial Board

R. Beig, Vienna, Austria
W. Beiglöck, Heidelberg, Germany
W. Domcke, Garching, Germany
B.-G. Englert, Singapore
U. Frisch, Nice, France
F. Guinea, Madrid, Spain
P. Hänggi, Augsburg, Germany
W. Hillebrandt, Garching, Germany
R. L. Jaffe, Cambridge, MA, USA
W. Janke, Leipzig, Germany
H. v. Löhneysen, Karlsruhe, Germany
M. Mangano, Geneva, Switzerland
J.-M. Raimond, Paris, France
D. Sornette, Zurich, Switzerland
S. Theisen, Potsdam, Germany
D. Vollhardt, Augsburg, Germany
W. Weise, Garching, Germany
J. Zittartz, Köln, Germany

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching – quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research and to serve three purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas
- to be a source of advanced teaching material for specialized seminars, courses and schools

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive being available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg / Germany
christian.caron@springer.com

P.J.V. Garcia
J.M. Ferreira (Eds.)

Jets from Young Stars IV

From Models to Observations and Experiments



Paulo J. V. Garcia
Universidade do Porto
Faculdade de Engenharia
SIM Unidade FCT 4006
Rua Dr. Roberto Frias, s/n
P-4200-465 Porto
Portugal
Laboratoire d'Astrophysique
Observatoire de Grenoble
BP 53
F-38041 Grenoble Cédex 9
France

João Miguel Ferreira
Universidade dos Açores
Angra do Heroísmo,
Açores, Portugal
Centro de Astrofísica
Rua das Estrelas, s/n
P-4150-762 Porto
Portugal

Garcia P.J.V., Ferreira J.M. (Eds.), *Jets from Young Stars IV: From Models to Observations and Experiments*, Lect. Notes Phys. 793 (Springer, Berlin Heidelberg 2010),
DOI 10.1007/978-3-642-02289-0

Lecture Notes in Physics ISSN 0075-8450 e-ISSN 1616-6361
ISBN 978-3-642-02288-3 e-ISBN 978-3-642-02289-0
DOI 10.1007/978-3-642-02289-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009937606

© Springer-Verlag Berlin Heidelberg 2010

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: Integra Software Services Pvt. Ltd., Pondicherry

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The way a star-disk system launches and collimates jets is one of the yet unsolved problems in contemporary astrophysics. This problem is relevant in active galactic nuclei, micro-quasars, gamma-ray bursts, planetary nebulae and young stars. Although there is agreement in that magnetic fields mediate the process, the exact details still remain elusive.

The larger angular size of the jet engines in young stars makes them important laboratories for direct observational model testing. In parallel, the recent advent of high-energy density facilities has added experimental control to jet studies.

Testing models for jet launching and collimation requires a high degree of interdisciplinary and sophistication. Indeed, closing the circle between pure MHD, thermo-chemical evolution, high angular resolution spectro-imaging and laboratory experiments is not trivial. This volume of *Lecture Notes in Physics* aims at bridging these gaps by providing a series of lectures bridging the foundations of the discipline.

The first two lectures by Rob Coker and Andrea Ciardi address studies of jets physics in the laboratory. Then the lecture by Nektarios Vlahakis introduces the magnetohydrodynamic theory of stationary jets. The heating mechanisms in these magnetohydrodynamic jets are largely unknown and the lecture on the coronal heating in the Sun presented by Alan Hood addresses physical aspects very relevant for jets. Another heating mechanism at work is shocks propagating along the jet. The lecture by David Flower addresses them with an emphasis on molecular material and chemistry. An atomic perspective on the microphysics of the shocked material in the jets is presented by Alex Raga. This book closes with the presentation of series of diagnostics allowing to recover basic physical quantities from jet emission lines by Catherine Dougados, Francesca Bacciotti, Sylvie Cabrit and Brunella Nisini.

The editors acknowledge financial support from the JETSET Marie Curie Research Training Network, the Universidade dos Açores, the Direção Regional da Ciência, Tecnologia e Comunicações dos Açores, the Luso-American Foundation, the British Council and the Fundação para a Ciência e Tecnologia.

The editors would like to thank all the lecturers for their excellent presentations and contributions to this book. We thank the referees, whose suggestions increased the clarity and accuracy of the contributions. We are also thankful to all school

participants. We warmly thank the work by Elsa Silva, Júlio Carreira, Manuel Monteiro, Eileen Flood, Samira Rajabi and Despoina Panoglou which guaranteed a smooth and efficient school.

Porto, Portugal
July, 2009

Paulo J.V. Garcia
João M. Ferreira

Contents

Laboratory Astrophysics and Scaling	1
Rob Coker	
1 Introduction	1
2 The Concept of Scaling	2
3 Experimental Facilities	15
4 Detailed Example and a Challenge	18
5 Summary	28
References	28
 Laboratory Studies of Astrophysical Jets	31
Andrea Ciardi	
1 Introduction	31
2 Plasma Conditions in z-Pinch and Laser Experiments	32
3 Relating Laboratory and Astrophysical Phenomena	33
4 Young Stellar Jets from z-Pinch Machines	35
5 Summary	47
References	48
 Output from MHD Models	51
Nektarios Vlahakis	
1 Steady, Axisymmetric, Nonrelativistic, Magnetized Outflows	52
2 Relativistic MHD	77
3 Conclusion	104
Appendix: Alfvén and Magnetosonic MHD Waves	105
References	107
 Coronal Heating	109
Alan William Hood	
1 The Sun	109
2 The Energy Budget	115
3 Observed Coronal Phenomena	117
4 Alternative Heating Observations	120

5	MHD Equations	122
6	Wave Heating Mechanisms	124
7	Magnetic Reconnection Theory: Two Dimensions	133
8	Magnetic Reconnection Theory: Three Dimensions	144
9	Applications to the Solar Corona	149
10	Plasma Response to Nanoflare Heating	156
11	Summary	157
	References	158
Flows in Molecular Media		161
David Flower		
1	Introduction	161
2	The MHD Conservation Equations	163
3	The Structure of Interstellar Shock Waves	172
4	Shock Waves in Dark Clouds	183
5	Shock Waves in the Presence of an Ultraviolet Radiation Field	187
6	Diagnostics of Shock Waves: The H ₂ Excitation Diagram	190
	References	194
The Ionisation and Excitation State of Stellar Outflows		195
A.C. Raga		
1	Introduction	195
2	The Gas dynamic Equations	196
3	The Ionisation State of the Gas	197
4	The Cooling Function	202
5	Recipes for Calculating the Cooling Function	206
6	Calculation of the Emitted Spectrum	208
7	Summary	211
	References	211
Deriving Physical Diagnostics from Observations		213
C. Dougados, F. Bacciotti, S. Cabrit, and B. Nisini		
1	Introduction	213
2	Model-Independent Diagnostics of Plasma Excitation Conditions	214
3	Deriving Excitation Conditions from Shock Diagnostics	230
4	Derivation of Mass-Loss Rates	236
5	Projection and Convolution Effects	240
	References	248
Index		251

Laboratory Astrophysics and Scaling

Rob Coker

Abstract This chapter is a summary of lectures intended to introduce the concept of scaling to graduate students. Specifically, we address whether astrophysical phenomena may be addressed using ‘scaled’ versions of the same processes in experiments in the laboratory. Recent (and ongoing) advances in laser facility capabilities, such as higher temperatures and Mach numbers, are making such scaled hydrodynamical and even magneto-hydrodynamical experiments possible. The general concept of scaling, relevant equations, and some applications are presented.

1 Introduction

Astrophysical data sets are inherently limited, be it temporally, spatially, or spectrographically. We are also limited to observations of 2D projections on the sky. In order to answer the increasingly complex questions raised by observations of complex astrophysical objects, we require repeatable and diagnosable experiments. Although the laboratory has long been used to assist astrophysics (e.g., the synchrotron), the concept of doing experiments that are ‘scaled-model’ versions of astrophysical objects is fairly new; at least, the *ability* to do so is new, due to advances in facilities that conduct high-energy density physics (HEDP) experiments.

In these lectures I will discuss scaling, first using the example of self-similarity. Then I will develop the concept for generic hydrodynamic (HD) and magneto-hydrodynamic (MHD) flows. This will lead to the idea of dimensionless parameters. Next I will briefly present the modern HEDP facilities and some of their diagnostics which permit scaled laboratory astrophysics. Finally, I will present some ongoing applications of scaling.

R. Coker (✉)

Los Alamos National Laboratory, Los Alamos, NM 87545, USA, robcc@lanl.gov

2 The Concept of Scaling

The premise behind scaling is to learn something about an astrophysical object using a laboratory experiment that represents the important properties of the object. An example of a ‘scaled experiment’ that I will discuss later is shown in Fig. 1. In this experiment, the Omega laser facility at the University of Rochester in New York is used to produce a jet intended to represent the astrophysical object HH110/270, a young stellar object (YSO) that is interesting because it is the only known instance of a jet from a young star being clearly deflected by a giant molecular cloud (GMC). The jets program discussed here is presently part of the National Laser Users’ Facility (NLUF) program, with P. Hartigan as the PI of the international collaboration. If things are ‘similar enough,’ then the laboratory experiment will behave in the ‘same way’ as the astrophysical object. The difficulty is defining ‘similar enough’ and determining which aspects of the two systems will behave in the same fashion and to what degree. Clearly, to capture the behavior of an astrophysical object, the experiment must be physically scaled down. However, scale does not only refer to reduced spatial scale (such as in a toy scale model), since astrophysical objects are generally not static. In fact, the term scaling refers to mapping of any characteristic of one system to another (e.g., velocity or Mach number) and not to just spatial or temporal scaling.

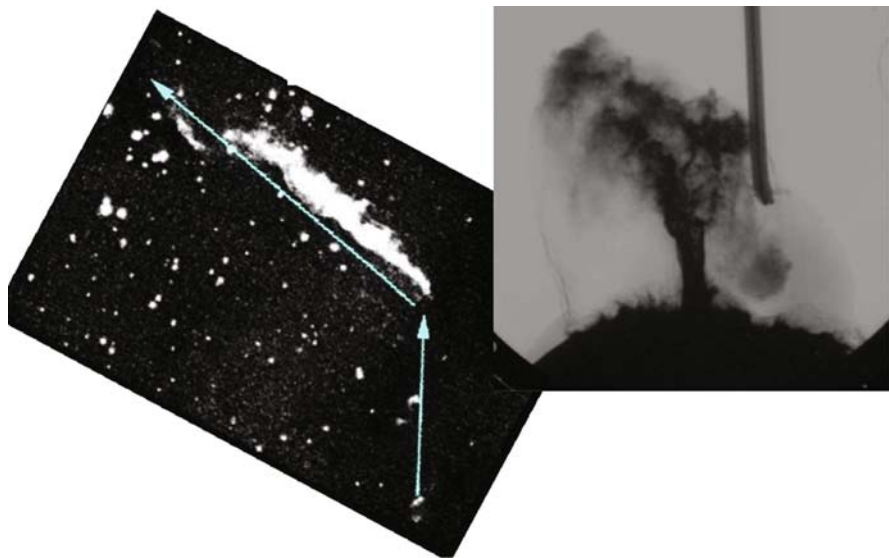


Fig. 1 What can be learned about the pc-scale HH110/270 (*left*, [1]) from an mm-scale Omega experiment (*right*, [2])?

2.1 Self-Similarity

Problems that are ‘self-similar’ are examples of scalable problems. In a self-similar problem, the solution does not depend explicitly on all variables but rather only on some combination of them. Here, I will use dimensional analysis to show solution dependencies; although dimensional analysis fails on some complex types of self-similar problems (see, e.g., [3]), I will not discuss them here. Three examples of self-similarity are filling a container with liquid, a piston-induced rarefaction, and the Sedov blast wave. These problems all scale perfectly such that an experiment on one ‘scale’ (temporal, spatial, etc.) will give the same results as an experiment (or astrophysical object) on some entirely different ‘scale.’

2.1.1 Filling a Container

Nearly a hundred years ago, experiments were done that measured the time τ it took to fill a bowl of volume V with a pipe having a pressure drop P , using various liquids with density ρ and viscosity μ [4]. A sketch of the experiments is shown in Fig. 2; the details of the pipe do not matter as long as it permits the flow to be fully turbulent. Plots of $P = f(\tau, V, \mu, \rho)$ versus τ show separate curves for the different fluids (see the left plot in Fig. 3). Examination of the variables in the problem shows

$$[P] = \frac{M}{LT^2}, \quad [\tau] = T, \quad [V] = L^3, \quad [\mu] = \frac{M}{LT}, \quad [\rho] = \frac{M}{L^3}, \quad (1)$$

where the square brackets are shorthand for ‘has the units of’ and M , L , and T are units of mass, length, and time, respectively. It can be seen that two dimensionless parameters can be created from the five variables:

$$\Pi_1 = \frac{P\tau}{\mu}, \quad \Pi_2 = \frac{\rho V^{2/3}}{\mu\tau}. \quad (2)$$

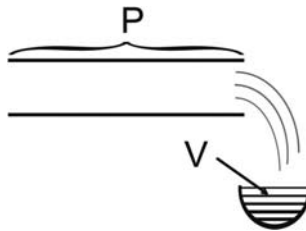


Fig. 2 A sketch of the simple flow experiments from [4], with a fluid from a pipe with a pressure drop P filling a bowl of volume V in time τ

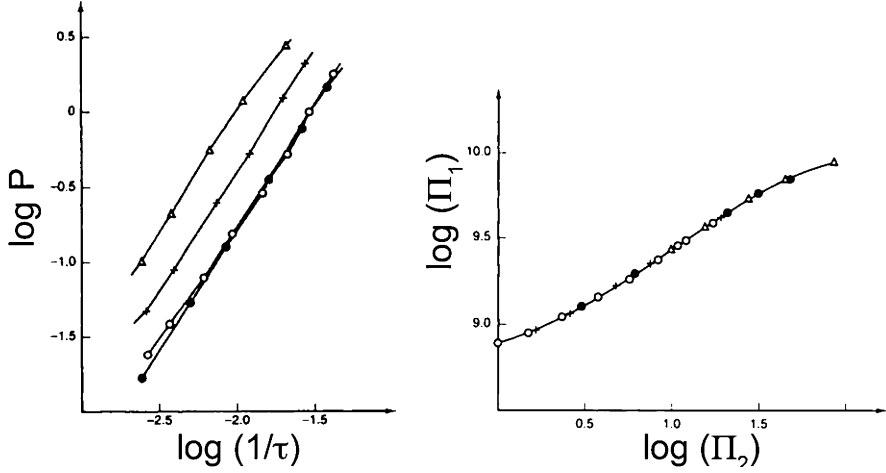


Fig. 3 Dimensional (*left*) and nondimensional (*right*) experimental results from [4] for water (*open circles*), chloroform (*closed circles*), bromoform (+), and mercury (Δ). Figures taken from [6]

A plot of Π_1 versus Π_2 results in a single curve for all the data [5], as shown in the right plot of Fig. 3. Therefore, the physical behavior of the fluid in this regime is independent of parameters such as composition of the liquid, temperature, or other considerations. Thus, if one wanted to do an experiment with a fluid that was dangerous or impractical (e.g., benzene), one could replace it with, say, water. This will result in a ‘scaled’ experiment that illustrates how benzene will behave without ever having to experiment with it. This illustrates the power of scaling.

2.1.2 A Rarefaction

A more idealized problem is that of a piston-induced rarefaction. The experiment starts with a uniform ideal gas in a tube with a piston at one end, which is withdrawn isentropically at some speed V that is less than the speed of sound, c_0 , in the gas (or, more strictly, that $V \leq 2c_0/(\gamma - 1)$, to avoid producing a vacuum behind the piston). The piston motion produces a rarefaction, a moving region of reduced pressure akin to the stretched portion of a spring. Ignoring the region where the piston accelerates from 0 to V , the solution for the resulting speed of the gas is [7]

$$u = 0, \quad \frac{2(c_0 - x/t)}{\gamma + 1}, \quad V, \quad (3)$$

where γ is the adiabatic index of the gas. Figure 4 shows a representation of this solution, where the three regions of the solution are evident: $u = 0$ for $x < tc_0 - \frac{\gamma+1}{2}Vt$ and $u = V$ for $x > tc_0$, where x and t are location and time, respectively. Notice there are no scales; the solution has the same shape for all time, deviating from a constant only between $x/t = c_0$ and $x/t = c_0 - \frac{\gamma+1}{2}V$. That is, the solution

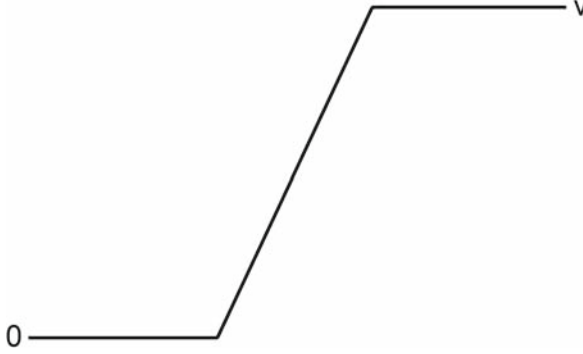


Fig. 4 Illustration of the solution for gas motion as a function of position in a piston-induced rarefaction problem. The front moves with the speed of sound, c_0 , while the back moves more slowly and may even be stationary

scales and looks like itself, so if the solution is known for any point in time, it is known for all time. Again, dimensional analysis also shows this. There are only four dimensional quantities in the problem, x , t , V , and c_0 . Since $[V] = [c_0] = L/T = [x]/[t]$, the solution is not going to explicitly depend on x and t but rather only on x/t .

2.1.3 Sedov Blast Wave

The Sedov blast wave problem [8] has some energy E delivered instantaneously to an infinitesimal point in an ideal gas with adiabatic index γ , density ρ_0 , and negligible pressure P_0 (there is a more general Sedov problem where the medium has a power-law density profile but I will not discuss that here). Since there is no explicit spatial or temporal scale or even a velocity scale, it is likely from just the problem description that the Sedov problem is self-similar.

Since the result of the Sedov problem is a strong spherical shock, the Hugoniot jump conditions apply and thus

$$\rho = \rho_0 \frac{\gamma + 1}{\gamma - 1}, \quad P = \frac{2}{\gamma + 1} \rho_0 V_{\text{sh}}^2, \quad u = \frac{2}{\gamma + 1} V_{\text{sh}}, \quad (4)$$

where ρ , P , and u are the density, pressure, and velocity, respectively, just behind the shock, and V_{sh} is the velocity of the shock. Dimensional analysis gives $[R_{\text{sh}}]^5 = [t]^2 [E]/[\rho_0]$. Since $[V_{\text{sh}}] = [R_{\text{sh}}]/[t]$, $V_{\text{sh}} = f\left(\sqrt{\frac{E}{\rho_0 R_{\text{sh}}^3}}\right)$ and one can then solve for ρ , P , and u behind the shock in terms of r/R_{sh} . This sort of dimensional analysis is only good up to a constant, but here it happens to be ~ 1 . Note that the Mach number will not be involved since $M \approx V_{\text{sh}} \sqrt{\rho_0/P_0} \rightarrow \infty$ for a Sedov problem.

Consider that $R_{\text{sh}} \sim (E/\rho_0)^{1/5} t^{2/5}$. Taking the log and rearranging gives

$$\frac{5}{2} \log R_{\text{sh}} = \frac{1}{2} \log \frac{E}{\rho_0} + \log t. \quad (5)$$

A plot of $\log t$ versus $5/2 \log R_{\text{sh}}$ thus gives a straight line whose intercept can then be used to find E/ρ_0 . This has been done, for example, for blasts of TNT to determine their explosive energy [9].

The solution of the Sedov problem far behind the shock is that $P(r)/P \rightarrow \text{constant}$, $u(r)/u \rightarrow r/R_{\text{sh}}$, and $\rho(r)/\rho \rightarrow$ a steep function of r/R_{sh} depending on γ . These curves are valid for all time and are illustrated in Fig. 5. Within the limits of the assumptions of the problem (δ -function source, uniform medium, small P_0 , and ideal gas), a ‘scaled’ experiment can mimic a problem with some E by adjusting ρ_0 (so that E/ρ_0 remains the same) and/or the temporal and spatial scales (so that $(R_{\text{sh}}/t)^{5/2}$ remains the same). An obvious possible application of this is to supernovae.

2.2 More General Scaling

A more common scaling approach is to look for dimensionless quantities that can be used to describe a specific problem (although note this does not work for the Sedov problem). Experimental approaches can then be created to represent these scalable properties.

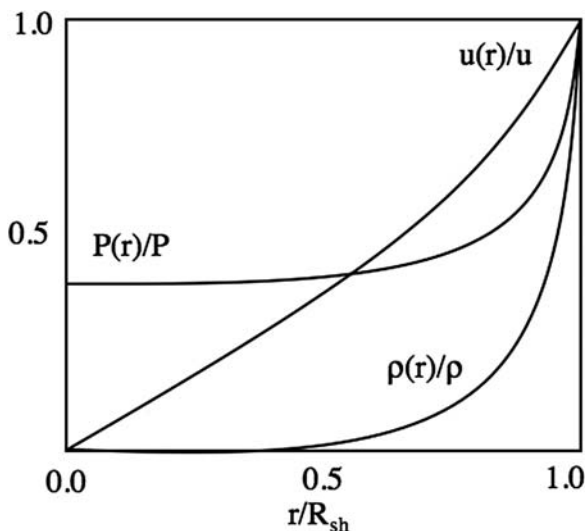


Fig. 5 Illustration of the solution for the uniform Sedov blastwave problem

2.2.1 Hydrodynamical Scaling

In the case of a problem described by the hydrodynamic equations, the dimensionless quantity that likely comes to mind is the Mach number. The polytropic hydrodynamic equations describing the evolution of momentum, mass, and energy, are, respectively [10],

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) + \nabla p = 0, \quad (6)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (7)$$

$$\frac{\partial p}{\partial t} + \gamma p \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla p = 0, \quad (8)$$

where ρ is the density, \mathbf{v} is the velocity, γ is the adiabatic index, and p is the gas pressure. There are no energy source or sinks so the flow is adiabatic ($ds/dt = 0$ where s is the entropy and d/dt is a total derivative), but the flow is also reversible, thus isentropic ($s = \text{constant} = c_v \log(p/\rho^\gamma)$, where c_v is the specific heat at constant volume).

The hydrodynamic equations have five dimensional variables: \mathbf{x} (through ∇), t , ρ , \mathbf{v} , and p . Dimensional analysis gives

$$[v] = [x]/[t], \quad [p] = [\rho][x^2]/[t^2]. \quad (9)$$

This means that after scaling \mathbf{x} and t by the same amount, or ρ and p by the same amount, the equations will give the same solution. This can also be seen by noting that multiplying ρ and p by the same constant in Eqs. (6), (7), and (8) does not change the result (the same goes for \mathbf{x} and t).

Somewhat like the case of a rarefaction, the other variables can all be cast in terms of $f(\mathbf{x}/t)$ rather than $f(\mathbf{x}, t)$. Let

$$\mathbf{x}' = a\mathbf{x} \quad (10)$$

$$t' = bt \quad (11)$$

$$\rho' = c\rho, \quad (12)$$

so, for example, the density evolution will scale as $\rho(\mathbf{x}, t) \rightarrow \rho/c(ax, bt)$. Also, from Eq. (9),

$$\mathbf{v}' = (a/b)\mathbf{v} \quad (13)$$

$$p' = (a/b)^2 cp. \quad (14)$$

This suggests defining

$$Eu = v\sqrt{\rho/p}, \quad (15)$$

so that

$$Eu' = (a/b) * (b/a) * \sqrt{c/c} = Eu. \quad (16)$$

Eu , called the Euler number, is, to within a constant involving γ , essentially the Mach number (and may be referred to as the ‘internal isothermal Mach number’). If Eu (and γ) is the same for two systems, they will have identical hydrodynamic solutions [11] regardless of ρ or p . However, initial and boundary conditions need to be properly scaled as well. For example, if one flow is confined by a wall and the other is freely expanding, clearly the solutions will differ. If the wall is ‘far enough’ away from the region of interest (such that no signal or wave from the wall has time to reach the region of interest), however, one can still apply scaling. Similarly, the initial conditions need to be matched to the degree that they affect the final solution. For example, if the density of the initially confining material for one system is $\rho/100$, the equivalent density of the other system may be smaller or larger than $\rho'/100$ and not impact scaling (as long as other complicating factors such as magnetic fields or heat conduction do not apply); *how much* smaller or larger is acceptable is problem dependent. If both systems are fully turbulent, they may have ‘forgotten’ their initial conditions, but the development of turbulence is sensitive to initial conditions.

In astronomy, the assumptions of the polytropic hydrodynamic equations are often valid, but in the laboratory they generally are not, since experimental materials are often not ideal. Therefore, even though the polytropic equations show the possibility of perfect scaling, in practice an experiment must choose a region of importance (e.g., the Mach disk at the head of a jet) to scale. The experiment then may fix some variables (e.g., Eu , spatial scale, and ρ) and the other variables (e.g., p , t , and \mathbf{v}) will be properly scaled.

If a system has shocks, it is not likely to be isentropic. Using curly brackets to denote differences across a shock, the jump conditions in 1D are

$$\{\rho v_{\perp}\} = 0, \quad (17)$$

$$\{p + \rho v_{\perp}^2\} = 0, \quad (18)$$

$$\left\{ v_{\perp} \left(\frac{\gamma p}{\gamma - 1} + \frac{\rho v^2}{2} \right) \right\} = 0, \quad (19)$$

where v_{\perp} is the component of the velocity perpendicular to the shock. One can still factor out a constant (e.g., $a/b * (a/b)^2 c$ for vp using the notation described above), so the hydrodynamic equations will still scale in the presence of a shock (ignoring for now dissipative mechanisms such as radiative cooling). Thus, scaling can be used in such diverse hydrodynamic applications as wind tunnels (a tunnel experiment with a full-sized airplane is expensive) and rocket turbomachinery (experiments pumping water are a lot easier than ones using liquid oxygen).

2.2.2 Ideal MHD or Euler–Alfvén Scaling

When adding idealized (i.e., nondissipative) magnetic fields to the polytropic hydrodynamic equations, the mass (Eq. (7)) and energy (Eq. (8)) equations do not change. However, the momentum equation has a magnetic force term and so becomes (in CGS units)

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) + \nabla p + \frac{\mathbf{B} \times \nabla \times \mathbf{B}}{4\pi} = 0, \quad (20)$$

where \mathbf{B} is the magnetic field. Similarly, the Faraday equation determines the evolution of the magnetic field:

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) = \mathbf{0}. \quad (21)$$

Since these are the equations for ideal MHD, there is no dissipation and infinite conductivity.

Since $[B] = [p^{1/2}] = [\rho^{1/2}][x]/[t]$, \mathbf{B} does not add anything new dimensionally, so the equations are still scalable [12]. However, for two systems to scale to each other, another dimensionless parameter, in addition to the Euler number, needs to be equal:

$$Al = \frac{B}{v\sqrt{\rho}}. \quad (22)$$

Instead of this choice, the Alfvén number, one could use the β parameter, the ratio of thermal pressure to magnetic pressure, $= 8\pi p/B^2$; with equal Euler numbers, the two are equivalent to within a constant. Thus, an ideal polytropic MHD problem can be perfectly scaled by matching Eu and β (and properly scaling boundary and initial conditions).

2.2.3 MHD Scaling with Viscosity and Ohmic Dissipation

If one includes a constant viscosity η and a constant conductivity σ in the polytropic MHD equations, again the energy and mass equations are not effected. The momentum equation becomes

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) + \nabla p + \frac{\mathbf{B} \times \nabla \times \mathbf{B}}{4\pi} - \eta \nabla^2 \mathbf{v} = \mathbf{0}, \quad (23)$$

and Faraday's law becomes

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) - \frac{c^2}{4\pi\sigma} \nabla^2 \mathbf{B} = \mathbf{0}, \quad (24)$$

where c is the speed of light (again using CGS units). Since $[\eta] = [\rho][x^2]/[t]$ and $[\sigma] = [t^{-1}]$, one can define two more dimensionless parameters, the Reynolds number

$$Re = \rho VL/\eta \quad (25)$$

and the magnetic Reynolds number

$$Re_M = 4\pi\sigma VL/c^2, \quad (26)$$

where V and L are characteristic velocity and length scales, respectively, for the problem at hand. This choice is somewhat arbitrary (e.g., one could use an inverse characteristic timescale instead of VL for both Re and Re_M), but it is a conventional choice (somewhat like Eu versus Mach number). In addition, for Re_M , the constant $4\pi/c^2$ is chosen so that $Re_M \sim 1$ when the dissipation term is comparable to the other terms in the Faraday equation. This is a common convention in defining dimensionless parameters: near unity, the relevant term in the governing equations is comparable to other terms, while if much larger than unity, the term is much less. Note that ideal MHD implicitly assumes that $Re \gg 1$ and $Re_M \gg 1$.

The choice of scales (that is, when and where to choose V and L in the problem) depends on the question at hand, but it is not rigorously defined. If one is interested in the collimated flow of a jet, but not the head, the details (scales and dimensionless parameters) of the latter do not matter (though they may of course affect boundary and initial conditions which would need to be scaled appropriately). The choice also depends on the physics of interest. For magnetic diffusion, the appropriate spatial scale is the scale on which \mathbf{B} changes ‘significantly,’ while for momentum diffusion (viscosity), it is the scale on which \mathbf{v} changes.

2.3 Scaling Comments

Adding more physics (e.g., molecular heat conduction) means adding at least one more dimensionless parameter (e.g., the Prandtl number). Consider the effect of including radiation in the MHD equations, which adds at least two additional dimensionless parameters. The first, the Boltzmann number, is the ratio of convective heat flux by hydrodynamics to that of the radiative flux:

$$Bo = \rho VC_p/\sigma_B T^3, \quad (27)$$

where C_p is the heat capacity at constant pressure, σ_B is the Boltzmann constant, T is the temperature, and V is a characteristic hydrodynamic velocity (e.g., the shock velocity if a shock is present). The second is the optical depth or the average number of mean free paths of a photon:

$$\tau = L/\lambda_p = L\kappa\rho, \quad (28)$$

where L is a characteristic spatial scale, λ_p is an average photon mean-free-path, and κ is the opacity. If the problem also has viscosity, a radiative Reynolds number needs to be considered. In general, any added physics needs to be evaluated for significance relative to all the physics already in the problem. The value of Bo can be used to determine whether radiative flux is important (relative to hydrodynamically driven convection); if $Bo \gg 1$, then radiative flux is unimportant. For two systems to scale, Bo only needs to be much greater than unity for both systems. If radiative flux is found instead to be important, then Bo needs to be the same for both systems. The optical depth is defined differently: if $\tau \ll 1$, then transport of the radiation is likely to be unimportant. In this limit, radiation can be treated simply as an energy sink. For scaling, in the absence of full radiative transport, τ needs to be in the same limit for both systems (much larger or much less than unity) or in the same value (if near unity). This illustrates the increased difficulty of scaling when dimensionless parameters are near unity.

It is difficult, if not impossible, to match all dimensionless parameters everywhere in an experiment, particularly boundary and initial conditions, so the experiment must define the ‘important’ regions while making sure the ‘unimportant’ regions do not significantly affect the solution. However, parameters that are significantly larger than unity (if properly defined as discussed above) do not have to be equal between systems, so long as such parameters are much larger than unity in both systems. There are caveats. For example, consider a viscous hydrostatic system with a large Re (so Eq. (23) applies with $\mathbf{B} = 0$). With η not identically 0, the viscosity term is potentially perturbative, such that two scaled hydrostatic systems that have different Re (though both much larger than unity) will behave significantly differently at late times. Still, in general, dimensionless parameters that are much larger than unity represent ‘ignorable physics’ for that problem.

2.4 Scaling Examples

Since perfect scaling of an astrophysical object in a laboratory experiment is not possible, the experiment usually focuses on a region of interest. Here, I detail one effort to experimentally scale the bow shock of a YSO.

Take a canonical YSO as having a Mach 30 jet with a scale $L_{\text{jet}} \sim 500 \text{ AU}$. With a pressure $P_{\text{jet}} \sim 10^{-9} \text{ dyn cm}^{-2}$ and a density $\rho_{\text{jet}} \sim 10^{-21} \text{ g cm}^{-3}$, we wish to scale this YSO jet to a laboratory experiment. A given facility will have a limit to P_{exp} ; here, we use the Omega facility, where $P_{\text{exp}} \sim 10^{11} \text{ dyn cm}^{-2}$ for a jet experiment. Since the Euler number must be matched, $Eu \sim 30$, the experiment must select v or ρ to achieve the match. Usually, the latter is more controllable, but it has less dynamic range. Choosing $\rho \sim 1 \text{ g cm}^{-3}$, we find that $v = Eu \sqrt{P/\rho} \sim 100 \text{ km s}^{-1}$. Finally, the experiment must have the correct spatial scale to achieve the correct temporal scale (or vice versa). Facilities tend to not have much dynamic range here either, so one or the other may be set by the facility itself. The Omega facility allows spatial scales of hundreds of microns, so in order to model the YSO jet evolution for $t_{\text{jet}} = L_{\text{jet}} \sqrt{\rho_{\text{jet}}/P_{\text{jet}}} \sim 200 \text{ years}$, the experiment needs to last for

$t_{\text{exp}} = L_{\text{exp}} \sqrt{\rho_{\text{exp}} / P_{\text{exp}}} \sim 60 \text{ ns}$. Fortunately, the diagnostics on Omega permit just such an experiment. However, this exercise shows that it is much easier to scale from an experiment (where choices are limited) to a similar astrophysical object than the other way around.

Neither the YSO jet nor the experiment will have the above canonical values everywhere, limiting the applicability of the scaling. However, scaling from experiment to experiment is considerably easier. For example, assume that two systems with large dimensionless parameters (say, Re) behave in the same way (with respect to the physics that parameter describes) as long as both systems have values of, e.g., $Re \gg 1$; it is not necessary for $Re_1 = Re_2$. It is possible to verify this assertion by doing two experiments where only the time and space scales are changed; in the case of such a ‘perfectly scaled’ hydrodynamic experiment, the equations are unchanged, but Re does change [13]. Consider Fig. 6, where an experiment with $Re \gg 1$ is done at two different times, giving results 1 at time $= t_1$ and 2 at time $= t_2$. If the experiment is temporally and spatially scaled by this same ratio (t_2/t_1), result 1 should now look just like 2. If, on the other hand, Re does matter (even though it is much larger than unity), the results of the scaled experiment might look like 3. This is a particularly important issue since Re for ‘numerical experiments’, though larger than unity, is typically much less than the experimental Re , so if Re still matters when much larger than unity, there is no reason to expect calculations to be able to reproduce the experiments.

A summary of some canonical dimensionless values for some astrophysical objects and experimental facilities is given in Table 1; the last line in the table refers to the Omega experiment discussed in the previous section. Experiment values of

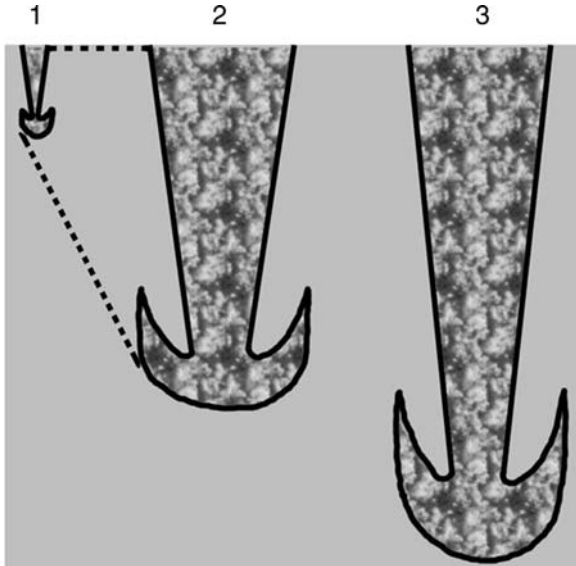


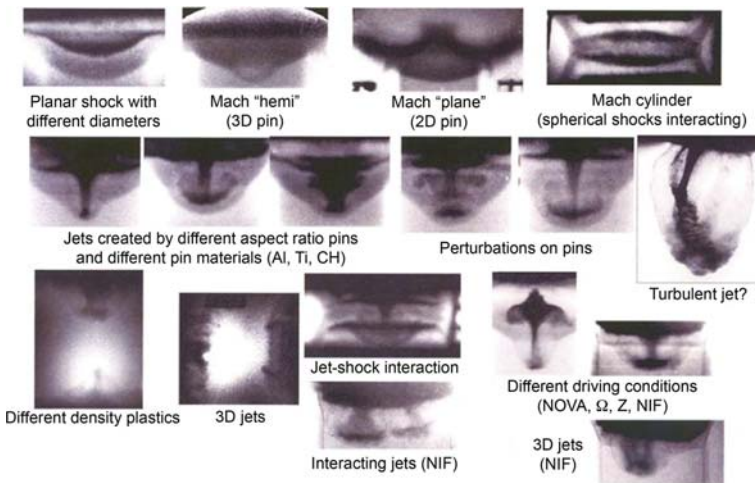
Fig. 6 Illustration of a perfectly scaled experiment. Adapted from [13]

Table 1 Notional dimensionless parameters (adapted from [14]).

Environment	Eu	β	Re	Re_M	Bo	τ
Warm ISM	1×10^1	3×10^{-1}	1×10^7	1×10^{19}	2×10^{-3}	1×10^{-4}
Dense cloud	6×10^0	3×10^{-2}	7×10^{13}	6×10^{14}	1×10^{-9}	2×10^{-5}
Stellar atmosphere	1×10^1	3×10^0	5×10^{12}	4×10^9	6×10^{-2}	1×10^1
Stellar environment	1×10^0	3×10^5	5×10^{11}	4×10^{13}	3×10^{-5}	3×10^3
AGN disk	1×10^1	3×10^{-8}	7×10^7	6×10^{13}	2×10^{-12}	5×10^0
XRB disk	1×10^0	8×10^1	1×10^9	4×10^{11}	5×10^{-4}	1×10^3
YSO jets	3×10^1	3×10^0	1×10^5	1×10^{12}	4×10^{-9}	1×10^{-6}
Omega hohlraum	1×10^0	3×10^1	3×10^3	5×10^1	4×10^{-1}	9×10^{-3}
NIF hohlraum	1×10^0	9×10^1	1×10^3	2×10^3	1×10^{-2}	1×10^{-3}
Z experiment	1×10^0	3×10^1	3×10^4	5×10^2	2×10^{-1}	9×10^{-2}
Short pulse laser	4×10^0	3×10^0	1×10^3	1×10^3	3×10^{-1}	1×10^{-2}
Omega jets (core)	5×10^0	5×10^{-1}	1×10^5	1×10^{-6}	1×10^9	1×10^0

Eu are somewhat lower than astrophysical objects. The magnetic β -parameter has a large range of values in astrophysical objects, but it is generally near unity in experiments (e.g., [15]). Both Re and Re_M tend to be larger in astrophysical objects than in laser experiments, but all tend to be larger than unity. Radiation tends to be important in astrophysical objects, while it is not so much true during hydrodynamic flow produced by the lasers. Finally, the optical depth varies considerably, but large values are not yet attainable in (diagnosable) laser experiments. Thus, transparent nonmagnetized low-Mach number flows are easily scalable to presently available experimental facilities, but other types of systems may have difficulties.

It can be seen from Table 1 that the YSO/Omega scaling is not fully ideal. Practical issues such as diagnostics, target fabrication, and metrology cause the experiment to evolve; it actually started as a scaled jet-induced supernova experiment based on [21]. The equation of state of the materials in the experiment are not

**Fig. 7** Examples of jet experiments on laser facilities. See [16–20]

an ideal gas (the Ti may even still be a liquid; see Fig. 18). Also, there are large gradients in Eu and other nonuniformities due to the impulsive nature of the laser-driven jet drive; that is, the initial and boundary conditions are not scaled to the YSO. However, the experiment does illustrate the possibility of conducting ‘mini-YSO’ experiments to elucidate the nature and evolution of true YSOs. Additionally, a much larger range of ‘jet-like experiments’ are available; some that have been conducted over the years on Omega, Z, NIF, and Nova are shown in Fig. 7. Other types of scaled astrophysics have also been done; one example investigating growth of Rayleigh–Taylor (RT) spikes in supernovae is shown in Fig. 8.

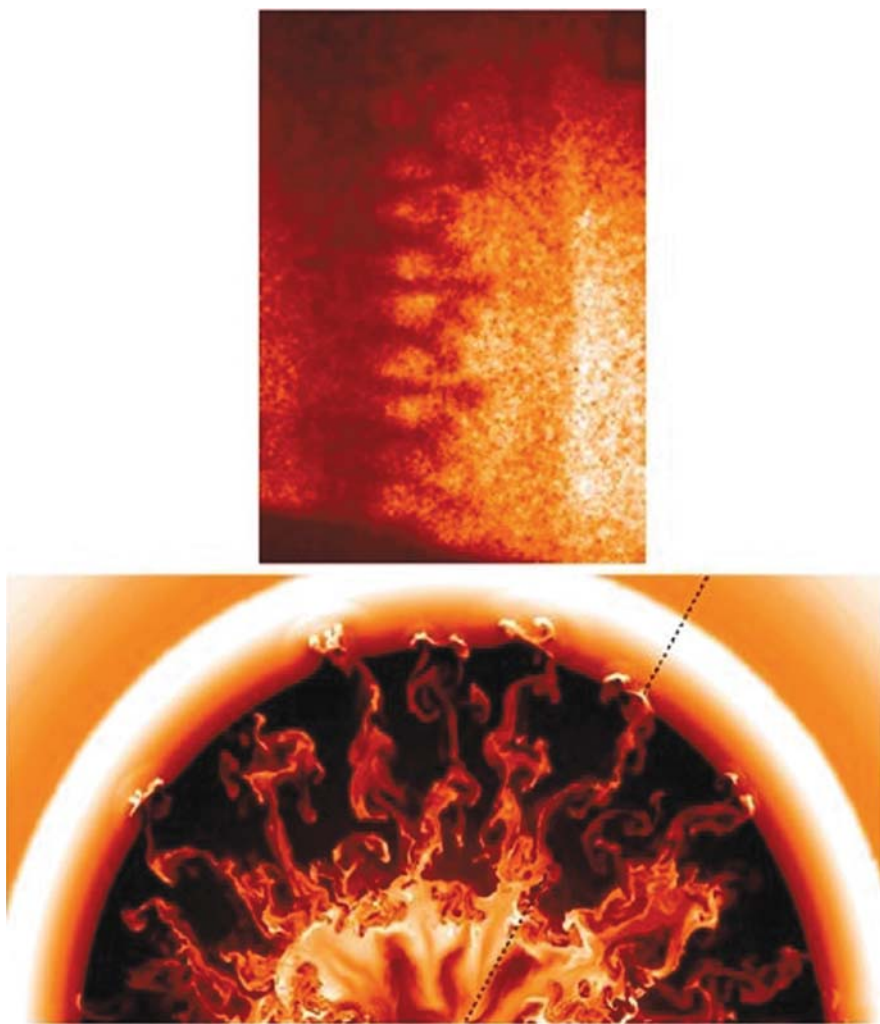


Fig. 8 *Bottom*: Simulation of RT spike growth in a supernova [22]. *Top*: Scaled Omega experiment of the same thing (e.g., [23–25]).

3 Experimental Facilities

Here, I give a high-level overview of presently and soon-to-be operational experimental facilities. The list is far from exhaustive and characteristics of individual facilities often change (though usually for the better).

3.1 *Omega*

The Omega laser is run by the Laboratory for Laser Energetics at the University of Rochester in New York in the United States. It has 60 beams available for spherical implosion. A total of 30 kJ at a frequency of 3ω can be delivered over 1 ns. Direct or indirect drive (using a hohlraum, described below) is available with a variety of backlighter and detector choices. An X-ray CCD is even available [26]. Groups with American collaborators can run experiments on Omega via the NLUF program (the EU has a similar ‘Laserlab’ program). In April 2008, Omega was expanded into ‘Omega-EP’ (extended performance). With a shot rate of 1 every 2 h and 2 new beams delivering in the optical 2.6 kJ in 10 ps or 1 kJ in 1 ps, Omega-EP is an ideal peta-watt (PW) laser for fast-ignition inertial confinement fusion (ICF) experiments.

3.2 *Gekko*

The Gekko XII laser at Osaka University in Japan has 12 beams with a total of 10 kJ of energy deliverable over 1 ns. Another 0.4 kJ 0.5 PW beam is available for fast ignition experiments, with another beam coming online to deliver a total of 10 kJ pulse in 10 ps. The Gekko is a green laser in that it uses frequency doubling rather than tripling, as is typical.

3.3 *NIKE*

The NIKE facility at the Naval Research Laboratory in Washington, DC, USA, is a KrF laser with 44 beams. Compared to typical Nd lasers, it has better beam uniformity, so it is a good facility for minimizing RT growth in an imploding fusion capsule. However, it only delivers 4–5 kJ of UV light in a 4-ns pulse.

3.4 *HELEN*

The HELEN laser is operated by AWE in the UK. It has two beams delivering 500 J each in 1 ns at a wavelength of 527 nm. The achieved intensity on the target is more than 10^{19} W cm⁻². A third beam, used as a backlighter, delivers 25 J in 500 fs. HELEN is being upgraded to ORION, a PW-scale laser.

3.5 *VULCAN*

VULCAN is run by the Rutherford Appleton Laboratory in the UK. It delivers up to 1 kJ, with a pulse duration between 700 fs and 5 ns and can achieve intensities of $10^{20} \text{ W cm}^{-2}$. It is being upgraded to a PW laser with intensities of $10^{21} \text{ W cm}^{-2}$. It has diagnostics including image plates for capturing X-rays and radiochromic films for capturing protons [27].

3.6 *LULI*

The lasers of the Laboratoire d'Utilisation des Lasers Intenses (LULI) are at the Ecole Polytechnique in France. They can produce a total of 30 J over 300 fs, for more than 100 TW of power. Soon they will be able to produce PW power levels, with resulting intensities of more than $10^{21} \text{ W cm}^{-2}$. For scaled jet experiments, a variety of diagnostics (e.g., VISAR and gated optical imaging) are available to infer quantities such as velocity, temperature, and emissivity [28, 29].

3.7 *Zebra, MAGPIE, and Z*

Zebra is located at the Nevada Terawatt Facility (NTF) at the University of Nevada in Reno, NV, in the United States. It has a 2 TW pulsed z-pinch with a 100 TW coupled laser. It is being upgraded to Leopard, a 35 J, 350 fs, $10^{19} \text{ W cm}^{-2}$ beam. The facility is ideal for scaled experiments of plasma flows interacting with magnetic fields [30]. The MAGPIE facility at Imperial College, London, is a TW pulsed power machine that can generate over an MA for more than 100 ns. It has an extensive suite of diagnostics, from Schlieren photography to XUV imaging, that has been used to observe z-pinch-induced jets interacting with a target [31]. The Z-machine at Sandia National Laboratory in New Mexico in the USA is another pulsed power z-pinch machine, using a 20 MA electrical discharge. The imploding plasma produces an X-ray pulse of 290 TW for a few ns. The Z-machine has just been refurbished (and renamed 'ZR') to produce 350 TW for a total X-ray output of 2.7 MJ. A variety of scaled experiments from jets [32] to supernova blastwaves [33] have been conducted on Z. Illustrations of how a z-pinch works are shown in Fig. 9. The parallel currents produce an inward $\mathbf{J} \times \mathbf{B}$ force that accelerates the wires. When the wires implode cylindrically on-axis, they release MJ of X-ray energy. This radiation can be used to drive a number of different experiments at different angles for the same shot. Note that z-pinch facilities are inherently cylindrical while direct drive facilities for use in ICF generally permit spherical implosions, with a resulting higher compression ratio.

3.8 *NIF and LMJ*

The National Ignition Facility (NIF) at LLNL in California in the USA is scheduled for full operation in 2010; it has already delivered 1.1 MJ in the ultraviolet. It will

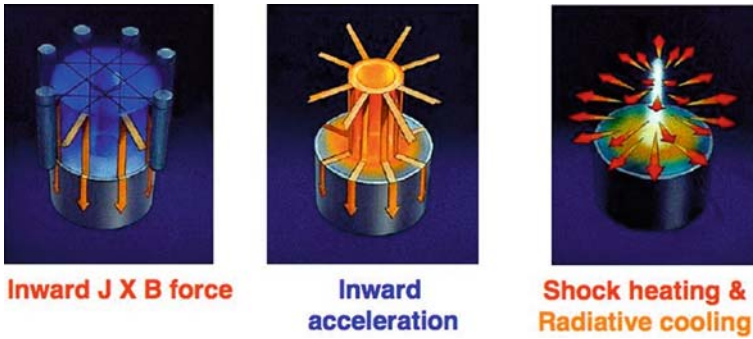


Fig. 9 Schematic of how a z-pinch works (from lectures by R. P. Drake)

eventually have 192 beams delivering 4.2 MJ in the infrared over a timescale of up to tens of ns. NIF is designed to ignite fusion capsules and presently has a slew of available diagnostics from VISAR to a gated X-ray CCD. The Laser Mega-Joule (LMJ) Facility in France will be a similar facility, with 240 beams delivering 2 MJ of energy with a peak power of 550 TW. LIL is the LMJ prototype; first light for the full 240 beams is scheduled for 2009. One of the first experiments done on NIF, in the early light phase, is shown in Fig. 10. The shock in the left image should have been flat; gold fiducial wires showed the misalignment of the drive, which was fixed for later jet shots, as shown in the right image.

Figure 13 illustrates the scale and complexity of modern laser facilities. The room housing the Z chamber is the size of a small house, with a relatively small attached control room. The entire building is a Faraday cage. The building holding NIF, on the other hand, is the size of three (American) football fields. Interestingly, the Z building resonates noticeably with each shot while NIF is eerily silent.

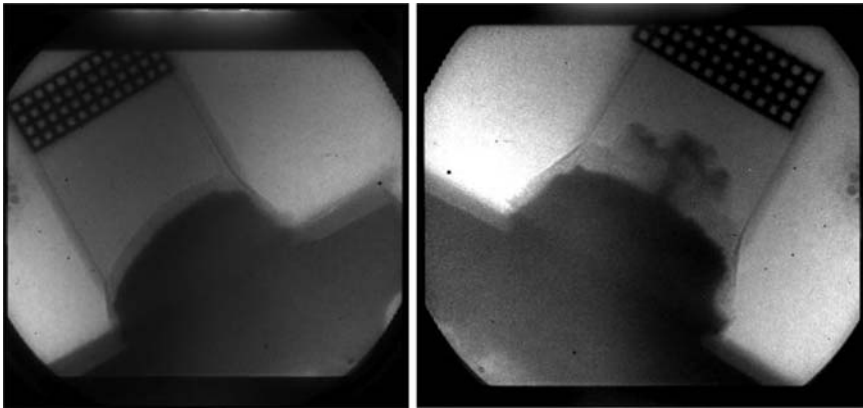


Fig. 10 NIF early light radiographic images of a planar shock (*left*) and a jet (*right*). See, e.g., [34]

4 Detailed Example and a Challenge

As mentioned above, we have conducted experiments on Omega. A schematic of the experiment target and assembly is shown in Fig. 11. For this particular experiment, we used indirect drive that is, the drive laser beams enter the hohlraum at an angle, ablating the gold which in turn produces a pressure pulse at the surface of the target. Figure 12 shows an older target design which used direct drive; the drive lasers directly ablate the surface of the target. Even though the Au shield and Ti target were coated with plastic to minimize background, the design resulted in a very high background due to photons leaking through the glue gap in the figure. However, we redesigned the gold shield to fix that problem, since we thought that direct drive was still needed to avoid a second jet at late times (we eventually moved to indirect drive due to unsolved target alignment issues). We needed to backlight the target several hundred nano seconds after the drive in order to capture late-time turbulent behavior. In order to get good contrast, we used a V backlighter with a Ti target. A schematic showing the ‘snout’ that contains the film is shown in Fig. 14. Details of the foam target as well as an example of a simulated radiograph are shown in Fig. 15. A 3D schematic representation of how the laser beams interact with the target is shown in Fig. 16.

Figure 17 shows spectral densities extracted from a film of a typical Omega jet shot. The background arises from X-ray photons of less than ~ 10 keV.

After subtraction, given the opacity and assuming cylindrical symmetry for the foam, a 3D density profile can be extracted using the unsaturated pixels.

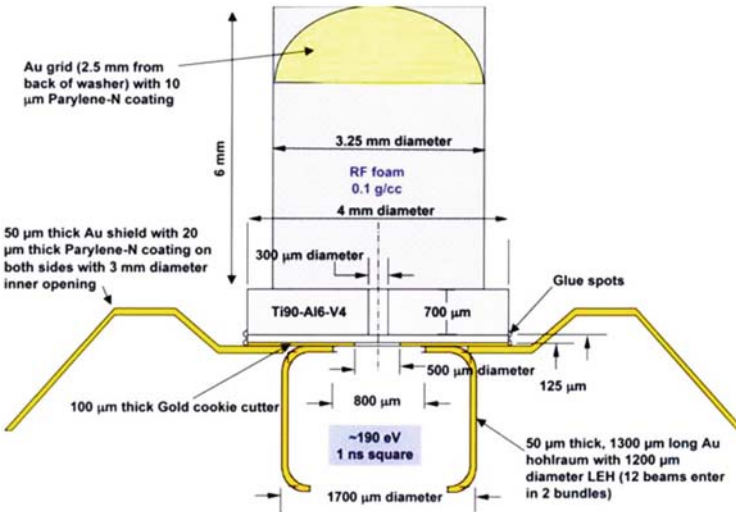


Fig. 11 Schematic of the indirect drive targets used on Omega for our jet experiments. Not all parts are shown to scale

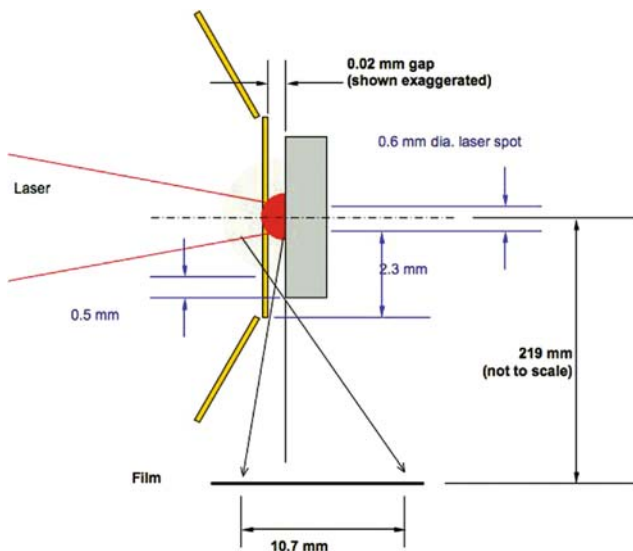


Fig. 12 Schematic of older direct drive targets used on Omega, showing the potential leaking of X-rays through the plastic and glue gap between Au and Ti. The foam (not shown) is to the right of the washer (*in gray*) and the backlighter is located off the top of the diagram

Figure 18 shows two radiographs from the Omega jet experiments taken at 200 ns; it is now possible to obtain two radiographs at the same time for the same shot (e.g., [35]). Uniform illumination is sometimes an issue, as can be seen in the left image in Fig. 18. The right image in Fig. 18 shows the gold fiducial grid, used as a spatial reference in determining resolution, and what appear to be liquid droplets of Ti. The latter is not too surprising given that the peak temperature during jet propagation is only tens of electron Volts. Note the Ti backlighter penetrates more deeply

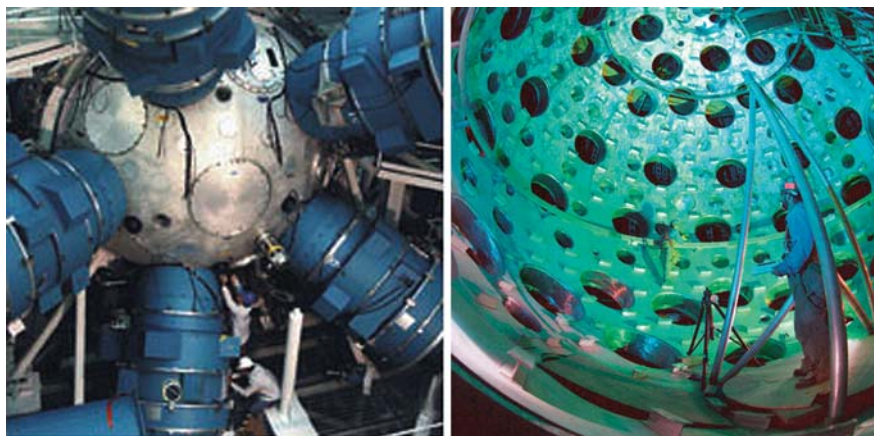


Fig. 13 Illustration of the scale and complexity of the Z (*left*) and NIF (*right*) chambers

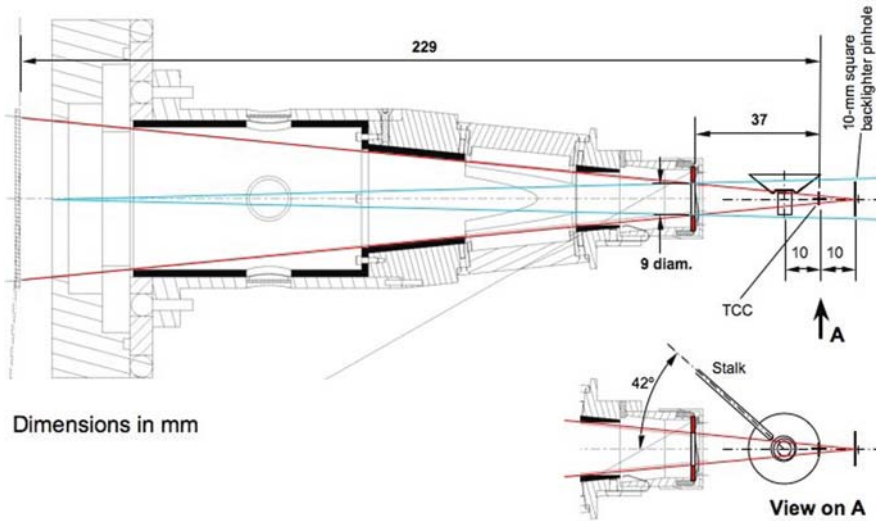


Fig. 14 Details of the snout (SPCA) setup. Filters are also in place to select for the backlighter energy

into the ‘flute’ of the jet, showing its inner structure, than the V backlighter does. Both images show the ‘pedestal,’ where the shock has broken out of the backside of the Ti washer, as well as the (variable) 3D nature of the jets.

The (unknown) initial conditions of the experiment (target roughness, drive uniformity, etc.) can have a large effect, essentially determining the fine structure of the jet. Issues of repeatability caused by variation in initial conditions made us rethink how the Ti washer was made. Figure 19 shows scanning electron microscope (SEM) images of our initial and redesigned targets. The redesign provided a much smoother surface, with less than $1\text{ }\mu\text{m}$ RMS surface roughness, enabling more repeatable initial conditions in the experiment. This feedback between the experiment designers and the target fabrication is essential.

Despite nonrepeatable effects caused by initial conditions, the displacement and diameter of the bow-shock can be effectively modeled. It is this type of experiment, which illustrates where codes succeed and where they fail, that is most useful. From

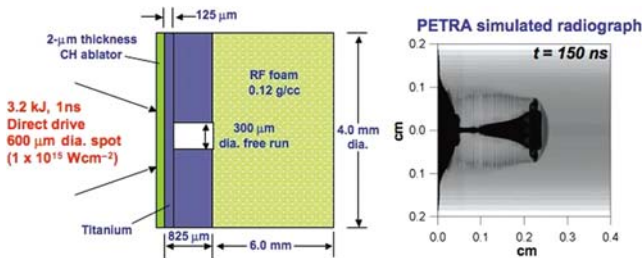


Fig. 15 *Left:* Details of the foam and target for the direct drive experiments on Omega. *Right:* A simulated radiograph from a PETRA simulation

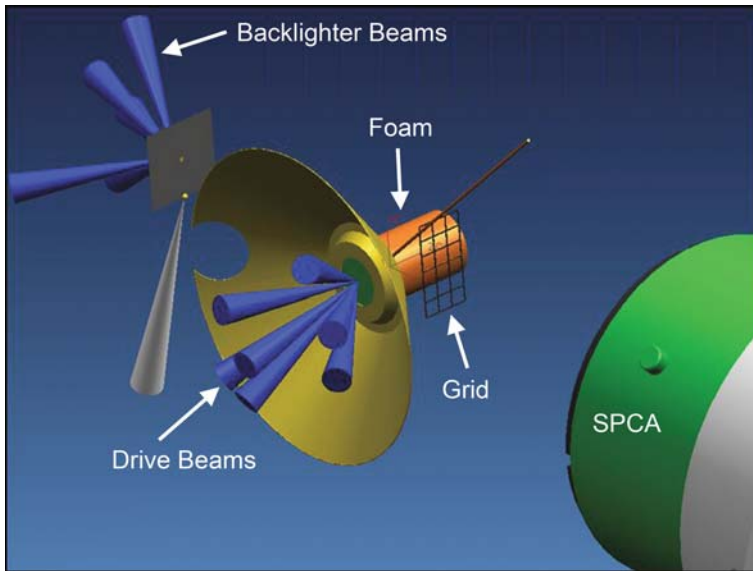


Fig. 16 A 3D schematic showing the drive and backlighter beams for the direct drive Omega jets. The grid used in the experiments was semi-circular rather than the rectangular grid shown here

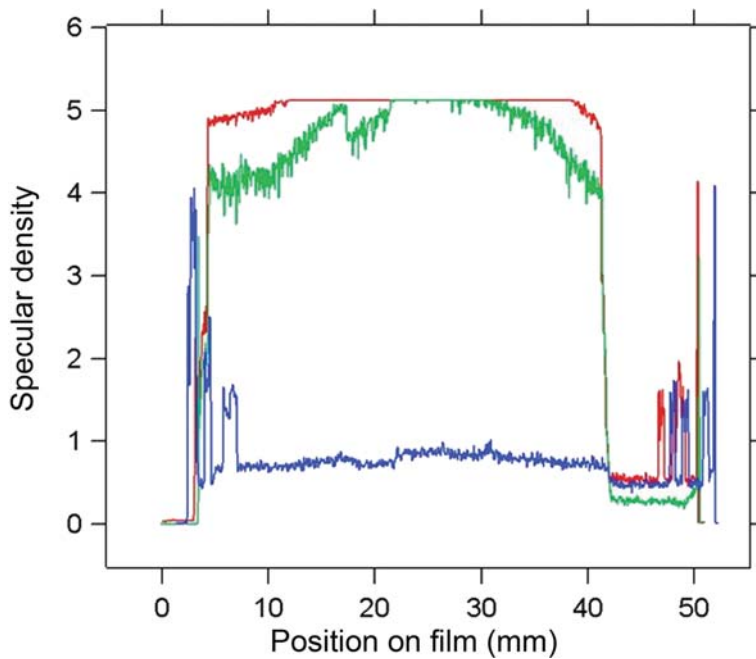


Fig. 17 Curves showing background (*bottom*), saturated (*top*), and unsaturated foam (*middle*) line-outs across the film of a typical Omega jet shot

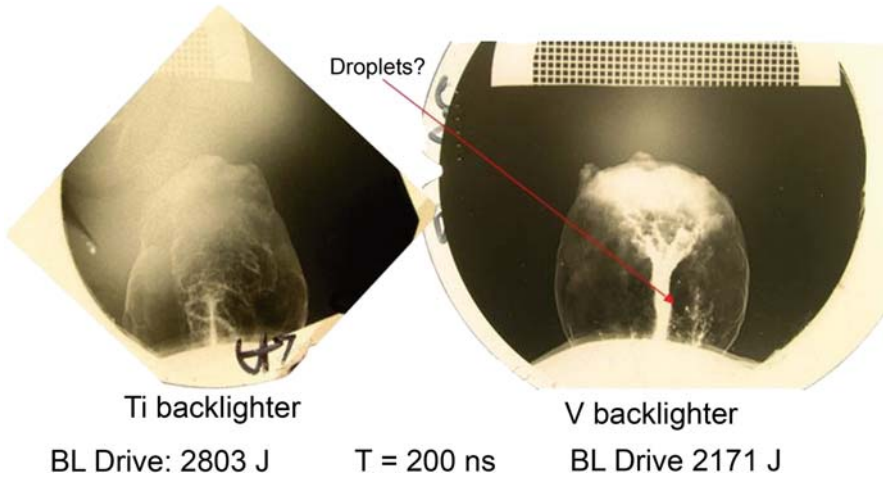


Fig. 18 Backlighter images from the jet experiments on Omega. The images are scanned in from the film. For scale, the grid lines, visible at the top of the images, are 60 Mm apart

a modeling perspective, this brings up a qualitative aspect of scaling: If a well-scaled experiment may be modeled by a numerical code, then the code may be applied more confidently to a similar astrophysical system; greater success modeling the experiment leads to greater confidence in the model output.

Results of simulations and experiments of a hohlraum-driven jet in a cylindrical foam are given in Fig. 20. The model successfully predicts the diameter of the bow-shock and describes the evolution of the pedestal. Deviations between the modeled fits and the experimental data are still under investigation. PETRA, an AWE Lagrangian code, underpredicts the displacement of the pedestal, while both PETRA and RAGE, a LANL/SAIC Eulerian AMR code, overpredict the displacement of the bow-shock. The latter may be due to equation-of-state issues (foam is notoriously difficult to model) or dissipation mechanisms that are not being properly captured (e.g., turbulence). Both codes model the bow-shock diameter to within the

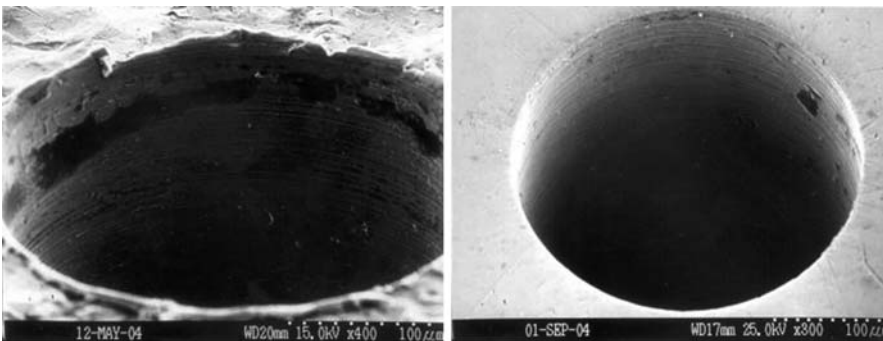


Fig. 19 SEM of a target using standard drill (*left*) and after burr removal and polishing (*right*).

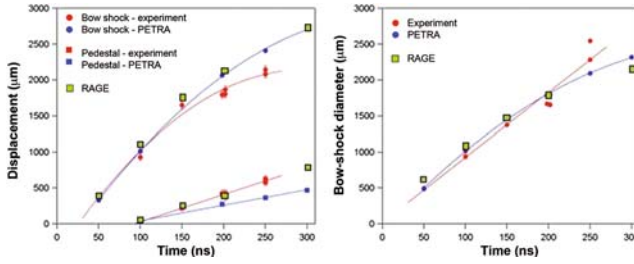


Fig. 20 Comparison of Omega jet data to simulations

shot-to-shot variation, though at late times the simulations suggest a flattening that may not be in the data.

The latest version of the jet experiments contained a ball of polystyrene in the 0.1 g cm^{-3} hydrocarbon foam. A total of 12 laser beams ablated the gold hohlraum. The resulting pressure pulse sent a cap of Ti down the ‘free run’ region (the hole in the washer). This collimated material collided with the polystyrene sphere, producing a deflected jet which was imaged by an X-ray backlighter. We investigated how the deflected jet behaved with different locations for the sphere. Figure 21 shows a schematic of the experiment, while Fig. 22 shows pre-shot images of a target. The foam is not optically transparent, so X-ray radiographs are used to locate the ball. The images show that the mating of the foam to the backside of the Ti washer is not perfectly flat, leading to perturbations not included in any of the simulations. Also, a stalk is required to hold the ball in place as the foam solidifies around it; this is not ideal for the scaling of the experiment, but the stalk can be included in the simulations.

Coarse resolution 3D simulations (with the finest resolution larger than $\sim 6 \mu\text{m}$) of the ball-in-foam experiments show a smooth intact ‘flute,’ while the data show significant small-scale structure and breakup. Finer resolution simulations begin to show this structure, but $1 \mu\text{m}$ resolution is required to capture the breakup of the Mach ring (the ‘grass’ on the pedestal visible in the right image of Fig. 18). This is

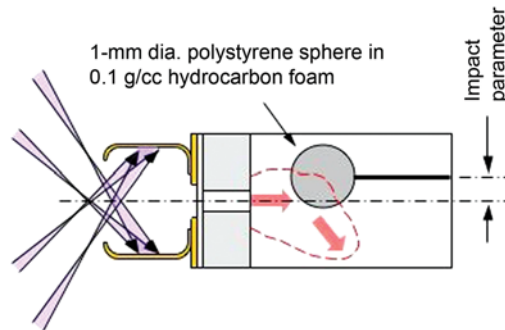


Fig. 21 Schematic showing the evolution of the ball-in-foam jet target. The impact parameter and displacement of the ball are varied from shot to shot

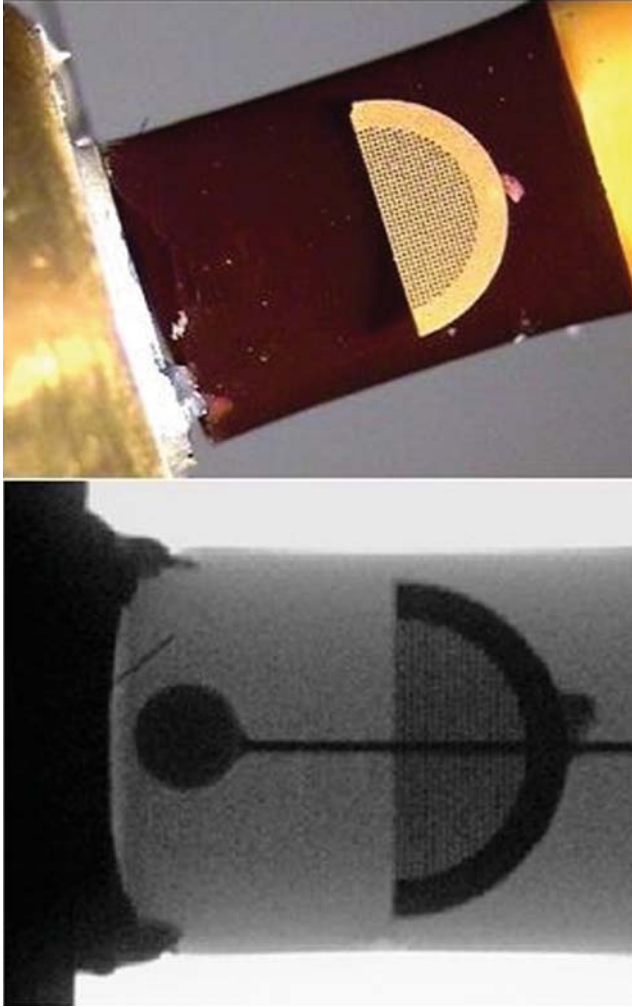


Fig. 22 Pre-shot optical (*top*) and X-ray (*bottom*) images for a ball-in-foam target

not unsurprising since the low-numerical Re is not likely to capture such ‘mixing’ behavior correctly. However, the low-resolution simulations still capture the large-scale structure correctly, just as they did in the original jet experiments (see Fig. 20). This is illustrated in Fig. 23. Lack of knowledge of initial conditions prevents the simulations, regardless of resolution, from reproducing in detail the fine-scale features of the data. Although the experiments are likely not turbulent in the sense of having a fully developed inertial range, they are certainly heavily ‘stirred’, and so the resulting structures still have some memory of the initial conditions.

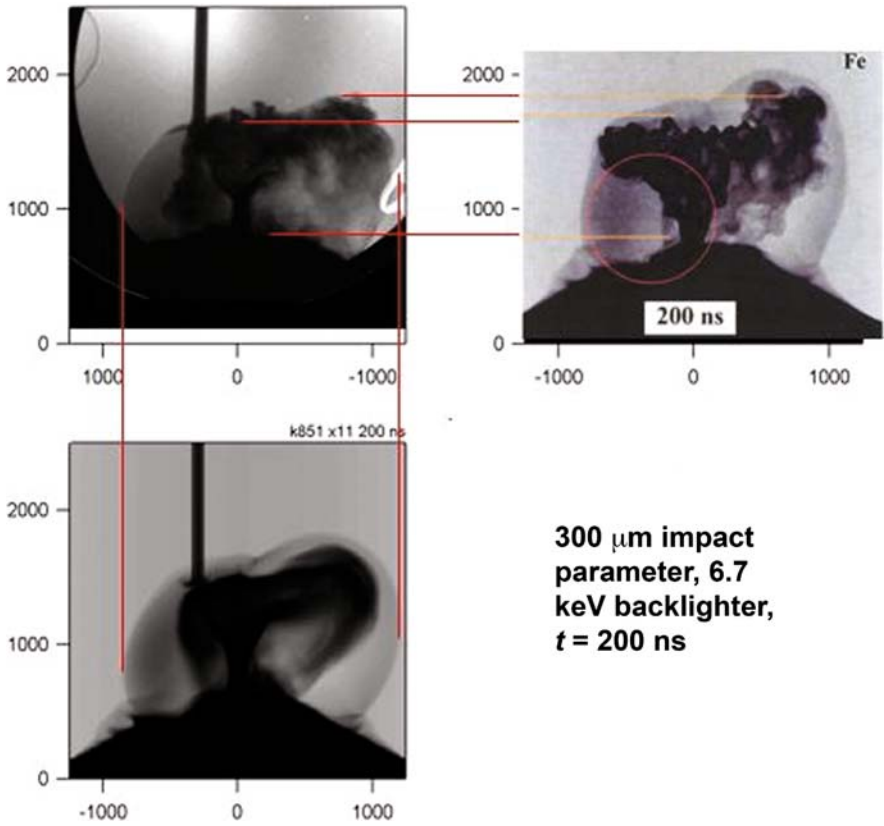


Fig. 23 Data (*upper left*) and RAGE simulations (12 μm resolution: *lower left*, 1 μm resolution: *upper right*) of a ball-in-foam Omega jet experiment. Spatial units are in microns. The circle shows the initial ball location, while the lines are guides showing different features of the jet

The ball is a scaled representation of the giant molecular cloud that HH270 is colliding with to produce HH110. The scaling is not perfect (e.g., the density ratio between the foam and the ball is an order of magnitude too low).

However, compare Fig. 24 with Fig. 1 of [36]. Although the jets are launched in different ways, the qualitative structure in density and velocity space is similar, illustrating the possibility of learning about how the YSO jet is being deflected from the laboratory experiments.

Our latest Omega shots are not jets or even explicitly scaled astrophysics experiments; rather, they explore a generic planar shock running into a spherical obstacle. Such shocks appear in astrophysical objects (e.g., red supergiant winds, planetary nebulae, etc.) but these are not specifically scaled to them. Density plots from a RAGE simulation of the shock experiment are shown in Fig. 25. The left image shows the initial setup. A layer of plastic laced with Br is needed to prevent X-ray radiation from reaching the Al ball directly; if this happens, the ball would expand

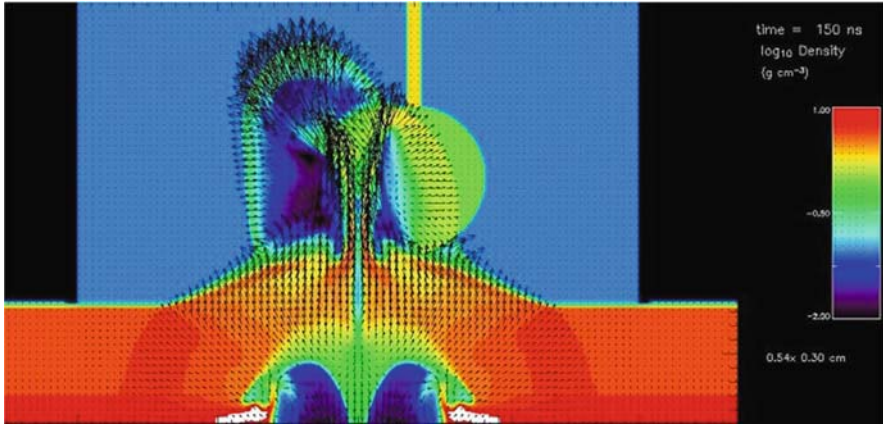


Fig. 24 RAGE simulation of the ball-in-foam Omega experiments that are scaled versions of HH110/270

prior to the shock arrival. The right image shows how the ball has been destroyed by 250 ns. We expect the downstream flow to show Widnall instabilities due to the nonaxisymmetric vortex ring [37]. Even without imposed perturbations, 3D simulations show breakup on scales of tens of microns; Ni or Zn backlighters with small enough pinholes should be able to image these features. We also plan experiments with multiple obstacles.

Figure 26 shows a lineout of Eu through the simulation at 250 ns. The flow is supersonic but not vastly so and highly variable in time and space. If this experiment is scalable to an astrophysical object, with correct initial conditions, the variations in

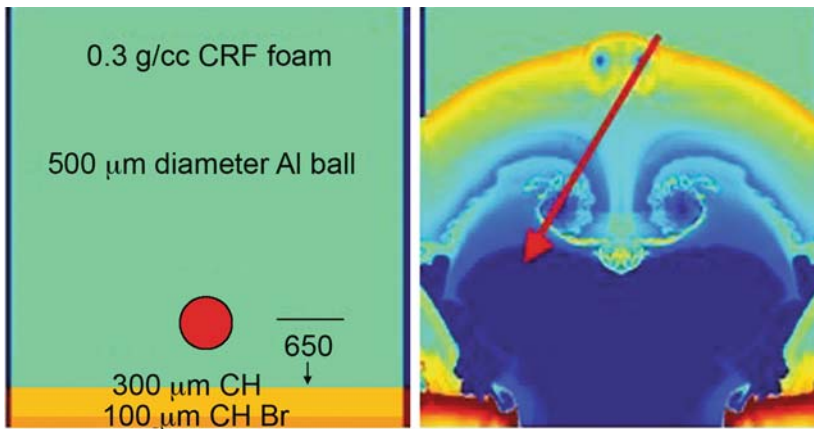


Fig. 25 RAGE simulation of the planar shock experiment at 0 time (*left*) and after 250 ns (*right*). The axisymmetric simulation had a resolution of 1.5 μm. The *large arrow* traces the lineout of Euler number shown in Fig. 26

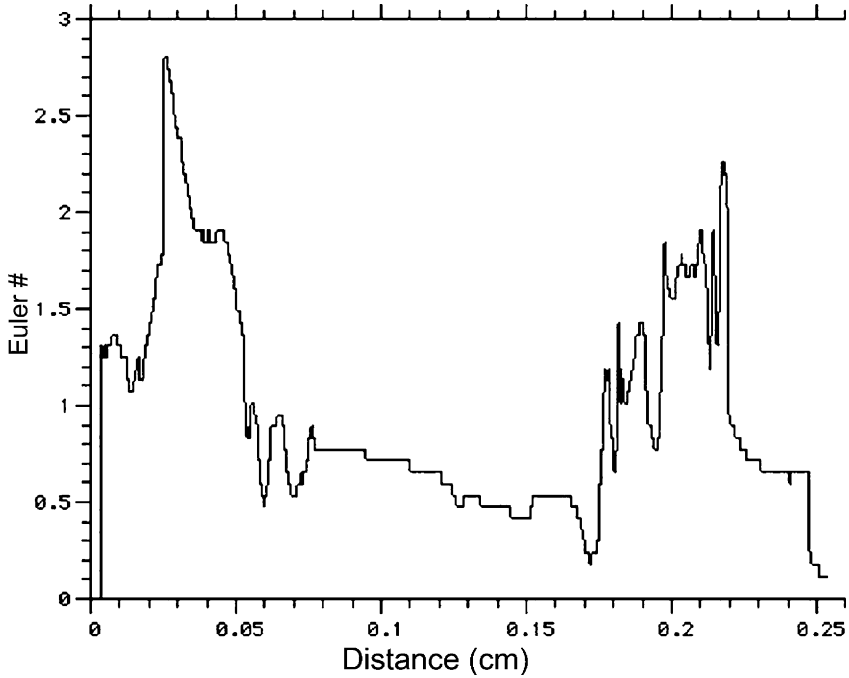


Fig. 26 Lineout of the Euler number for the *arrow* shown in the *right* image of Fig. 25

Eu shown in Fig. 26 will also be present in the astrophysical object. This illustrates the predictive possibilities of laboratory astrophysics.

4.1 A Scaled AGN Jet

To my knowledge, no scaled AGN jet experiment has ever been attempted, so this is a potential example of future work. Taking numbers based very roughly on surveys of radio loud objects with large lobes and jets, one can determine representative AGN jet characteristics. Unlike the YSO jets, it is likely to be inaccurate to assume infinite conductivity, so nonidealized magnetic fields need to be included. The question is Can present or near-future facilities do experiments that are scaled AGN jets? Take, for example, the following values: $L = 10^{19}$ cm, $V = 10^9$ cm s $^{-1}$, $\rho = 10^{-27}$ g cm $^{-3}$, $p = 10^{-11}$ dyn cm $^{-2}$, and $B = 15$ μ G. With these canonical values, one gets for the Euler and Alfvén scaling parameters: $Eu = V\sqrt{\rho/p} \sim 20$ and $\beta = 8\pi p/B^2 \sim 1$, respectively. Leaving aside issues of thermal conduction, viscosity, and radiation, one can look at the ideal MHD scaling. Given experimental facility limitations of $p_{\text{exp}} \sim 10^{12}$ dyn cm $^{-2}$ and $\rho \sim 5$ g cm $^{-3}$, in order to Euler scale, one needs $v_{\text{exp}} \sim 300$ km s $^{-1}$. This sets an experimental timescale of ~ 20 ns. These constraints are feasible with today’s facilities. However, to Alfvén

scale, one needs $B_{\text{exp}} \sim 5$ MG. Upgrades at the NTF and the Prague Asterix Laser System in the Czech Republic [38] are achieving fields of more than 1 MG, so a field of 5 MG is not far out of reach. The difficulty is likely to be in scaling the other pieces of physics, such as radiation. I leave it to the next generation of experimental astrophysicists to develop an experiment that accurately represents an AGN jet.

5 Summary

Many astrophysical phenomena are difficult or impossible to observe directly with sufficient detail or for long enough times to understand them. The inability to conduct reproducible experiments of these phenomena is also limiting. However, they may be investigated using appropriately scaled laboratory experiments. If all dimensionless parameters describing two systems are the same, they will behave identically since the equations describing their behavior will be the same. In practice, limitations on achievable laboratory conditions restrict investigations to specific aspects of the relevant systems (radiative transfer in supernovae, Mach disks in YSOs, etc.) However, with upgrades and new facilities, scaling using laboratory astrophysics is an expanding field with many opportunities.

Acknowledgement Much of the work presented here was done by members of the NLUF Jet Group, particularly B. Wilde, P. Rosen, J. Foster, B. Blue, and P. Hartigan. The Omega experiments would not be possible without the assistance of the LLE support staff and AWE, GA, and LLNL target fabrication teams.

References

1. Reipurth, B., et al.: A&Ap. **311**, 989 (1996) 2
2. Coker, R. F., et al.: ApSS. **307**, 57 (2007) 2
3. Barenblatt, G. I. R.: *textitScaling*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, UK 3
4. Bose, E., Rauert, D.: Phys. Zeit. **10**, 406 (1909) 3, 4
5. von Karman, Th.: Phys. Zeit. **12**, 283 (1911) 4
6. von Karman, Th.: *TextitAerodynamics*, Cornell University Press, Ithaca (1957) 4
7. Landau, L. D., Lifshitz, E. M.: *TextitFluid Mechanics*, 2nd edn., Vol. 6, Butterworth-Heinemann, Oxford (1997) 4, 241
8. Sedov, L.: Prikl. Mat. Mekh. **10**, 241 (1946) 5
9. Taylor, G. I.: Proc. Roy. Soc. **201**, 159 (1950) 6
10. Zeldovich, Y., Raizer Yu, Y.: *Textitphysics of Shock Waves and High Temperature Hydrodynamic Phenomena*, Academic, New York (1967) 7
11. Ryutov, D., et al.: ApJ. **518**, 821 (1999) 8
12. Ryutov, D., et al.: ApJS. **127**, 465 (2000) 9
13. Ryutov, D., Remington, B. A.: Physics of Plasmas. **10**, 2629 (2003) 12
14. Castor, J. I.: ApSS. **307**, 207 (2007) 13
15. Presura, R., et al.: ApSS. **307**, 307 (2007) 13
16. Foster, J. M., et al.: Physics of Plasmas. **9**, 2251 (2002) 13
17. Foster, J. M., et al.: ApJL. **634**, L77 (2005) 13

18. Blue, B. E., et al.: *Physics of Plasmas*. **12**, 6313 (2005) 13
19. Blue, B. E., et al.: *Phys. Rev. Lett.* **94**, 095005 (2005) 13
20. Blue, B. E., et al.: *Journal de Physique IV*. **133**, 107 (2006) 13
21. Khokhlov, A. M., et al.: *ApJL*. **524**, L107 (1999) 13
22. Kifonidis, K., et al.: *A&Ap*. **408**, 621 (2003) 14
23. Drake, R. P.: *Plasma Phys. Contr. Fusion*. **47**, 419 (2005) 14
24. Drake, R. P., et al.: *ApJ*. **564**, 896 (2002) 14
25. Kuranz, C. C., et al.: *ApSS*. **307**, 115 (2007) 14
26. Sublett, S., et al.: *ApSS*. **307**, 47 (2007) 15
27. Koenig, M., et al.: *ApSS*. **307**, 257 (2007) 16
28. Loupiau, B., et al.: *ApSS*. **307**, 103 (2007) 16
29. Michaut, C., et al.: *ApSS*. **307**, 159 (2007) 16
30. Sotnikov, V. I., et al.: *ApSS*. **307**, 99 (2007) 16
31. Ampleford, D. J., et al.: *Phys. Rev. Lett.* **100**, 035001 (2008) 16
32. Bennett, G. R., et al.: *APS Meeting Abstracts*, 1015P (2004) 16
33. Edens, A. D., et al.: *ApSS*. **307**, 127 (2007) 16
34. Landen, O. L., et al.: *Eur. Phys. J. D*. **44**, 273 (2007) 17
35. Kuranz, C. C., et al.: *Rev. Sci. Instrum.* **77**, 327 (2006) 19
36. de Gouveia Dal Pino, E. M.: *ApJ*. **526**, 862 (1999) 25
37. Widnall, S. E., et al.: *J. Fluid Mech.* **66**, 35 (1974) 26
38. Nicolai, P., et al.: *ApSS*. **307**, 87 (2007) 28

Laboratory Studies of Astrophysical Jets

Andrea Ciardi

Abstract Jets and outflows produced during star-formation are observed on many scales: from the “micro-jets” which extend over a few hundred Astronomical Units to the “super-jets” which propagate over distances of a few parsecs. Recently, a new “class” of short-lived (hundreds of nano-seconds) centimetre-long jets has emerged in the laboratory as a complementary tool to study the physics of astrophysical jets. Here I will discuss and review the work aimed at “simulating” protostellar jets in the laboratory using z-pinch machines.

1 Introduction

The connection between astrophysics and laboratory experiments is often made in the context of atomic physics, and in particular spectroscopic studies. These generally concentrate on the micro-physics of astrophysical plasmas, providing important data on cross sections, opacities, grain chemistry and other processes that are vital to the modelling and interpretation of astronomical observations. Indeed, detailed laboratory opacity measurements of iron absorption lines have greatly improved models of pulsation periods in Cepheid variables [17].

Recreating large-scale astrophysical phenomena in the laboratory has also inspired scientists. A pioneer in the use of laboratory experiments to study space physics was the Norwegian scientist Kristian Birkeland (1867–1917), who used a terrella¹ and gas discharges to investigate auroral phenomena [22]. Indeed, as early as the seventeenth century a terrella was used by William Gilbert to study the Earth’s magnetism [40]. More recently, the idea that physical problems in the cosmos may be elucidated by more mundane Earth-based events, such as fluid flows and nowadays laboratory plasmas, was taken up in a series of meetings held just after World War II, with the aim: ‘To bring together workers from astrophysics and from aerodynamics; ...to consider which developments in fluid mechanics may be

A. Ciardi (✉)

Université Pierre et Marie Curie, LERMA, UMR 8112, Observatoire de Paris, 5 Place Jules Janssen, 92195 Meudon, France, andrea.ciardi@obspm.fr

¹ A terrella is a small (tens of centimetres) magnetized model Earth.

applicable to astrophysical problems, and to arrive at a formulation of these problems in such a way that mathematicians and fluid mechanics people may find a way of attack' [55]. Topics of discussion included accretion, gas and dust in the interstellar medium, shocks and turbulence; the latter being very popular with those working in fluid mechanics. An indication that this was taken rather seriously can be gleaned from the list of scientists who took part in the meetings. These included the likes of Heisenberg, Sciama, Bondi, Hoyle, Bok, Burgers, Seaton and Cowling. At the meeting held in Cambridge in 1954, Chandrasekar and Fermi contributed a preliminary paper but did not actually attend; while Hubble, although listed as participant, had died in fact 9 months earlier.²

In the last 10 years or so, experiments on pulsed power facilities (z-pinches) and high-power lasers are leading the way in studying astrophysical phenomena in the laboratory. Emerging areas of research have been aimed at producing complex dynamical phenomena, such as compressible hydrodynamic mixing, hypersonic jets, shock physics, radiation hydrodynamics and photoionized plasmas, to name a few. These can help to understand the physics of phenomena associated with a wide range of astrophysical objects, including protostellar and AGNs jets, supernovae explosions and the subsequent generation of remnants and photoevaporated molecular clouds. Here we will restrict the discussion to the jets produced on z-pinch facilities and in particular to the work performed on the MAGPIE generator at Imperial College. For a broader and more detailed discussion of laboratory astrophysics on lasers and z-pinches we refer the reader to the review by Remington et al. 2006 [46] and the book by Drake 2006 [21].

2 Plasma Conditions in z-Pinch and Laser Experiments

Typical plasmas produced on z-pinch and laser facilities have pressures of \sim Mbar, corresponding to energy densities of $\sim 10^{12}$ erg cm⁻³, at a fraction of solid density. An overview of the plasma conditions attainable on experimental installations together with some of those found in space is given in Fig. 1. z-Pinch facilities rely on stored electrical energy (hundreds of kilojoules) to deliver large currents (\sim of a few mega amperes) over a short time (~ 100 – 1000 ns) to a 'load' usually consisting of a gas or thin metallic wires. These facilities typically produce volumes of plasma of ~ 1 cm⁻³ (for a review see [48]). Laser facilities instead rely on focusing onto a solid or gaseous target, single or multiple high-power laser beams. These concentrate several kilojoules of energy, over timescales \sim pico- to nano-second, into plasma volumes of ~ 1 mm⁻³. The future arrival of the Laser Megajoule (LMJ) Facility in France and the National Ignition Facility (NIF) in the USA will produce fusion plasmas under conditions similar to stellar interiors.

For the present discussion (i.e. jets) it is more interesting to look at the dynamical conditions that can be obtained in the laboratory. When the energy available on z-pinches and lasers is partially converted into kinetic energy, it can generate hyper-

² As a final piece of trivia, Hubble's funeral was never held and the location of his resting body never disclosed.

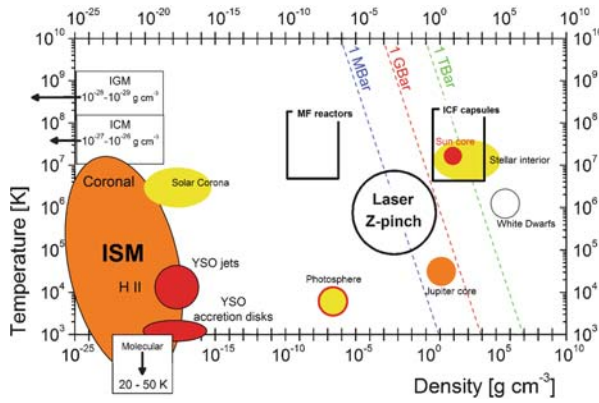


Fig. 1 Plot of the temperature versus density for a variety of laboratory and astrophysical plasmas. Lines of constant pressure are shown for fully ionized hydrogen. High-energy density laboratory astrophysics is in the regime of pressures ≥ 1 Mbar and the typical conditions currently obtained in laser and z-pinch experiments are easily in this range. The conditions that will be accessible in future magnetic fusion (MF) reactors and inertial confinement fusion (ICF) laser experiments are also indicated. The main phases of the interstellar medium (ISM) are shown. The intergalactic medium (IGM) and the intracluster medium (ICM) lie outside the plot at lower densities

sonic (Mach numbers $M > 5$), radiatively cooled flows with characteristic velocities of the order of $100\text{--}1000 \text{ km s}^{-1}$. These flows can include dynamically important magnetic fields, \sim several 10^6 Gauss, and have a large range of plasma- β (the ratio of thermal to magnetic pressure), $1 \gg \beta \gg 1$. In such cases, our inability to obtain the adequate astrophysical plasma conditions (see Fig. 1) may be overcome by producing scaled ‘conditions’ of the phenomena of interest. These are discussed in the next section.

3 Relating Laboratory and Astrophysical Phenomena

The framework to relate experiments to astrophysical phenomena, within magneto-hydrodynamics, was developed in a series of papers by Ryutov et al. [47, 49–51]. A general discussion of hydrodynamic scaling can also be found in [56], and in the present volume it is reviewed in detail by Cocker. An interesting duality between imploding and exploding systems is given in [43]. For collisionless plasmas the scaling conditions are described in [20].

Here we will qualitatively introduce some of the ideas behind scaling and shall be concerned with systems which are to a good degree magnetofluids. As a specific example we take the MHD jet launching, which will be later discussed in more detail in the context of laboratory experiments. Figure 2 shows schematically the comparison of the ‘environment’ producing astrophysical and laboratory jets. The details of the experimental set-up depicted are not important at this stage; it is sufficient to say that the physical processes leading to the distributions of magnetic fields, plasma density, pressure and velocity inside the ‘modelling box’ (MB) are completely different for the two systems. However, because we are interested in studying experimentally the collimation and launching of the jet in the MB, we are

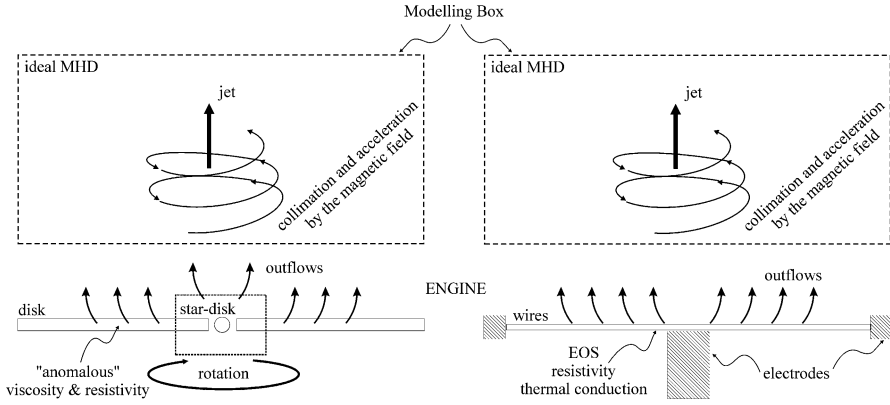


Fig. 2 The lower part of each plot shows schematically the ‘engine’ that produces astrophysical (*left*) and laboratory jets (*right*). The physical mechanisms are clearly very different; however, we are interested in the scaling of the flows inside the modelling boxes

concerned only with the differences that may be present there. The physical model to be employed in the MB is that of ideal MHD (for an introduction to MHD, see [44]). While this is a good approximation for the astrophysical case, it is not usually the case for the experiments, which have to deal with nonideal effects arising, for example, from thermal conduction, viscosity and resistivity. Nevertheless the experiments can be *designed* so that the ideal MHD approximation is valid at least inside the MB, thus making the set of equations describing the plasma evolution in the two MB the same. For scaling purposes this is, however, not sufficient, it is also fundamental for the initial (and boundary) conditions in the two MB to be geometrically similar. Meaning that at some arbitrary time, the initial spatial distributions of all physical quantities (density ρ , velocity V , pressure P and magnetic field B) are the same. Finally, the actual scaling factors relating the physical variable in the two systems and their initial distributions should obey the following constraints:

$$\tilde{V}_L \sqrt{\frac{\tilde{\rho}_L}{\tilde{P}_L}} = \tilde{V}_A \sqrt{\frac{\tilde{\rho}_A}{\tilde{P}_A}}, \quad \frac{\tilde{B}_L}{\sqrt{\tilde{P}_L}} = \frac{\tilde{B}_A}{\sqrt{\tilde{P}_A}},$$

where the tilde denotes characteristic (dimensional) values for the laboratory (L) and astrophysical (A) plasmas. If the aforementioned conditions are met, namely the ideal MHD equations hold, the initial conditions are geometrically similar and the constraints on the scaling factors hold, then the evolution in the modelling boxes, from some initial time and over the scaled time span, will be *indistinguishable*. The addition of extra physical processes, for example, radiation, increases the constraints to be satisfied and hence the difficulty in obtaining an exact scaling (see, for example, [7]).

Two points need to be clarified. The first regards obtaining geometrically similar initial conditions in the laboratory. This is clearly very difficult and it has been done exactly in only a few cases [45]. The question also concerns which initial

condition should be used. Obviously observations cannot be expected to provide 3D distributions of all the quantities of interest, if they did the problem would be almost solved! Then, with observational constraints in mind, we resort to models, which are in general as difficult to reproduce. The usefulness of laboratory astrophysics, however, goes beyond strict scaling. By using ‘reasonable’ initial conditions, as it is usually done in numerical simulations, we can study in detail complex astrophysical phenomena in a repeatable and accessible manner. In addition the possibility of validating astrophysical codes using laboratory data is also very important. The second point is on the accuracy of the approximation of ideal MHD for laboratory plasmas, and more specifically the ones produced on z-pinch installations. Neglecting the effects associated with viscosity, thermal conduction and resistivity (the ideal MHD approximation) means that the Reynolds (Re), Peclet (Pe) and magnetic Reynolds (Re_m) numbers are $Re, Pe, Re_m \gg 1$. While for YSO jets these dimensionless numbers are very large (see Table 1), for laboratory jets they are generally many orders of magnitude smaller, thus calling into question the validity of the ideal MHD approximation. However, the values obtained in the experiments should be compared to astrophysical numerical simulations, where because of the finite accuracy of the numerical schemes employed, the ideal MHD approximation also breaks down. Unless included explicitly in the simulations, viscous and resistive dissipation and thermal conduction occur at the grid level through unphysical numerical truncation errors. Existing ideal MHD simulations of jets have dimensionless numbers in the range 50–1000, well within the reach of laboratory experiments. In some case, as for the Reynolds number, the laboratory values can be orders of magnitude larger. Thus laboratory astrophysics experiments can be very useful in studying complex and intrinsically nonlinear 3D phenomena, in regimes where accurate numerical simulations are particularly difficult to perform, and thus it represents a powerful complementary tool to astrophysical modelling. As a final justification: ‘Real experiments are also irreplaceable in providing new insights into subtle physics issues and in stirring the creative imagination of scientists’ [50].

4 Young Stellar Jets from z-Pinch Machines

In this section we shall discuss the z-pinch studies of astrophysical jets. It is worth mentioning that the time for an experiment to reach ‘maturity’: from the conception, design and first experiments to the modelling, data analysis and then the application to astrophysical models is of the order of 3–5 years, depending on its complexity. Considering that most of the experiments, both on laser and on z-pinch, are less than 10 years old, it should not come as a surprise that in many instances the applications to astrophysics are not yet fully developed or very clear. Although we shall restrict our discussion to jets, there are other areas of laboratory astrophysics research on z-pinch that are well developed, such as equations of state studies for planetary interiors and photoionized plasmas relevant to accretion disks around compact objects. These are thoroughly reviewed in [46] and we refer the interested reader to that paper.

Table 1 Characteristic conditions in laboratory (z-pinch) and YSO jets

	Laboratory	YSO
Fluid velocity [km s ⁻¹]	100–400	100–500
Density [g cm ⁻³]	10 ⁻⁴ –10 ⁻⁶	10 ⁻¹⁸ –10 ⁻²⁰
Temperature [eV]	5–200	0.5–100
Magnetic field [G]	10 ⁴ –10 ⁶	10 ⁻³ –10 ³
Dynamical age [ns]	200–400	10 ²²
Length [cm]	2–4	10 ¹⁷
Radius [cm]	0.5	10 ¹⁵
Mass flux [M _⊙ year ⁻¹]	10 ⁻³³	10 ⁻⁷ –10 ⁻⁸
Mean ionization	5–10	10 ⁻³ –1
Sound speed [km s ⁻¹]	10 ³ –10 ⁴	10 ³ –10 ⁴
Radiative cooling time [ns]	4–40	10 ¹⁸
Mean free path [cm]	10 ⁻⁵	10 ⁹
Magnetic diffusivity [cm ² s ⁻¹]	10 ⁴	10 ⁸
Kinematic viscosity [cm ² s ⁻¹]	10 ⁻³ –10	10 ¹⁴
Thermal diffusivity [cm ² s ⁻¹]	10 ³ –10 ⁶	10 ¹⁵
Mach number	5–40	>>5
Re _m	10–10 ³	>10 ¹⁵
Re	>10 ⁴	>10 ⁸
Pe	50–10 ⁴	>10 ⁷
Density contrast	0.1–10	>1
Cooling parameter	0.01–10	<1
Localization parameter	<10 ⁻⁴	<10 ⁻⁶
Plasma-β	0.01–100	0.01–100

4.1 Hydrodynamic Jets from Conical Wire Arrays

Conically converging flows were investigated in astrophysics as a possible mechanism for converting wide angle winds into collimated jets. Such models do not require magnetic fields, at least to collimate and launch the jet, and rely instead on purely hydrodynamic means [52, 6]. Within this framework and with the aim of producing hydrodynamic jets to be used for interaction studies, a series of experiments [31, 10] were developed on the z-pinch generator MAGPIE. The schematic of the experimental configuration is shown in Fig. 3. It consists of a conical array of micron-sized metallic wires driven by a current of 1 MA rising to its peak value in 240 ns. The basic mechanism of plasma formation in wire arrays is the following: resistive heating rapidly converts the wires into a heterogeneous structure consisting of a cold (< 1 eV) and dense liquid–vapour core, surrounded by a relatively hot (10–20 eV) and low-density ($\sim 10^{17}$ cm⁻³) plasma. Most of the current flows in the latter, where the resistivity is lower, which undergoes acceleration by the $\mathbf{J} \times \mathbf{B}$ force towards the array axis. These streams of plasma have characteristic velocities of ~ 100 – 150 km s⁻¹ and corresponding Mach numbers $M \sim 5$. The wire cores act as a reservoir of plasma, replenishing the streams during the entire duration of the experiment (several hundreds of nano-seconds). The converging plasma is virtually magnetically field-free and the interaction on axis is hydrodynamic in character. The collision produces a standing conical shock where part of the kinetic energy of the

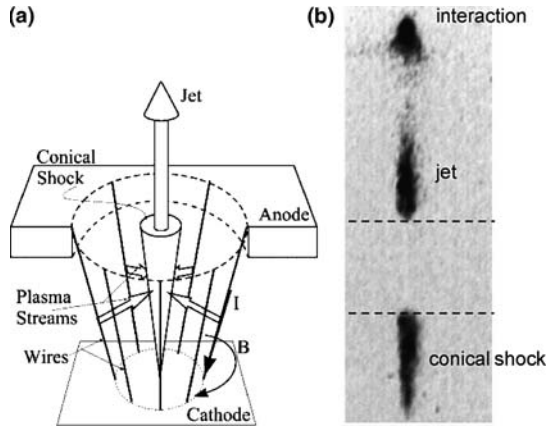


Fig. 3 (a) Typical arrays are made with 16 tungsten wires, each with a diameter of $18\ \mu\text{m}$. The smaller radius of the array is 8 mm and the wires are inclined at an angle of 30° with respect to the axis. The axial length of the array is 12 mm. Continuous plasma streams converge on axis producing a conical shock which redirects the flow axially. (b) Time-resolved, filtered XUV emission from a laboratory jet. Emission from the region in between the *dotted lines* is screened by the anode. As the jet propagates it cools down and its emission decays; however, it is again visible where the jet interacts with a background plasma

streams is thermalized. However, it is important to note that the plasma streams are not perpendicular to the surface of the conical shock. Thus the component of the velocity parallel to the shock is continuous across the shock and the flow is redirected upwards into a jet. Typical jet velocities are $\sim 100\text{--}200\ \text{km s}^{-1}$ and hypersonic jets with $M > 10$ can be produced by this mechanism. The jet collimation and Mach numbers depend predominantly on the level of radiation cooling in the plasma, which can be changed experimentally by varying the wire material (Al, Fe, W and so on). Increasing the atomic number of the wires increases the rate of cooling from the plasma, lowers its temperature and leads to the formation of more collimated jets (with higher Mach numbers) [10, 31, 32]. These jets are used to study the propagation and interaction with an ambient medium, which are described in Sect. 4.3. The characteristic conditions and dimensionless parameters obtained are shown in Table 1.

It is possible to design and modify the experiments to include additional physics, such as dynamically dominant magnetic fields and rotation. Indeed, for accretion onto the forming star to occur, angular momentum needs to be removed from the in-falling material. In combination with the processes present in the accretion disk, such as MHD instabilities and turbulent transport, jets and winds can also remove a considerable fraction of the excess angular momentum from the accreting flow [16]. One of the obvious implications is that jet will be rotating and some confirmation has arrived with recent observations of rotation in a number of YSO jets [15]. Supersonically rotating laboratory jets and flows of astrophysical relevance were recently produced for the first time [2] using a variant of the conical wire array. Rotation in

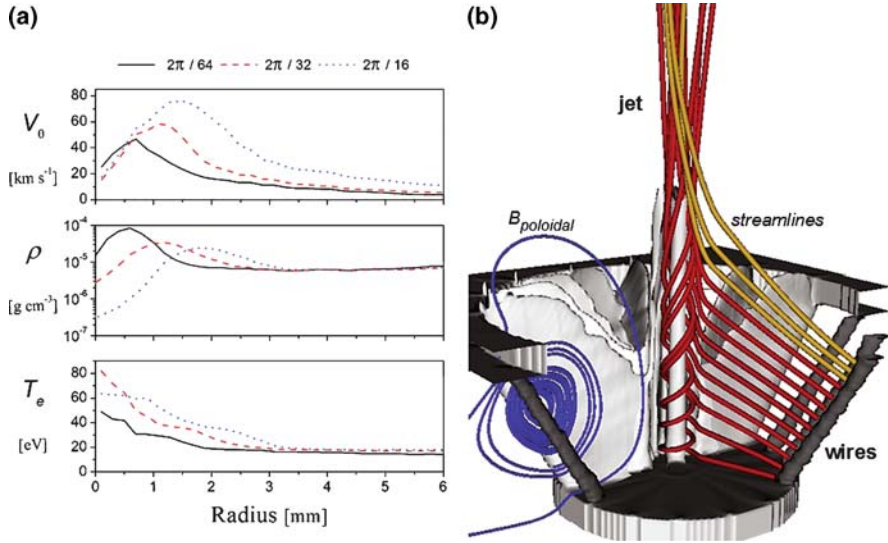


Fig. 4 The experimental set-up is similar to that shown in Fig. 3, in this case, however, the wires are twisted in the azimuthal direction (see main text). **(a)** Azimuthally averaged profiles of azimuthal velocity, density and electron temperature as a function of radial position, taken 6 mm above the cathode 200 ns after the start of the current pulse. The profiles are for different twist angles of the array, which can be used to modify the angular rotation in the jet. **(b)** Isodensity surfaces at 270 ns show the dense plasma around the wires ($10^{-3} \text{ g cm}^{-3}$, dark grey) and plasma streams ($2.5 \times 10^{-5} \text{ g cm}^{-3}$, light grey); velocity streamlines from one of the wires are shown with the oppositely rotating flows separated visually by red and orange streamlines (the top 3 of which represent the counter-rotating flow). The oppositely rotating flow is due to electrode's effects and it is dynamically unimportant. The azimuthally averaged poloidal magnetic field lines are shown in blue [2]

the flow is accomplished by slightly twisting the wires in the azimuthal direction. This results in a poloidal magnetic field and an azimuthal component of the Lorentz force, giving a nonzero torque on the plasma streams (Fig. 4). The level of angular momentum introduced in the system can be controlled by changing the twist angle and in general the jets ejected have rotation velocities $\sim 100\text{--}200 \text{ km s}^{-1}$, corresponding to $\sim 0.1\text{--}0.2$ of the jet propagation velocity. One of the applications of these proof-of-principle experiments will be to study the effects of rotation on the propagation of jets and on the growth of the Rayleigh–Taylor instability in curved jets (see Sect. 4.3)

4.2 Magnetohydrodynamic Jets

Protostellar (and galactic jets) are thought to be powered by the combination of rotation and magnetic fields, which extract the rotational energy from an accreting system and create magnetic stresses which accelerate and collimate the flow (see the lectures notes [23, 53]. Depending on the details of the models, the winding of

an initially poloidal magnetic field results in a flow pattern dominated by a toroidal field. A similar situation is also attained when the foot-points of a field line, connecting the disc to a central compact object or connecting different parts of a disc, rotate with different angular velocities. In such cases, the relative angular displacement of the foot-points causes one of them to move ahead of the other and the field loop to twist. The induced toroidal component results in an increase of the magnetic pressure which drives the expansion of the loop itself [34]. In the magnetic tower scenario [37, 36], the outcome is a magnetic cavity consisting of a highly wound-up toroidal field which accelerates the flow. In this case, the presence of an external plasma medium was shown to be necessary to confine the magnetic cavity, which would otherwise splay out to infinity within a few rotations [35]. The basic picture of magnetic tower evolution has also been confirmed numerically by several authors [28, 29, 41, 42, 39].

4.2.1 Magnetically Driven Jets from Radial Wire Arrays

The study of magnetically collimated and accelerated jets on z-pinches was developed in the last few years [9, 33, 11, 12] using a modified wire array configuration. The basic astrophysical mechanism studied in the experiments is the interaction of a toroidal magnetic field with a plasma ambient medium, leading to the formation of jets and magnetic ‘bubbles’. The schematic of the experimental set-up, a radial wire array, is shown in Fig. 5. The formation of plasma is similar to that discussed in conical wire arrays, however, the plasma is now accelerated vertically filling the space (few centimetres) above the array. Below the wires there is only a toroidal magnetic field. The formation of the jet and its time evolution is shown in Fig. 6. The initial formation of the magnetic cavity and jet occurs at the time when the magnetic pressure is large enough to break through the wires. This occurs only over a small region close to the central electrode, where the toroidal magnetic field B_G is strongest. The results show the system evolving into a structure consisting of an approximately cylindrical magnetic cavity with an embedded jet on its axis

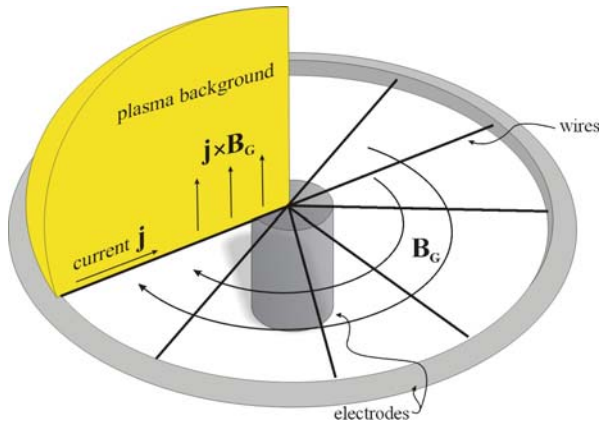


Fig. 5 Schematic of a radial wire array

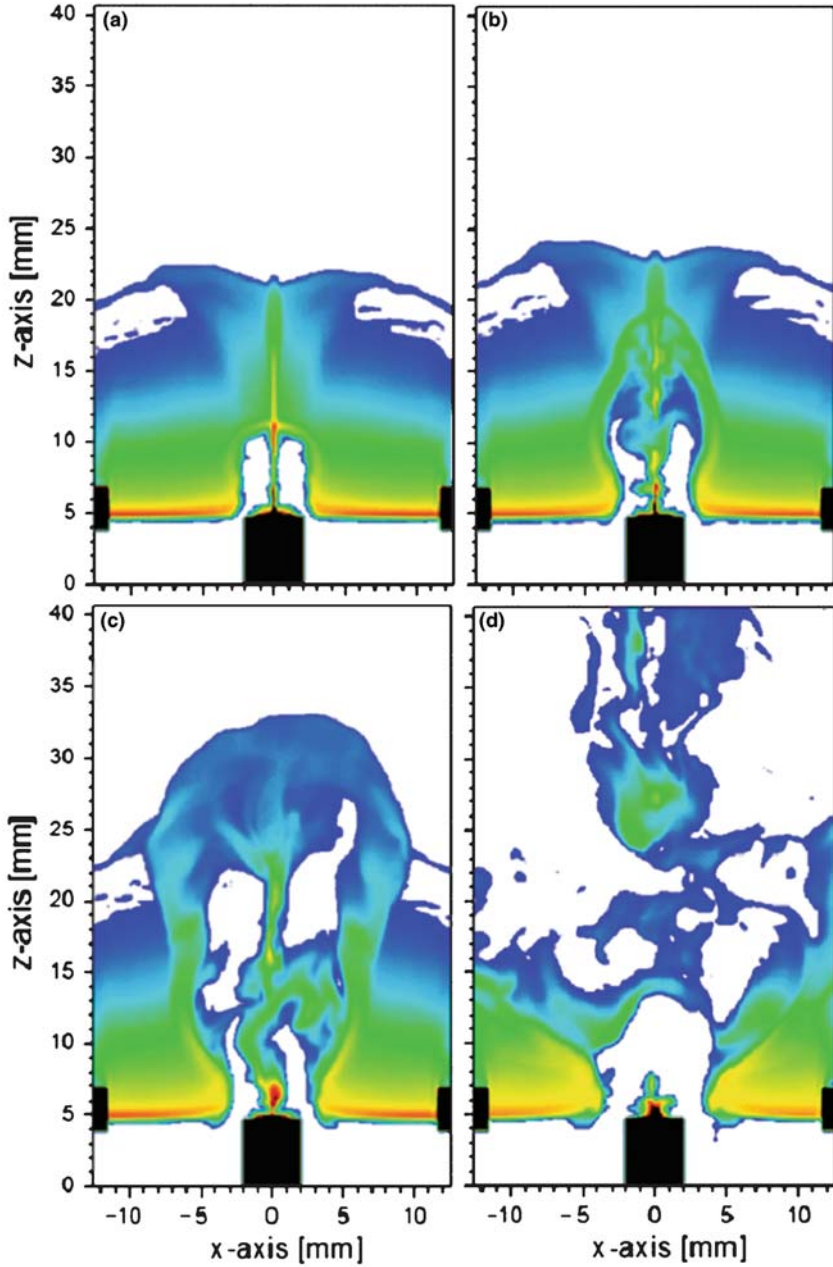


Fig. 6 Slices of mass density from a 3D MHD simulation of radial wire arrays. The evolution is shown at times (a) 165 ns, (b) 175 ns, (c) 185 ns and (d) 205 ns. The logarithmic density scale is from $10^{-7} \text{ g cm}^{-3}$ (blue or light grey in the black and white version) to $10^{-1} \text{ g cm}^{-3}$ (red or dark grey in the black and white version). Regions in white are void of plasma, but not electromagnetic fields, and are essentially a computational ‘vacuum’. The square, black regions are the electrodes [11]

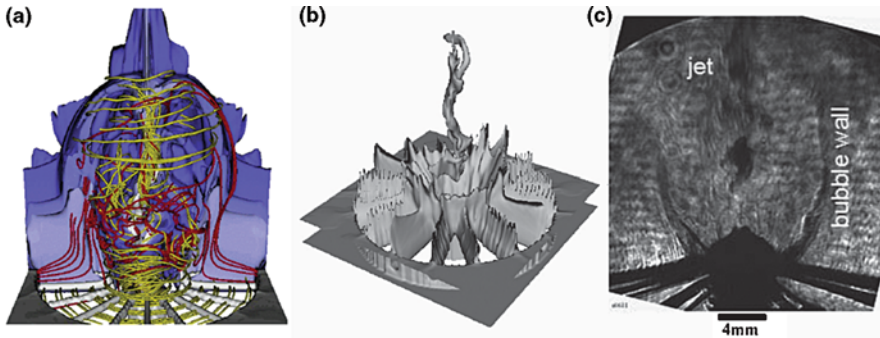


Fig. 7 (a) Magnetic field lines (yellow or light grey in the black and white version) and current density (red or dark grey in the black and white version) distribution inside the magnetic cavity at 245 ns [11]. To show the inside of the magnetic cavity the isodensity contours (shades of blue or grey in the black and white version) are sliced vertically. With the onset of the kink instability the magnetic field wraps tightly around the jet, which is seen more clearly in (b) without the cavity walls [12]. (c) Experimental shadowgraph showing the bubble wall and the clumpy jet launched along the axis [33]

confined by the magnetic ‘pinching’ force. A shell of swept-up plasma surrounds and partially confines the magnetic bubble. The subsequent evolution is dominated by current-driven instabilities and the development of the asymmetric ‘kink’ mode ($m = 1$) which leads to a distortion of the jet and a re-arrangement of the magnetic field. In Fig. 7a, the magnetic field lines can be seen to twist inside the jet, an effect caused by the instability which turns toroidal into poloidal magnetic flux. The end result of the instabilities, however, is not to destroy the jet, but to produce an inhomogeneous or ‘knotty’ jet, shown in Fig. 7b, c. The resulting jet has typical super-fast magnetosonic Mach numbers in excess of 5, it is kinetically dominated and its opening angle $< 20^\circ$.

The relatively simple initial conditions implemented experimentally produce a very complex and rich dynamics which share many important features with astrophysical models. One important example is the presence of an envelope surrounding the magnetic cavity and confining it. Although this is discussed in the astrophysical literature [37, 54], it has so far only been observed in a laboratory experiment. The stability and dynamics of the envelope, which determine the collimation of the cavity itself, can thus be directly studied in the laboratory before astronomical observations may become available. Finally, it is worth pointing out that while 2D, axisymmetric MHD simulations reproduce very well the experimental results, up to the development of the nonaxisymmetric current-driven instabilities. There are fundamental differences in the long-term evolution of the system, which can only be reproduced by fully 3D simulations.

4.2.2 Episodic Ejection of Magnetic Bubbles and Jets

Protostellar jets are characterized by the presence of knots and multiple bow-shocks, tracing their propagation [25]. These are often interpreted as internal shocks driven by relatively small perturbations in a steady ejection process, and which occur on

typical timescale between ~ 5 and 20 years. For example in [18] it was shown that the temporal variability of the jet velocity may be associated with a time-varying stellar magnetic field. Episodic jet ejection behaviour may also be associated with variation in the accretion rates or an inflating stellar magnetosphere.

Recent experiments have studied for the first time the episodic ejection of magnetic bubbles and jets, and its effects on the overall propagation of the outflow[13]. The experimental set-up is similar to that shown in Fig. 4; however, the wires are replaced by a 6- μm thick metallic foil (usually aluminium). A 3D MHD simulation of the experiments is shown in Fig. 8. The evolution of the first bubble is similar to that of radial wire arrays. However, the total mass in the plasma source, as a function of radius, is larger for a foil than for a radial wire array. Thus after the first magnetic cavity and jet are formed, there is a larger quantity of plasma available to refill the ‘gap’ between the central electrode and the left-over foil; the presence of this gap is visible for example in Fig. 5 for radial arrays, and it is produced by the magnetic field pressure breaking through the wires (or foil). Once the gap is refilled with plasma, the currents can flow once again across the base of the magnetic cavity, thus re-establishing the initial configuration. When the magnetic pressure is large enough to break through this newly deposited mass, a new jet/bubble ejection cycle can begin. Typical flow velocities observed are $\sim 100\text{--}400\text{ km s}^{-1}$, the simulated sonic and the alfvénic Mach numbers in the jet, defined as the ratios of the flow speed to the sound and Alfvén speed, respectively, are $M_s \sim M_A \sim 3\text{--}10$. The resulting flow is heterogeneous and clumpy, and it is injected into a long lasting and well-collimated channel made of nested cavities. Each jet/outflow episode propagates, interacts and substantially alters the surrounding environment by injecting mass, momentum, energy and magnetic flux into it. An important aspect of the episodic ejection process is, broadly speaking, its self-collimation. Since the initial ambient medium is swept away after a few ejections, newly formed magnetic cavities are confined solely by the environment left by earlier episodes, thus making the collimation process insensitive to the initial ambient conditions. An experimental

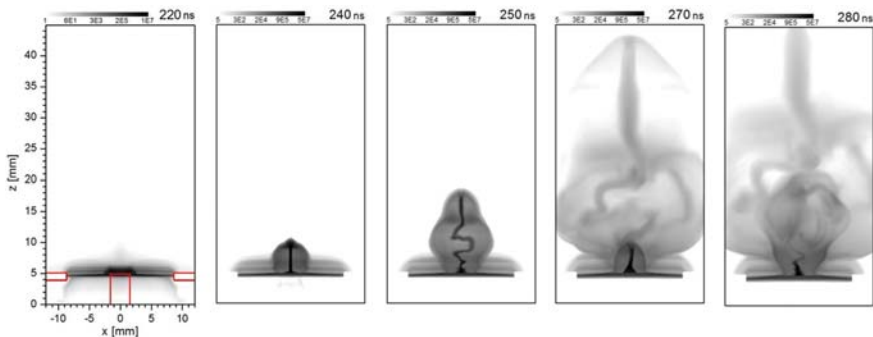


Fig. 8 Three-dimensional MHD simulation of a radial foil experiment. The images show the line-of-sight integrated emission (in arbitrary units) at different times. The red line (black in the black and white version) in the first panel shows the position of the electrodes

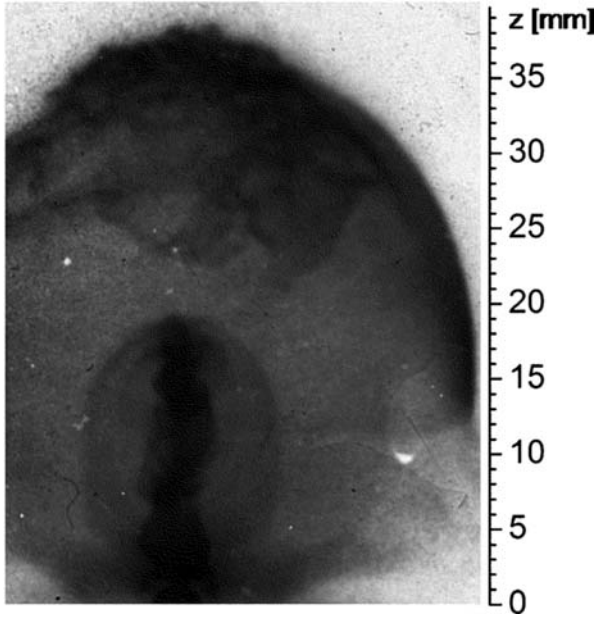


Fig. 9 Experimental image showing the presence of nested magnetic cavities with an embedded jet. The image shows the self-emission in the XUV range [13]

image of the evolution of the system is shown in Fig. 9. In the magnetic cavities $Re_M > 100$, and each bubble expands with its own ‘frozen-in’ magnetic flux; in the experiments this is confirmed by the magnetic probe measurements of the trapped magnetic field at the outer edge of the bubbles, $B \sim 1\text{--}5$ kG. The collimation is then determined not only by the pressure of the left-over plasma but also by the pressure of the tangled magnetic field trapped in the bubbles, where the plasma- β is in the range $0.1 < \beta < 1$. A high level of symmetry is maintained after many ejections (5 in the current experiments), the number being limited only by the duration of the current pulse delivered by the generator. Overall, the experiments demonstrate that magnetic acceleration and collimation, occurring within a framework of strongly episodic outflow activity, can be effective in producing well-collimated and heterogeneous jets.

By drawing a parallel with the dynamics observed in the experiments, one can gain useful insights and a qualitative view of the possible evolution of astrophysical jets. In the experiments there are two timescales which determine the magnetic bubbles/jets development: the current-driven (CD) instability timescale τ_I and the episodic bubble ejection timescale τ_B . For conditions applicable to the formation region of protostellar jets [26], we can estimate the growth time of the CD kink mode as the Alfvén crossing time $\tau_I \sim 1$ year; corresponding to a few nanoseconds in the experiments. The second timescale is linked to the temporal variability of the Poynting flux feeding the bubbles, and for astrophysical sources τ_B should be associated with a substantial variation in the outflow launching activity; observations

of knots kinematics suggest characteristic times $\tau_B \sim 5\text{--}20$ years; the experiments are in a similar regime of τ_B/τ_I . Because both timescales are relatively longer than the characteristic Keplerian period of rotation at the inner disk radius, jet launching should have ample time to reach steady state. The characteristic astrophysical flow velocities can be taken to be $v \sim 200 \text{ km s}^{-1}$. With these conditions, the presence of multiple bubble-like features should be observed on scales ranging from a few tens to a few hundreds of AU from the source. Indeed ejection variability, limb-brightened bubble-like structures and the presence of wiggles in the optical DG Tau jet are evident on scales ranging from a few tens to a few hundreds of AU the source [4, 19]. The experiments also indicate that asymmetries in the flow can be produced by instabilities that do not destroy the collimation, and because of their relatively short growth time, jets should develop nonaxisymmetric features already within a few tens of AU from the source and become more heterogeneous and clumpy as they move further away to hundreds of AU. It was recently reported for a number of TTauri jets, including DG Tau, that already within 100 AU from the source the jet's physical conditions show considerable asymmetries with respect to the axis [14]. Finally over the same length scales the experiments suggest magnetic energy dissipation, heating of the plasma and a transition to a kinetically dominated jet which propagates ballistically. X-ray emission from the DG Tau jet was recently detected on the same length scales and it was proposed that magnetic energy dissipation may be behind the heating mechanism [24]. As in the experiments, instabilities and the tangling of the magnetic field may provide a compelling route to efficient heating of such plasmas.

4.3 Interaction with the Interstellar Medium

4.3.1 Curved Jets

A number of bipolar Herbig-Haro (HH) jets exhibit a distinguishing C-shape morphology indicative of a steady bending [5]. Less regular curvature is also observed in a number of other HH jets; for example, in HH 30 a small side drift close to the jet source is followed further away by a sudden bending [3]. In general, the curvature in jets has been linked either with the motion of the jet sources relative to the ambient medium or with the presence of a widespread outflow; both cases giving rise to an effective transverse ram pressure (cross-wind) which curves the jet. Expected wind velocities vary from a few kilometres per second for the jet-wind interaction associated with relative motions of TTauri stars with respect to the surrounding environment (see, for example, [27]) to typically higher velocities for irradiated jets, where best fits to HH505 H α emission maps were obtained for a wind velocity of 15 km s^{-1} [38] and estimates in [5] give wind velocities in the Orion nebula and in NGC1333 of $\sim 10\text{--}20 \text{ km s}^{-1}$.

The hydrodynamic laboratory jets described in Sect. 4.1 are ideally suited to study the interaction with an ambient medium. The region into which the jet is launched (cf. Fig. 3) is a large vacuum that can be easily filled with different types

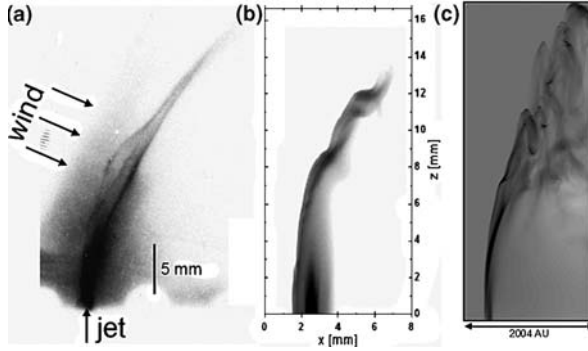


Fig. 10 (a) Experimental time-resolved XUV image of a curved jet. (b) Synthetic XUV image of a curved laboratory jet from a 3D simulation. (c) Column density from a 3D simulation of an astrophysical jet [8]

of background gases. To investigate the dynamics of curved HH jets, a cross-wind was produced by a radiatively ablated foil appropriately placed in the jet propagation region [1, 32, 30]. Typical wind velocities $\sim 30\text{--}50 \text{ km s}^{-1}$ can be produced in the laboratory, with the important parameters characterizing the interaction in the range $V_{\text{jet}}/V_{\text{wind}} \sim 2\text{--}4$ and $n_{\text{jet}}/n_{\text{wind}} \sim 0.1\text{--}10$. Figure 10 shows an example of the experiments and simulations of curved jets [8]. The characteristic dynamics of the interaction is similar for the laboratory and astrophysical systems, showing notably the formation of new working surface in the jet and ‘knotty’ structure in the flow. Curved jets are also Rayleigh–Taylor (RT) unstable, with the growth of such mode disrupting their propagation. However, it was shown that jet rotation may partially suppress the instability by shearing the RT modes and confining the perturbations to a narrower layer of the jet body. Nevertheless this promotes the development of the Kelvin–Helmholtz instability (at least for the subsonically rotating jets) which is later responsible for disrupting the jet. Experimentally the RT growth time is of the order of the dynamical time over which the interaction can be produced and new, longer timescale experiments will be needed to observe its full development.

4.3.2 Clump Propagation

We now return to the scaling issue of some of the laboratory flows and in particular the MHD jets. As we have seen their evolution is dominated by current-driven instabilities, and the resulting flow is inherently time dependent and inhomogeneous. To study the propagation of such flows in an astrophysical setting, the data obtained from laboratory MHD jet simulations can be scaled up and used as initial conditions to model astrophysical clumpy jets. There is clearly some arbitrariness on the choice of some of the scaling parameters and for the case presented here we assume the flow to be close to the YSO source. Noting that the laboratory and astrophysical jet velocities are of the same order, we choose the following three scaling: $V_{\text{lab}} = V_{\text{astro}}$,

$L_{\text{lab}} = 1 \text{ mm} \rightarrow L_{\text{astro}} = 10 \text{ AU}$ and $\rho_{\text{lab}} = 10^{-3} \text{ g cm}^{-3} \rightarrow \rho_{\text{astro}} = 10^{-18} \text{ g cm}^{-3}$. The choice of scale length gives an initial jet radius $\sim 20\text{--}30 \text{ AU}$ and we take the maximum jet density in the laboratory to scale to a maximum astrophysical jet density of $\sim 10^6 \text{ cm}^{-3}$. The constraints on the scaling discussed in Sect. 3 give $1 \text{ ns} \rightarrow 0.05 \text{ years}$, $50 \text{ eV} \rightarrow 3000 \text{ K}$ and $50 \text{ T} \rightarrow 15 \text{ mG}$. The specific scaling applied

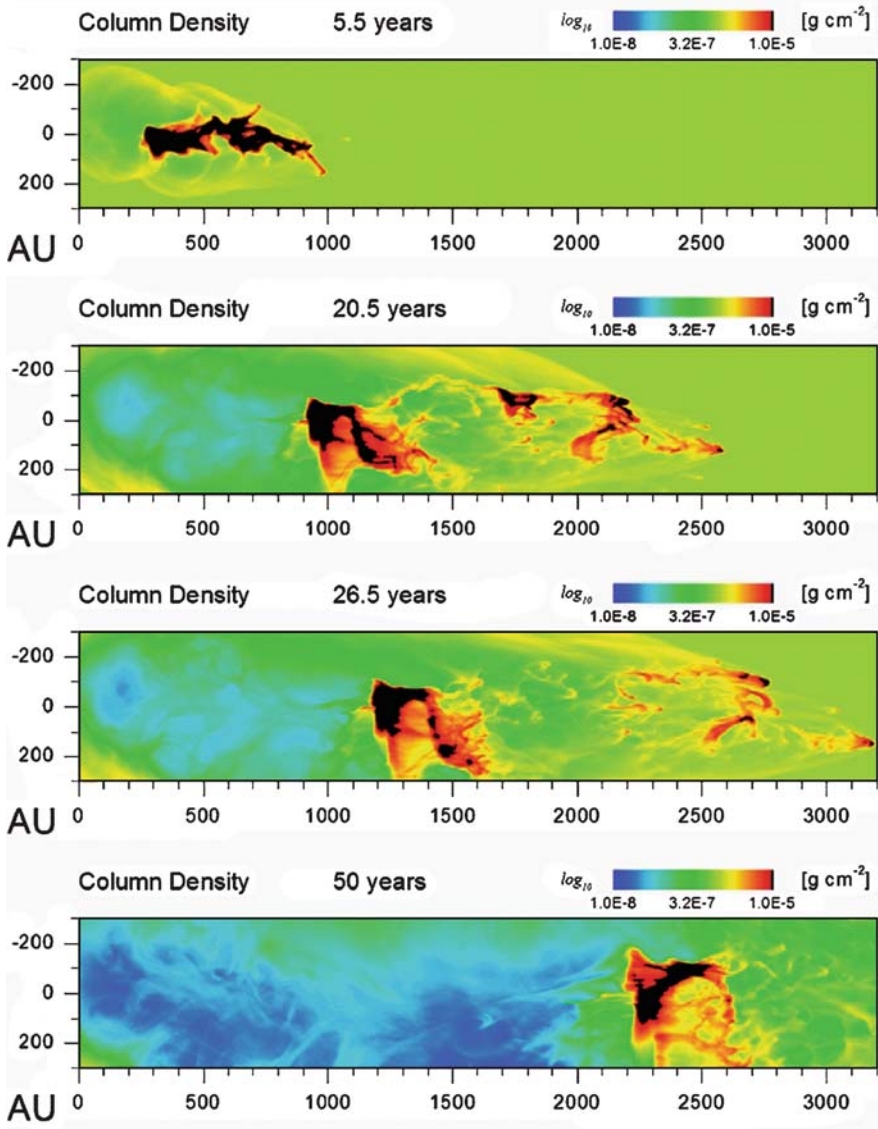


Fig. 11 Propagation of an episodic protostellar jet showing the break-up into small and large clumps

in this case implies that the experimental flow, which lasts ~ 200 ns, corresponds to an astrophysical outflow lasting ~ 10 years. Such short times, when compared to the lifetimes of tens of thousands of years for protostellar outflows, may correspond to the ejection of a single ‘clump’ as part of a more extended jet. The laboratory jet profiles to be scaled up are taken at a time approximately corresponding to the image in Fig. 7b. The simulated astrophysical jet was evolved with the inclusion of cooling over ~ 50 years on a Cartesian grid of 400×10^6 cells with a resolution of 2 AU. In these simulations the magnetic field is not included, and for the regime modelled here we would expect their inclusion to modify somewhat the dynamics. The flow dynamics is shown Fig. 11. Initially the jet elongates because of the velocity variations imparted by the current-driven instability. The stretching of the jet is then followed by a rapid break-up into smaller clumps which move at different velocities. The structure appearing in the large knot forming in the flow appears to be the result of Rayleigh–Taylor instabilities. In general, a single dense clump is produced by an ejection event like this, with the resulting outflow remaining well collimated over the propagation across ~ 3000 AU.

5 Summary

Progress in high-energy density plasma experiments on lasers and z-pinch facilities has permitted in the last 10 years to start investigating a range of ‘large-scale’ astrophysical phenomena in the laboratory, extending the traditional domain of laboratory astrophysics beyond the work on micro-physics. Through careful design of the experiments, the plasma produced can be scaled to the astrophysical environment, allowing complex, intrinsically nonlinear, 3D phenomena to be accessed in a controlled manner. An important outcome being the validation of astrophysical codes on the laboratory data. Although work on astrophysical jets has been performed on both lasers and z-pinch facilities, we have focused here only on the studies of jets produced on the MAGPIE z-pinch facility. Two main ‘types’ of jets were developed: hydrodynamic jets, to be used for propagation studies, and magnetohydrodynamic jets of interest to the launching phase. In general, there is some considerable control on the experiments: the initial condition can be partially modified, for example, the density and magnetic field distributions; more complex physics, such as rotation, can be added and different plasma condition can be produced by modifying, for example, the cooling rates. Overall the combination of laboratory experiments and simulations can provide some very important insights on the physics of astrophysics, and as technology advances we can expect evermore exotic phenomena to be reproduced in the laboratory.

Acknowledgement I would like to thank C. Stehlé (Observatoire de Paris), S.V. Lebedev (Imperial College) and A. Frank (University of Rochester) for many useful discussions. This work was supported in part by the European Community’s Marie Curie Actions-Human Resource and Mobility within the JETSET (Jet Simulations Experiments and Theory) network under contract RTN-CT-2004 005592. Access to the Marenostrum supercomputer, at the Barcelona Supercomputing

Centre (Spain), was granted through the HPC-EUROPA project (RII3-CT-2003-506079), with the support of the European Community – Research Infrastructure Action under the FP6 ‘Structuring the European Research Area’ Programme. Finally, the author acknowledges the London e-Science Centre (LESC) for the provision of computational facilities and support.

References

1. Ampleford, D. J., et al.: Laboratory modeling of standing shocks and radiatively cooled jets with angular momentum. *Astrophys. Space Sci.* **307**, 51–56 (2007). 45
2. Ampleford, D. J., et al.: Supersonic radiatively cooled rotating jets and jets in the laboratory. *Phys. Rev. Lett.* **100**, 35001 (2008). 37, 38
3. Anglada, G., et al.: Proper motions of the jets in the region of hh 30 and hl/xz tau: Evidence for a binary exciting source of the hh 30 jet. *Astron. J.* **133**, 2799–2814 (2007). 44
4. Bacciotti, F., et al.: Hubble space telescope stis spectroscopy of the optical outflow from dg tauri: Structure and kinematics on subarcsecond scales. *Astrophys. J.* **537**, L49–L52 (2000). 44
5. Bally, J., Reipurth, B.: Irradiated herbig-haro jets in the orion nebula and near ngc 1333. *Astrophys. J.* **546**, 299–323 (2001). 44
6. Canto, J., et al.: The formation of interstellar jets by the convergence of supersonic conical jets. *Astron. Astrophys.* **192**, 287–294 (1988). 36
7. Castor, J. I.: Astrophysical radiation dynamics: The prospects for scaling. *Astrophys Space Sci.* **307**, 207–211 (2007). 34
8. Ciardi, A., et al.: Curved herbig-haro jets: Simulations and experiments. *Astrophys. J.* **678**, 968–973 (2008). 45
9. Ciardi, A., et al.: Modeling magnetic tower jets in the laboratory. *Astrophys. Space Sci.* **298** (1–2), 277–286 (2005). 39
10. Ciardi, A., et al.: Modeling of supersonic jet formation in conical wire array z-pinch. *Laser Part. Beams.* **20**(2), 255–262 (2002). 36, 37
11. Ciardi, A., et al.: 3d mhd simulations of laboratory plasma jets. *Astrophys. Space Sci.* **307**(1), 17–22 (2007). 39, 40, 41
12. Ciardi, A., et al.: The evolution of magnetic tower jets in the laboratory. *Physics of Plasmas.* **14**, 056501 (2007). 39, 41
13. Ciardi, A. et al.: Episodic Magnetic Bubbles and Jets: Astrophysical Implications from Laboratory Experiments. submitted to *ApJ* (ArXiv 0811.2736), November 2008. 42, 43
14. Coffey, D., et al.: T tauri jet physics re-solved near the launching region with the hubble space telescope. to be published in *The Astrophysical Journal*, 2008. 33 pages, 16 figures, accepted by *ApJ*. 44
15. Coffey, D., et al.: Rotation of jets from young stars: New clues from the hubble space telescope imaging spectrograph. *Astrophys. J.* **604**, 758–765 (2004). 37
16. Combet, C., J. Ferreira. The radial structure of protostellar accretion disks: influence of jets. *Astron. Astrophys.* **479**, 481–491 (2008). 37
17. da Silva, L. B., et al.: Absorption measurements demonstrating the importance of $\delta n = 0$ transitions in the opacity of iron. *Phys. Rev. Lett.* **69**, 438–441 (1992). 31
18. De Colle, F., et al.: The effect of a stellar magnetic variation on the Jet velocity. *Astrophys. J.* **688**, 1137–1141, December 2008. 42
19. Dougados, C., et al.: T tauri stars microjets resolved by adaptive optics. *Astron. Astrophys.* **357**, L61–L64 (2000). 44
20. Drake, R. P.: The design of laboratory experiments to produce collisionless shocks of cosmic relevance. *Physics of Plasmas.* **7**, 4690–4698 (2000). 33
21. Drake, R. P.: *High-Energy-Density Physics: Fundamentals, Inertial Fusion, and Experimental Astrophysics*. Springer, Berlin (2006). 32

22. Egeland, A., Burke, W. J.: Kristian Birkeland, the First Space Scientist. In *Astrophysics and Space Science Library*, vol. 325, 2005. Springer, Dordrecht. 31
23. Ferreira, J.: Mhd Disc Winds. In *Lecture Notes in Physics*, vol. 723, p.181. Springer Verlag, Berlin (2007). 38
24. Gdel, M., et al.: Discovery of a bipolar x-ray jet from the t tauri star dg tauri. *Astron. Astrophys.* **478**, 797–807 (2008). 44
25. Hartigan, P., et al.: Proper motions of the HH 47 jet observed with the hubble space telescope. *Astron. J.* **130**, 2197–2205, November 2005. 41
26. Hartigan, P., et al.: Magnetic fields in stellar jets. *Astrophys. J.* **661**, 910–918 (2007). 43
27. Jones, B. F., Herbig, G. H.: Proper motions of t tauri variables and other stars associated with the taurus-auriga dark clouds. *Astron. J.* **84**:1872–1889 (1979). 44
28. Kato, Y., et al.: Formation of semirelativistic jets from magnetospheres of accreting neutron stars: Injection of hot bubbles into a magnetic tower. *Astrophys. J.* **600**:338–342 (2004). 39
29. Kato, Y., et al.: Magnetohydrodynamic accretion ows: Formation of magnetic tower jet and subsequent quasi-steady state. *Astrophys. J.* **605**(1):307–320 (2004). 39
30. Lebedev, S. V. et al.: Jet detection via crosswinds: Laboratory astrophysical studies. *Astrophys. J.* **616**:988–997 (2004). 45
31. Lebedev, S. V., et al.: Laboratory astrophysics and collimated stellar outflows: The production of radiatively cooled hypersonic plasma jets. *Astrophys. J.* **564**(1), 113–119 (2002). 36, 37
32. Lebedev, S. V., et al.: Production of radiatively cooled hypersonic plasma jets and links to astrophysical jets. *Plasma Phys. Contr. Fusion.* **47**, 465–B479 (2005). 37, 45
33. Lebedev, S. V., et al.: Magnetic tower outflows from a radial wire array z-pinch. *Mon. Not. R. Astron. Soc.* **361**, 97–108 (2005). 39, 41
34. Lovelace, R. V. E., et al.: Spin-up/spin-down of magnetized stars with accretion discs and outflows. *Mon. Not. R. Astron. Soc.* **275**, 244–254 (1995). 39
35. Lynden-Bell, D.: Magnetic collimation by accretion discs of quasars and stars. *Mon. Not. R. Astron. Soc.* **279**(2), 389–401 (1996). 39
36. Lynden-Bell, D.: On why discs generate magnetic towers and collimate jets. *Mon. Not. R. Astron. Soc.* **341**(4):1360–1372 (2003). 39
37. Lynden-Bell, D.: Magnetic jets from swirling discs. *Mon. Not. R. Astron. Soc.* **369**, 1167–1188 (2006). 39, 41
38. Masciadri, E., Raga, A. C.: A jet-side wind interaction model for the curved jets in the orion nebula. *Astron J.* **121**, 408–412 (2001). 44
39. Matt, S., et al.: Astrophysical explosions driven by a rotating, magnetized, gravitating sphere. *Astrophys. J.* **647**:L45–L48 (2006). 39
40. Mottelay, P. F.: William Gilbert of Colchester, Physician of London, On the Load Stone and Magnetic Bodies, and on the Great Magnet the Earth. J. Wiley and Sons, New York (1893). 31
41. Nakamura, M. et al.: Structure of magnetic tower jets in stratified atmospheres. *Astrophys. J.* **652**, 1059–1067 (2006). 39
42. Nakamura, M. et al.: Stability properties of magnetic tower jets. *Astrophys. J.* **656**, 721–732 (2007). 39
43. O’C Drury, L., Mendonca, J. T.: Explosion implosion duality and the laboratory simulation of astrophysical systems. *Physics of Plasmas.* **7**, 5148–5152, December 2000. 33
44. Pelletier, G.: Introduction to Magneto-Hydrodynamics. In *Lecture Notes in Physics*, vol. 723, p 77, Springer Verlag, Berlin (2007). 34
45. Remington, B. A. et al.: Supernova hydrodynamics experiments on the nova laser. *Physics of Plasmas.* **4**, 1994–2003 (1997). 34
46. Remington, B. A. et al.: Experimental astrophysics with high power lasers and z pinches. *Reviews of Modern Physics*, **78**:755–807 (2006). 32, 35
47. Ryutov, D. et al.: Similarity criteria for the laboratory simulation of supernova hydrodynamics. *Astrophys. J.* **518**, 821–832 (1999). 33

48. Ryutov, D. D., et al.: The physics of fast z pinches. *Reviews of Modern Physics*. **72**, 167–223 (2000). 32
49. Ryutov, D. D., et al.: Criteria for scaled laboratory simulations of astrophysical mhd phenomena. *Astrophys. J. Supplement Series*. **127**, 465–468 (2000). 33
50. Ryutov, D. D., Remington, B. A.: Scaling astrophysical phenomena to high-energy-density laboratory experiments. *Plasma Phys Contr Fusion*. **44**, 407 (2002). 33, 35
51. Ryutov, D. D., Remington, B. A.: A “perfect” hydrodynamic similarity and effect of the Reynolds number on the global scale motion. *Physics of Plasmas*. **10**, 2629–2632 (2003). 33
52. Tenorio-Tagle, G., et al.: The formation of interstellar jets. *Astron. Astrophys.* **202**, 256–266 (1988). 36
53. Tsinganos, K.: Theory of mhd Jets and Outflows. In *Lecture Notes in Physics*, vol. 723, p. 117, Springer Verlag, Berlin (2007). 38
54. Uzdensky, D. A., MacFadyen, A. I.: Stellar explosions by magnetic towers. *Astrophys. J.* **647**, 1192–1212 (2006). 41
55. van-de Hulst, H. C., Burgers, J. M., (eds.) *Gas Dynamics of Cosmic Clouds*. North-Holland Publishing Company, Amsterdam (1955). 32
56. Zel’Dovich, Ya. B., Raizer, Yu. P.: *Physics of shock waves and high-temperature hydrodynamic phenomena*. 1966/1967, Hayes, W. D.; Probstein, R. F (eds.). Academic Press, New York (1967). 33

Output from MHD Models

Nektarios Vlahakis

Abstract Outflows emanating from the environment of stellar or galactic objects are a widespread phenomenon in astrophysics. Their morphology ranges from nearly spherically symmetric winds to highly collimated jets. In some cases, e.g., in jets associated with young stellar objects, the bulk outflow speeds are nonrelativistic, while in others, e.g., in jets associated with active galactic nuclei or gamma-ray bursts, it can even be highly relativistic. The main driving mechanism of collimated outflows is likely related to magnetic fields. These fields are able to tap the rotational energy of the compact object or disk, accelerate, and collimate matter ejecta. To zeroth order these outflows can be described by the highly intractable theory of magnetohydrodynamics (MHD). Even in systems where the assumptions of zero resistivity (ideal MHD), steady state, axisymmetry, one fluid description, and polytropic equation of state are applicable, the problem remains difficult. In this case the problem reduces to only two equations, corresponding to the two components of the momentum equation along the flow and in the direction perpendicular to the magnetic field (transfield direction). The latter equation is the most difficult to solve, but also the most important. It answers the question on the degree of the collimation, but also crucially affects the solution of the first, the acceleration efficiency and the bulk velocity of the flow. The first and second parts of this chapter refer to nonrelativistic and relativistic flows, respectively. These parts can be read independently. In each one, the governing equations are presented and discussed, focusing on the case of flows that are magnetically dominated near the central source. The general characteristics of the solutions in relation to the acceleration and collimation mechanisms are analyzed. As specific examples of exact solutions of the full system of the MHD equations that satisfy all the analyzed general characteristics, self-similar models are presented.

N. Vlahakis (✉)

Section of Astrophysics, Astronomy & Mechanics, Department of Physics, University of Athens, Panepistimiopolis, 15784 Zografos Athens, Greece, vlahakis@phys.uoa.gr

1 Steady, Axisymmetric, Nonrelativistic, Magnetized Outflows

A magnetized flow can be described by

- The density of the flow ρ . We assume that the plasma consists of positive and negative charges with masses per particle m_+ and m_- and number densities n_+ and n_- , respectively.¹ The mass density is then $\rho = n_+m_+ + n_-m_-$. The charge density is $\rho_q = n_+q_+ + n_-q_-$, where q_+ and q_- are the positive and negative charges (in most astrophysical cases $q_+ = e$, $q_- = -e$). Note that the plasma is assumed quasi-neutral, meaning that $(n_+ - n_-)/(n_+ + n_-) \ll 1$, but not exactly neutral. The charge density ρ_q is connected to the appearance of the electric field through Gauss' law (the electric field is small in the nonrelativistic case, as will be explained later, but in general nonzero).
- The flow velocity \mathbf{V} is the velocity of the center of mass. If \mathbf{V}_+ and \mathbf{V}_- are the velocities of the positively and negatively charged particles, respectively, then $\mathbf{V} = (n_+m_+\mathbf{V}_+ + n_-m_-\mathbf{V}_-)/(n_+m_+ + n_-m_-)$. This is the so-called bulk velocity of the fluid and should not be confused with the random motion of particles in the comoving frame (the frame in which the momentum of the flow vanishes; note that a different comoving frame is defined at each point of the flow). The random motion of particles is connected to the thermodynamic quantities, internal energy, temperature, and pressure. In reality the positive and negative charges move with slightly different velocities \mathbf{V}_+ and \mathbf{V}_- . Both are very close to the velocity of the center of mass \mathbf{V} . The difference between the velocities of positive and negative charges is connected to the current density $\mathbf{J} = n_+q_+\mathbf{V}_+ + n_-q_-\mathbf{V}_-$, which, in principle, is nonzero (its value is connected to the magnetic field through Ampère's law).
- The magnetic field \mathbf{B} and the electric field \mathbf{E} , as we measure them in the frame of the central object. This frame is also called the laboratory frame.
- The current density \mathbf{J} and the charge density ρ_q , as we measure them in the laboratory frame.
- The thermodynamic quantities pressure P , internal energy per mass e , and temperature T (these are defined in the comoving frame). The temperature is a given function of density and pressure $k_B T = P/(n_+ + n_-)$ (we assume one-temperature flow). For example, in electron–proton plasma where $n_- \approx n_+$ and $\rho \approx n_+m_p$, we get $k_B T = (m_p/2)(P/\rho)$. For electron–positron plasma where $n_- \approx n_+$ and $\rho \approx 2n_+m_e$ we get $k_B T = m_e(P/\rho)$.

The above quantities are connected through equations that we describe next.

¹ A generalization in the case of different types of positive/negative charges is straightforward. However, the present analysis does not cover the cases where neutrals are also present in the flow and have significantly different velocities compared to the charged components.

1.1 Governing Equations

1.1.1 Maxwell Equations

The electric and magnetic fields are connected to the charge and current densities through Gauss law

$$\rho_q = \frac{1}{4\pi} \nabla \cdot \mathbf{E}, \quad (1)$$

and Ampère's law

$$\mathbf{J} = \frac{c}{4\pi} \nabla \times \mathbf{B} - \frac{1}{4\pi} \frac{\partial \mathbf{E}}{\partial t}. \quad (2)$$

The previous two equations can be seen as the ones that give the charge and current densities as functions of (\mathbf{B}, \mathbf{E}) .

When we discuss the relation between fields (\mathbf{B}, \mathbf{E}) and charge and current densities (ρ_q, \mathbf{J}) , it is always correct to say that the (\mathbf{B}, \mathbf{E}) coexist with (ρ_q, \mathbf{J}) , in a way that the Gauss and Ampère laws are satisfied. However, it is not correct to say that (ρ_q, \mathbf{J}) create the fields (\mathbf{B}, \mathbf{E}) . If we want to answer the question of what creates the other, it is better to say the opposite. There is a large number of charges in the plasma and the appearance of the fields moves them in a way such as to create the necessary charge and current densities (ρ_q, \mathbf{J}) through small deviations in number densities $(n_+ - n_-)$ and velocities $(\mathbf{V}_+ - \mathbf{V}_-)$. Note also that a problem of a magnetohydrodynamic outflow in astrophysics is a boundary-value problem. The electromagnetic field inside the flow should satisfy the given boundary conditions. Thus, it is better to work with fields rather than with charge and current densities. The use of the current density help us to better understand some mechanisms, though. Examples are the acceleration and the collimation in astrophysical jets, as will be explained later.

The (\mathbf{B}, \mathbf{E}) fields should also obey the homogeneous Maxwell's equations, namely the solenoidal condition

$$\nabla \cdot \mathbf{B} = 0 \quad (3)$$

(which means that no magnetic monopoles can exist) and the Faraday's equation

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = 0. \quad (4)$$

1.1.2 Ohm's Law

Ohm's law, in its simplest form, connects the charge density with the electric field through the resistivity $\mathbf{E}_{\text{co}} = \rho_c \mathbf{J}_{\text{co}}$, where \mathbf{E}_{co} is the electric field and \mathbf{J}_{co} the current density measured in the comoving frame. In most cases of astrophysical

interest, the resistivity is very low, allowing to replace the above equation with $\mathbf{E}_{\text{co}} = 0$. This is a consequence of the ability of charges to move fast and screen every induced electric field, leading to the condition that no electric field can exist in the comoving frame. Using Lorentz transformations of the electromagnetic field² we find the following relation between the electric and the magnetic fields in the laboratory frame

$$\mathbf{E} = -\frac{\mathbf{V}}{c} \times \mathbf{B}. \quad (5)$$

It is evident that for nonrelativistic flows $E \ll B$, and for this reason we neglect all the effects of the electric field when we compare them with analogous effects of the magnetic field. For example, we neglect the displacement current in Eq. (2). Also we neglect the electric force in the momentum equation as we will see later.

1.1.3 Mass Conservation

Assume a constant volume $\delta\tau$ in space, bounded by the closed surface S . If δm is the mass inside this volume, then $-d(\delta m)/dt = -\partial(\delta m)/\partial t$ is the rate at which the mass inside the volume $\delta\tau$ decreases with time. Due to mass conservation, this rate equals to the mass flow rate that passes the surface S , which is $\oint (n_+ m_+ \mathbf{V}_+ + n_- m_- \mathbf{V}_-) \cdot d\mathbf{S} = \oint \rho \mathbf{V} \cdot d\mathbf{S} = \iiint \nabla \cdot (\rho \mathbf{V}) d\tau$. By writing $\delta m = \iiint \rho d\tau$ and equating the two expressions we find the so-called continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0. \quad (6)$$

1.1.4 Momentum Equation

The various force densities that act on the fluid are

- pressure gradient $-\nabla P$,
- electric force per volume $(n_+ q_+ + n_- q_-) \mathbf{E} = \rho_q \mathbf{E}$,
- magnetic force per volume $(n_+ q_+ \mathbf{V}_+ / c + n_- q_- \mathbf{V}_- / c) \times \mathbf{B} = (1/c) \mathbf{J} \times \mathbf{B}$,
- gravitational force density $-\rho \nabla \Phi_g$ which becomes $-\rho (\mathcal{G} \mathcal{M} / r^2) \hat{\mathbf{r}}$ in the case of a central mass \mathcal{M} (r is the spherical distance from the central mass and $\hat{\mathbf{r}}$ the corresponding unit vector).

The sum of these force densities equals the inertial force density $\rho d\mathbf{V}/dt = \rho(\partial/\partial t + \mathbf{V} \cdot \nabla) \mathbf{V}$, giving the following momentum equation (after substituting \mathbf{J} from Ampère's law and dropping the electric field terms which are negligible, since $E \ll B$ as explained in Sect. 1.1.2):

² $\mathbf{E}_{\text{co}} = \gamma (\mathbf{E} + \frac{\mathbf{V}}{c} \times \mathbf{B}) - (\gamma - 1) (\mathbf{E} \cdot \frac{\mathbf{V}}{V}) \frac{\mathbf{V}}{V}$, where γ the Lorentz factor $\gamma = (1 - \mathbf{V}^2/c^2)^{-1/2}$.

$$\rho \left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \right) \mathbf{V} = -\nabla P + \frac{(\nabla \times \mathbf{B}) \times \mathbf{B}}{4\pi} - \rho \frac{\mathcal{GM}}{r^2} \hat{\mathbf{r}}. \quad (7)$$

Note that the change in any quantity of the fluid is due to two factors: (1) the possible direct dependence on time and (2) the fluid motion. Mathematically this is expressed as

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla$$

and can be seen as the property for any function Φ of (t, \mathbf{r})

$$\frac{d\Phi(t, \mathbf{r})}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Phi(t + \Delta t, \mathbf{r} + \mathbf{V}\Delta t) - \Phi(t, \mathbf{r})}{\Delta t} = \frac{\partial \Phi}{\partial t} + \mathbf{V} \cdot \nabla \Phi.$$

1.1.5 Thermodynamic Relations

The temperature of an ideal gas, isotropic in its rest frame, is a known function of the pressure and density given by

$$k_B T = \frac{P}{n_+ + n_-} = \frac{P}{\rho/m}, \quad (8)$$

where m is the mean mass per particle (e.g., $m = m_p/2$ for electron–proton plasma and $m = m_e$ for electron–positron plasma).

According to the kinetic theory of gases, in a gas with f degrees of freedom ($f = 3$ for monatomic gas), the internal energy per particle is $(f/2)k_B T$. Thus, the internal energy per mass is $e = (f/2)(P/\rho)$. Defining the polytropic index as $\Gamma = 1 + 2/f$ we can write

$$e = \frac{1}{\Gamma - 1} \frac{P}{\rho}. \quad (9)$$

The previous holds for nonrelativistic thermal motions in which the random motion of particles in the comoving frame is nonrelativistic, or, in other words, when $mc^2 \gg k_B T$ holds.³

Another useful thermodynamic quantity is the enthalpy. Since it is defined as the sum of the internal energy plus the product pressure times volume, the enthalpy per mass is

³ The other limit, $k_B T \gg mc^2$, is in principle possible in relativistic MHD flows. In that case, the energy of a particle is proportional to its momentum $\sqrt{p^2 c^2 + m^2 c^2} \approx pc$, and thus, the particles behave as a photon fluid with energy density $3P$. This translates to energy per mass $e = 3P/\rho$, again a relation of the form (9), with $\Gamma = 4/3$. (This value of Γ is not related to degrees of freedom.)

$$h = e + \frac{P}{\rho} = \frac{\Gamma}{\Gamma - 1} \frac{P}{\rho}. \quad (10)$$

1.1.6 The First Law of Thermodynamics

A mass δm of gas has a volume $\delta m/\rho$. The expansion of the mass is connected to the production of work $P d(\delta m/\rho)$. This mass can be heated or cooled. If q is the volumetric rate of net energy input/output, then energy $q(\delta m/\rho)dt$ is added or subtracted (when $q > 0$ the mass is heated, while $q < 0$ means that the system is cooled). The first law of thermodynamics describes the energy conservation $d(e\delta m) = q(\delta m/\rho)dt - P d(\delta m/\rho)$. Thus, we can write

$$\frac{de}{dt} + P \frac{d(1/\rho)}{dt} = \frac{q}{\rho}, \quad \text{or,} \quad \frac{dh}{dt} - \frac{1}{\rho} \frac{dP}{dt} = \frac{q}{\rho}. \quad (11)$$

Using Eq. (9) we find

$$\frac{d}{dt} \left(\frac{1}{\Gamma - 1} \frac{P}{\rho} \right) + P \frac{d(1/\rho)}{dt} = \frac{q}{\rho}, \quad \text{or,} \quad \frac{\rho^\Gamma}{\Gamma - 1} \frac{d}{dt} \left(\frac{P}{\rho^\Gamma} \right) = q, \quad (12)$$

a relation which provides a connection with the entropy. In adiabatic flows ($q = 0$), the above equation yields the conservation of specific entropy ($P/\rho^\Gamma = \text{constant}$ at each fluid parcel).

The first law of thermodynamics applies only for reversible processes and not, for example, in the presence of shocks. As will become clear later, it is equivalent to the energy equation in conservative form. The latter implies Eq. (11) only if all the involved quantities are differentiable, i.e., in the absence of shocks.

Polytropic flows

In cases where the analysis is focused on the flow dynamics (rather than on the energetics) and/or the heating/cooling mechanisms are poorly known or technically difficult to be described, the polytropic assumption is often used. The idea is to mimic the heating/colling process with an adiabatic evolution. This can be done by rearranging Eq. (12) to a similar equation for an adiabatic flow, but with an effective polytropic index Γ_{eff} . Equivalently, the function q can be modeled as

$$q = \frac{\Gamma - \Gamma_{\text{eff}}}{\Gamma - 1} \rho P \frac{d(1/\rho)}{dt} = \frac{\Gamma - \Gamma_{\text{eff}}}{\Gamma - 1} P \nabla \cdot \mathbf{V}, \quad (13)$$

and Eq. (12) can be simplified to

$$\frac{d}{dt} \left(\frac{1}{\Gamma_{\text{eff}} - 1} \frac{P}{\rho} \right) + P \frac{d(1/\rho)}{dt} = 0, \quad \text{or,} \quad \frac{d}{dt} \left(\frac{P}{\rho^{\Gamma_{\text{eff}}}} \right) = 0, \quad (14)$$

which gives the integral of the effective specific entropy $P/\rho^{\Gamma_{\text{eff}}} = \text{constant}$ at each fluid parcel. In other words, the evolution of a gas with $2/(\Gamma - 1)$ degrees of freedom

and nonzero heating is modeled as the adiabatic evolution of a gas with $2/(\Gamma_{\text{eff}} - 1)$ degrees of freedom (with $\Gamma_{\text{eff}} < \Gamma$). The external heating is in essence replaced by the energy that is stored initially in the additional degrees of freedom and released as the flow expands. Similarly, a cooled flow can be modeled with an effective polytropic index $\Gamma_{\text{eff}} > \Gamma$.

Note that the effective specific entropy $P/\rho^{\Gamma_{\text{eff}}}$ which is constant at each fluid parcel is not necessarily a global constant. Instead, it may take different values for different parcels of the flow. The same applies to the P/ρ^{Γ} in adiabatic flows, which can be seen as subcases of polytropic flows corresponding to $\Gamma_{\text{eff}} = \Gamma$.

Nonpolytropic flows

An alternative way of modeling magnetized outflows is the so-called nonpolytropic approach (see Sauty et al. [46] and references therein). The idea is to simply ignore the energy equation, ignore all the thermodynamic quantities,⁴ and find solutions for the flow speed, density, and magnetic field that are consistent with the rest of the MHD equations. Hence, in such a treatment the heating sources that produce some specific solution are not known a priori; instead, they can be determined only a posteriori, by calculating q from Eq. (12). This method has the advantage of being more general compared to the polytropic approach, since the latter corresponds to a special functional relation of q with the density and pressure. However, it has the disadvantage of not allowing to control the energetics, since these can be examined only a posteriori. This means that the heating/cooling required by a solution may be unreasonable making the solution unphysical. Nevertheless, nonpolytropic solutions with reasonable heating/cooling are perfectly acceptable and are by no means worse than the polytropic solutions.

1.1.7 Conservative Forms

In a conservative equation of the form $(\partial/\partial t)\mathcal{D} + (\partial/\partial x_j)\mathcal{F}_j = \mathcal{S}$, written for some quantity, we can identify the density of this quantity \mathcal{D} , the corresponding flux \mathcal{F}_j , and possible source terms \mathcal{S} . For example, the continuity Eq. (6) is the conservative form for the mass, with ρ being the mass density and ρV_j being the mass flux (no source terms appear since the mass is conserved).

We can also write conservative equations for the momentum and energy. In fact these are the equations that are found from first principles (either using the stress-energy tensor as we do in the description of relativistic flows, see Sect. 2.1, or if we use the kinetic theory). These equations yield Eqs. (7) and (12) in the case where shocks are not present.

The conservative form of the momentum equation is $\left(\text{in Cartesian coordinates } [x_j, j = 1, 2, 3] \right)$

⁴ The pressure enters the momentum equation through the ∇P term, but this can be eliminated by taking the $\nabla \times$ of that equation.

$$\begin{aligned}
& \frac{\partial}{\partial t} \left(\rho \mathbf{V} + \frac{\mathbf{E} \times \mathbf{B}}{4\pi c} \right) \\
& + \sum_{i,j=1}^3 \frac{\partial}{\partial x_j} \left(\rho V_i V_j + P \delta_{ij} + \frac{B^2 + E^2}{8\pi} \delta_{ij} - \frac{B_i B_j + E_i E_j}{4\pi} \right) \hat{\mathbf{x}}_i \\
& = -\rho \frac{\mathcal{GM}}{r^2} \hat{\mathbf{r}}. \quad (15)
\end{aligned}$$

We identify the momentum density for the fluid $\rho \mathbf{V}$ and the electromagnetic field $\mathbf{E} \times \mathbf{B}/4\pi c$, the flux of ‘ i ’ momentum through the ‘ j ’ direction for the fluid $\rho V_i V_j + P \delta_{ij}$ and the electromagnetic field $(B^2 + E^2) \delta_{ij}/8\pi - (B_i B_j + E_i E_j)/4\pi$, and the source term $-\rho \mathcal{GM} \hat{\mathbf{r}}/r^2$ due to the gravitational field.

The conservative form of the energy equation is

$$\begin{aligned}
& \frac{\partial}{\partial t} \left(\frac{1}{2} \rho V^2 + \frac{1}{\Gamma - 1} P + \frac{B^2 + E^2}{8\pi} \right) \\
& + \nabla \cdot \left(\frac{1}{2} \rho V^2 \mathbf{V} + \frac{\Gamma}{\Gamma - 1} P \mathbf{V} + \frac{c}{4\pi} \mathbf{E} \times \mathbf{B} \right) = q - \rho \frac{\mathcal{GM}}{r^2} \hat{\mathbf{r}} \cdot \mathbf{V}. \quad (16)
\end{aligned}$$

Since $-\rho \mathbf{V} \cdot \nabla \Phi_g = -\partial(\rho \Phi_g)/\partial t - \nabla \cdot (\rho \Phi_g \mathbf{V})$, we can include the gravitational term of the right-hand side inside the energy density and flux and write

$$\begin{aligned}
& \frac{\partial}{\partial t} \left(\frac{1}{2} \rho V^2 + \frac{1}{\Gamma - 1} P + \frac{B^2 + E^2}{8\pi} - \rho \frac{\mathcal{GM}}{r} \right) \\
& + \nabla \cdot \left(\frac{1}{2} \rho V^2 \mathbf{V} + \frac{\Gamma}{\Gamma - 1} P \mathbf{V} + \frac{c}{4\pi} \mathbf{E} \times \mathbf{B} - \rho \frac{\mathcal{GM}}{r} \mathbf{V} \right) = q. \quad (17)
\end{aligned}$$

Thus, the energy density is the sum of the corresponding kinetic energy part $\rho V^2/2$, the internal energy part $P/(\Gamma - 1)$, the electromagnetic part $(B^2 + E^2)/8\pi$, and the gravitational energy density $\rho \Phi_g$. The terms inside the total energy flux correspond to the kinetic energy part $(\rho V^2/2) \mathbf{V}$, the enthalpy flux $[\Gamma/(\Gamma - 1)] P \mathbf{V}$, the Poynting flux $(c/4\pi) \mathbf{E} \times \mathbf{B}$, and the gravity term $\rho \Phi_g \mathbf{V}$.

It is straightforward to prove that if we multiply Eq. (15) with \mathbf{V} and then subtract it from Eq. (16), we arrive at the expression (12) for the first law of thermodynamics.

One can also prove that in polytropic flows Eq. (16) becomes

$$\begin{aligned}
& \frac{\partial}{\partial t} \left(\frac{1}{2} \rho V^2 + \frac{1}{\Gamma_{\text{eff}} - 1} P + \frac{B^2 + E^2}{8\pi} - \rho \frac{\mathcal{GM}}{r} \right) \\
& + \nabla \cdot \left(\frac{1}{2} \rho V^2 \mathbf{V} + \frac{\Gamma_{\text{eff}}}{\Gamma_{\text{eff}} - 1} P \mathbf{V} + \frac{c}{4\pi} \mathbf{E} \times \mathbf{B} - \rho \frac{\mathcal{GM}}{r} \mathbf{V} \right) = 0, \quad (18)
\end{aligned}$$

i.e., takes the form of the energy equation for an adiabatic evolution of a gas with polytropic index Γ_{eff} .⁵

1.2 Integrals of Motion

Since the outflows that we examine originate from the environment of rotating objects (central masses or disks), axisymmetry is a reasonable assumption. In spherical (r, θ, ϕ) or cylindrical (z, ϖ, ϕ) coordinates, with the z -axis the axis of rotation, $\partial/\partial\phi = 0$ holds. In addition, steady-state $\partial/\partial t = 0$ is also assumed, meaning that the flows do not change much on the timescales of interest. Thus, all the unknown functions of the problem depend on a pair of variables (z, ϖ) or (r, θ) depending on the choice of coordinates.

These assumptions make the partial integration of most equations possible. The resulting integrals give important information on the characteristics of outflows as will become clear in the following.

We will analyze polytropic flows, but the analysis can be easily generalized to cover nonpolytropic flows as well, as we discuss later.

1.2.1 The Magnetic Flux Function

The most important unknown of the problem is the magnetic flux function A . We define the plane (z, ϖ) in cylindrical (or $[r, \theta]$ in spherical) coordinates as the poloidal plane, see Fig. 1, and split all vectors in poloidal and azimuthal components. The velocity has a poloidal \mathbf{V}_p and an azimuthal \mathbf{V}_ϕ part. Similarly, the magnetic field is decomposed to \mathbf{B}_p and \mathbf{B}_ϕ .

The Maxwells' equation $\nabla \cdot \mathbf{B} = 0$ can be simplified to $\nabla \cdot \mathbf{B}_p = 0$ (since $\nabla \cdot \mathbf{B}_\phi$ is identically 0 due to axisymmetry). So, there is a function $A(\varpi, z)$ such that

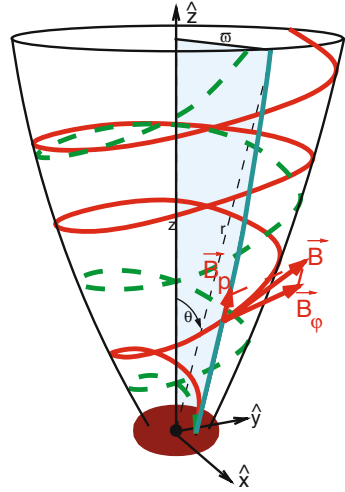
$$\mathbf{B}_p = \frac{\nabla A \times \hat{\phi}}{\varpi} \text{ or } \mathbf{B}_p = \nabla \times \left(\frac{A \hat{\phi}}{\varpi} \right). \quad (19)$$

Obviously the function A is related to the vector potential of the poloidal magnetic field. Its name comes from the relation with the magnetic flux

$$A = \frac{1}{2\pi} \iint \mathbf{B}_p \cdot d\mathbf{S}. \quad (20)$$

⁵ Note that the polytropic assumption introduces the following error in the presence of shocks. The polytropic assumption is based on the replacement of the heating function q with the form given in Eq. (13). This expression gives a delta function at the position of the shock where the velocity is discontinuous, introducing an artificial cooling if $\Gamma_{\text{eff}} < \Gamma$ (or artificial heating for $\Gamma_{\text{eff}} > \Gamma$). This has important consequences, e.g., the compression ratio of a strong shock is now $(\Gamma_{\text{eff}} + 1) / (\Gamma_{\text{eff}} - 1)$ instead of the correct value $(\Gamma + 1) / (\Gamma - 1)$.

Fig. 1 Sketch of a streamline (dashed line) and a field line (solid line). The projection of both lines on the poloidal plane (shadowed plane) coincide. The equation of this line is $A = \text{constant}$



Thus, the equation of the poloidal magnetic field line (which is the projection of the magnetic field line on the poloidal plane) is $A = \text{constant}$, see Fig. 1.

1.2.2 The Field Angular Velocity

Faraday's equation, in the case of steady magnetic field, gives that the electric field comes from a scalar potential $\mathbf{E} = -\nabla\Phi_E$. Due to the axisymmetry, $E_\phi = 0$. In that case, Ohm's law (5) gives $(\mathbf{V} \times \mathbf{B}) \cdot \hat{\phi} = 0 \Leftrightarrow (\mathbf{V}_p \times \mathbf{B}_p) \cdot \hat{\phi} = 0$, meaning that the \mathbf{V}_p and \mathbf{B}_p are parallel. Thus, the plasma flow on the poloidal plane follows the magnetic field, or, the magnetic field is frozen into the plasma. The 3D streamline and field line do not in general coincide, though, as seen in Fig. 1.

Since \mathbf{V}_p and \mathbf{B}_p are parallel, there are functions Ψ_A and Ω such that

$$\mathbf{V} = \frac{\Psi_A}{4\pi\rho} \mathbf{B} + \varpi \Omega \hat{\phi}, \quad \frac{\Psi_A}{4\pi\rho} = \frac{V_p}{B_p}. \quad (21)$$

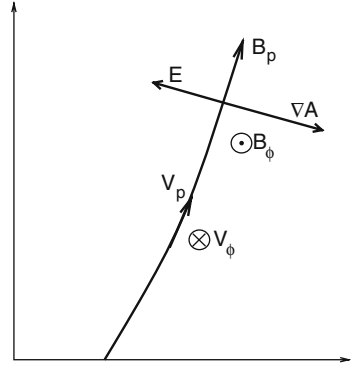
Substituting the previous relation in Ohm's law (5) we get $\mathbf{E} = -(\varpi \Omega / c) \hat{\phi} \times \mathbf{B}_p$. Using Eq. (19) we find that the electric field is

$$\mathbf{E} = -\frac{\Omega}{c} \nabla A, \quad \text{with magnitude } E = \frac{\varpi \Omega}{c} B_p. \quad (22)$$

\mathbf{E} is normal to the poloidal magnetic field lines $A = \text{constant}$ and points toward the rotation axis, as seen in Fig. 2.

Using Eq. (22) Faraday's law (4) gives $\nabla \times (\Omega \nabla A) = 0 \Leftrightarrow \nabla \Omega \times \nabla A = 0$. Thus, $\Omega = \Omega(A)$. In other words, Ω is a constant of motion, in other words it is an integral. From field line to field line this function can in principle be different, but

Fig. 2 Sketch of a poloidal field/streamline $A = \text{constant}$ and the directions of the various vector quantities



since the plasma flows along field lines $A = \text{constant}$, Ω remains constant through the motion of each fluid parcel.

If near the central object the density is relatively high, or, the poloidal field is strong such that $|\Psi_A B_p / 4\pi\rho| \ll |\varpi\Omega|$, or, $|V_p B_\phi / B_p| \ll |\varpi\Omega|$, Eq. (21) yields that $V_\phi \approx \varpi\Omega$, meaning that Ω is the fluid angular velocity near the base of the flow. At larger distances this no longer holds, as will become clear in the following. In fact, the azimuthal velocity becomes smaller compared to the corotation value, a result of rotation combined with the inertia of the field lines. If the roots of poloidal magnetic field lines with $B_z > 0$ are rotating with $\Omega > 0$ (i.e., in the $+\hat{\phi}$ direction), the inertia of the field creates an azimuthal field component in the $-\hat{\phi}$ direction. If the roots are rotating in the $-\hat{\phi}$ direction, or, the field has $B_z < 0$, a positive B_ϕ is created. In the following we examine the case with $B_z > 0$ and $\Omega > 0$, in which $B_\phi < 0$ (all other cases are straightforward generalizations).

1.2.3 The Mass-to-Magnetic Flux Ratio

Substituting Eq. (21) in the continuity Eq. (6), and since $\nabla \cdot \mathbf{B} = 0$, we get $\mathbf{B}_p \cdot \nabla \Psi_A = 0$. Thus, the derivative of Ψ_A along the poloidal field line is 0, meaning that Ψ_A is a constant of motion, $\Psi_A = \Psi_A(A)$. From Eq. (21), $\Psi_A = 2(\rho V_p \delta S) / (B_p \delta S / 2\pi)$. The denominator is the magnetic flux (see Eq. [20]), while the nominator is the mass flux (counting both hemispheres). So, Ψ_A is the mass-to-magnetic flux ratio, and it is a constant of motion (a consequence of the fact that $\mathbf{V}_p \parallel \mathbf{B}_p$).

1.2.4 The Adiat

The integration of Eq. (14) gives another constant of motion, the adiat

$$\frac{P}{\rho r_{\text{eff}}} = Q(A). \quad (23)$$

1.2.5 The Angular Momentum-to-Mass Flux Ratio

As a consequence of the assumed axisymmetry, a constant of motion related to the angular momentum should exist. Indeed, the $\hat{\phi}$ component of the momentum Eq. (7) can be integrated⁶ giving

$$\varpi V_\phi - \frac{\varpi B_\phi}{\Psi_A} = L(A). \quad (24)$$

The first term corresponds to the angular momentum-to-mass flux ratio for the matter, while the second to the angular momentum of the electromagnetic field over the mass flux. This becomes clear if we rewrite the second term as $(-\varpi B_p B_\phi / 4\pi) / (\rho V_p)$, the ratio of the angular momentum flux of the field $-\varpi B_p B_\phi / 4\pi$ over the mass flux ρV_p . Only the sum of the two angular momentum-to-mass flux ratios is constant along the flow; angular momentum can be transferred from the field to the matter and vice versa. In fact both terms have the same sign (since $B_\phi < 0$ as explained earlier). We expect that for strongly magnetized flows near the source the electromagnetic term is the dominant, while at large distances a significant part has been transferred to the matter.

1.2.6 The Energy-to-Mass Flux Ratio

An integral related to the energy can be found if we project the momentum equation along the flow, i.e., project Eq. (7) along \mathbf{V} (or use the equivalent Eq. (18)). The result is

$$\frac{1}{2} V^2 + \frac{\Gamma_{\text{eff}}}{\Gamma_{\text{eff}} - 1} \frac{P}{\rho} - \frac{\varpi \Omega B_\phi}{\Psi_A} - \frac{\mathcal{GM}}{r} = \mathcal{E}(A), \quad (25)$$

meaning that the total energy-to-mass flux ratio is a constant of motion. Note that the third term is positive (since $B_\phi < 0$) and is related to the electromagnetic field. Indeed, the Poynting flux $(c/4\pi)\mathbf{E} \times \mathbf{B}$ can be split into two parts. The part $(c/4\pi)\mathbf{E} \times \mathbf{B}_p$ is in the azimuthal direction and describes Poynting flux that cannot escape. The part $(c/4\pi)\mathbf{E} \times \mathbf{B}_\phi$ is in the direction of \mathbf{V}_p and points outward for $B_\phi < 0$. Its magnitude is $(c/4\pi)E|B_\phi|$, or, using Eq. (22), $(c/4\pi)(\varpi \Omega / c)B_p|B_\phi|$. Dividing with the mass flux ρV_p and using that $V_p/B_p = \Psi_A/4\pi\rho$ (see Eq. (21)), we find $\varpi \Omega |B_\phi|/\Psi_A$, i.e., the term of the integral (25) corresponding to the electromagnetic field.

For strongly magnetized flows near the source we expect that the dominant part of the energy integral is the electromagnetic term, while at larger distances the kinetic energy term $V^2/2$ becomes important.

⁶ In the proof we use the identity $(\mathbf{F} \cdot \nabla)\mathbf{F} = \nabla(F^2/2) + (\nabla \times \mathbf{F}) \times \mathbf{F}$. We also use that for any axisymmetric vector quantity $\mathbf{F} = \mathbf{F}_p + F_\phi \hat{\phi}$, the $\hat{\phi}$ component of $(\nabla \times \mathbf{F}) \times \mathbf{F}$ equals $[\mathbf{F}_p \cdot \nabla (\varpi F_\phi)](\hat{\phi}/\varpi)$.

Note that in the general, nonpolytropic case, Eq. (17) gives

$$\rho V_p \cdot \nabla \left(\frac{1}{2} V^2 + \frac{\Gamma}{\Gamma - 1} \frac{P}{\rho} - \frac{\varpi \Omega B_\phi}{\Psi_A} - \frac{\mathcal{GM}}{r} \right) = q, \quad (26)$$

resulting in the expression

$$\frac{1}{2} V^2 + \frac{\Gamma}{\Gamma - 1} \frac{P}{\rho} - \frac{\varpi \Omega B_\phi}{\Psi_A} - \frac{\mathcal{GM}}{r} = \mathcal{E}_0(A) + \int_{\ell_0}^{\ell} \frac{q}{\rho V_p} d\ell, \quad (27)$$

where ℓ is the arclength along each streamline (e.g., Sauty et al. [46]; Ferreira and Casse [22]). The integral $\int_{\ell_0}^{\ell} (q/\rho V_p) d\ell$ represents the heating added (or cooling for $q < 0$) per unit mass between arclengths ℓ_0 and ℓ , on a streamline $A = \text{constant}$.

1.3 The Use of the Integrals

Using the expressions for the integrals Ψ_A , Ω , and L (Eqs. (21), (24)), the Alfvénic Mach number

$$M = \frac{V_p}{B_p / \sqrt{4\pi\rho}}, \quad (28)$$

and the function G defined as

$$G = \frac{\varpi}{\varpi_A}, \text{ with } \varpi_A = \left(\frac{L}{\Omega} \right)^{1/2}, \quad (29)$$

we can express the MHD quantities as

$$\rho = \frac{\Psi_A^2}{4\pi M^2}, \quad (30)$$

$$\mathbf{B} = \frac{\nabla A \times \hat{\boldsymbol{\phi}}}{\varpi} - \frac{L\Psi_A}{\varpi} \frac{1 - G^2}{1 - M^2} \hat{\boldsymbol{\phi}}, \quad (31)$$

$$\mathbf{V} = \frac{M^2 \nabla A \times \hat{\boldsymbol{\phi}}}{\Psi_A \varpi} + \frac{L}{\varpi} \frac{G^2 - M^2}{1 - M^2} \hat{\boldsymbol{\phi}}. \quad (32)$$

In the case of a polytropic flow Eq. (23) gives

$$P = Q \left(\frac{\Psi_A^2}{4\pi} \right)^{\Gamma_{\text{eff}}} \frac{1}{M^{2\Gamma_{\text{eff}}}}. \quad (33)$$

Thus, all quantities are expressed in terms of the two remaining unknowns of the problem, A and M . (Note that the integrals are regarded as given functions of A , constrained by the boundary conditions of the outflow near its source.) The two remaining equations that must be solved with respect to A and M are the two components of the momentum equation on the poloidal plane, one along the flow and the other in the transfield direction, i.e., along ∇A .

The former can be simplified in the case of a polytropic flow, since it can be replaced by the energy integral (25). This relation in turn can be rewritten as (using Eqs. (30), (31), (32), and (33))

$$\frac{M^4 |\nabla A|^2}{2\Psi_A^2 \varpi^2} + \frac{L^2}{2\varpi^2} \left(\frac{G^2 - M^2}{1 - M^2} \right)^2 + \frac{\Gamma_{\text{eff}}}{\Gamma_{\text{eff}} - 1} Q \left(\frac{\Psi_A^2}{4\pi M^2} \right)^{\Gamma_{\text{eff}} - 1} + L\Omega \frac{1 - G^2}{1 - M^2} - \frac{\mathcal{GM}}{r} = \mathcal{E}, \quad (34)$$

an algebraic relation giving M as function of A and the magnitude of its gradient $|\nabla A|$.

The component of the momentum equation (7) in the transfield direction (along ∇A) gives the so-called Grad–Shafranov, or transfield equation

$$\begin{aligned} & (1 - M^2) \left[\varpi \frac{\partial}{\partial \varpi} \left(\frac{1}{\varpi} \frac{\partial A}{\partial \varpi} \right) + \frac{\partial^2 A}{\partial z^2} \right] - \frac{\partial A}{\partial \varpi} \frac{\partial M^2}{\partial \varpi} - \frac{\partial A}{\partial z} \frac{\partial M^2}{\partial z} \\ & + \frac{\varpi^2 \Psi_A^2}{M^2} \frac{d\mathcal{E}}{dA} - \frac{L^2 \Psi_A^2}{M^2} \frac{G^2 - M^2}{1 - M^2} \frac{d \ln L}{dA} - \frac{L^2 \Psi_A^2 G^2}{M^2} \frac{1 - G^2}{1 - M^2} \frac{d \ln \Omega}{dA} \\ & + \left[M^2 \left(\frac{\partial A}{\partial \varpi} \right)^2 + M^2 \left(\frac{\partial A}{\partial z} \right)^2 + L^2 \Psi_A^2 \left(\frac{1 - G^2}{1 - M^2} \right)^2 \right] \frac{d \ln \Psi_A}{dA} = 0 \end{aligned} \quad (35)$$

(e.g., Tsinganos [52]; Heyvaerts and Norman [27]; Heyvaerts [26]).

An alternative way of writing Eq. (35) is (after eliminating the derivatives of M^2 using Eq. (34))

$$\frac{V_p^4 - V_p^2 (C_s^2 + V_A^2) + C_s^2 V_{Ap}^2}{V_p^4} \nabla^2 A - \nabla A \cdot \nabla \ln |\nabla A| = \frac{\mathcal{F}}{1 - M^2}, \quad (36)$$

where \mathcal{F} is a function of r , M , A , and first order derivatives of A . Here $C_s \equiv \sqrt{\Gamma_{\text{eff}} P / \rho}$ is the effective sound speed, $V_{Ap} \equiv B_p / \sqrt{4\pi \rho}$ is the poloidal Alfvén speed, and $V_A \equiv B / \sqrt{4\pi \rho}$ is the total Alfvén speed. The Grad–Shafranov Eq. (36) is a highly nonlinear, second-order partial differential equation for A , linear in second-order derivatives. It is of mixed type, elliptic or hyperbolic depending on if the quantity

$$\left[V_p^4 - V_p^2 (C_s^2 + V_A^2) + C_s^2 V_{Ap}^2 \right] \left[V_p^2 (C_s^2 + V_A^2) - C_s^2 V_{Ap}^2 \right]$$

is negative or positive, respectively (for details see e.g., Tsinganos et al. [54]; Vlahakis [56]).

1.3.1 The Alfvénic Surface

An important surface in magnetized outflows is the Alfvénic surface, defined as the locus of points where $M = 1$. It is expected that near the source the poloidal magnetic field is strong, corresponding to $M < 1$, while at large distances $M > 1$, meaning that the flow poloidal speed V_p is larger than V_{Ap} (or equivalently the poloidal kinetic energy density $\rho V_p^2/2$ is larger than the poloidal magnetic energy density $B_p^2/8\pi$). Thus, the surface $M = 1$ is crossed at some point on each field/stream line. The $\hat{\phi}$ parts of Eqs. (31), (32) imply that at the Alfvén surface $G = 1$ should also hold. In other words, $\varpi_A(A)$ which is defined as $\sqrt{L/\Omega^2}$ (see Eq. (29)) is the cylindrical distance of the Alfvén point on each field/stream line. The ratios appearing in the $\hat{\phi}$ parts of Eqs. (31), (32) are then finite.

The Alfvén surface is also a singularity for the Grad–Shafranov Eq. (36). This implies that a regularity condition $\mathcal{F}|_{M=1} = 0$ must be satisfied.

The density at the Alfvén surface is directly connected to the integral Ψ_A . Indeed, Eq. (30) yields $\rho_A = \Psi_A^2/4\pi$. The square of the Alfvénic Mach number at any point can be written as $M^2 = \rho_A/\rho$ (so it is $1/\rho$ normalized to its value at the Alfvén surface, $1/\rho_A$).

1.3.2 The Bernoulli Equation

The energy integral relation (34) is also called Bernoulli equation. It can be written as $\mathcal{B}(A, G, M) = \mathcal{E}(A)$ with

$$\begin{aligned} \mathcal{B}(A, G, M) = & \frac{|\nabla A|^2}{2\varpi_A^2 \Psi_A^2 G^2} M^4 + \frac{\Gamma_{\text{eff}}}{\Gamma_{\text{eff}} - 1} \mathcal{Q} \left(\frac{\Psi_A^2}{4\pi M^2} \right)^{\Gamma_{\text{eff}} - 1} + \\ & + \varpi_A^2 \Omega^2 \left[\frac{(G^2 - M^2)^2}{2G^2 (1 - M^2)^2} + \frac{1 - G^2}{1 - M^2} \right] + \Phi_g. \end{aligned} \quad (37)$$

The term $|\nabla A|$ and the gravitational potential Φ_g are regarded as functions of G and A only.

If we want to find M at each G and on a constant line A , we must find the intersection of the function \mathcal{B} (regarded as a function of M only) with $\mathcal{E}(A)$. For each $G \neq 1$ we have

$$\frac{\partial \mathcal{B}(A, G, M)}{\partial M^2} = \frac{|\nabla A|^2 M^2}{\varpi_A^2 \Psi_A^2 G^2} - \frac{\Gamma_{\text{eff}} Q}{M^2 \Gamma_{\text{eff}}} \left(\frac{\Psi_A^2}{4\pi} \right)^{\Gamma_{\text{eff}}-1} + \varpi_A^2 \Omega^2 \frac{M^2 (1-G^2)^2}{G^2 (1-M^2)^3},$$

$$\frac{\partial^2 \mathcal{B}(A, G, M)}{\partial (M^2)^2} = \frac{|\nabla A|^2}{\varpi_A^2 \Psi_A^2 G^2} + \Gamma_{\text{eff}}^2 Q \frac{\left(\frac{\Psi_A^2}{4\pi} \right)^{\Gamma_{\text{eff}}-1}}{M^2 \Gamma_{\text{eff}}+2} + \varpi_A^2 \Omega^2 \frac{(1-G^2)^2}{G^2} \frac{2M^2+1}{(1-M^2)^4}.$$

We see that

$$\frac{\partial^2 \mathcal{B}(A, G, M)}{\partial (M^2)^2} > 0, \forall M^2 \text{ and } \lim_{M^2 \rightarrow 1^\pm} \frac{\partial^2 \mathcal{B}(A, G, M)}{\partial (M^2)^2} = +\infty.$$

So the first derivative $\frac{\partial \mathcal{B}(A, G, M)}{\partial M^2}$ monotonically increases from $-\infty$ (when $M^2 \rightarrow 0^+$) to ∞ (when $M^2 \rightarrow 1^-$) and from $-\infty$ (when $M^2 \rightarrow 1^+$) to ∞ (when $M^2 \rightarrow \infty$). Therefore, $\exists M_1^2 \in (0, 1)$, $M_2^2 \in (1, \infty)$ such that $\left(\frac{\partial \mathcal{B}(A, G, M)}{\partial M^2} \right)_{M^2=M_{1,2}^2} = 0$.

As we see in general there are four roots (see Fig. 3) which correspond to super- or sub-Alfvénic (depending on if $G > 1$ or $G < 1$), in- or outflow (depending on the sign of M). So, as we change G the function \mathcal{B} consists of two grooves which connect with each other at $G = 1$ (the singularity $M = 1$ disappears when $G = 1$) and then separate again (see Fig. 4). The two local minima of these grooves move, and when $\mathcal{B} = \mathcal{E}$ at this minimum we have a critical point because at this point: $\partial \mathcal{B}(A, G, M) / \partial M = 0$ and $\partial \mathcal{B}(A, G, M) / \partial G = 0$ (since immediately after this point the minimum of the function \mathcal{B} decreases in order to be solvable the equation $\mathcal{B} = \mathcal{E}$). This point corresponds to a saddle point of the Bernoulli surface (Fig. 4).

After some manipulation, we find that

$$M^2 \frac{\partial \mathcal{B}(A, G, M)}{\partial M^2} = \frac{V_p^4 - V_p^2 (C_s^2 + V_A^2) + C_s^2 V_{\text{Ap}}^2}{V_p^2 - V_{\text{Ap}}^2}. \text{ So when at this point } \mathcal{B} = \mathcal{E}$$

(this minimum is a point of the solution), we have $V_p^2 = V_s^2$ or $V_p^2 = V_f^2$, where V_s^2, V_f^2 are the two roots of the equation $V_{s,f}^4 - V_{s,f}^2 (C_s^2 + V_A^2) + C_s^2 V_{\text{Ap}}^2 = 0$ (with $V_s^2 < V_f^2$).

The V_s and V_f are the effective phase speeds of the slow and fast magnetosonic waves that have wavevectors parallel to \mathbf{V}_p (for details, see, e.g., Vlahakis [56]). For this reason, the loci of points where $V_p = V_s$ or $V_p = V_f$ are called slow or fast magnetosonic surfaces, respectively.

We note that if we knew the field line shape (the function A), the Bernoulli equation fully determines the flow characteristics, since it gives the remaining unknown function M . However, the function A must satisfy the transfield (or Grad-Shafranov) equation, which also depends on M . Thus, the system of the Bernoulli and transfield equations is coupled and must be solved simultaneously. Nevertheless, the critical points of the Bernoulli equation are important reference points that help us understand the general characteristics of a magnetized outflow.

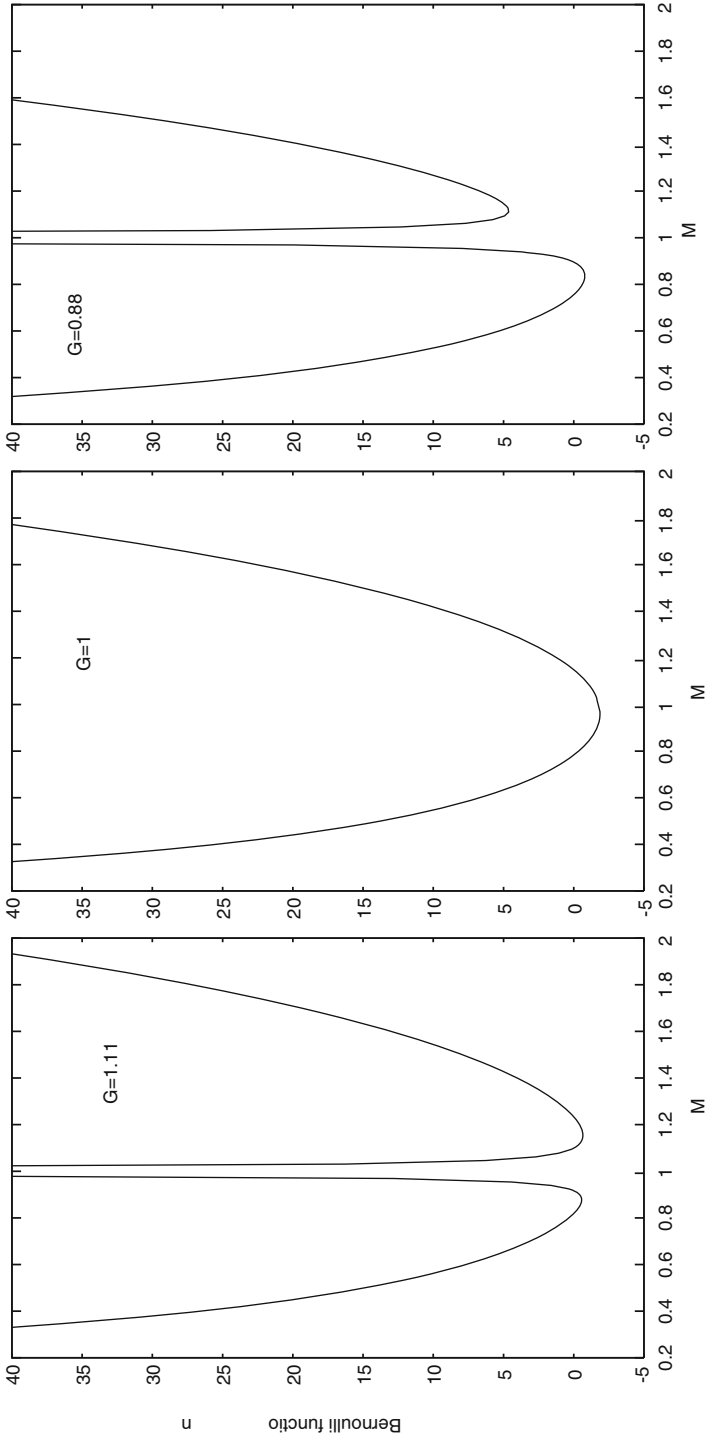


Fig. 3 The function B for various values of G . The example is taken from an exact polytropic, r self-similar solution, from Vlahakis [56]

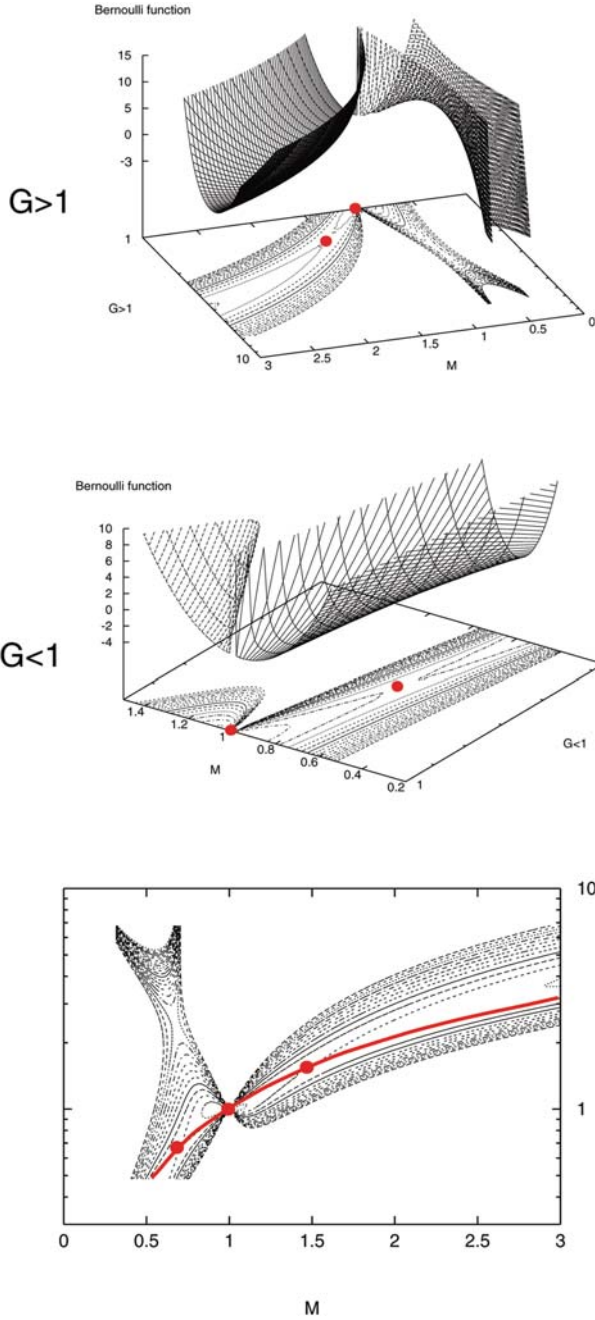


Fig. 4 The 3D diagram for the function \mathcal{B} and the isocontours on the G, M plane. We see the critical solution (*thick line*) which passes through the three critical points: the slow magnetosonic (where $V_p = V_s$), the Alfvén (where $V_p = V_{Ap}$), and the fast magnetosonic (where $V_p = V_f$). The example is taken from an exact polytropic, r self-similar solution which is analyzed in detail in Vlahakis [56]

The crossing of the fast magnetosonic point is important since it is related to the causal connection of the flow at large distances with the base of the flow and the source. In particular, in the superfast regime of the outflow all signals propagate inside a fast Mach cone, defined by the characteristic curves (which exist since in this regime the Grad–Shafranov equation (36) is a hyperbolic partial differential equation). There is a limiting characteristic (or modified fast magnetosonic surface) which is the horizon for the propagation of the MHD waves, in the sense that all disturbances from this surface downstream cannot travel back and reach the base of the flow (for details see, e.g., Bogovalov [9]; Tsinganos et al. [54]; Vlahakis [56]; Sauty et al. [47]). Since this surface is inside the superfast regime, a necessary (but not sufficient) condition is to have a superfast magnetosonic outflow.

1.3.3 The Flow Near the Central Source

Near the source of the flow (central object or disk) the poloidal magnetic field is expected to be strong. Thus, at the ejection surface (denoted by subscript ‘ i ’) it is expected that $B_{pi} \gtrsim |B_{\phi i}|$. In addition, the magnetic energy density is much larger than the poloidal kinetic energy density, meaning that $M_i \ll 1$.

A consequence of these conditions is that the azimuthal speed near the base of the flow is $V_\phi \approx \varpi \Omega$ (as can be found by applying $M_i \ll 1$ in Eq. (32) and using the definition of G from Eq. (29); this relation was also discussed in Sect. 1.2.2, as a consequence of Eq. (21)). The flow rotates with angular velocity Ω . Since Ω is constant along the field/stream lines, the flow corotates with its base, and plasma parcels move like ‘beads on rotating wires.’ In the case of disk-driven flows, the rotation of the base is roughly Keplerian $\Omega \approx \Omega_K = \sqrt{GM/\varpi_i^3}$, while for flows originating from a central object Ω is the angular velocity of that object.

Since the flow moves as a solid body up to the Alfvén surface whose lever arm is ϖ_A , the extracted angular momentum-to-mass flux ratio is not $\varpi_i V_{\phi i} \approx \varpi_i^2 \Omega$, but instead $\varpi_A^2 \Omega$. This is an exact result, since the integral L gives the total angular momentum-to-mass flux ratio (see Sect. 1.2.5), and the regularity condition at the Alfvén surface gives $L = \varpi_A^2 \Omega$ (see Eq. (29)).

In the sub-Alfvénic part of the flow, $\varpi_i \ll \varpi_A$, i.e., $G_i \ll 1$. The azimuthal magnetic field is then $B_\phi \approx -L\Psi_A/\varpi$ (as can be found by applying $M_i \ll 1$ and $G_i \ll 1$ in Eq. (31)), meaning that the main part of the angular momentum is carried by the electromagnetic field (the dominant part of L is the $-\varpi B_\phi/\Psi_A$ term, see Eq. (24)). Thus, the product ϖB_ϕ is roughly constant along the flow, near the source (in the sub-Alfvénic region). This product is proportional to the poloidal current I , since $\mathbf{J}_p = (c/4\pi)\nabla \times \mathbf{B}_\phi = (1/2\pi\varpi)\nabla I \times \hat{\phi}$, with $I = \iint \mathbf{J}_p \cdot d\mathbf{S} = (c/2)\varpi B_\phi$. The current lines on the poloidal plane represent the loci of constant poloidal current ($I = \text{constant} \Leftrightarrow \varpi B_\phi = \text{constant}$). The constancy of ϖB_ϕ along the flow means that the current lines and the streamlines are parallel, with consequences on the acceleration that will be discussed later.

An interesting subclass of magnetized outflows are the ones which are Poynting flux dominated near their sources. In these flows the main acceleration mechanisms

have magnetic origin (contrary to a thermal acceleration scenario). The dominant term of the energy integral (25) is the Poynting, which – using the approximate expression of B_ϕ – gives

$$\mathcal{E} \approx L\Omega = (\varpi_A\Omega)^2. \quad (38)$$

1.3.4 Rough Scalings

From magnetic flux conservation, the poloidal magnetic field drops as $\propto \varpi^{-2}$ along any field/stream line.

The azimuthal part of the field is connected to the poloidal current, which is roughly constant in the sub-Alfvénic region, implying that $B_\phi \propto -\varpi^{-1}$. Even after the Alfvénic surface the B_ϕ continues to drop roughly as $\propto -\varpi^{-1}$ (looking the energy integral (25), the Poynting-to-mass flux ratio is proportional to ϖB_ϕ ; this quantity drops only by 1/2 at the distance where half of the energy carried by the field has been transferred to the matter).

Since the poloidal component of the magnetic field drops faster than the azimuthal part, after some distance $B \approx |B_\phi|$ (although near the origin $B_p \gg |B_\phi|$).

From the constancy of the mass-to-magnetic flux, $\rho V_p \propto B_p$, and thus $\rho V_p \propto 1/\varpi^2$. For increasing V_p along the flow, the density drops faster than ϖ^{-2} , and the pressure drops faster than $\varpi^{-2\Gamma_{\text{eff}}}$.

1.3.5 The Fast Magnetosonic Point

As already discussed, the fast magnetosonic point is defined as the locus where $V_p = V_f$, where V_f is the larger solution of $V_f^4 - V_f^2(C_s^2 + V_A^2) + C_s^2 V_{Ap}^2 = 0$. At the distance of that point the flow is expected to be practically cold, since the thermal pressure drops faster than B^2 . Using Eq. (21) we can write the poloidal kinetic energy-to-Poynting flux ratio (the ratio of the corresponding parts in the energy integral (25))

$$\left(\frac{V_p^2/2}{-\varpi\Omega B_\phi/\Psi_A} \right)_f \approx \frac{1}{2} \left[\left(1 - \frac{V_\phi}{\varpi\Omega} \right) \frac{B^2}{B_\phi^2} \right]_f \approx \frac{1}{2} \quad (39)$$

(since at the fast point the azimuthal magnetic field is the dominant component and the azimuthal velocity much smaller than $\varpi\Omega$). Thus, roughly 1/3 of the total energy flux has been transferred to the matter up to the fast point (with the remaining 2/3 still carried by the electromagnetic field). Thus, $V_f^2/2 = \mathcal{E}/3$, or,

$$V_f = \left(\frac{2}{3} \mathcal{E} \right)^{1/2}. \quad (40)$$

At the fast point $M \gg 1$, $G \gg 1$, and the expression (31) for the azimuthal magnetic field gives $-\varpi\Omega B_\phi/\Psi_A \approx L\Omega G^2/M^2 = (\Omega^2 B_p \varpi^2)/(\Psi_A V_p)$, since

$V_p = M^2 B_p / \Psi_A$ from Eq. (32). Combining the above relation with Eq.(39) we find the velocity at the fast magnetosonic point

$$V_f = \left[\frac{\Omega^2 (B_p \varpi^2)_f}{\Psi_A} \right]^{1/3}. \quad (41)$$

A combination of the previous equation with Eq. (40) yields that $(B_p \varpi^2 / A)_f = (\Psi_A / A \Omega^2) (2\mathcal{E}/3)^{3/2}$, since the Poynting flux at the fast point equals 2/3 of the total energy flux $(-\varpi \Omega B_\phi)_f / \Psi_A = (2/3)\mathcal{E}$. Combining the previous two relations we find the ratio of azimuthal-to-poloidal magnetic field at the fast point $(-B_\phi / B_p)_f = \varpi_f \Omega / (2/3\mathcal{E})^{1/2}$. For strongly magnetized flows near the source $\mathcal{E} \approx L\Omega = (\varpi_A \Omega)^2$, and thus $(-B_\phi / B_p)_f = (3/2)^{1/2} \varpi_f / \varpi_A$. Clearly the azimuthal component of the magnetic field is the dominant component, as was previously stated.

As we have already explained, at small distances from the source almost all the angular momentum is carried by the electromagnetic field, resulting in $L = \varpi_A^2 \Omega \gg \varpi_i^2 \Omega \approx \varpi_i V_{\phi i}$. The ratio of the angular momentum flux carried by the field over the total angular momentum flux decreases with distance, since angular momentum is transferred to the matter. This ratio is $(-\varpi B_\phi / \Psi_A) / L$, and for strongly magnetized flows in which $L\Omega \approx \mathcal{E}$, it equals the ratio of the Poynting flux over the total energy flux $(-\varpi \Omega B_\phi / \Psi_A) / \mathcal{E}$. Since at the fast magnetosonic point the energy flux carried by the matter is 1/3 of the total, the angular momentum flux carried by the matter is also 1/3 of the total (with the remaining 2/3 still in the field). Equivalently, the angular velocity at the fast is $V_{\phi f} \approx L / (3\varpi_f)$. The ratio $V_{\phi f} / (\varpi_f \Omega)$ becomes $\varpi_A^2 / (3\varpi_f^2)$ and is much smaller than unity as was assumed in Eq. (39).

Using Eq. (32) the $V_{\phi f} \approx L / (3\varpi_f)$ yields the ratio $(G^2 / M^2)_f = 2/3$. Since $G = \varpi / \varpi_A$ and $M = \rho_A / \rho$, we find $(\rho_f \varpi_f^2) / (\rho_A \varpi_A^2) = 2/3$.

1.3.6 The Flow at Large Distances

The flow velocity in principle continues to increase after the fast magnetosonic surface. Its maximum possible value is $V_{\max} = \sqrt{2\mathcal{E}}$. For strongly magnetized flows near the source $\mathcal{E} \approx L\Omega = \varpi_A^2 \Omega^2$, and thus $V_{\max} \approx \sqrt{2L\Omega} = \varpi_A \Omega \sqrt{2}$. Since $V_{\phi i} \approx \varpi_i \Omega$, the ratio $V_{\max} / V_{\phi i} = (\varpi_A / \varpi_i) \sqrt{2}$.

However, it is not always possible to have a full transformation of Poynting to kinetic energy flux, and the terminal velocity V_∞ may be smaller than V_{\max} . Defining the function

$$\zeta = \frac{V_p^2 / 2}{\mathcal{E}}, \quad (42)$$

the efficiency of the acceleration is the terminal value of this function ζ_∞ , and the terminal velocity is $V_\infty = \zeta_\infty^{1/2} V_{\max}$.

At large distances the two dominant parts of the energy-to-mass flux ratio are the kinetic $\zeta \mathcal{E}$ and the Poynting $-\varpi \Omega B_\phi / \Psi_A = (1 - \zeta) \mathcal{E}$. Using Eq. (31) we find that $\zeta \approx 1 - G^2/M^2$. Equation (32) then implies that

$$V_\phi \approx \zeta \frac{L}{\varpi}, \quad (43)$$

meaning that the efficiency of the transfer of angular momentum from the field to matter is also ζ .

Substituting the definition of ζ , Eq. (42), and using $\mathcal{E} \approx L\Omega$ we find

$$\frac{V_p^2}{2\varpi V_\phi} = \Omega. \quad (44)$$

Thus, the above combination of V_p and ϖV_ϕ remains constant as long as the flow is accelerated. The terminal values of these quantities, which are observable quantities, also satisfy relation (44). For disk-driven flows with $\Omega = \sqrt{\mathcal{G}\mathcal{M}/\varpi_i^3}$ we can estimate the launching region of the wind (Anderson et al. [1])

$$\varpi_i \approx \left(\frac{4\mathcal{G}\mathcal{M}\varpi_\infty^2 V_{\phi\infty}^2}{V_\infty^4} \right)^{1/3}. \quad (45)$$

Equation (44), with $L = \varpi_A^2 \sqrt{\mathcal{G}\mathcal{M}/\varpi_i^3}$ and ϖ_i from Eq. (45), gives the value of the quantity $\zeta_\infty \varpi_A^2 / \varpi_i^2$ as a function of observable quantities (Ferreira et al. [26])

$$\zeta_\infty \frac{\varpi_A^2}{\varpi_i^2} = \left(\frac{\varpi_\infty V_{\phi\infty} V_\infty}{\sqrt{2}\mathcal{G}\mathcal{M}} \right)^{2/3}. \quad (46)$$

1.4 The Poloidal Components of the Momentum Equation

By analyzing the component of the momentum equation along the flow, we can identify the acceleration mechanisms and estimate the resulting bulk velocities that each mechanism can give. The full analysis can be found in the second part of the chapter, in Sect. 2.2.2. We only need to make the following substitutions since now we are examining nonrelativistic flows: $x \rightarrow 0$, $\gamma \rightarrow 1$, $\xi \rightarrow 1$, $\mu \rightarrow 1 + \mathcal{E}/c^2$, while we drop the electric field and the related force.⁷ Summarizing, the acceleration mechanisms are

⁷ In the second part of the chapter we ignore gravity, which, however, does not play a significant role in the case where the flow is strongly magnetized near the source.

1. The pressure-gradient force $f_{p\parallel}$ that gives velocities of the order of the sound speed.
2. The magnetocentrifugal mechanism that gives velocities of the order of the initial Keplerian speed.
3. The magnetic acceleration that is based on the decline in the poloidal current $|I| = (c/2)\varpi B_\phi$ and gives final speeds at most $\sqrt{2\mathcal{E}}$.

Similarly, the projection of the momentum equation in the transfield direction is analyzed in Sect. 2.2.4. Contrary to the relativistic case where the electric force is important and almost balances the collimating magnetic force, here the curvature radius can be $\mathcal{R} \gtrsim \varpi$, meaning that the magnetic self-collimation is very effective.

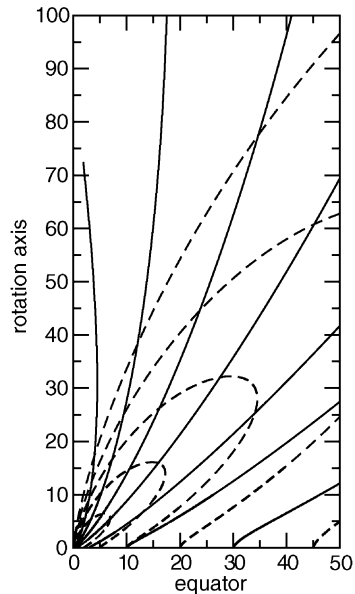
1.5 A Sample Solution

There exist a large number of exact solutions (by means of solving the full momentum equation and not only its component along the flow) of the nonrelativistic, polytropic, axisymmetric, ideal MHD equations. They belong to two categories: semianalytic works (e.g., Blandford and Payne [6]; Contopoulos and Lovelace [15]; Rosso and Pelletier [45]; Ferreira [21]; Vlahakis et al. [62]; Ferreira and Casse [22]) and end-states of time-dependent axisymmetric simulations (e.g., Ouyed and Pudritz [43]; Krasnopolsky et al. [33]; Bogovalov and Tsinganos [8]; Kato et al. [29]; Casse and Keppens [12]; Gracia et al. [25]; Meliani et al. [38]; Fendt [18]; Zanni et al. [64]; Matsakos et al. [37]), to name just a few of them. There also exist 3D simulations with similar results (Ouyed et al. [43]; Anderson et al. [2]). On the other hand, nonpolytropic studies are reviewed in Tsinganos [53] and Sauty [48].

All these solutions obey the analytical results presented in the previous sections. Details such as if the outflow is driven from the inner part of an accretion disk, from a more extended part of the disk, or from a stellar surface, do not substantially change the derived results. What matters is the relative contribution of the various energy flux terms at the ejection surface and the magnetic flux distribution. If the thermal part dominates (this could happen, e.g., in the vicinity of the rotation axis where the Poynting flux vanishes) the flow is thermally-driven. If the Poynting part dominates (this is the case for disk-driven jets, but could also apply to stellar jets, at distances not so close to the rotation axis) the flow is magnetically driven. As an example of a magnetically disk-driven flow, the semi-analytical solution of Vlahakis et al. [62] is presented below.

Figure 5 shows the flow shape and the current lines. Obviously collimation is very efficient, a result mainly of the $\mathbf{J}_p \times \mathbf{B}_\phi/c$ force density. In the region close to the rotation axis, this force density is normal to the current line and points toward the axis. Its projection in the transfield direction collimates the flow, while the projection along the flow has the direction of the flow speed (resulting in acceleration). Note that close to the axis the poloidal current has negative component along \hat{z} (we remind that the $B_\phi < 0$). At large distances from the rotation axis, where the component of \mathbf{J}_p along \hat{z} is positive, the $\mathbf{J}_p \times \mathbf{B}_\phi/c$ force contributes to the

Fig. 5 Field/stream lines (solid lines) and current lines (dashed lines) on the poloidal plane



acceleration, but its transfield component tries to decollimate the flow. Collimation in this regime is mainly due to the poloidal magnetic field force density $\mathbf{J}_\phi \times \mathbf{B}_p/c$.

Figure 6 shows the angle between the flow direction and the equator. Close to the disk the flow makes an angle $\sim 67^\circ$ with the equator, while at large distances it becomes more or less cylindrical. Note that the angle near the disk does not satisfy the criterion of Blandford and Payne [6] according to which plasma parcels move out along the rotating field lines only if the angle between the lines and the equatorial plane is larger than 60° . This is so, because this criterion holds only for cold flows; with a small contribution from the thermal pressure plasma can move out even on lines that make angle larger than 60° with the equatorial plane.

Figure 7 shows the components of the flow speed. Near the disk surface the azimuthal part dominates. In fact V_ϕ increases, in agreement with the expected corotation up to distances comparable to the Alfvén surface. The two poloidal components are comparable up to some distance ($z \sim 10\varpi[z=0]$), after which collimation results in $V_z \gg V_\varpi$. The asymptotic value of the velocity is $\sim 10V_{\phi i}$.

Figure 8 shows the various Mach numbers. The flow crosses the slow magnetosonic and Alfvén surfaces, and becomes superfast at $z \sim 20\varpi(z=0)$.

Figure 9 shows the various contributions of the energy-to-mass flux ratio. Close to the disk the flow is Poynting dominated. The acceleration results in increasing poloidal kinetic energy part and decreasing electromagnetic part. At the fast magnetosonic surface ($z \sim 20\varpi[z=0]$), one-third of the total energy has been transferred to the matter, in accord with the analytical estimation. The acceleration continues after the fast magnetosonic surface and reaches efficiency $\sim 100\%$ asymptotically.

The peculiar behavior at very large distances (e.g., increasing azimuthal speed as seen in Fig. 7, or overfocussing field/stream lines as seen in Fig. 5, marks the

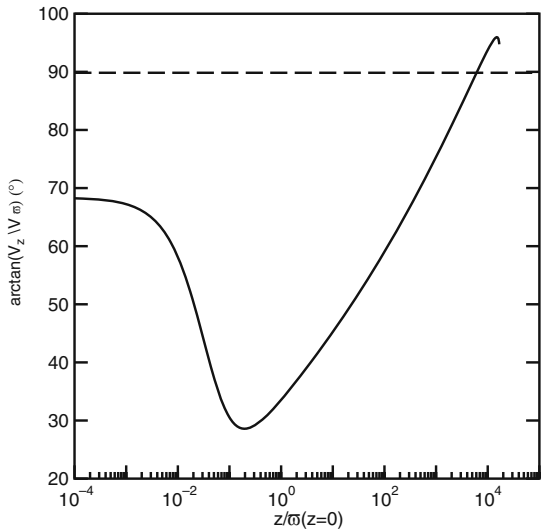


Fig. 6 The angle between the flow direction and the equator, in degrees, as a function of height above the midplane z normalized to the cylindrical distance of the root-point of each field/stream line $\varpi(z = 0)$

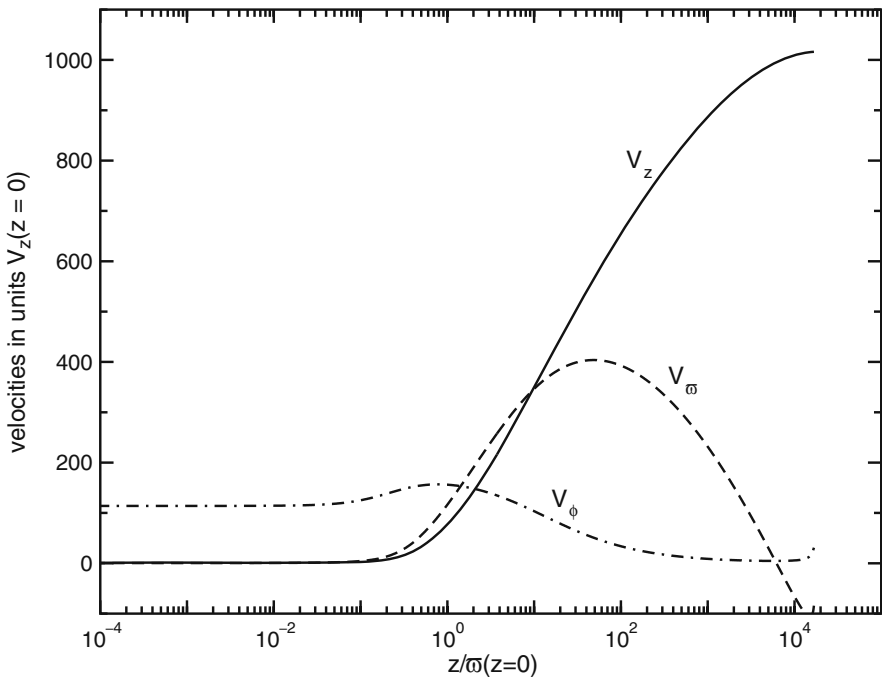


Fig. 7 Components of the flow speed normalized to the value of V_z at the root-point of each field/stream line

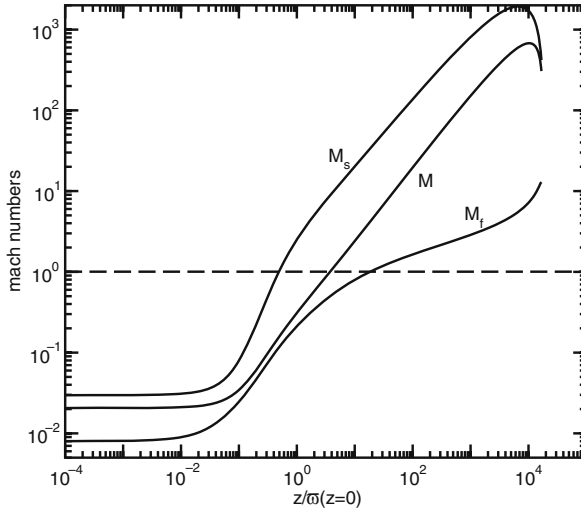


Fig. 8 Slow magnetosonic ($M_s = V_p/V_s$), Alfvén ($M = V_p/V_{Ap}$), and fast magnetosonic ($M_f = V_p/V_f$) Mach numbers. When each of these numbers become unity the flow crosses the corresponding critical surface

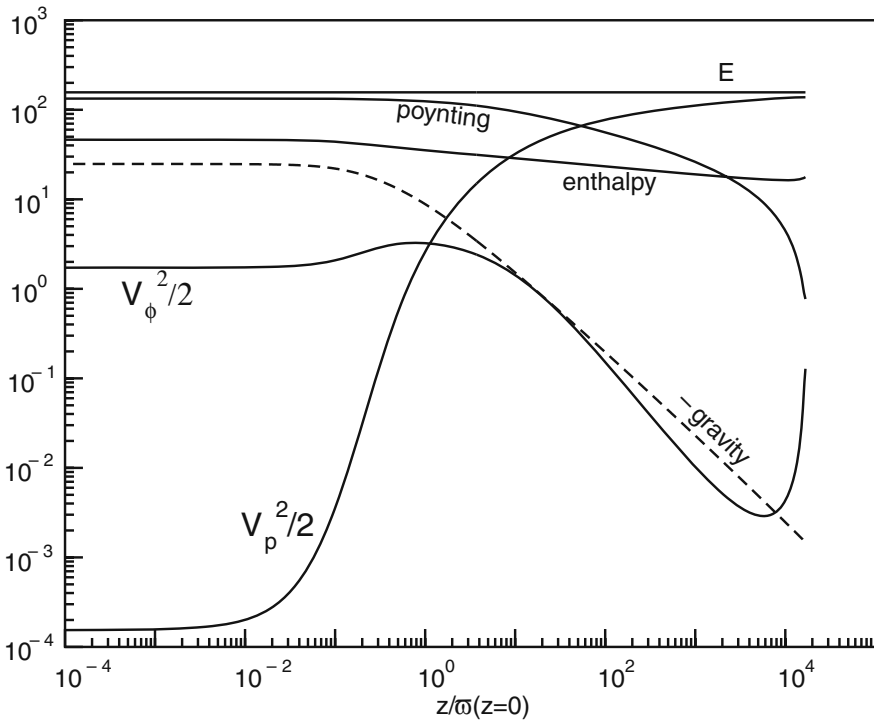


Fig. 9 The various contributions of the energy-to-mass flux ratio

regime where radially self-similarity no longer holds. Another limitation of radially self-similar models is that they give infinite values for various physical quantities near the rotation axis (see e.g., Ferreira [23]). Nevertheless, the solution captures all the basic characteristics of a magnetized disk-driven outflow. In fact, simulations extending this particular self-similar solution by removing the singularities on the rotation axis show that the flow remains very close to the analytical result (Gracia et al. [25]; Matsakos et al. [37]).

2 Relativistic MHD

A flow can be relativistic in four ways:

1. The bulk velocity can be relativistic, $V \lesssim c$.
2. The particle's motion in the rest frame of the flow can be relativistic, or equivalently, the particle's kinetic energy can become comparable or larger than their rest energy. Since the particle's kinetic energy is $\sim k_B T$, this regime is characterized by relativistic temperatures $k_B T \gtrsim mc^2$.
3. Relativistic effects for the electromagnetic field are measured through the light surface whose cylindrical distance is $\varpi = c/\Omega$. The flow can reach and cross this surface.
4. General relativistic effects become important if the distances from the central mass \mathcal{M} are $r \sim \mathcal{G}\mathcal{M}/c^2$.

In the following we analyze relativistic flows covering the first three cases.⁸ For this purpose, it is enough to use special relativity.

2.1 Equations of Steady, Axisymmetric, Special Relativistic, Magnetized Outflows

The stress-energy tensor of relativistic MHD consists of two parts – matter (subscript M) and electromagnetic fields (subscript EM): $T^{\kappa\nu} = T_M^{\kappa\nu} + T_{EM}^{\kappa\nu}$ ($\kappa, \nu = 0, 1, 2, 3$). The former is given by $T_M^{\kappa\nu} = (\rho_0 c^2 + \rho_0 e + P) U^\kappa U^\nu / c^2 + P \eta^{\kappa\nu}$, where P is the pressure, $U^\nu = (\gamma c, \gamma \mathbf{V})$ is the fluid four-velocity, and $\eta^{\kappa\nu} = \text{diag}(-1 \ 1 \ 1 \ 1)$ is the metric tensor (assuming a flat spacetime and Cartesian space coordinates x_j , $j = 1, 2, 3$). Here ρ_0 is the rest-mass density, $\rho_0 e = P/(\Gamma - 1)$ is the internal energy density, with Γ denoting the polytropic index ($= 4/3$ or $5/3$ in the limit of an ultrarelativistic or a nonrelativistic temperature, respectively), \mathbf{V} is the three-velocity measured in the frame of the central object, and $\gamma = 1/(1 - V^2/c^2)^{1/2}$ is the Lorentz factor.

⁸ The last one can be treated in the framework of general relativity, either in a fixed background spacetime (Schwarzschild or Kerr), or taking into account the influence of the stress energy of the matter and the electromagnetic field to the spacetime (i.e., solving also Einstein's equations).

Introducing the specific relativistic enthalpy ξc^2 , where

$$\xi = \frac{\rho_0 c^2 + \rho_0 e + P}{\rho_0 c^2} = 1 + \frac{\Gamma}{\Gamma - 1} \frac{P}{\rho_0 c^2}, \quad (47)$$

and including the contribution of the electric (\mathbf{E}) and magnetic (\mathbf{B}) fields (measured in the central object frame), the components of the total stress-energy tensor take the form ($j, k = 1, 2, 3$)

$$T^{00} = \gamma^2 \xi \rho_0 c^2 - P + \frac{E^2 + B^2}{8\pi}, \quad (48)$$

$$T^{0j} = T^{j0} = \left(\xi \rho_0 c \gamma^2 \mathbf{V} + \frac{\mathbf{E} \times \mathbf{B}}{4\pi} \right) \cdot \mathbf{x}_j, \quad (49)$$

$$T^{jk} = \xi \rho_0 \gamma^2 V_j V_k - \frac{E_j E_k + B_j B_k}{4\pi} + \left(P + \frac{E^2 + B^2}{8\pi} \right) \eta^{jk}, \quad (50)$$

with T^{00} , $cT^{0j}x_j$, and T^{jk} representing the energy density, energy flux, and spatial stress contributions, respectively.

The steady electromagnetic field obeys Maxwell's equations

$$\nabla \cdot \mathbf{B} = 0, \quad \nabla \cdot \mathbf{E} = \frac{4\pi}{c} J^0, \quad \nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{J}, \quad \nabla \times \mathbf{E} = 0, \quad (51)$$

where $J^\nu = (J^0, \mathbf{J})$ is the four-current (with J^0/c representing the charge density). Under the assumption of ideal MHD, the comoving electric field is 0, which implies

$$\mathbf{E} = -\frac{\mathbf{V}}{c} \times \mathbf{B}. \quad (52)$$

The mass conservation equation is $(\rho_0 U^\nu)_{,\nu} = 0$, or

$$\nabla \cdot (\gamma \rho_0 \mathbf{V}) = 0. \quad (53)$$

In the absence of a gravitational field or any other external force, the equations of motion are $T^{\kappa\nu}_{,\nu} = 0$. The entropy conservation equation (the first law of thermodynamics) is obtained by setting $U_\kappa T^{\kappa\nu}_{,\nu} = 0 \Leftrightarrow \mathbf{V} \cdot \nabla e + P \mathbf{V} \cdot \nabla (1/\rho_0) = 0$, which can be rewritten (using the equation of state $\rho_0 e = P/[\Gamma - 1]$) as

$$\mathbf{V} \cdot \nabla (P/\rho_0^\Gamma) = 0. \quad (54)$$

An equivalent expression for the specific entropy conservation in adiabatic flows, independent on the equation of state, is $\mathbf{V} \cdot \nabla P = \rho_0 c^2 \mathbf{V} \cdot \nabla \xi$.

The momentum conservation equation is given by the $\kappa = 1, 2, 3$ components of $T^{\kappa\nu}_{,\nu} = 0$,

$$\gamma\rho_0(\mathbf{V} \cdot \nabla)(\xi\gamma\mathbf{V}) = -\nabla P + (\mathbf{J}^0\mathbf{E} + \mathbf{J} \times \mathbf{B})/c. \quad (55)$$

One can carry out a partial integration of Eqs. (51), (52), (53), (54) and (55) under the assumptions of axisymmetry [in cylindrical coordinates (ϖ, ϕ, z) , $\partial/\partial\phi = 0$] and of a zero azimuthal electric field ($E_\phi = 0$) (e.g., Bekenstein and Oron [4]; Lovelace et al. [35]; Okamoto [42]; Camenzind [10]).

The solenoidal condition on the magnetic field, $\nabla \cdot \mathbf{B} = 0$, implies that there is a poloidal magnetic flux function $A(\varpi, z)$, defined by $2\pi A = \iint \mathbf{B}_p \cdot d\mathbf{S}$, which satisfies

$$\mathbf{B} = \mathbf{B}_p + \mathbf{B}_\phi, \quad \mathbf{B}_p = \nabla A \times \hat{\phi}/\varpi, \quad (56)$$

where the subscripts p and ϕ denote the poloidal and azimuthal components, respectively. Furthermore, Eq. (52) together with the condition $E_\phi = 0$ implies $\mathbf{V}_p \parallel \mathbf{B}_p$, from which it follows that there are functions Ψ_A and Ω (whose coordinate dependence we discuss below) such that

$$\mathbf{V} = \frac{\Psi_A}{4\pi\gamma\rho_0}\mathbf{B} + \varpi\Omega\hat{\phi}, \quad \frac{\Psi_A}{4\pi\gamma\rho_0} = \frac{V_p}{B_p}. \quad (57)$$

The continuity equation (53), using Eq. (57), yields $\mathbf{B}_p \cdot \nabla\Psi_A = 0 \Leftrightarrow \Psi_A = \Psi_A(A)$, i.e., the quantity Ψ_A is a constant of motion.

The combination of Eqs. (52) and (57) gives

$$\mathbf{E} = -\frac{\Omega}{c}\nabla A, \quad E = \frac{\varpi\Omega}{c}B_p. \quad (58)$$

Substituting the previous expression for the electric field in Faraday's law we find $\nabla\Omega \times \nabla A = 0 \Leftrightarrow \Omega = \Omega(A)$, i.e., the quantity Ω is a constant of motion.

The entropy conservation equation (54) gives the adiabat $P/\rho_0^\Gamma = Q(A)$, i.e., P/ρ_0^Γ is a conserved quantity along the flow, related to the specific entropy.

Two more conserved quantities can be found from the integration of the two components of the momentum equation (55) along the azimuthal direction $\hat{\phi}$, and along the flow \mathbf{V} . The first is related to the total specific angular momentum and the second to the total energy-to-mass flux ratio. The following properties are helpful in deriving these two remaining integrals: (1) For an axisymmetric vector \mathbf{F} , the poloidal part of $\nabla \times \mathbf{F}$ is $\nabla \times (F_\phi \hat{\phi}) = \nabla(\varpi F_\phi) \times \hat{\phi}/\varpi$ and its azimuthal part is $\nabla \times \mathbf{F}_p$. (2) The equality $(\mathbf{F} \cdot \nabla)\mathbf{F} = \nabla(F^2/2) + (\nabla \times \mathbf{F}) \times \mathbf{F}$ is an identity. We first apply property (2) for the vector $\mathbf{F} = \xi\gamma\mathbf{V}$ in the momentum equation (55). Then, dotting with $\hat{\phi}$ yields $\gamma\rho_0[\nabla \times (\xi\gamma V_\phi \hat{\phi})] \cdot \mathbf{V}_p = \mathbf{J}_p \times \mathbf{B}_p/c$. Substituting $\mathbf{J}_p = (c/4\pi)\nabla \times (B_\phi \hat{\phi})$ and using property (1) for the vectors $\xi\gamma V_\phi \hat{\phi}$ and $B_\phi \hat{\phi}$ yields $4\pi\gamma\rho_0 \mathbf{V}_p \cdot \nabla(\xi\gamma\varpi V_\phi) = \mathbf{B}_p \cdot \nabla(\varpi B_\phi)$. Using the poloidal part of Eq. (57) we finally get that the quantity $\xi\gamma\varpi V_\phi - \varpi B_\phi/\Psi_A$ is a field line constant. This integral represents the total specific angular momentum conservation. For the integral related to energy conservation, we apply property (2) for the vector $\mathbf{F} = \xi\gamma\mathbf{V}$ in the momentum

equation (55) and then, by dotting with \mathbf{V} , we get $(\rho_0/\xi)\mathbf{V} \cdot \nabla(\xi^2\gamma^2V^2/2) + \mathbf{V} \cdot \nabla P = \mathbf{V} \cdot (\mathbf{J} \times \mathbf{B}/c)$. On the left-hand side we use the identity $\gamma^2V^2/c^2 = \gamma^2 - 1$, and also the $\mathbf{V} \cdot \nabla P = \rho_0c^2\mathbf{V} \cdot \nabla\xi$ (see the comment after Eq. (54)). The left-hand side becomes then $\gamma\rho_0\mathbf{V}_p \cdot \nabla(\xi\gamma c^2)$. The right-hand side, using Eq. (57), can be written as $\varpi\Omega\hat{\phi} \cdot (\mathbf{J}_p \times \mathbf{B}_p/c)$. Substituting $\mathbf{J}_p = (c/4\pi)\nabla \times (B_\phi\hat{\phi})$ and using property (1) we get $\mathbf{B}_p \cdot \nabla(\varpi\Omega B_\phi/4\pi)$. Equating the left- and right-hand sides and using the poloidal part of Eq. (57) we finally get that the quantity $\xi\gamma c^2 - \varpi\Omega B_\phi/\Psi_A$ is a field line constant. This integral represents the total energy-to-mass flux ratio.

Summarizing, the full set of steady-steady equations can be partially integrated to yield five field line constants:

- (a) The mass-to-magnetic flux ratio

$$\Psi_A = \Psi_A(A) = 4\pi\gamma\rho_0V_p/B_p. \quad (59)$$

- (b) The field angular velocity, which equals the matter angular velocity at the foot-point of the field line at the midplane of the disk,

$$\Omega = \Omega(A) = \frac{V_\phi}{\varpi} - \frac{V_p}{\varpi} \frac{B_\phi}{B_p}. \quad (60)$$

- (c) The total (matter + magnetic) specific angular momentum,

$$L = L(A) = \xi\gamma\varpi V_\phi - \varpi B_\phi/\Psi_A. \quad (61)$$

- (d) The total energy-to-mass flux ratio μc^2 , where

$$\mu = \mu(A) = \xi\gamma - \frac{\varpi\Omega B_\phi}{\Psi_A c^2}. \quad (62)$$

- (e) The adiabat

$$Q = Q(A) = \frac{P}{\rho_0^\Gamma}. \quad (63)$$

Equation (63) is the usual polytropic relation between density and pressure. The polytropic index takes the value $4/3$ if the temperatures are relativistic, and $5/3$ if not. Any value of Γ other than $4/3$ or $5/3$ would imply a nonadiabatic evolution and hence require the incorporation of heating/cooling terms into the entropy and momentum equations.

Two integrals remain to be performed, involving the Bernoulli and transfield equations. There are correspondingly two unknown functions, which we choose to be the magnetic flux function A , and the “Alfvénic” Mach number (e.g., Michel [40])

$$M \equiv \left(\frac{4\pi\rho_0\xi\gamma^2V_p^2}{B_p^2} \right)^{1/2} = \left(\frac{\xi\Psi_A^2}{4\pi\rho_0} \right)^{1/2}. \quad (64)$$

We define the cylindrical radius of the field line in units of the light surface radius,

$$x \equiv \frac{\varpi\Omega}{c}. \quad (65)$$

We also define the Alfvénic lever arm by $\varpi_A \equiv (L/\mu\Omega)^{1/2}$ [and correspondingly $x_A \equiv \varpi_A\Omega/c = (L\Omega/\mu c^2)^{1/2}$] and use it to scale the cylindrical radius of the field line by introducing

$$G \equiv \frac{\varpi}{\varpi_A} = \frac{x}{x_A}. \quad (66)$$

We now give the expressions for the physical quantities in terms of the defined variables and the explicit expressions for the Bernoulli and transfield equations.

We may write for the physical quantities

$$\rho_0 = \frac{\xi\Psi_A^2}{4\pi M^2}, \quad P = Q \left(\frac{\xi\Psi_A^2}{4\pi M^2} \right)^\Gamma, \quad \xi = 1 + \frac{\Gamma}{\Gamma-1} \frac{Q}{c^2} \left(\frac{\xi\Psi_A^2}{4\pi M^2} \right)^{\Gamma-1}. \quad (67)$$

$$\mathbf{B} = \frac{\nabla A \times \hat{\phi}}{\varpi} - \frac{\mu c \Psi_A x_A^2}{x} \frac{1-G^2}{1-M^2-x^2} \hat{\phi}, \quad (68)$$

$$\mathbf{E} = -\frac{\Omega}{c} \nabla A, \quad (69)$$

$$\gamma = \frac{\mu}{\xi} \frac{1-M^2-x_A^2}{1-M^2-x^2}, \quad (70)$$

$$\gamma \frac{V}{c} = \frac{M^2}{c\xi\Psi_A} \frac{\nabla A \times \hat{\phi}}{\varpi} + \frac{\mu x_A}{\xi G} \frac{G^2-M^2-x^2}{1-M^2-x^2} \hat{\phi}, \quad (71)$$

where the expressions for γ , V_ϕ , and B_ϕ come from a combination of Eqs. (60), (61), and (62).

Knowing the five field line constants (Ω , Ψ_A , L , μ , Q), or equivalently the (ϖ_A , x_A , σ_M , μ , q), where

$$\sigma_M \equiv \frac{A\Omega^2}{c^3\Psi_A}, \quad q \equiv \frac{\Psi_A^2}{4\pi} \left(\frac{\Gamma}{\Gamma-1} \frac{Q}{c^2} \right)^{\frac{1}{\Gamma-1}}, \quad (72)$$

we can find the quantities A , M , x , G , ξ by solving the following system of equations:

$$G = \frac{\varpi}{\varpi_A}, \quad x = x_A G, \quad M^2 = q \frac{\xi}{(\xi-1)^{\frac{1}{\Gamma-1}}}, \quad (73)$$

the Bernoulli equation

$$\frac{\mu^2}{\xi^2} \frac{G^2 (1 - M^2 - x_A^2)^2 - x_A^2 (G^2 - M^2 - x^2)^2}{G^2 (1 - M^2 - x^2)^2} = 1 + \left[\frac{\sigma_M M^2 \varpi \nabla A}{\xi x^2 A} \right]^2, \quad (74)$$

which is obtained after substituting all quantities in the identity $\gamma^2 - (\gamma V_\phi/c)^2 = 1 + (\gamma V_p/c)^2$ using Eqs. (70) and (71), and which in the nonrelativistic limit takes the familiar form (after Taylor expanding in $1/c^2$)

$$\frac{V^2}{2} + \frac{\Gamma}{\Gamma - 1} \frac{P}{\rho_0} - \frac{\varpi \Omega B_\phi}{\Psi_A} = (\mu - 1)c^2,$$

and the transfield equation (obtained from the component of the momentum equation along $-\nabla A$)

$$\begin{aligned} & \left[x^2 (\nabla A)^2 \frac{d \ln (x_A / \varpi_A)}{dA} - \bar{L} A (1 - M^2 - x^2) \right] \left(\frac{\nabla A}{\varpi} \right)^2 + \\ & + \left[\frac{2x_A^2}{\varpi_A^3 G} (\nabla A)^2 + \frac{\mu^2 x_A^6 A^2}{\varpi_A^5 \sigma_M^2 M^2 G^3} \left(\frac{G^2 - M^2 - x^2}{1 - M^2 - x^2} \right)^2 \right] \hat{\omega} \cdot \nabla A - \\ & - \frac{M^2}{2} \nabla \left[\left(\frac{\nabla A}{\varpi} \right)^2 \right] \cdot \nabla A - \frac{\Gamma - 1}{\Gamma} \nabla \left[\frac{\xi(\xi - 1)}{M^2} \frac{A^2 x_A^4}{\sigma_M^2 \varpi_A^4} \right] \cdot \nabla A - \\ & - \frac{1}{2\varpi^2} \nabla \left[\frac{\mu^2 A^2 x_A^6}{\sigma_M^2 \varpi_A^2} \left(\frac{1 - G^2}{1 - M^2 - x^2} \right)^2 \right] \cdot \nabla A = 0, \end{aligned} \quad (75)$$

where the operator $\bar{L} \equiv \nabla^2 - (2/\varpi) \hat{\omega} \cdot \nabla$ is related to the curvature radius of the poloidal field lines $\mathcal{R} = |\nabla A| (\bar{L} A - \nabla A \cdot \nabla \ln |\nabla A / \varpi|)^{-1}$.

Solving the MHD equations requires the specification of seven constraints (corresponding to the seven unknowns A , B_ϕ , V_ϖ , V_z , V_ϕ , ρ_0 , and P), of which four are associated with boundary conditions at the source and three are determined by the regularity requirements at three singular surfaces that are related to the Alfvén, slow, and fast magnetosonic waves. In general, if we start the integration downstream from some surface, the number of boundary conditions which we must give on this surface is equal to the number of waves which can be emitted from this surface downstream. (These waves which can propagate from a given surface are: one entropy, two Alfvén, two slow and two fast magnetosonic waves). The remaining conditions (until reach the number seven) are specified in order to pass through singular surfaces. When we pass through these surfaces the number of outgoing waves is changed. Thus, the component of the velocity perpendicular to a singular surface equals the phase velocity of the Alfvén, slow or fast magnetosonic wave. For example, if we begin the integration from some surface with the component of the velocity perpendicular to this surface smaller than the cor-

responding component of the slow speed, then we must give four boundary conditions (related to the entropy wave, one Alfvén, one slow and one fast wave). The other three are specified in order to pass the solution through the three singular surfaces associated to the slow magnetosonic, Alfvén, and fast magnetosonic waves. If, on the other hand, the initial velocity is super-slow magnetosonic but sub-Alfvénic, the number of outgoing waves is five (one entropy, two slow, one Alfvén, and one fast), and we should give five boundary conditions at the base of the flow.

2.2 General Characteristics of Magnetized Outflows

We are interested in examining initially Poynting-dominated outflows, meaning that the electromagnetic part of the energy-to-mass flux ratio is much larger than the corresponding part for the matter, i.e., initially (subscript i) $\varpi_i \Omega |B_\phi|_i / \Psi_A c^2 \approx \mu \gg \xi_i \gamma_i$ (see Eq. (62)). The value of the field line constant μ should be $\gg 1$ such that the maximum possible Lorentz factor is $\mu \gg 1$, corresponding to relativistic flows.

We also restrict our analysis to outflows that initially have sub-Alfvénic velocities. The Alfvén surface is defined as the locus of points where static Alfvén waves with wavevector in any direction in the meridional plane (in the central object's frame) can exist (i.e., Eq. (100) with $\omega = 0$ and $k_\phi = 0$ is satisfied). An equivalent statement is that the Alfvén surface marks the locus of points where the flow proper velocity in any direction in the meridional plane is equal to the comoving proper phase speed of an Alfvén wave that propagates in that direction (see Eq. (102)). Equivalently, at the Alfvénic surface $4\pi\rho_0\xi\gamma^2V_p^2 = B_p^2(1-x^2) \Leftrightarrow M^2 = 1 - x^2$. The expressions for B_ϕ , γ , V_ϕ (see Eqs. (68), (70), (71)) imply that at the Alfvén surface $G = 1$, or $x = x_A$, and $\varpi = \varpi_A$, justifying the definition of $\varpi_A = (L/\mu\Omega)^{1/2}$ being the Alfvénic lever arm. Thus, sub-Alfvénic flow means that $x_i \ll x_A$ and $M_i^2 \ll 1 - x_A^2$. By employing Eq. (70) we find that $\xi_i \gamma_i \approx \mu(1 - x_A^2)/(1 - x_i^2) \approx \mu(1 - x_A^2)$. Since $\mu \gg 1$, the value of the field line constant $x_A^2 \approx 1 - \xi_i \gamma_i / \mu \approx 1^-$, meaning that the Alfvén and light surfaces ($x = 1$) almost coincide (the Alfvénic surface is, however, always located inside the light surface).

2.2.1 Forces in the Poloidal Plane

The momentum equation (55) can be written as the sum of the following force densities (for simplicity we use hereafter the term force):

$$f_G + f_T + f_C + f_I + f_P + f_E + f_B = 0, \quad (76)$$

where

$$\begin{aligned}
f_G &= -\gamma \rho_0 \xi (\mathbf{V} \cdot \nabla \gamma) \mathbf{V} \\
f_T &= -\gamma^2 \rho_0 (\mathbf{V} \cdot \nabla \xi) \mathbf{V} && : \text{temperature force} \\
f_C &= \hat{\omega} \gamma^2 \rho_0 \xi V_\phi^2 / \varpi && : \text{centrifugal force} \\
f_I &= -\gamma^2 \rho_0 \xi (\mathbf{V} \cdot \nabla) \mathbf{V} - f_C \\
f_P &= -\nabla P && : \text{pressure force} \\
f_E &= (\nabla \cdot \mathbf{E}) \mathbf{E} / 4\pi && : \text{electric force} \\
f_B &= (\nabla \times \mathbf{B}) \times \mathbf{B} / 4\pi && : \text{magnetic force}
\end{aligned}$$

The ‘gamma’ force f_G further decomposes into two terms: $f_G = f_{G_p} + f_{G_\phi}$, with

$$f_{G_p} = -\frac{\gamma^4 \rho_0 \xi}{2c^2} (\mathbf{V} \cdot \nabla V_p^2) \mathbf{V}, \quad f_{G_\phi} = -\frac{\gamma^4 \rho_0 \xi}{2c^2} (\mathbf{V} \cdot \nabla V_\phi^2) \mathbf{V}.$$

The poloidal part of the f_I force is

$$-\gamma^2 \rho_0 \xi (\mathbf{V}_p \cdot \nabla) \mathbf{V}_p = -\gamma^2 \rho_0 \xi \left(V_p^2 \frac{\partial \vartheta}{\partial \ell} \frac{\nabla A}{|\nabla A|} + \frac{\partial V_p}{\partial \ell} \mathbf{V}_p \right),$$

where ℓ is the arclength along the poloidal field line, ϑ is the angle between the poloidal magnetic field and the rotation axis ($\sin \vartheta = V_\omega / V_p$), and the derivative $\partial / \partial \ell = \sin \vartheta \partial / \partial \varpi$ is taken keeping A constant. The radius of curvature of a poloidal field line is $\mathcal{R} = -(\partial \vartheta / \partial \ell)^{-1}$ (positive when the field line bends toward the axis, i.e., when ϑ decreases along the poloidal flow line).

2.2.2 Acceleration

The projection of Eq. (76) along the poloidal flow, i.e., along $\hat{\mathbf{b}} \equiv \mathbf{B}_p / B_p = \sin \vartheta \hat{\omega} + \cos \vartheta \hat{\mathbf{z}}$, is

$$\begin{aligned}
\frac{\gamma^2 \rho_0 \xi}{2} \frac{\partial V_p^2}{\partial \ell} + \frac{\gamma^4 \rho_0 \xi V_p^2}{2c^2} \frac{\partial V_p^2}{\partial \ell} &= -\frac{\gamma^4 \rho_0 \xi V_p^2}{2c^2} \frac{\partial V_\phi^2}{\partial \ell} - \gamma^2 \rho_0 V_p^2 \frac{\partial \xi}{\partial \ell} + \\
+ \gamma^2 \rho_0 \xi \frac{V_\phi^2}{\varpi} \sin \vartheta - \rho_0 c^2 \frac{\partial \xi}{\partial \ell} - \frac{B_\phi}{4\pi \varpi} \frac{\partial (\varpi B_\phi)}{\partial \ell}.
\end{aligned} \tag{77}$$

The terms on the right-hand side of Eq. (77) are recognized as $f_{G_\phi \parallel}$, $f_{T \parallel}$, $f_{C \parallel}$, $f_{P \parallel}$, and $f_{B \parallel}$, respectively, where a subscript \parallel denotes the component of a vector along the poloidal field line. The first term on the left-hand side of Eq. (77) is $-f_{I \parallel}$, whereas the second term is $-f_{G_p \parallel}$ (note that the electric force has no component along the flow, $f_{E \parallel} = 0$). The magnetic force component $f_{B \parallel}$ decomposes into the azimuthal magnetic pressure gradient $-\partial (B_\phi^2 / 8\pi) / \partial \ell$ and the magnetic tension $-B_\phi^2 \sin \vartheta / 4\pi \varpi$. These two parts cancel each other when $B_\phi^2(A, \varpi) \propto 1/\varpi^2$; if $B_\phi^2(A, \varpi)$ decreases faster than ϖ^{-2} then the gradient of the azimuthal magnetic pressure exceeds the magnetic tension, resulting in a positive $f_{B \parallel}$. Thus, the important quantity that determines the magnetic acceleration is the product ϖB_ϕ ,

which is proportional to the Poynting-to-mass flux ratio, since $(c/4\pi)(\mathbf{E} \times \mathbf{B}) \cdot \hat{\mathbf{b}}/(\gamma\rho_0 V_p) = -(\Omega/\Psi_A)(\varpi B_\phi)$ (see also Eq. (62)). The product ϖB_ϕ is also proportional to the total poloidal current $\mathbf{J}_p = (c/4\pi)\nabla \times \mathbf{B}_\phi = (1/2\pi\varpi)\nabla I \times \hat{\phi}$, with $I = \iint \mathbf{J}_p \cdot d\mathbf{S} = (c/2)\varpi B_\phi$; the meridional current lines represent the loci of constant total poloidal current ($I = \text{const} \Leftrightarrow \varpi B_\phi = \text{const}$).

In the nonrelativistic regime ($V \ll c$, $x \ll 1$, $\xi \approx 1$, $f_{G\phi} = f_{T\parallel} = 0$), the pressure force $f_{P\parallel}$ dominates up to the slow magnetosonic point, but the bulk of the acceleration is either magnetocentrifugal – corresponding to the $f_{C\parallel}$ term, which can be interpreted in the ‘bead on a wire’ picture (e.g., Blandford and Payne [6])⁹ – or a consequence of the magnetic pressure-gradient force $f_{B\parallel}$ (which near the surface of the disk can be interpreted in the ‘uncoiling spring’ picture; e.g., Uchida and Shibata [55]).¹⁰

The magnetic force generally becomes important also in flows where the centrifugal acceleration initially dominates: in this case the inertia of the centrifugally accelerated gas amplifies the B_ϕ component, and eventually (beyond the Alfvén point) $f_{B\parallel}$ becomes the main driving force. This force continues to accelerate the flow beyond the fast magnetosonic point (which separates the elliptic and hyperbolic regimes of the MHD partial differential equations).¹¹

In the case where the outflow attains a highly relativistic speed, the centrifugal acceleration cannot play an important role. This is because the non-negligible V_ϕ that would be required in this case would constrain the maximum value of the poloidal speed: $V_p^2 = c^2(1 - 1/\gamma^2) - V_\phi^2 < c^2 - V_\phi^2$. Therefore, in Eq. (77), $f_{C\parallel} \approx 0$ (and also $f_{G\phi} \approx 0$). The $f_{P\parallel}$ force can be neglected since $f_{T\parallel}/f_{P\parallel} = (\gamma V_p/c)^2 \gg 1$. The two remaining terms are $f_{T\parallel}$ (a force with a relativistic origin) and $f_{B\parallel}$. The expressions for these terms in Eq. (77) (or, equivalently, Eq. (62) for the total energy-to-mass flux ratio) indicate that the bulk Lorentz factor can increase in response to the decline in either the enthalpy-to-rest-mass ratio ξ (the thermal acceleration case) or the Poynting-to-mass flux ratio $\propto -\varpi B_\phi$ (the magnetic acceleration case) along the flow. When the temperature is relativistic, the initial acceleration is dominated by the temperature force, but after ξ drops to ~ 1 the magnetic force takes over.

⁹ The strong poloidal magnetic field line plays the role of the wire. In the cold limit one has $M \ll 1$ and $x \ll x_A$ (with $x_A \lesssim 1$) near the base of the flow, implying that the azimuthal field satisfies $-\varpi B_\phi/L\Omega \approx 1 - G^2(1 - x_A^2) + M^2 \approx 1$ and hence that the $f_{B\parallel}$ term is negligible and that a near-corotation ($V_\phi/\varpi\Omega \approx 1 - M^2/G^2 \approx 1$) holds. The small value of M in turn implies a large density and hence a measurable thermal pressure, resulting in a nonnegligible pressure force at the base.

¹⁰ In this picture, the winding-up of the field lines by the disk rotation produces a large azimuthal magnetic field component that is antiparallel to V_ϕ in the northern hemisphere (and parallel to V_ϕ in the southern hemisphere), and a corresponding outward-directed magnetic pressure gradient $-\nabla(B_\phi^2/8\pi)$.

¹¹ At this point, static fast magnetosonic waves with wavevectors parallel to \mathbf{V}_p in the central object’s frame can exist (i.e., Eq. (101) with $\omega = 0$, $\mathbf{k} \parallel \mathbf{V}_p$ is satisfied). The fast magnetosonic point is equivalently defined by the condition that the poloidal proper speed equals the comoving proper phase speed of a fast magnetosonic wave whose wavevector is parallel to \mathbf{V}_p (see Eq. (102)).

When the outflow speed is only mildly relativistic, the magnetocentrifugal force may be important during the initial acceleration phase, especially if the temperatures are nonrelativistic. It is, however, also conceivable that the magnetic pressure-gradient force dominates from the start, as might be the case if the azimuthal field component at the disk surface is large enough.

We can obtain an expression for $f_{C\parallel}$ as follows. By eliminating $\xi\gamma$ between Eqs. (61) and (62), we obtain a relation between ϖV_ϕ and ϖB_ϕ :

$$\frac{\varpi \Omega V_\phi}{c} = 1 - \frac{\mu(1 - x_A^2)}{\mu + \varpi \Omega B_\phi / \Psi_A c^2}, \quad (78)$$

whose divergence along the flow implies

$$\frac{\partial(\varpi V_\phi)}{\partial \ell} = \frac{\mu(1 - x_A^2)}{\Psi_A \xi^2 \gamma^2} \frac{\partial(\varpi B_\phi)}{\partial \ell}. \quad (79)$$

Employing the relations for $f_{B\parallel}$, $f_{C\parallel}$ (see Eq. (77)), and Eqs. (67), (68), (69), (70) and (71), we obtain

$$f_{C\parallel} = -\frac{\xi \rho_0 \gamma^2}{2} \frac{\partial V_\phi^2}{\partial \ell} + \frac{(1 - x_A^2)(1 - M^2 - x^2)}{1 - M^2 - x_A^2} \frac{V_\phi}{V_p} \frac{B_p}{(-B_\phi)} f_{B\parallel}. \quad (80)$$

The first term on the right-hand side of Eq. (80) can give rise to either acceleration (when V_ϕ decreases along the flow line) or deceleration (when V_ϕ increases, as in the corotation regime at the base of the outflow). This term, together with $f_{G_\phi\parallel}$, can lead to a situation in which V_p increases (resp., decreases) and V_ϕ decreases (resp., increases) while the Lorentz factor remains roughly constant. The second term on the right-hand side of Eq. (80), which is proportional to $f_{B\parallel}$, demonstrates that the centrifugal force also has a magnetic component and accounts for the Poynting-to-kinetic energy transfer that underlies the magnetocentrifugal acceleration process (see also Contopoulos and Lovelace [15] for a related discussion). The form of this term makes it clear why the centrifugal force exceeds the magnetic force during the initial stage of the acceleration, when the flow is still nonrelativistic ($x \ll 1$, $M \ll 1$, with $B_p > |B_\phi|$, $V_\phi > V_p$).

The conclusion from the above analysis is that, even though centrifugal and thermal effects could dominate initially, the magnetic force eventually takes over and is responsible for the bulk of the acceleration to high terminal speeds, corresponding to terminal Lorentz factors of the order of the energy integral μ . Li et al. [34] described the efficient conversion of Poynting-to-kinetic energy fluxes in relativistic MHD outflows in terms of a ‘magnetic nozzle’ (see also Vlahakis and Königl [58]). This effect is not inherently relativistic — this conclusion has, in fact, been verified explicitly in the case of the nonrelativistic self-similar solutions constructed by Vlahakis et al. [62].

In general, besides the acceleration due to the decline in the specific enthalpy, the only other available acceleration mechanism is based on the decline of the poloidal

current, i.e., the quantity $-\varpi B_\phi$ should be a decreasing function along the flow. In other words, as the flow moves along a particular field line it crosses current lines with decreasing $|I| = (c/2)(-\varpi B_\phi)$, meaning that it is accelerated and the Poynting-to-matter energy flux ratio decreases. Since the quantity ϖV_ϕ is a given function of ϖB_ϕ (see Eqs. (78), (79)), it always follows the change of the poloidal current and increases whenever the Lorentz force accelerates the flow.

2.2.3 The Function $\varpi^2 B_p/A$ and the Asymptotic Lorentz Factor γ_∞

An important combination of the field line constants is the ‘Michel’s magnetization parameter’ $\sigma_M(A) = A\Omega^2/\Psi_A c^3$, already defined in Eq. (72). In terms of σ_M , and using Eqs. (68), (69) and (70), we may write the exact expression

$$\frac{\mu - \xi\gamma}{\mu} = \left(\frac{\sigma_M}{\mu}\right) \left(\frac{-B_\phi}{E}\right) \left(\frac{B_p \varpi^2}{A}\right) \quad (81)$$

The left-hand side represents the Poynting-to-total energy flux. As long as the flow is Poynting dominated ($\gamma \ll \mu$), this ratio is close to unity. This continues to be the case in the neighborhood of the fast magnetosonic surface, where $\gamma \approx \mu^{1/3}$ (e.g., Camenzind [11]).¹² As a result, the superfast regime of the flow is the only place where a transition from a Poynting- to a matter-dominated flow is possible. In this regime, and for extremely relativistic flows, the term $(-B_\phi/E)$ is very close to unity.¹³ Hence, Eq. (81) gives a simple relation between the Lorentz factor and the function $B_p \varpi^2/A$:

$$\frac{\mu - \gamma}{\mu} \approx \frac{\sigma_M}{\mu} \frac{B_p \varpi^2}{A}. \quad (82)$$

Suppose that the value of the function $B_p \varpi^2/A$ near the fast surface is $(B_p \varpi^2/A)_f$. Since $\gamma \ll \mu$ at this point, Eq. (82) implies that the constant of motion $\sigma_M/\mu \approx 1/(B_p \varpi^2/A)_f$.

¹² For a cold flow (as it is practically the case near the fast magnetosonic surface) the fast waves have phase velocities $(\omega_{co}/k_{co})^2 = v_A^2$, as measured in the comoving frame, with $v_A^2/(1 - v_A^2/c^2) = B_{co}^2/4\pi\rho_0$ (Appendix). In the super-Alfvénic regime where the fast surface is located, the azimuthal component of the magnetic field dominates over the poloidal part, and $|B_\phi| \approx E = x B_p$. So, the comoving magnetic field is $B_{co} \approx x B_p/\gamma$. Thus $(\omega_{co}/ck_{co})^2 [1 - (\omega_{co}/ck_{co})^2]^{-1} \approx x^2 B_p^2/4\pi\rho_0 c^2 \gamma^2$. At the fast surface, one of the two fast waves propagating along V_p is static (zero phase velocity) in the central object’s frame. Equivalently $\omega_{co}/k_{co} = V_p \Leftrightarrow (\gamma V_p/c)^2 \approx x^2 B_p^2/4\pi\rho_0 c^2 \gamma^2$ or $\gamma^3 \approx \gamma x^2/M^2 \approx \mu$.

¹³ The requirement that the Lorentz invariant $B^2 - E^2 > 0$, using $E = x B_p$, gives $B_\phi^2/E^2 > 1 - 1/x^2$. In addition, Eq. (60) gives $V_\phi/c = x + (V_p/c)(B_\phi/B_p)$, and the condition $V_\phi \gtrsim 0$ implies $-B_\phi/E \lesssim c/V_p$. Thus, $(1 - 1/x^2)^{1/2} < -B_\phi/E \lesssim c/V_p$. A careful analysis gives the approximate expression $-B_\phi/E \approx [(1 - 1/x^2)/(1 - 1/\gamma^2)]^{1/2}$, for $x \gg 1$ (see Eq. (13) in Vlahakis [57]).

Denoting with $\delta\ell_\perp$ the distance between two neighboring poloidal field lines A and $A + \delta A$, magnetic flux conservation implies $B_p \varpi^2 = (\varpi/\delta\ell_\perp)\delta A$. Thus, a decreasing $B_p \varpi^2$ – and hence, from Eq. (82), an accelerating flow – corresponds to poloidal field lines expanding in a way such that their distance $\delta\ell_\perp$ increases faster than ϖ . How fast the field lines expand is determined by the trans-field force-balance equation; thus, Eq. (75) indirectly determines the flow acceleration. Since the available solid angle for expansion of the field lines is finite, there is a minimum value of the $B_p \varpi^2/A$ function. The field lines asymptotically have a shape $z \approx z_0(A) + \varpi/\tan\vartheta(A)$, where $\vartheta(A)$ is their opening angle (Vlahakis [57]). Differentiating the latter equation we get a decreasing function $B_p \varpi^2/A = (A\vartheta'/\sin\vartheta - Az'_0 \sin\vartheta/\varpi)^{-1}$, reaching a minimum value $\sin\vartheta/A\vartheta'$ at $\varpi \gg z'_0 \sin^2\vartheta/\vartheta'$. Since the factor $\sin\vartheta/A\vartheta'$ is ~ 1 , the minimum value of the $B_p \varpi^2/A$ function is ~ 1 , corresponding to¹⁴

$$\frac{\mu - \gamma_\infty}{\mu} \approx \frac{\sigma_M}{\mu} \left(\frac{B_p \varpi^2}{A} \right)_\infty \approx \frac{(B_p \varpi^2)_\infty}{(B_p \varpi^2)_f} \sim \frac{\sigma_M}{\mu} \sim \frac{1}{(B_p \varpi^2/A)_f}. \quad (83)$$

Equivalently, the asymptotic Lorentz factor is $\gamma_\infty \sim \mu - \sigma_M$, and the asymptotic Poynting-to-mass flux ratio is $\sim \sigma_M c^2$.

Another interesting connection with the boundary conditions near the source can be found, by noting that, as long as $|B_\phi| \approx E = x B_p$, $B_p \varpi^2/A \approx 2|I|/A\Omega$. Thus, $(B_p \varpi^2/A)_f \approx 2|I|_f/A\Omega$, and since $|I|$ remains practically constant of motion inside the force-free subfast regime, $(B_p \varpi^2/A)_f \approx 2|I|_i/A\Omega$, and $\mu/\sigma_M \approx 2|I|_i/A\Omega$. Hence, Eq. (83) implies a direct connection of the acceleration efficiency and the asymptotic Lorentz factor to the ejection characteristics

$$\frac{\mu - \gamma_\infty}{\mu} \approx \frac{A\Omega}{2|I|_i} \left(\frac{B_p \varpi^2}{A} \right)_\infty \sim \frac{A\Omega}{2|I|_i}, \quad \gamma_\infty \sim \mu \left(1 - \frac{A\Omega}{2|I|_i} \right) \quad (84)$$

Summarizing, the acceleration mechanisms are as follows:

1. The pressure-gradient force $f_{P\parallel}$ that works in the mildly relativistic regime and gives velocities of the order of the sound speed.
2. The magnetocentrifugal mechanism that gives velocities of the order of the initial Keplerian speed.
3. In case of relativistic initial temperatures, i.e., if ξ_i is not very close to unity, the temperature force $f_{T\parallel}$ gives final Lorentz factor ξ_i .

¹⁴ The only exception to this general result is to have asymptotically $(\mu - \gamma)/\mu \ll \sigma_M/\mu$ in some finite solid angle regions, combined with other regions with bunched field lines [in which $B_p \varpi^2 \gg A$ and $(\mu - \gamma)/\mu \gg \sigma_M/\mu$]. Note also that the most general asymptotic field line shape slightly deviates from straight lines, resulting in a logarithmic acceleration (Chiueh et al. [13]; Okamoto [41]; Vlahakis [57]). However, this acceleration can happen in exponentially large distances and hence is physically irrelevant.

4. The magnetic acceleration that is based on the decline in the poloidal current $|I| = (c/2)\varpi B_\phi$ and gives final Lorentz factors at most $\sim \mu - \sigma_M \sim \mu - \mu A\Omega/2|I|_i$.

2.2.4 Collimation

The projection of Eq. (76) in the direction perpendicular to the poloidal flow, i.e., along $\hat{n} \equiv \mathbf{E}/E = -\cos \vartheta \hat{\boldsymbol{\omega}} + \sin \vartheta \hat{\mathbf{z}}$, gives the transfield force-balance equation

$$f_{C\perp} + f_{I\perp} + f_{P\perp} + f_{E\perp} + f_{B\perp} = 0,$$

where the subscript \perp denotes the vector component along \hat{n} . The various terms are as follows:

- The azimuthal centrifugal term

$$f_{C\perp} = -\xi \gamma^2 \rho_0 \frac{V_\phi^2}{\varpi} \cos \vartheta = -\frac{B_p^2}{4\pi \varpi} \left(\frac{M V_\phi}{V_p} \right)^2 \cos \vartheta.$$

- The rest of the inertial force along \hat{n}

$$f_{I\perp} = -\gamma^2 \rho_0 \xi \hat{n} \cdot [(\mathbf{V} \cdot \nabla) \mathbf{V}] - f_{C\perp} = -\frac{B_p^2}{4\pi \mathcal{R}} M^2$$

is the poloidal centrifugal term.

- The pressure-gradient force along \hat{n}

$$f_{P\perp} = -\hat{n} \cdot \nabla P.$$

- The ‘electric field’ force

$$f_{E\perp} = \frac{1}{8\pi \varpi^2} \hat{n} \cdot \nabla (\varpi^2 E^2) - \frac{E^2}{4\pi \mathcal{R}}.$$

- The ‘magnetic field’ force along \hat{n}

$$f_{B\perp} = -\frac{1}{8\pi \varpi^2} \hat{n} \cdot \nabla (\varpi^2 B^2) + \frac{B_p^2}{4\pi \mathcal{R}} - \frac{B_p^2}{4\pi \varpi} \cos \vartheta.$$

The total electromagnetic force in the transfield direction ($f_{E\perp} + f_{B\perp}$) can be decomposed as

$$\underbrace{-\frac{1}{8\pi \varpi^2} \hat{n} \cdot \nabla [\varpi^2 (B^2 - E^2)]}_{f_{EM1}} + \underbrace{\frac{B_p^2(1 - x^2)}{4\pi \mathcal{R}}}_{f_{EM2}} + \underbrace{\frac{B_p^2}{4\pi \varpi} \cos \vartheta}_{f_{EM3}} \quad (85)$$

Altogether, they give the following form of the transfield force-balance equation:

$$\begin{aligned} \frac{B_p^2}{4\pi\mathcal{R}} (M^2 + x^2 - 1) = & -\frac{1}{8\pi\varpi^2} \hat{\mathbf{n}} \cdot \nabla [\varpi^2 (B^2 - E^2)] \\ & - \frac{B_p^2}{4\pi\varpi} \cos \vartheta - \frac{B_p^2}{4\pi\varpi} \left(\frac{MV_\phi}{V_p} \right)^2 \cos \vartheta - \hat{\mathbf{n}} \cdot \nabla P. \end{aligned} \quad (86)$$

(Using Eqs. (67), (68), (69), (70) and, the latter gives Eq. (75), a second-order partial differential equation for A .)

The terms on the right-hand side of Eq. (86) are recognized as f_{EM1} , f_{EM3} , $f_{C\perp}$, and $f_{P\perp}$, respectively. The left-hand side consists of the sum of $-f_{I\perp}$ and $-f_{EM2}$ terms that both are proportional to the curvature of the poloidal field lines.

In the sub-Alfvénic regime ($M^2 + x^2 - 1 < 0$) the left-hand side of Eq. (86) has a sign opposite to \mathcal{R} . As explained below, this means that the larger the forces that point away from the rotation axis the faster the collimation!¹⁵ Since the inertial force is negligible, the requirement that the remaining forces cancel each other determines the curvature radius. Besides the pressure-gradient component that is expected to play a significant role only very close to the outflow base, the right-hand side of Eq. (86) consists of the negative terms f_{EM3} , $f_{C\perp}$, and the term f_{EM1} . Exactly at the Alfvénic surface the algebraic sum of these three terms should vanish (since the left-hand side of Eq. (86) vanish; this is the Alfvénic regularity condition). Thus, the term f_{EM1} is positive. In the sub-Alfvénic part, for a collimating flow ($\mathcal{R} > 0$), it is $|f_{EM3} + f_{C\perp}| > f_{EM1}$. The larger the $|f_{EM3}|$, $|f_{C\perp}|$ forces that point away from the axis, the smaller the curvature and the faster the collimation.

In the super-Alfvénic part the positive term f_{EM1} takes over (since the terms related to the poloidal magnetic field as well as the centrifugal force become smaller than the azimuthal magnetic field term¹⁶) and the collimation continues due to this term. By employing the comoving magnetic field $B_{co} = (B^2 - E^2)^{1/2} \approx |B_\phi|/\gamma$, the force $f_{EM1} = -(B_{co}/4\pi\varpi) \hat{\mathbf{n}} \cdot \nabla (\varpi B_{co}) \sim B_\phi^2/4\pi\varpi\gamma^2$ (assuming $|\hat{\mathbf{n}} \cdot \nabla| \sim 1/\varpi$). Thus, Eq. (86) implies (using $|B_\phi| \approx x B_p$)

$$\frac{\varpi}{\mathcal{R}} \leq \frac{1}{\gamma^2} \frac{x^2}{x^2 + M^2} \lesssim \frac{1}{\gamma^2}.$$

This is a well-known result, meaning that the field lines are difficult to bend when the flow is relativistic, because (1) the electric force almost cancel the transfield component of the magnetic force and (2) the effective matter inertia is larger (e.g., Bogovalov [7]). The result is negligible curvature, $\varpi/\mathcal{R} \approx 0$ (e.g., Chiueh et al.

¹⁵ The situation is similar to a steady, hydrodynamic wind, where the stronger gravity contributes to the faster acceleration in the sub-sonic regime.

¹⁶ See, however, the analysis of cases where the term $|f_{EM3} + f_{C\perp}|$ is comparable, or even dominates over the f_{EM1} term, in Komissarov et al. [31]. A detailed study on the flow shape in connection with the external pressure that confines the jet, the spatial growth of the Lorentz factor, and the causality across the jet, can also be found in that paper.

[14]), and only close to the origin where the Lorentz factor is relatively small is collimation efficient. The relation $\varpi/\mathcal{R} \lesssim 1/\gamma^2$, however, does not mean that the field lines are exactly straight. For example, for a Lorentz factor $\gamma \propto \varpi^\beta$, the shape of the poloidal field lines could be $z \propto \varpi^\alpha$, with $\alpha \sim \beta + 1$.

On the other hand, even if the poloidal lines are exactly straight ($\varpi/\mathcal{R} = 0$), acceleration is possible. The most general line shape with zero curvature is $z = z_0(A) + \varpi/\tan \vartheta(A)$, corresponding (as we explained in Sect. 2.2.3) to an accelerating flow.¹⁷

2.3 Quasi-monopolar Solutions

Michel's [40] solution, the first try to obtain relativistic MHD solutions, is a generalization of the Weber and Davis [63] classical paper. Exactly as the Weber–Davis model, it applies only to the equatorial plane and ignores the transfield force balance. By assuming a cold flow, a monopole magnetic field $A = B_0 \varpi_0^2 (1 - \cos \theta) \Leftrightarrow \mathbf{B}_p = B_0 (\varpi_0/r)^2 \hat{\mathbf{r}}$, and by restricting the analysis to the plane $\theta = \pi/2$ (when $\varpi|\nabla A|/A = 1$), Eq. (74) becomes

$$\frac{[\sigma_M \mu (1 - x_A^2) - \mu x^2 U]^2 - x^2 [\sigma_M \mu (1 - x_A^2) - \mu x_A^2 U]^2}{[\sigma_M (1 - x^2) - x^2 U]^2} = 1 + U^2, \quad (87)$$

where $U = \gamma V_p/c = \sigma_M M^2/x^2$ is the proper velocity (see Eq. (71)). Equation (87) can be solved for the function $U(x)$, i.e., the flow speed as function of the distance, if the constants of motion σ_M , μ , and x_A are known. From the requirement to have $U \rightarrow 0$ in the limit $x \rightarrow 0$, we get a relation between the two constants $\mu(1 - x_A^2) = 1$, so only σ_M , μ are free. For each μ we can find the solution $U(x)$ that starts from the point $U(x = 0) = 0$. For small enough values of μ the solution does not extend to infinity (see the curve ' $\mu < \mu_c$ ' in Fig. 10). The smaller μ for which the solution goes to $x \rightarrow \infty$ is the critical value $\mu_c = \left(1 + \sigma_M^{2/3}\right)^{3/2}$. For this value the asymptotic proper velocity is $U_\infty = \sigma_M^{1/3}$, and since for $\sigma_M \gg 1$ the critical value is $\mu_c \approx \sigma_M$, it is $U_\infty \approx \mu^{1/3}$. For larger values of μ the solution extends up to infinity, but the asymptotic velocity is smaller than $\sigma_M^{1/3}$ (see Fig. 10). The critical solution is the 'minimum torque' solution that represents the appropriate physical solution (Michel [40]). The solution $U_\infty = \sigma_M^{1/3}$ is a double root of Eq. (87) at $x \rightarrow \infty$, meaning that the fast magnetosonic surface is located at infinity.

Although Michel's solution was the first step toward understanding relativistic magnetized flows, we should always remember that it is a solution if the Bernoulli equation alone, and the important transfield force-balance equation is

¹⁷ This shows how the arguments of Chiueh et al. [14] against the efficient magnetic acceleration in relativistic outflows can be circumvented (see Vlahakis [57], Komissarov et al. [30, 31] for more detail).

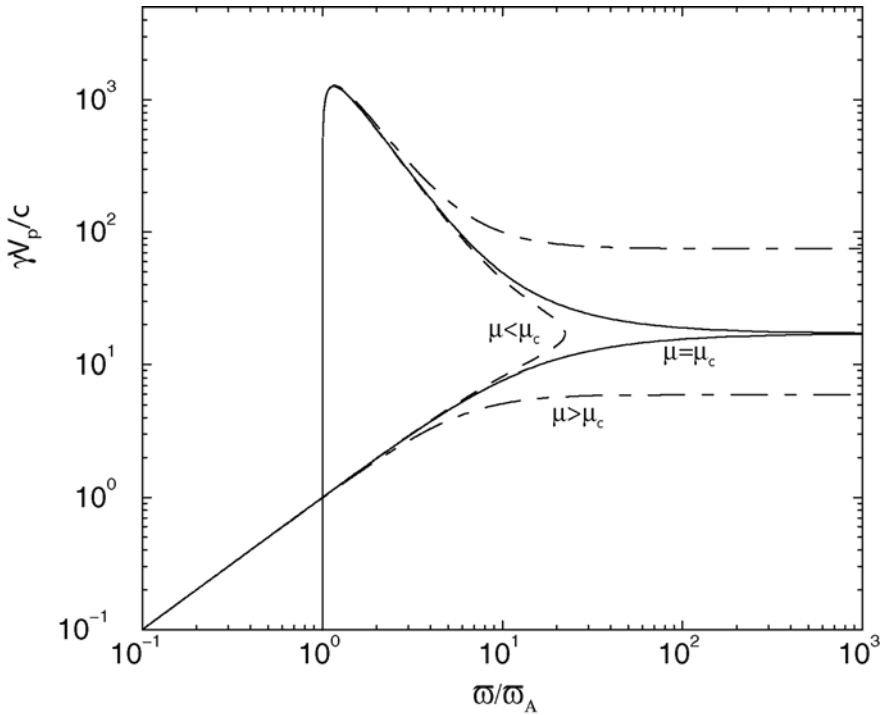


Fig. 10 Michel's solution for $\sigma_M = 5000$. The critical value for μ is $\mu_c = (1 + \sigma_M^{2/3})^{3/2} = 5075.67$ and the requirement to have $\gamma \rightarrow 1$ for $x \rightarrow 0$ gives $x_A = 0.9998$. In the figure three contours are shown: the critical $\mu = \mu_c$ and the $\mu = \mu_c - 5 < \mu_c$, $\mu = \mu_c + 50 > \mu_c$. The asymptotic Lorentz factor for the critical solution is $\sigma_M^{1/3} = 17.1$

simply ignored. The finding that the asymptotic Lorentz factor is just $\mu^{1/3}$, much smaller than the maximum value $\gamma_\infty = \mu$ that corresponds to matter-dominated flow, is a direct consequence of the assumed field line shape. According to the analysis of Sect. 2.2.3, since the function $B_p \omega^2/A$ is by assumption constant, the magnetic acceleration is expected to be completely inefficient. This does not mean that the magnetic acceleration is in general inefficient and the Michel's scaling $\gamma_\infty = \mu^{1/3}$ is not a general result.

Beskin et al. [5] examined deviations from the monopole magnetic field. They found that the fast magnetosonic point is moved at finite distance; however, the acceleration was still inefficient. This result agrees with the analysis of Sect. 2.2.3. The reason of the magnetic acceleration being inefficient is not the position of the fast point alone. Definitely the main part of the magnetic acceleration happens from that point downstream, so for solutions with the fast point at infinity there is no chance to have efficient acceleration. However, even in cases where the flow

becomes superfast at finite distance, the function $B_p \varpi^2 / A$ should not be forced to be constant. In the perturbations around the monopole solution, this function is by construction roughly constant, since the zeroth-order solution is the monopole one in which $B_p \varpi^2$ is exactly constant. In fact there is a very slow, logarithmic acceleration in the superfast regime, as Lyubarsky and Eichler [36] showed. This acceleration could give asymptotically $\gamma_\infty = \mu$, but over completely unrealistic (exponentially large) distances.

2.4 The Prescribed Field Line Shape

In principle, if we knew the field line shape, the Bernoulli equation (74) yields the flow velocity. The Bernoulli equation is in fact quite easy to solve, since it is algebraic. The only difficulty is to cross the slow and fast magnetosonic points (where the solution becomes double) in a similar way as in Michel's solution where the acceptable solution becomes double at infinity. In the superfast regime the simple expression (82) actually gives directly the flow speed.

For this reason, many people assumed a prescribed field line shape and then concentrated on solving the Bernoulli equation alone, by ignoring the transfield equation. Michel's solution belongs to this category. More sophisticated line shapes were assumed (e.g., Takahashi et al. [50]; Takahashi and Shibata [49]; Fendt and Greiner [19]; Daigne and Drenkhahn [17]; Fendt and Ouyed [20]), without, however, to check a posteriori the force-balance in the transfield direction and the induced error. Although these analyses are useful in providing information on how the flow reacts in various field line shapes, further work is necessary in order to find the line shape that is consistent with the solution of the Bernoulli equation.

2.5 Asymptotic Solutions

Another way to get useful insights on relativistic MHD is to analyze the asymptotic behavior of the flows. We already used in Sect. 2.2.3 a simplified asymptotic form of the Bernoulli equation. More extensive analysis include the work of Heyvaerts and Norman [27] for nonrelativistic flows, its relativistic generalization by Chiueh et al. [13], and more recently, the works by Tomimatsu and Takahashi [51], Heyvaerts and Norman [28], and Vlahakis [57]. We note, however, that asymptotic analysis of the highly nonlinear system of the MHD equations always involves a risk. The dropped terms – either terms of order $1/x$ or curvature radius terms – sometimes surprisingly modify the full numerical solutions; thus, we thought it is better not to analyze in detail the previously cited works, but instead concentrate to the only available exact (by means of solving the full system of equations, the transfield force-balance included) semianalytic work, valid from the base of the flow to infinity (contrary to asymptotic studies), which is the r self-similar model.

2.6 The r Self-similar Model

2.6.1 Model Construction

To obtain semianalytic solutions of the highly nonlinear system of Eqs. (74) and (75), it is necessary to make additional assumptions: in particular, we look for a way to effect a separation of variables.

The most complicated expression is the one for B_ϕ (Eq. (68)). In view of the importance of the azimuthal field component, which plays a crucial and varied role as part of the magnetic pressure gradient, magnetic tension, and centrifugal acceleration terms in the momentum equation, the only realistic possibility of deriving exact semianalytic solutions is to assume that the $M = \text{const}$, $G = \text{const}$, and $x = \text{const}$ surfaces coincide, i.e., $M = M(\chi)$, $G = G(\chi)$, $x = x(\chi)$ (Vlahakis [56]). We aim to find appropriate forms for the functions of A such that expressions (74) and (75) become ordinary differential equations. From an inspection of the Bernoulli equation (74) we conclude that, in order to separate the variables χ and A and get a *single* equation that only has a χ dependence, it is necessary to assume that the $(\nabla A)^2$ term is a product of a function of A times a function of χ . As A is a function of $\varpi/G(\chi)$ (see Eq. (66)), there must exist functions \mathcal{H}_1 , \mathcal{H}_2 such that

$$\left[\nabla \left(\frac{\varpi}{G} \right) \right]^2 = \mathcal{H}_1 \left(\frac{\varpi}{G} \right) \mathcal{H}_2(G).$$

There always exist the trivial possibilities $G \propto r$ in spherical coordinates (r, θ, ϕ) [$A = A(\theta)$ when the field is radial], and $G = G(\varpi)$ ($A = A(\varpi)$ for a cylindrical field), which are not of interest. After some algebra one can prove that the only nontrivial case is to have $G = G(\theta)$, i.e., $\chi = \theta$. It thus appears that, to obtain an semianalytic adiabatic solution, it is necessary to assume r self-similarity.¹⁸

The remaining assumptions for constructing an r self-similar solution are that the cylindrical distance (in units of the Alfvénic lever arm), the poloidal Alfvénic Mach number, and the relativistic specific enthalpy are also functions of θ only: $x = x(\theta)$, $M = M(\theta)$, $\xi = \xi(\theta)$ (with the result for ξ following from the nonlinearity of the expression for M ; see Eq. (73)).

Following the algorithm described in Vlahakis and Tsinganos [61], we change variables from (r, θ) to (ϖ_A, θ) and obtain the forms of the integrals under the assumption of separability in ϖ_A and θ in Eqs. (74) and (75). The results are¹⁹

¹⁸ For nonadiabatic flows there exist, in principle, other solutions as well, exactly as in the nonrelativistic case, (e.g., Meliani et al. [39]). Here, however, we concentrate on adiabatic flows where the total energy-to-mass flux ratio is conserved.

¹⁹ The nonrelativistic limit of our model is *not* the generalization of the Blandford and Payne [6] model, examined, e.g., in Vlahakis et al. [62]. The nonrelativistic limit can, however, be obtained from the analysis of Vlahakis and Tsinganos [61]: it corresponds to the third line of their Table 3 (setting $x_1 = F - 2$, $x_2 = F - 5/2$, $E_2 = C_1 = D_2 = 0$, and ignoring gravity, so it is possible to assume a polytropic equation of state).

$$\sigma_M, \mu, x_A, q = \text{const}, \quad A = \frac{B_0 \varpi_0^2}{F} \left(\frac{\varpi_A}{\varpi_0} \right)^F, \quad (88)$$

where B_0 and ϖ_0 are some reference magnetic field and distance, chosen for adimensionalization, while F is the only parameter of the model (besides Γ that is always given [4/3 or 5/3 depending on the temperatures]).

The r self-similar character of the poloidal field line shape is shown in Fig. 11.

The physical quantities can be recovered using

$$\frac{\mathbf{B}}{B_0(\varpi_A/\varpi_0)^{F-2}} = \frac{\sin \theta}{G^2 \sin(\theta - \vartheta)} \hat{\mathbf{b}} - \frac{\mu x_A^4 (1 - G^2)}{F \sigma_M x (1 - M^2 - x^2)} \hat{\boldsymbol{\phi}}, \quad (89)$$

$$\frac{\mathbf{E}}{B_0(\varpi_A/\varpi_0)^{F-2}} = \frac{x_A \sin \theta}{G \sin(\theta - \vartheta)} \hat{\mathbf{n}}, \quad (90)$$

$$\frac{V}{c} = \frac{F \sigma_M M^2 \sin \theta}{\gamma \xi x^2 \sin(\theta - \vartheta)} \hat{\mathbf{b}} + \frac{x_A \mu (G^2 - M^2 - x^2)}{\gamma \xi G (1 - M^2 - x^2)} \hat{\boldsymbol{\phi}}, \quad (91)$$

$$\gamma = \frac{\mu}{\xi} \frac{1 - M^2 - x_A^2}{1 - M^2 - x^2}, \quad \rho_0 = \frac{B_0^2 x_A^4 \xi}{4\pi c^2 F^2 \sigma_M^2 M^2} \left(\frac{\varpi_A}{\varpi_0} \right)^{2F-4}, \quad (92)$$

$$P = \frac{B_0^2}{4\pi} \frac{\Gamma - 1}{\Gamma} \frac{x_A^4}{F^2 \sigma_M^2} \frac{\xi (\xi - 1)}{M^2} \left(\frac{\varpi_A}{\varpi_0} \right)^{2F-4}, \quad (93)$$

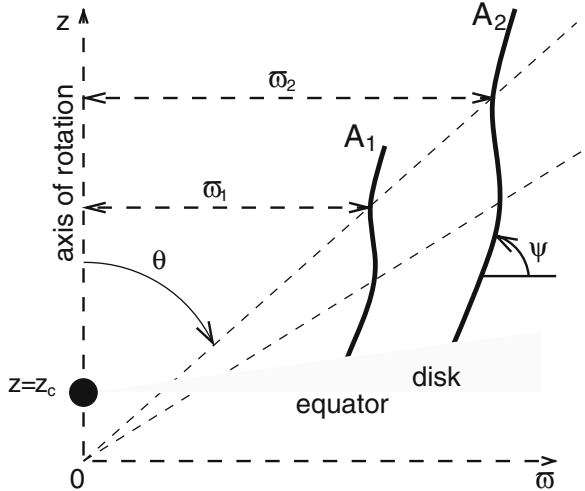


Fig. 11 Sketch of r self-similar field lines in the meridional plane. For any two field lines A_1 and A_2 , the ratio of cylindrical distances for points corresponding to a given value of θ is the same for all the cones $\theta = \text{const}$: $\varpi_1/\varpi_2 = \varpi_{A1}/\varpi_{A2} = (A_1/A_2)^{1/F}$

where the functions $x(\theta)$, $G(\theta)$, $\vartheta(\theta)$, $M(\theta)$, $\xi(\theta)$ are the solution of Eqs. (73), (74), and (75). In particular, Eq. (73) remain the same. Equation (74) gives

$$\frac{\mu^2}{\xi^2} \frac{G^4 (1 - M^2 - x_A^2)^2 - x^2 (G^2 - M^2 - x^2)^2}{G^4 (1 - M^2 - x^2)^2} = 1 + \frac{F^2 \sigma_M^2 M^4 \sin^2 \theta}{\xi^2 x^4 \cos^2 (\theta - \vartheta)}. \quad (94)$$

Using for the slope of the poloidal field line $\tan \vartheta = (\partial \varpi / \partial z)_A$ with $\varpi = \varpi_A G$ and $z = \varpi / \tan \theta$ we find the equation that connects the angle ϑ with the derivative of G

$$\frac{dG^2}{d\theta} = -\frac{2G^2 \sin \vartheta}{\sin \theta \sin (\theta - \vartheta)}. \quad (95)$$

The last equation that closes the system is the transfield force-balance equation (75) that becomes

$$\begin{aligned} \frac{d}{d\theta} \left[\frac{1 - M^2 - x^2}{G \tan (\theta - \vartheta)} \right] &= \frac{(F - 1) x_A^4 \mu^2 x^2}{F^2 \sigma_M^2 G \sin^2 \theta} \left[\frac{1 - G^2}{1 - M^2 - x^2} \right]^2 \\ &- \frac{x_A^4 \mu^2 x^2}{F^2 \sigma_M^2 M^2 G \sin^2 \theta} \left[\frac{G^2 - M^2 - x^2}{1 - M^2 - x^2} \right]^2 - \frac{M^2 + F x^2 - F + 1}{G \sin^2 (\theta - \vartheta)} \\ &+ \frac{2(\Gamma - 1)(F - 2)\xi(\xi - 1)x^4}{\Gamma F^2 \sigma_M^2 G M^2 \sin^2 \theta}. \end{aligned} \quad (96)$$

The r dependence of all the physical quantities can be inferred from expressions (89), (90), (91), (92), and (93) on the basis of the known r dependence of ϖ_A ($\propto r$; since $\varpi_A = \varpi / G(\theta) = r \sin \theta / G(\theta)$). This is a general characteristic of r self-similar models. We may say that r self-similar solutions correspond to boundary conditions in a conical surface ($\theta = \theta_i$) of the form $B_r = C_1 r^{F-2}$, $B_\phi = -C_2 r^{F-2}$, $V_r = C_3$, $V_\theta = -C_4$, $V_\phi = C_5$, $\rho_0 = C_6 r^{2F-4}$, $P = C_7 r^{2F-4}$, with constant C_1, \dots, C_7 . If we start the integration from a cone $\theta = \theta_i$, the separability of the equations means that in a subsequent cone $\theta = \theta_i + d\theta$ the quantities have the same form, with different, however, C_1, \dots, C_7 .

The parameter of the model F controls the initial current distribution: $-\varpi B_\phi = C_2 \sin \theta_i r^{F-1}$ is an increasing or decreasing function of r for $F > 1$ or < 1 ; see Vlahakis and Königl [58] for details; we note, however, that the $F > 1$ solutions correspond to zero current near the rotation axis, while the $F < 1$ to an infinite current at $r \rightarrow 0$, like the Blandford and Payne [6] nonrelativistic solution. Despite the assumed form of the boundary conditions, the assumption that gravity is negligible, and the absence of intrinsic scale, r self-similar remain the only self-consistent adiabatic relativistic MHD solutions.

The model described above has been analyzed in Vlahakis and Königl [58, 59] in the context of GRB outflows and in Vlahakis and Königl [60] in the context of AGN

outflows. It is the generalization to a ‘hot’ ($\xi > 1$) gas of the ‘cold’ r self-similar wind solution found independently by Li et al. [34] and Contopoulos [16].²⁰

2.6.2 A Sample Relativistic Solution

We now present a specific example of an exact solution of the MHD equations in order to compare with the analysis of Sect. 2.2. It is easier to focus on the magnetic effects when the flow is practically cold; thus, we choose to analyze the solution b in Vlahakis and Königl [58].²¹ Figure 12 shows the poloidal field-streamline shape in logarithmic scale. The plotted curve holds for every poloidal field line (i.e., for every value of A); however, as we move to lines away from the rotation axis (i.e., at larger A) the Alfvénic lever arm ($\varpi_A \propto A^{1/F}$) increases, so both ϖ and z scale according to $\varpi = \varpi_A G(\theta)$ and $z = \varpi_A G(\theta) / \tan \theta$.

In the main part of the flow the line shape is parabolic $z \propto \varpi^2$, while asymptotically it becomes cylindrical (the slope $d \ln z / d \ln \varpi$ progressively increases near $\varpi / \varpi_A \sim 10^4$, as shown in Fig. 12). The half-opening angle of the outflow is shown in Fig. 13. The angle $\vartheta = \arctan(V_\varpi / V_z)$ that the poloidal flow makes with the

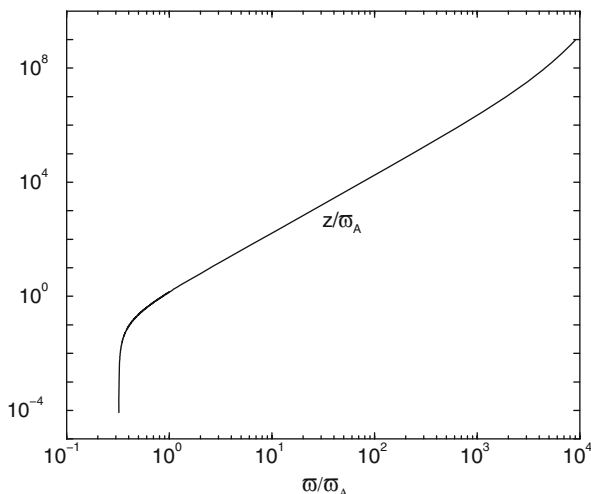


Fig. 12 Poloidal field-streamline shape in logarithmic scale

²⁰ The r self-similarity was first employed by Bardeen and Berger [3], who examined purely HD flows, but it has become well known only after Blandford and Payne [6] used it to construct a nonrelativistic MHD disk-wind model. The latter work has subsequently been generalized by many authors (see Vlahakis et al. [62] and references therein).

²¹ For this solution the parameters/boundary conditions are $F = 1.01$, $x_A^2 = 0.9999$, $\sigma_M = 5000$, $q = 0$, while the Alfvén surface is located at polar angle $\theta_A = 35^\circ$. From the requirement that the flow crosses smoothly the Alfvén surface and also it continues to large distances without being decelerated (this is the magnetic nozzle effect) we find $\mu = 9997.4$, and the half-opening angle of the poloidal lines at the Alfvén surface $\vartheta_A = 17.3^\circ$.

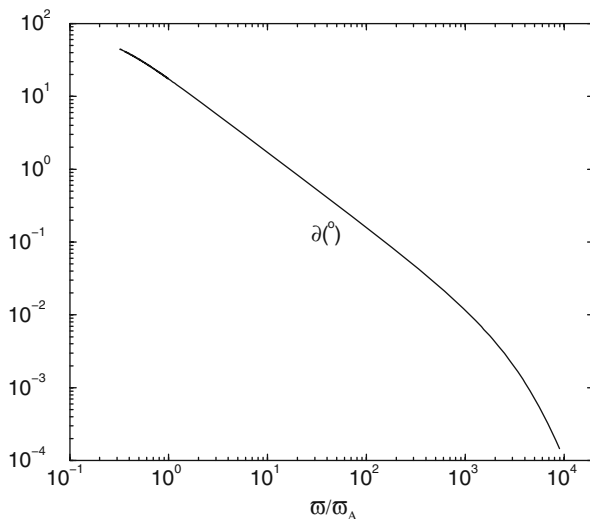


Fig. 13 Half-opening angle of the poloidal flow, $\vartheta = \arctan(V_{\varpi}/V_z)$

rotation axis, initially has the value 44.6° , and decreases, for the main part of the flow, as $\vartheta \propto \varpi^{-1}$, a characteristic of the parabolic line shape. Near $\varpi/\varpi_A \sim 10^4$ it decreases faster than $\propto \varpi^{-1}$, approaching the asymptotic value ≈ 0 when the flow becomes cylindrical.

Figure 14 shows $x = \varpi\Omega/c$ and $M = (\gamma V_p/B_p)\sqrt{4\pi\rho_0\xi}$ (the light surface radius and the Alfvénic Mach number, respectively) as well as the ‘fast magnetosonic proper Mach number’ $M_f \equiv \gamma V_p/U_f$, where U_f is the larger solution of the quadratic

$$\left(\frac{U_f}{c}\right)^4 - \left(\frac{U_f}{c}\right)^2 \left(\frac{U_s^2}{c^2} + \frac{B^2 - E^2}{4\pi\rho_0\xi c^2}\right) + \frac{U_s^2}{c^2} \frac{B_p^2(1-x^2)}{4\pi\rho_0\xi c^2} = 0. \quad (97)$$

The point where $x = 1$ corresponds to the light surface, which practically coincides with the Alfvén surface $\varpi/\varpi_A = 1$. At the Alfvén surface $x = x_A$ and $M = (1 - x_A^2)^{1/2}$. The point $M_f = 1$ (at $\varpi/\varpi_A \approx 22.6$) is the fast magnetosonic point.²² It is seen that for the main part of the flow $M \approx M_f$. Since for cold flows the solution of Eq. (97) is $U_f^2 = (B^2 - E^2)/(4\pi\rho_0)$ and $M^2 = 4\pi\rho_0(\gamma V_p/B_p)^2$, this means that $B^2 - E^2 \approx B_p^2$. Equivalently, since $B^2 - E^2 \approx B_p^2/\gamma^2 \approx B_p^2(x^2/\gamma^2)$, the Lorentz factor increases almost linearly with the cylindrical distance $\gamma \sim x$ (this is indeed the numerical result as we discuss below).

²² The function M_f shows where the fast magnetosonic surface is located. Besides that, M_f is related to the Mach cone of the propagation of fast magnetosonic waves in the superfast regime. For practically cold flow, this cone has half-opening angle $\arctan(M_f^2 - 1)^{-1/2}$ (the cone’s symmetry axis coincides with the poloidal field line).

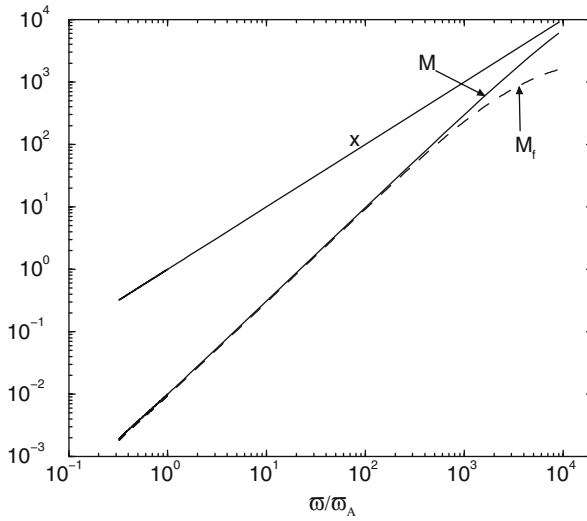


Fig. 14 Cylindrical distance in units of the ‘light surface’ distance x and the Mach numbers M and M_f

Figure 15 shows the two components of the magnetic field. It is seen that close to the base the poloidal part dominates, near the Alfvén surface the two parts become comparable, while at large distances the azimuthal part dominates. The two components scale as $B_p \propto 1/\varpi^2$ and $-B_\phi \propto 1/\varpi$. However, the small deviation of the product ϖB_ϕ from a constant is connected to the acceleration, since the Poynting-to-mass flux ratio (over c^2 for adimensionalization) is $-\varpi \Omega B_\phi / \Psi_A c^2$; see Fig. 16. Also the deviation of $B_p \varpi^2$ is shown in Fig. 17 and plays an important role for the

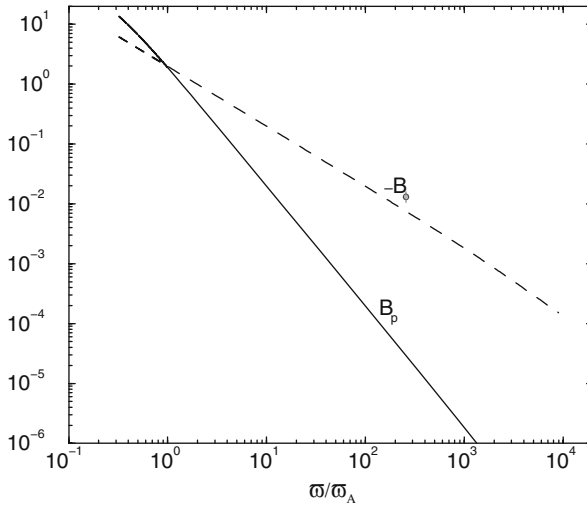


Fig. 15 Poloidal and azimuthal components of the magnetic field

acceleration (since $-B_\phi \approx x B_p$, the quantities ϖB_ϕ and $B_p \varpi^2$ are proportional to each-other).

Figure 16 shows the two parts of the total energy-to-mass flux ratio, see Eq. (62). The matter part γ is almost linearly increasing with the cylindrical distance ($\gamma \propto \varpi$) in expense on the electromagnetic part $-\varpi \Omega B_\phi / \Psi_A c^2$. The small deviation of the ϖB_ϕ results in significant acceleration. A variation from $-\varpi \Omega B_\phi / \Psi_A c^2 = \mu$ near the base to $\sim \mu/2$ asymptotically corresponds to acceleration up to $\gamma \sim \mu/2$, leading to equipartition between matter and electromagnetic energy fluxes at large distances. As we discussed in Sect. 2.2.3, the acceleration efficiency depends on the value of $B_p \varpi^2 / A$ at the fast magnetosonic surface. Since this value is 2 (see Fig. 17) we expect $(\mu - \gamma_\infty) / \mu \approx 1/2$ (see Eq. (83)), and this is indeed the case. Figure 17 also verifies the results of Sect. 2.2.3. In particular it shows that the analytic expression (82) is almost exact in the superfast regime.

Figure 18 shows that the electric field is slightly less than the absolute value of the azimuthal magnetic field. To an almost perfect accuracy, their difference is given by the approximate expression $-B_\phi / E \approx \sqrt{(1 - 1/x^2)/(1 - 1/\gamma^2)}$, valid in the superfast regime (see footnote 13).

The poloidal and azimuthal parts of the flow velocity are shown in Fig. 19. The azimuthal part has an initial value of the order of $\varpi_i \Omega$ and keeps increasing inside the roughly corotating sub-Alfvénic regime, while at large distances decreases as $\propto 1/\varpi \gamma$ from angular momentum conservation.

Figure 20 shows the various force components along the poloidal flow. It is seen that the magneto-centrifugal term $f_{C\parallel}$ dominates near the base of the flow, but the magnetic force $f_{B\parallel}$ soon takes over (for non- or mildly relativistic flows the regime

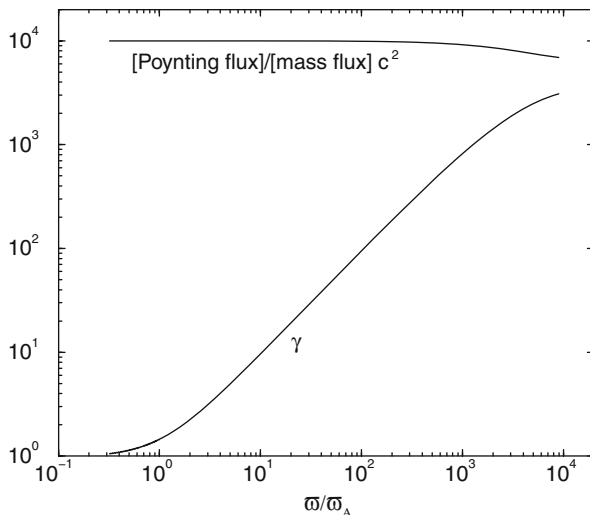


Fig. 16 Lorentz factor and the quantity $-\varpi \Omega B_\phi / \Psi_A c^2 = \mu - \xi \gamma$, representing the Poynting-to-mass flux ratio (over c^2). These are the two parts of the total energy-to-mass flux ratio (over c^2); their sum equals the field line constant μ

where the magneto-centrifugal mechanism dominates is expected to be larger). In response to the accelerating forces, the inertia terms $-f_{I\parallel} = \gamma^2 \rho_0 \xi V_p \partial V_p / \partial \ell$ and $-f_{G\parallel} = \gamma \rho_0 \xi V_p^2 \partial \gamma / \partial \ell$ are positive, resulting in an accelerating flow. As long as the flow remains mildly relativistic the term $-f_{I\parallel}$ dominates, while at large distances the $-f_{G\parallel}$ is the main inertial force in the V_p direction.

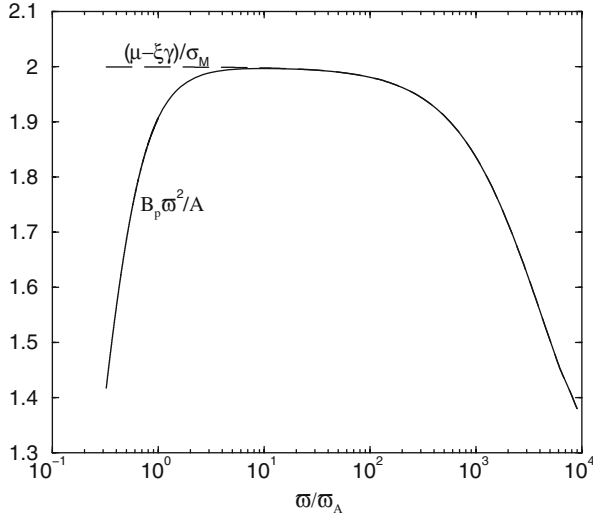


Fig. 17 Function $B_p \varpi^2 / A$ and the electromagnetic part of the energy-to-mass flux ratio, over $\sigma_M c^2$. The two quantities become practically equal in the superfast regime

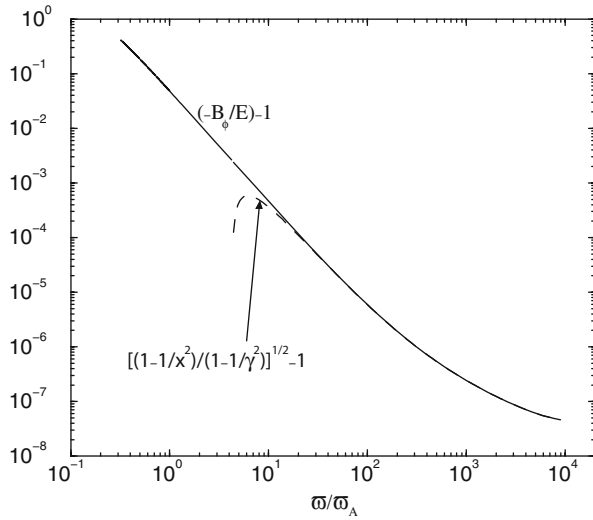


Fig. 18 The $-B_\phi / E$ ratio and its approximate expression

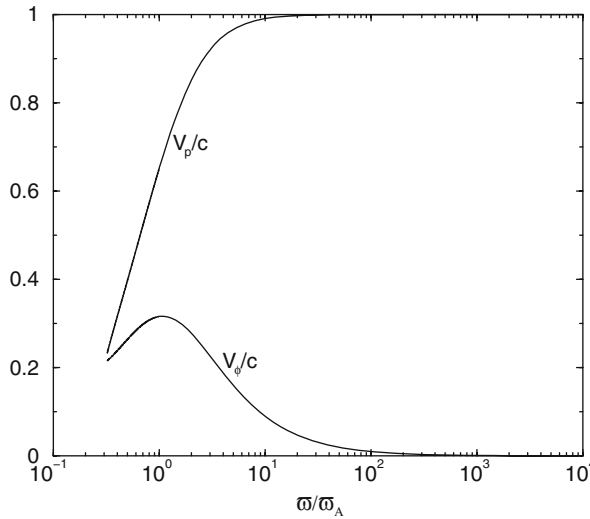


Fig. 19 Poloidal and azimuthal velocity components

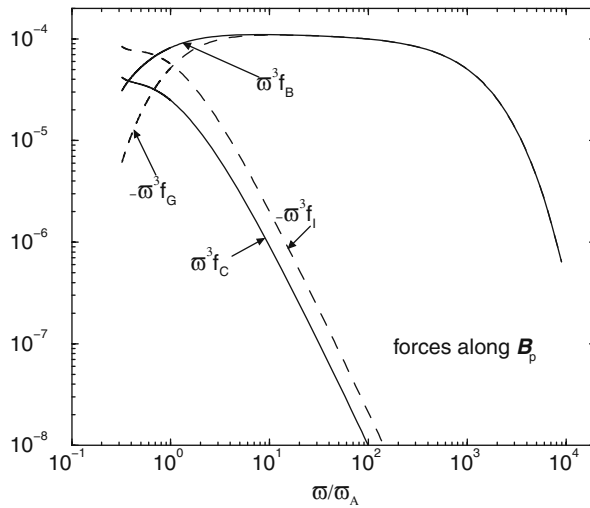


Fig. 20 Force density components along the poloidal flow (multiplied with ω^3). The numbers in the vertical axis are in units of $(B_0^2 \omega_0^2 / 4\pi)(\omega_A / \omega_0)^{2F-2}$

The force components in the transfield direction on the poloidal plane (i.e., along the direction of the electric field) are shown in Fig. 21. The force balance in this direction determines the poloidal field line shape and indirectly the acceleration, as we discussed in Sect. 2.2.4. In the sub-Afvénic regime the force $-f_{EM3}$ dominates, while in the super-Afvénic part of the flow the force f_{EM1} takes over. However, the $-f_{EM3}$ part remains comparable to the f_{EM1} ; thus, it is not a correct approxima-

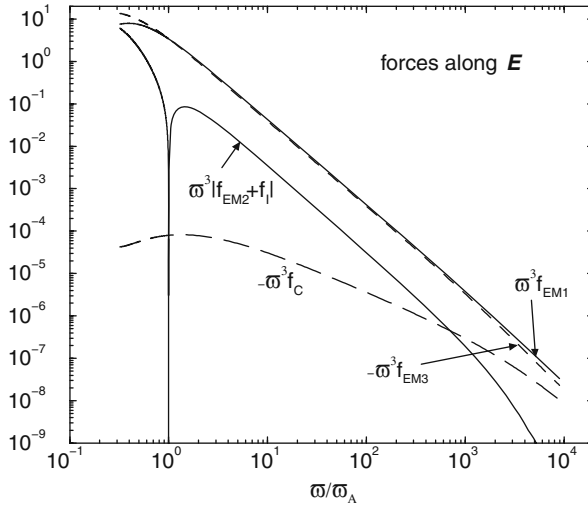


Fig. 21 Force density components along the electric field direction (multiplied with ω^3). The numbers in the vertical axis are in units of $(B_0^2 \omega_0^2 / 4\pi)(\omega_A / \omega_0)^{2F-2}$

tion to ignore the poloidal field terms even at large distances. For $\omega / \omega_A < 5 \times 10^2$ the centrifugal force is negligible. As a result, the difference between the f_{EM1} and $-f_{EM3}$ terms determines the poloidal curvature (this difference equals the $|f_{EM2} + f_{I\perp}|$ term. For $\omega / \omega_A > 5 \times 10^2$, however, the centrifugal term $-f_{C\perp}$ is nonnegligible and equals the difference between the f_{EM1} and $-f_{EM3}$ terms (the curvature term $|f_{EM2} + f_{I\perp}|$ is much smaller, a characteristic of cylindrical asymptotics).

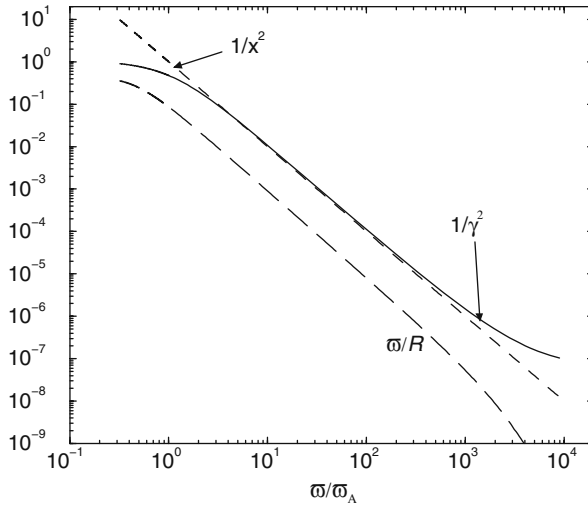


Fig. 22 The quantity ω / \mathcal{R} , where \mathcal{R} is the curvature of the poloidal field lines, and the $1/x^2$, $1/\gamma^2$

Figure 22 shows the curvature of the poloidal field lines. As expected from the analysis of Sect. 2.2.4, $\varpi/\mathcal{R} < 1/\gamma^2$. However, the approximation $\varpi/\mathcal{R} \sim 1/\gamma^2$ does not hold, because the force f_{EM3} related to the poloidal field is non-negligible as we discussed in the previous paragraph.²³

The solution that we analyzed above was presented by Vlahakis and Königl [58] in the context of GRB outflows. In that paper the reader could find ‘hot’ solutions as well; however, the main characteristics of the outflows remain the same. One could also find solutions of the same model in the context of AGN outflows in Vlahakis and Königl [60] and in the cold subcase in Li et al. [34] Contopoulos [16].

2.7 Simulations

Komissarov et al. [30] have taken a major step toward solving the numerical problems associated with relativistic magnetized flows, simulating special relativistic, axisymmetric, ideal MHD flows with asymptotic Lorentz factors up to ~ 15 . In a more recent work Komissarov et al. [31] they simulate flows that reach asymptotic Lorentz factors up to 300. These solutions are in a very good agreement with the presented analytical results, as can be seen in the corresponding discussion sections (Sect. 5 in Komissarov et al. [30] and in Komissarov et al. [31]).

3 Conclusion

Summarizing, it is important to note that in order to solve for the acceleration it is absolutely necessary to solve for the poloidal field line shape as well. The Bernoulli and transfield force-balance equations are interrelated and we cannot solve them separately, especially in the superfast regime.

We made qualitative arguments about the various acceleration and collimation mechanisms. We could briefly review the basic results that are confirmed by all known exact solutions of the MHD equations. The acceleration mechanisms are as follows:

- Pressure-gradient force giving velocities of the order of the sound speed.
- Magneto-centrifugal acceleration which transforms part of the Poynting flux to matter energy flux and gives velocities of the order of the azimuthal speeds near the base of the flow.
- Temperature (or enthalpy) gradient, resulting in matter kinetic energies of the order of the initial enthalpy (this mechanism works only in cases where the temperatures are relativistic).
- Magnetic acceleration. In nonrelativistic flows this mechanism gives flow speeds at the fast magnetosonic surface that are already $\sqrt{1/3}$ of the maximum possible

²³ In the case of negligible poloidal magnetic field (this is, for example, the case of an initial super-Alfvénic flow) it is $\varpi/\mathcal{R} \approx 1/\gamma^2$ (see Fig. 2b in Vlahakis and Königl [59]).

velocity (the maximum velocity corresponds to the full transfer of the electromagnetic energy to the matter, meaning that $V^2/2$ equals the total energy-to-mass flux ratio). The acceleration may continue after the fast surface reaching efficiencies even close to 100%. In relativistic flows the main part of the magnetic acceleration is realized from the fast magnetosonic surface downstream. The asymptotic Lorentz factor depends on the initial total energy-to-mass flux ratio μc^2 and the value of the Michel's magnetization parameter σ_M . According to the discussion in Sect. 2.2.3, an analytic expression for the asymptotic Lorentz factor is $\gamma_\infty \sim \mu - \sigma_M$. (Possible acceleration to $\gamma_\infty \rightarrow \mu$ can only happen in exponentially large – and thus physically irrelevant – distances.) Another expression connecting the asymptotic value γ_∞ with the conditions near the origin of an initially Poynting-dominated flow is $\gamma_\infty \sim \mu(1 - A\Omega/2|I|_i)$. The asymptotic Poynting-to-mass flux ratio is $\sim \sigma_M c^2 \approx \mu c^2 A\Omega/2|I|_i$.

Regarding the collimation, in nonrelativistic flows the $\mathbf{J}_p \times \mathbf{B}_\phi/c$ force can easily collimate the flow to more or less cylindrical geometries. It is relatively inefficient in relativistic flows in comparison to nonrelativistic ones; however, it is still possible. Practically, the main part of the confinement happens closer to the base. At large distances the poloidal curvature radius satisfies $\varpi/\mathcal{R} \lesssim 1/\gamma^2$, a relation that is consistent with parabolic line shapes $z \propto \varpi^\alpha$. In these cases the Lorentz factor scales as $\gamma \propto \varpi^{\alpha-1}$.

Semianalytic results as well as simulations of the problem agree very well with the simple analytical scalings described in this chapter (relativistic and nonrelativistic). We can say that our knowledge on the theory of MHD is at a level sufficient to understand global properties of magnetized flows.

Appendix: Alfvén and Magnetosonic MHD Waves

Suppose that we have obtained a solution of the axisymmetric, relativistic, ideal-MHD equations (51), (52), (53), (54) and (55). If we consider localized, fast varying, axisymmetric disturbances, then we may assume that the unperturbed solution is uniform and time independent and neglect all its space and time derivatives. We may then look for perturbations having a Fourier dependence $\exp[i(\omega t - \mathbf{k} \cdot \mathbf{r})] = \exp[i(\omega_{co} t_{co} - \mathbf{k}_{co} \cdot \mathbf{r}_{co})]$, where by using the Lorentz transformations between the comoving $(\mathbf{r}_{co}, t_{co})$ and observer's frame (\mathbf{r}, t) , we get $\omega_{co} = \gamma(\omega - \mathbf{k} \cdot \mathbf{V})$, $\mathbf{k}_{co} = \mathbf{k} - \mathbf{V}[\gamma\omega/c^2 - (\gamma - 1)\mathbf{k} \cdot \mathbf{V}/V^2]$.²⁴

It is more convenient to analyze the disturbances in the (comoving) flow frame. Define a local Cartesian system of coordinates (x, y, z) such that $\mathbf{B}_{co} = B_{co}\hat{z}$, and $\mathbf{k}_{co} = k_{co}(\hat{z} \cos \theta_0 + \hat{x} \sin \theta_0)$. After linearizing Eqs. (51), (52), (53), (54) and (55), we may express all the perturbed quantities in terms of the perturbation $\delta \mathbf{V}_{co}$. After some manipulation, we obtain

²⁴ Note that, for axisymmetric disturbances in the observer's frame, the wavevector \mathbf{k} lies in the meridional plane, but that, nonetheless, the comoving wavevector \mathbf{k}_{co} has an azimuthal component.

$$\begin{aligned}
& \begin{pmatrix} D_{11} & 0 & D_{13} \\ 0 & D_{22} & 0 \\ D_{13} & 0 & D_{33} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{x}} \cdot \delta \mathbf{V}_{\text{co}} \\ \hat{\mathbf{y}} \cdot \delta \mathbf{V}_{\text{co}} \\ \hat{\mathbf{z}} \cdot \delta \mathbf{V}_{\text{co}} \end{pmatrix} = 0, \quad \text{where} \\
D_{11} &= \frac{c_s^2}{c^2} \sin^2 \theta_0 + \left(\frac{v_A^2}{c^2} - \frac{\omega_{\text{co}}^2}{c^2 k_{\text{co}}^2} \right) \left(1 - \frac{v_A^2}{c^2} \right)^{-1}, \\
D_{13} &= \frac{c_s^2}{c^2} \sin \theta_0 \cos \theta_0, \\
D_{22} &= \left(\frac{v_A^2}{c^2} \cos^2 \theta_0 - \frac{\omega_{\text{co}}^2}{c^2 k_{\text{co}}^2} \right) \left(1 - \frac{v_A^2}{c^2} \right)^{-1}, \\
D_{33} &= \frac{c_s^2}{c^2} \cos^2 \theta_0 - \frac{\omega_{\text{co}}^2}{c^2 k_{\text{co}}^2}.
\end{aligned}$$

The Alfvén speed can be expressed in terms of the corresponding proper speed U_A , which satisfies

$$U_A^2 \equiv \frac{v_A^2}{1 - v_A^2/c^2} = \frac{B_{\text{co}}^2}{4\pi\rho_0\xi}. \quad (98)$$

The square of the proper sound speed is given by

$$U_s^2 = c^2 \frac{(\Gamma - 1)(\xi - 1)}{(2 - \Gamma)\xi + \Gamma - 1} = \frac{c_s^2}{1 - c_s^2/c^2}, \quad c_s^2 = \Gamma \frac{P}{\rho_0\xi}. \quad (99)$$

Besides the trivial entropy wave $\omega_{\text{co}} = 0$ (which, however, corresponds to $\omega \neq 0$), the dispersion relation $|D| = 0$ yields the wave modes listed below.

- Alfvén waves: $D_{22} = 0$, or $\omega_{\text{co}}/k_{\text{co}} = \pm v_A \cos \theta_0$, corresponding to a displacement $\delta \mathbf{V}_{\text{co}}$ normal to the $\{\mathbf{B}_{\text{co}}, \mathbf{k}_{\text{co}}\}$ plane. Transforming the dispersion relation to the observer's frame, we get

$$\begin{aligned}
& \left(\gamma \frac{\omega - \mathbf{k} \cdot \mathbf{V}}{ck} \right)^2 = \\
& \frac{(\mathbf{B} \cdot \mathbf{k}/k)^2}{4\pi\rho_0\xi c^2} \left[1 - \left(x + \frac{\omega B_\phi}{c\mathbf{B} \cdot \mathbf{k}} \right)^2 - \left(\frac{\omega B_p}{c\mathbf{B} \cdot \mathbf{k}} \right)^2 \right]. \quad (100)
\end{aligned}$$

- Slow/fast magnetosonic waves: $D_{11}D_{33} = D_{13}^2$, or,

$$\left(\frac{\omega_{\text{co}}}{ck_{\text{co}}} \right)^4 - \left(\frac{\omega_{\text{co}}}{ck_{\text{co}}} \right)^2 \left(\frac{c_s^2}{c^2} + \frac{v_A^2}{c^2} - \frac{c_s^2 v_A^2}{c^4} \sin^2 \theta_0 \right) + \frac{c_s^2 v_A^2}{c^4} \cos^2 \theta_0 = 0,$$

corresponding to a displacement $\delta \mathbf{V}_{\text{co}}$ in the $\{\mathbf{B}_{\text{co}}, \mathbf{k}_{\text{co}}\}$ plane. In the observer's frame, we have

$$\begin{aligned}
& \left(1 - \frac{\omega^2}{c^2 k^2}\right)^{-1} \left(\gamma \frac{\omega - \mathbf{k} \cdot \mathbf{V}}{ck}\right)^4 - \\
& \left(\gamma \frac{\omega - \mathbf{k} \cdot \mathbf{V}}{ck}\right)^2 \left(\frac{U_s^2}{c^2} + \frac{B^2 - E^2}{4\pi\rho_0\xi c^2}\right) + \\
& \frac{U_s^2}{c^2} \frac{(\mathbf{B} \cdot \mathbf{k}/k)^2}{4\pi\rho_0\xi c^2} \left[1 - \left(x + \frac{\omega B_\phi}{c\mathbf{B} \cdot \mathbf{k}}\right)^2 - \left(\frac{\omega B_p}{c\mathbf{B} \cdot \mathbf{k}}\right)^2\right] = 0. \quad (101)
\end{aligned}$$

An interesting property of the waves, related to the discussion on critical/singular surfaces of steady-state MHD, is the following: If the component of the flow proper velocity along the wavevector equals in magnitude, but is opposite in sign, to the comoving proper phase velocity of the wave, then $\omega = 0$ and the wave is static. (The converse is also true.) Thus,

$$\omega = 0 \quad \Leftrightarrow \quad \gamma \mathbf{V} \cdot \frac{\mathbf{k}}{k} = -\frac{\omega_{co}/k_{co}}{(1 - \omega_{co}^2/c^2 k_{co}^2)^{1/2}}. \quad (102)$$

This statement is easily proved by combining $\omega_{co} = \gamma(\omega - \mathbf{k} \cdot \mathbf{V})$ with the Lorentz invariant $\omega_{co}^2 - c^2 k_{co}^2 = \omega^2 - c^2 k^2$ and solving for ω .

Equation (102) is the generalization of the property of nonrelativistic static waves $\mathbf{V} \cdot \mathbf{k} + \omega_{co} = \omega = 0$. It is consistent with the result that proper speeds are the appropriate generalization of ordinary speeds in relativistic theory (e.g., Königl [32]).

References

1. Anderson, J. M., et al.: ApJ **590**, L107 (2003) 72
2. Anderson, J. M., et al.: ApJ **653**, L33 (2006) 73
3. Bardeen, J., Berger, B.: ApJ **221**, 105 (1978) 97
4. Bekenstein, J. D., Oron, E.: Phys. Rev. D **18**, 1809 (1978) 79
5. Beskin, V. S., et al.: MNRAS **299**, 341 (1998) 92
6. Blandford, R. D., Payne, D. G.: MNRAS **199**, 883 (1982) 73, 74, 85, 94, 96, 97
7. Bogovalov, S. V.: A&A 371 (2001), 1155 90
8. Bogovalov, S., Tsinganos, K.: MNRAS **305**, 211 (1999) 73
9. Bogovalov, S.: A&A **323**, 634 (1997) 69
10. Camenzind, M.: A&A **156**, 137 (1986a) 79
11. Camenzind, M.: A&A **162**, 32 (1986b) 87
12. Casse, F., Keppens, R.: ApJ **601**, 90 (2004) 73
13. Chiueh, T., et al.: ApJ **377**, 462 (1991) 88, 93
14. Chiueh, T., et al.: ApJ **505**, 835 (1998) 91
15. Contopoulos, J., Lovelace, R.: ApJ **429**, 139 (1994) 73, 86
16. Contopoulos, J.: ApJ **432**, 508 (1994) 97, 104
17. Daigne, F., G. Drenkhahn: A&A **381**, 1066 (2002) 93
18. Fendt, C.: ApJ **651**, 272 (2006) 73
19. Fendt, C., Greiner, J.: A&A **369**, 308 (2001) 93
20. Fendt, C., Ouyed, R.: ApJ **608**, 378 (2004) 93
21. Ferreira, J.: A&A **319**, 340 (1997) 73
22. Ferreira, J., Casse, F.: ApJ **601**, L139 (2004) 63, 73

23. Ferreira, J.: In LNP Vol. 723, J. Ferreira, C. Dougados, E. Whelan (Eds.), 181 (2007) 77
24. Ferreira, J., et al.: A&A **453**, 785 (2006)
25. Gracia, J., et al.: MNRAS **367**, 201 (2006) 73, 77
26. Heyvaerts, J.: In Plasma Astrophysics, (EADN Astrophysics School VII, San Miniato, Italy 1994). C. Ghiuderi & G. Einaudi (Eds.), Springer, p. 31 (1996) 64, 72
27. Heyvaerts, J., Norman, C.: ApJ **347**, 1055 (1989) 64, 93
28. Heyvaerts, J., Norman, C.: ApJ **596**, 1240 (2003) 93
29. Kato, S. X., et al.: ApJ **565**, 1035 (2002) 73
30. Komissarov, S. S., et al.: MNRAS **380**, 51 (2007) 91, 104
31. Komissarov, S. S., et al.: MNRAS **394**, 1182 (2009) 90, 91, 104
32. Konigl, A.: Phys. Fluids **23**, 1083 (1980) 107
33. Krasnopolsky, R., et al.: ApJ **526**, 631 (1999) 73
34. Li, Z.-Y., et al.: ApJ **394**, 459 (1992) 86, 97, 104
35. Lovelace, R. V. E. et al.: ApJS **62**, 1 (1986) 79
36. Lyubarsky, Y., Eichler, D.: ApJ **562**, 494 (2001) 93
37. Matsakos, T., et al.: A&A **477**, 521 (2008) 73, 77
38. Meliani, Z., et al.: A&A **460**, 1 (2006a) 73
39. Meliani, Z., et al.: A&A **447**, 797 (2006b) 94
40. Michel, F. C.: ApJ **158**, 727 (1969) 80, 91
41. Okamoto, I.: ApJ **573**, L31 (2002) 88
42. Okamoto, I.: MNRAS **185**, 69 (1978) 79
43. Ouyed, R., Pudritz, R.: ApJ **482**, 712 (1997) 73
44. Ouyed, R., et al.: ApJ **582**, 292 (2003)
45. Rosso, F., Pelletier, G.: A&A **287**, 325 (1994) 73
46. Sauty, C., et al.: A&A **389**, 1068 (2002a) 57, 63
47. Sauty, C., et al.: In LNP Vol. 589, A. W. Guthmann, M. Georganopoulos, A. Marcowith, K. Manolakou (Eds.), 41 (2002b) 69
48. Sauty, C.: In LNP Vol. 723, J. Ferreira, C. Dougados, E. Whelan (Eds.), 209 (2007) 73
49. Takahashi, M., Shibata, S.: PASJ **50**, 271 (1998) 93
50. Takahashi, M., et al.: ApJ **363**, 206 (1990) 93
51. Tomimatsu, A., Takahashi, M.: ApJ **592**, 321 (2003) 93
52. Tsinganos, K.: ApJ **252**, 775 (1982) 64
53. Tsinganos, K.: In LNP Vol. 723, J. Ferreira, C. Dougados, E. Whelan (Eds.), 117 (2007) 73
54. Tsinganos, K., et al.: MNRAS **283**, 811 (1996) 65, 69
55. Uchida, Y., Shibata, K.: PASJ **37**, 515 (1985) 85
56. Vlahakis, N.: Analytical Modeling of Cosmic Winds and Jets. Ph.D. thesis, University of Crete, Heraklion (1998) 65, 66, 67, 68, 69, 94
57. Vlahakis, N.: ApJ **600**, 324 (2004) 87, 88, 91, 93
58. Vlahakis, N., Konigl, A.: ApJ **596**, 1080 (2003a) 86, 96, 97, 104
59. Vlahakis, N., Konigl, A.: ApJ **596**, 1104 (2003b) 96, 104
60. Vlahakis, N., Konigl, A.: ApJ **605**, 656 (2004) 96, 104
61. Vlahakis, N., Tsinganos, K.: MNRAS **298**, 777 (1998) 94
62. Vlahakis, N., et al.: MNRAS **318**, 417 (2000) 73, 86, 94, 97
63. Weber, E., Davis, L.: ApJ **148**, 217 (1967) 91
64. Zanni, C., et al.: A&A **469**, 811 (2007) 73

Coronal Heating

Alan William Hood

1 The Sun

1.1 Introduction

The Sun is, of course, the only star we can see in detail and the recent space missions and advances in ground based telescopes means that we can now analyse it in more detail than ever before. Looking at smaller lengthscales and with higher time and spectral resolution, it has become clear that there are many phenomena that are still not clearly understood.

Why should we study the Sun? Firstly, it has a major influence on the Earth, both its environment and its climate, and predicting space weather will be essential if manned missions are sent to Mars. Secondly, it is important for astronomy because if we do not understand the Sun, we cannot understand how stars work. Finally, it is a laboratory for understanding fundamental plasma processes, which has implications for understanding the Cosmos.

The structure of the solar interior is illustrated in Fig. 1. There are three main regions, namely the core, the radiative and the convective zones. The fusion of hydrogen into helium occurs in the *core*, which extends out to about a quarter of the radius of the Sun. Next is the *radiative zone* where the photons are repeatedly emitted and absorbed and the wavelengths lengthen from gamma rays to appear at the surface as visible light. The top of the radiative zone is around 0.7 of the solar radius and now the temperature gradients are sufficiently steep that convection sets in. What is clear from surface observations is that the equator rotates faster than the poles. One of the key results from helioseismology is that this differential rotation is constant along constant latitude inside the convection zone. However, the radiative zone is almost rotating with a constant rate and at the interface between these two zones is an interface layer called the *tachocline* where there is a high radial shear

A.W. Hood (✉)

School of Mathematical and Computational Sciences, University of St. Andrews, St. Andrews, KY16 9SS, UK, alan@mcs.st-and.ac.uk

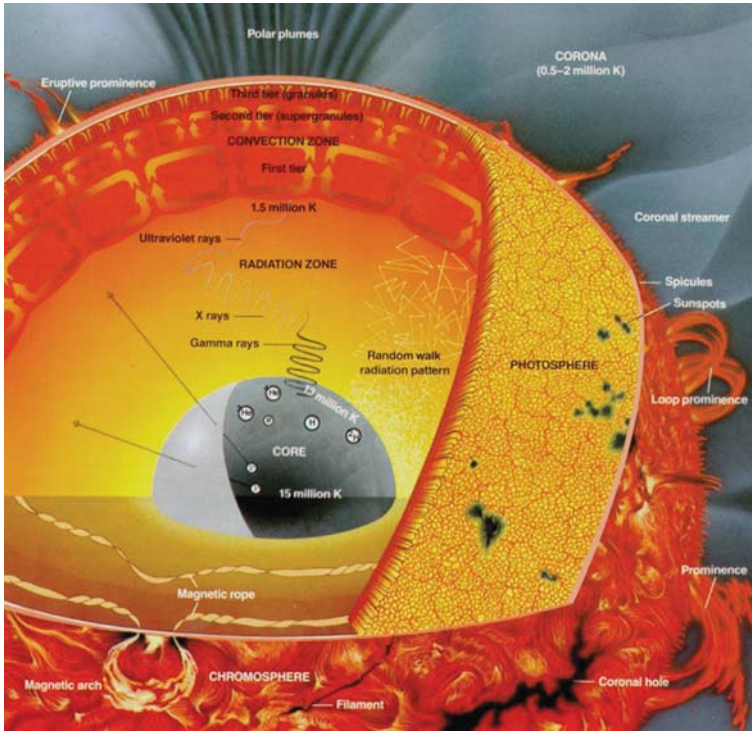


Fig. 1 A cutaway showing the nuclear core, the radiative zone and the convection zone in the interior. The interface layer lies between the radiative and convection zones and is the location where the solar dynamo operates. The visible surface is the photosphere and above it lie the chromosphere and the corona

in the rate of rotation. It is thought that the location of the solar dynamo is in this interface layer.

There are three distinct regions in the atmosphere of the Sun, as illustrated in Fig. 1. The lowest level is the *photosphere*, with a temperature of the order of 6,000 K and this extends up to a height of about 5×10^5 m or, as is more commonly used in Solar Physics, 0.5 Mm. The temperature rises slowly through the *chromosphere* until there is a rapid rise through a narrow region, known as the *transition region*, to the *corona* where the temperature is somewhat above 10^6 K.

1.2 History of the Solar Corona

How do we know that the corona is hot? Our information about the solar atmosphere comes from spectroscopic data and interpretation. Below is a potted history that led to the determination that the corona is hot:

- 1814 – Fraunhofer interprets the spectrum of the Sun.
- 1820 – Invention of photography.
- 1869 – New emission line discovered in the corona and the element is named *coronium*.
- 1875 – Secchi realises that the form of the corona changes during the solar cycle.
- 1908 – Hale discovered magnetic fields in sunspots.
- 1913 – Bohr interpreted the Hydrogen spectrum.
- 1929 – Russell obtained the chemical composition of the Sun by spectroscopic methods.
- 1940 – Edlen establishes that coronium is a highly ionised known element implying the corona has a temperature above 10^6 K.
- 1948 – Biermann and Schwartzchild propose heating by sound waves (they may heat the chromosphere but not the corona).
- 1952 – Magnetograph invented by Babcock.
- 1962 – Orbiting Solar Observatory Series launched and produced first disk images of the corona.

There is a substantial amount of material relating to the coronal heating problem in the literature. One very useful source is the ESA proceedings of SOHO15, a major international meeting to discuss coronal heating held in 2004 [32]. Virtually all the contributions are available online, through the ADS service.

1.3 The Solar Corona

Everything in the solar corona is controlled by the local magnetic field. In the photosphere, the most obvious features of the Sun's magnetism are the sunspots, the number of which varies in time over an approximate 11 year cycle due to magnetic dynamo action in the solar interior. Sunspots are objects of intense research and recent observations have thrown up new problems as to their origin, formation and structure. Figure 2 shows a detailed picture of the central sunspot umbra and the surrounding penumbra. The magnetic field shows up different structures in the chromosphere and corona. Using a Hydrogen alpha filter, we can see the plasma at chromospheric temperatures. The detailed structures also outline the local magnetic field lines.

In Fig. 3, cool dense structures, known as prominences, are seen in absorption as thin dark ribbons on the solar disk and in emission as bright structures on the limb. They are located at typical coronal heights but consist of plasma about 100 times cooler and about 100 times denser than the surrounding plasma. They are relatively long-lived (of the order of days to months) and they are supported against gravity by the local magnetic field. They form when the plasma radiation exceeds the local heating and the plasma cools catastrophically.

White light eclipse images show the overall structure of the solar corona and the difference between solar minimum and solar maximum. However, the structure of the solar corona can only be seen on the disk when the plasma is studied in

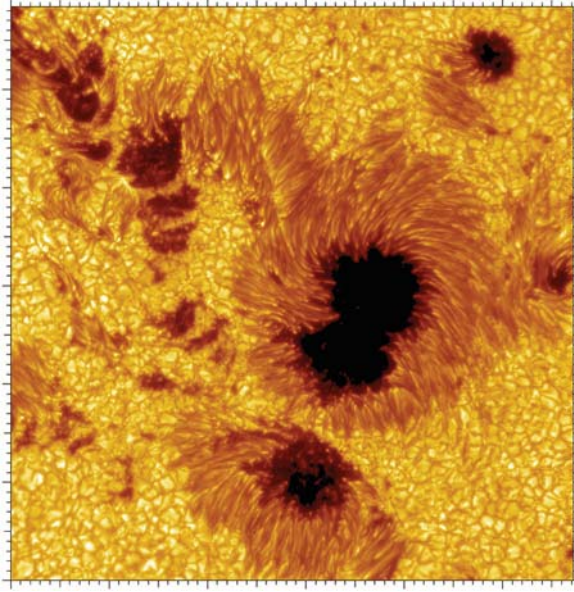


Fig. 2 A sunspot (from the Swedish 1 m Solar Telescope). The *dark* central region is the umbra and the *stripey* region surrounding it is the penumbra

EUV and x-ray. This requires space based instruments, since the Earth's atmosphere will block radiation at these wavelengths. On the top row, Fig. 4 shows soft x-ray images (from Yohkoh) of the corona over a 3 year period. Clearly, the structure of the corona changes quite dramatically over this time. The middle row shows the number of sunspots as a function of time. The bottom row shows four white light

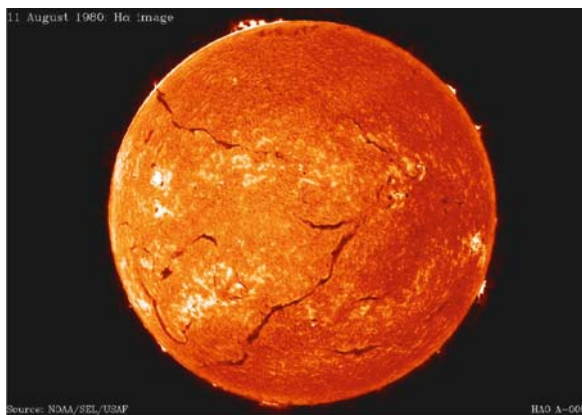


Fig. 3 Prominences appear dark on the disk and bright on the limb (courtesy of the High Altitude Observatory, <http://www.hao.ucar.edu/education/slides/slide6.php>)

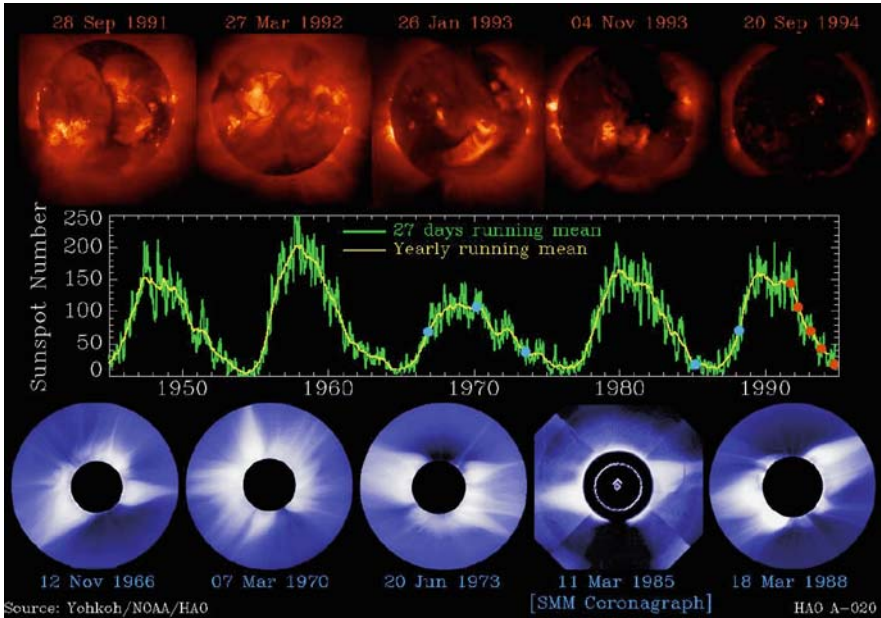


Fig. 4 *Top row:* Soft x-ray images of the solar corona showing the change in the magnetic field over a 3 year period. *Middle row:* The sunspot number showing the solar cycle. The *red dots* refer to the *top row* and the *blue dots* to the *bottom row*. *Bottom row:* Eclipse images showing the shape of magnetic fields in the solar corona at solar minimum and solar maximum (courtesy of the High Altitude Observatory, <http://www.hao.ucar.edu/education/slides/slide20.php>)

eclipse photographs and a white light coronagraph images from the Solar Maximum Mission. The red dots indicate when the soft x-ray images were taken. The blue dots refer to dates of the eclipse and coronagraph images. It is clear that the shape of the coronal magnetic field varies from sunspot maximum to sunspot minimum.

Figure 5 shows an image from the Extreme ultra violet Imaging Telescope (EIT) on SOHO. The observed plasma is at a temperature of around 1.5×10^6 K. The variation in the structure of the corona is quite noticeable. The figure on the right shows the corona in 1999 as the solar cycle is heading towards sunspot maximum. The very bright regions are called active regions and consist of closed magnetic loops (closed in the sense that one end rises from a sunspot into the corona and then closes back down at the photosphere at another sunspot). In 1997 (left hand figure,) just after sunspot minimum, there are very few active regions. What is clear is the large number of transient bright points called x-ray bright points. They are small magnetic loops and they occur equally during sunspot maximum but appear less visible due to the active regions. Also clear in the left hand image are two dark regions at the poles of the Sun. These *coronal holes* consist of open magnetic field lines that connect to interplanetary space and are the main source of the fast solar wind. The coronal holes are still present in the right hand image but they are not confined to the polar regions and holes may form at low latitudes. Away from the

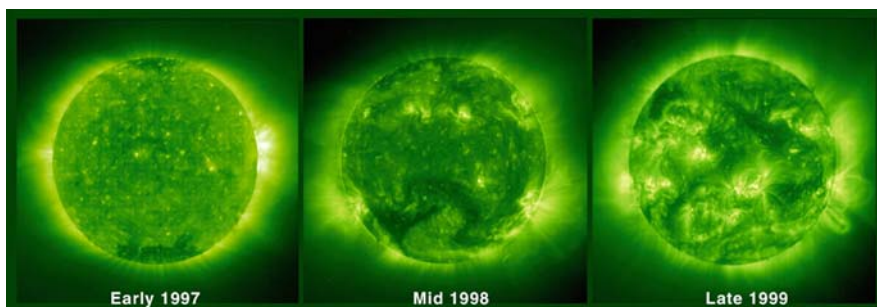


Fig. 5 Three images showing the plasma at 1.5×10^6 K over a 3 year period (courtesy of the SOHO website, http://sohowww.nascom.nasa.gov/gallery/images/large/tricomp_prev.jpg)

active regions and the coronal holes, the corona has a general fuzzy emission called the *quiet sun*.

Coronal loops can be seen in amazing clarity with TRACE satellite. Figure 6 shows the emission from the coronal plasma at a temperature of around 10^6 K. As will be discussed below, the ionised coronal plasma is ‘tied’ to the local magnetic field and it is generally assumed that the observed plasma is outlining the coronal magnetic field. Clearly the field is highly filamentary and it is likely that these fine strands are not yet fully resolved. Why should the plasma on some magnetic fields appear hotter or denser than neighbouring ones remains a mystery and is something that must be explained.

What is clear is that the corona consists of several fundamental building blocks, namely coronal holes, active regions (and their coronal loops) and quiet sun regions, within which transient x-ray bright points are observed. Other transient brightenings are observed including nanoflares and microflares and these will be discussed later.



Fig. 6 Coronal loops above an active region at about 10^6 K (courtesy of TRACE website, http://trace.lmsal.com/POD/images/TRACE171_991106_023044.gif). The plasma emission outlines the local magnetic field

This is an excellent time to study the Sun since there are a larger number of dedicated solar space missions. These include the SOHO, TRACE, RHESSI, Hinode and Stereo missions. When the exciting new high resolution data are becoming available, we expect to make many major breakthroughs in our fundamental understanding of how the solar corona works.

The remainder of this chapter is organised in the following manner. A brief discussion of the energy requirements is given, followed by an estimation of the role various observed coronal phenomena may play in the heating process. To understand the theoretical ideas used later, an introduction to the basic ideas in MHD is presented. Then the main energy conversion mechanisms, namely MHD wave damping and magnetic reconnection, are investigated. Next, applications to the solar atmosphere are given. Finally, the response of the plasma to a sudden input of heat is shown.

2 The Energy Budget

So why is there a coronal heating problem? Firstly, it seems to fly in the face of physical intuition that the temperature should rise as one moves away from the heat source at the centre of the Sun. However, given that the corona is at a temperature above one million degrees, why does it remain hot? The main cooling mechanisms at this temperature and low density are optically thin radiation and thermal conduction parallel to the magnetic field. The timescale for radiative cooling is the order of 3,000 s and the conduction timescale is the order of 500 s. Hence, heat must be fed into the corona on at least this timescale and so there must be a mechanism that provides a continual input of heat to the corona. Withbroe and Noyes [33] estimated that 300 W m^{-2} (or $3 \times 10^5 \text{ erg cm}^{-2} \text{ s}^{-1}$) are required to heat the quiet Sun and coronal holes and $5,000 \text{ W m}^{-2}$ to heat active regions. In addition, the energy required to drive the solar wind is linked with the heating of the corona and there needs to be enough energy available to do both.

Klimchuk [18] summarised the *coronal heating problem* as a flowchart, which is shown in Fig. 7. Firstly, there must be a source for the energy necessary to heat the corona and this must lie in the solar interior (and ultimately in the nuclear core). It is normally agreed that the energy for heating must be stored in the magnetic field in the corona. So is there a sufficient flow of magnetic energy from the interior to the corona? This can be estimated by calculating the Poynting flux through the photospheric surface. Using $\mathbf{E} = -\mathbf{v} \times \mathbf{B}$, the Poynting flux is

$$\frac{\mathbf{E} \times \mathbf{B}}{\mu} \approx \frac{v B_h B_v}{\mu} \approx 10^4 \text{ W m}^{-2},$$

where v is the horizontal photospheric velocity, B_h and B_v are the horizontal and vertical components of the photospheric magnetic field. Here $v = 1.2 \text{ kms}^{-1} = 1200 \text{ ms}^{-1}$, $B_h = 10 \text{ G} = 10^{-3} \text{ T}$, $B_v = 100 \text{ G} = 10^{-2} \text{ T}$ and $\mu = 4\pi \times 10^{-7} \text{ H m}^{-1}$. This suggests that there is sufficient magnetic energy being transported into the corona

to keep it hot. Secondly, there must be a mechanism that can convert this stored magnetic energy into heat. The mechanism responsible for releasing the stored energy is the main part of this chapter and we will investigate MHD waves and magnetic reconnection as two possibilities. Note that this part is a purely theoretical investigation and is entirely independent of the specific application to the coronal heating problem. It simply investigates how the stored magnetic energy is converted into heat. Hence, these conversion mechanisms represent a study of basic plasma processes. Thirdly, after the appropriate mechanism has extracted the stored magnetic energy, the plasma will heat and its temperature will rise to the order of a million degrees. This is investigated by following the thermodynamic evolution of the coronal plasma in response to the rapid deposition of heat by the conversion mechanism. Fourthly, the plasma now radiates its energy in the extreme ultraviolet and soft x-ray part of the electromagnetic spectrum. This plasma radiation can be detected by suitable space based telescopes as the fifth part of the chain. The question now remains as to whether the conversion mechanism actually gives an accurate representation of the observed plasma radiation. The sixth part studies whether the properties of the observed plasma radiation matches the theoretical predictions. If they do not agree, then the loop is repeated and the conversion mechanisms must be modified. At present, the coronal heating problem is divided up into the individual parts and solved separately.

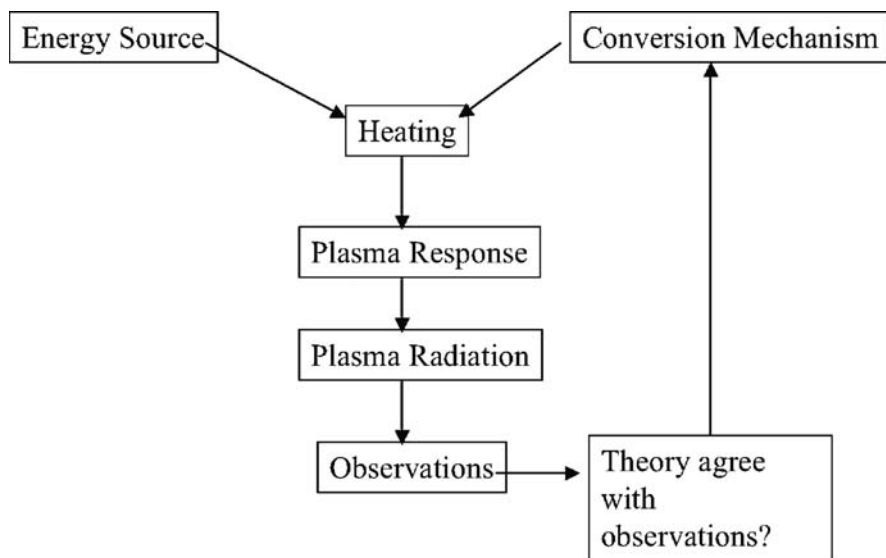


Fig. 7 A flowchart, adapted from Klimchuk [18], showing how the coronal heating problem may be solved. Taking the energy stored in the magnetic field, the conversion mechanism converts this energy into heat. The heated plasma now radiates and this is the radiation that is observed. Comparison between the plasma radiation predicted by the conversion mechanism and the observed plasma radiation can be used to reject or modify the conversion mechanism

3 Observed Coronal Phenomena

This section looks at a variety of observed dynamic solar coronal phenomena that release energy and result in heated plasma. One can use estimates of the energy released during the event to see if it could contribute to the coronal heating. This, of course, does not answer the question of how the stored magnetic energy is released but simply if there is enough energy being released in the event.

3.1 Solar Flares

A solar flare is the result of a rapid release of magnetic energy over a relatively short period of time. Can they contribute to the heating of the corona? They are large, with an area of up to $3 \times 10^{15} \text{ m}^2$ (although this has to be compared to the surface area of the Sun of $6 \times 10^{18} \text{ m}^2$) and the main energy release has a duration of minutes to hours. Simple estimates of the total energy released vary between 10^{28} and 10^{32} ergs and it is fairly easy to see that this amount of energy can be stored in the in situ magnetic field. The temperature of flare heated plasma is normally over $5 \times 10^6 \text{ K}$. The number of flares occurring on the Sun in a day depends on the status of the solar cycle but they may be up to 2–3 on average.

Their contribution to coronal heating is limited, since there are too few events, and thermal conduction across field lines is negligible. Hence, it is difficult to see how the energy could be provided to quiet sun regions.

3.2 Coronal Mass Ejections (CMEs)

Coronal Mass Ejections, more commonly known as CMEs, are large-scale dynamic and highly energetic events. They eject a large amount of mass, up to 10^{13} kg per event, from the Sun into interplanetary space. The initial area is around 10^{15} m^2 but this rapidly expands as they move up and away from the initiation site. The typical lifetime of CMEs is the order of minutes to hours and the energy involved in a CME eruption is 10^{29} – 10^{32} ergs, making them very energetic. The frequency does depend on the solar cycle but there are up to 2 or 3 a day on average. Some, but certainly not all, CMEs are clearly linked to solar flares and some to erupting prominences.

However, the contribution to coronal heating is like that of solar flares, limited. There are just too few events to be useful. In fact, most of the energy released is ejected from the Sun.

3.3 Active-Region Transient Brightenings

These are much smaller and more numerous events. Yet they are still dynamic, covering an area of 10^{12} – 10^{13} m^2 and releasing 10^{25} – 10^{29} ergs in about 2 – 7 min.

Active-Region Transient Brightenings occur every 3 min on average in active regions and every hour in quieter active regions. The energy released in these events could contribute up to 20% of the energy required to heat active regions.

3.4 X-Ray Bright Points

Although called x-ray bright points, due to the poorer resolution of the early x-ray telescopes, they are actually small magnetic loops that are either emerging or cancelling with neighbouring magnetic elements. They are seen in EUV and soft x-rays and cover an area of $5\text{--}10 \times 10^7 \text{ km}^2$, releasing $10^{25}\text{--}10^{28}$ ergs during a lifetime of between 1 h and 2 days. There are the order of 1,500 new x-ray bright points every day, with about 200 present on the Sun at any one time. The energy released in these events could contribute the order of 20–30% of the energy required to heat the quiet sun.

3.5 Nanoflares and Microflares

Nanoflares and microflares are scaled down versions of flares. They are small scale, dynamic events that are observed in EUV and soft x-ray wavelengths [3, 19, 24]. The energy released is less than 10^{27} ergs and their duration is about 10 min. The area covered is about $2.7 \times 10^7 \text{ km}^2$. Although small, these events are very frequent with a frequency between 140 and 475 s^{-1} measured over the disk of the Sun. Hence, they could contribute up to 30% of the energy necessary to heat the corona. The exact amount will depend on the frequency of smaller events and so the distribution of nanoflares is extremely important. Plotting the observed event frequency, $f(E)$ as a function of event energy, E [17] gives a power law distribution as shown in Fig. 8. Estimating the negative power, α , in the power law is extremely important to decide whether small flares (including those too small to be seen at present) dominate the heating or whether large flares dominate. The total energy, P , released from all events is given by integrating the event frequency, $f(E) = f_0 \times [E/E_0]^{-\alpha}$ times the event size, E , over all sizes from the minimum, N , to the maximum, X , detected. Thus,

$$P = \int_N^X f(E) E dE = \frac{f_0 E_0^2}{2 - \alpha} \left[\left(\frac{X}{E_0} \right)^{2-\alpha} - \left(\frac{N}{E_0} \right)^{2-\alpha} \right]. \quad (1)$$

As can be seen, the small events dominate if $\alpha > 2$ and the large events dominate if $\alpha < 2$. Estimating the value of α from figures similar to Fig. 8 is difficult and the result depends on the assumptions used. Typically the value, based on maximum likelihood methods, is slightly below 2. However, there is sufficient uncertainty that it is difficult to conclude that the large events dominate.

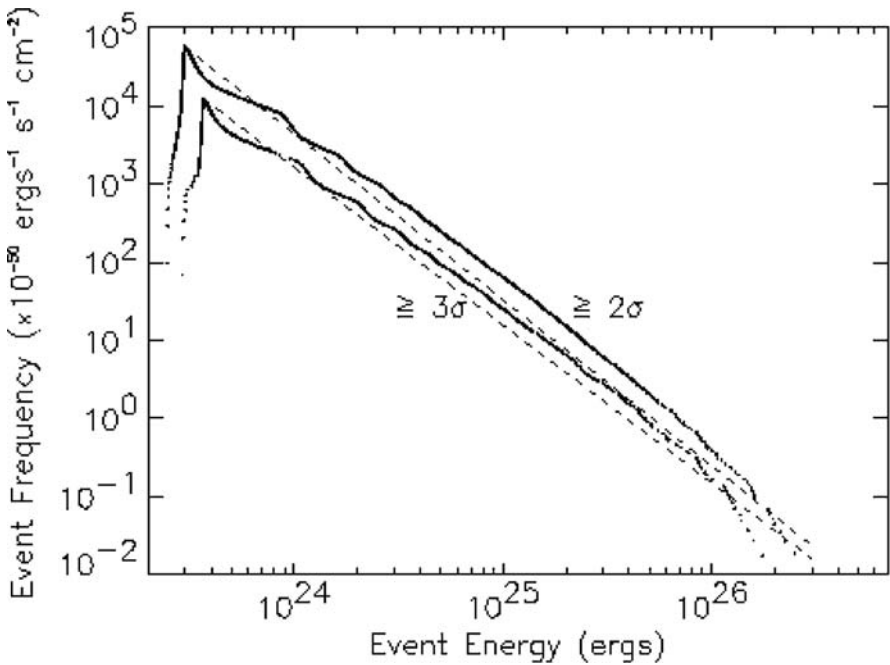


Fig. 8 The frequency of observed events as a function of the event energy. The slope indicates whether the main energy release is due to small scale or large scale energy events

3.6 Summary

The likely contributions to coronal heating from the above observed coronal phenomena are listed in Table 1.

The main limitations in determining what is heating the corona by observing coronal events are as follows:

- Thermal conduction across magnetic field lines is negligible, so that each event can only heat locally. Need something that applies everywhere.
- If an event is identifiable, then it stands out from the background. Since the background is the background hot corona this does not allow one to identify the background heating mechanism.

Table 1 Observed coronal phenomena and their possible contribution to heating active regions (ARs), quiet sun regions (QRs) and coronal holes (CH)

	Regions heated	Heating contribution
Solar flares	ARs	Limited
CMEs	CH	Limited
ARTBs	ARs	20%
X-ray bright points	QRs/CH	20–30%
Nanoflares/microflares	ARs/QRs/CH	30%

4 Alternative Heating Observations

In this section, we mention some alternative observations that can help provide some measure of the form of the coronal heating term in the energy equation and the amount of heat entering the corona through the photospheric surface.

4.1 *Temperature Profile of Coronal Arcades*

Using soft x-ray images of large scale coronal loops, it is possible to estimate the temperature and density profiles along an individual loop. The actual profiles will result from a balance of thermal conduction, plasma radiation and the unknown coronal heating function. The first attempt by Priest et al. [27] suggested that uniform heating (the same heating everywhere along the loop) fitted the data best. However, Mackay et al. [20] included a transition region and found heating located near the footpoints provided the best fit. Finally, Reale [29] considered background subtraction to isolate the loop first and found that heating localised about the apex of the loop gave the best fit. Despite the contradictory results, this method does have potential and just needs better observations of the temperature and density along the loop.

4.2 *Indirect Detection from Magnetograms*

A handle on the energy flow into the corona can be obtained by looking at the photospheric magnetic field. High resolution magnetograms, such as Fig. 9, show the magnetic field in active regions clearly but they also show the quiet sun field as well. An enlargement of a magnetogram, see Fig. 10 left, shows how the magnetic field is clumped into small fragments. These fragments are continually emerging, moving in response to the convective granular motions, fragmenting as they break up, coalescing as two fragments of the same polarity collide and cancelling as two fragments of opposite polarity meet. Potential field extrapolations using these magnetograms show that each source connects to several sources. If the clumps of magnetic field are represented by point sources having the same flux, the potential field can be readily calculated and Fig. 10 right illustrates the domains of connectivity. The key point to get from this is that the connectivity is very complex.

Following these fragments in time, it is possible to show that the photospheric magnetic field is completely recycled in 15 h. This means that all the magnetic field line connections at one time are completely changed every 15 h. This implies that there are significant changes in connectivity and thus magnetic reconnection must be important. Magnetic reconnection will be discussed later. The small magnetic fragments cover the quiet sun photosphere and this is called the *Magnetic carpet*. Interestingly, estimates of the coronal recycling time, the time for the coronal field to completely change all its connections, are much shorter at 1.4 h. Magnetic recon-

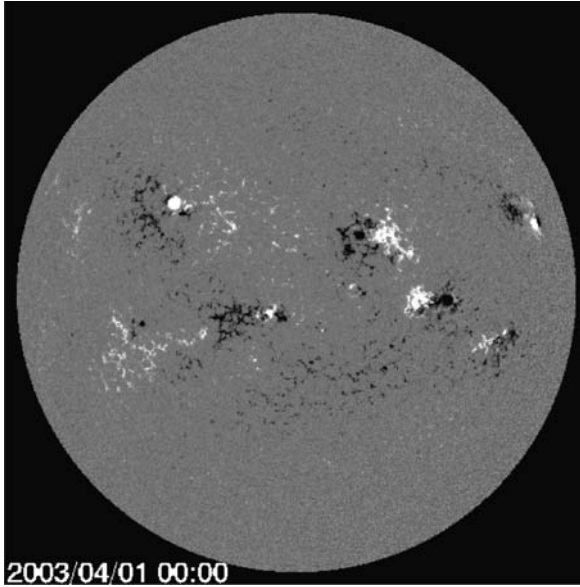


Fig. 9 A magnetogram of the solar disk. The white regions shows positive polarity and the black regions negative. The active regions and sunspots are clearly seen

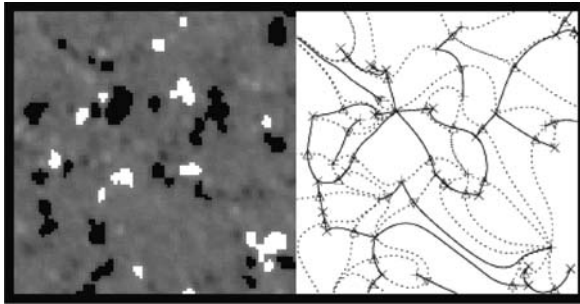


Fig. 10 *Left.* An enlargement of a magnetogram showing a region of the quiet sun. The photospheric field forms discrete fragments that move in response to convective motions. *Right.* The regions of connectivity based on representing the magnetogram by point sources. Within one region, the field connects from one source to one sink (after Close et al. [7])

nection must be efficient. Thus, it can be deduced that motion of the magnetic carpet will contribute to between 90 and 95% of the energy needed to heat the corona in the quiet sun.

5 MHD Equations

The equations of Magnetohydrodynamics (MHD) consist of conservation laws for momentum, mass, energy and magnetic flux, the ideal gas law for the equation of state and Maxwell's equations. They are

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla p + \mathbf{j} \times \mathbf{B} + \rho \mathbf{g} + \mathbf{F}, \quad (2)$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (3)$$

$$\frac{\rho^\gamma}{\gamma - 1} \frac{D}{Dt} \left(\frac{p}{\rho^\gamma} \right) = \nabla \cdot (\kappa \nabla T) - \rho^2 Q(T) + \frac{j^2}{\sigma} + H, \quad (4)$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}, \quad (5)$$

$$p = \frac{1}{\tilde{\mu}} \rho R T, \quad (6)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (7)$$

$$\mathbf{j} = \frac{1}{\mu} \nabla \times \mathbf{B}. \quad (8)$$

ρ is the mass density, \mathbf{v} the fluid velocity, p the gas pressure, \mathbf{j} the current density, \mathbf{B} the magnetic induction (although commonly called the magnetic field), \mathbf{g} the gravitational acceleration, \mathbf{F} other body forces such as viscosity, γ the ratio of specific heats and is commonly taken as 5/3, κ the thermal conductivity tensor, T the temperature, $Q(T)$ the optically thin radiative loss function, σ the electrical conductivity, H any additional heating terms (such as viscous heating), $\tilde{\mu}$ the mean molecular weight (normally about 0.6 in the solar corona), R the gas constant, μ the magnetic permeability and $\eta = 1/\mu\sigma$ the magnetic diffusivity. The convective time derivative is defined as

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla.$$

Note that the Coriolis force and variable rotation are ignored in the equation of motion (2). The assumptions used in deriving the MHD equations are that lengths are much larger than the collision length so that there are many collisions and a single fluid is valid and the plasma is fully ionised and neutral. The primary variables in MHD are the velocity \mathbf{v} and the magnetic field \mathbf{B} , while the secondary variables are the electric field \mathbf{E} and the current density \mathbf{j} . So the two main equations are the induction equation (5) and the equation of motion (2).

It can help interpret phenomena by splitting the Lorentz force in the following manner.

$$\mathbf{j} \times \mathbf{B} = \frac{1}{\mu} (\nabla \times \mathbf{B}) \times \mathbf{B} = (\mathbf{B} \cdot \nabla) \mathbf{B} - \nabla \left(\frac{B^2}{2\mu} \right). \quad (9)$$

Here the magnetic tension is given by $(\mathbf{B} \cdot \nabla) \mathbf{B} / \mu$ and the magnetic pressure is $B^2 / 2\mu$. The tension force will appear whenever the magnetic field lines are curved and the magnetic pressure force will act from regions of strong field strength to regions of weak field strength.

We can estimate the importance of the various terms in (2) by assuming

$$\frac{\partial}{\partial t} \approx \frac{1}{\tau} \text{ and } \nabla \approx \frac{1}{L},$$

where τ and L are a typical timescale and lengthscale of the system. This defines a typical velocity as $v = L / \tau$. In many quasi-static situations, the equation of motion (2) reduces to the *force free* equation

$$\mathbf{j} \times \mathbf{B} = 0. \quad (10)$$

Hence, the Lorentz force dominates all the other terms provided

$$v \ll v_A = \frac{B}{\sqrt{\mu\rho}}, \quad \beta = \frac{2\mu p}{B^2} \ll 1 \text{ and } v_A^2 \gg gL. \quad (11)$$

So provided the plasma β (defined as the ratio of the gas pressure to the magnetic pressure) is small and the typical velocities (both v and the free fall speed) are smaller than the Alfvén speed, v_A , the magnetic field is essentially in force balance and we call this a *force free field*. Note that the Lorentz force is perpendicular to the magnetic field and it can play no role in the balance of forces in the direction along \mathbf{B} . For example, if the magnetic field is vertical and the plasma is static and isothermal, (2) reduces to

$$\frac{dp}{dz} = -\rho g = -\frac{\tilde{\mu} p g}{RT} = -\frac{p}{H}, \quad (12)$$

where the pressure scale height is given by $H = RT / \tilde{\mu} g$. So along the direction of the magnetic field, the plasma is in hydrostatic balance.

To a very good approximation, quasi-static magnetic fields in the solar corona are force free and so \mathbf{j} is parallel to \mathbf{B} . Hence, the simplest solution for a force free field is

$$\nabla \times \mathbf{B} = \alpha \mathbf{B}. \quad (13)$$

This simple looking equation is actually non-linear and complicated to solve, except for a few cases. Two cases are when $\alpha = 0$, and the field is potential, and $\alpha = \text{constant}$, and the field is a linear force free field. The coronal magnetic field is

not actually potential and it is likely that α is not constant either. The coronal magnetic field can be clearly seen in EUV and x-ray images of the Sun. Figure 6 shows coronal loops, as seen by the TRACE satellite, which outline the local magnetic field. The magnetic field structure is approximated by a force free field solution to (13). The loops are almost isothermal since the parallel thermal conductivity, that is parallel to the magnetic field, is substantially larger than the perpendicular value. Thus, heat can readily flow along magnetic field lines but not across them.

6 Wave Heating Mechanisms

6.1 Introduction

In recent years oscillations in the solar corona have been detected and these will be discussed below. Previously, it was assumed that there was insufficient energy in waves to heat the corona but recent high resolution and high time cadence observations have indicated that waves are much more pervasive than originally thought. The estimates of the energy contained within these wave motions is now significant. In addition, waves are observed almost everywhere in the solar corona and are not confined to individual local structures. Waves still remain the only likely candidate for heating the open field regions in coronal holes.

However, before describing observations of them, it is important to ensure that they are correctly identified. So we will start with a description of MHD waves in a uniform medium before discussing the nomenclature used in a structured magnetic field. After the observations, we will present ideas behind possible dissipation mechanisms, namely phase-mixing and resonant absorption.

6.2 Waves in a Uniform Plasma

Consider an equilibrium consisting of a static, uniform plasma with a uniform vertical magnetic field

$$\mathbf{B}_0 = (0, 0, B_0), \quad p_0 \text{ constant}, \quad \rho_0 \text{ constant}, \quad \mathbf{v}_0 = \mathbf{0}. \quad (14)$$

If the equilibrium is disturbed by a small amount, then damped, propagating linear waves will satisfy the linearised resistive MHD equations, where the linear terms have subscript 1,

$$\rho_0 \frac{\partial}{\partial t} \mathbf{v}_1 = -\nabla p_1 + \frac{1}{\mu} (\nabla \times \mathbf{B}_1) \times \mathbf{B}_0, \quad (15)$$

$$\frac{\partial \rho_1}{\partial t} = -\nabla \cdot (\rho_0 \mathbf{v}_1), \quad (16)$$

$$\frac{\partial \mathbf{B}_1}{\partial t} = \nabla \times (\mathbf{v}_1 \times \mathbf{B}_0) + \eta \nabla^2 \mathbf{B}_1, \quad (17)$$

$$\frac{\partial p_1}{\partial t} = -\gamma p_0 \nabla \cdot \mathbf{v}_1. \quad (18)$$

Simple Fourier components of the form $\exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t)$ can be used to determine the possible harmonic wave solutions. There are three types of MHD waves, namely the *fast* and *slow* magnetoacoustic waves and the *Alfvén* wave, and, in the absence of dissipation $\eta = 0$, they satisfy the following dispersion relations. The Alfvén wave satisfies

$$\omega^2 = \frac{(\mathbf{k} \cdot \mathbf{B}_0)^2}{\mu \rho_0} = k_z^2 v_A^2, \quad (20)$$

where the square of the Alfvén speed is defined by $v_A^2 = B_0^2/\mu\rho_0$, and the fast and slow magnetoacoustic waves satisfy

$$\omega^4 - \omega^2 k^2 (c_s^2 + v_A^2) + k^2 \frac{(\mathbf{k} \cdot \mathbf{B}_0)^2}{\mu \rho_0} c_s^2 = 0, \quad (21)$$

where the square of the sound speed is $c_s^2 = \gamma p_0/\rho_0$ and $k^2 = |\mathbf{k}|^2$. The terms fast and slow refer to the phase speed of the two solutions to the quadratic in ω^2 . The fast wave has a faster phase speed than the slow wave. Note that if the phase speed, ω/k , depends on the wavenumber, then an initial wave pulse that is made up of several individual modes will break up (or disperse) since they all travel at different speeds. If, on the other hand, the phase speed is constant, then the wave is non-dispersive and the initial pulse maintains its shape and propagates as a single entity.

In the solar corona the slow and fast phase speeds are approximately given by $c_s \approx 150$ km/s and $v_A \approx 1,000$ km/s. Since the plasma β is small in the corona, the fast wave has a phase speed close to the Alfvén speed and the wave is almost isotropic. However, the slow wave has a phase speed called the tube speed, $c_T^2 = c_s^2 v_A^2 / (c_s^2 + v_A^2)$ that is approximately equal to the sound speed and the wave is almost field aligned. It cannot propagate perpendicular to the equilibrium magnetic field.

Coronal observations always show loop like structure and not a uniform plasma. Hence, we consider a structured plasma, either a magnetic slab or a magnetic tube as shown in Fig. 11. Inside the slab or tube the magnetic field strength is B_o , the temperature T_o and the density ρ_o . In the exterior the values are B_e , T_e and ρ_e respectively. Observations of coronal loops illustrate the inhomogeneous nature of the plasma. Since the magnetic field is almost uniform in a low beta plasma, the Alfvén speed will also be inhomogeneous. In addition, the sound speeds will be less than the corresponding Alfvén speeds. When the internal density is higher than the surrounding plasma the Alfvén speed in the interior will be smaller than the value in the exterior and the structure will act as a wave guide. If the internal density is lower than the surrounding plasma, the tube or slab will not support fast magnetoacoustic

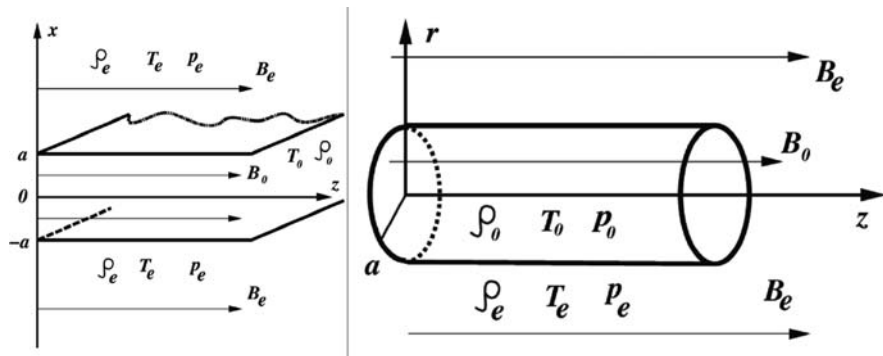


Fig. 11 *Left: A magnetic slab model. Right: A flux tube model*

waves. We will assume a uniform higher density plasma inside the tube or slab surrounded by a uniform lower density plasma.

As each region is uniform, the equilibrium is given by continuity of total pressure at the interface, namely

$$\frac{\rho_0 R T_0}{\tilde{\mu}} + \frac{B_0^2}{2\mu} = \frac{\rho_e R T_e}{\tilde{\mu}} + \frac{B_e^2}{2\mu}. \quad (22)$$

The disturbances in the tube case are defined in terms of cylindrical coordinates and all perturbations are expressed in the form

$$f(r)e^{(im\theta + ikz)}.$$

Two major classes of waves are defined by the form of the oscillations of the boundary, namely the *sausage* mode for which the azimuthal wavenumber is $m = 0$ and the *kink* mode with $m = 1$. The sausage mode does not displace the magnetic axis while the kink mode does. In slab geometry there is no equivalent to the $m \geq 2$ cylindrical modes and so they will not be discussed. A good animation of the two types of waves can be found at www2.warwick.ac.uk/fac/sci/physics/research/cfsa/people/erwin/research/vis/.

6.3 Dispersion Curves for a Coronal Loop

Determining the dispersion relation for a coronal loop (or slab) is more complicated than the simple uniform medium case given by (21). The equilibrium no longer consists of a uniform plasma and are instead functions of, for example, the horizontal coordinate x . Hence, it is not possible to Fourier analyse in the x direction. The resulting differential equation for the velocity component, v_x , can only be solved for particular cases, one such case being two uniform plasmas separated by an interface. Assuming that the solutions are exponentially decaying at large distances, the

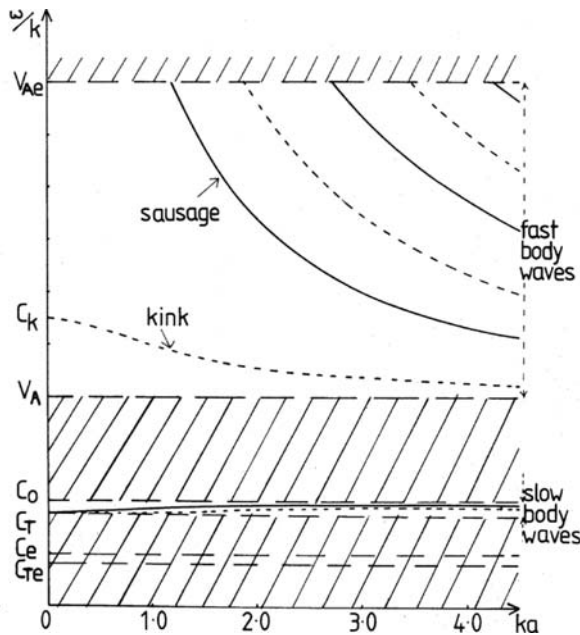


Fig. 12 The variation of the phase speed, ω/k , as a function of the dimensionless wavenumber ka . The *solid curves* correspond to sausage modes and the *dashed curves* to kink modes (after Edwin and Roberts [11])

dispersion relation is obtained by imposing continuity of normal velocity and continuity of total pressure at the interface. The dispersion relation is a transcendental equation involving Bessel functions and needs to be solved numerically. The phase speed as a function of dimensionless wavenumber is shown in Fig. 12. There are two main bands of propagation corresponding to the fast body waves (both sausage and kink modes), lying between the external and the internal Alfvén speeds V_{Ae} and V_A respectively, and the slow body modes, lying between the external sound speed c_0 and the internal tube speed c_T . The internal tube speed is defined as

$$c_T^2 = \frac{V_A^2 c_0^2}{V_A^2 + c_0^2}, \quad (23)$$

with a similar expression for the external tube speed, c_{Te} . There is very little variation in the phase speed for the slow body modes and so they propagate with very little dispersion. The Alfvén modes are torsional modes with a non-dispersive phase speed of v_A .

The fast body modes all exhibit a cut-off in the wavenumber below which the modes are no longer trapped and are now propagating into the external medium. These are called *leaky modes*. The only exception is the fundamental fast kink mode, whose phase speed tends to the kink speed, c_k , as $ka \rightarrow 0$. c_k is defined by

$$c_k = \left(\frac{\rho_0}{\rho_0 + \rho_e} \right)^{1/2} V_A. \quad (24)$$

The fact that it remains for small ka , while the other fast body waves leak away, means that transverse oscillations observed in coronal loops are likely to be kink waves.

The shaded regions in Fig. 12 are where there are no solutions to the differential equation that satisfy the imposed boundary conditions for real ω/k . For example, the region above V_e corresponds to waves propagating in the external region. Since they do not satisfy decaying solutions at large distances, they are excluded. However, they correspond to *leaky* waves.

The key speeds for a coronal loop are summarised in Table 2.

Table 2 Key speeds for the various MHD modes in a cylindrical loop

Sound speed	$c_s^2 = \frac{\gamma p}{\rho}$
Alfvén speed	$v_A^2 = \frac{B_z^2}{\mu \rho_0}$
Slow modes	$c_t^2 = \frac{c_s^2 v_A^2}{c_s^2 + v_A^2}$
Fast modes	$v_A < c < v_{Ae}$
Global kink	$c_k = \left(\frac{\rho_0}{\rho_0 + \rho_e} \right)^{1/2} v_A$

6.4 Observations of Waves in the Solar Corona

Radio observations of coronal oscillations have been known for many years, with good temporal resolution but poor spatial information. The majority of observations of different wave modes are recent, using spectrographs and imagers to determine wave properties. This is a short summary of some of the different instruments that have detected oscillations.

The SUMER spectrograph on the SOHO satellite detected oscillations through the Doppler effect. They were found in hot loops with a temperature of between 6–10 million Kelvin and the periods were in the range 7–31 min. These waves rapidly decayed over 1–2 periods.

Eclipse observations with the SECIS camera detected rapid oscillations with periods in the range 0.5–10 s. There was no apparent decay. The phase speed of one example of a wave propagating along a coronal loop was 2100 km s^{-1} with a period of 6 s. Since these observations involve density variations, the mode must be compressible and, hence, a form of fast wave.

Observations of TRACE intensity variations, with a magnitude of about 4% above the background values, show a longitudinal oscillation propagating up from the lower atmosphere. The modes are damped after a few wavelengths. For fields above sunspots the typical periods are the order of 3 min, while fields on the edge of active regions have periods of around 5 min. The latter is hard to understand since

the 5 min oscillations of the solar interior are evanescent in the lower atmosphere. In addition, TRACE has also seen flare excited transverse oscillations of coronal loops that are rapidly damped.

More recent observations of atmospheric oscillations have been observed by the Hinode and Stereo satellites and I am sure more exciting discoveries will be published shortly.

6.5 Classical Alfvén Wave Damping

To understand the difficulty in dissipating these MHD waves, we now consider damped Alfvén waves with

$$\mathbf{v}_1 = (0, v_y, 0) \text{ and } \mathbf{B}_1 = (0, B_y, 0).$$

Again using Fourier components of the form, $\exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t)$, we have solutions to Eqs. (15) and (17), with $\eta \neq 0$, provided the dispersion relation

$$\omega^2 = k_y^2 v_A^2 - i\omega\eta k_y^2, \quad (25)$$

is satisfied, where the Alfvén speed is defined by

$$v_A^2 = \frac{B_0^2}{\mu\rho_0}. \quad (26)$$

Assuming that the frequency, ω , is fixed, the wavenumber for upward propagating waves satisfies

$$k_y = \frac{\omega}{\sqrt{v_A^2 - i\omega\eta}}.$$

Assuming $\omega\eta \ll v_A^2$, the wavenumber is approximately given by

$$k_y = \frac{\omega}{v_A} \left(1 + i \frac{\omega\eta}{2v_A^2} \right).$$

Thus, the wavenumber is complex with $k_y = k_r + ik_i$. Substituting into the exponential form $e^{ik_y z} = e^{ik_r z} e^{-k_i z}$, it is clear that the Alfvén waves are damped and the damping length, L_d , is the reciprocal of the imaginary part of the wavenumber, namely

$$L_d = \frac{2v_A^3}{\eta\omega^2}. \quad (27)$$

Consider a wave with a period of 5 min, a typical coronal Alfvén speed of $1,000 \text{ km s}^{-1}$ and a resistivity of $1 \text{ m}^2 \text{ s}^{-1}$. The damping length is estimated as $L_d = 5 \times 10^{21} \text{ m}$. This is approximately 10 million light years! Thus, Alfvén waves are extremely difficult to damp with classical resistivity.

6.6 Waves in a Non-uniform Plasma: Phase Mixing

A mechanism for damping Alfvén waves in a more realistic distance was suggested by Heyvaerts and Priest [15]. Since the solar corona consists of many loop structures, with fine scale threads in them, it is clear that the coronal plasma is non-uniform. Variations in the intensity observed, at a particular spectral line, indicate variations in the plasma density. As a simple model, we assume that the background density varies in the horizontal direction as $\rho_0(x)$. The same equations as (15), (16), (17) and (18) apply but this time the wave equation reduces to

$$\frac{\partial^2 B_y}{\partial t^2} = v_A^2(x) \frac{\partial^2 B_y}{\partial z^2} + \eta \frac{\partial}{\partial t} \left(\frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_y}{\partial z^2} \right). \quad (28)$$

The only difference from above is that the Alfvén speed now depends on x . If we initially neglect the resistive damping term ($\eta = 0$), the general harmonic solution can be expressed as

$$B_y = \sin(\omega t - k(x)z),$$

where $k(x) = \omega/v_A(x)$ is the wavenumber and it is a function of x . It is clear that an initially horizontal wave front will appear to turn away from horizontal as part of the wave travels faster where $v_A(x)$ is large and slower where $v_A(x)$ is smaller. Even if there are no variations with x at the lower boundary, x variations will build up as the wave propagates higher into the atmosphere. This is illustrated in Fig. 13 by the

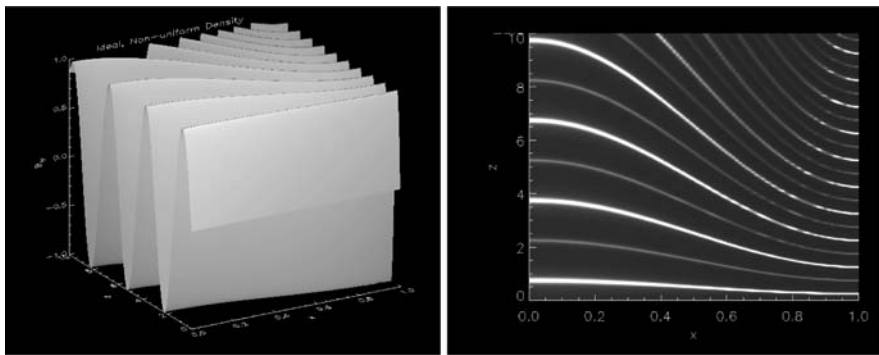


Fig. 13 The variation in Alfvén speed causes the wave fronts to turn. Variations in the x direction build up indefinitely in the absence of η

surface and contour plots of B_y for the Alfvén speed profile $v_A(x) = 1 + 0.5 \cos(\pi x)$. The turning of the wave fronts is maximal where the gradient of the Alfvén speed profile is largest and minimal where the gradient is zero at $x = 0$ and $x = 1$.

Note that the neglected diffusion term, $\partial^2 B_y / \partial x^2$, is given by

$$\frac{\partial^2 B_y}{\partial x^2} = -z^2 \left[\left(\frac{1}{v_A(x)} \right)' \right]^2 \sin(\omega t - k(x)z) - z \left(\frac{1}{v_A(x)} \right)'' \cos(\omega t - k(x)z).$$

Because of the z^2 term, this term must become important as the wave propagates higher into the atmosphere. So, even if the diffusion term is initially small, it must be included at higher heights.

Returning to (28) but now assuming that η is non-zero, we can solve this equation either numerically or obtain an approximate solution by the method of multiple scales. The approximate solution, correct to $O(\eta)$, is [15]

$$B_y = \exp \left[-\frac{(k'(x))^2 k(x)}{6\omega} \eta z^3 \right] \sin(t - k(x)z). \quad (29)$$

The numerical solution is shown in Fig. 14 and the effect of enhanced damping is clearly seen, in that the magnitude of B_y is dramatically reduced in the regions of steep Alfvén speed gradient. Taking a cut at $x = 0.5$, Fig. 15 shows $B_y(0.5, z)$ as a function of z . The damping is clearly seen to be faster than a simple exponential. In fact, it is given by (29), exponential of $-z^3$.

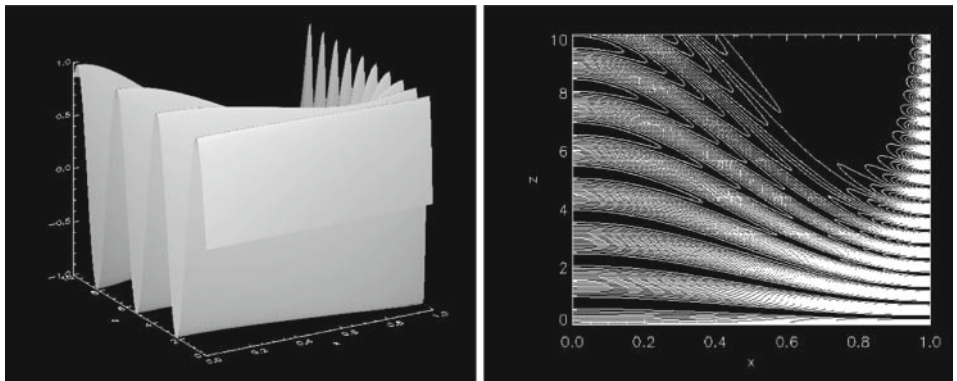


Fig. 14 The variation in Alfvén speed causes the wave fronts to turn, generating gradients in the x direction. A finite value of η means that the field is quickly damped where the Alfvén speed gradient is largest

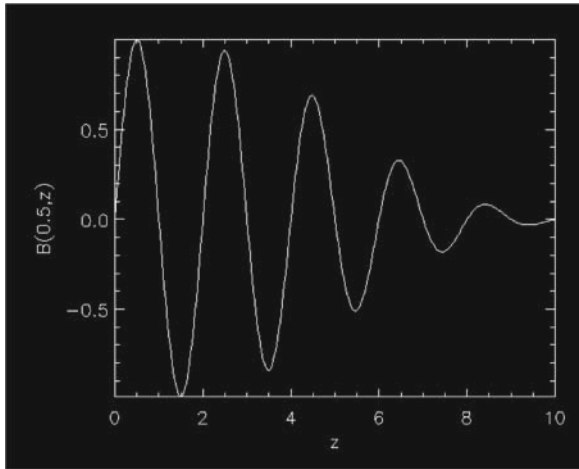


Fig. 15 The variation of B_y as a function of height at $x = 0.5$. The field is rapidly damped with height, showing the $\exp(-z^3)$ envelope

6.7 Waves in a Non-uniform Plasma: Resonant Absorption

As in the last section, consider a background uniform vertical magnetic field and a density that varies in the horizontal direction as $\rho_0(x)$. This time we consider fast MHD waves that are propagating in the horizontal direction. The velocity perturbations $(v_x, 0, v_z)$ are written in terms of Fourier components of the form $f(x)e^{ik_z z + i\omega t}$ and the linearised equation of motion can be expressed in the form

$$\frac{d}{dx} \left(\frac{N}{D} \frac{dv_x}{dx} \right) + (\omega^2 - k_z^2 v_A^2(x)) v_x = 0, \quad (30)$$

where the numerator and denominator are given by

$$N = (v_A^2(x) + c_s^2(x)) (\omega^2 - k_z^2 v_A^2(x)) \left(\omega^2 - k_z^2 \frac{v_A^2(x) c_s^2(x)}{v_A^2(x) + c_s^2(x)} \right), \quad (31)$$

$$D = (\omega^2 - \omega_I^2(x)) (\omega^2 - \omega_{II}^2(x)). \quad (32)$$

The zeros of the numerator give rise to the *continuous spectra* for the Alfvén and slow modes. The zeros of the denominator give turning points, where the nature of the solution changes from oscillatory to exponential or vice-versa.

If the fixed frequency ω is such that the numerator is zero at a point, say x_r , then the equation has a singularity called the *resonance point*. Here the energy in the fast wave can be effectively damped by non-ideal dissipation. In a resistive plasma, note that the magnitude of the perturbed magnetic field B_y scales inversely with the resistivity as $\eta^{-1/3}$ and the width of the resonant layer is approximately $\Delta x \approx \eta^{1/3}$.

Hence, the magnitude of the current in the layer is $j \approx B_y/\Delta x \approx \eta^{-2/3}$. The total Ohmic heating is given by $\eta j^2 \approx \text{constant}$. Thus, the total Ohmic heating is independent of the size of the resistivity.

6.8 Comparison with Observations

Rapid damping of coronal waves may be due to several possible mechanisms in addition to phase-mixing and resonant absorption. Possibilities are *thermal conduction*, *mode conversion*, *area divergence* and *leakage*. Each damping mechanism has its own predictable characteristics that should be detectable in future observations. For example, we have seen that the decay time due to phase mixing is

$$t_{\text{decay}} \approx \left(\frac{6L^2 l^2}{\eta \pi^2 v_A^2} \right)^{1/3},$$

where L is the loop length, l the loop width and v_A the Alfvén speed. If l is proportional to L , then $t_{\text{decay}} \propto (LP)^{2/3}$, where the period P is defined by the time travel time $P = L/v_A$. Finally, since the period is related to the length, we have

$$t_{\text{decay}} \propto P^{4/3}.$$

If the damping was due to wave leakage through the boundary we would expect

$$t_{\text{decay}} \propto LP.$$

Finally, if resonant absorption is responsible for the damping, then assuming that the width of the resonant layer is proportional to the width of the structure we have

$$t_{\text{decay}} \propto P.$$

Analysing several TRACE observations of oscillations the decay time can be plotted against both LP and P and the slope of the curve determined. As shown in Fig. 16, the best fit gives a power of 0.67 ± 0.11 for dependence of t_{decay} on LP and 1.3 ± 0.21 for the dependence on P . This appears to indicate that phase mixing is responsible for the damping. However, this is not really clear since the waves detected are probably not Alfvén waves and it is clear that more observations are needed.

7 Magnetic Reconnection Theory: Two Dimensions

In simple terms, magnetic reconnection is the process that breaks and rejoins magnetic lines of force. In the process of changing the magnetic topology, it converts magnetic energy into kinetic energy and heat and also accelerates electrons and ions

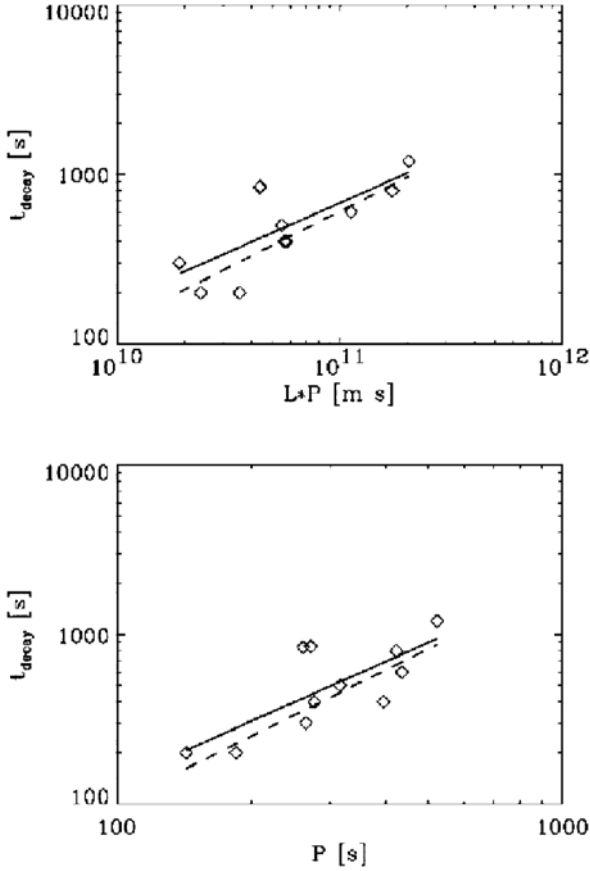


Fig. 16 *Upper:* The decay time t_{decay} as a function of L^*P . *Lower:* The decay time t_{decay} as a function of P

to extremely high speeds. In the solar corona, magnetic reconnection provides a possible mechanism for the energy released in solar flares, CMEs and for the energy required to heat the corona. For a detailed description of MHD, Hall MHD and collisionless reconnection see the book by Birn and Priest [4].

7.1 Brief History

This section gives a brief account of the history of reconnection, without going into any details at this stage. Some key papers are mentioned that can provide the necessary details of the concepts mentioned but not described.

Giovanelli [13] predicted that flares were occurring near neutral points, where the magnetic field was very weak. Cowling [8] concluded that if the energy released during flares was due to Ohmic heating, then the current sheets must be only a few

metres thick. Dungey [9, 10] realised that current sheets could form as the result of an instability near a null point and that magnetic field lines can be broken. He called this breaking of lines of force *reconnection*.

The first simple model of reconnection was due to Sweet [31] and Parker [22] and introduced basic scaling laws for the merging of magnetic field lines. Parker investigated the internal structure of the current sheet due to magnetic annihilation but found that the energy release was too slow by a factor of at least 100 to explain solar flares. The first detailed description of the linear development of resistive instabilities was due to Furth et al. [12] and the *tearing mode instability* remains the only mathematically rigorous description of magnetic reconnection. This will be discussed in detail below. However, the growth rate of the linear instability is still too slow to explain coronal heating.

The next major advance in reconnection theory came when Petschek [25] realised that magnetic energy could be released in slow MHD shocks. He imagined that the current sheet was small in comparison to the length scales of the external plasma and that four slow shocks extended from the corners of the current sheet. This results in fast magnetic reconnection with the rate of reconnection only depending on the logarithm of the magnetic Reynolds number. However, the current sheet is not treated properly.

Numerical simulations of magnetic reconnection started in the 1980s. Biskamp [5] performed some simulations and found that the current sheet always lengthened to form Sweet–Parker reconnection and never to produce Petschek reconnection. However, Priest and Forbes [30] showed that fast reconnection can occur depending on the particular boundary conditions chosen.

The first investigation of reconnection in 3D was presented by Schindler et al. [30]. We now look at several of the above ideas in more detail.

7.2 Basic Ideas

The main equations to describe reconnection are the equation of motion

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho \mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \mathbf{j} \times \mathbf{B}, \quad (33)$$

and the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{v} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}. \quad (34)$$

The induction equation describes how the magnetic field changes due to transport by the plasma velocity and magnetic diffusion. The importance of the terms on the right hand side of (34) can be estimated by looking at the ratio of the advection term to the diffusion term. Using simple order of magnitude estimates, we have

$$\frac{\nabla \times (\mathbf{v} \times \mathbf{B})}{\eta \nabla^2 \mathbf{B}} \approx \frac{V B}{L} \frac{L^2}{\eta B} = \frac{V L}{\eta} = R_m. \quad (35)$$

The ratio of the two terms is called the *magnetic Reynolds number*, R_m , and its size determines which terms are important in the induction equation. Here V is a typical plasma velocity and L a typical length scale. If we replace V by the Alfvén speed v_A , then the ratio is called the *Lundquist number*, $S = v_A L / \eta$. If $R_m \gg 1$, then the advection terms dominate the diffusion terms. The plasma is called ideal and the magnetic field lines are ‘frozen into’ the plasma flow (Alfvén’s frozen-in theorem). The magnetic field lines move with the plasma flow. While there are ideal MHD instabilities that evolve on a fast Alfvénic timescale and can convert some of the magnetic energy into kinetic energy, there can be no change to the topology of the field. Magnetic field lines cannot break and rejoin in ideal MHD. On the other hand, if $R_m \ll 1$, then the diffusion term dominates and the magnetic field lines can slip through the plasma and changes in field topology can now occur. Hence, the plasma is able to access lower energy states and can release more magnetic energy. It is this change in topology that has now been observed and which is generally seen as circumstantial evidence that reconnection is taking place.

What is a typical value for the magnetic Reynolds number? Assume that the typical length scale in the solar corona is the order of $50 \text{ Mm} = 5 \times 10^7 \text{ m}$, the velocity is the order of $100 \text{ km s}^{-1} = 10^5 \text{ ms}^{-1}$ and the magnetic diffusivity is $\eta = 1 \text{ m}^2 \text{ s}^{-1}$ so that the magnetic Reynolds number is $R_m = 5 \times 10^{12}$. This would suggest that magnetic diffusion is negligible, unless the length scales in the current sheets are extremely small (the order of 10 m). The next subsection deals with a possible process by which a current sheet and short length scales may be generated.

7.3 Current Sheet Formation

One place where current sheets can readily form are at magnetic null points, points where the magnetic field is actually zero. There are two types of null points in 2D, namely *X-type* points and *O-type* points and they represent weaknesses in the field. The X-type point can collapse to form a current sheet, as indicated in Fig. 17. The O-type and X-type points can form together, as we will see, from the break up of a current sheet. Consider the initial X-point neutral point that is given by the potential magnetic field

$$\mathbf{B} = (y, x, 0) \quad \Rightarrow \quad \mathbf{j} = \left(0, 0, \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) = (0, 0, 0).$$

Assume that the magnetic field is perturbed in the following manner, similar to a pair of scissors being closed, so that

$$B_x = y, \quad B_y = \alpha^2 x.$$

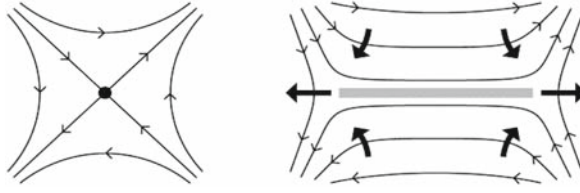


Fig. 17 The X-point in the *left hand* figure can collapse to form a current sheet with the plasma flows indicated by *thick arrows* in the *right hand* figure

The equation of the field lines is

$$\frac{dx}{B_x} = \frac{dy}{B_y}, \quad \Rightarrow \quad \frac{dy}{dx} = \frac{B_y}{B_x} = \frac{\alpha^2 x}{y}.$$

Integrating shows that the field lines lie along $y^2 - \alpha^2 x^2 = \text{constant}$ and $j_z = (\alpha^2 - 1)/\mu$. So the Lorentz force is no longer zero but instead is given by

$$\mu \mathbf{j} \times \mathbf{B} = (\alpha^2 - 1) (-\alpha^2 x, y, 0). \quad (36)$$

Assume that $\alpha^2 < 1$, then the plasma acceleration is positive in the x direction and negative in the y direction (for positive x and y). This means that the force is enhancing the disturbance and the X-point is closing up in the manner indicated in Fig. 17.

7.4 Diffusion of a Current Sheet

Once the X-point has closed up, it forms a current sheet. This section looks at the diffusion of a current sheet, specifically in the limit when $R_m \ll 1$. Hence, the induction equation reduces to

$$\frac{\partial \mathbf{B}}{\partial t} = \eta \nabla^2 \mathbf{B}. \quad (37)$$

The magnetic field is now governed by a diffusion equation with a characteristic time to diffuse, $\tau_d = l^2/\eta$, where l is the width of the current sheet, and a typical diffusion speed, $v_d = l/\tau_d = \eta/l$. Since a current sheet is extremely narrow and to first approximation 1D, the simplest model of a vertical current sheet is $\mathbf{B} = (0, B(x, t), 0)$ where $B(x, t)$ satisfies

$$\frac{\partial B}{\partial t} = \eta \frac{\partial^2 B}{\partial x^2}, \quad (38)$$

with the initial condition

$$B(x, 0) = \begin{cases} B_0 & \text{for } x > 0, \\ -B_0 & \text{for } x < 0. \end{cases}$$

Since there is no physical length scale in this problem (i.e. no boundary conditions at, say, $x = \pm l$), we can obtain a similarity solution that reduces the partial differential equation to an ordinary differential equation. This can be solved exactly in terms of the error function, *erf*, as

$$B(x, t) = B_0 \text{erf}(\xi), \quad \text{where } \xi = \frac{x}{\sqrt{4\eta t}}.$$

This can be expressed in terms of the integral

$$B(x, t) = \frac{2B_0}{\sqrt{\pi}} \int_0^\xi e^{-u^2} du, \quad \text{where } \xi = x/\sqrt{4\eta t}.$$

Note that for $x > 0$ as $t \rightarrow 0$, $\xi \rightarrow \infty$ and so the integral is simply $\sqrt{\pi}/2$ and $B(x, 0) = B_0$. Similarly for $x < 0$. For $x \gg \sqrt{4\eta t}$, $B(x, t) \approx B_0$ and for $x \ll \sqrt{4\eta t}$, $B(x, t) \approx B_0 x / \sqrt{\pi\eta t}$. Thus, the field reduces to a straight line with a gradient that decreases in time.

Taking the diffusion equation (38), multiplying by B/μ and integrating over space, gives an expression for the rate of change of the total magnetic energy. Thus,

$$\begin{aligned} \frac{B}{\mu} \frac{\partial B}{\partial t} &= \frac{\partial}{\partial t} \left(\frac{B^2}{2\mu} \right), \\ \int_{-\infty}^{+\infty} \frac{\partial}{\partial t} \left(\frac{B^2}{2\mu} \right) dx &= \eta \int_{-\infty}^{+\infty} B \frac{\partial^2 B}{\partial x^2} ds \\ &= \left[\frac{\eta B}{\mu} \frac{\partial B}{\partial x} \right]_{-\infty}^{+\infty} - \frac{\eta}{\mu} \int_{-\infty}^{+\infty} \left(\frac{\partial B}{\partial x} \right)^2 dx. \end{aligned}$$

Thus, we have

$$\frac{d}{dt} \int_{-\infty}^{+\infty} \frac{B^2}{2\mu} dx = - \int_{-\infty}^{+\infty} \frac{j^2}{\sigma} dx, \quad (39)$$

where $\mu j = \partial B / \partial x$ and $\eta = 1/\mu\sigma$. Hence, we have shown that the magnetic energy decreases in time (since the right hand side is negative). As energy must be conserved, this magnetic energy is being converted into Ohmic heating. There is a non-zero Lorentz force but this is only bringing more magnetic field into the diffusion region.

We now investigate how an inflow into the current sheet and the diffusion in the sheet can reach a steady state. Consider the stagnation point flow

$$\mathbf{v} = \left(-\frac{U}{a}x, \frac{U}{a}y, 0 \right),$$

where U is a typical velocity and a a typical length scale, and a magnetic field given by

$$\mathbf{B} = (0, B(x), 0).$$

Given this velocity, we can solve the induction equation to get $B(x)$. However, it turns out that this solution is also a solution of the equation of motion and we have an exact solution of the non-linear, incompressible, resistive MHD equations. So the magnetic field satisfies the steady induction equation

$$0 = \frac{d}{dx} \left(\frac{U}{a} x B \right) + \eta \frac{d^2 B}{dx^2}. \quad (40)$$

The first integral gives

$$E = \left(\frac{Ux}{a} \right) B + \eta \frac{dB}{dx}, \quad (41)$$

where E is the constant electric field. For small values of x the left hand side is balanced by the resistive term and

$$B \approx \frac{Ex}{\eta},$$

where for large values of x it is the advection term that balances E and so

$$B \approx \frac{E}{U} \frac{a}{x}.$$

The full solution is given by solving the first order linear differential equation as

$$B = \frac{E}{\eta} e^{-x^2/2l^2} \int_0^x e^{u^2/2l^2} du, \quad (42)$$

where the width of the diffusion layer, l , is defined by $l^2 = \eta a / U$. The full solution and the approximations are shown in Fig. 18.

7.5 Tearing Mode Instability

Ideal MHD conserves the magnetic fieldline connectivity so that field lines maintain their identity, are frozen into the plasma and must move with the plasma flow. In a resistive, non-ideal plasma field lines need not maintain their identity. This section

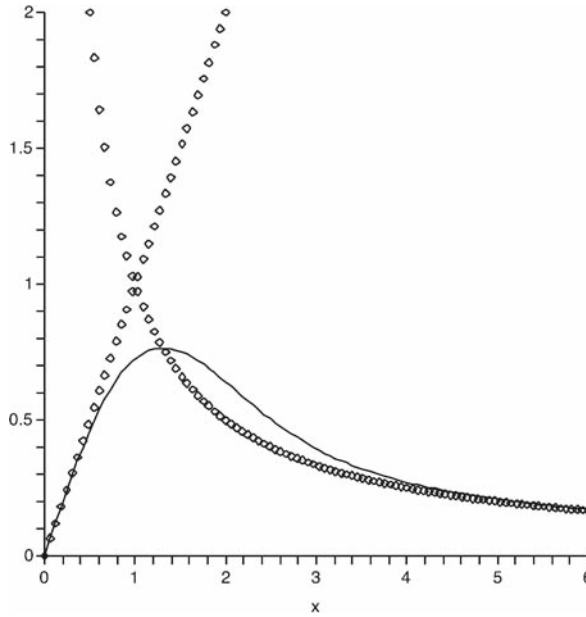


Fig. 18 The full steady state solution (42) as a function of x and the two approximations, $B \approx \frac{Ex}{\eta}$ and $B \approx \frac{E}{U} \frac{a}{x}$

will investigate the effects of finite resistivity. Now the topology of the magnetic field can change, from straight fieldlines to closed fieldlines, and create magnetic islands.

If we have a magnetic field $\mathbf{B} = (0, 0, B_{z0}(x))$ in equilibrium with a gas pressure $p_0(x)$, such that

$$p_0 + \frac{B_{z0}^2}{2\mu} = \text{constant}, \quad (43)$$

then in ideal MHD a perturbation that squeezes the field lines together will simply build up a magnetic pressure that opposes the motion. However, in a resistive plasma the field lines can break and reconnect, accessing a different magnetic topology with a lower magnetic energy, as indicated in Fig. 19. The *tearing mode* is a linear instability that can grow exponentially in time and cause reconnection to occur.

First of all we must linearise the non-linear MHD equations about the equilibrium. Thus, we set

$$\begin{aligned} p &= p_0 + p_1(x)e^{ikz}e^{\sigma t}, \\ \mathbf{B} &= \mathbf{B}_0 + \mathbf{B}_1e^{ikz}e^{\sigma t}, \\ \mathbf{v} &= \mathbf{v}_1e^{ikz}e^{\sigma t}, \end{aligned}$$

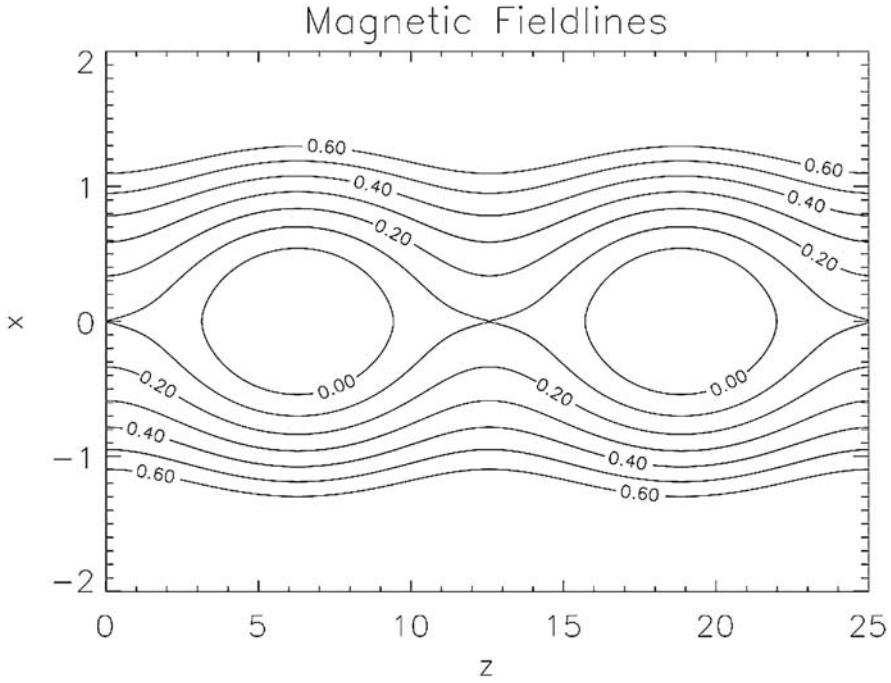


Fig. 19 The tearing mode instability forms X-type and O-type neutral points. The creation of closed fieldlines illustrates change in topology from straight lines and indicates that reconnection has occurred

and so the linearised, incompressible, resistive MHD equations are

$$\rho_0 \frac{\partial \mathbf{v}_1}{\partial t} = -\nabla p_1 + \mathbf{j}_1 \times \mathbf{B}_0 + \mathbf{j}_0 \times \mathbf{B}_1, \quad (44)$$

$$\nabla \cdot \mathbf{v}_1 = 0, \quad (45)$$

$$\frac{\partial \mathbf{B}_1}{\partial t} = \nabla \times (\mathbf{v}_1 \times \mathbf{B}_0) + \eta \nabla^2 \mathbf{B}_1. \quad (46)$$

We have chosen incompressibility ($\nabla \cdot \mathbf{v}_1 = 0$) to simplify the analysis. Thus, because the divergence of both \mathbf{B}_1 and \mathbf{v}_1 are zero, it is convenient to set

$$\mathbf{B}_1 = \nabla A_1 \times \mathbf{e}_y, \quad \mathbf{v}_1 = \nabla \psi_1 \times \mathbf{e}_y,$$

where the perturbed flux function is $A_1(x)e^{ikz}e^{\sigma t}$ and the streamfunction is $\psi_1(x)e^{ikz}e^{\sigma t}$. We now have

$$\sigma A_1 - ik B_{z0} \psi_1 = \eta \left(\frac{d^2 A_1}{dx^2} - k^2 A_1 \right), \quad (47)$$

$$\sigma \rho_0 \left(\frac{d^2 \psi_1}{dx^2} - k^2 \psi_1 \right) = \frac{ik}{\mu} \left[\frac{d^2 B_{z0}}{dx^2} - B_{z0} \left(\frac{d^2}{dx^2} - k^2 \right) \right] A_1. \quad (48)$$

Equation (48) is derived by taking the curl of the linearised equation of motion. This eliminates the perturbed pressure gradient. Because η is small, there are boundary layers at the location where $B_{z0}(x) = 0$, say $x = 0$. At the field reversal reconnection can occur. Depending on the values of k and the value of $d^2 B_{z0}/dx^2$, the plasma can be unstable with a growth rate that scales with fractional powers of the Alfvén timescale $\tau_A = 1/v_A k$ and the diffusion timescale $\tau_r = \eta/L^2$. For Cartesian problems it can be shown that

$$\sigma \approx \tau_A^{-2/5} \tau_r^{-3/5}. \quad (49)$$

A consequence of the tearing mode is the break up of the current sheet into O-type and X-type null points, where the magnetic field is zero, as shown in Fig. 19. At the O-type points the current is enhanced and at the X-type points it is reduced. We can think of the current sheet as being represented by many line currents in the y direction, lying along the z axis. The instability causes the line currents to attract at the O-type points, increasing the current there. Note that the instability only occurs when resistivity is non-zero.

Finally, it is possible to extend this description to include a magnetic field component in the y direction. The same X-type and O-type geometry is seen but these points are not null points since the magnetic field no longer vanishes at those points. The null points are *not essential* for the tearing mode to occur. All that is needed is a concentrated current sheet, where the current is large. If a is the width of the current sheet and k is the wavenumber of the disturbances along the sheet, then tearing will occur when ka is typically less than unity. Note that the magnitude of the current is the order of $B_0/\mu a$, so as a is reduced the current increases.

7.6 Reconnection Models

Having looked at the formation of a current sheet and how an instability can start reconnection, we look at some models that describe how steady reconnection can develop. The first model was the Sweet–Parker model [31, 22]. This is simply an order of magnitude model that assumes the diffusion layer has a length of $2L$ and a width of $2l$ as shown in Fig. 20. The in-flow velocity is v_i and the in-flow magnetic field is B_i . The out-flow velocity and magnetic field are v_o and B_o . If the plasma is incompressible, conservation of mass gives

$$v_i L = v_o l,$$

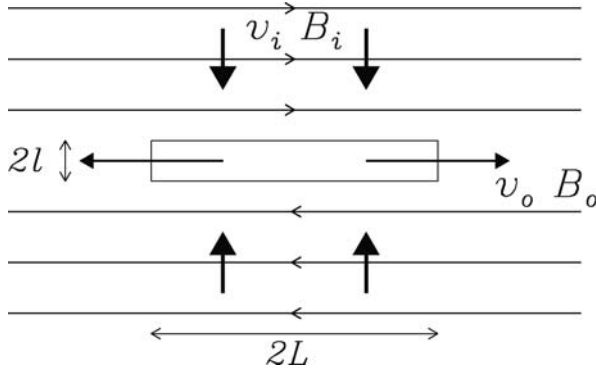


Fig. 20 The Sweet–Parker model for steady state, 2D reconnection. *Solid arrows* indicate the plasma velocity

and balancing the in-flow velocity with an outward diffusion velocity defines v_i as

$$v_i = \frac{\eta}{l}.$$

Estimating the acceleration due to the Lorentz force along the sheet and into the out-flow region gives

$$v_o = v_{Ai},$$

where v_{Ai} is the Alfvén speed associated with the in-flow magnetic field, that is the magnetic field component that actually reconnects. Thus, the reconnection rate can be determined as

$$v_i = \frac{v_{Ai}}{R_{mi}^{1/2}}, \text{ where } R_{mi} = \frac{Lv_{Ai}}{\eta}. \quad (50)$$

It should be noted that this model is still just a cartoon and, although backed up by numerical simulation, there is no detailed mathematical description of the diffusion region. Nonetheless, what is clear from this model is that the reconnection rate depends on the inverse square root of the magnetic Reynolds number in the in-flow region. For large values of R_m this reconnection rate is too slow to explain the energy released in solar flares or indeed for heating the solar corona. The only way to speed things up would be to reduce the value of L and have a much shorter current sheet.

Petschek [25] suggested that slow MHD shocks could extend from the corners of a much smaller diffusion region, as shown in Fig. 21, and that the majority of the magnetic energy could be released in the slow shocks rather than in the current sheet. A detailed analysis of the external region showed that the reconnection rate could be as large as

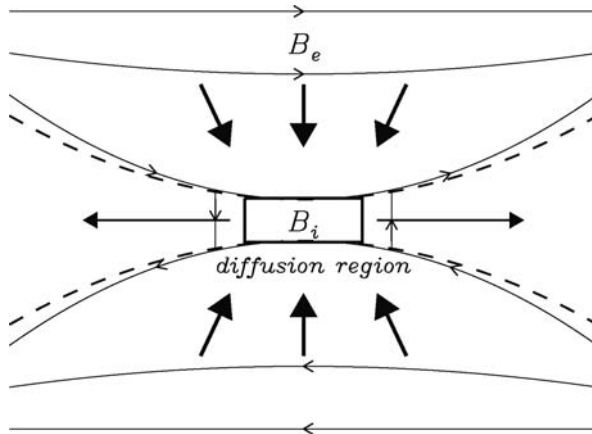


Fig. 21 Petschek reconnection, showing the diffusion region and the four slow mode shocks emanating from the corners (shown as *dashed curves*). *Solid arrows* indicate the plasma velocity

$$v_i = \frac{\pi v_{Ai}}{8 \log R_{mi}}. \quad (51)$$

The logarithmic dependence on the magnetic Reynolds number is very weak and the reconnection rate is now the order of a tenth of the Alfvén speed. This was the first fast reconnection model.

Priest and Forbes [28] have presented a new generation of fast reconnection regimes, of which Petschek is only one. The theory agrees with numerical simulations provided the same boundary conditions are used.

The length scales inside the current sheet can become extremely small so that MHD is no longer appropriate and additional physical terms must be included. Hall MHD, in which the Hall term in Ohm's law is retained, has been quite successful in obtaining a high reconnection rate. In addition, kinetic theory is important when collisions are no longer dominant. Details of these approaches can be found in Birn and Priest [4].

8 Magnetic Reconnection Theory: Three Dimensions

Reconnection in 3D opens up many new features that are just not present in 2D. The detailed theory of how magnetic field lines break and reconnect is much more recent and less well developed. Before looking at some of the ideas, it is useful to introduce some of the building blocks that will be needed later on.

8.1 Magnetic Null Points

Magnetic null points are locations where the magnetic field is actually zero and these represent weaknesses in the field. In 3D the simplest null point has a magnetic field of the form

$$\mathbf{B} = (x, y, -2z). \quad (52)$$

Here the null point is at the origin and nearby the field can be approximated by linear expressions in x , y and z . The field lines are indicated in Fig. 22. It is clear that there are two families of special field lines that pass through the null point. The single field line along the z axis is called the *spine* and the group of field lines lying in the x - y plane are called the *fan plane*. In 2D there are only a pair separatrix curves passing through the null and one of the separatrix curves becomes the spine while the other forms the fan plane when considering 3D null points. The fan plane, when extended away from the null point, forms a surface that separates regions of different magnetic connectivity and so this surface is called a *separatrix surface*. Note that, although the magnetic field strength is zero at the null point, it is not zero on either the separatrix surface/fan plane or spine. If there are two 3D null

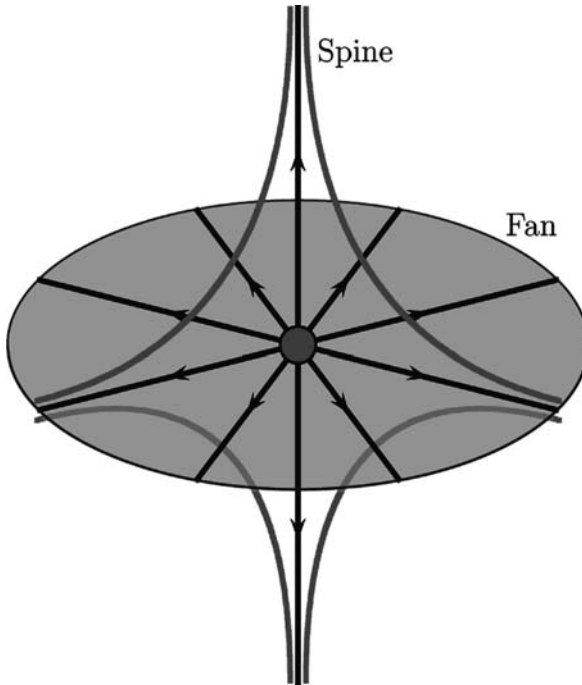


Fig. 22 Magnetic field lines near a simple 3D null point. Special field lines coming from the null show the spine and the fan plane

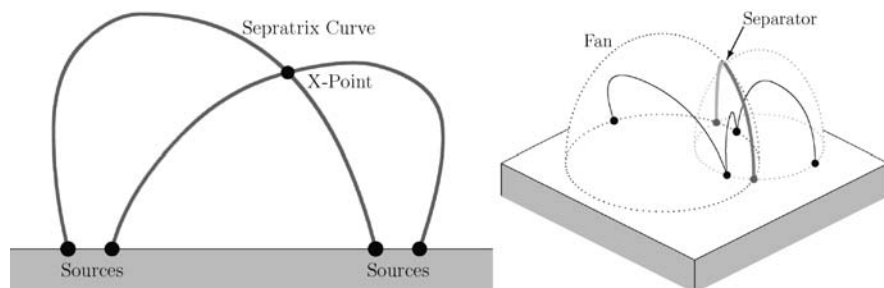


Fig. 23 *Left:* Magnetic field lines passing through a 2D null point are the separatrices. These separate regions of different connectivity between the four sources. *Right:* In 3D the intersection of two separatrix surfaces is a single field line called the separator that connects two null points

points, then the separatrix surfaces may intersect giving rise to a special field line that connects the two nulls, called a *separator*. Projecting the magnetic field onto a plane perpendicular to the separator will show a magnetic field structure that is topologically similar to the 2D picture of an X-point (see Fig. 23). Of course, since there is a guide field, the field strength is only zero at the null points.

In 2D, magnetic reconnection transfers magnetic flux from one 2D region to another through the null point. In 3D reconnection, the transfer of magnetic flux from one 3D region to another can occur along the separator.

8.2 The Magnetic Skeleton

The magnetic field in the solar corona is highly complex because there are many distinct sources in the photosphere. To understand how this complexity is linked to ways of heating the solar corona, we need to understand where the currents are likely to form and, hence, where the heating will occur. Detailed numerical experiments have been performed but to see what is happening pictorially requires a detailed understanding of the magnetic topology. One way of doing this is to plot many field lines and where they end up. This is a little random and it is better to determine the special field lines that outline the *magnetic skeleton*. To do this, it is important to locate the nulls and then determine the spine and fan planes. Next, if there are two nulls and their separatrix surfaces intersect, then there will be a separator field lines connecting the nulls. This is illustrated in Fig. 24.

The topology can be very complex, with more than one separator coming from a null point, with null points in the corona as well as in the photospheric boundary. However, what is clear is that there can be many separatrix surfaces in the corona. As we shall see in the sections below, separatrix surfaces are ideal locations for current sheets to form and, hence, they are likely to indicate the locations where coronal heating and magnetic reconnection will occur. Assuming that the photospheric sources are slowly moved by the changing convection pattern, we can investigate how the topology of the magnetic field can also change. If there is a dramatic

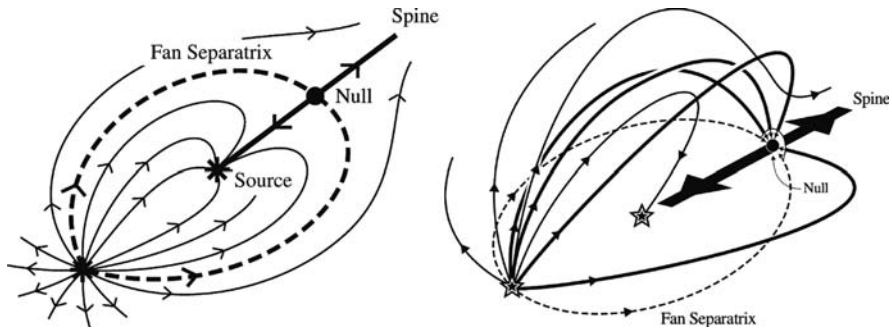


Fig. 24 Sample field lines, projected onto the lower boundary, for two unbalanced sources (*left*) and the corresponding 3D magnetic skeleton (*right*). There is one null point lying on the line joining the two sources, beyond the weaker source

change in topology of these potential field calculations through bifurcations, then it is likely that there will be a dramatic change in the evolution of the real field. A bifurcation will require reconnection in order for it to occur.

8.3 Examples of Reconnection in 3D

Reconnection can occur in three dimensions either at a null point or in the absence of a null. At a null point there are three different types of reconnection, namely spine reconnection, fan reconnection and separator reconnection.

One of the surprising differences between 2D and 3D concerns the field line velocity. In 2D, away from the null point, the electric field, the magnetic field and the velocity satisfy

$$\mathbf{E} + \mathbf{w} \times \mathbf{B} = 0,$$

where we have defined the field line velocity \mathbf{w} . In the ideal MHD region, away from the null point, \mathbf{w} is identical to the plasma velocity \mathbf{v} . In 2D, \mathbf{w} exists everywhere except at the null point, where the magnetic field lines change their connectivity. If two identical flux tubes are brought together to reconnect at the null point, they will join up with their partner and form another two identical tubes. The flux tubes rejoin perfectly. In 3D, if there is an isolated diffusion region surrounding a null point, then there is no \mathbf{w} that satisfies the above ideal Ohm's law. What happens is that field lines continuously change their connectivity in 3D. If two flux tubes are brought into the diffusion region, they will split up but will *not*, in general, rejoin perfectly.

The non-linear evolution of a kink instability has been studied by Browning et al. [6] and Hood et al. [16]. An initially twisted magnetic loop becomes unstable to an ideal kink instability. The kink forms a current sheet and magnetic reconnection

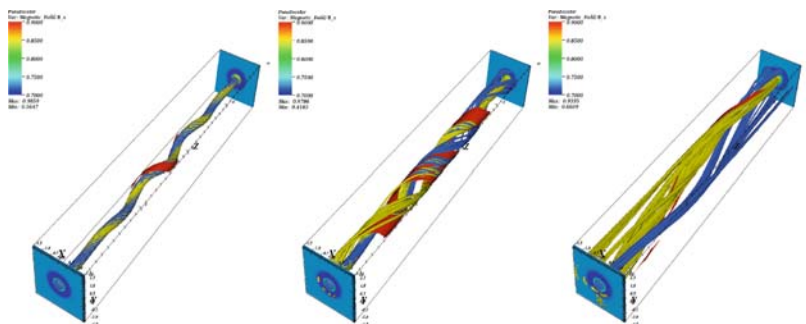


Fig. 25 *Left:* The twisted loop becomes unstable to the linear kink instability. *Middle:* A helical current sheet wraps around the loop and reconnection starts to occur. *Right:* Towards the end of the simulation, substantial reconnection has resulted in two almost unconnected loops

occurs when there is no null point in the system. The tube eventually untwists and relaxes towards a lower energy state. Figure 25 shows the evolution at three different times. In the left hand figure, the tube has begun to kink and a small current sheet (red) is forming. The blue fieldlines are traced from the centre of the loop at the left hand end and the yellow fieldlines from the right hand end. There is an axial field everywhere but only the fieldlines near the loop centre are shown. Obviously the blue and yellow fieldlines wrap around each other. As the instability develops, the current sheet extends along the loop, as shown in the middle figure, and the blue and yellow fieldlines are no longer closely linked. This illustrates that reconnection has occurred. In the right hand figure, the instability is nearly finished and the current sheet is almost entirely dissipated. The blue and yellow fieldlines now form two almost separate loops that slowly wind around each other.

In simulations looking at how magnetic fields emerge through the photosphere and rise into the corona, examples of 3D reconnection have been clearly seen. For example, Archontis et al. [2] looked at how one tube emerged and formed a non-uniform corona. A second tube then emerged into this more complex field and a current sheet formed at the separatrix surface/interface between the two flux systems, as shown in Fig. 26. Although there is no magnetic field outside the initial tubes, they rapidly expand and fill the volume above the solar surface. Since there is a field component in the ignorable direction, there is no null point associated with the current sheet. As the sheet strengthens and thins, a tearing mode is excited and plasmoid-like structures are formed, even though there is no null point in this region. These plasmoids are ejected from the current sheet, allowing reconnection to occur more rapidly. Eventually, a vertical current sheet forms and a hot jet is created that causes the plasma to flow along the separatrix curve.

Reconnection jets have also been seen in simulations investigating how an emerging flux tube interacts with a horizontal uniform magnetic field in the corona. The initial set-up is shown in Fig. 27 (left). Depending on the orientation of the overlying field, the emerging field can meet the coronal field at any angle from anti-parallel

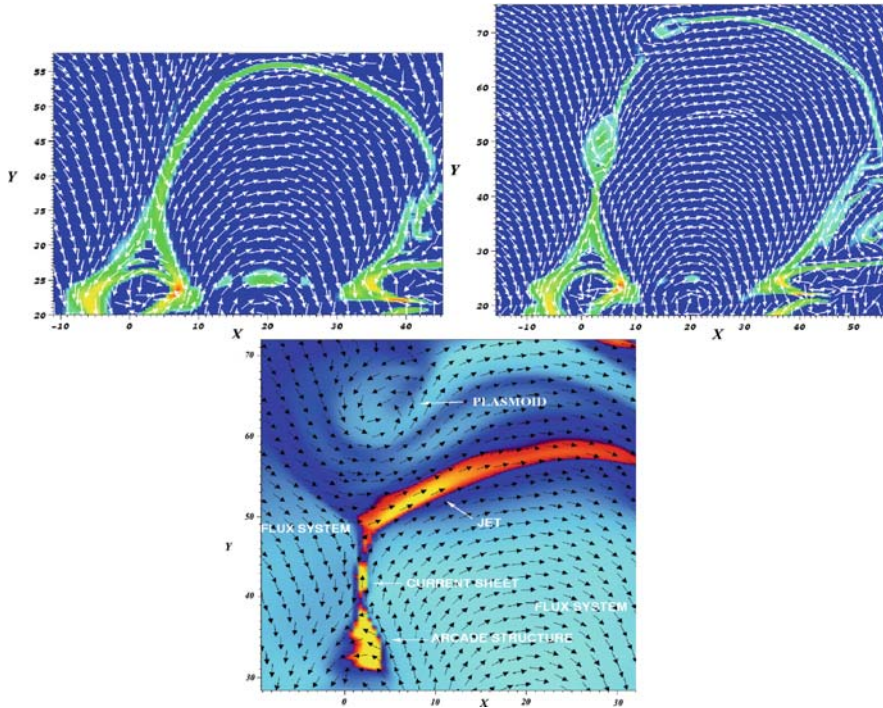


Fig. 26 *Upper:* The arrows show field lines and the magnitude of the current is shown in colours. The current sheet forms between the two flux systems. *Middle:* The current sheet tears, forming plasmoids. *Lower:* The colour contours indicate the plasma temperature and arrows the direction of the local magnetic field. High temperature jets are formed along the separatrix

to parallel. A current sheet forms along the separatrix surface between the two flux systems (shown in blue in Fig. 27 (right) and reconnection can occur. Strong reconnection outflows are seen as jets, shown as green isosurfaces in Fig. 27 (right). In the majority of the cases analysed, there is no null point related to the reconnection site but it does occur along separatrix surfaces. Further details can be found in Archontis et al. [1].

9 Applications to the Solar Corona

The theories of reconnection require current sheets that are much narrower than the typical lengthscales of the corona and if we are going to use these current sheets to heat the corona there must be many such sheets in a single coronal loop. How can a large number of current sheets form in the corona? Parker [23] suggested that random motions at the photosphere, due to granulation, would slowly braid magnetic field lines in the corona resulting in the formation of many current sheets.

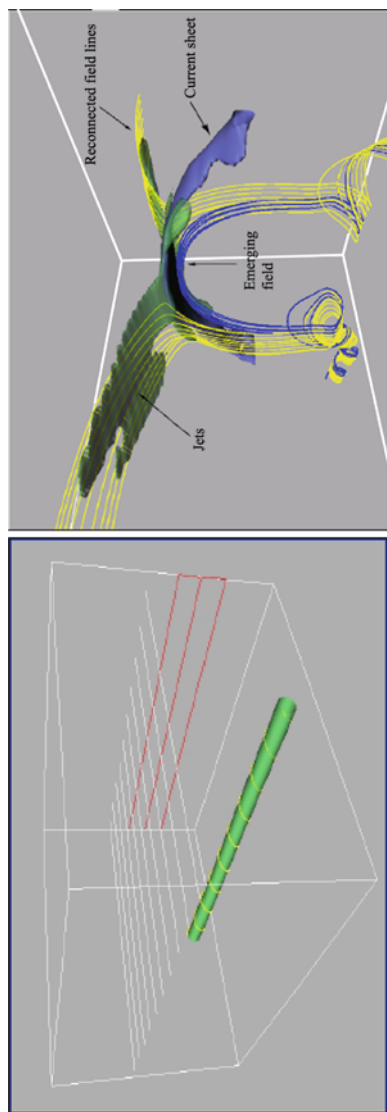


Fig. 27 *Left:* The initial configuration consists of a stratified atmosphere, a flux tube in the solar interior and a uniform horizontal coronal magnetic field. *Right:* High temperature jets (*green isosurface*) are formed due to reconnection between the emerging field and the overlying coronal field

Using observed magnetograms, it is possible to track the movement of magnetic fragments at the photosphere. What is seen is that fragments can either emerge (as a new fragment appears), merge (as two fragments of the same polarity merge), split (as a fragment breaks into two or more components) or cancel (as two opposite polarities merge and cancel). Hagenaar [14] estimated that the photospheric magnetic flux is completely re-cycled every 15 h. Close et al. [7] constructed coronal potential fields from photospheric magnetograms and studied the connectivity of the resulting field. They found that the coronal magnetic field is re-cycled every 1.5 h. This suggests that reconnection must be efficient in the corona and that it must occur at many distinct locations. Hence, the Parker braiding idea can really form many discrete current sheets. Each one will allow reconnection, in the form of a nanoflare, and will release magnetic energy as heat throughout the corona. To heat a single coronal loop, as observed by TRACE, requires 1 nanoflare every second. However, each coronal loop may in fact have around 100 different photospheric flux sources in what is called the *magnetic carpet*.

This section will show how current sheets can develop with two simple examples, both involving the simple motion of discrete magnetic sources.

9.1 Flux Tube Tectonics

Parker's idea of braiding involved complex photospheric motions of an initially uniform magnetic field. However, Priest et al. [26] realised that the same effect can occur with a simple flow pattern and discrete magnetic footpoints in the photosphere. Neighbouring sources can move past each other and currents can build up on the separatrix surfaces between them, in a manner similar to plate tectonics on the Earth. Thus, *flux tube tectonics* will cause currents to build up in the quasi-separatrix surfaces between the sources and eventually reconnection will occur and heat the loop. The initial build up of currents was demonstrated in the numerical simulations of Mellor et al. [21]. Four positive magnetic sources are placed on the left hand boundary representing the photosphere and four negative source on the right hand boundary, representing where the sources return to the photosphere. An imposed photospheric flow moves two of the sources on each boundary between the other two stationary sources. The initial set up at one of the boundaries is similar to the cartoon in Fig. 28. The central source is driven between the outer two with a smooth velocity profile. Thus, although the velocity and the motion of the footpoints is smooth, the field distribution is not smooth and discontinuities can develop in the magnetic field, which will appear as current sheets.

Currents sheets rapidly build up on the separatrix surfaces as shown schematically in Fig. 28 and from the simulations in Fig. 29. What is clear from the results is that the current sheets are almost one dimensional and this allows one to develop a simple theory to explain the main features in the simulation. The magnetic field strength builds up quadratically in time, as shown in Fig. 30 for both our simple model and the numerical solution. Note that the numerical solution has an oscillation

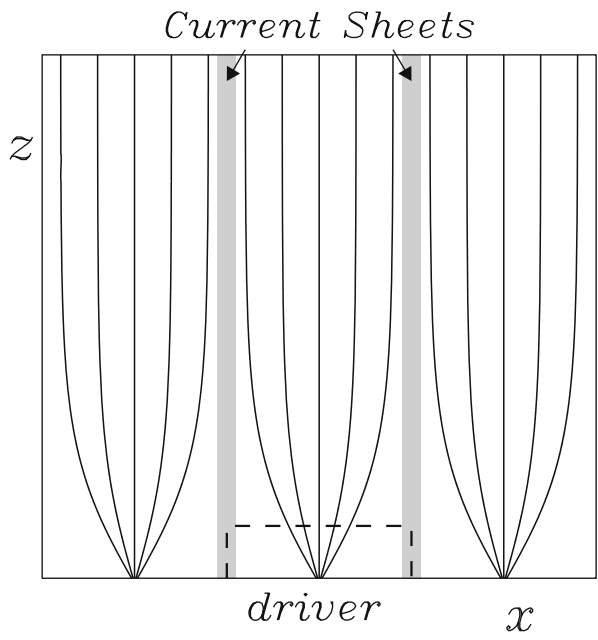


Fig. 28 A smooth motion on the photospheric boundary (only non-zero within the dashed region) causes current sheets to develop along the separatrix surfaces between the central source and the two outer sources

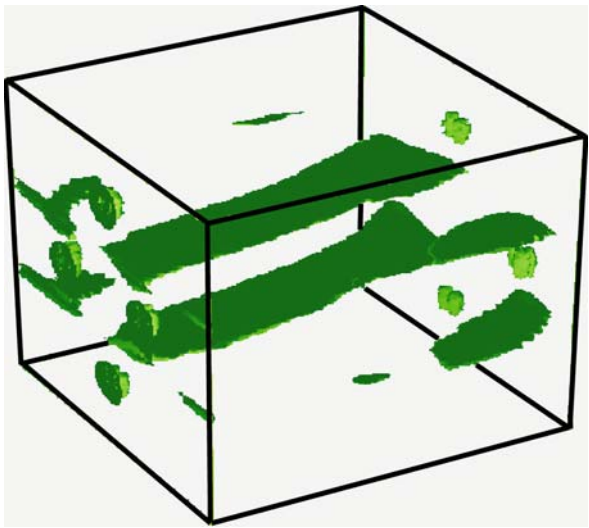


Fig. 29 Isosurfaces of current density demonstrate the formation of current sheets along the separatrix surface between the outer sources and inner magnetic sources

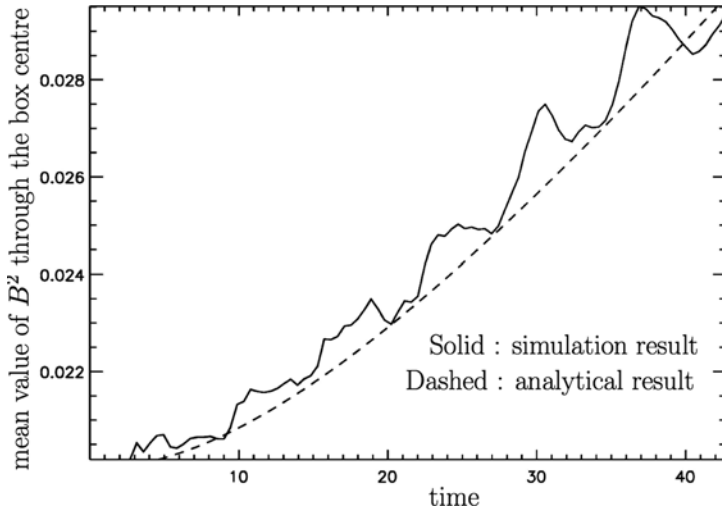


Fig. 30 The mean of B^2 (solid curve) at $x = 0$, $y = 0$, for a driving velocity of 0.03, as a function of time. An analytical approximation for B^2 demonstrates a good agreement between theory and simulations (after Mellor et al. [21])

on top of the general rise in field strength due to propagation of Alfvén waves along the field. The location of the current sheets can be calculated from the jump in B_y and this is shown in Fig. 31. Note that the value of B_y and the current sheet location increase in time.

To understand how this can be modelled by a simple expression, the fact that the field rapidly expands from the sources on the boundaries and then remains almost uniform means it can be approximated by a simple flux function, $A(x, t)$, where

$$\mathbf{B} = \left(0, B_y(A), \frac{\partial A}{\partial x} \right).$$

This ensures that $\nabla \cdot \mathbf{B}$ is always zero. Given the prescribed boundary velocity, which can be expressed in terms of the flux function as $v_p(A)$, the footpoint location can be determined and expressed in terms of B_y by integrating along a field line. Thus,

$$\delta(A) = v_p(A)t = \int_{z=0}^{z=l} \frac{B_y(A)}{B_z} dz \quad \Rightarrow \quad \delta(A) = \frac{B_y(A)l}{\partial A / \partial x}.$$

From this it is clear that the magnetic energy due to the increase in B_y will increase quadratically in time. Finally, the condition for the vanishing of the Lorentz force gives total magnetic pressure constant in space, but not in time, and so

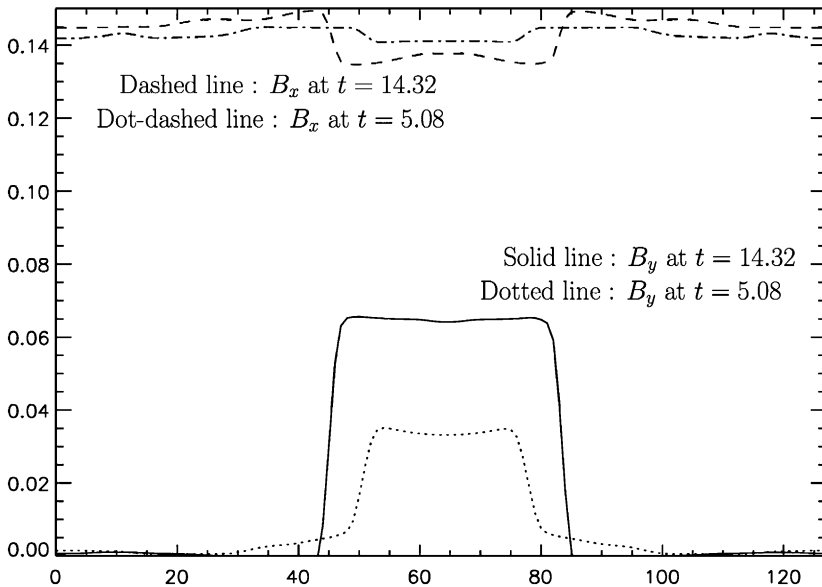


Fig. 31 Line plots through the current sheets clearly indicate the position of the current sheets and that they are moving apart in time. The positions are predicted by the simple analytical model described in the text

$$\left(\frac{\partial A}{\partial x}\right)^2 + B_y^2(A) = k^2(t).$$

This can be solved to get $A(x, t)$, B_y and so the current sheet location can be determined as a function of time. This can be compared with the numerical solution as shown in Fig. 31, and good agreement is obtained.

9.2 Fly By

Now we consider how the simple relative motion of two single sources, with an overlying uniform coronal magnetic field, can lead to the formation of current sheets and reconnection. The initial set up is shown in Figure 32. The photospheric sources do not initially connect with each other, only connecting to the overlying magnetic field. The magnetic skeleton is shown in Fig. 32. The yellow spine field line is clearly seen coming from the photospheric null point. The fan plane sweeps round, above the positive white source and forms the separatrix surface. All the field lines from the white source lie inside this surface. Similarly, the same null point, spine, fan plane and separatrix surface are seen for the negative, black source. What is clear is that the two surfaces do not intersect, indicating that the sources are disconnected initially.

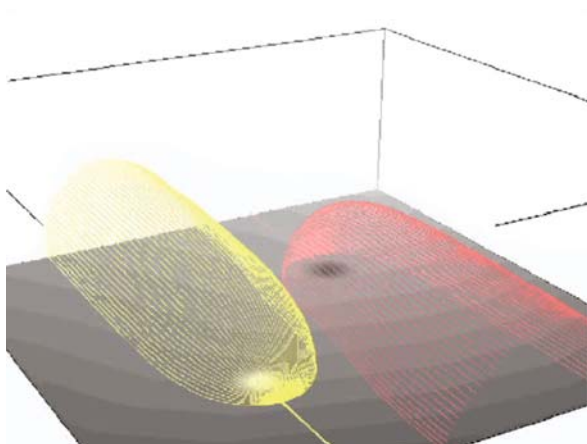


Fig. 32 The initial set up as described in the text. The *white source* and *yellow field lines* will move to the right and the *black* and *red field lines* will move to the left in response to photospheric motions

An imposed photospheric motion is imposed on each source, driving the white source to the right and the black source to the left. This causes the two surfaces to interact with each other, forming a current sheet in between (Fig. 33). Here reconnection, and heating, can occur.

What this illustrates is how a simple motion of discrete magnetic sources can produce current sheets in the corona and, when reconnection occurs in these sheets, heat the plasma.

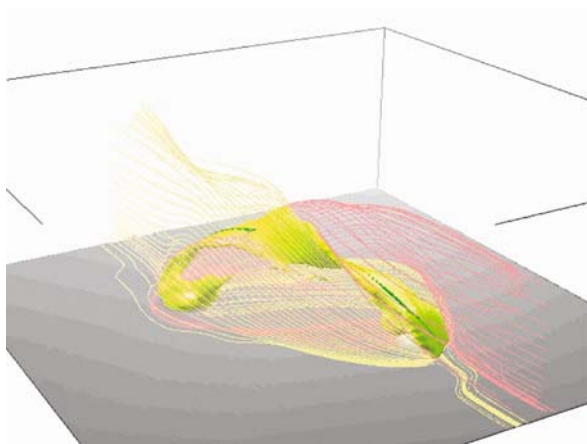


Fig. 33 Once the separatrix surfaces come into contact with each other, a current sheet forms between them

10 Plasma Response to Nanoflare Heating

The final piece of the coronal heating problem is to understand how the plasma responds to the deposition of heat. Is it possible to determine the form of the heating function by analysing the observed temperature structure along a coronal loop? Based on Yohkoh soft x-ray observations of a large coronal loop, Mackay et al. [20] calculated the plasma temperature along the loop, $T(s)$, where s is the distance along the loop. The idea was to choose the form of the coronal heating function, $H(s)$, to give the best fit with this temperature profile. The results were somewhat contradictory.

However, a steady state is unlikely and if the heating is due to nanoflares then the heating function must depend on both position and time, $H(s, t)$. Functional forms for the heating function can be picked and the response of the plasma followed using a one dimensional hydrodynamic simulation code.

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial s}(\rho v) = 0, \quad (53)$$

$$\frac{\partial}{\partial t}(\rho v) + \frac{\partial}{\partial s}(\rho v^2) = -\frac{\partial p}{\partial s} + \rho g(s) + F_v, \quad (54)$$

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial s}(pv) = (\gamma - 1) \left(\frac{\partial}{\partial s} \left(\kappa \frac{\partial T}{\partial s} \right) + H(s, t) - \rho^2 Q(T) - p \frac{\partial v}{\partial s} + \epsilon_v \right) \quad (55)$$

where the equation of state is the ideal gas law, the coronal heating function is assumed separable, F_v is the viscous force and ϵ_v is the viscous heating term and so

$$p = \rho RT,$$

$$H(s, t) = f(s)g(t),$$

$$F_v = \frac{\partial}{\partial s} \left(\frac{4}{3} \nu \frac{\partial v}{\partial s} \right),$$

$$\epsilon_v = \frac{4}{3} \nu \left(\frac{\partial v}{\partial s} \right)^2.$$

$Q(T)$ represents the optically thin radiative loss function, κ is the parallel electron heat conductivity, γ is the ratio of specific heats and $g(s)$ gives the variation of gravity along the magnetic field line.

The simulations of Klimchuk [18] assume that the plasma is heated for 250 s and this heating is repeated after 3,800 s. The spatial forms of the heating is varied in different runs using uniform heating, heating near the base, heating at the top and random heating locations. The temperature remains over 1 million K but the temperature profile does vary depending on the form of the heating. These differences

are sufficiently large to be detectable with the new EIS observations from Hinode (Fig. 34).

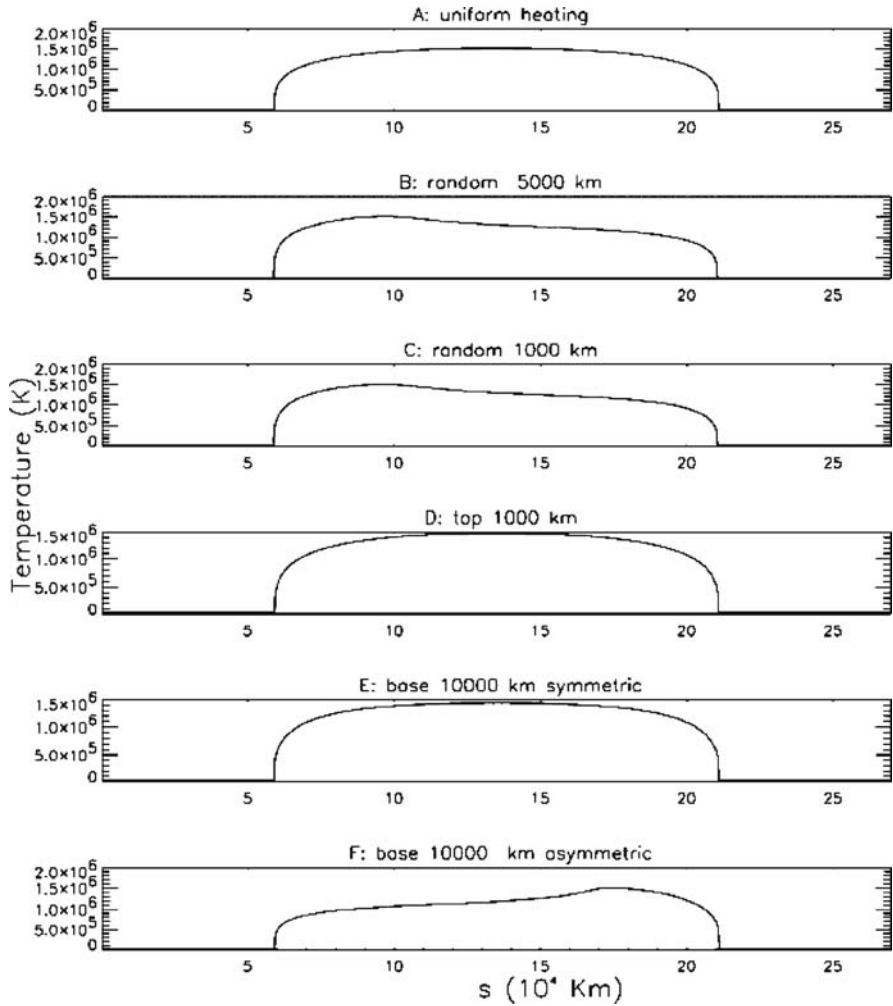


Fig. 34 The temperature profile along a coronal loop at the same time for various different forms of heating

11 Summary

To make progress on understanding the mechanism that heats the corona, the problem can be split into several distinct sections. Firstly, it is important to understand how much energy is required and how observable phenomena can provide that amount of energy. What is certainly clear is that the motion of the many discrete

photospheric sources in the quiet sun can provide ample supply of energy through the Poynting flux. This energy may be transmitted to the corona by wave motions or in a quasi static manner that results in the build up of many current sheets. It is not clear which dominates at present.

The damping of MHD waves is difficult and in a uniform plasma the lengthscales and timescales are totally inappropriate for the heating problem. However, damping is dramatically enhanced in a non-uniform plasma due to phase mixing or resonant absorption.

The formation and disruption of current sheets was introduced in 2D and the basic ideas behind reconnection explained. In 3D things are more complicated! However, an understanding of where current sheets are likely to form and where reconnection may occur can be obtained by looking at the magnetic skeleton and identifying the null points, separators, separatrix surfaces, fan planes and spines.

But what happens to the plasma once the relevant mechanism deposits the heat in the corona? Here studies have restricted attention to high resolution 1D hydrodynamic simulations, confined to individual field lines. Heat is deposited at various locations and the temperature and density of the plasma determined. This can be compared with observations with a view to determining the heating profile. Finally, the heating mechanism must be able to reproduce the same heating profile.

We are not yet able to say exactly how the corona is heated but it is due to the coronal magnetic field. Progress will be made over the coming years with the amazing high resolutions observations from the recently launched Hinode and STEREO missions and the SDO mission due for launch in 2010. Exciting times are ahead.

Acknowledgement The SOHO/LASCO data used here are produced by a consortium of the Naval Research Laboratory (USA), Max-Planck-Institut fuer Aeronomie (Germany), Laboratoire d'Astronomie (France) and the University of Birmingham (UK). SOHO is a project of international cooperation between ESA and NASA. The Transition Region and Coronal Explorer, TRACE, is a mission of the Stanford-Lockheed Institute for Space Research, and part of the NASA Small Explorer programme.

References

1. Archontis, V., et al.: Emergence of magnetic flux from the convection zone into the corona. *Astron and Astrophys.* **426**, 1047–1063 (2004). 149
2. Archontis, V., et al.: Emergence and interaction of twisted flux tubes in the sun, *Astron. and Astrophys.* **466**, 367–376 (2007). 148
3. Aschwanden, M. J.: Do EUV nanoflares account for coronal heating? *Solar Phys.* **190**, 233–247 (1999). 118
4. Birn, J., Priest, E. R.: *Reconnection of Magnetic Fields: MHD and Collisionless Theory and Observations*, Cambridge University Press, Cambridge (2007) 134, 144
5. Biskamp, D.: Magnetic reconnection via current sheets. *Phys Fluids* . **29**, 1520–1531 (1986). 135
6. Browning, P. K., et al.: Coronal heating by magnetic reconnection in loops with zero net current. *Astron. Astrophys.* **485**, 837 (2008). 147
7. Close, R., et al.: Coronal flux recycling times. *Solar Phys.* **231**, 45–70 (2005). 121, 151
8. Cowling, T. G.: *Solar Electrodynamics*, The Sun edited by Gerard P. Kuiper, University of Chicago Press, Chicago (1953). 134

9. Dungey, J. W.: The motion of magnetic fields. *Mon Not R Astron Soc*, **113**, 679 (1953). 135
10. Dungey, J. W.: The Neutral Point Discharge Theory of Solar Flares. a Reply to Cowling's Criticism, *Electromagnetic Phenomena in Cosmical Physics*, Proceedings from IAU Symposium no. 6. Edited by Bo Lehnert. International Astronomical Union. Symposium no. 6, Cambridge University Press, 135 (1958). 135
11. Edwin P., Roberts, B.: Wave propagation in a magnetic cylinder, *Solar Phys.* **88**, 179 (1983). 127
12. Furth, H. P., et al.: Finite-Resistivity Instabilities of a Sheet Pinch. *Phys. Fluids*, **6**, 459–484 (1963). 135
13. Giovanelli, R. G.: Magnetic and electric phenomena in the sun's atmosphere associated with sunspots. *MNRAS*, **107**, 338 (1947). 134
14. Hagenaar, H. J.: Ephemeral regions on a sequence of full-disk michelson doppler imager magnetograms. *Astrophys J*, **555**, 448–461 (2001). 151
15. Heyvaerts, J., Priest, E. R.: Coronal heating by phase-mixed shear Alfvén waves. *Astron. Astrophys.* **117**, 220–234 (1983). 130, 131
16. Hood, A. W., et al.: Heating the corona by nanoflares: Simulations of energy release triggered by a kink instability. *Astron. Astrophys.*, in press (2009). 10-1051/0004-6361/200912285. 147
17. Hudson, H. S., Solar flares, microflares, nanoflares, and coronal heating. *Solar Physics*, **133**, 357–369. (1991). 118
18. Klimchuk, J. A.: How do we solve the coronal heating problem? Proceedings of the SOHO 15 Workshop - Coronal Heating. 6–9 September 2004, St. Andrews, Scotland, UK (ESA SP-575). Editors: R. W. Walsh, J. Ireland, D. Danesy, B. Fleck. Paris: European Space Agency (2004). 115, 116, 156
19. Krucker, S., Benz, A. O.: Are heating events in the quiet solar corona small flares? multiwavelength observations of individual events. *Solar Physics*, **191**, 341–358 (2000). 118
20. Mackay, D. H., et al.: How accurately can we determine the coronal heating mechanism in the large-scale solar corona? *Solar Physics*, **193**, 93–116 (2000). 120, 156
21. Mellor, C., et al.: Numerical simulations of the flux tube tectonics model for coronal heating. *Solar Phys.* **227**, 39–60 (2005). 151, 153
22. Parker, E. N.: Sweet's mechanism for merging magnetic fields in conducting fluids. *JGR*, **62**, 509–520 (1957). 135, 142
23. Parker, E. N.: Magnetic reconnection and magnetic activity, *Magnetic reconnection in space and laboratory plasmas*, Proceedings of the Chapman Conference on Magnetic Reconnection, Los Alamos, NM, October 3–7, 1983 (A85-43501 21–46). Washington, DC, American Geophysical Union, 1984, pp. 32–38; Discussion, p. 38. 149
24. Parnell, C. E., Jupp, P. E.: Statistical analysis of the energy distribution of nanoflares in the quiet sun, *Astrophys J*, **529**, 554–569, 2000. 118
25. Petschek, H. E.: Magnetic Field Annihilation, *The Physics of Solar Flares*, Proceedings of the AAS-NASA Symposium held 28–30 October, 1963 at the Goddard Space Flight Center, Greenbelt, MD. Edited by Wilmot N. Hess. Washington, DC: National Aeronautics and Space Administration, Science and Technical Information Division, 1964., p.425 (1964). 135, 143
26. Priest, E. R., et al.: A flux-tube tectonics model for solar coronal heating driven by the magnetic carpet. *Astrophys J*, **576**, 533–551 (2002). 151
27. Priest, E. R., et al.: Nature of the heating mechanism for the diffuse solar corona. *Nature*, **393**, 545 (1998). 120
28. Priest, E. R., Forbes, T. G.: New models for fast steady state magnetic reconnection. *JGR*, **91**, 5579–5588 (1986). 144
29. Reale, F.: More on the determination of the coronal heating function from Yohkoh data. *Astrophys J*, **580**, 566–573 (2002). 120
30. Schindler, K., et al.: General magnetic reconnection, parallel electric fields, and helicity. *JGR*, **93**, 5547–5557 (1988). 135
31. Sweet, P. A.: The Neutral Point Theory of Solar Flares, *Electromagnetic Phenomena in Cosmical Physics*, Proceedings from IAU Symposium no. 6. Edited by Bo Lehnert. Cambridge University Press, 123 (1958). 135, 142

32. Walsh, R. W., et al.: Proceedings of the SOHO 15 Workshop – Coronal Heating, 6–9 September 2004, St. Andrews, Scotland, UK (ESA SP-575). Editors: R. W. Walsh, J. Ireland, D. Danesy, B. Fleck. Paris: European Space Agency (2004). 111
33. Withbroe, G. L., Noyes, R. W. Mass and energy flow in the solar chromosphere and corona, *Annu Rev Astron Astrophys.* **15**, 363–387 (1977). 115

Flows in Molecular Media

David Flower

Abstract We consider the hydrodynamic properties of shock waves that are propagating under the very low-density conditions of interstellar clouds. The macroscopic and microscopic aspects of the flows are discussed, including the feedback of the chemistry and of radiative losses on the structure of the shock waves. The diagnostic potential of the rovibrational spectrum of H_2 , excited in outflows that are associated with low-mass star formation, is given particular attention. Most of the analysis is restricted to steady-state, one-dimensional structures, but some consideration is given to the temporal evolution of shock waves, which it is necessary to consider in the context of outflow sources.

1 Introduction

The generation of shock waves is an inevitable consequence of the impact of a protostellar jet on the surrounding ambient material – dust and molecular gas – from which the protostar formed. The speeds of the jets, as deduced from the widths of spectral lines, are strongly supersonic, and a shock wave will propagate into the molecular gas. In these lectures, we shall be concerned with techniques for modelling the physical and chemical structure of such shock waves and for predicting their spectroscopic signatures. The ultimate objective is to relate the intensities and profiles of emission lines produced by the shock waves, which are observable quantities, to the dynamical characteristics of the jets and thence to the process of star formation.

In practice, not one but two shock waves are produced by the interaction of a jet with the ambient material. In the rest frame of the ambient gas, the jet is seen to impact at its outflow speed. However, if a Galilean transformation is made into the rest frame of the jet, the ambient gas appears to impact the jet with the same speed, generating a ‘reverse’ shock wave. The appropriate reference frame in which to view the collision is the centre of mass frame; but if the mass of the ambient molecular cloud is much greater than that of the jet – a condition that is very likely to be

D. Flower (✉)

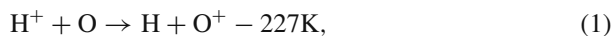
Physics Department, University of Durham, DH1 3LE, UK, david.flower@durham.ac.uk

satisfied – the centre of mass frame is almost coincident with that of the ambient medium. The consequences of the heating and compression of the ambient molecular gas and grains by a shock wave will be considered below.

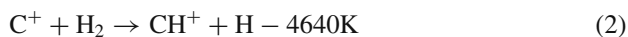
In early work on hydromagnetic shock wave in the interstellar medium [1], the ionized and neutral fluids were assumed to be fully coupled. In this case, the action of the magnetic field on the ionized gas is transmitted simultaneously to the neutral gas. If the degree of ionization is low, decoupling of the flows of the ionized and the neutral fluids can occur. The magnetic field may be considered to be ‘frozen’ in the ionized fluid, which is electrically conducting. When the magnetic field has a component perpendicular to the flow direction, the compression of the ionized gas is accompanied by the compression of the magnetic field, but the effects of this compression are felt by the neutrals, via collisions with the ions, only after a delay. Differences develop in the flow speeds of the charged and neutral fluids, a phenomenon which is termed ‘ion-neutral drift’, from the viewpoint of the fluids, or ‘ambipolar diffusion’, from the viewpoint of the magnetic field, which diffuses through the neutrals, along with the ions.

The effects which are associated with the partial decoupling of the charged and the neutral fluid were investigated subsequently by Mullan [2]. As a consequence of ambipolar diffusion, the neutral fluid is compressed and heated *in advance* of the shock discontinuity. This heating mechanism was incorporated, in an approximate manner, into the shock model of Hollenbach and McKee [3], but it was not until the work of Draine [4] that a quantitative model of such magnetohydrodynamical (MHD) shock waves became available. The model of Draine, like its predecessors, assumed that steady state had been attained; only recently have time-dependent models (which describe the temporal evolution and the spatial structure of shock waves) found their way into the literature [5–8].

In the ambient medium, the kinetic temperature of the gas is low ($T \lesssim 100\text{K}$) and even slightly endothermic processes, such as the important charge exchange reaction



are effectively inhibited. Consequently, the gas-phase chemistry of the ambient medium is dominated by reactions with no endothermicity and no reaction barrier. However, as a result of shock wave heating, chemical reactions which are endothermic, or which present reaction barriers attaining a few tenths of an electron volt, become significant. The requisite energy derives from the motion of the shock wave (i.e. from the ‘piston’ which drives it) and is transmitted to the gas in the form of thermal energy and, in MHD shocks, through the ion-neutral drift. Then, reactions which are unimportant in the ambient medium, such as



and



(which has a barrier of 2980K), assume significance.

Whilst the shock structure is important in determining the chemistry of the medium, the chemistry is important to the structure of the shock wave. Chemical reactions affect directly the degree of ionization of the medium and hence the interaction with the magnetic field. Furthermore, chemical reactions influence the abundances of the atomic and molecular species, which cool the gas, principally through the collisional excitation of rovibrational and fine structure transitions. The dynamical and chemical conservation equations are interdependent and should be solved in parallel. So too should the equations for the population densities of the rovibrational levels of the H_2 molecule, which is the main coolant of shocked molecular gas. There is a delay in the response of the level populations to changes in the density and the kinetic temperature of the gas; this delay is taken into account only when the equations for the level populations are integrated *in parallel* with the MHD and chemical conservation equations, to which we now turn.

2 The MHD Conservation Equations

The physical quantities with which we shall be concerned are the numbers of particles, their mass, momentum and energy. It has already been mentioned that the ionized and neutral fluids can develop different flow velocities; their temperatures may also differ. Furthermore, the temperatures of the ions and the electrons at any given point in the flow may not be equal. The development of differences in the velocities of the positively and negatively charged fluids is resisted by large electrical forces; these ensure that the velocities and the number densities of the positive and negative particles are effectively equal everywhere.

2.1 The Conservation Equations in One Dimension

In the following analysis, it will be assumed that a stationary state has been attained, in which case $\partial/\partial t = 0$. The total time derivative may then be written as

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla = \mathbf{u} \cdot \nabla, \quad (4)$$

where \mathbf{u} is the flow velocity and $\nabla = \hat{\mathbf{i}}\partial/\partial x + \hat{\mathbf{j}}\partial/\partial y + \hat{\mathbf{k}}\partial/\partial z$ is the gradient operator. If the flow is plane-parallel in the z -direction, then $\nabla = \hat{\mathbf{k}}\partial/\partial z$ and

$$\frac{d}{dt} = u \frac{d}{dz}. \quad (5)$$

By making these assumptions, we exclude the possibility of studying rigorously both the temporal evolution of the shock wave and its structure in more than one spatial dimension; but we simplify considerably the numerical aspects of the problem, which reduces to solving coupled *ordinary* (rather than partial) differential equations.

Subject to the assumptions above, the equation for the number density of neutral particles states that

$$\frac{d}{dz}(\rho_n u_n / \mu_n) = \mathcal{N}_n \quad (6)$$

where ρ_n is the mass density of the neutrals at the point z , μ_n their mean molecular weight and u_n is their flow speed in the z -direction. The ‘source’ term on the right-hand side of Eq. (6) is the rate of creation (or destruction, if negative) of neutral particles per unit volume through recombination and ionization processes. For example, the formation of molecular from atomic hydrogen results in a net reduction in the number of neutral particles, and so $\mathcal{N}_n < 0$.

An analogous equation holds for the positively (and the negatively) charged fluid:

$$\frac{d}{dz}(\rho_+ u_+ / \mu_+) = \mathcal{N}_+. \quad (7)$$

The positively charged fluid, denoted by ‘+’, comprises the positive ions and positively charged grains. In general, $\mathcal{N}_n \neq -\mathcal{N}_+$: dissociative recombination processes, for example,



result in the destruction of *one* ion but create *two* neutrals.

The equation of mass conservation for the neutrals may be written as

$$\frac{d}{dz}(\rho_n u_n) = \mathcal{S}_n, \quad (9)$$

where \mathcal{S}_n denotes the rate per unit volume at which neutral mass is created or destroyed. The corresponding equation for the positively charged fluid is

$$\frac{d}{dz}(\rho_+ u_+) = \mathcal{S}_+. \quad (10)$$

Because neutral mass may be created only through the destruction of charged mass, by recombination of positively and negatively charged particles, $\mathcal{S}_+ + \mathcal{S}_- = -\mathcal{S}_n$, where the subscript ‘-’ refers to the negatively charged fluid, which comprises the electrons, negative ions and the negatively charged grains. When the negatively charged particle in the recombination reaction is an electron, which is the most frequent case, \mathcal{S}_- is negligible.

Momentum has to be conserved also. For the neutral fluid, the equation of momentum conservation takes the form

$$\frac{d}{dz} \left(\rho_n u_n^2 + \frac{\rho_n k_B T_n}{\mu_n} \right) = \mathcal{A}_n, \quad (11)$$

where T_n is the temperature of the neutral fluid at the point z , and \mathcal{A}_n determines the rate at which momentum is being gained or lost by unit volume of the neutral gas; k_B is Boltzmann’s constant. Momentum transfer occurs principally in collisions of the neutrals with ions and with charged grains. As the positively and the negatively

charged particles have the same number density and the same flow speed, their combined equation of momentum conservation may be written as

$$\frac{d}{dz} \left[(\rho_+ + \rho_-)u_+^2 + \frac{\rho_+ k_B (T_+ + T_-)}{\mu_+} + \frac{B^2}{8\pi} \right] = -\mathcal{A}_n, \quad (12)$$

where we have used the fact that $\rho_+/\mu_+ = \rho_-/\mu_-$, owing to the overall charge neutrality. In Eq. (12), B is the component of the magnetic field perpendicular to the z -direction, and $B^2/(8\pi)$ is the magnetic pressure term. The magnetic field acts directly on the charged fluid, which communicates its influence on the neutral fluid through collision processes. If the magnetic field is assumed to be ‘frozen’ into the charged fluid, the condition

$$Bu_+ = B_0 u_s \quad (13)$$

is satisfied, where B_0 is the value of the magnetic field strength upstream of the shock wave, in the ‘preshock’ gas, and u_s is the shock speed. In the reference frame of the shock wave, in which the conservation equations are formulated, the initial (‘upstream’ or ‘preshock’) flow speeds of the neutral and charged fluids are both equal to the shock speed. Taking the flow to be in the positive z -direction implies that the shock wave is propagating in the negative z -direction. We recall that the discussion here applies to planar shock waves. Oblique MHD shock waves have been considered by Wardle and Draine [9].

The condition of energy conservation, applied to the neutral fluid, yields

$$\frac{d}{dz} \left[\frac{\rho_n u_n^3}{2} + \frac{5\rho_n u_n k_B T_n}{2\mu_n} + \frac{\rho_n u_n U_n}{\mu_n} \right] = \mathcal{B}_n, \quad (14)$$

where U_n denotes the mean internal energy per neutral particle, and \mathcal{B}_n is the rate of gain (or loss, if negative) of energy per unit volume of the neutral fluid. The internal energy consists essentially of the rovibrational excitation energy of the H_2 molecule, whose excited state population densities can become appreciable as the temperature and the density increase, owing to the passage of the shock wave.

In order to discriminate between the temperatures of the positively and negatively charged fluids, it is necessary to separate their energy conservation relations. Draine [10] derived these equations for the case of an MHD shock wave. However, the difference between T_+ and T_- is difficult to evaluate accurately. In general, it has only a minor influence on the dynamical and chemical structure of the shock wave. Accordingly, we consider the combined equation of energy conservation of the (positively and negatively) charged fluid, which takes the form

$$\begin{aligned} \frac{d}{dz} \left[\frac{(\rho_+ + \rho_-)u_+^3}{2} + \frac{5\rho_+ u_+ k_B (T_+ + T_-)}{2\mu_+} + \frac{u_+ B^2}{4\pi} \right] \\ = \mathcal{B}_+ + \mathcal{B}_-, \end{aligned} \quad (15)$$

where $(\mathcal{B}_+ + \mathcal{B}_-)$ is the rate of energy gain per unit volume of the charged fluid.

2.2 The Role of the Magnetic Field

We have seen that the selective action of the magnetic field on the charged particles can give rise to different flow speeds for the charged and the neutral fluids. Consider first the neutral fluid. If a sonic point occurs in the flow, at the point where

$$u_n^2 = \frac{5k_B T_n}{3\mu_n} \equiv c_s^2 \quad (16)$$

and c_s is the adiabatic sound speed, then the neutral flow becomes discontinuous and a shock occurs. In fact, this shock ‘discontinuity’ has a finite thickness, owing to viscous forces, which is of the same order as the length scale which characterizes elastic collisions between the neutral particles.

It is possible to integrate the conservation equations through the shock ‘discontinuity’ by introducing *artificial viscosity* terms [11]. Providing the transition from the pre- to the postshock gas occurs adiabatically, i.e. processes of energy transfer (notably radiative losses) to the surrounding environment are negligible, integration of the conservation equations through the ‘discontinuity’ automatically satisfies the Rankine–Hugoniot relations. The latter specify the compression ratio and the temperature ratio across the shock front in the limit in which the shock front may be treated as a discontinuity; the Rankine–Hugoniot relations are derived in Sect. 3.1 below. We note that, in the case of a shock discontinuity, ‘adiabatically’ is not synonymous with ‘isentropically’, i.e. with constant entropy. Within the ‘discontinuity’, *irreversible* processes (viscous heating and thermal conduction) assume importance, and the entropy increases from the preshock to the postshock gas – even though this transition takes place adiabatically. Entropy may be considered to be a measure of the degree of disorder in a medium, and the gas undergoes a transition, across the discontinuity, from an ordered flow (at the shock speed, in the frame of the shock wave) of the low-temperature preshock gas to a more slowly flowing, high-temperature postshock gas. In essence, the kinetic energy associated with the bulk flow velocity is partially converted into the kinetic energy associated with random thermal motions.

Similarly, a discontinuity occurs in the flow of the charged fluid at the point at which

$$u_+^2 = \frac{5k_B(T_+ + T_-)}{3(\mu_+ + \mu_-)} + \frac{B^2}{4\pi(\rho_+ + \rho_-)} \equiv c_m^2, \quad (17)$$

where c_m is the magnetosonic speed in the charged fluid. As $u_+ \leq u_s$ everywhere, in the reference frame of the shock wave, a discontinuity in the flow of the charged fluid cannot occur if

$$u_s < c_m,$$

a condition which is satisfied in magnetically dominated flows. It follows that flows can exist which are discontinuous in only the neutral flow variables.

In practice, modest values of the magnetic field strength are sufficient to suppress the discontinuity in the flow variables of the charged fluid when the degree of ionization of the medium is low. For example, if $u_s = 10 \text{ km s}^{-1}$ and the charged mass density $\rho_+ + \rho_- = 2 \times 10^{-25} \text{ g cm}^{-3}$ (corresponding to C^+ ions in a medium with a total particle density of about 50 cm^{-3}), $B \approx 1 \text{ } \mu\text{G}$ is all that is required.

The region upstream of the discontinuity, in which the charged fluid has been compressed along with the magnetic field, has been termed the ‘magnetic precursor’ or ‘acceleration zone’; the width of the precursor increases with the magnetic field strength. As the shock evolves, the discontinuity in the neutral flow moves progressively downstream and weakens, until finally the discontinuity is suppressed. Thus, for sufficiently large field strengths, the shock wave evolves from ‘jump’ or J-type, to J-type with a magnetic precursor, to ‘continuous’ or C-type [6]. To describe this evolution rigorously, a time-dependent MHD code must be used; but the evolution can be simulated by means of calculations which are not explicitly time-dependent, as will be seen below. The mass density of the charged fluid, $\rho_+ + \rho_-$, is a factor determining the magnetosonic speed and hence the width of the acceleration zone. The endothermic reaction (2) of C^+ ions with H_2 molecules can be initiated by ion–neutral drift in this zone, yielding CH^+ . The rapid exothermic reactions



and



then lead to the formation of CH_2^+ and CH_3^+ . The molecular ions are destroyed by dissociative recombination reactions, such as



which are believed to be rapid at the relevant temperatures. The net result is the neutralization of an important part of the ionized component of the gas. This process, of partial neutralization, can occur on a distance scale which is comparable with the dimensions of the acceleration zone, resulting in a significant enhancement of the width of this zone through the increase in the magnetic field term in Eq. (17).

In diffuse clouds, the situation is rendered more complicated by photoionization processes, notably



which restore ions to the medium and which also occur over a characteristic distance scale which is comparable with the width of the acceleration zone. However, before

considering further the chemistry in shock waves, in both diffuse and dense clouds, it is appropriate to consider in more detail the ‘source terms’ (\mathcal{N} , \mathcal{S} , \mathcal{A} , \mathcal{B}) appearing in the MHD conservation equations.

2.3 The Source Terms

The ‘source terms’ which appear on the right-hand sides of the MHD conservation equations in Sect. 2.1 contain the micro-physics of the problem. These terms describe the interactions between the particles of the medium, including the grains, and the ways in which these interactions modify the number and mass densities, momentum and energy of the charged and neutral fluids. The principal terms will be introduced below; their hierarchy of importance depends on the context of the problem being considered. For a detailed discussion of the source terms, the reader may consult the paper of Draine [10].

Let us denote the net rate at which a particular atomic or molecular species, α , is produced per unit volume by \mathcal{C}_α ; a net destruction rate corresponds to $\mathcal{C}_\alpha < 0$. The total number of neutral particles produced per unit volume and time is

$$\mathcal{N}_n = \sum_{\alpha_n} \mathcal{C}_{\alpha_n}, \quad (22)$$

where the subscript ‘n’ identifies the species as being neutral. Similarly, for the positive ions

$$\mathcal{N}_+ = \sum_{\alpha_+} \mathcal{C}_{\alpha_+}. \quad (23)$$

As already noted, $\mathcal{N}_+ \neq -\mathcal{N}_n$, in general.

Creation of neutral mass proceeds at a rate per unit volume:

$$\mathcal{S}_n = \sum_{\alpha_n} \mathcal{C}_{\alpha_n} m_{\alpha_n}, \quad (24)$$

where m_{α_n} is the mass of the neutral species α_n ; the corresponding expression for the positive ions is

$$\mathcal{S}_+ = \sum_{\alpha_+} \mathcal{C}_{\alpha_+} m_{\alpha_+}. \quad (25)$$

In this case, $\mathcal{S}_+ + \mathcal{S}_- = -\mathcal{S}_n$ must be satisfied.

Momentum is transferred between the charged and the neutral fluids through ion–neutral reactions and elastic scattering. Denoting a particular ion–neutral reaction by β , we may write

$$\mathcal{C}_\alpha = \sum_{\beta} \mathcal{C}_{\alpha\beta} \quad (26)$$

and the associated rate of momentum transfer from the charged to the neutral fluid is

$$\mathcal{A}_n^{(i)} = \sum_{\alpha_n \beta} \left[\sum_{\mathcal{C}_{\alpha_n \beta} > 0} \mathcal{C}_{\alpha_n \beta} m_{\alpha_n} u_{\beta}(\text{CM}) + \sum_{\mathcal{C}_{\alpha_n \beta} < 0} \mathcal{C}_{\alpha_n \beta} m_{\alpha_n} u_n \right], \quad (27)$$

where the centre-of-mass collision velocity is given by

$$u_{\beta}(\text{CM}) = \frac{m_i u_{\pm} + m_n u_n}{m_i + m_n}, \quad (28)$$

where m_i is the mass of a (positive or negative) ion and m_n is the mass of the neutral. Equation (27) expresses the fact that, when a neutral is a product and hence $\mathcal{C}_{\alpha_n \beta} > 0$, it is created at the centre-of-mass velocity of the reaction β . On the other hand, when a neutral is a reactant ($\mathcal{C}_{\alpha_n \beta} < 0$), it is removed whilst moving with the velocity of the neutral fluid, u_n .

Osterbrock [12] derived an expression for the cross section for momentum transfer in a collision between a charged and a neutral particle from considerations of the long-range charge-induced dipole interaction between the colliding pair; the expression which Osterbrock derived is

$$\sigma_{\text{in}} = 2.41\pi \left[\frac{e^2 \alpha_n}{m_{\text{in}} v_{\text{in}}^2} \right]^{\frac{1}{2}}, \quad (29)$$

where e is the electron charge, α_n is the polarizability of the neutral, $m_{\text{in}} = m_i m_n / (m_i + m_n)$ is the reduced mass of the ion–neutral pair and v_{in} is the relative collision speed. Expression (29) exceeds by 20% the ‘Langevin’ cross section which is often used to estimate the rates of ion–neutral chemical reactions.

Equation (29) has been shown to be valid at low collision speeds but to underestimate the momentum transfer at high collision speeds [13]. The polarizabilities of the principal constituents of the neutral fluid are: atomic hydrogen, $\alpha_{\text{H}} = 4.5 a_0^3$; molecular hydrogen, $\alpha_{\text{H}_2} = 5.2 a_0^3$; and helium, $\alpha_{\text{He}} = 1.4 a_0^3$, where a_0 is the Bohr radius, the atomic unit of length; the polarizabilities of H and H_2 are similar in magnitude and substantially larger than that of He.

The rate of transfer of momentum to the neutral fluid, owing to elastic scattering on the ions, is given by

$$\mathcal{A}_n^{(ii)} = \frac{\rho_n \rho_i}{\mu_n + \mu_i} \langle \sigma v \rangle_{\text{in}} (u_i - u_n), \quad (30)$$

where

$$\langle \sigma v \rangle_{\text{in}} = 2.41\pi \left(\frac{e^2 \alpha_n}{\mu_{\text{in}}} \right)^{\frac{1}{2}} \quad (31)$$

is the corresponding rate coefficient; $\mu_{in} = \mu_i \mu_n / (\mu_i + \mu_n)$ is the reduced mass, evaluated using the mean molecular weights of the ionized and neutral fluids.

Momentum transfer between the neutral gas and the charged grains is important in dense clouds, where the degree of ionization of the gas is low. In this case, the cross section may be taken approximately equal to the geometrical cross section of the grain, πa_g^2 , where a_g is the grain radius. (There is a correction to the geometrical cross section, significant at low collision speeds, arising from the polarization of the neutral by the charged grain [14]). As the collision speed is, to a good approximation, equal to the ion–neutral drift speed, $|u_i - u_n|$, and $\mu_g \gg \mu_n$, the rate of momentum transfer between the neutral gas and the charged grains (in dense clouds, most of the grains have a single negative charge) is

$$\mathcal{A}_n^{(iii)} = \rho_n n_g \pi a_g^2 |u_i - u_n| (u_i - u_n). \quad (32)$$

The total rate of momentum transfer is $\mathcal{A}_n = \mathcal{A}_n^{(i)} + \mathcal{A}_n^{(ii)} + \mathcal{A}_n^{(iii)}$.

Various physical processes lead to energy exchange between the charged and the neutral fluids. Chemical reactions are responsible for kinetic energy transfer from the charged to the neutral fluid at a rate per unit volume:

$$\mathcal{B}_n^{(i)} = \sum_{\alpha_n \beta} \left[\sum_{\mathcal{C}_{\alpha_n \beta} > 0} \mathcal{C}_{\alpha_n \beta} \frac{1}{2} m_{\alpha_n} u_{\beta}^2 (\text{CM}) + \sum_{\mathcal{C}_{\alpha_n \beta} < 0} \mathcal{C}_{\alpha_n \beta} \frac{1}{2} m_{\alpha_n} u_n^2 \right]. \quad (33)$$

An analogous expression holds for transfer of kinetic energy from the neutral to the charged fluid; the kinetic energy of the electrons may be neglected, in comparison with that of the ions.

When considering heat transfer between fluids, a distinction has to be made once again between formation ($\mathcal{C}_{\alpha_n \beta} > 0$) and destruction ($\mathcal{C}_{\alpha_n \beta} < 0$) processes. When an ion and an electron at temperatures T_+ and T_- , respectively, dissociatively recombine to form two neutrals, as in the reaction



an amount of heat $3k_B(T_+ + T_-)/2$ is transferred to the neutral fluid. On the other hand, an amount of heat $3k_B T_n/2$ is lost by the neutral fluid through photoionization as in



The net rate of thermal energy transfer to the neutral fluid is

$$\mathcal{B}_n^{(ii)} = \sum_{\alpha_n \beta} \left[\sum_{\mathcal{C}_{\alpha_n \beta} > 0} \mathcal{C}_{\alpha_n \beta} \frac{3}{2} k_B \frac{T_+ + T_-}{2} + \sum_{\mathcal{C}_{\alpha_n \beta} < 0} \mathcal{C}_{\alpha_n \beta} \frac{3}{2} k_B \frac{T_n}{2} \right]. \quad (36)$$

Dissociative recombination and photoionization are the most rapid and important processes determining the degree of ionization in shocks propagating in diffuse interstellar clouds. In the interiors of dense clouds, the degree of ionization is lower, owing to the absence of ionizing photons, which are absorbed and scattered by dust in the outer layers; cosmic ray ionization takes over but is a slow process.

When photoionization does occur, the corresponding heating rate is given by

$$\mathcal{B}_{-}^{(\text{iii})} = \sum_{\alpha} n_{\alpha} \int_{\nu_{\alpha}}^{\nu_{\text{H}}} \frac{4\pi J_{\nu}}{h\nu} a_{\nu}(\alpha)(h\nu - h\nu_{\alpha}) d\nu, \quad (37)$$

where J_{ν} is the mean radiation intensity at frequency ν , $a_{\nu}(\alpha)$ is the frequency-dependent photoionization cross section of species α and ν_{α} is the photoionization threshold frequency. The integral extends to the Lyman limit in atomic hydrogen, $h\nu = 13.598$ eV; photons of higher energy tend to be absorbed (by atomic hydrogen) in the immediate vicinities of the sources of the ultraviolet radiation.

Chemical reactions also affect the thermal balance of the medium by virtue of their energy defects, ΔE . The corresponding rate of heating of the neutral fluid is

$$\mathcal{B}_{\text{n}}^{(\text{iv})} = \sum_{\alpha\beta} \sum_{C_{\alpha\beta} > 0} C_{\alpha\beta} \frac{M_{\beta} - m_{\alpha_n}}{M_{\beta}} \Delta E_{\beta}, \quad (38)$$

where M_{β} is the total mass of the products of reaction β , including m_{α_n} . The factor $(M_{\beta} - m_{\alpha_n})/M_{\beta}$ determines the partition of energy amongst the reaction products, with the lighter products carrying off more of the energy, ΔE_{β} , released in the reaction.

Elastic scattering of the neutrals on the ions results in the exchange of energy between the fluids. The neutral fluid is heated through this process at a rate given by

$$\mathcal{B}_{\text{n}}^{(\text{v})} = \frac{\rho_{\text{n}}\rho_{\text{i}}}{\mu_{\text{n}}\mu_{\text{i}}} \langle \sigma v \rangle_{\text{in}} \frac{2\mu_{\text{n}}\mu_{\text{i}}}{(\mu_{\text{n}} + \mu_{\text{i}})^2} \left[\frac{3}{2} k_{\text{B}}(T_{\text{i}} - T_{\text{n}}) + \frac{1}{2}(u_{\text{i}} - u_{\text{n}})(\mu_{\text{i}}u_{\text{i}} + \mu_{\text{n}}u_{\text{n}}) \right], \quad (39)$$

where the rate coefficient for ion–neutral elastic scattering, $\langle \sigma v \rangle_{\text{in}}$, is given by Eq. (31). Inspection of Eq. (39) shows that $\mathcal{B}_{\text{i}}^{(\text{v})} = -\mathcal{B}_{\text{n}}^{(\text{v})}$. The corresponding rate of energy transfer from the charged grains to the neutral fluid is

$$\mathcal{B}_{\text{n}}^{(\text{vi})} = \rho_{\text{n}}n_{\text{g}}\pi a_{\text{g}}^2 |u_{\text{i}} - u_{\text{n}}|(u_{\text{i}} - u_{\text{n}})u_{\text{i}}, \quad (40)$$

where we assume $\mu_{\text{g}} \gg \mu_{\text{n}}$.

Collisional excitation, followed by radiative decay at an optically thin wavelength, is an important source of energy loss from the gas and must be taken into account. Particularly significant is the collisional excitation of rovibrational transitions in molecules and of fine structure transitions in atoms and ions, discussed in [15]. The rates of cooling processes are proportional to the number densities of the coolants, which depend in turn on the chemical reactions occurring within the shocked gas.

In shocks which give rise to appreciable collisional dissociation of molecular hydrogen, the reformation of H_2 in the cooling flow of the shock wave represents a significant heating process. Molecular hydrogen forms on grains, and the H_2 molecules are returned to the gas phase with a finite amount of translational energy; this is subsequently converted to heat through elastic collisions with the other constituents of the gas. The rate of heating is proportional to the rate of formation of H_2 and to the translational energy of the molecule as it leaves the grain. The total energy released when a molecule of hydrogen forms is 4.48 eV, the molecular binding energy. It is often assumed that this energy is partitioned, in equal fractions of $\frac{1}{3}$, as internal (rovibrational) and translational energies of the molecule, and with $\frac{1}{3}$ being recovered by the grain in phonon excitation. Whether this assumption is correct remains to be established, probably by means of experiments in the laboratory.

3 The Structure of Interstellar Shock Waves

In the previous section, we introduced the MHD conservation equations applicable to one-dimensional, steady-state, multi-fluid flows; these equations enable the structure of C-type shock waves to be calculated. However, shock waves in the interstellar medium are not necessarily, perhaps not normally, of C-type. When a shock wave is produced, for example, in a collision between interstellar clouds at supersonic relative speed, the shock wave is initially of J-type. Depending on the shock speed and the magnetic field strength, this J-type shock wave may develop a magnetic precursor and ultimately become C-type.

The shock speed, u_s , is an important parameter. The kinetic energy flux associated with the shock wave, $\frac{1}{2}\rho u_s^3$, increases rapidly with u_s . Some of this energy is used to heat the gas. When the temperature, T , exceeds a few thousand degrees, molecular hydrogen begins to be collisionally dissociated. Because H_2 is a major coolant, its destruction leads to a further increase in T . Ultimately, the adiabatic sound speed $c_s = (\gamma k_B T / \mu)^{\frac{1}{2}}$, where γ is the ratio of specific heats at constant pressure and volume, approaches the flow speed and a discontinuity occurs in the flow: see Sect. 2.2.

From the conservation relations presented in the previous section, the equations which are applicable to the ‘discontinuity’ in a J-type shock wave and to the cooling flow behind the discontinuity, may be derived. As we have already noted, the so-called discontinuity has a finite width, owing to the effects of viscosity and thermal conduction, which are characterized by length scales comparable with the mean free path for elastic scattering. The process of elastic scattering tends to equalize the values of parameters, such as u and T , associated with the flows of the various fluids. Accordingly, we shall assume single-fluid flow in what follows; but we note that this assumption is not valid for the grains, particularly the more massive grains, which have large inertia.

The conservation equations for single-fluid flow are obtained by adding the equations derived in Sect. 2.1 for multi-fluid flow, i.e. for the neutral, positively and negatively charged fluids. The sums of the source terms appearing on the right-hand sides of the resulting equations of conservation of mass and momentum are identically 0. However, the number density and the energy of the flow are not conserved, in general. The number density can vary because of reactions, such as the collisional dissociation of H_2 :



in which there are two reactants but three products. Energy is lost from the flow in the form of radiation, as mentioned already. Thus, the conservation equations may be written in the form

$$\frac{d}{dz} \left(\frac{\rho u}{\mu} \right) = \mathcal{N}, \quad (42)$$

$$\frac{d}{dz}(\rho u) = 0, \quad (43)$$

$$\frac{d}{dz} \left[\rho u^2 + \frac{\rho k_B T}{\mu} + \frac{B^2}{8\pi} \right] = 0 \quad (44)$$

and

$$\frac{d}{dz} \left[\frac{\rho u^3}{2} + \frac{5\rho u k_B T}{2\mu} + \frac{\rho u U}{\mu} + \frac{u B^2}{4\pi} \right] = \mathcal{B}, \quad (45)$$

where

$$Bu = B_0 u_s$$

and where the subscript ‘0’ denotes quantities in the preshock gas. Equivalently, we have [using Eq. (43)] that

$$\frac{d}{dz} \left(\frac{B}{\rho} \right) = 0 \quad (46)$$

when the magnetic field is frozen in the fluid.

The solution of Eq. (43), (44), (45) and (46) for the flow variables u , ρ , T and B , across the discontinuity and in the cooling flow, was considered by Field et al. [1]. If the molecular weight of the gas varies owing to reactions such as (41) above, Eq. (42) must also be included. We consider first the discontinuity, then the cooling flow.

3.1 The ‘Discontinuity’ in a J-type Shock Wave

We have already noted that the width of the ‘discontinuity’ is determined by viscosity and thermal conduction, and hence by the distance scale for elastic scattering processes. Chemical reactions, including collisional dissociation (41), and collisional processes leading to the emission of radiation, and hence cooling, are all *inelastic* processes, for which the characteristic distance scales are larger, by at least an order of magnitude, than the corresponding elastic scattering processes. Thus, within the shock ‘discontinuity’, the source terms \mathcal{N} and \mathcal{B} in Eq. (42) and (45), respectively, can be taken equal to 0. Furthermore, the flux of internal energy is constant, as the populations of internal energy states do not change. In the context of the equation of energy conservation, the shock transition (‘discontinuity’) may be qualified as *adiabatic*, i.e. there is no exchange of energy with the shock’s environment. Viscosity and thermal conduction are significant only within the shock transition, where the velocity and thermal gradients are large; these processes (viscosity and thermal conduction) are not included in the above equations, as they can be neglected on either side of the ‘discontinuity’. Thus, relations can be obtained between the flow variables immediately downstream and upstream of the discontinuity. These equations – commonly referred to as the *Rankine–Hugoniot relations* – are

$$\rho_1 u_1 = \rho_0 u_s, \quad (47)$$

$$\rho_1 u_1^2 + \frac{\rho_1 k_B T_1}{\mu} + \frac{B_1^2}{8\pi} = \rho_0 u_s^2 + \frac{\rho_0 k_B T_0}{\mu} + \frac{B_0^2}{8\pi}, \quad (48)$$

$$\begin{aligned} & \frac{\rho_1 u_1^3}{2} + \frac{5\rho_1 u_1 k_B T_1}{2\mu} + \frac{u_1 B_1^2}{4\pi} \\ &= \frac{\rho_0 u_s^3}{2} + \frac{5\rho_0 u_s k_B T_0}{2\mu} + \frac{u_s B_0^2}{4\pi}, \end{aligned} \quad (49)$$

$$\frac{B_1}{\rho_1} = \frac{B_0}{\rho_0}, \quad (50)$$

where the subscript ‘0’ denotes the preshock gas, upstream of the discontinuity, ‘1’ denotes the postshock gas, downstream of the discontinuity and the molecular weight μ is constant. Equations (47), (48), (49) and (50) may be combined to yield a quadratic equation for the compression ratio, ρ_1/ρ_0 , across the discontinuity:

$$2(2 - \gamma)b \left(\frac{\rho_1}{\rho_0} \right)^2 + [(\gamma - 1)M^2 + 2\gamma(1 + b)] \frac{\rho_1}{\rho_0} - (\gamma + 1)M^2 = 0. \quad (51)$$

In Eq. (51), M is the shock Mach number, the ratio of the shock speed, u_s , to the isothermal sound speed in the preshock gas, $(k_B T_0/\mu)^{1/2}$; $b = B_0^2/(8\pi p_0)$ is the ratio of the magnetic pressure, $B_0^2/(8\pi)$, to the gas pressure, $p_0 = \rho_0 k_B T_0/\mu$, in the preshock gas. The ratio of specific heats at constant pressure and volume, γ , should be taken equal to 5/3, the value appropriate to a gas with only translational degrees

of freedom; this is because the internal energy of molecules such as H_2 does not have time to adjust to the changes in the temperature and density across the shock discontinuity. This adjustment occurs in the cooling flow, where the equations for the populations of the rovibrational levels of H_2 should be solved in parallel with the hydrodynamic conservation equations in order to follow correctly the variation in the internal energy, U .

In the absence of a magnetic field ($B_0 = 0 = B_1$), the compression ratio across the shock discontinuity is given by

$$\frac{\rho_1}{\rho_0} = \frac{p_0 + h^2 p_1}{p_1 + h^2 p_0}, \quad (52)$$

where $h^2 = (\gamma + 1)/(\gamma - 1)$. The pressure ratio across the shock discontinuity, p_1/p_0 , is related to the isothermal Mach number in the preshock gas by

$$M^2 = \frac{\gamma + 1}{2} \frac{p_1}{p_0} + \frac{\gamma - 1}{2}. \quad (53)$$

The processes which determine the thickness of the shock transition – viscosity and thermal conduction – are irreversible and give rise to an increase in entropy across the shock front. It can be shown that, as a consequence, the condition $p_1 > p_0$ must apply; it follows from (52) that $\rho_1/\rho_0 > 1$. Thus, both the pressure and the density of the gas *increase* as the gas traverses the discontinuity. The quadratic Eq. (51) has two roots, in general, but only one of the solutions corresponds to a compression shock, in which $\rho_1/\rho_0 > 1$.

Equation (52) shows that, in the limit of $p_1 \gg p_0$, $\rho_1/\rho_0 \rightarrow h^2 = 4$ for $\gamma = 5/3$. From Eq. (47), we see that, in this limit,

$$\frac{u_1}{u_s} = \frac{\rho_0}{\rho_1} = \frac{1}{4}.$$

Thus, in the reference frame of the shock wave, the gas flows into the shock front at speed u_s and out at speed $u_s/4$. In an inertial frame in which the preshock gas is at rest, the gas is accelerated at the shock front to a speed $3u_s/4$.

The temperature rise at the discontinuity is given by

$$\frac{T_1}{T_0} = \frac{(p_1 + h^2 p_0) p_1}{(p_0 + h^2 p_1) p_0}. \quad (54)$$

Thus, in the limit of $p_1 \gg p_0$, $T_1/T_0 \rightarrow (p_1/h^2 p_0)$. Unlike the compression ratio, the temperature ratio across the shock wave is unlimited.

The presence of a transverse magnetic field moderates the compression and the increase in temperature which occur at the shock front. As we have already seen, if the magnetic field is sufficiently strong, the discontinuity can be suppressed altogether, in C-type shock waves.

3.2 The Cooling Flow of a J-type Shock Wave

If sufficiently hot, the compressed gas which flows out of a shock discontinuity is able to excite molecules, atoms and ions. These ‘cooling’ processes cause the temperature of the gas to fall whilst it continues to be compressed. By the time that the gas has cooled to its equilibrium (postshock) temperature, the total compression ratio, relative to the preshock gas, can be much greater than the maximum value of 4 at the shock discontinuity. The presence of a finite, transverse magnetic field limits the degree of compression of the gas.

The temperature profile computed for a J-type shock wave with a speed $u_s = 25 \text{ km s}^{-1}$, propagating into gas of (preshock) density $n_H = n(\text{H}) + n(\text{H}_2) + n(\text{H}^+) = 10^4 \text{ cm}^{-3}$, in the absence of a magnetic field, is shown in Fig. 1. The independent variable in this figure is the flow time,

$$t = \int \frac{1}{u} dz,$$

where z is the direction of flow and u is the flow speed, in the shock frame. In these calculations, the shock ‘discontinuity’ has a small but finite width, owing to artificial viscosity terms having been introduced into the conservation equations; these equations can then be integrated from the preshock through to the postshock, equilibrium gas.

The initial jump in temperature at the discontinuity is sufficient for collisional dissociation to take place, as may be seen from Fig. 1, where the abundances of H, H_2 and H^+ , relative to n_H , are plotted. In fact, molecular hydrogen is rapidly collisionally dissociated in the hot gas behind the ‘discontinuity’, on a flow timescale of

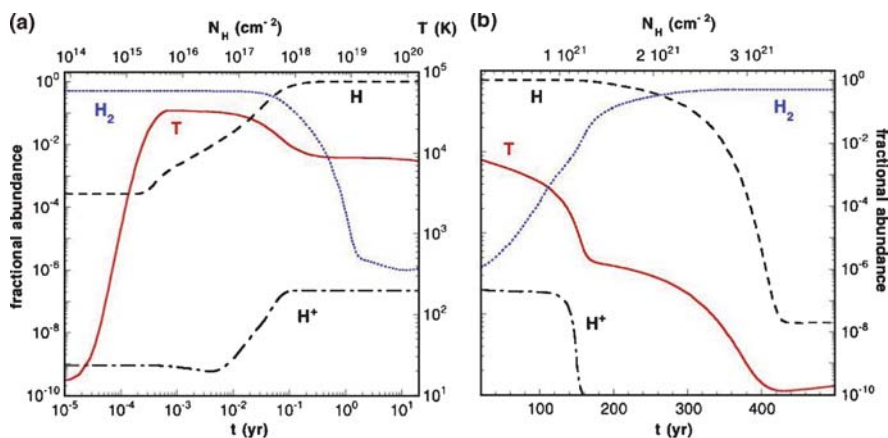


Fig. 1 The temperature profile computed for a J-type shock wave with a speed $u_s = 25 \text{ km s}^{-1}$, propagating into gas of (preshock) density $n_H = n(\text{H}) + n(\text{H}_2) + n(\text{H}^+) = 10^4 \text{ cm}^{-3}$; N_H is the corresponding column density. The fractional abundances of H, H_2 and H^+ are also plotted. In the left-hand panel, the abscissa (the flow time) is on a logarithmic scale

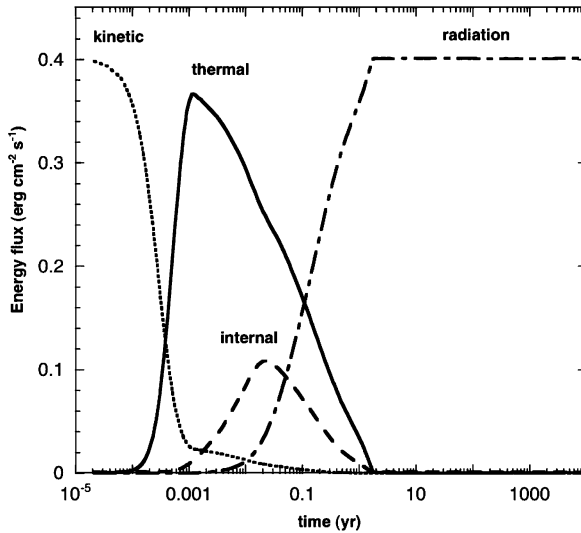


Fig. 2 Contributions to the total energy flux as functions of the flow time, for a J-type shock wave with a speed $u_s = 15 \text{ km s}^{-1}$, propagating into gas of (pres shock) density $n_H = 10^5 \text{ cm}^{-3}$ in the absence of a magnetic field

the order of 1 year. The main coolants of the gas are then atoms and ions, notably atomic oxygen, through its fine structure transitions at 63 and 147 μm . In the cooling flow, H_2 reforms (on grains), and the associated heating of the gas gives rise to the ‘knee’ in the temperature profile, apparent in the right-hand panel of Fig. 1. Finally, after approximately 500 year, the kinetic temperature reaches its postshock equilibrium value, of the order of 10K. The total compression ratio in this case approaches 10^4 : there is no magnetic field to moderate the compression of the gas.

Referring to Eq. (45), we see that, in the absence of a magnetic field, contributions to the energy flux arise from: (i) the kinetic energy of the flow; (ii) the thermal energy of the gas; (iii) the internal energy of the gas; and (iv) radiative losses (incorporated in \mathcal{B}). In the preshock gas, (i) dominates. Immediately behind the ‘discontinuity’, (ii) is the major term, with some contribution from (iii) at the beginning of the cooling flow. Finally, radiative losses take over and the gas cools to its postshock, equilibrium state. This transformation of one form of energy into another through a J-type shock wave is illustrated in Fig. 2.

3.3 C-Type Shock Waves

The interaction of the gas and the grains with the magnetic field is crucially important for the development of C-type shock waves. The field couples directly to the charged fluid and thence to the neutral fluid, which contains most of the mass, via collision processes (cf. Sect. 2). The strength of the coupling between the charged

and the neutral fluids depends on the degree of ionization of the gas and hence on the rates of chemical reactions which modify the degree of ionization within the shock wave. The coupling between the charged and neutral fluids also depends on the fraction of the grains which is (principally negatively) charged. Although the *number density* of the grains is much smaller than that of the gaseous ions, their *mass density* is, in dark clouds, much larger.

By way of illustration of the importance of chemical reactions in this context, Fig. 3 compares the fractional ionization of the gas and the thermal profiles computed for a C-type shock wave of speed $u_s = 10 \text{ km s}^{-1}$ which propagates into gas of preshock density $n_H = 10^3 \text{ cm}^{-3}$ and in which the transverse magnetic field strength is $B_0 = 25 \mu\text{G}$; in one calculation, chemical reactions were neglected, and, in the other, they were included. As may be seen from this figure, the degree of ionization is modified considerably by the chemistry, both within the shock wave and in the postshock, equilibrium gas. Endothermic reactions between atomic ions

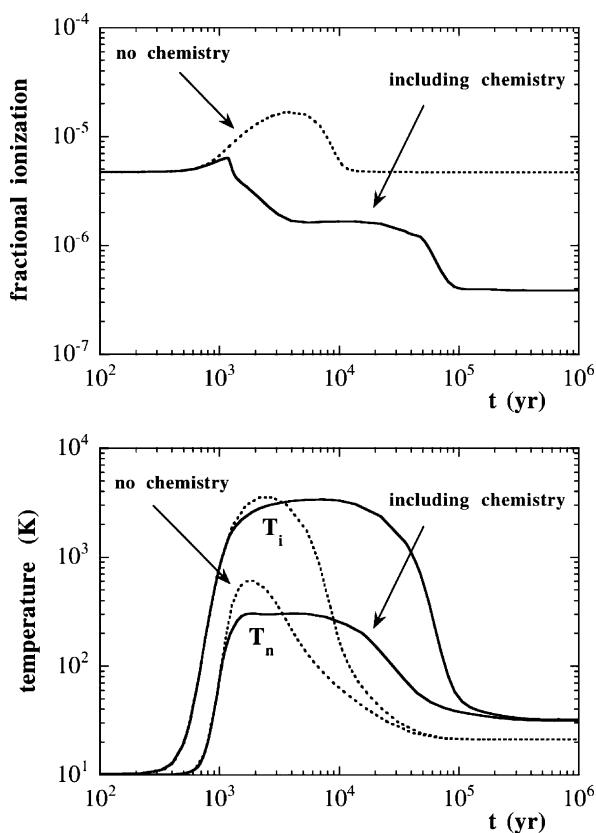


Fig. 3 The fractional ionization of the gas and the thermal profiles in a C-type shock wave of speed $u_s = 10 \text{ km s}^{-1}$, which propagates into gas of preshock density $n_H = 10^3 \text{ cm}^{-3}$ in which the transverse magnetic field strength $B_0 = 25 \mu\text{G}$

and H_2 are activated by the ion–neutral drift speed within the shock wave, and the molecular ions which are produced are able to recombine rapidly with electrons. Thus, the fractional ionization *falls* within the shock wave when the chemistry is included. On the other hand, when the chemistry is neglected, the fractional ionization increases, owing to the differential compression of the ions, before falling back to its equilibrium value in the postshock gas. To a lower degree of ionization corresponds a more extended shock wave, in which the neutral fluid has a lower maximum temperature.

Figure 3 also shows that the time of flow through a C-type shock wave, from the preshock to the postshock, equilibrium gas, is of the order of 10^5 year; this is very much greater than in a J-type shock wave of comparable speed. As the time to reach steady state cannot be less than the time of flow through the shock wave, it follows that C-type shock waves do not attain steady state under conditions, such as those in jets, in which the dynamical timescales are much less than 10^5 year.

In order to contrast the flow times through J- and C-type shock waves, we plot, in Fig. 4, the thermal and density profiles computed in the case of a shock wave of speed $u_s = 30 \text{ km s}^{-1}$, propagating into gas of preshock density $n_H = 10^4 \text{ cm}^{-3}$; in the upper panel, the transverse magnetic field strength, $B_0 = 0$, and the shock wave is J-type, in the lower panel, $B_0 = 100 \mu\text{G}$, and the shock wave is C-type, in steady state. The difference in the times of flow through the shock wave and into the postshock region is striking: the flow time is 2 orders of magnitude larger in the case of the C-type shock wave. On the other hand, the peak kinetic temperature is an order of magnitude higher and the density of the postshock gas is 3 orders of magnitude larger in the J-type shock wave. We conclude that the presence of a magnetic field transforms the physical – and hence the chemical – steady-state structure of a shock wave.

In order to determine rigorously the structure of shock waves in their evolution to a steady state, a time-dependent MHD code is necessary. Such codes have been developed, but they are still restricted in the range and complexity of the physico-chemical processes which can be incorporated. An alternative approach, which provides an approximation to the time-dependent shock structure that is acceptable in the context of many applications, will be presented below.

When there is a disturbance which propagates at supersonic speed in a medium, a shock wave can be produced. Stellar winds, jets, turbulence and collisions between interstellar clouds, for example, are all susceptible to generating shock waves. The concept of ‘steady state’ is relevant only if the mechanism responsible for producing the shock wave endures for at least the time required for the shock wave to attain its equilibrium state. Evanescent phenomena must, by their nature, be studied by means of an explicitly time-dependent model. The energy source which creates and maintains a shock wave may be compared with the ‘piston’ that can be used to generate shock waves in the laboratory. Let us denote the speed of propagation of the piston by u_p . We shall assume that u_p is constant and show, from considerations of the continuity of the flow, that the shock front propagates at a speed, u_s , which somewhat exceeds u_p .

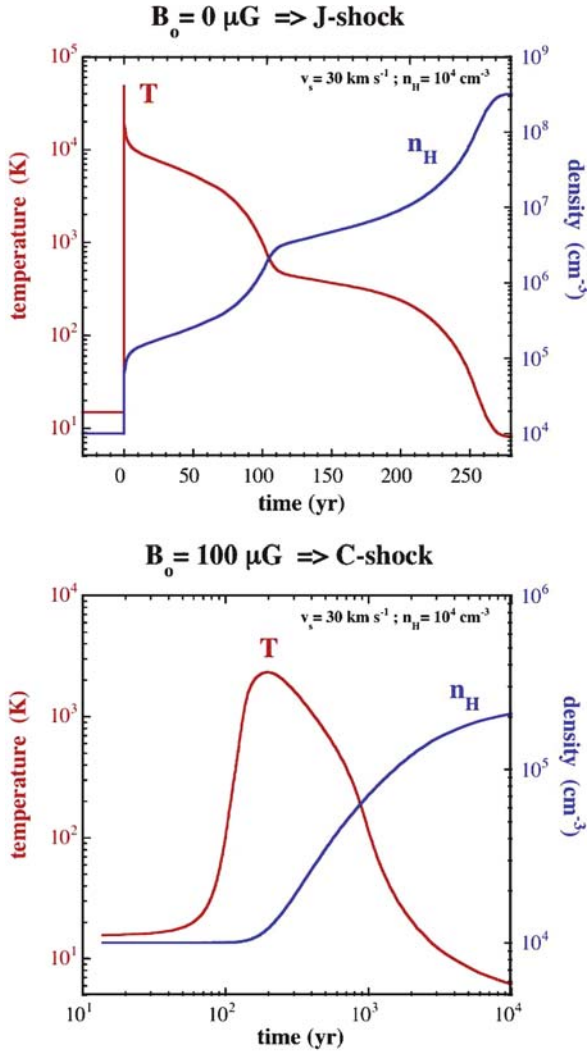


Fig. 4 Comparing the thermal and density profiles, in steady state, of a shock wave of speed $u_s = 30 \text{ km s}^{-1}$ which propagates into gas of preshock density $n_H = 10^4 \text{ cm}^{-3}$, in which the transverse magnetic field strength $B_0 = 0$ in the *upper panel* and $B_0 = 100 \text{ } \mu\text{G}$ in the *lower panel*

As is customary, we apply the equation of continuity in the frame of the shock wave, i.e. the frame in which the shock front is at rest; this is achieved by subtracting the shock velocity, \mathbf{u}_s , from velocities in an inertial frame, usually taken to be the frame of the preshock gas. Referring to Eq. (47), we see that

$$u_1 = \frac{u_s}{(\rho_1/\rho_0)}, \quad (55)$$

where ρ_0 , ρ_1 denote the preshock and postshock gas density, respectively, and u_1 is the postshock flow speed; the ratio (ρ_1/ρ_0) is the compression factor. In the inertial frame, the preshock gas is at rest and the postshock gas flows at speed

$$u_s - \frac{u_s}{(\rho_1/\rho_0)}$$

in the direction of propagation of the shock front. At the surface of the piston, the gas is moving with the speed of the piston, u_p , and hence

$$u_p = u_s \left[1 - \frac{1}{(\rho_p/\rho_0)} \right], \quad (56)$$

where ρ_p is the value of ρ_1 at the surface of the piston. It follows from (56) that $u_s > u_p$.

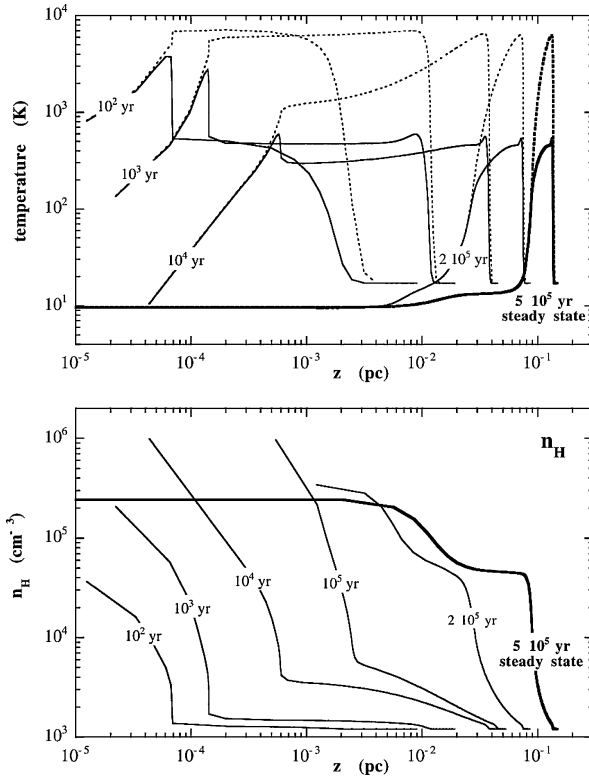


Fig. 5 Temperature and density profiles as functions of position and time for a shock wave of speed $u_s = 10 \text{ km s}^{-1}$, propagating into gas of density $n_H = 10^3 \text{ cm}^{-3}$ in which the transverse magnetic field strength $B_0 = 25 \mu\text{G}$. The shock wave advances from left to right until a stationary state is finally attained. The origin of the position coordinate, z , is at the ‘piston’

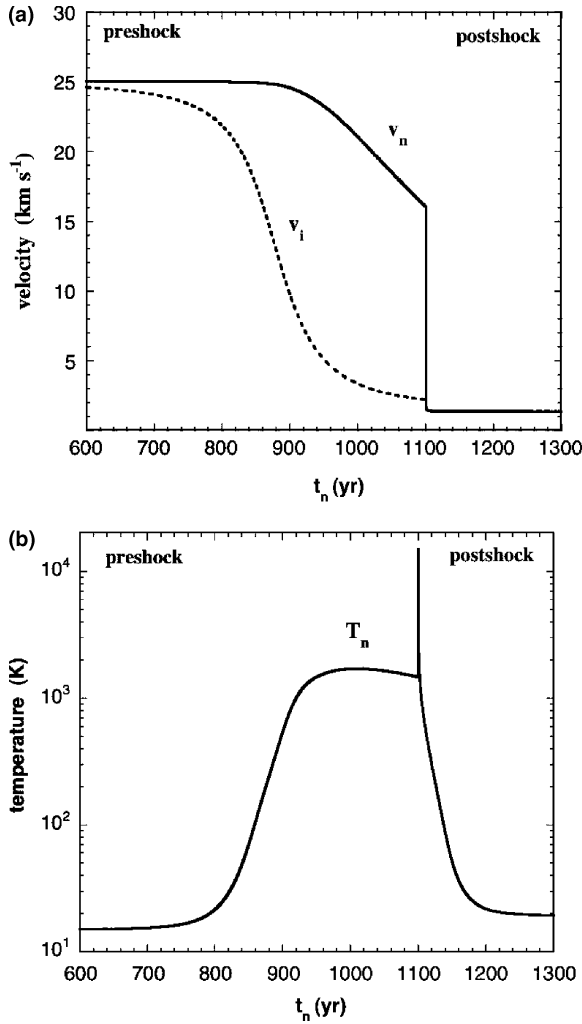


Fig. 6 Velocity and temperature as functions of flow time for a J-type shock wave with a magnetic precursor: the shock speed $u_s = 25 \text{ km s}^{-1}$; the preshock gas density $n_H = 10^4 \text{ cm}^{-3}$; and the transverse magnetic field strength in the preshock gas $B_0 = 100 \mu\text{G}$

If the shock is initially J-type, the compression factor at the discontinuity is $(\rho_1/\rho_0) = 4$, in the limit of large Mach numbers. Then, $u_s = 4u_p/3$, and the shock discontinuity moves away from the piston with which it was initially in contact. In the cooling flow which develops between the discontinuity and the piston, the gas is compressed further and the speed of the shock front decreases towards that of the piston (which is assumed constant). By the time that the compression factor at the surface of the piston has become large $[(\rho_p/\rho_0) \gg 1]$, the piston and the shock front are travelling at the same speed. From Eq. (56), it may be seen that the speed

with which the shock front moves away from the piston, $u_s - u_p = u_s/(\rho_p/\rho_0)$, is also the speed of fluid flow (in the reference frame of the shock wave) at the surface of the piston.

If the transverse magnetic field strength is sufficiently large, a magnetic precursor develops upstream of the shock discontinuity and preheats the gas. As a consequence, the sound speed in the gas immediately upstream of the discontinuity increases and the Mach number falls, i.e. the shock discontinuity weakens. By the time that steady state is attained, the discontinuity may have disappeared altogether, in which case the structure has become pure C-type; this evolution is illustrated in Fig. 5. This figure shows also the shock front gradually separating from the ‘piston’ as time progresses; the speed of the shock front, relative to the piston, decreases as the compression factor increases.

The evolution of the shock wave from initially pure J-type to J-type with a magnetic precursor, seen in Fig. 5, to finally C-type, may be simulated by introducing a discontinuity in the flow at a point in the steady-state profile which is located increasingly downstream as time advances. The time is given by the time of flow of a fluid particle through the precursor to the discontinuity,

$$t = \int \frac{1}{u} dz.$$

The steady-state structure of the shock wave ‘unfolds’ as time progresses. Comparisons with the results of explicitly time-dependent MHD calculations ([7, 8]) have shown that the evolution of the shock wave is satisfactorily described by means of the approximation outlined above. An example of such a composite shock structure is given in Fig. 6.

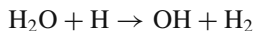
4 Shock Waves in Dark Clouds

The characteristics and spectroscopic signatures of J-type shock waves propagating in molecular media have been studied for many years. In the cooling flow behind the discontinuity, molecules, atoms and ions can be collisionally excited in the shock-heated gas. Rovibrational molecular transitions, fine structure and other ‘forbidden’ atomic and ionic transitions are emitted and, when detected, provide diagnostic information on the medium.

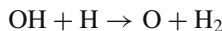
The physical state and chemical composition of the cooling flow depend on the shock speed, the transverse magnetic field strength in, and the density of, the preshock gas. With increasing shock speed, u_s , the maximum postshock temperature increases to values (of the order of 10^3 K) at which the collisional excitation and dissociation of molecular hydrogen begin to be significant. As u_s increases further and the temperature reaches approximately 10^4 K, electronic excitation of atomic hydrogen occurs and, ultimately, H is collisionally ionized. The ultraviolet radiation produced by the radiative decay of the electronically excited states of H, which are populated either by collisional excitation or by radiative recombination of H^+

with electrons, becomes sufficiently intense to pre-ionize the gas upstream of the shock discontinuity; the shock wave is then said to have a *radiative precursor*. If the transverse magnetic field is weak or absent, shock speeds $u_s \approx 25 \text{ km s}^{-1}$ are sufficient to cause almost complete dissociation of H_2 immediately downstream of the discontinuity, where the kinetic temperature is highest.

In Fig. 7 are illustrated the contributions of various atomic and molecular species to the cooling of a J-type shock wave of speed $u_s = 25 \text{ km s}^{-1}$, preshock density $n_{\text{H}} = 10^4 \text{ cm}^{-3}$ and transverse magnetic field strength $B_0 = 0$. The collisional dissociation of molecular hydrogen releases atomic hydrogen into the hot gas, leading to the destruction of other molecular species. For example, the reactions



and



return oxygen to atomic form. Through its fine structure transitions at $63 \mu\text{m}$ and $147 \mu\text{m}$, O then becomes the principal coolant of the gas. As the gas cools down, first H_2 reforms, and then other molecules, such as H_2O and CO. The time required for the medium to reach its postshock equilibrium temperature is approximately 500 year. Thus, in dynamically young objects, even J-type shock waves may not have had sufficient time to reach steady state.

As the shock speed increases beyond $u_s = 25 \text{ km s}^{-1}$, the extent of the cooling flow (and the time required for the gas to attain its postshock equilibrium temperature) begins to *decrease*: see Fig. 8. This reversal occurs because the degree of ionization of atomic hydrogen, and consequently the fractional electron abundance,

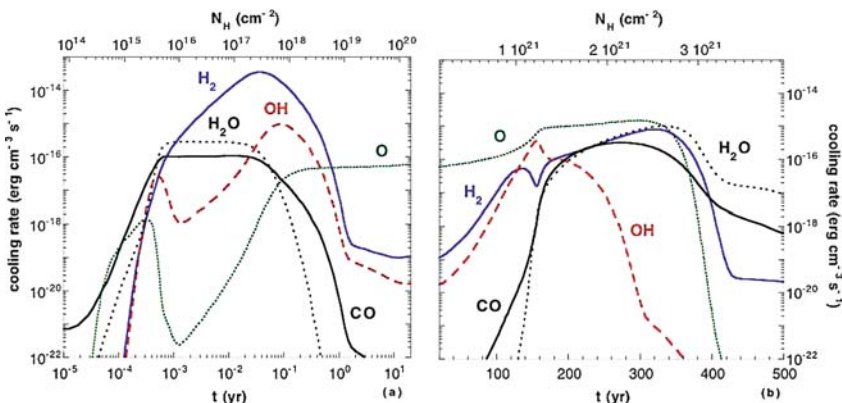


Fig. 7 The rates of cooling by the principal coolants for a J-type shock wave with a speed $u_s = 25 \text{ km s}^{-1}$, propagating into gas of (preshock) density $n_{\text{H}} = n(\text{H}) + n(\text{H}_2) + n(\text{H}^+) = 10^4 \text{ cm}^{-3}$, in the absence of a magnetic field; N_{H} is the corresponding column density. In the *left-hand panel*, the abscissa (the flow time) is on a logarithmic scale

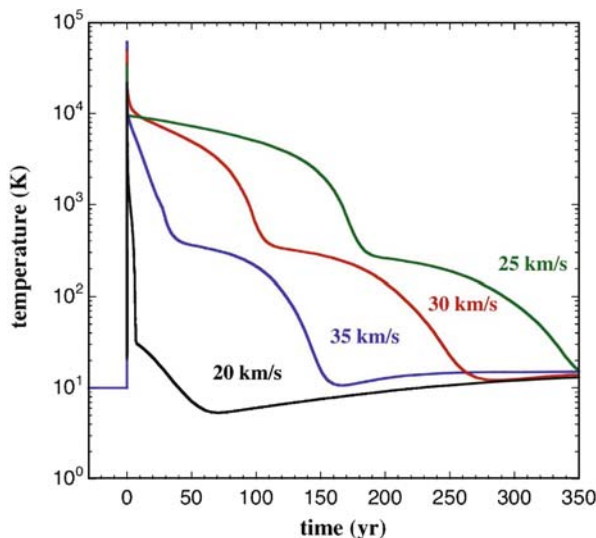


Fig. 8 Dependence of the duration of the cooling flow on the shock speed, for gas of (pres shock) density $n_H = 10^4 \text{ cm}^{-3}$, in the absence of a magnetic field

increase rapidly with u_s above approximately 25 km s^{-1} . Cooling of the medium owing to electron collisional excitation of atomic hydrogen, principally the Ly α transition, then becomes important.

The populations of the rovibrational levels of H_2 do not respond instantaneously to the changes in temperature and density which occur at and behind the shock discontinuity. Indeed, as we have already seen, excitation of the internal degrees of freedom of H_2 is insignificant within the shock 'discontinuity', where the flow variables change adiabatically. In the cooling flow, immediately behind the discontinuity, the level populations respond to changes in the temperature and density on a timescale which, by definition, is comparable with the local cooling time (on which the temperature changes significantly), as H_2 is the principal coolant of the gas. Under these conditions, it is essential to integrate the differential equations governing the H_2 level populations in parallel with the dynamical equations and the chemical rate equations.

In the presence of a transverse magnetic field of sufficient strength, an initially J-type shock wave develops a magnetic precursor and can ultimately become C-type. In order for a precursor to develop, the magnetosonic speed, c_m [Eq. (17)], must exceed the shock speed, u_s . As the physical conditions in the preshock gas determine the value of the magnetosonic speed, the requirement that $u_s < c_m$ sets an upper limit on the speed of shock waves which can become C-type.

The physical conditions in the preshock gas, notably the degree of ionization, depend on the rate of cosmic ray ionization of hydrogen, ζ . Most of the positive charge is associated with atomic and molecular ions in the gas. However, contributions to the negative charge arise not only from the free electrons

but also from negatively charged grains and, possibly, from anions of polycyclic aromatic hydrocarbons (PAH). The distribution of the negative charge amongst free electrons, grains and PAH depends on the fractional abundance of the PAH molecules and on the rates of electron attachment and detachment reactions, as well as the rates of recombination of positive ions with electrons on the surfaces of negatively charged grains; all these parameters are subject to significant uncertainties.

In Fig. 9 are plotted the values of the magnetosonic speed computed for a range of densities of the preshock gas and of fractional abundances of PAH molecules; the cosmic ray ionization rate is $\zeta = 1 \times 10^{-17} \text{ s}^{-1}$ and the transverse magnetic field strength $B_0 = [n_H]^{0.5}$, where n_H is in units of cm^{-3} and B_0 is in μG ; this functional dependence of B_0 on n_H ensures that the magnetic energy density in the preshock gas remains proportional to the thermal energy density, at a given temperature.

Also plotted in Fig. 9 is the critical shock speed at which the degree of collisional dissociation of H_2 , the principal coolant, becomes sufficient for a thermal runaway to occur. There results a sonic point in the flow (owing to the rapidly rising temperature and hence sound speed), and the shock becomes J-type. It may be seen from Fig. 9 that, for $n_{\text{PAH}}/n_H = 10^{-6}$ and $n_H > 10^4 \text{ cm}^{-3}$, the upper limit to the possible speed of a C-type shock wave is determined by the collisional dissociation of H_2 , whereas, for $n_H < 10^4 \text{ cm}^{-3}$, it is determined by the magnetosonic speed in the preshock gas. The limit imposed by the magnetosonic speed becomes more stringent for $n_{\text{PAH}}/n_H < 10^{-6}$. Although still uncertain, the fraction of carbon which is

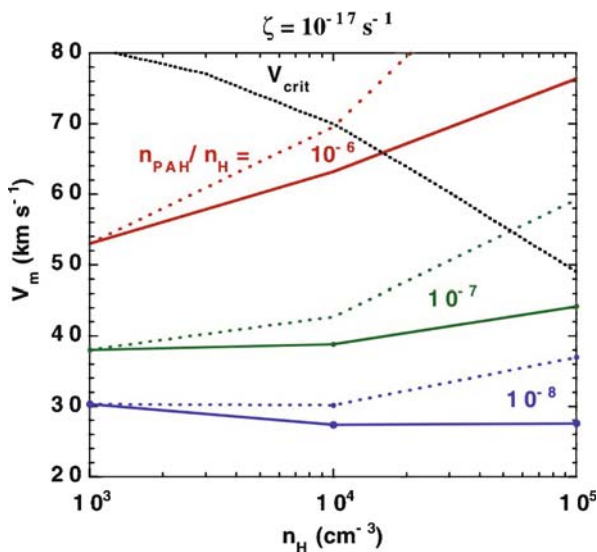


Fig. 9 The magnetosonic speed in preshock gas of density $10^3 \leq n_H \leq 10^5 \text{ cm}^{-3}$ and fractional abundance of PAH $10^{-8} \leq n_{\text{PAH}}/n_H \leq 10^{-6}$; the cosmic ray ionization rate is $\zeta = 1 \times 10^{-17} \text{ s}^{-1}$ and the transverse magnetic field strength $B_0 \mu\text{G} = [n_H]^{0.5}$, where n_H is in units of cm^{-3}

believed to exist in ‘very small grains’ sets an upper limit to the fractional abundance of the PAH, namely $n_{\text{PAH}}/n_{\text{H}} \lesssim 10^{-6}$.

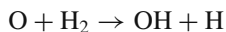
The steady-state thermal, velocity and density profiles of the neutral and charged fluids for an illustrative C-type shock wave are shown in Fig. 10. These profiles display a number of characteristics of C-type shock structure: the initially rapid decoupling of the velocity of the charged fluid from that of the neutrals, followed by their recoupling on a timescale of the order of 10^3 year for the model shown; the initial increase in the fractional ionization, owing to the differential compression of the charged fluid, followed by a decrease (by approximately 3 orders of magnitude in the model shown) to its postshock value; the decoupling of the temperatures of the charged fluids from the temperature of the neutrals and, over a more restricted time range, the decoupling of the temperature of the ions from that of the electrons. The electron temperature lies between the temperatures of the ions and the neutrals: the much greater abundance of the neutrals compensates for the stronger thermal coupling of the electrons to the positive ions (which is mediated by the attractive Coulomb force). Owing to the differential compression of the charged fluid and the initial rise in T_e , the rate of electron attachment to grains is sufficient to ensure that the grains are predominantly negatively charged within the shock wave. The rate equations which determine the grain charge must be solved in parallel with the equations which describe the dynamical structure of the shock wave, because the chemistry and the dynamics interact strongly.

5 Shock Waves in the Presence of an Ultraviolet Radiation Field

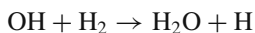
The interstellar background ultraviolet radiation field permeates the diffuse interstellar gas and ionizes species with ionization potentials less than that of atomic hydrogen. Consequently, in diffuse gas, the most abundant ion is C^+ . It has already been mentioned that endothermic reactions, such as



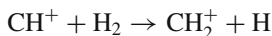
can become significant in shocked gas; this is true also of reactions such as



which is endothermic and has a barrier, and of



which has a barrier. In a medium which is rich in molecular hydrogen, the reaction of C^+ with H_2 is followed rapidly by the exothermic hydrogenation reactions



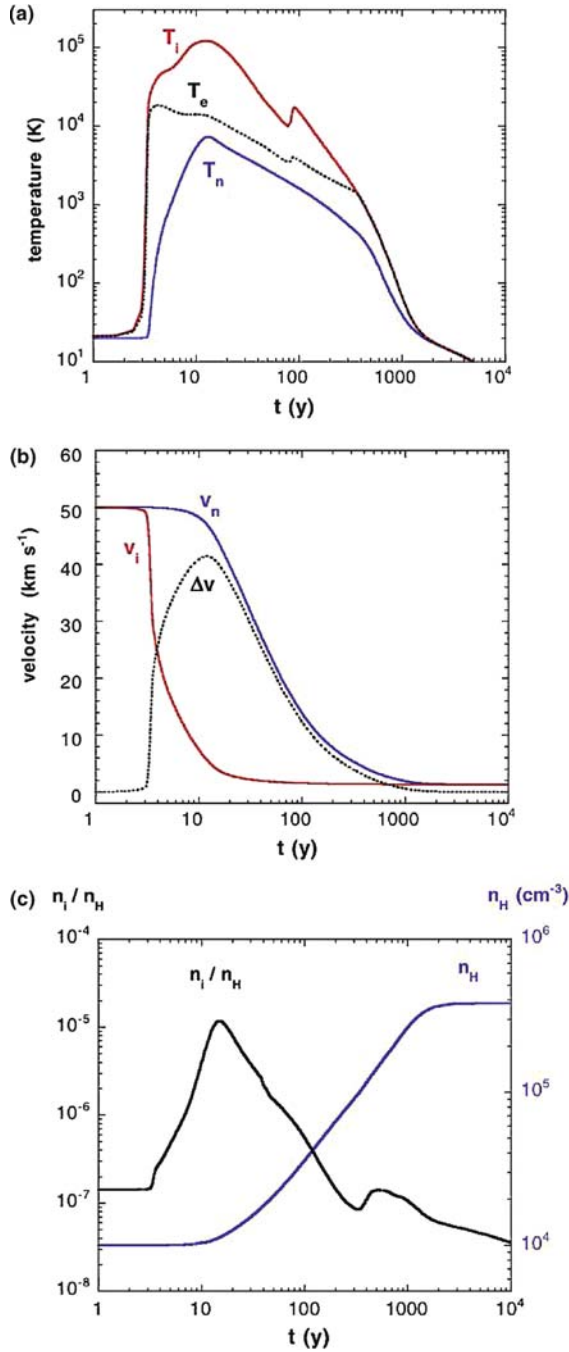
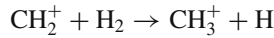
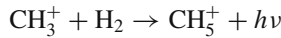


Fig. 10 The steady-state profiles of (a) temperature, (b) velocity ($\Delta v = v_n - v_i$) and (c) density of the neutral and charged fluids, for an illustrative C-type shock wave of speed $v_s = 50 \text{ km s}^{-1}$, preshock density $n_H = 10^4 \text{ cm}^{-3}$ and transverse magnetic field strength $B_0 = 100 \mu\text{G}$

and



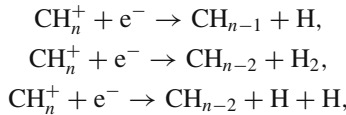
beyond which the sequence proceeds much more slowly, either by radiative association



or by the strongly endothermic reaction



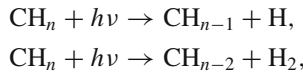
All of the hydrocarbon ions which are formed undergo dissociative recombination reactions with electrons, such as



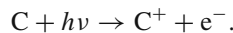
the net effect being the neutralization of C^+ ions in the gas.

The key reaction in the above hydrogenation cycle is $\text{C}^+(\text{H}_2, \text{H})\text{CH}^+$, which is endothermic by 4640K. Ambipolar diffusion in shock waves will drive this reaction once the relative kinetic energy of the ions and the neutrals is comparable with the endothermicity, i.e. once $\mu_{\text{in}}(u_{\text{i}} - u_{\text{n}})^2 / (2k_{\text{B}}) \approx 4640\text{K}$, where $\mu_{\text{in}} = \mu_{\text{i}}\mu_{\text{n}} / (\mu_{\text{i}} + \mu_{\text{n}})$ is the reduced mass of the $\text{C}^+\text{-H}_2$ pair. This relation implies that the relative drift speed, $(u_{\text{i}} - u_{\text{n}})$, should be at least 6 or 7 km s^{-1} . Such speeds are readily attained in shock waves with speeds $u_{\text{s}} \gtrsim 10 \text{ km s}^{-1}$ in which the magnetic field strength in the preshock gas, B_0 , is a few μG .

Photoreactions prevent ambipolar diffusion leading to the complete neutralization of the C^+ component of the ionized gas. The CH_n molecules which are the products of the above cycle are photodissociated:



and the atomic carbon which is produced is then photoionized:



As a result of these reactions, C^+ ions are restored to the gas over a distance scale which is comparable with the MHD shock width. Thus, the C^+ density in such a shock wave first rises owing to the compression of the ionized gas, then falls as a result of ion-molecule reactions and dissociative recombination, and finally rises again as photoreactions take over. This behaviour is illustrated in Fig. 11. The pre- and postshock values of the density of C^+ ions are determined by the equilibrium

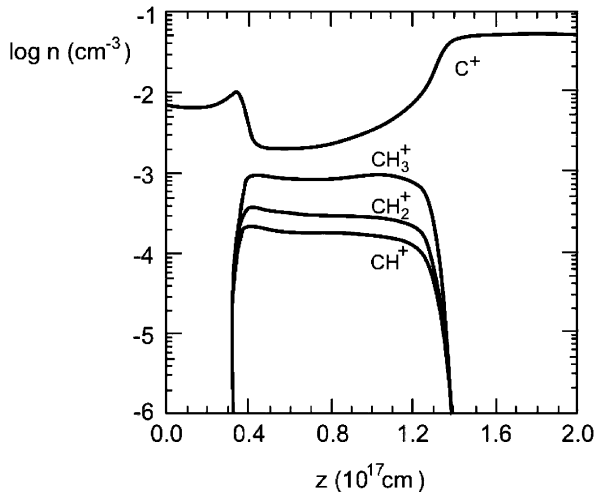


Fig. 11 The densities of C-bearing ions through a C-type shock wave of speed 12 km s^{-1} , propagating into a diffuse interstellar cloud in which the preshock density $n_{\text{H}} = 20 \text{ cm}^{-3}$ and the transverse magnetic field strength $B_0 = 5 \text{ } \mu\text{G}$

between the rate of photoionization of C and the reverse process, namely, radiative recombination of C^+ with electrons.

It has already been mentioned that, in shock-heated gas, the chemistry of oxygen is initiated by the reaction $\text{O}(\text{H}_2, \text{H})\text{OH}$ which has a barrier of 2980K (which is larger than the endothermicity of approximately 960K). The subsequent reaction, $\text{OH}(\text{H}_2, \text{H})\text{H}_2\text{O}$, has a smaller barrier of 1490K. Photodissociation of H_2O and OH eventually returns O to the gas. Reactions of C^+ and CH_n^+ with O and OH_n lead to the formation of CO^+ and H_nCO^+ , which recombine dissociatively with electrons. The molecules which are so produced are ultimately photodissociated into C and O. The end result is the restoration of oxygen to its atomic form in the postshock gas. However, the facility with which water is produced in shock waves, through the reactions $\text{O}(\text{H}_2, \text{H})\text{OH}$ and $\text{OH}(\text{H}_2, \text{H})\text{H}_2\text{O}$, results in fractional abundances of H_2O which exceed that observed in shocked molecular gas associated with IC443 (a supernova remnant) by the SWAS satellite [16]. Thus, although the transformation of atomic oxygen into water in shock waves was believed to be well understood, the models failed their first observational test. Other surprises of this type undoubtedly await us.

6 Diagnostics of Shock Waves: The H_2 Excitation Diagram

Because of the high abundance of H_2 in the ambient molecular gas into which interstellar jets propagate, the simulation of the rovibrational spectrum of H_2 by means of shock models is a unique diagnostic tool for outflow sources. Observations and cal-

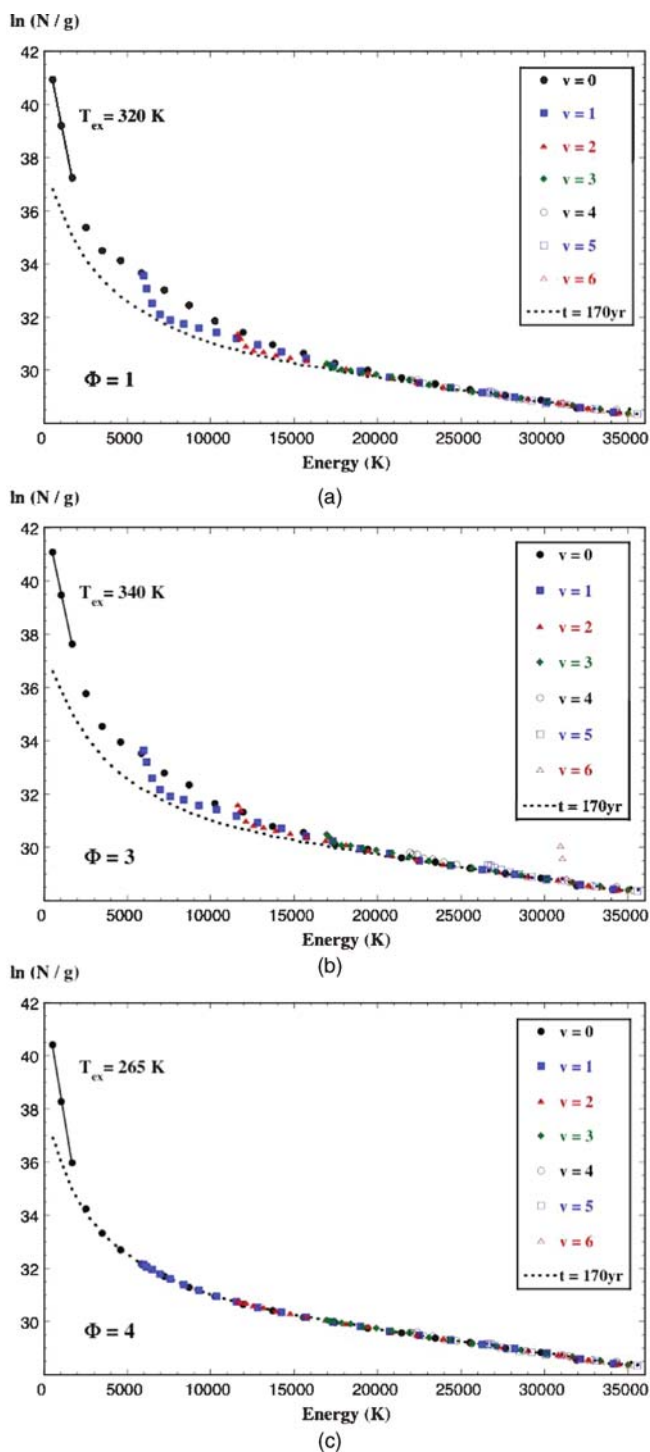


Fig. 12 (continued)

culations of the emission line intensities are usually confronted through the medium of the ‘excitation diagram’, which derives from considerations of thermodynamic equilibrium, when the population n of the rovibrational level (v, J) , relative to the ground state $(0, J')$, is given by a Boltzmann distribution:

$$\frac{n_{vJ}}{n_{0J'}} = \frac{g_J}{g_{J'}} \exp\left(-\frac{E_{vJ} - E_{0J'}}{k_B T}\right). \quad (57)$$

In Eq. (57), E denotes the energy of the level and g its statistical weight; T is the kinetic temperature. Molecular hydrogen occurs in either the ortho (odd values of J) or the para (even values of J) form, whose respective ground states are $J' = 1$ and $J' = 0$. Allowing for the nuclear spin degeneracies, the statistical weights of the ortho levels are $g_J = 3(2J + 1)$ and of the para levels $g_J = (2J + 1)$. Integrating along the line of sight and taking natural logarithms, Eq. (57) may be written in the form

$$\ln\left(\frac{N_{vJ}}{g_J}\right) = \ln\left(\frac{N_{0J'}}{g_{J'}}\right) - \frac{E_{vJ} - E_{0J'}}{k_B T}, \quad (58)$$

where N is a column density and T is assumed to be constant. Then, a plot of $\ln(N_{vJ}/g_J)$ against $(E_{vJ} - E_{0J'})/k_B$ has gradient $-1/T$.

In practice, thermodynamic equilibrium rarely, if ever, applies to H_2 in the interstellar medium. The more highly excited levels have larger radiative transition probabilities and become underpopulated, relative to a Boltzmann distribution, whereas the lower levels become overpopulated. In addition, the kinetic temperature, T , varies along the line of sight, in general. Consequently, the plot is not linear but curved, as shown in Fig. 12. Nonetheless, the excitation diagram provides a convenient means of comparison of the observed and calculated column densities of the rovibrational levels. The local tangent to the curve provides a measure of the kinetic temperature in the region of formation of the line, with the temperature increasing (gradient becoming more positive) with increasing excitation energy of the emitting level.

We recall that H_2 has no permanent dipole moment. Its rovibrational transitions arise from a quadrupole moment, and the associated radiative emission and absorption probabilities are small. Consequently, the rovibrational transitions of H_2 are optically thin, and the interpretation of its spectrum is not complicated by radiative



Fig. 12 The H_2 excitation diagram for a J-type shock wave in which $u_s = 25 \text{ km s}^{-1}$ and $n_H = 10^4 \text{ cm}^{-3}$ in the preshock gas; $B_0 = 0$. It was assumed in the calculations [17] that the reformation of H_2 in the cooling flow leads to, from top to bottom: (a) the rovibrational levels being populated in proportion to $g_J \exp(-E_{vJ}/17316)$; (b) population of the $J = 0$ and $J = 1$ rotational levels of the $v = 6$ manifold (1/4 in the para and 3/4 in the ortho level); (c) the levels are populated in proportion to their local population densities. For the purposes of comparison, the broken curves exclude the contribution of the process of reformation of H_2 to the column densities (This figure is presented in color in the electronic version of the book on springerlink.com.)

transfer effects, which have to be taken into account when studying less abundant but dipolar molecules.

Figure 12 illustrates also the effects of the reformation of H_2 , on grain surfaces, in the cooling flow of the shock wave. The molecule is not produced, in general, in its ortho or para ground state ($J = 1$ and $J = 0$, respectively), but in excited rovibrational states, with what is still an unknown distribution of population. In the top panel of Fig. 12, it is assumed that $1/3$ of the binding energy of H_2 (4.48 eV or 51947K) is retained as internal excitation energy of the molecule, with the level populations following a Boltzmann distribution at $T = 51947/3 = 17316\text{K}$; the remainder of the binding energy is absorbed by the grain or transformed into translational kinetic energy of the molecule. The broken curves show the result obtained neglecting the contribution of the reformation process to the level populations. It may be seen that the different assumptions lead to distinguishable results and that the reformation process makes a significant contribution to the populations of the lower rovibrational levels.

Because of the differences in their dynamical structure, J- and C-type shock waves present distinct signatures in the H_2 excitation diagram. In Fig. 13 are shown the predictions for a shock wave with the specified parameters at different evolutionary times. When the shock wave is ‘young’, it retains J-type characteristics, with relatively large column densities of levels with high excitation energy. As the shock wave evolves, it acquires a magnetic precursor and ultimately evolves into C-type, with much smaller column densities of high-excitation levels (and larger column densities of the low-excitation levels). As may be seen from the figure, at an intermediate age of $t = 1500$ year, when the shock structure is J-type with a

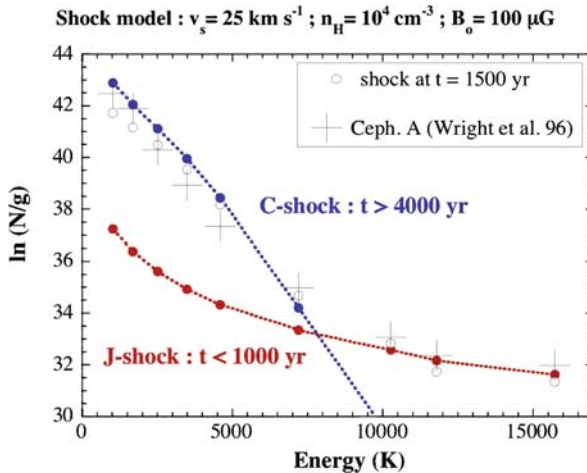


Fig. 13 The H_2 excitation diagram for Cep A West, observed by Wright et al. [18], compared with the predictions of shock model in which $u_s = 25 \text{ km s}^{-1}$ and $n_{\text{H}} = 10^4 \text{ cm}^{-3}$, $B_0 = 100 \text{ } \mu\text{G}$ in the preshock gas. The predictions of the model are shown for different evolutionary times, t . For $t < 1000$ year, the shock wave retains J-type characteristics, whereas, for $t > 4000$ year, it has become C-type. At $t = 1500$ year, its structure is J-type with a magnetic precursor

magnetic precursor, excellent agreement is obtained with the observations of the outflow source Cepheus A West, made with the infrared space observatory (ISO) satellite [18]. In this case, the lines observed are rotational transitions within the $v = 0$ vibrational ground state. Such comparisons confirm that outflows are dynamically young objects and even provide a means of estimating their evolutionary ages.

References

1. Field, G. B., et al.: Hydromagnetic shock waves and their infrared emission in H I regions. *Astrophys. J.* **151**, 953–975 (1968). 162, 173
2. Mullan, D. J.: The structure of transverse hydromagnetic shocks in regions of low ionization. *Mon. Not. Roy. Astron. Soc.* **153**, 145–170 (1971). 162
3. Hollenbach, D. J., McKee, C. F. Molecule formation and infrared emission in fast interstellar shocks I. Physical processes. *Astrophys. J. Suppl.* **41**, 555–592 (1979). 162
4. Draine, B. T.: Interstellar shock waves with magnetic precursors, *Astro-phys. J.* **241**, 1021–1038 (1980). 162
5. Smith, M. D., Mac Low, M.-M.: The formation of C-shocks: structure and signatures. *Astron. Astrophys.* **326**, 801–810 (1997). 162
6. Chièze, J. -P., et al.: Temporal evolution of MHD shocks in the interstellar medium. *Mon. Not. Roy. Astron. Soc.* **295**, 672–682 (1998). 162, 167
7. Lesaffre, P., et al.: Temporal evolution of magnetic molecular shocks I. Moving grid simulations. *Astron. Astrophys.* **427**, 147–155 (2004a). 162, 183
8. Lesaffre, P., et al.: Temporal evolution of magnetic molecular shocks II. Analytics of the steady state and semi-analytical construction of intermediate ages. *Astron. Astrophys.* **427**, 157–167 (2004b). 162, 183
9. Wardle, M., Draine, B. T.: Oblique magnetohydrodynamic shock waves in molecular clouds. *Astrophys. J.* **321**, 321–333 (1987). 165
10. Draine, B. T.: Multicomponent, reacting MHD flows, *Mon. Not. Roy. Astron. Soc.* **220**, 133–148 (1986). 165, 168
11. Richtmyer, R. D., Morton, K. W.: *Difference Methods for Initial-value Problems*. John Wiley & Sons, New York (1957). 166
12. Osterbrock, D. E.: On ambipolar diffusion in H I regions, *Astrophys. J.* **134**, 270–272 (1961). 169
13. Flower, D. R.: Momentum transfer between ions and neutrals in molecular gas. *Mon. Not. Roy. Astron. Soc.* **313**, L19–L21 (2000). 169
14. Draine, B. T., Sutin, B.: Collisional charging of interstellar grains. *Astrophys. J.* **320**, 803–817 (1987). 170
15. Flower, D. R.: *Molecular Collisions in the Interstellar Medium*, 2nd edn. Cambridge University Press, Cambridge (2007). 171
16. Snell, R. L., et al.: Detection of water in the shocked gas associated with IC 443: constraints on shock models, *Astrophys. J.* **620**, 758–773 (2005). 190
17. Flower, D. R., et al.: The contributions of J-type shocks to the H₂ emission from molecular outflow sources. *Mon. Not. Roy. Astron. Soc.* **341**, 70–80 (2003). 192
18. Wright, C. M., et al.: Molecular hydrogen observations of Cepheus A West, *Astron. Astrophys.* **315**, L301–L304 (1996). 193, 194

The Ionisation and Excitation State of Stellar Outflows

A.C. Raga

Abstract This chapter describes how to include ‘microphysics’ (i.e. the relevant physical and chemical processes) in dynamical models of stellar outflows. The gas dynamic equations (including energy gains and losses) and the equations for a system of atomic/ionic species are introduced. The calculation of the collisional rate coefficients (recombination and ionisation) is described in detail, and a summary of how to include photoionisation processes is also done. A quite detailed description of how to compute the cooling function is presented. Finally, a summary is made of how to compute the emitted spectrum.

1 Introduction

The course which covered the material in these lecture notes was meant as an introduction on how to include ‘microphysics’ (i.e. the relevant physical and/or chemical processes) in dynamical calculations of HH jets. At the more basic level, this has to be done in order to have a more or less realistic calculation of the radiative cooling term, which is fundamental for obtaining a correct description of the dynamics of the gas. A more detailed description of the microphysics is necessary if one aims at obtaining predictions (e.g. emission line maps and line profiles) that can be compared directly with observations.

These notes cover the basics of how to introduce atomic microphysics in numerical models of outflows, covering from the most simple possible approach to the level of complexity necessary for obtaining predictions of the emission line spectrum. The introduction of chemical species is not directly discussed, but many of the considerations and methods described in these notes can be applied with only minor modifications to chemical networks.

Many of the publications resulting from the Jetset project cover topics related to these notes. For example, the contribution of David Flower on the chemistry of outflows ([5], in these Lecture Notes) is very closely related to the topics covered

A.C. Raga (✉)

Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ap. 70-543,
04510 DF, México, raga@nucleares.unam.mx

below. The Ph.D thesis of Ovidiu Tesileanu [20] has an extended description of the treatment of ionisation and cooling appropriate for jet simulations. With respect to the dynamics of HH jets, Raga [12] gives an elementary introduction to the subject. Finally, the book of Aller [3] is a detailed guide on the calculation of the ionisation and excitation state of atoms and ions in the ISM.

2 The Gas dynamic Equations

The 3D, Cartesian equations of gas dynamics can be written in the following way: continuity equation \rightarrow

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j) = 0, \quad (1)$$

momentum equation (three components) \rightarrow

$$\frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j + P \delta_{ij}) = 0, \quad (2)$$

energy equation \rightarrow

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial x_j} [u_j (E + P)] = G - L, \quad (3)$$

with $E = \rho u_j u_j / 2 + P / (\gamma - 1)$ (γ is the specific heat ratio = 5/3 for our discussion of a non-relativistic, atomic/ionic gas). In these equations, the implicit Einstein summation notation is used. We have used the standard notation: ρ = density, u_i = i th component of the flow velocity, x_i = i th spatial coordinate, t = time and P = pressure.

In an analogous way, one could write the MHD equations, which have extra terms due to the Lorentz force, and also include the vector induction equation (which describes the evolution of the magnetic field). The rest of the discussion in these notes mostly applies also to the MHD equations.

In the energy equation, we have the terms G and L , which are the thermal energy gain and loss (respectively) per unit volume and time. These two terms are the root of all microphysical evil.

Both of these terms depend on the atomic/ionic structure of the gas. There is both an energy gain and an energy loss associated with each of the species (for example, HI, HII, HeI, HeII, HeIII, CI, CII, CIII, ...) present in the gas, so that we have to compute them as

$$G = \sum_a G_a; \quad L = \sum_a L_a, \quad (4)$$

where a is an index that numbers all of the species included in the model. In order to compute the G_a and L_a we need to know the value of the number densities n_a of all of the species. These number densities follow continuity equations of the form

$$\frac{\partial n_a}{\partial t} + \frac{\partial}{\partial x_j}(n_a u_j) = S_a, \quad (5)$$

where S_a is the source/sink term due to ionisations and recombinations (as well as charge exchange for some species). By combining Eq. (5) with Eq. (1) one can find the alternative set of equations

$$\frac{\partial y_a}{\partial t} + u_j \frac{\partial y_a}{\partial x_j} = \frac{S_a}{n}, \quad (6)$$

where $y_a = n_a/n$ is the fractional abundance of the species a (with $n = \rho/m$, m being the average mass per atom or ion).

Therefore, in order to solve the gas dynamic+microphysical problem one has to integrate the five gas dynamic Eq. (1), (2) and (3) together with the set of equations for the ionisation network (Eq. 5, or alternatively Eq. 6). The network equations are coupled with the gas dynamic equations through the radiative energy gain and loss terms (see Eq. 4).

3 The Ionisation State of the Gas

3.1 The Rate Equations

For the sake of simplicity, in this section we consider the case of a constant density gas, for which the equations for an atomic/ionic network are

$$\frac{dn_{a,z}}{dt} = S_{a,z}^c + S_{a,z}^{\text{ph}}, \quad (7)$$

with the collisional (S^c) and photoionisation (S^{ph}) source terms being given by

$$S_{a,z}^c = n_e [n_{a,z-1} c_{a,z-1} + n_{a,z+1} \alpha_{a,z+1} - n_{a,z} (c_{a,z} + \alpha_{a,z})], \quad (8)$$

$$S_{a,z}^{\text{ph}} = n_{a,z-1} \phi_{a,z-1} - n_{a,z} \phi_{a,z}. \quad (9)$$

The collisional source term (S^c) counts the contributions from the collisional ionisation of the ions (of atoms a) of charge $z - 1$ ($n_e n_{a,z-1} c_{a,z-1}$), the recombinations of $z + 1$ ions ($n_e n_{a,z+1} \alpha_{a,z+1}$) and the collisional ionisation ($n_e n_{a,z} c_{a,z}$) and recombination ($n_e n_{a,z} \alpha_{a,z}$) of z ions. The photoionisation source term (S^{ph}) counts the contributions from the photoionisation of $z - 1$ and z ions.

If the $n_{a,z}$ are known, the electron density can be computed as

$$n_e = \sum_a \sum_z z n_{a,z} . \quad (10)$$

3.2 The Collisional Rate Coefficients

The collisional ionisation (c) and recombination (α) coefficients are functions of temperature T given by integrals of the form

$$\alpha(T), c(T) = \int_0^\infty f(v, T) \sigma_v v dv , \quad (11)$$

where $f(v, T)$ is the Maxwell–Boltzmann distribution for the electrons and σ_v is the velocity-dependent collision cross section for the appropriate process. The recombination rate takes into account two processes: radiative recombination (in which the extra energy of the free electron is carried away by an emitted photon) and dielectronic recombination (in which the extra energy of the free electron is distributed as excitation energy of the inner electrons of a multi-electron ion).

It is common to give analytic fits to these coefficients in the ‘Arrhenius interpolation’ form:

$$r(T) = b_1 T^{b_2} e^{b_3/T} . \quad (12)$$

As an example of other formulae, Aldrovandi and Péquignot [1, 2] have used an interpolation

$$r(T) = b_1 \left(\frac{T}{10^4} \right)^{-b_2} + b_3 T^{-3/2} \exp(-b_4/T) [1 + b_5 \exp(-b_6/T)] , \quad (13)$$

for recombination coefficients.

Then, the b_1, b_2, \dots coefficients are tabulated for all of the ionisation and recombination processes that have to be considered. An example of such a tabulation (taken from [14]) is given below.

Finally, we should note that for some atoms and ions, in the collision source term $S_{a,z}^c$ (see Eq. (8)) one also has to include ‘charge exchange’ reactions with H. These are processes in which an atom or ion collides with HI or HII, and interchanges an electron during the collision with the H atom or ion. Examples of important charge exchange reactions are the processes $\text{HII} + \text{OI} \rightarrow \text{HI} + \text{OII}$ and $\text{HI} + \text{OII} \rightarrow \text{HII} + \text{OI}$ (see Table 1). These processes have rates of the form $n_{\text{HII}} n_{\text{OI}} \xi(T)$ and $n_{\text{HI}} n_{\text{OII}} \xi'(T)$, respectively, with the rate coefficients $\xi(T)$ and $\xi'(T)$ also given by Arrhenius interpolations (see Table 1).

Table 1 Ionisation, recombination and charge exchange coefficients

Reaction	Coefficients ^a
$e + \text{HI} \rightarrow 2e + \text{HII}$	1: 5.83×10^{-11} , 0.5, -157800
$e + \text{HII} \rightarrow \text{HI}$	1: 3.69×10^{-10} , -0.79, 0
$e + \text{HeI} \rightarrow 2e + \text{HeII}$	1: 2.707×10^{-11} , 0.5, -285400
$e + \text{HeII} \rightarrow \text{HeI}$	2: 4.3×10^{-13} , 0.672, 0.0019, 4.7×10^5 , 0.3, 94000
$e + \text{HeII} \rightarrow 2e + \text{HeIII}$	1: 5.711×10^{-12} , 0.5, -631000
$e + \text{HeIII} \rightarrow \text{HeII}$	1: 2.21×10^{-9} , -0.79, 0
$e + \text{CII} \rightarrow 2e + \text{CIII}$	1: 3.93×10^{-11} , 0.5, -283000
$e + \text{CIII} \rightarrow \text{CII}$	2: 3.2×10^{-12} , 0.770, 0.038, 9.1×10^4 , 2.0, 3.7×10^5
$e + \text{CIII} \rightarrow 2e + \text{CIV}$	1: 2.04×10^{-11} , 0.5, -555600
$e + \text{CIV} \rightarrow \text{CIII}$	2: 2.3×10^{-12} , 0.645, 7.03×10^{-3} , 1.5×10^5 , 0.5, 2.3×10^5
$e + \text{NI} \rightarrow 2e + \text{NII}$	1: 6.18×10^{-11} , 0.5 -168200
$e + \text{NII} \rightarrow \text{NI}$	2: 1.5×10^{-12} , 0.693, 0.0031, 2.9×10^5 , 0.6, 1.7×10^5
$e + \text{NII} \rightarrow 2e + \text{NIII}$	1: 4.21×10^{-11} , 0.5, -343360
$e + \text{NIII} \rightarrow \text{NII}$	2: 4.4×10^{-12} , 0.675, 0.0075, 2.6×10^5 , 0.7, 4.5×10^5
$e + \text{OI} \rightarrow 2e + \text{OII}$	1: 1.054×10^{-10} , 0.5, -157800
$e + \text{OII} \rightarrow \text{OI}$	2: 2.0×10^{-12} , 0.646, 0.0014, 1.7×10^5 , 3.3, 5.8×10^4
$e + \text{OII} \rightarrow 2e + \text{OIII}$	1: 3.53×10^{-11} , 0.5, -407200
$e + \text{OIII} \rightarrow \text{OII}$	2: 3.1×10^{-13} , 0.678, 0.0014, 1.7×10^5 , 2.5, 1.3×10^5
$e + \text{OIII} \rightarrow 2e + \text{OIV}$	1: 1.656×10^{-11} , 0.5, -636900
$e + \text{OIV} \rightarrow \text{OIII}$	2: 5.1×10^{-12} , 0.666, 0.0028, 1.8×10^5 , 6.0, 91000
$e + \text{SII} \rightarrow 2e + \text{SIII}$	1: 7.12×10^{-11} , 0.5, -271440
$e + \text{SIII} \rightarrow \text{SII}$	2: 1.8×10^{-12} , 0.686, 0.0049, 1.2×10^5 , 2.5, 88000
$\text{HI} + \text{NII} \rightarrow \text{HII} + \text{NI}$	1: 1.1×10^{-12} , 0, 0
$\text{HII} + \text{NI} \rightarrow \text{HI} + \text{NII}$	1: 4.95×10^{-12} , 0, -10440
$\text{HI} + \text{OII} \rightarrow \text{HII} + \text{OI}$	1: 2.0×10^{-9} , 0, 0
$\text{HII} + \text{OI} \rightarrow \text{HI} + \text{OII}$	1: 1.778×10^{-9} , 0, -220

^aThe interpolation formulae are of the form 1: Arrhenius or 2: Aldrovandi and Péquignot [1], see Eq. (12) and (13)

3.3 The Photoionisation Rates

The photoionisation rates are calculated as a function of the average intensity J_ν of the radiative field through the frequency integrals

$$\phi_{a,z} = \int_{\nu_{a,z}}^{\infty} \frac{4\pi J_\nu}{h\nu} \sigma_{a,z}(\nu) d\nu, \quad (14)$$

where ν is the frequency, $\nu_{a,z} = \chi_{a,z}/h$ is the frequency of the ionisation edge and $\sigma_{a,z}(\nu)$ is the photoionisation cross section. Here again, one can find tabulations of coefficients of power law interpolations for the $\sigma_{a,z}(\nu)$ (see, e.g. any of the versions of the book of Osterbrock).

A more complex process for photoionisation of HI, important in regions of the jets close to the outflow sources, also takes place. This ionisation occurs in two stages: first an $n = 1 \rightarrow 2$ excitation (from the ground state of H to the first excited level, through the absorption of a Ly α photon or through a collision with a free electron), then followed by a photoionisation of the $n = 2$ electron. This $n = 2$

photoionisation requires photons of only 25% of the energy necessary to photoionise electrons directly from the $n = 1$ level.

The problem of course is that one has to solve a system of radiative transfer equations of the form

$$\frac{dI_v}{dl} = j_v - \kappa_v I_v \quad (15)$$

for the specific intensity I_v (j_v and κ_v being the emission and absorption coefficients, respectively). This is a system of equations, since one has to solve this for many propagation directions (l being the distance element along a given direction) and frequencies.

After integrating these radiative transfer equations, one can then carry out the appropriate angular average:

$$J_v = \frac{1}{4\pi} \oint I_v d\Omega \quad (16)$$

in order to obtain the average intensity of the radiative field.

Even though it is common practice to include photoionisation processes in 1D shock models (a practice started by Raymond [18]), very little work has been done in multi-dimensional simulations. There is the sole paper of Raga et al. [16] in which the ‘diffuse’ ionising photon field (i.e. the ionising photons produced by the shocked flow itself) was considered in a 3D jet model, and there are a few papers in which an external ionising photon field was introduced (in order to model HH jets within HII regions, see, e.g. Raga and Reipurth [13]). The more detailed descriptions of how to introduce photoionisation in multi-dimensional simulations have been given by Mellema et al. [8, 9].

However, for models of jets in neutral or molecular regions, the photoionisation processes are only of secondary importance and are normally not included in the models. Basically, it is normally assumed that CII and SII do not recombine to their neutral states because of the presence of enough photons below the Lyman limit within the region, and the problem of photoionisation is left at this. In these lecture notes, the complex problem of photoionisation is abandoned here and now.

3.4 Coronal Ionisation Equilibrium

If the temperature is kept constant, the ionisation state of the gas evolves to the point where $S_{a,z} = 0$ for all i and z . This condition (see Eq. (8)) gives a system of linear equations which are equivalent to the system:

$$n_{a,z} c_{a,z} = n_{a,z+1} \alpha_{a,z+1} \cdot \quad (17)$$

This equation reflects the simple statement that in coronal ionisation equilibrium the collisional ionisations from $z \rightarrow z + 1$ are balanced by recombinations from $z + 1 \rightarrow z$.

Because c and α depend only on T , this system gives ionisation fractions $y_{a,z} = n_{a,z}/n_a$ (where $n_a = \sum_z n_{a,z}$) which are exclusively a function of T . This result holds if charge exchange reactions are included. The resulting ionisation fractions $y_{a,z}^c(T)$ are called the ‘coronal ionisation equilibrium’ ionisation state.

At high densities, the coronal ionisation equilibrium condition (Eq. (17)) has to be modified by including the so-called ‘three-body recombination’ process (in which one electron recombines with an ion, and a second electron carries away the required energy difference). This process is the one that allows the system of Eq. (17) to approach the Saha, thermodynamic equilibrium ionisation distribution. However, densities above 10^9 cm^{-3} are necessary for three-body recombination to become a relevant process.

Clearly, if we have an ionisation fraction $y_{a,z} > y_{a,z}^c(T)$, $y_{a,z}$ will evolve to a lower value with time, and the reverse is true for $y_{a,z} < y_{a,z}^c(T)$.

As an example, let us consider the coronal ionisation equilibrium for H. Equation (17) is

$$n_{\text{HI}}c(T) = n_{\text{HII}}\alpha(T). \quad (18)$$

We can combine this equation with $n_{\text{H}} = n_{\text{HI}} + n_{\text{HII}}$ to obtain

$$y_{\text{HII}} = \frac{n_{\text{HII}}}{n_{\text{H}}} = \frac{1}{1 + \alpha(T)/c(T)}, \quad (19)$$

where the Arrhenius interpolations for the coefficients are $\alpha(T) = 3.69 \times 10^{-10} T^{-0.79}$ and $c(T) = 5.83 \times 10^{-11} T^{0.5} e^{-157800/T}$ (see Eq. (12) and the first two lines of Table 1). The hydrogen ionisation fraction y_{HII} as a function of temperature T obtained from Eq. (19) is shown in Fig. 1.

Life would be easy if the gas in shocked interstellar flows were in coronal ionisation equilibrium, as one could then make a tabulation of ionisation fractions as a function of temperature and use this to obtain a tabulation of the cooling function as a function of density and temperature. However, the timescale for relaxation to coronal equilibrium is comparable to the cooling timescale of the flow, and therefore in many regions the ionisation state of the gas is well away from the coronal equilibrium state. Because of this, in order to obtain an accurate cooling function it is indeed necessary to integrate continuity equations for all of the species relevant for the cooling term (see Eq. (5)).

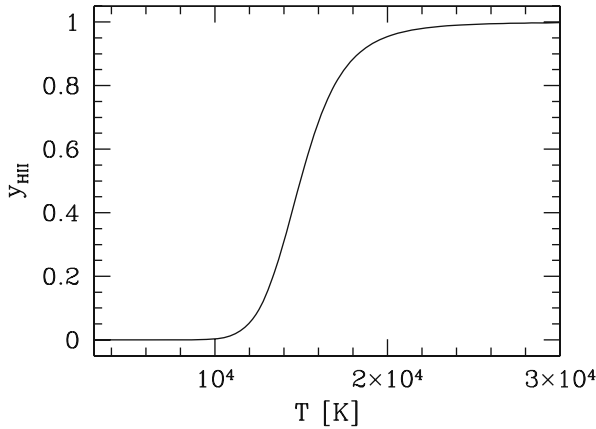


Fig. 1 Coronal ionisation fraction of H as a function of temperature

4 The Cooling Function

4.1 Introduction

In this section, we describe in some detail how to include the more important contributions to the cooling due to different processes. Actually, the cooling is dominated by collisional ionisation of HI and by collisional excitation of emission lines. We describe the radiative recombination and free–free losses only because it is very simple to include them, though they do not make an important contribution to the cooling function (at least for a gas with solar abundances) for temperatures lower than $\sim 10^8$ K. For temperatures above $\sim 10^8$ K, the cooling function is actually dominated by the free–free losses.

4.2 Recombination and Free–Free Cooling

When an electron passes by an ion and recombines, all of the kinetic energy of the electron is lost from the thermal energy reservoir. Analogously, when an electron loses kinetic energy in an inelastic collision (with the resulting emission of a photon), the thermal energy of the gas is reduced. The resulting energy losses are called the ‘recombination’ and ‘free–free’ cooling rates, respectively.

The free–free energy loss (per unit time and volume) due to the interaction of H ions and free electrons is given by

$$L_{\text{ff}}(HII) = n_e n_{\text{HII}} \beta_{\text{ff}}(T), \quad (20)$$

where the interpolation formula

$$\beta_{\text{ff}}(z, T) = 1.846 \times 10^{-27} z^2 T^{1/2} \quad (21)$$

can be used (see the book of Osterbrock). In this interpolation function, one has to set a charge $z = 1$ for HII. One can use the same function (i.e. with $z = 1$) for calculating the free-free losses due to HeII ions (this of course being only an approximation because HeI is not a hydrogenic ion) and the function with $z = 2$ for HeIII ions.

The radiative losses due to recombination of HII are given by

$$L_{\text{rec}}(\text{HII}) = n_e n_{\text{HII}} \beta_{\text{rec}}(T), \quad (22)$$

where the interpolation formula

$$\beta_{\text{rec}}(t) = 1.133 \times 10^{-24} t^{-1/2} (-0.0713 + 0.5 \ln t + 0.640 t^{-1/3}), \quad (23)$$

with $t = 157890/T$ (see Seaton [19]). The contribution of the recombination of He ions can be computed with the scaling

$$\beta_{\text{rec}}(z, T) = z \beta_{\text{rec}}(1, T/z^2), \quad (24)$$

with $z = 1$ for HeII and $z = 2$ for HeIII.

4.3 Collisional Ionisation

The energy loss due to collisional ionisation of the ion a , z can be written as

$$L_{a,z}^{\text{ion}} = n_e n_{a,z} c_{a,z}(T) \chi_{a,z}, \quad (25)$$

where $c_{a,z}(T)$ is the collisional ionisation coefficient and $\chi_{a,z}$ the ionisation potential of the ion a , z .

The terms that dominate the collisional ionisation cooling are the ionisation of HI, HeI and HeII. These terms dominate the cooling function for a neutral gas that is suddenly shocked to a temperature above a few times 10^4K .

4.4 Collisionally Excited Lines

For each atom or ion that contributes substantially to the cooling function, one has to solve a system of equations giving the populations of the excited levels n_l ($l = 1, 2, \dots, N$ numbering in order of increasing energy all of the relevant levels) of the ion a , z . Of course, we have the relation

$$n_{a,z} = \sum_{l=1}^N n_l. \quad (26)$$

Once we have computed the n_l populations, we can compute the energy loss due to the collisional excitation of all of the relevant levels of the ion a, z as

$$L_{a,z}^{\text{col}} = \sum_{l=1}^N n_l \sum_{m < l} A_{l,m} h\nu_{l,m}, \quad (27)$$

where $A_{l,m}$ is the Einstein A coefficient for the spontaneous transition $l \rightarrow m$ and $h\nu_{l,m}$ is the energy associated with this transition.

The non-trivial problem of finding the populations n_l of the excited levels of course still remains. Because the relaxation time for the excitation/de-excitation of the levels is generally much shorter than the cooling and/or dynamical timescales of the flow, the calculation of the n_l is usually done under a statistical equilibrium assumption. The equilibrium condition results in the set of equations:

$$\sum_{m > l} n_m A_{m,l} + n_e \sum_{m \neq l} n_m q_{m,l}(T) = n_l \left[\sum_{m < l} A_{m,l} + n_e \sum_{m \neq l} q_{l,m}(T) \right], \quad (28)$$

where we have introduced the $q_{m,l}(T)$ which are the radiative excitation ($m < l$) or de-excitation ($m > l$) coefficients. The left-hand side of this equation represents the number of radiative ($n_m A_{m,l}$) and collisional ($n_e n_m q_{m,l}$) transitions into state l per unit volume and time. The right-hand side represents the radiative and collisional transitions from level l into all of the other levels of the ion or atom.

For $m > l$, the $q_{m,l}(T)$ coefficients are given by the expression

$$q_{m,l}(T) = \frac{8.629 \times 10^{-6}}{T^{1/2}} \frac{\Omega_{ml}(T)}{g_m}, \quad (29)$$

where g_m is the statistical weight of the level at which the transition begins. The collisional excitation coefficients (i.e. $l \rightarrow m$ with $m > l$) are given by the relation

$$q_{l,m}(T) = \frac{g_m}{g_l} e^{-h\nu_{m,l}/kT} q_{m,l}(T). \quad (30)$$

The function $\Omega_{ml}(T)$ has a value of order 1 and is only slowly dependent on T . In many calculations, these ‘collision strengths’ are considered to be constant and taken from tabulations such as the classical one of Mendoza [10]. However, in our modern electronic world there is the database of the Arcetri/Cambridge/NRL ‘CHI-ANTI’ atomic database collaboration (<http://www.arcetri.astro.it/science/chianti/database/>) which has temperature-dependent interpolations of the collision strengths and the Einstein A coefficients of all of the transitions I have ever known.

We now proceed as follows. We assume that the A and q coefficients are known, as well as the temperature T (obtained, e.g. from a dynamical simulation) the electron density n_e and the ionic number density $n_{a,z}$ (both obtained from a solution of the non-equilibrium ionisation rate equations). We can then invert the system of linear Eq. (26) and (28) in order to find the populations n_l ($l = 1 \rightarrow N$) of the excited levels of the ion a, z .

Let us show how the equations look for a three-level atom ($N = 3$). Equation (26) takes the form

$$n_1 + n_2 + n_3 = n_{a,z} . \quad (31)$$

Now, for $l = 1$, from Eq. (28) we obtain

$$n_1 [-n_e(q_{12} + q_{13})] + n_2(A_{21} + n_e q_{21}) + n_3(A_{31} + n_e q_{31}) = 0 , \quad (32)$$

and for $l = 2$, we obtain

$$n_1(n_e q_{12}) + n_2 [-A_{21} - n_e(q_{23} + q_{21})] + n_3(A_{32} + n_e q_{32}) = 0 . \quad (33)$$

For $l = 3$, we obtain an equation that is a linear combination of Eq. (32) and (33).

Now, the system of three linear Eq. (31), (32) and (33) can be inverted in order to find n_1 , n_2 and n_3 as a function of $n_{a,z}$, n_e and T .

4.5 The Two-Level Atom

Let us now consider a two-level atom ($N = 2$). Equation (26) takes the form

$$n_{a,z} = n_1 + n_2 , \quad (34)$$

and Eq. (28) takes the form

$$n_1 n_e q_{12} = n_2 (n_e q_{21} + A_{21}) . \quad (35)$$

The left-hand side of this equation is the rate of transitions into level 2 (the only possible route being collisional excitations up from level 1) and the right-hand side represents the rate of transitions from 2 to 1 (with both collisional and radiative transitions).

Equations (34) and (35) can be combined with Eq. (30) to obtain

$$n_2 = \frac{n_{a,z}}{(g_1/g_2)e^{E_{21}/kT} + 1 + n_c/n_e} , \quad (36)$$

where the critical density is defined as $n_c \equiv A_{21}/q_{12}$. Then, for $n_e \ll n_c$ (the ‘low-density regime’), we have

$$n_2 = \frac{n_{a,z} n_e q_{12}}{A_{21}}, \quad (37)$$

and for $n_e \gg n_c$ (the ‘high-density regime’), we have

$$n_2 = \frac{n_{a,z} g_2 e^{-E_{21}/kT}}{g_1 + g_2 e^{-E_{21}/kT}}, \quad (38)$$

which is the Boltzmann distribution (LTE) population of level n_2 .

The critical density n_c therefore represents the density at which the transition between the low- and high-density regimes occurs. For atoms or ions with many excited levels, it is possible to define critical densities for the different transitions, which represent the densities around which the associated emission lines have the low/high-density regime transition.

The energy loss associated with the transition between levels 2 and 1 is

$$L_{21} = n_2 A_{21} h \nu_{21}, \quad (39)$$

which for the low-density regime then takes the form

$$L_{21} = n_{a,z} n_e q_{12} h \nu_{21}, \quad (40)$$

and for the high-density regime becomes

$$L_{21} = n_2 (LTE) A_{21} h \nu_{21}, \quad (41)$$

These equations illustrate the well-known fact that the collisionally excited line cooling function depends quadratically on the density (namely, on the product $n_e n_{a,z}$) for the ‘low-density regime’ and linearly for the ‘high-density regime’.

5 Recipes for Calculating the Cooling Function

5.1 Isochoric and Coronal Ionisation Cooling Functions

The most simple way to include a cooling function is to include in the energy Eq. (3) either the coronal ionisation equilibrium or the isochoric cooling.

The coronal ionisation equilibrium cooling is calculated as a function of density and temperature from the ionisation fractions calculated from the equilibrium Eq. (17). The cooling function is normally calculated for an atom+ion number density of 1 cm^{-3} , and the ‘user’ scales it to higher (or lower) densities by assuming that it is in the ‘low-density regime’ (in which the cooling is proportional to the square of the density). As can be seen from tabulations of cooling functions for different ions (see Raga et al. 1997), low-density regime cooling can be safely assumed up to densities of $\sim 10^5 \text{ cm}^{-3}$.

The isochoric cooling function is usually computed by considering a parcel with a time-independent density (of atoms+ions, usually of 1 cm^{-3}) which cools from a given initial temperature (typically 10^6 K). The non-equilibrium ionisation rate equations (7) are then integrated in time together with an energy equation of the form

$$\frac{3}{2} \frac{d}{dt} (n + n_e) kT = -L, \quad (42)$$

where n and n_e are the atom+ion and electron (respectively) number densities (n being time-independent, but not n_e). From the integration one obtains T and L as a function of time, and then one plots the radiative energy loss as a function of temperature in order to obtain the so-called isochoric cooling function.

The coronal equilibrium and isochoric cooling functions are tabulated in Table 2 (taken from De Colle [4]). These tables can be introduced in a code (with an appropriate interpolation in $\log T$ in order to obtain a continuous function) and multiplied by the square of the number density of atoms+ions in order to obtain a cooling function to be used in the energy equation.

As far as I am aware, there are no publications yet describing the differences that are obtained in the dynamics of the flow when using one or the other of these cooling functions. Also, there is no discussion in the literature of how different are the resulting flow structures when one introduces a more accurate, many species non-equilibrium treatment of the ionisation and cooling.

5.2 Tabulations and Interpolation Functions

A detailed description of how to introduce an ionisation network with enough species so as to give a realistic cooling function is given by Raga et al. [17]. These authors tabulate the cooling functions for a number of atoms and ions as a function of temperature and number density, and one can carry out an appropriate interpolation to obtain the numerical values of the cooling function for the desired densities and temperatures.

A reduced description of the cooling is proposed by Raga et al. [14, 15], who give a series of interpolation formulae which reproduce the low-density regime of the ionic cooling functions of Raga et al. [17] For example, the interpolations of the cooling due to the OI and OII ions are

$$\log_{10} \left(\frac{L_{\text{OI}}}{n_e n_{\text{OI}}} \right) = -45.0 + 1.23t_1 + 0.5t_1^{10} + 1.2t_2 + 1.2 [\max(t_2, 0)], \quad (43)$$

$$\log_{10} \left(\frac{L_{\text{OII}}}{n_e n_{\text{OII}}} \right) = -47.3 + 7.9t_3 + 1.9 \frac{t_4}{|t_4|^{0.5}}, \quad (44)$$

with $t_1 = 1 - 100/T$, $t_2 = 1 - 10^4/T$, $t_3 = 1 - 2000/T$, $t_4 = 1 - 5 \times 10^4/T$.

Table 2 Coronal ionisation equilibrium and isochoric cooling functions¹

$\log_{10} T$ [K] [erg]	$\log_{10} L_{ce}$ [cm ⁻³]	$\log_{10} L_{iso}$ [s ⁻¹]
3.9106	-26.0784	-25.5891
4.0006	-24.0161	-23.7582
4.0282	-23.5181	-22.8345
4.0601	-23.0615	-22.5186
4.0956	-22.6889	-22.3633
4.2008	-22.1165	-22.2669
4.4103	-22.0975	-22.2346
4.6313	-22.1336	-22.2015
4.7165	-22.0483	-22.1604
4.9465	-21.5183	-22.0718
5.1332	-21.2607	-21.7326
5.3338	-21.0870	-21.4906
5.4942	-21.0875	-21.2639
5.6010	-21.1587	-21.1075
5.7289	-21.2863	-21.0778
5.8089	-21.3616	-21.1810
5.9180	-21.4217	-21.3489
5.9809	-21.4289	-21.4227

¹The coronal equilibrium and isochoric cooling functions are given for an atom+ion number density of 1 cm⁻³

5.3 Cooling Function Including a Single Species Ionisation Network

A possibility first suggested by Hartigan and Raymond [6] is to solve a single ionisation rate equation for HII (or for HI) and assume that $n_{\text{OII}}/n_{\text{O}} = n_{\text{HII}}/n_{\text{H}}$ (which approximately holds due to the similarity of the ionisation potentials of OI and HI, and the strong charge exchange reaction which couples O and H). Then, the cooling function can be computed including the collisional ionisation of H (see Eq. (25)) and the cooling due to collisional excitation of OI and OII (Eq. (43) and (44)).

Such a cooling function is appropriate only for temperatures below $\sim 2 \rightarrow 4 \times 10^4 \text{K}$ (as in this temperature regime the cooling rate is indeed dominated by OI and OII). Therefore, one has to incorporate a switch to either the coronal or the isochoric cooling function for higher temperatures.

6 Calculation of the Emitted Spectrum

6.1 General Considerations

Once we have calculated a dynamical model of an outflow (using a cooling function and an ionisation network such as the ones described above), we can proceed to compute the emission from the model flow. Typically, one wants to compute the emission in a set of lines which have been observed.

6.2 The Emission Coefficient

If we know the number density n_2 of the excited level which gives rise to a given transition $2 \rightarrow 1$, the total line emission coefficient (integrated over all frequencies within the line profile) can be computed as

$$j_{21}^{(L)} = \frac{n_2 A_{21} h \nu_{21}}{4\pi}. \quad (45)$$

The emission coefficient as a function of frequency ν is

$$j_{21}(\nu) = \frac{j_{21}^{(L)}}{\sqrt{\pi} \Delta \nu_D} e^{-(\nu - \nu_{21})^2 / \Delta \nu_D^2}, \quad (46)$$

where $\Delta \nu_D = \nu_{21} v_T / c$ with $v_T = \sqrt{2kT/m_a}$ (m_a being the mass of the element which gives rise to the line). Using the standard Doppler effect, one can obtain the emission coefficient as a function of radial velocity v_r (along the line of sight):

$$j_{21}(v_r) = \frac{j_{21}^{(L)}}{\sqrt{\pi} v_T} e^{-v_r^2 / v_T^2}. \quad (47)$$

Equation (47) can be derived from Eq. (46) by considering that $v_r/v_T = (\nu - \nu_{21})/\Delta \nu_D$ and that $d\nu = (\Delta \nu_D/v_T)dv_r$.

6.3 Optically Thin Lines

For an optically thin emission line, one can obtain an emission map by integrating $j_{21}^{(L)}$ (see Eq. (45)) along lines of sight:

$$I_{21}(x, y) = \int j_{21}^{(L)}(x, y, z) dz, \quad (48)$$

where (x, y) are the plane-of-the-sky coordinates and z lies along the line of sight. In order to compare the predicted emission map with an observed map, a convolution with a simulated ‘seeing profile’ or ‘point spread function’ might be necessary.

Also, maps at a given radial velocity can be computed by integrating along lines of sight the radial velocity profile (Eq. 47) of the line:

$$I_{21}(v_r, x, y) = \int j_{21}(v_r - v_r^{(f)}, x, y, z) dz, \quad (49)$$

where $v_r^{(f)}(x, y, z)$ is the component along the line of sight of the velocity of the flow at position (x, y, z) .

From Eq. (49), one can compute line profiles at a given position (x, y) on the plane of the sky. In order to compare with an observation, a convolution with the spectral response function of the spectrograph might be necessary, as well as a plane-of-the-sky spatial integration or convolution in order to simulate the effect of the seeing and of the width (and/or length) of the spectrograph slit.

Also, one can carry out an integration of (49) over given radial velocity bins in order to obtain ‘radial velocity channel maps’ that can be directly compared with Fabry–Pérot or with 2D field spectroscopic observations (again, a convolution with a ‘seeing function’ might be appropriate).

6.4 Optically Thick Lines

For an optically thick line, one has to first calculate the optical depth to the position z along the line of sight:

$$\tau_{21}(v_r, x, y, z) = \int_{-\infty}^z \kappa_{21}(v_r - v_r^{(f)}, x, y, z') dz', \quad (50)$$

where κ_{21} is the absorption coefficient:

$$\kappa_{21}(v_r, x, y, z) = n_2(x, y, z) \left(\frac{g_2}{g_1} \right) \frac{c^2 A_{21}}{8\pi^{3/2} v_{21}^2 v_T} e^{(-v_r/v_T)^2}, \quad (51)$$

where c is the velocity of light, and v_T is the spatially dependent thermal velocity (see the text following Eq. (46)).

One should point out that for an optically thick line the population of the excited level (n_2 in Eq. (51)) will be affected by the absorption of the photons of the emission line. This effect produces a coupling between the radiative transfer and the calculation of the level populations.

The emission line map at a given radial velocity v_r can then be calculated as

$$I_{21}(v_r, x, y) = \int j_{21}(v_r - v_r^{(f)}, x, y, z) \exp[-\tau_{21}(v_r, x, y, z)] dz. \quad (52)$$

This is the kind of treatment necessary to calculate the P-Cygni profiles which are obtained for stellar winds. The calculation of optically thick lines is of course a problem which has been developed in detail in the field of stellar atmospheres [7].

Somewhat more simple is the inclusion of the effect of dust extinction, because in this case the absorption coefficient has no radial velocity dependence (being approximately constant across the width of an emission line).

7 Summary

When one computes models of stellar outflows, one has to include a cooling function. This can be done in several different degrees of approximation:

- using a tabulated coronal ionisation equilibrium or isochoric cooling function (Sect. 5.1),
- including one extra continuity equation for HI or for HII (Sect. 5.3), and
- including several atoms/ions (Sect. 5.2).

These three approaches give cooling functions which differ from each other by relatively large factors, and the effect of using the more simplified cooling functions should be evaluated quantitatively for the flow which one wants to simulate.

In order to obtain a prediction of the emitted spectrum, one has to have a prediction of the ionisation state of the gas. This can again be done by taking the coronal equilibrium or isochoric ionisation fractions (which would be consistent for a model calculated with the corresponding cooling functions), or by using the non-equilibrium ionisation state which is obtained from a full gas dynamic+ionisation network simulation.

One then has to compute the populations of the excited levels giving rise to the lines of interest and compute the corresponding emission coefficients (Sect. 6.2). In this course we have given the details of how to compute the populations of levels excited by collisions (Sect. 4.4), but we have not given a description of how to do this calculation for levels populated by a recombination cascade (this can be found, e.g. in the book of Osterbrock [11]). This latter case of course applies for the H Balmer lines. With the emission line coefficients one can then carry out predictions of intensity maps, line profiles and radial velocity channel maps (Sect. 6.3) which can be compared directly with observations.

References

1. Aldrovandi, S. M. V., Péquignot, D.: *A&A*. **25**, 137 (1973). 198, 199
2. Aldrovandi, S. M. V., Péquignot, D.: *A&A*. **47**, 321 (1976). 198
3. Aller, L. H.: *Physics of thermal gaseous nebulae* (Reidel) (1984). 196
4. De Colle, F.: private communication (2007). 207
5. Flower, D.: *These Lecture Notes* (2009). 195
6. Hartigan, P., Raymond, J. C.: *ApJ*. **409**, 705 (1993). 208
7. Mihalas, D.: *Stellar Atmospheres* (Freeman) (1978). 210
8. Mellema, G., et al.: *NewA*. **11**, 374 (2006). 200
9. Mellema, G., et al.: *A&A*. **331**, 335 (1988). 200
10. Mendoza, C.: In *Planetary Nebulae* (A83-49126 24-89). Reidel, Dordrecht, p. 143 (1983). 204
11. Osterbrock, D. E.: *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*. Univ. Sci. Books, Mill Valley (1989). 211
12. Raga, A. C.: *Astroph. Sp. Sci.* **208**, 163 (1993). 196
13. Raga, A. C., Reipurth, B.: *RMxAA*. **40**, 15 (2004). 200
14. Raga, A. C., et al.: *A&A*. **465**, 879 (2007). 198, 207
15. Raga, A. C., et al.: *A&A*, **392**, 267 (2002). 207

16. Raga, A. C., et al.: RMxAA. **35**, 123 (1999). 200
17. Raga, A. C., et al.: ApJS. **109**, 517 (1997). 207
18. Raymond, J. C.: ApJS. **39**, 1 (1979) 200
19. Seaton, M. J.: MNRAS. 119, 81 (1959) 203
20. Tesileanu, O.: PhD Thesis (Univ. de Torino) (2008) 196

Deriving Physical Diagnostics from Observations

C. Dougados, F. Bacciotti, S. Cabrit, and B. Nisini

Abstract We review in this course diagnostics for the physical conditions in the atomic component of the flow (electronic densities n_e , electronic temperatures T_e and hydrogen ionization fraction x_e) based on the most prominent optical and near-infrared forbidden emission lines. We discuss both diagnostics independent of the excitation process and methods based on the comparison with radiative shock models. We then detail the different techniques used to derive jet mass-loss rates, an important parameter for launching models, and their associated uncertainties. Finally, we describe important biases introduced by projection and convolution effects which can critically affect the translation of observed quantities into meaningful physical quantities of the flow.

1 Introduction

The purpose of this course is to review the main methods and associated biases used to derive gas physical quantities from emission line observations in the context of protostellar jet studies. We will concentrate here on jet physical parameters critical to constrain launching models.

We will specifically focus on diagnostics for the atomic component of the flow based on the most prominent optical and near-infrared forbidden emission lines. These optically thin and spatially resolved tracers allow to derive model-independent physical parameters for the inner regions of the jets. Other important jet tracers not treated here include near-and mid-infrared H_2 lines (see [33] and D.

C. Dougados (✉)

Laboratoire d'Astrophysique de Grenoble, Catherine.Dougados@obs.ujf-grenoble.fr

F. Bacciotti

INAF-Osservatorio Astrofisico di Arcetri, fran@astro.arcetri.it

S. Cabrit

LERMA, UMR 8112 du CNRS, Observatoire de Paris, sylvie.cabrit@obspm.fr

B. Nisini

INAF-Osservatorio Astrofisico di Roma, nisini@oa-roma.inaf.it

Flower this book), permitted emission lines such as the ones from atomic hydrogen which require radiative transfer modelling, mid- and far-infrared emission lines such as [O I] λ 63 μ m (also treated in [33]), radio millimetric and centimetric emissions (see [11]) and X-ray emission (see [35]).

This chapter is organized as follows. We first discuss model-independent diagnostics for plasma excitation conditions (electronic densities n_e , electronic temperatures T_e and hydrogen ionization fraction x_e) based on the analysis of the strong optical emission lines of O⁰, S⁺, N⁺ and the infrared emission lines of Fe⁺. The derivation of these jet quantities is critical to constrain mass-loss rates, an important parameter for launching models, and identify the main heating and ionization sources. The second section describes an alternative method to derive hydrogen ionization fractions and jet densities by comparison of the radiative properties of shock models with observed line emission. Indeed, shocks are likely the main heating and ionization mechanisms in young jets. We discuss in the third section the different methods used to derive mass-loss rates for atomic jets and their associated uncertainties. Finally, we describe in the last section important biases introduced by projection and convolution effects, which affect the translation of observed quantities into physically meaningful ones.

2 Model-Independent Diagnostics of Plasma Excitation Conditions

We describe in this section methods for deriving plasma conditions (electronic density n_e , electronic temperature T_e and hydrogen ionization fraction x_e) without any a priori assumptions on the nature of the heating and ionization mechanisms (shocks, ambipolar diffusion, turbulent mixing layers, etc.). The main hypothesis underlying the different diagnostics discussed below are the following:

- The lines are collisionally excited, i.e. photo-excitation and photoionization processes are neglected. This assumption may not be valid in the very inner regions of the wind (see below for a more detailed discussion of photoionization processes),
- Only collisions with electrons are considered since they are an order of magnitude more efficient than with H atoms in partially ionized atomic plasma,
- The emission is assumed optically thin (no radiative transfer effects),
- Excitation conditions are assumed homogeneous within an elementary emissivity volume element (the smallest spatial resolution element of current observations is on the order of $0.1'' = 14$ AU at $d = 140$ pc, the distance of the Taurus SFR),
- The level population of atoms is given by statistical equilibrium: for the studied optical and near-IR transitions the timescales to establish statistical equilibrium are 10^4 – 10^7 s (for n_e in the range 10^2 – 10^5 cm⁻³). This is smaller than the typical dynamical timescale for the flow to cross the elementary resolution element ($\text{tdyn} > 7 \times 10^6$ s for flow velocities < 300 km sv⁻¹ in a $0.1''$ pixel at $d = 140$ pc).

- S and Fe atoms are assumed fully singly ionized, while N and O are in ionization equilibrium with H. We will see below that this is achieved in the low-excitation conditions typically observed in young stellar jets.

2.1 Forbidden Line Emission: Emissivity and Critical Density

We first recall some fundamental concepts and definitions regarding forbidden emission lines (for a more detailed description see [24, 33]). In the case of the multi-level atom, the emissivity of an optically thin, collisionally excited, emission line resulting from the transition between levels j (upper) and i (lower) of the atom X in the ionization state X^k is given by

$$\begin{aligned}\epsilon_{ji} &= n_j A_{ji} h\nu_{ji} = n_H \left(\frac{n_X}{n_H}\right) \left(\frac{n_{X^k}}{n_X}\right) \left(\frac{n_j}{n_{X^k}}\right) A_{ji} h\nu_{ji} \\ &= f(n_H, [X/H], [X^k/X], n_e, T_e),\end{aligned}$$

where A_{ji} is the Einstein coefficient for spontaneous radiative de-excitation, n_j the number volume density of atoms X^k in the level j and $h\nu_{ij} = E_j - E_i$ with E_i the excitation energy of level i . The line emissivity therefore depends on the elemental abundance ($[X/H] = \frac{n_X}{n_H}$), the ionization state of element X ($[X^k/X] = \frac{n_{X^k}}{n_X}$) and the excitation state of element X^k ($\frac{n_j}{n_{X^k}}$), where n stands for the number volume density.

The excitation state is derived from the assumption of statistical equilibrium and collisionally dominated excitation with electrons. Taking into account a multi-level atom, the statistical equilibrium for level j reads as

$$n_j \left(\sum_{i \neq j} n_e q_{ji} + \sum_{i < j} A_{ji} \right) = \sum_{i \neq j} n_e n_i q_{ij} + \sum_{i > j} A_{ij} n_i, \quad (1)$$

where the sums are performed over all contributing levels i .

For a Maxwellian distribution of velocities, the collisional de-excitation coefficient q_{ji} is given by

$$q_{ji} = \frac{8.63 \times 10^{-6} \Omega_{ji}(T_e)}{g_j T_e^{1/2}} \text{ cm}^3 \text{ s}^{-1} \quad (2)$$

with Ω_{ji} ($=\Omega_{ij}$) the collision strength of the transition and g_j the statistical weight of level j . Collision strengths are dimensionless and typically of order unity. The relationship between the excitation and de-excitation collisional coefficients is given by the detailed balance equation

$$\frac{q_{ij}}{q_{ji}} = \frac{g_j}{g_i} e^{-(E_j - E_i)/kT} \quad (3)$$

with E_i the excitation energy of level i .

The critical (electronic) density of level j is defined by

$$n_{\text{cr}}(j) = \frac{\sum_{i < j} A_{ji}}{\sum_{i \neq j} q_{ji}}. \quad (4)$$

Note that the critical density depends on the electronic temperature (via the collisional coefficients). The critical density defines two regimes for the emissivity of the $j - i$ line.

- For electronic densities much larger than the critical density of the upper level j ($n_e \gg n_{\text{cr}}$, *high-density regime*), collisional processes dominate over radiative processes on the left-hand side of Eq. (1) and the population number density of level j approaches its local thermodynamic equilibrium (LTE) value:

$$\frac{n_j}{n} \simeq \frac{g_j e^{-E_j/kT}}{\sum g_i e^{-E_i/kT}}, \quad (5)$$

where $n = \sum n_i$ is the total number density of element X^k and E_i is the excitation energy of level i . In this limit, the level population does not depend on n_e and the line emissivity ϵ_{ji} increases linearly with the total density n_{H} ($\epsilon_{ji} \propto n_j \propto n_{\text{H}} f(T_e)$).

- For electronic densities much lower than the critical density ($n_e \ll n_{\text{cr}}$, *low-density regime*), collisional processes from level j are negligible and the population number density of level j can be approximated as

$$\frac{n_j}{n} \simeq \frac{n_e}{A_{ji}} \sum_{i \geq j} q_{0i}, \quad (6)$$

where q_{0i} is the collisional excitation coefficient to state i from the ground state. Now $\epsilon_{ji} \propto n_e n_{\text{H}} f(T_e) \propto x_e n_{\text{H}}^2 f(T_e)$, and the line emissivity increases as the square of the total density.

Critical densities for the main optical and near-IR forbidden emission lines investigated in jet studies are given in Table 1. These lines will probe emitting regions with n_e in the range $10^3 - 10^6 \text{ cm}^{-3}$.

2.2 Electronic Density Diagnostics

Electronic density diagnostics are obtained from line ratios involving transitions from the same element, in the same ionization state and originating from upper

Table 1 Critical densities

Line	$n_{\text{cr}} (T_e = 10^4 \text{K})$
[Fe II]1.644, 1.257 μm	$5.6 \times 10^4 \text{ cm}^{-3}$
[Fe II]1.533 μm	$4.6 \times 10^4 \text{ cm}^{-3}$
[O I] λ 6300 \AA	$1.8 \times 10^6 \text{ cm}^{-3}$
[S II] λ 6731 \AA	$3.5 \times 10^3 \text{ cm}^{-3}$
[N II] λ 6583 \AA	$1.2 \times 10^5 \text{ cm}^{-3}$

levels of similar excitation energies to minimize the effect of electronic temperatures but different critical densities.

An example of such a density-sensitive line ratio is the [S II] λ 6716 \AA /[S II] λ 6731 \AA ratio. These two lines involve transitions between the $^2D_{3/2}$ and $^2D_{5/2}$ terms and the ground level (see Fig. 1). If we denote by 1 the ground level of both transitions and by 2 and 3 the upper levels of the [S II] λ 6716 \AA and [S II] λ 6731 \AA transitions, respectively, the ratio of these two lines is given by

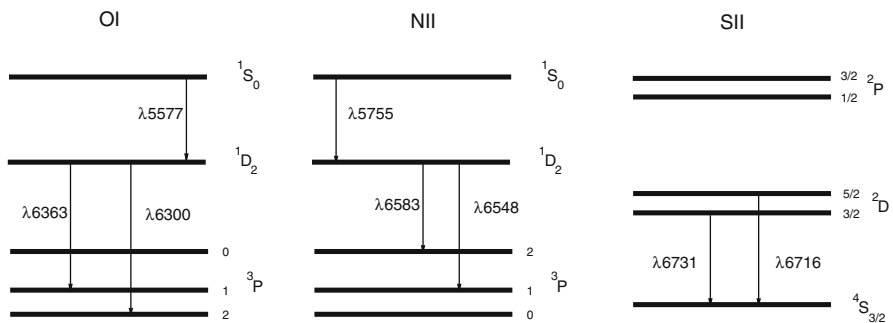
$$\frac{\epsilon_{31}}{\epsilon_{21}} = \frac{A_{31}}{A_{21}} \times \frac{h\nu_{31}}{h\nu_{21}} \times \frac{n_3}{n_2}, \quad (7)$$

with n_2 and n_3 the population number densities of levels $^2D_{3/2}$ and $^2D_{5/2}$, respectively. Levels 2 and 3 having similar excitation energies, the derived ratio will depend mostly on n_e and little on T_e .

Figure 3 shows the computed variation of the [S II] ratio as a function of n_e , taking into account the first five levels of the S^+ atom with excitation temperatures less than $2 \times 10^4 \text{K}$.

Another very useful density diagnostic is the ratio involving the strong near-infrared lines of [Fe II]1.533 and 1.644 μm (Fig. 2). For excitation temperatures less than $2 \times 10^4 \text{K}$, Fig. 4 shows that a simplified model involving the first 16 levels of the Fe^+ atom gives a very good approximation of the level population distribution. This ratio provides a good diagnostic for n_e in the range 10^3 – 10^5 cm^{-3} .

[S II] and near-IR [Fe II] line ratios trace electron densities up to 10^4 and 10^5 cm^{-3} , respectively. Higher densities can be probed by additional ratios of lines

**Fig. 1** Energy diagrams of the main optical forbidden emission lines. Adapted from [28]

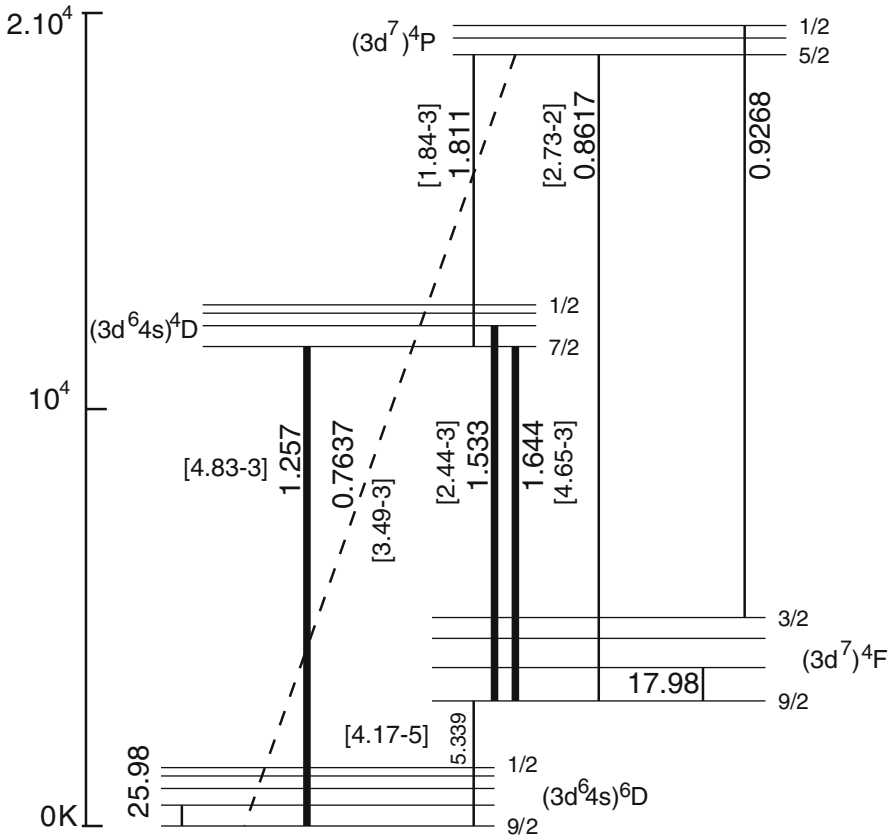


Fig. 2 Energy diagrams of the main [Fe II] emission lines. Adapted from [37]

falling in the optical regime. In particular, the ratio between the two red lines [FeII] $\lambda 7155\text{\AA}$ and 8617\AA is sensitive to electron densities in the range $n_e \sim 10^5\text{--}10^7\text{ cm}^{-3}$, as shown in [10, 22]. This ratio has been recently employed on several T Tauri and Class I to evidence densities larger than 10^6 cm^{-3} at the base of their jets [22, 32, 39]. Another interesting tracer of high-density gas is represented by the ratio between the permitted CaII $\lambda 8542\text{\AA}$ and the forbidden [CaII] $\lambda 7291\text{\AA}$ transitions. Although fluorescence may be in general important in the population of the levels giving rise to these transitions, in HH objects the dominant excitation mechanism appears to be through collisions [22, 32]. In such a case, their ratio is sensitive to electron densities larger than 10^6 cm^{-3} (see Fig. 5).

2.3 Temperature Diagnostics

Electron temperatures can be measured through the ratio of lines of the same element with different excitation energies. In the optical, sensitive ratios are those

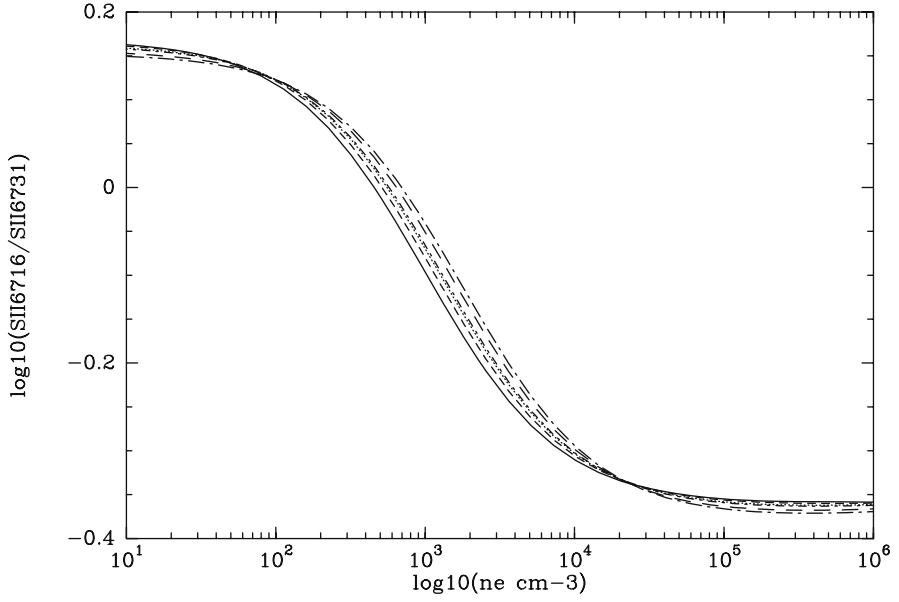


Fig. 3 Variation of the [S II] $\lambda 6716\text{\AA}/[\text{S II}]\lambda 6731\text{\AA}$ line ratio as a function of n_e . The different curves correspond to T_e varying between 7×10^3 and 2×10^4 K. Adapted from [28]

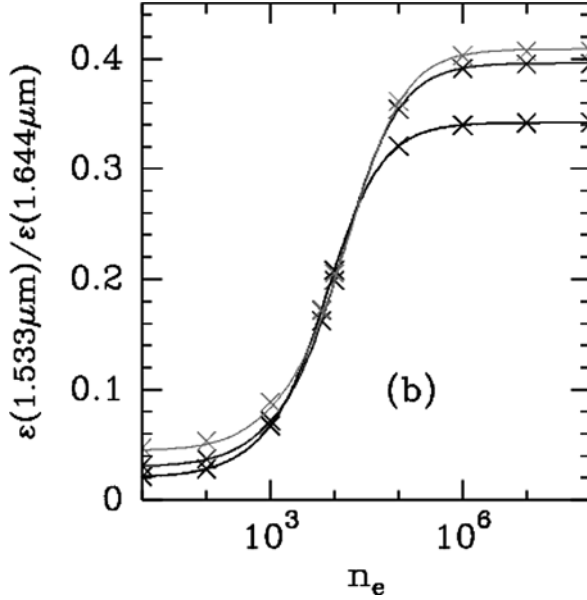


Fig. 4 Variation of the [Fe II] $\lambda 1.533\text{ }\mu\text{m}/[\text{Fe II}]\lambda 1.644\text{ }\mu\text{m}$ line ratio as a function of n_e . The different curves correspond to T_e values of 3,000, 10^4 and 2×10^4 K. Adapted from [37]

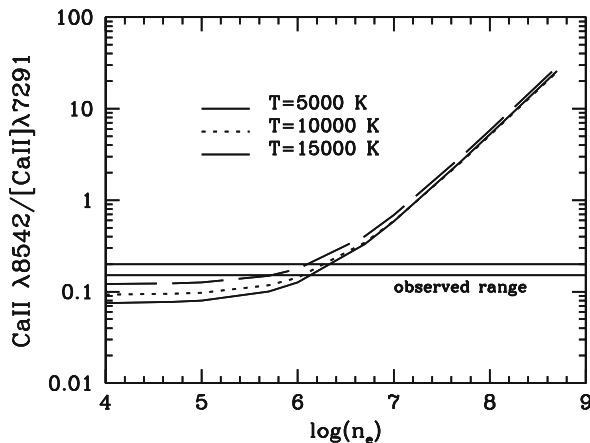


Fig. 5 Plot of the high-density tracer ratio $\text{Ca II } \lambda 8542 / [\text{Ca II}] \lambda 7291 \text{ \AA}$, calculated assuming collisional excitation. The solid lines indicate the range of values observed in the HH 1 jet ([32])

between nebular transitions (those connecting the 1D_2 with the 3P terms) and auroral transitions (connecting the 1S_0 and 1D_2 terms) in elements having a 2p or 4p configuration (cf. [24]). Examples of these ratios are $[\text{OI}] \lambda (6300+6363) \text{ \AA} / [\text{OI}] \lambda 5577 \text{ \AA}$ or $[\text{N II}] \lambda (6548+6583) \text{ \AA} / [\text{N II}] \lambda 5755 \text{ \AA}$ (see Fig. 6). These ratios can in principle probe a large range of temperatures, between 5000 and 20,000 K. However, the nebular lines are much stronger than auroral lines in jet conditions, thus their ratio can be very large (in the range 100 – 1000) and difficult to be observed.

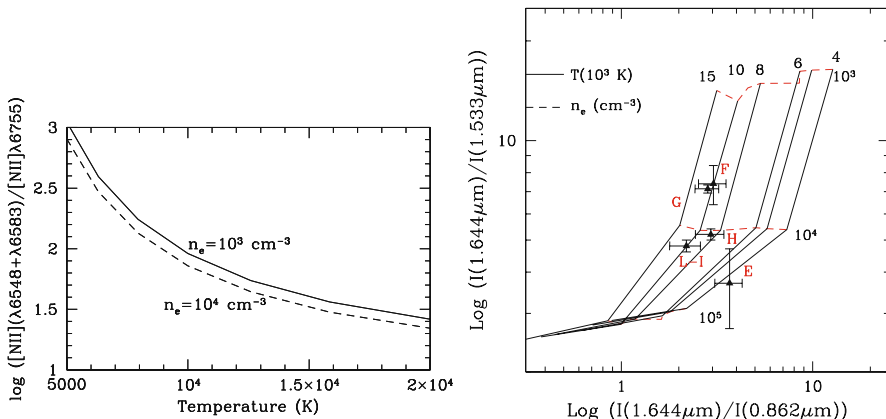


Fig. 6 *Left*: intensity ratio of optical [N II] transitions as a function of temperature and for two values of electron density, computed adopting the approximated analytical solution given in [36]. *Right*: diagnostic diagrams for n_e and T_e determination based on different [Fe II] lines. On the *right*, a diagram constructed combining the near-IR density-sensitive ratio $1.64 \mu\text{m}/1.53 \mu\text{m}$ and the temperature-sensitive ratio between the $1.64 \mu\text{m}$ and the optical 8620 \AA line. The symbols indicate the line ratios observed in the HH 1 jet knots (from [32]).

Also in the rich spectrum of Fe^+ there are several potential transitions that can be used as temperature tracers. To this aim it is, for example, possible to combine one of the near-IR transitions from the 4D term ($T_{\text{ex}} \sim 11\,000\text{K}$), with one of the transitions from the 4P term, excited at $\sim 19\,000\text{K}$. These latter transitions, however, either fall in regions of poor atmospheric transmission (those connecting the 4P – 4D terms) or fall in the optical range (the 4P – 4F transitions). An important temperature tracer is, for example, given by the ratio $[\text{Fe II}]1.644\,\mu\text{m}/[\text{Fe II}]8620\text{\AA}$ that probes temperatures ranging from ~ 4000 to $\sim 15\,000\text{K}$ (e.g. [37, 32]), which use requires to combine lines observed with different instrumentation.

Through multi-wavelength observations of lines of the same element it is possible to combine ratios sensible to density variations with ratios sensible to temperature variations, constructing diagnostic diagrams from which one can simultaneously derive the two physical parameters. An example of such a kind of diagrams is given in Fig. 6, constructed with the IR and optical $[\text{Fe II}]$ diagnostic lines.

2.4 Deriving the Ionization Fraction from Optical Line Ratios: The BE99 Technique

As explained above, there are several possibilities to derive in a relatively simple way the electron density in stellar jets. The knowledge of n_e , however, is not sufficient to determine the dynamical properties of the flow, because the bulk of the moving material might be only partially ionized or even predominantly neutral. To effectively compare observations and theoretical models one actually needs to know the total hydrogen density n_H . The total density cannot be retrieved directly from the observations, but it can be derived from the electron density if one has an estimate of the hydrogen ionization fraction $x_e = n_e/n_H$. To find x_e observationally, however, is not trivial. In this paragraph we examine the difficulties connected with the determination of the ionization fraction in jets, and we describe an original method, the so-called *BE technique*, that allows one to overcome these problems and derive x_e from commonly observed emission lines.

In many astrophysical situations the ionization fraction is computed via the so-called ‘coronal equilibrium’ [36], in which the collisional ionization of hydrogen is balanced by recombination at a given temperature. In jets, however, one cannot apply this method, because at the derived electron densities the hydrogen recombination time is about the same as the total transit time of a gas particle along the bright jet beam ($t_{\text{rec}} \sim 1/n_e\alpha(H, T) \sim 10^3$ year, where $\alpha(H, T) \sim 3 \times 10^{-13}\text{cm}^3\text{s}^{-1}$ for $T_e \sim 10^4\text{K}$). Therefore, once it is created, the hydrogen ionization decouples from the thermal conditions of the gas, ‘freezing’ at its initial value no matter the temperature variations of the gas, and it is subsequently carried along the jet beam in a process of slow recombination (see [4]).

It follows that the hydrogen ionization has to be computed in another way. A method applied by several authors in the past has been the comparison with grids of

shock models that simulate the line intensities [42, 19, 20] (see the next section). In this way the hydrogen ionization fraction was indirectly inferred by the best fit to the line fluxes. This method, however, is extremely complex and implies a number of assumptions on the physical mechanism that produces the gas heating (i.e., a shock of given velocity and geometry, with a given pre-shock density and pre-shock magnetic field, etc.) which is something that one would rather avoid when trying to establish the nature of the observed emission.

A simple and powerful technique to derive x_e and T_e from the ratio of optical lines of S^+ , O^0 and N^+ observed in jets was first presented in [4] and subsequently refined in [5]. The method, referred to as the ‘BE’ technique in the jet community, was built upon the recognition of the fact that charge exchange between O and N with H is the dominant mechanism in the ionization balance of these species. This process, in fact, is favoured by the circumstance that the ionization potential of N and O is very close to that of hydrogen ($E_{\text{ion}} = 13.6$ eV for H, 13.61 for O and 14.54 for N). In addition, even if the hydrogen ionization fraction is not in equilibrium with the local temperature, it can be considered at any time in equilibrium with the N and O ionization fractions, because the processes that regulate the ionization balance of N and O are much faster than those that regulate the ionization balance of hydrogen. Thus, x_e could be derived indirectly from the ionization of O and N, which in turn could be inferred from the observed forbidden lines from these species.

The method just assumes that the gas is collisionally excited, making no hypothesis regarding the heating agent. This is clearly an advantage if the results are used to validate a given thermal model. The range of applicability of the technique, however, is limited to the case of low-excitation conditions, which applies in the beam section of stellar jets. In this case, all sulphur can be assumed to be singly ionized and nitrogen and oxygen not more than once ionized. These assumptions are confirmed by the fact that S^{++} , N^{++} and O^{++} lines are never detected in the spectra of HH jets (with the exception, sometimes, of strong bow-shocks at the head of the jet).

The ratios used in the technique are $[O\text{ I}]\lambda 6300\lambda\lambda(6300 + 6363)\text{\AA} / N\text{ II}\lambda\lambda(6548 + 6583)\text{\AA}$ and $[S\text{ II}]\lambda 6731\lambda\lambda(6716 + 6731)\text{\AA} / [O\text{ I}]\lambda 6300\lambda\lambda(6300 + 6363)\text{\AA}$. These depend not only on the electron density and temperature, but also on the ionization fractions and the abundances of the involved species.

As mentioned above, due to its low ionization potential, sulphur can be assumed to be all singly ionized in the jet beams. Thus, to compute N^+/N^0 and O^+/O^0 one can consider only the processes of collisional ionization, the simple and dielectronic recombination and charge exchange with hydrogen. The latter mechanism, in particular, provides a mean to connect the observed ratios with the hydrogen ionization fraction. In fact, the relationship between x_e and N^+/N^0 , O^+/O^0 is given by the following equations:

$$\frac{O^+}{O^0} = \frac{x_e(C(O, T) + \delta_O(T))}{x_e(\alpha(O, T) - \delta'_O(T)) + \delta'_O(T)}, \quad (8)$$

$$\frac{N^+}{N^0} = \frac{x_e(C(N, T) + \delta_N(T))}{x_e(\alpha(N, T) - \delta'_N(T)) + \delta'_N(T)}, \quad (9)$$

where $C(O, T)$, $C(N, T)$ are the collisional ionization rates for oxygen and nitrogen, $\alpha(O, T)$, $\alpha(N, T)$ are the direct plus dielectronic recombination rates, $\delta_O(T)$, $\delta_N(T)$ and $\delta'_O(T)$, $\delta'_N(T)$ are the direct and inverse charge exchange ionization rates, respectively.

Thus, assuming a given abundance set (see below for a discussion of this point), the considered line ratios, written for simplicity as $R_1 = [S \text{ II}]\lambda 6731\text{\AA}/[O \text{ I}]\lambda 6300\text{\AA}$ and $R_2 = [O \text{ I}]\lambda 6300\text{\AA}/[N \text{ II}]\lambda 6583\text{\AA}$, can be calculated numerically by solving the ionization balance Eqs. (8), (9) and the system of equations for the population of the levels in statistical equilibrium.

The ratios are functions of the electron density, n_e , temperature, T_e , and hydrogen ionization fraction, x_e . Once n_e is determined from the sulphur doublet, both R_1 and R_2 are functions of x_e and T_e only. It follows that in the (x_e, T_e) parameter space the values of R_1 and R_2 are found on two surfaces, and the observed values of the ratios will define a curve on each surface. The intersection between the two curves

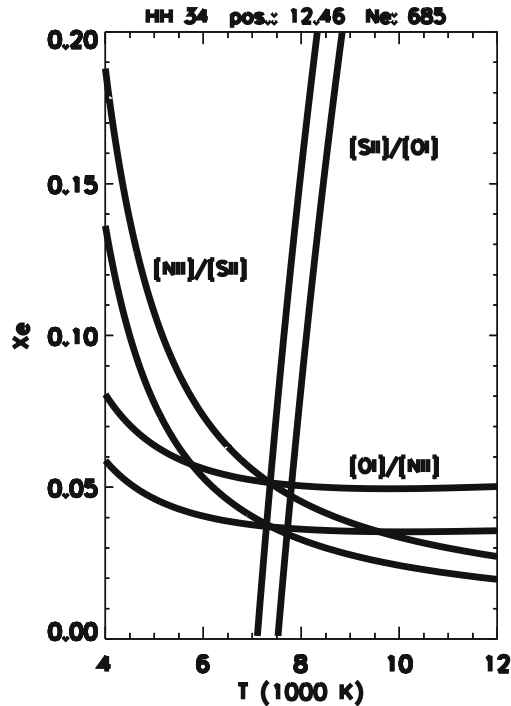


Fig. 7 Example of an application of the BE technique from [5]: the diagnostic diagram for the position at 12.5'' along the HH 34 jet beam. Each stripe defines the loci of the (x_e, T_e) values for which the predicted line ratio equals the observed one $\pm 1\sigma$ rms error. The intersection of the crossing stripes provides $T_e = 7550 \pm 250\text{K}$ and $x_e = 0.044 \pm 0.007$

projected onto the (x_e, T_e) plane will then indicate the values that these physical quantities must have to produce the observed line ratios. If the errors of the measured ratios are included, the curves will widen into two ‘stripes’ in the (x_e, T_e) plane, providing also an estimate of the error that affects x_e and T_e .

This concept is illustrated graphically in Fig. 7, taken from [5]. Here the crossing of the ‘stripes’ indicate the values of x_e and T_e (with the derived errors) for one position in the HH 34 jet.

It can be noticed that while R_1 depends on both T_e and x_e , R_2 is essentially only sensitive to x_e variations. The BE technique is thus an optimal tool to investigate the ionization of the flow while giving also an estimate of the gas temperature. In practice, however, the calculation of the physical quantities from the line ratios is calculated with an automated inversion code that allows one to investigate big data sets in a limited time. In this code, the level populations are determined with a five-level model for the atoms, using transition probabilities and coefficients for electron collisions from [30], while the ionization fractions for O and N are found using collisional ionization and radiative recombination rates from [27], dielectronic recombination rates from [27, 34] and charge exchange rates from [26, 2].

The method has been applied successfully to the study of the emission along the beams of several HH jets, unresolved across their width. A typical diagnostic result of this type is illustrated in Fig. 8 for the HH 111 jet. The application of the method has shown that jets are only partially ionized, with average x_e values between 0.01 and 0.6. The variation of n_e and x_e along the flow *is consistent with ionization freezing close to the source*, followed by slow non-equilibrium recombination. The mechanism that produces a high ionization at the base of the jet, however, is not known, although it seems to have similar efficiencies in all cases: Lower ionization fractions are found in the densest jets, i.e. those that recombine faster. In any event, the realization that stellar jets are only partially ionized has provided new, accurate estimates of the total density $n_H = n_e/x_e$. These estimates are much larger than previously thought and typically, on large scales, range from 10^3 to 10^5 cm^{-3} . The total density estimate, in turn, allows one to derive the jet mass, energy and momentum fluxes that are critical parameters to understand the relationship between the jet and its environment, the jet launching mechanism and the properties of the flow propagation.

Finally, it is worth mentioning that being the procedure fast and easy to apply (with respect, for example, to a grid of shock models), it turns out to be well suited for the analysis of the large data sets provided by high angular resolution observations. Application of the technique to spectra of jets taken with the hubble space telescope can be found, for example, in [6, 8, 31]. In particular, running the diagnostic code on input 3D datacubes taken at high spatial *and* spectral resolution with the space telescope imaging spectrograph (STIS) it is possible to investigate the physical conditions of interesting regions of the flows in separated velocity intervals, as illustrated in Fig. 9 for the first 200 AU of the jet from the T Tauri star DG Tau.

HH 111 DIAGNOSTICS

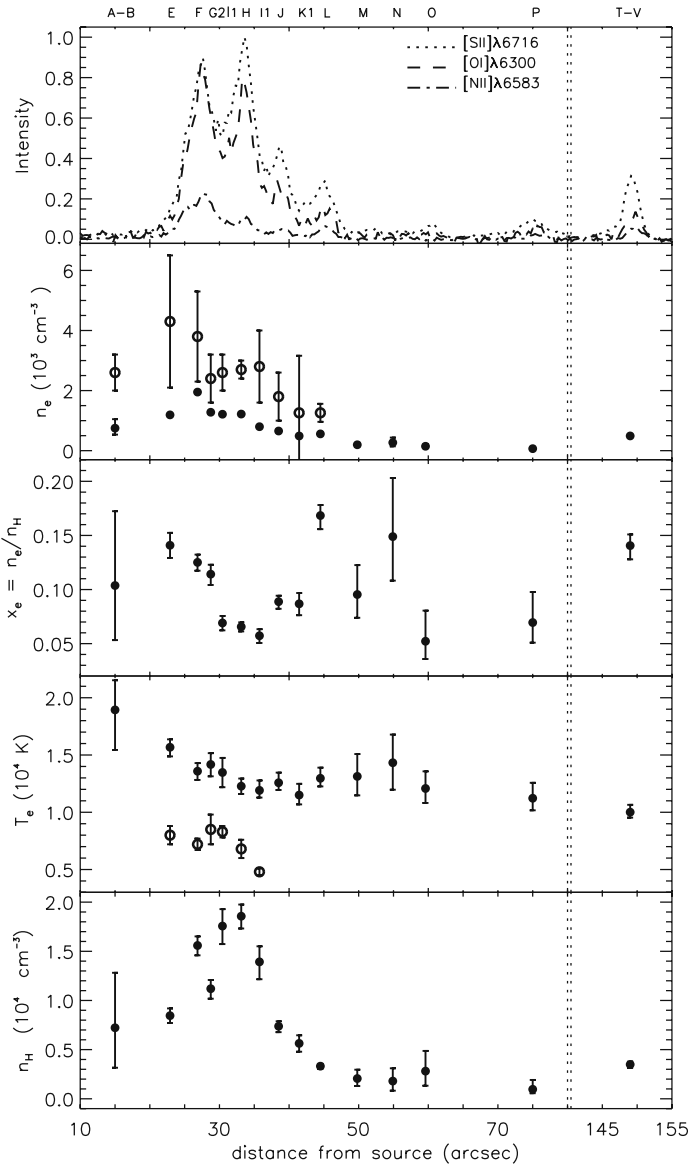


Fig. 8 Physical conditions along the HH 111 jets diagnosed with the BE technique, taken from [39]. From *top to bottom*: intensity profiles of the optical lines, the electronic density, n_e , in units of 10^3 cm^{-3} , the ionization fraction, x_e , the temperature, T_e , in units of 10^4 K and the total density, n_H , in units of 10^4 cm^{-3} . The *open circles* are the values derived from the [Fe II] lines, while the *filled circles* are parameters inferred from the optical S⁺, N⁺ and O⁰ lines using the BE technique. The zero point of the spatial scale is the driving source HH 111 IRS

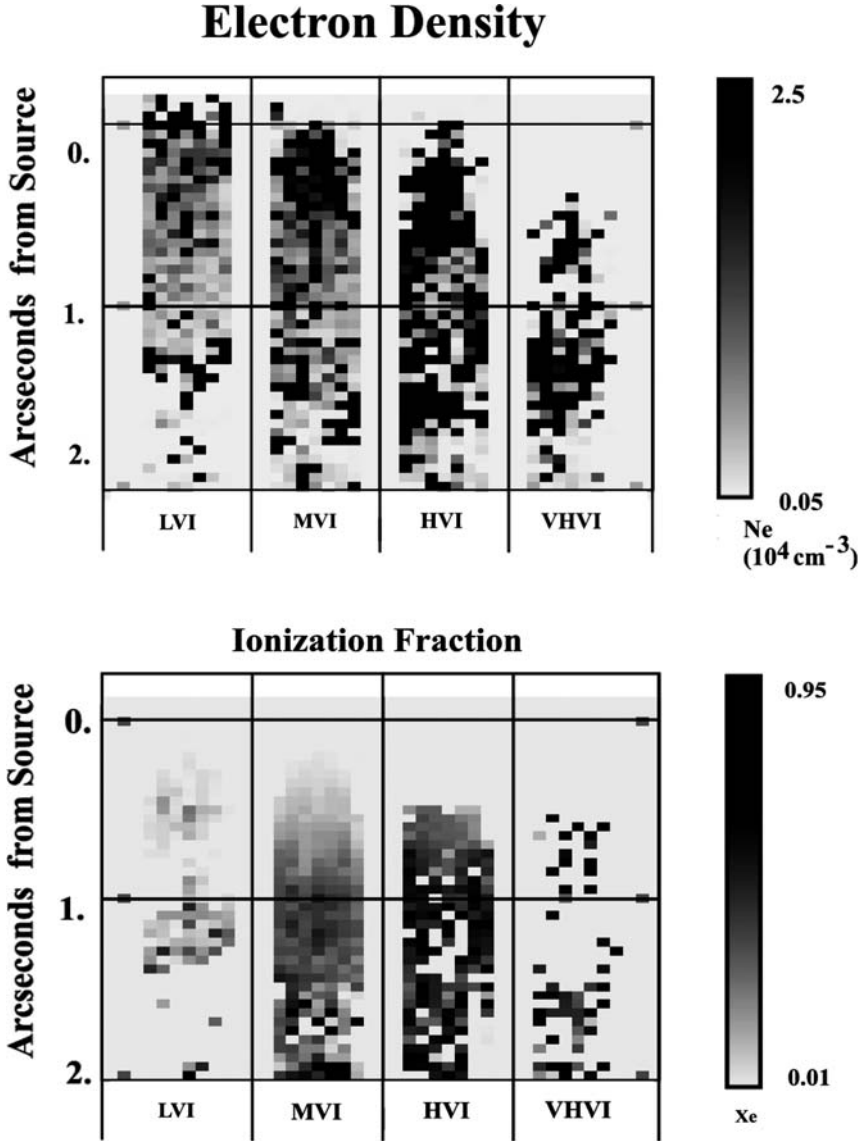


Fig. 9 Electron density and ionization fraction derived with the BE diagnostics for the first 200 AU of the jet from DG Tau in four different velocity channels. The input high angular resolution data were obtained from HST/STIS multi-slit optical spectra. The low, medium, high and very high velocity intervals are approximately, for DG Tau, from $[+60, -70 \text{ km s}^{-1}]$ (LV), $[-70, -195 \text{ km s}^{-1}]$ (MV), $[-195, -320 \text{ km s}^{-1}]$ (HV), and $[-320, -450 \text{ km s}^{-1}]$ (VHV), respectively [41].

2.5 Main Uncertainties and Limitations

The sensibility of the BE diagnostic results on the adopted parameters (e.g. abundances, atomic parameters) and on the observational uncertainties (extinction, critical densities, etc.) has been analysed in BE99, [6, 28]. These studies show that in average signal-to-noise conditions the accuracy is generally very good for the estimates of n_e ($\leq 3\%$), far from the [S II] $\lambda 6731$ critical density ($\sim 2 \times 10^4 \text{ cm}^{-3}$), while for T_e and x_e the accuracy is about 15–20% and 20–30%, respectively. Lavalley-Fouquet [28], using independent routines, also investigates the reliability of the predictions of the BE technique as the conditions of applicability progressively fail to hold. In the following paragraphs we discuss in some detail the influence of photoionization, excitation by shocks, choice of the abundance set and reddening.

2.5.1 Uncertainties Introduced by Neglecting Photoionization

One important limitation of the technique is that photoionization is not included in computing the ionization fraction of N and O. This simplification was suggested by the fact that in the case of HH jets, the central source, which is a low-mass young star, produces only mildly energetic photons that are not capable to ionize N and O. Bacciotti and Eistöffel [5] and other authors, e.g. [28, 23] checked carefully the validity of this approximation and showed that it is justified as long as x_e is lower than 0.8. Ionizing photons, however, can be produced in the strong bow-shocks at the ‘head’ of the jet, where the flow impacts the interstellar medium. The typical mean free path of ionizing photons in the jet gas, however, is $\sim 10^{16} \text{ cm}$, i.e. one order of magnitude smaller than the typical distance of the bow-shocks from the jet beam where the diagnostic technique is usually applied. Nevertheless, the inclusion of photoionization should be regarded among the first priorities among the future developments of the diagnostics, together with the treatment of higher ionization states of the considered elements. This will allow one to analyze the physical conditions in the highly excited terminal bow-shocks, as well as to examine objects of different nature, as peculiar regions of planetary nebulae or the ‘irradiated jets’ recently discovered in the Orion nebula.

2.5.2 Uncertainties Introduced by Shock Structure

The derivation of jet physical parameters from line ratios assumes that the observed region of the jet is filled with gas at a single excitation condition. However, this is usually not true as one is observing the cooling zone behind a shock front, having strong spatial gradients in density, ionization and temperature. This zone is only few AU in length, for shock velocities between 35 and 70 km s^{-1} and a pre-shock density of 10^3 cm^{-3} [20], and thus it is generally not resolved by the observations. As a consequence, biases in the parameters derived from integrated line ratios may be introduced in a diagnostic that use lines emitted in regions at different excitation.

Figure 10 plots the emission profiles (normalized to their peak value) in the cooling region of a 70 km s^{-1} shock, of a number of optical and IR diagnostic lines. The profiles of x_e , T_e and n_e used to compute the intensity profiles are taken

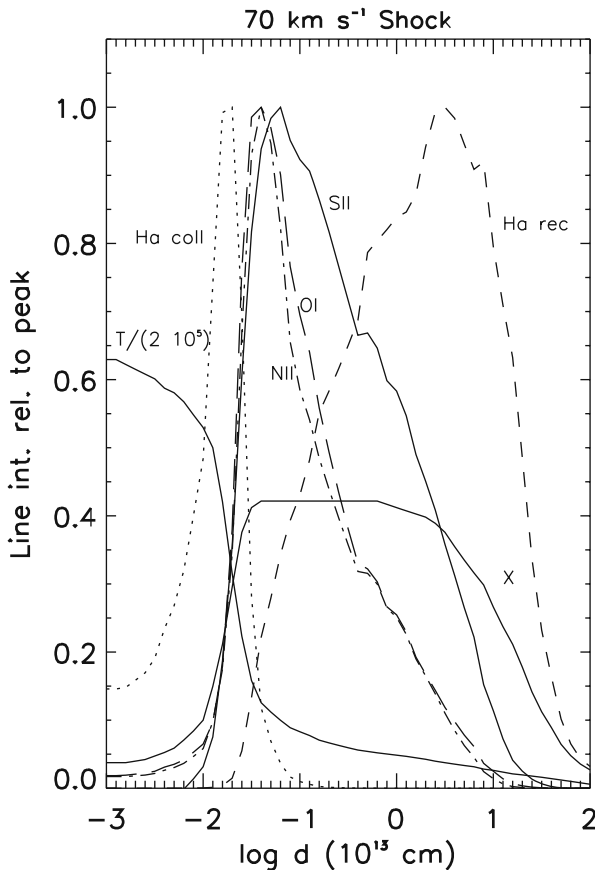


Fig. 10 Relative intensity profiles of several optical and infrared lines as a function of distance from the shock front for a 70 km s^{-1} shock (from [32]). These profiles have been calculated taking the temperature, density and ionization fraction profiles from Fig. 1 of [20]. Figure adapted from [5]

from the shock model of [20]. This figure illustrates that the optical [SII], [NII] and [OI] transitions used in the BE analysis all originate from similar regions and probe intermediate temperatures and ionization fractions, thus biases introduced by spatial gradients are minimized in this technique. The limit of applicability of the BE analysis has been addressed by [28], who shows that this technique reproduces the correct average physical conditions in shocks having velocities $< 50 \text{ km s}^{-1}$ and a pre-shock density $< 10^4 \text{ cm}^{-3}$.

Nisini [32] also shows that IR [Fe II] lines, such as the $1.644 \mu\text{m}$ transition, are emitted from a zone broader than the optical lines, that covers most of the post-shocked cooling region. As a consequence, from a diagnostic based on [Fe II] lines one usually derives electron densities and temperatures higher and lower, respectively, with respect to the values derived from the BE technique [32, 39]. This is particularly true for high-density shocks, where the [SII] density diag-

nostic lines used in the BE technique reach their high-density limit becoming saturated.

The stratification of temperature and densities present in the entire cooling zone can be traced using a larger sample of lines, that map different excitation conditions, as it is shown in Fig. 8.

2.5.3 Uncertainties Introduced by Elemental Abundance

To estimate the ionization fraction, x_e , and the temperature, T_e , with the BE technique one has to assume the relative abundances of S, O, N used in the diagnostic. As a consequence, the derived values may depend on the adopted set of abundances, an issue that has to be investigated in more detail. Originally the diagnostic analysis was done adopting solar abundances from [18]. A large number of HH jets, however, are located, for example, in the Orion star forming cloud. Moreover, there are many lines of evidence that suggest solar system abundances may not even be representative of the local ISM 4.6 Gyr ago, at the time the Sun formed. It follows that solar abundances may not be adequate to study the properties of the jets tout-court. To elucidate the influence of the choice of abundances, several cases have been tested recently [39] by repeating the diagnostics for different sets of abundances adopted by various authors [19, 5, 16, 3] and cross-checking with the expected values of other theoretical line ratios. The results indicate that the differences in the diagnosed values of the physical quantities adopting different abundance sets can amount to 30–40% and that the most reliable values are obtained, not surprisingly, adopting the most recently determined abundance set for the interstellar medium (ISM) in the cloud to which the examined object belongs.

2.5.4 Uncertainties Introduced by Reddening

In general, when working with line ratios one should first deredden the spectra and thus eliminate the effect of wavelength-dependent absorption of dust located along the line of sight to the observed object. The amount of dust in the intervening ISM is, however, generally unknown and may even vary from knot to knot in a single jet. It is thus quite difficult to correct the data for dust extinction in jets, although one expects dust being substantial in a star formation cloud. The lines used in the BE technique are quite close in wavelength. This on one hand makes it impossible to determine the extinction from these lines themselves. On the other hand, one can expect that the proximity in wavelength makes the reddening contribution small on the line ratios. The possible effect on the determination of the physical quantities of not taking reddening into account was estimated quantitatively by [5], who found that, for example, an extinction of $A_V = 3$, quite typical for a T Tauri Star, was producing a variation in the estimated x_e and T_e of at most 10 and 15%, respectively, which is in any case smaller than the error due to the measurement of the fluxes in good signal-to-noise conditions. The reddening uncertainty is of course completely eliminated if one has a mean to determine the extinction, either from the literature or by direct determination using jet emission lines far apart in wavelength. The latter case has been investigated recently with a combined optical-NIR approach (see below).

2.6 Generalizations of the BE Diagnostics

Finally, it is worth mentioning two important generalizations of the BE method that have been developed recently.

First, in two recent papers [32, 39] it is illustrated how the scientific outcome of the BE diagnostic method can be greatly increased if it is combined with intercalibrated NIR spectral analysis. First of all, lines far apart in wavelength can be used to determine the extinction in any spatial resolution element of the spectra, thus correcting it exactly for the dust differential absorption. Then, for example, one can employ ratios between lines from the same species well separated in wavelength to provide more stringent constraints on excitation conditions. This is the case, for example, for the $[\text{Fe II}]\lambda 8617\text{\AA}/1.64\text{ }\mu\text{m}$ and the $[\text{S II}]\lambda 6716,6731\text{\AA}/1.03\text{ }\mu\text{m}$ ratios, which are very sensitive to the gas electron temperature. Also, as mentioned above, such an approach gives the possibility of probing spatially non-resolved different components of the jet cooling layers using ratios of various lines of different critical density. In an optical/NIR global spectrum, NIR lines from H_2 , can be used to compare the properties of the molecular and atomic components of the jet, giving insights on the relative importance of mass flux in the two and on the prevalence of slow C-shocks with respect to J-shocks at the borders of the flow. In addition, a wide wavelength coverage gives one the possibility of observing lines of less abundant species, that, nevertheless, contain important diagnostic information. In particular, lines from many refractory species (Ca, Ni, Cr in addition to Fe and C) can be used to derive the amount of gas depletion in the jet, thus setting constraints to the degree of dust grain destruction by shocks [39].

The second important generalization of the BE technique has appeared in [23], where diagnostics of HST slitless spectra of the HH 30 jet at high angular resolution has been attempted. Here the authors have tested the validity of the basic principles of the method, but also have developed a new diagnostic code that uses all the lines available in the optical range (i.e. O I, O II, N I, N II, S II lines) in a procedure that minimizes the differences between observed and calculated ratios on a large grid of values of the density, temperature and ionization fraction. The net result is to extend the range of validity of the technique to regions characterized by electron densities larger than the $[\text{S II}]\lambda 6731$ critical density and by high temperatures. This, for example, makes it possible to investigate the physics of the dense and excited regions close to the star, thus providing a mean to directly test the validity of the models proposed for the jet acceleration.

3 Deriving Excitation Conditions from Shock Diagnostics

An alternative approach for deriving excitation conditions in jets is the comparison of observed line emission properties (intensities, line ratios) with shock models predictions. It is beyond the scope of this chapter to review the arguments in

favour of shocks as the main line excitation mechanism in jets from young stars. Line emission in jets appears highly structured with small-scale knots separated by a few 100 AU on scales $\leq 10^4$ AU and larger structures (so-called Herbig–Haro objects) separated by a few 0.1 pc on scales up to a few parsecs. Two main origins for the observed structure of jets are considered: instabilities versus internal shocks due to time variable ejection. As we will see below these latter models have proven successful at reproducing the kinematical and excitation properties of knots in young stellar jets. However, it is fair to say that the debate is still going on, especially regarding the small-scale knots. We show below that the comparison of line emission characteristics with predictions from the radiative properties of shock models allows to constrain shock velocities and hydrogen ionization fractions in the post-shock cooling zone. This method therefore provides an alternate way to derive jet densities and mass-loss rates.

3.1 J-Type Shock Radiative Models

We will concentrate here on the analysis of atomic line emission in jets from comparison with the radiative properties of the post-shock cooling regions of J-type (hydrodynamic) shocks. If the shock velocity is low enough (typically less than 30 km s^{-1}) and the transverse magnetic field strength large enough, the discontinuity will be smoothed and a C-type shock may develop. Time-dependent MHD emission, especially H_2 , are detailed in the chapter by David Flower.

Radiative properties of (steady-state) shock models depend on essentially three parameters: the shock velocity, the pre-shock gas density and the pre-shock magnetic field (note that only the component of the magnetic field parallel to the shock surface influences the shock structure). To illustrate the method presented here we will use the 1D radiative J-type shock models developed by [20, 21]. These models are computed for shock velocities V_{shock} ranging from 15 to 100 km s^{-1} , pre-shock atomic hydrogen densities $n_{\text{pre-shock}}$ ranging from 10^2 to 10^6 cm^{-3} and pre-shock magnetic fields $B_0 = 0.1\text{--}3000 \mu\text{G}$, which correspond to the expected conditions in jets on scales of a few 100 AUs. The ionization state of the gas entering the shock affects the amount of Lyman continuum emitted by the shock. Hence a pre-ionization equilibrium is computed through an iterative procedure ensuring that the H pre-ionized fraction agrees with the ionization fraction at the end of the simulation. The flow is allowed to cool for $t = 200$ years behind the shock, which ensures to fully resolve the cooling zone for most of the studied initial conditions (the exception being at the lowest shock velocities and lowest pre-shock densities). Fully 2D bow-shock models have been also developed by different authors (see for, e.g. [19, 43]). These may be more representative of reality, however, a detailed comparison with observations requires a good knowledge of the shock geometry. Planar shock models should apply to any portion of the jet characterized by a single effective shock velocity regardless of the true (certainly complex) geometry of the shock.

3.2 Diagnostics for x_e and V_{shock} from Line Ratios

Hartigan et al. [20] present a method to derive ionization fractions and shock velocities from comparison of observed line ratios involving the strongest optical emission lines with predictions from radiative planar J-type shocks. Figure 11 illustrates the

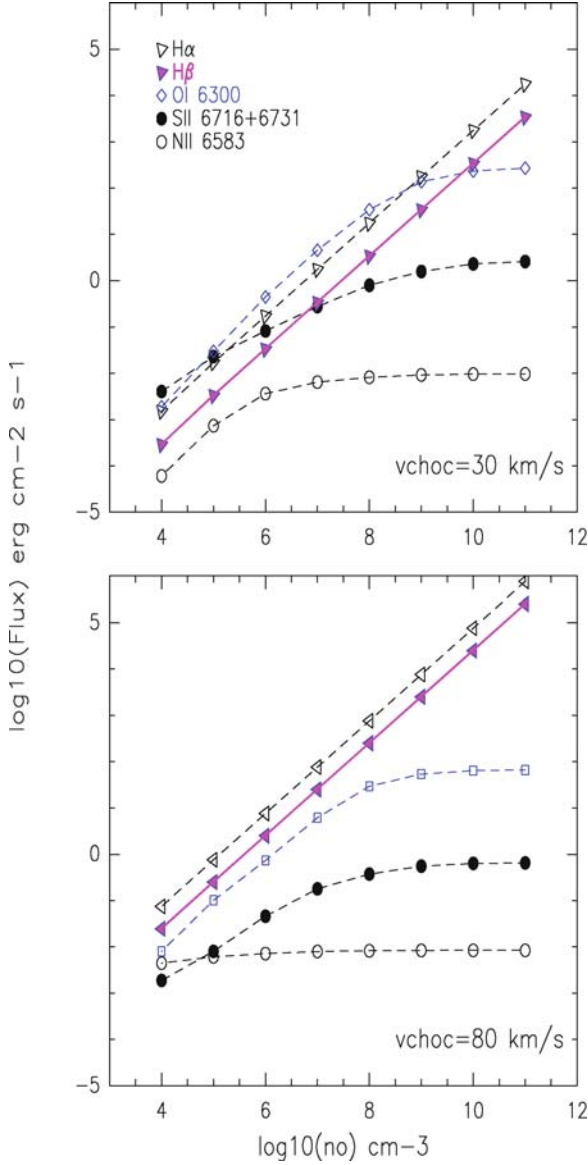


Fig. 11 Line flux radiated per unit surface of the shock front for the most prominent optical emission lines as a function of pre-shock density for two different shock velocities: $V_{\text{shock}} = 30$ and 80 km s^{-1} . Shock models are taken from [20, 21]. Figure adapted from [28]

variation of the most prominent optical lines fluxes as a function of pre-shock density and shock velocity. Under the assumption of optically thin emission, line fluxes produced in the cooling region of a shock front can be approximated as

$$F_{21} \propto \epsilon_{21} d_{\text{cool}} = \epsilon_{21} V_{\text{shock}} t_{\text{cool}}, \quad (10)$$

where ϵ_{21} is the line emissivity corresponding to the transition from level 2 to level 1 (cf §2.1) and d_{cool} the longitudinal size of the post-shock cooling region. The cooling time behind the shock front is roughly given by the total thermal energy divided by the cooling rate, i.e. $t_{\text{cool}} \simeq (3/2 n_{\text{H}} kT) / (n_{\text{post-shock}}^2 \Lambda(T))$ where $\Lambda(T)$ is the cooling function. $n_{\text{post-shock}}$, the H density in the post-shock cooling gaz, is given by $n_{\text{post-shock}} = C n_{\text{pre-shock}}$, C being the shock compression factor. Hence,

$$\begin{aligned} F_{21} &\propto \epsilon_{21} V_{\text{shock}} (3/2 n_{\text{post-shock}} kT) / (n_{\text{post-shock}}^2 \Lambda(T)) \\ F_{21} &\propto \frac{\epsilon_{21}}{C n_{\text{pre-shock}}} V_{\text{shock}} f(T). \end{aligned}$$

Line fluxes from the shock increase linearly with the pre-shock density in the low-density regime (when $n_e < n_{\text{cr}}$) and become almost insensitive to $n_{\text{pre-shock}}$ at high electronic densities (corresponding to large $n_{\text{pre-shock}}$ or large V_{shock}). The combination of lines with different critical densities will be strongly dependent on $n_{\text{pre-shock}}$, while the combination of lines with similar critical densities will be more sensitive to V_{shock} .

In particular the ratio $[\text{N II}]\lambda 6583\text{\AA} / [\text{O I}]\lambda 6300\text{\AA}$ appears as a good diagnostic of the intensity weighted ionization fraction in the post-shock cooling zone. This ratio is insensitive to pre-shock densities (the two lines having similar critical densities) and pre-shock magnetic field. The post-shock ionization fraction increases with shock velocity and decreases with increasing pre-shock density.

From comparison of five independent optical line ratios with predictions from their planar J-type shock models, Hartigan et al. [20] derive accurate estimates of post-shock ionization fractions and shock velocities in three HH jets (HH 34, HH 47 and HH 111). From knowledge of the post-shock electronic densities (derived from the $[\text{S II}]\lambda 6716 / \lambda 6731\text{\AA}$ ratio, for example) and using the compression factor given by the required shock model, one can further derive both post and pre-shock total gaz densities. Hartigan et al. [20] argue that the best estimate of the average jet total H density in presence of ejection variability is given by $n_{\text{H}} = \sqrt{n_{\text{pre-shock}} n_{\text{post-shock}}} = C^{-1/2} n_{\text{post-shock}}$. The low derived ionization fractions (in the range 1–5%) lead the authors to significantly increase previous estimates of the jet average density hence the mass-loss rates in these flows, bringing them in better agreement with mass-loss rates derived from line luminosities and suggesting that the atomic component of these jets is powerful enough to drive associated molecular outflows.

3.3 Line Diagnostic Diagrams

Another interesting tool proposed by [29] relies on line ratio diagrams involving the strong optical forbidden emission lines of [O I] λ 6300, [N II] λ 6583, [S II] λ 6731 and [S II] λ 6716. We plot in Fig. 12 the locus occupied by J-type planar shock

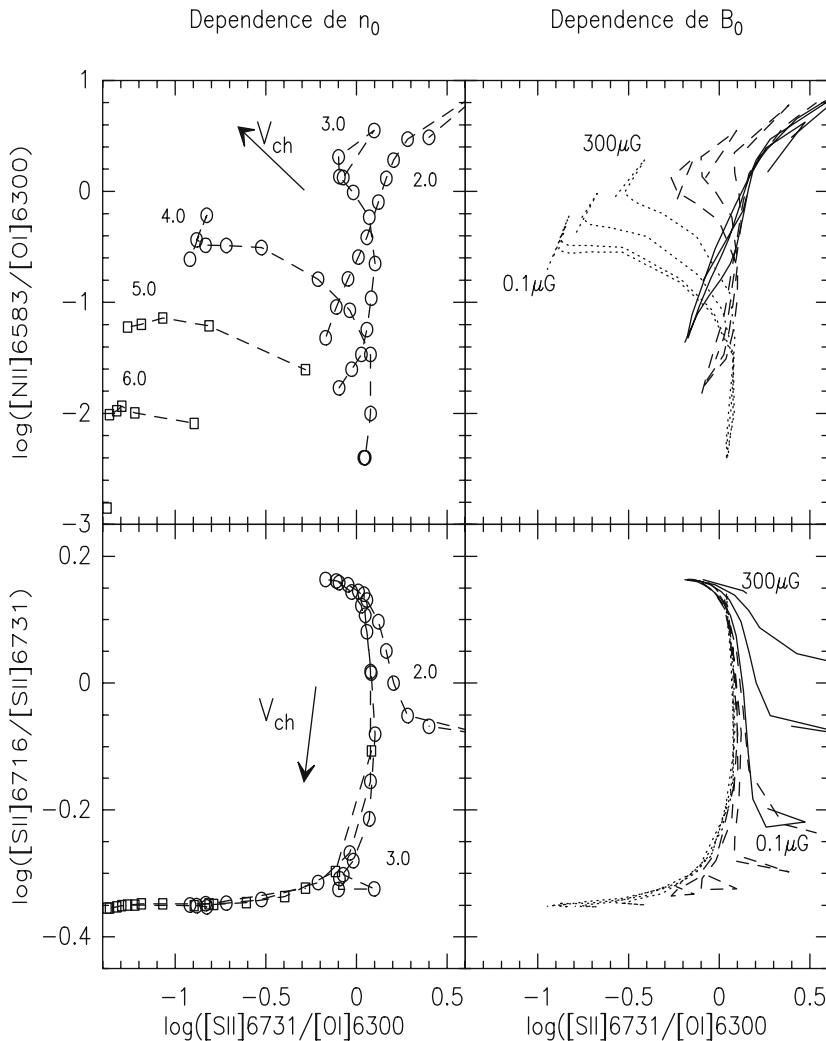


Fig. 12 Optical line ratio diagrams predicted by the planar J-type shock models of [20, 21]. Left panels: $B_0 = 100 \mu\text{G}$ and $n_{\text{pre-shock}}$ from 10^2 to 10^6 cm^{-3} . The curves connect models with the same pre-shock density. Shock speeds increase with [N II] λ 6583/[O I] λ 6300 along each curve from 15 to 100 km s^{-1} for $n_{\text{pre-shock}} \leq 10^4 \text{ cm}^{-3}$, from 30 to 80 km s^{-1} for $n_{\text{pre-shock}} = 10^5, 10^6 \text{ cm}^{-3}$. Right panels: The variation of the curves with a pre-shock magnetic field from 0.1 to $300 \mu\text{G}$ is shown for pre-shock densities $n_{\text{pre-shock}} = 10^3, 10^4 \text{ cm}^{-3}$. Figure taken from [28]

models in the [N II] λ 6583/[O I] λ 6300 versus [S II] λ 6731/[O I] λ 6300 and in the [S II] λ 6716/ λ 6731 versus [S II] λ 6731/[O I] λ 6300 diagrams.

All shock models are found to follow a well-defined sequence in the [S II] λ 6716/ λ 6731 versus [S II] λ 6731/[O I] λ 6300 diagram, not affected by the magnetic field. The post-shock forbidden line emitting regions have all a similar excitation temperature and hence a similar [S II] λ 6731/[O I] λ 6300 ratio. The [S II] λ 6716/ λ 6731 decreases as the post-shock electronic densities increase, i.e. as the shock velocity and/or pre-shock densities increase. When post-shock electronic densities become larger than the critical density of [S II] λ 6731, the [S II] λ 6731/[O I] λ 6300 ratio decreases and the [S II] λ 6716/ λ 6731 reaches its high-density limit. On the other hand, in the [N II] λ 6583/[O I] λ 6300 versus [S II] λ 6731/[O I] λ 6300 curves for differing pre-shock densities $n_{\text{pre-shock}}$ are clearly separated for $n_{\text{pre-shock}} \geq 10^3 \text{ cm}^{-3}$. This again results from the fact that post-shock electronic densities reach the critical densities of the [S II] lines. The [N II] λ 6583/[O I] λ 6300 ratio, sensitive to post-shock ionization fraction, combined with the [S II] λ 6731/[O I] λ 6300 ratio, sensitive to post-shock electronic densities, provides a tool to constrain the shock parameters.

These two diagrams therefore allow (1) to clearly identify shocks as the excitation mechanism if line ratios follow the *shock sequence* in the [S II] λ 6716/ λ 6731 versus [S II] λ 6731/[O I] λ 6300 diagram and (2) to derive estimates of $n_{\text{post-shock}}$ and V_{shock} in the inner regions of T Tauri microjets from the location of their line ratios in the [N II] λ 6583/[O I] λ 6300 versus [S II] λ 6731/[O I] λ 6300 diagram, provided an assumption is made on the strength of the magnetic field.

This tool has been applied to the DG Tau and RW Aur microjets ([29, 15]). Optical forbidden line ratios in these two jets closely follow the shock sequence at all distances along the jet and at all flow velocities except the lowest ones in DG Tau. All other investigated heating mechanisms (turbulent mixing layers, ambipolar diffusion) fail to reproduce the observed optical line ratios in these jets. In DG Tau, inferred shock speeds increase with flow velocity, ranging from 50 to 100 km s $^{-1}$ and pre-shock densities decrease along the jet from 10^5 cm^{-3} close to the star to 10^3 cm^{-3} at 1.5".

3.4 Constraining Pre-shock Magnetic Fields

Even a moderate component of the magnetic field parallel to the shock front in the pre-shock gas lowers the compression of post-shock material and changes the shock excitation conditions and radiative properties. In principle estimates of the magnetic field parallel to the shock front can be derived from the comparison of observed line emission with predictions of magnetized shock models. However, Hartigan et al. [20] demonstrate that the optical line ratios of shocks with magnetic field closely resemble the ones of shocks with lower shock velocities and no B (see Fig. 12). The H β line flux could help discriminate between a magnetized and non-magnetized case but this requires knowledge of the total area of the shocks contributing to the observed line emission.

The presence of magnetic fields in jets can be indirectly inferred from their observed excitation properties assuming this is due to shocks. From a detailed HST/STIS study of the HH30 jet, Hartigan et al. [23] find high-ionization knots forming at around 100 AU in the flow and propagating outwards. These knots are not accompanied by corresponding increase in the density, suggesting that the compression is lowered by strong internal magnetic pressure which also lengthens recombination times.

4 Derivation of Mass-Loss Rates

Mass-loss rates, \dot{M}_{jet} , are critical jet parameters to derive. First, the mass ejection to accretion rates ratio is a constraining parameter for MHD launching models. The flux of angular momentum ($\dot{M}_{\text{jet}} \times r v_{\phi}$) carried by the jet may play an important role in the regulation of stellar rotation. Last, the jet thrust or flux of linear momentum ($\dot{M}_{\text{jet}} V_{\text{jet}}$) is important to establish whether atomic jets are powerful enough to drive molecular outflows and to evaluate the amount of turbulence injected in the interstellar medium. We concentrate here on methods deriving \dot{M}_{jet} from direct atomic jet observations. The assumptions and limitations of each method are presented in detail. This section is adapted from [11].

A second approach for estimating wind dynamics uses its interaction with ambient material, manifested in the form of slow molecular outflows and shock-excited emission lines from the wind/cloud interface. Unlike direct jet observations, these estimates are integrated over all solid angles and independent of the wind excitation conditions or chemical composition. This second class of methods will not be discussed here and is described in detail in [11].

4.1 Methods Based on the Jet Density and Cross Section

These methods, which are free of extinction corrections, determine the mass-loss rate \dot{M}_{jet} from estimates of the mean jet density n_{jet} , velocity V_{jet} and radius r_{jet} (when spatially resolved):

$$\dot{M}_{\text{jet}} = 10^{-8} \left(\frac{n_{\text{jet}}}{10^4 \text{ cm}^{-3}} \right) \left(\frac{r_{\text{jet}}}{50 \text{ AU}} \right)^2 \left(\frac{V_{\text{jet}}}{200 \text{ km s}^{-1}} \right) M_{\odot} \text{ year}^{-1}. \quad (11)$$

Two different assumptions can be used to derive the mean jet density n_{jet} from the measured density $n_{\text{H}} = n_{\text{e}}/x_{\text{e}}$, estimated from emission lines as discussed in the previous sections.

(A.1) *Physical conditions are uniform within the observing pixel:* Then, no compression correction is necessary and $n_{\text{jet}} = n_{\text{H}}$ ([5]). This method is probably an *upper limit* to the actual mass-loss rate: If the beam contains a mixture of high

and low density regions, n_{H} will probably be dominated by the high density, high emissivity ones.

(A.2) *Emission comes from a shock wave:* Then, n_{jet} can be taken as the pre-shock density, i.e. $n_{\text{jet}} = n_{\text{H}} \times C^{-1}$ where C is the compression factor in the shocked layer where the lines are emitted, estimated from comparison of observed line ratios with shock models [20]. Actually, Hartigan et al. [20] recommend to use $(n_{\text{post-shock}} n_{\text{pre-shock}})^{1/2}$ as a better estimate of n_{jet} and thus apply $\sqrt{C^{-1}}$ rather than C^{-1} as compression correction: 1D hydrodynamical simulations suggest that it gives a better estimate of the average jet mass flux in the presence of velocity variability; ejected material tends to accumulate in the shocks, diminishing the mass-flux in the freely flowing segments between shocks. For shock speeds of 30 km s^{-1} , compression factors are on the order of 15–20 and correction factors of $\simeq 4$ should be applied to the measured jet density.

4.2 Methods Based on the Jet Luminosity

Luminosity-based methods have the advantage of being independent of the jet diameter (provided that it is smaller than the beam of observations), but at the cost of an uncertainty in the extinction correction. The two methods are presented in detail in Appendix A of [21].

(B.1) *Physical conditions are uniform within the observing pixel:* The mass-loss rate is simply given by

$$\dot{M}_{\text{jet}} = \mu' 2m_{\text{H}} N_{\text{H}} V_{\perp} / l_{\perp}, \quad (12)$$

where μ' is the mean particle weight per H nucleus and is related to μ (the mean atomic weight) by $\mu' = \mu N_{\text{tot}} / N_{\text{H}}$ ($\mu' = 1.35$ for atomic gas of solar abundance), N_{H} is the total number of hydrogen atoms within the aperture, derived from the (optically thin) forbidden line luminosity ($N_{\text{H}} = n_{\text{H}} \times V$), and V_{\perp} and l_{\perp} are the projected jet velocity and length in the plane of the sky. For the [S II] $\lambda 6731 \text{ \AA}$ line one finds, assuming all sulfur is singly ionized (Eq. A10 of [21]):

$$\begin{aligned} \dot{M}_{\text{jet}} = 3 \times 10^{-8} \left(1 + \frac{n_{\text{c}}(6731)}{n_{\text{e}}} \right) & \left(\frac{L(6731)}{10^{-4} L_{\odot}} \right) \\ & \left(\frac{V_{\perp}}{150 \text{ km s}^{-1}} \right) \left(\frac{l_{\perp}}{2 \times 10^{15} \text{ cm}} \right)^{-1} M_{\odot} \text{ year}^{-1}, \end{aligned} \quad (13)$$

where $n_{\text{c}}(6731) = 1.3 \times 10^4 \text{ cm}^{-3}$ is the critical density for collisional de-excitation. Similarly, for the [O I] $\lambda 6300 \text{ \AA}$ line, $n_{\text{c}}(6300) = 2 \times 10^6 \text{ cm}^{-3}$ at a representative temperature of 8200K, and

$$\dot{M}_{\text{jet}} = 2 \times 10^{-10} \left(1 + \frac{n_c(6300)}{n_e} \right) \left(\frac{L(6300)}{10^{-4} L_{\odot}} \right) \left(\frac{1}{1 - x_c} \right) \left(\frac{V_{\perp}}{150 \text{ km s}^{-1}} \right) \left(\frac{l_{\perp}}{2 \times 10^{15} \text{ cm}} \right)^{-1} M_{\odot} \text{ year}^{-1}. \quad (14)$$

The above expression is derived from Eq. A8 of [21], with an additional correction factor for oxygen ionization (assuming that, through charge exchange, $\text{O}^+/\text{O} = \text{H}^+/\text{H}$). Similar expressions can be derived for mass-loss rates estimated from the fluxes of the near-infrared [Fe II] emission lines ([32]). In that case, however, one must take into account the possible strong Fe depletion in the gas phase due to the presence of dust in the flow (see the chapter by Brunella Nisini in Elba school proceedings).

(B.2) *Emission traces cooling in a shock wave:* Hollenbach et al. [25] find a quasi-proportionality between the [O I] $63 \mu\text{m}$ line luminosity and the mass-flux through a shock, \dot{M}_{shock} . The same roughly applies to the [O I] $\lambda 6300$ line ([21]). This behaviour is understood with the following simple argument: Let us assume that the [O I] $\lambda 6300$ line represents a fixed fraction $1/f$ of the shock cooling in the temperature range $T_2 - T_1$. Then, if $N(T_2, T_1)$ denotes the number of atoms of post-shock gas in this same range,

$$f L(\text{line}) = L_{\text{cool}} \simeq \frac{3}{2} k(T_2 - T_1) \times \frac{N(T_2, T_1)}{t_{\text{cool}}(T_2, T_1)} \simeq \frac{3}{2} k(T_2 - T_1) \frac{\dot{M}_{\text{shock}}}{\mu m_H}. \quad (15)$$

A proportionality between $L(\text{line})$ and \dot{M}_{shock} is thus predicted as long as f does not vary greatly with shock conditions (which is justified at pre-shock densities $\geq 10^5 \text{ cm}^{-3}$). To derive \dot{M}_{jet} from \dot{M}_{shock} , an assumption must be made about the shock geometry. For a moving shock intercepting the whole jet beam, and perpendicular to the jet flow, $\dot{M}_{\text{jet}} = (V_{\text{jet}}/V_{\text{shock}}) \dot{M}_{\text{shock}}$. Quantitatively, Hartigan et al. [21] find

$$\dot{M}_{\text{jet}} = 5 \times 10^{-7} \left(\frac{L(6300)}{10^{-4} L_{\odot}} \right) \left(\frac{\cos \theta}{N_{\text{shock}}} \right) \left(\frac{V_{\text{jet}}/300 \text{ km s}^{-1}}{V_{\text{shock}}/40 \text{ km s}^{-1}} \right) M_{\odot} \text{ year}^{-1} \quad (16)$$

The above expression is derived from Eq. (A17) of [21], with an additional correction factor $\cos \theta / N_{\text{shock}}$, where N_{shock} is the number of shocks within the beam, and μ is the angle of the shock from the perpendicular to the jet flow. In the particular case of a stationary oblique shock, $V_{\text{shock}} = \cos \theta V_{\text{jet}}$ and $\dot{M}_{\text{jet}} = \dot{M}_{\text{shock}}$.

4.3 Comparison of the four Methods: Class I Jets and the DG Tau Microjet

Since the above four methods all use the same tracer: forbidden lines, it is instructive to compare the resulting \dot{M}_{jet} , as an indicator of uncertainties. Table 2 presents such

Table 2 Comparison of four mass-loss rate derivations from optical lines in Class I jets

Jet	n_e (cm^{-3})	$x_e(BE)$	$\langle x_e \rangle$	\sqrt{C}	$V_{\text{jet}}/V_{\text{shock}}$ (km s^{-1})	Cross section		[O I] luminosity	
						Uniform (A.1)	Shock (A.2)	Uniform (B.1)	Shock (B.2)
HH 34	650	0.027	0.016	4.0	220/28	$4.0 \cdot 10^{-7}$	$1.7 \cdot 10^{-7}$	$1.5 \cdot 10^{-7}$	$4.5 \cdot 10^{-6}$
HH 47	250	0.070	0.036	4.9	350/34	$9.8 \cdot 10^{-7}$	$3.9 \cdot 10^{-7}$	$4.2 \cdot 10^{-7}$	$4.6 \cdot 10^{-5}$
HH 111	900	0.064	0.027	4.7	320/32	$3.6 \cdot 10^{-7}$	$1.8 \cdot 10^{-7}$	$5.8 \cdot 10^{-7}$	$2.9 \cdot 10^{-5}$

a comparison for the three bright Class I jets: HH 34, HH 47 and HH 111. The table lists planar shock model parameters $\langle x_e \rangle$, \sqrt{C} and V_{shock} determined from five different line ratios by Hartigan et al. [20] and their published values of \dot{M}_{jet} using methods (A.2) and (B.1). In addition, the table also lists \dot{M}_{jet} for method (A.1), using the values of $x_e(BE)$ derived by [5] with their procedure, and \dot{M}_{jet} for method (B.2), assuming $\cos \mu / N_{\text{shock}} = 1$.

As x_e and n_e actually vary along the jet (e.g. [5]) and with radial velocity inside the line profile (e.g. [29]) it is also instructive to perform a comparison as a function of distance and flow speed. This is done in Fig. 13 for the microjet from the Class II source DG Tau, using 80 AU pixels ([29]).

Among the two methods assuming uniform emission, the cross section-based estimate (A.1) is always higher than that expected from the observed [O I] λ 6300 luminosity (B.1), for both Class I jets and the DG Tau microjet. It suggests that the local n_H is higher than the average density in the jet volume. A filling factor <1 is similarly derived by [5] for the HH 30 jet. Hence (A.1) tends to give an upper limit to the mass-loss rate.

Among the two methods assuming shocks, (B.2) derived from the total luminosity is generally higher than (A.2). The difference amounts to 10–100 in Class I jets, where line luminosity was integrated over a section of 10^3 – 10^4 AU along the jet ($2''$ – $20''$ at 500 pc; [20]). It would suggest that $N_{\text{shock}} / \cos \mu \simeq 1$ per 100 AU of jet length. The difference is less severe in DG Tau, a factor of 1–4 typically; the smaller pixel of 80 AU includes less shocks. Overall, (B.2) does not appear reliable unless small pixels ≤ 100 AU are used.

The most reliable methods thus appear to be (A.2) and (B.1). Interestingly, they give very similar results within a factor 3 in Class I jets, as well as in the DG Tau microjet beyond 100 AU of the source. Closer to DG Tau, however, the two methods become discrepant, with $\dot{M}_{\text{jet}}(A.2)$ increasing by almost a factor 10 while $\dot{M}_{\text{jet}}(B.1)$ decreases. The latter could be due to extinction, and the former to an overestimated jet radius (close to the resolution limit). Extinction correction of (B.1) using the stellar A_V (3.5 mag towards DG Tau, [21]) will give the correct value if the increase in $\dot{M}_{\text{jet}}(A.2)$ is real, and an upper limit otherwise, if the larger scale value is more representative.

In addition, it should be kept in mind that, whatever the method, the *observable* mass-loss rate could be either an upper limit (if part of the ‘jet’ is tracing entrained ambient material) or a lower limit (if some ejected material is too cool and/or too diffuse to emit).

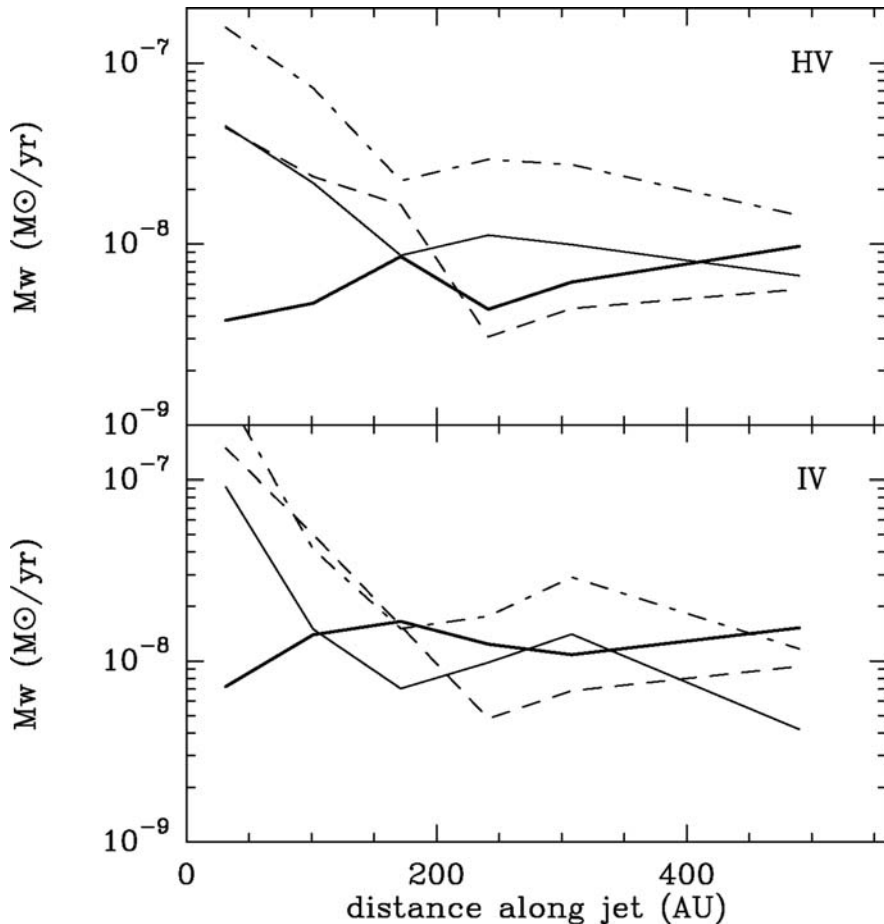


Fig. 13 Mass-loss rate in the DG Tau microjet as a function of distance for two different velocity ranges: *Top panel*: high-velocity $[-400, -250] \text{ km s}^{-1}$, *bottom panel*: intermediate velocity $[-250, -100] \text{ km s}^{-1}$. Curve styles refer to the four different methods: Using the jet cross section, and assuming uniform emission (A.1, *dash-dot*) or a shock (A.2, *solid line*); $[\text{O I}]\lambda 6300$ luminosity, assuming uniform emission (B.1, *grey line*) or a single shock (B.2, *dashed*, with $\mu/N_{\text{shock}} = 1$). Adapted from [28].

5 Projection and Convolution Effects

We discuss in this section some of the prominent observational biases introduced by line of sight projection and instrumental convolution effects. Indeed, observables correspond at best to a 2D spatial projection onto the plane of the sky of an underlying 3D structure. In addition, the instrumental configuration introduces a transfer function, which affects both the spatial and spectral domains and further modifies the brightness distribution. Taking into account these two effects can be critical to derive physically meaningful parameters from observations.

5.1 Deriving Inclination Angles

We will first start with a cautionary remark about deriving jet inclination angles. This parameter is critical to take into account projection effects but notoriously difficult to constrain. The first possibility is a priori knowledge of the inclination angle of the rotation axis of the central star-disc system. For all sources with known close to edge-on geometry, jets are observed to emerge perpendicular to the plane of the circumstellar disc. On larger scales, however, changes of jet directions may be observed. For stars where rotation periods are determined, through periodic photometric variations, rotation axis inclination angles can be derived from the broadening of photospheric lines which is proportional to $V_{\text{eq}} \sin(i)$, where V_{eq} is the equatorial stellar rotation velocity. Unfortunately, most active T Tauri stars driving jets show irregular photometric variations, preventing accurate determination of rotation periods. The circumstellar disc rotation axis can only be accurately determined from fitting simultaneously fully resolved images and spectral energy distributions.

An alternative way is to derive jet inclination angles from proper motion measurements. In principle, $\tan(i) = V_{\perp} / V_{\parallel}$ where i = inclination angle of the jet axis to the line of sight, V_{\parallel} and V_{\perp} , respectively, the line of sight and tangential velocities measured at the same position in the jet. However, the brightness distribution in jets is dominated by knots of emission, most likely due to shocks arising from variable ejection at the source. Eisloffel and Mundt [17] have shown that internal knots travel at a fraction of the space velocity of jet material (between 40 and 100 %). Thus jet inclination angle can be safely derived with this method only for the knot located at the apex of the terminal shock where jet material meets interstellar medium at rest. Unfortunately, terminal shocks are not always clearly identified ...

5.2 Flow Line of Sight Velocities

Jet flow velocities are usually derived from peak or centroid line of sight velocities of emission line profiles. Knowing a priori either the jet inclination angle or associated proper motions, true space velocity vectors can be reconstructed. This method will give an accurate description of the poloidal velocity field provided that no strong velocity gradients occur within one spatial or spectral resolution element. This is generally the case for studies of jets on large distances from their driving source ($> a \text{ few } 1000 \text{ AU}$). For studies of jets in their collimation and acceleration regions, however, this approximation may fail. Indeed, close to the source the flow is strongly stratified, both in velocities and in excitation conditions ([7]). The line of sight velocities where the maximum of line emission occurs will be a complex convolution of the true underlying velocity field with the excitation conditions of the flow surfaces controlling the line emissivity (T_e, n_e, x_e), with projection effects onto the plane of the sky and with observational parameters (spectral and spatial resolutions). An accurate comparison with model predictions requires to take into account all of these effects. To illustrate this effect we show in Fig. 14 the predicted position–velocity diagram obtained for a self-similar disc wind solution with mag-

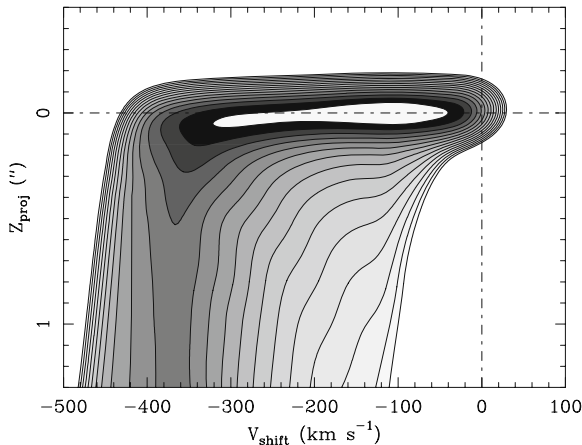


Fig. 14 Synthetic position–velocity diagram along the jet computed for a self-similar disc wind solution with $\lambda = 50$, inner and outer launching radii $r_{\text{in}} = 0.1$ AU and $r_{\text{out}} = 1$ AU, and $\dot{M}_{\text{acc}} = 10^{-6} M_{\odot} \text{ year}^{-1}$. The observational parameters are: $i = 45^\circ$, spatial resolution 22 AU and spectral resolution 30 km s^{-1} . Adapted from [37]

netic lever arm $\lambda = 50$, inner and outer launching radii 0.1 and 1 AU, respectively. True terminal poloidal velocities range from 680 km s^{-1} for the inner streamline to 170 km s^{-1} for the outer one. The predicted terminal peak line of sight velocity is $\simeq 520 \text{ km s}^{-1}$, which corresponds to the true flow velocity of an intermediate streamline and depends on both the excitation conditions and the resolution parameters.

5.3 Rotation Velocities

One of the most exciting findings in recent years has been the detection of an asymmetry in the radial velocity across the jet section, that might be interpreted as rotation of YSO jets around their symmetry axes. Indications for rotation have been obtained in the first 100–200 AU of the jet channel through high angular resolution observations, both from space and the ground at wavelengths ranging from the near-infrared to the near-UV [9, 13, 14, 45, 12].

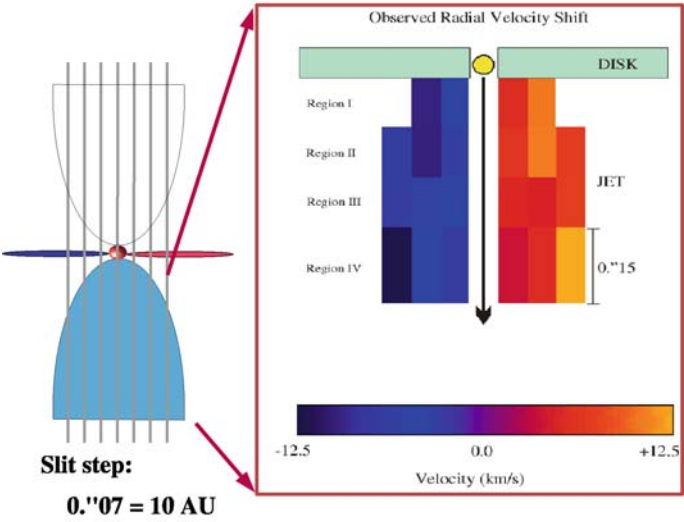
The principle of measurement of jet rotation is to search for differences in the Doppler shift of the lines emitted by the flow at equal distances from the symmetry axis, and in particular at the opposite borders of the jet. The rotation of the jet, in fact, will cause the two borders of the flow to move towards the observer at slightly different velocities with respect to the material seen towards the symmetry axis. To detect rotation one has to use instruments working at subarcsecond resolution, to separate spatially the information coming from the two sides of the jets, and at moderate/high spectral resolution, to identify expected velocity differences of the order of $5\text{--}15 \text{ km s}^{-1}$. The radial velocity (i.e. towards the observer) of a given line can be measured by fitting a Gaussian to its profile and subtracting with respect to the system velocity of the star. Rotation will then be searched by comparing the

velocities of two measured profiles. In the case of rotation studies, however, one can directly measure the *difference* of the radial velocity of the line profiles of interest by applying cross-correlation routines that measure the displacement of two line profiles independently of their particular shape. Note that using the above techniques of data analysis one can improve upon the nominal spectral resolution by more than ten times. For example, the velocity resolution of HST/STIS in the optical is 55 km s^{-1} , but applying gaussian fitting and cross-correlation routines to the line profiles it is possible to detect shifts in spatially resolved locations down to about 5 km s^{-1} at high signal over noise. Indeed, the standard deviation of the centroid estimate for a gaussian distribution of rms standard deviation σ_{line} and peak signal-to-noise SNR is given by $\sigma_{\text{cent}} = \frac{\sigma_{\text{line}}}{\text{SNR}}$ [40]. Uncertainties in velocity shifts derived from gaussian fitting procedure are thus typically: $\sigma_V = \frac{\sqrt{2}\text{FWHM}/2.35}{\text{SNR}}$ where FWHM is the spectral FWHM of the profile and SNR the signal over noise ratio at the line peak.

As illustrated in Fig. 15, one can use both a ‘parallel’ and a ‘perpendicular’ configuration of the slit(s) with respect to the flow axis. The ‘parallel’ configuration allows one to directly measure the velocity of the line peak in different spectra and thus map the differential velocity in a larger region of the flow. This method, however, has the disadvantage of requiring a careful subtraction of spurious effects caused by uneven slit illumination that could mimic rotation signatures [9, 45]. The ‘perpendicular’ configuration allows one to get the spectrum of the entire jet section in one go, and spurious illumination effects are minimized. The analysis of the line profile, however, is more difficult, as one needs to search for a ‘tilt’ in the brightness distribution with respect to space and wavelength directions, caused by the fact that the borders of the flow move towards the observer at slightly different velocities. Such a tilt, represented by the oblique line in the bottom panel of Fig. 15, is not always easily detectable, and its measurement requires velocity analysis of each pixel row in the spectrum [13]. In both cases, to connect the observed velocity shift to actual measurements of flow rotation, one has to consider the (sometimes unknown) inclination of the flow with respect to the plane of the sky and the influence of projection effects through the flow itself, as it will be discussed below.

Rotation signatures have been seen in all the T Tauri jets observed with HST/STIS (DG Tau, RW Aur, CW Tau, Th 28 and HH 30) in different emission lines and using slit orientations both parallel and perpendicular to the outflow axis. At optical wavelengths, systematic shifts in radial velocity, typically $5\text{--}25 \pm 5 \text{ km s}^{-1}$, were found at jet positions displaced symmetrically with respect to the outflow axis, at 50–60 AU from the source and 20–30 AU from the axis (see Fig. 15). Very recently, velocity asymmetries compatible with jet rotation have also been detected in the near ultra-violet (NUV) lines of $\text{Mg}^+ \lambda\lambda 2796, 2803$ in the jets from Th 28 and DG Tau [14] and in the H_2 $2.12 \mu\text{m}$ lines using ISAAC on the VLT in two small-scale jets, HH 26 and HH 72, driven by Class I sources [12]. The detection of rotation is interesting per se, as it supports the idea that jets are centrifugally launched presumably through the action of a magnetic ‘lever-arm’. An important quantity derived from the observed rotation signatures, assuming steady magneto-hydrodynamic ejection, is

Jet rotation with 'parallel' slits : DG Tau jet



Jet rotation with 'perpendicular' slits: TH 28 jet

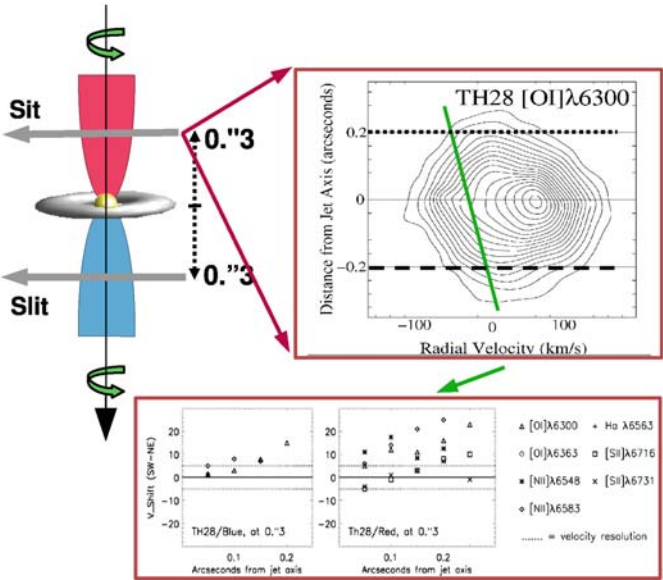


Fig. 15 (continued)

the ‘footpoint radius’, i.e. the location in the accretion disc from where the observed portion of the jet is launched [1]. The observations would be consistent with footpoint radii between 0.5 and 5 AU from the star, i.e. originating from the inner disc regions. Another interesting quantity derived from the observed rotation signatures is the amount of angular momentum carried by the jet. In the two systems for which sufficient information is available (namely, DG Tau and RW Aur), this amounts to between 60 and 100% of the angular momentum that the inner disc has to lose to accrete at the observed rate. Thus, the fundamental physical implication of the observed rotation is that jets are likely to be the major agent for extracting excess angular momentum from the inner disc and that they do so throughout their lives.

However, care must be taken in translating the observed velocity shifts into jet azimuthal velocities. For an annulus of radius r rotating at V_ϕ the measured velocity shift is related to the azimuthal velocity by $V_{\text{shift}}(r) = 2V_\phi(r) \sin(i)$ with r cylindrical radius in the jet and i inclination of the jet axis to the line of sight. A true jet, observed at a given spectral and spatial resolution, will, however, suffer from projection and convolution effects such as the simple formula above can no longer be used. These effects are particularly critical in the measurements of jet azimuthal velocities close to their driving source since both the currently best achievable spatial ($\simeq 0.1''$) and spectral resolutions ($\simeq 30 \text{ km s}^{-1}$) are comparable to jet widths and expected transverse velocity gradients, respectively. These effects have been studied in detail by Pesenti et al. [38] and applied to analysis of the transverse velocity shifts measurements in the DG Tau microjet. The authors have shown that the velocity shifts measured at a given transverse distance, d_\perp , in the jet always underestimate the true rotation velocities at d_\perp and that the discrepancy strongly increases towards the jet axis. The observed velocity shifts will only be a reliable measure of V_ϕ on transverse distances greater than $\simeq 2$ spatial beams.

This result arises from the combination of two effects. The first one, illustrated in Fig. 16, comes from projection along the line of sight. At a given projected transverse distance d_\perp from the jet axis, the emission integrated along the line of sight will arise from concentric rings of true cylindrical radius $R \geq d_\perp$. The contribution to the measured velocity shift at $d_\perp = r$ from the ring of true radius R will be: $V_{\text{shift}}(d_\perp = r) = 2 \times V_\phi(R) \times (r/R) \times \sin(i) < 2 \times V_\phi(r) \times \sin(i)$ if as expected from ejection models the rotation profile decreases with cylindrical radius r . In addition, convolution by a spatial beam will tend to further decrease the measured V_{shift} , since the contribution from rings located at true cylindrical radii $> d_\perp$ will be increased.

Full analysis of biases introduced by observational parameters has been conducted in [38] and is illustrated in Fig. 17. Synthetic predictions of measured trans-

Fig. 15 Transverse velocity shifts in optical emission lines detected with HST/STIS across the jets from DG Tau (*top*) and Th 28 (*bottom*), at about 50–60 AU from the source and 20–30 AU from the outflow axis, in ‘parallel’ and ‘perpendicular’ slit modes, respectively. Applying gaussian fitting and cross-correlation routines to the line profiles in locations opposed with respects to the jet axis velocity shifts of 5–25 km s^{−1} could be detected. The values for the shifts obtained in this way suggest toroidal speeds of 10–20 km/s at the jet boundaries

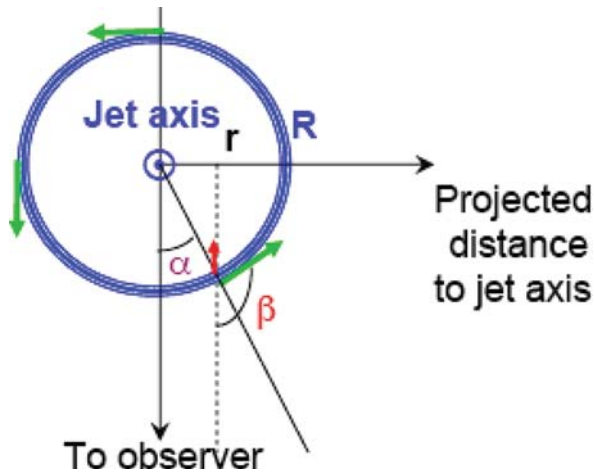


Fig. 16 Illustration of projection effects on measured velocity shifts. The jet is assumed cylindrical with its axis perpendicular to the line of sight. The contribution from annulus of true cylindrical radius R to the measured velocity shift at a projected transverse distance $d_{\perp} = r$ from the axis is $V_{\text{shift}}(r) = 2 \times V_{\phi}(R) \times \sin(\alpha + \pi/2) = 2 \times V_{\phi}(R) \times (r/R)$

verse velocity shifts have been computed for a warm self-similar disc wind solution, varying the spatial resolution, the range of velocities examined and the excitation conditions in the jet (the spectral resolution does not play a major role, provided it is ≥ 3000). Clearly, all these parameters have a strong influence on the detected velocity shifts. In particular, we note that with parameters comparable to the HST/STIS observations of the DG Tau microjet by Bacciotti et al. [9] (spatial resolution 14 AU, velocity shifts computed for intermediate velocity component $|V| \leq 100 \text{ km s}^{-1}$), the velocity shifts are predicted to increase with distance from the jet axis (which is what is observed), while the underlying intrinsic rotation profile does the opposite as expected. The measured velocity shifts in these conditions will give an accurate measure of V_{ϕ} only on transverse distances $> 30 \text{ AU}$, i.e. on the outer jet radius streamline. The self-similar disc wind solution shown here does indeed reproduce the velocity shifts observed with HST/STIS in the DG Tau microjet (see [38]). Full detection of the underlying rotation profile will require spatial resolution $\leq 5 \text{ AU}$ ($0.035''$ at the distance of Taurus). This will be a critical test of the different classes of MHD launching models.

5.4 Excitation Conditions and Mass-Loss Rate Determinations

Projection and convolution effects may also have an effect on excitation conditions and mass-loss rates derived from line ratios/intensities. The effect of stratification of the emitting volume over integrated line ratios has been first pointed

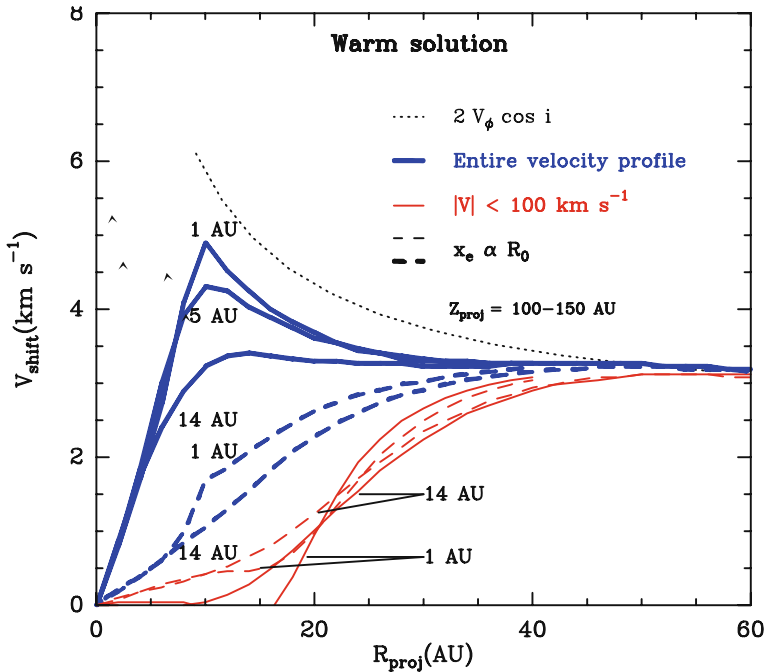


Fig. 17 Illustration of projection and convolution effects on measured transverse velocity shifts. The dotted curve shows the true underlying rotation profile ($2V_\phi \cos i$) of the warm self-similar disc wind solution as a function of the cylindrical radius. Thick curves show the predicted velocity shifts computed using the entire velocity profile as a function of the projected cylindrical radius R_{proj} . Thin curves show predicted transverse velocity shifts for the low-velocity component only (for $|V| < 100 \text{ km s}^{-1}$). Dashed curves show the effect of adopting an ionization law $x_e \propto r_0$ where r_0 is the streamline launching radius. Three spatial resolutions of 1 AU, 5 AU and 14 AU are considered. All model predictions are computed at a projected altitude above the disc $z_{\text{proj}} = 100 - 150 \text{ AU}$. Figure adapted from [38]

out by Safier [44] and is illustrated in Fig. 18. In this study, the author considers different power law distributions for the total element and electronic densities in the emitting volume in both cylindrical and spherical geometries and computes emergent line ratios. The comparison with line ratios expected from homogeneous conditions shows that stratification can have a significant effect on the integrated line ratios.

Complex non-linear effects may thus be introduced by stratification in the elementary emitting volume in previous derivations of excitation conditions (n_e , T_e , x_e) and mass-loss rates from integrated line ratios and emissivities. Of course these effects will be more critical when strong gradients of excitation conditions occur within the elementary emitting volume considered, for example, in studies of the very inner regions of the jets at unsufficient angular and velocity resolution.

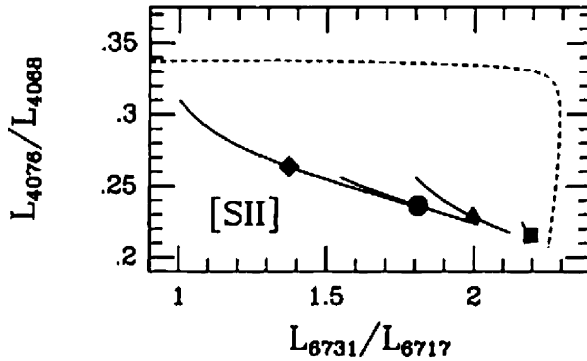


Fig. 18 Effect of stratification of the emitting volume over the [S II] line ratio L_{4076}/L_{4088} versus the [S II] line ratio L_{6731}/L_{6717} . The *thick* line shows the emergent line ratios for different cylindrical and spherical stratification laws of the total and electronic densities in the emitting volume. The *dashed* line shows the values obtained in the case of a spatially homogeneous emitting volume of varying electronic densities (n_e in the range $10\text{--}10^{10}\text{ cm}^{-3}$). Figure from [44]

5.5 Summary

Projection and convolution effects may introduce complex non-linear effects on the derivation of physical quantities from observations. This is particularly critical in the study of the inner regions of jets where strong spatial and kinematical gradients are observed and where both the observational spatial and spectral resolutions are currently limited. A detailed comparison with models involving synthetic observational predictions (taking into account plane of the sky projection effects and convolution with the observational parameters) may often be required to derive meaningful constraints. Alternatively, high angular resolution studies at ≤ 15 AU spatial resolution, where the transverse structure of the jet is well resolved, may offer the prospects for tomographic reconstruction of the underlying 3D structure.

References

1. Anderson, J. M., et al.: *Astrophys. J.* **590**, L107–L110 (2003) 245
2. Arnaud, M., Rothenflug, R.: *Astron. Astrophys. Suppl.* **60**, 425–457 (1985) 224
3. Asplund, M., et al.: In ASP conference Series Vol XXX (astroph/0410214). 229
4. Bacciotti, F., et al.: *Astron. Astrophys.* **296**, 185–200 (1995) 221, 222
5. Bacciotti, F., Eisloffel, J.: *Astron. Astrophys.* **342**, 717–735 (1999) 222, 223, 224, 227, 228, 229, 236, 239
6. Bacciotti, F., et al.: *Astron. Astrophys.* **350**, 917–927 (1999) 224, 227
7. Bacciotti, F., et al.: *Astrophys. J.* **537**, L49–L52 (2000)
8. Bacciotti, F.: *Rev. Mex. Astr. Ap.* **13**, 8 (2002) 224
9. Bacciotti, F., et al.: *Astrophys. J.* **576**, 222–231 (2002) 242, 243, 246
10. Bautista, M. A., Pradhan, A. K.: *Astrophys. J.* **492**, 650 (1998) 218
11. Cabrit, S.: EAS Publications Series, Volume 3, Proceedings of Star Formation and the Physics of Young Stars, held 18–22 September, 2000 in Aussois France. J. Bouvier and J.-P. Zahn (eds.). EDP Sciences, 147–182 (2002) 214, 236
12. Chrysostomou, A.: *Astron. Astrophys.* **482**, 575 (2008) 242, 243

13. Coffey, D., et al.: *Astrophys. J.* **604**, 758–765 (2004) 242, 243
14. Coffey, D., et al.: *Astrophys. J.* **663**, 350 (2007) 242, 243
15. Dougados, C., et al.: *Revista Mexicana de Astronomía y Astrofísica (Serie de Conferencias)* Vol. 13, pp. 43–48 (2002) 235
16. Esteban, C., et al.: *M.N.R.A.S.* **335**, 229 (2004) 229
17. Eisloffel, J., Mundt, R.: *Astron. Astrophys.* **263**, 292 (1992) 241
18. Grevesse, N., Sauval, A. J.: *Space Sci. Rev.* **85**, 161 (1998) 229
19. Hartigan, P., et al.: *Astrophys. J.* **316**, 323–348 (1987) 222, 229, 231
20. Hartigan, P., et al.: *Astrophys. J.* **444**, 943 (1994) 222, 227, 228, 231, 232, 233, 234, 235, 237, 239
21. Hartigan, P., et al.: *Astrophys. J.* **452**, 736 (1995) 231, 232, 234, 237, 238, 239
22. Hartigan, P., et al.: *Astroph. J.* **614**, L69 (2004) 218
23. Hartigan, P., Morse, J.: *Astroph. J.* **660**, 426–440 (2007) 227, 230, 236
24. Hartigan P.: *Jets from Young Stars II, Lecture Notes in Physics*, Vol. 742, Springer-Verlag, Berlin, Heidelberg, p. 15 (2008) 215, 220
25. Hollenbach, D., McKee, C. F.: *Astroph. J.* **342**, 306 (1989) 238
26. Kingdon, J. D., Ferland, G. J.: *Astrophys. J. Lett.*, 106, 205 (1996) 224
27. Landini, M., Monsignori Fossi, B. C.: *Astron. Astrophys. Suppl.* **82**, 229 (1990) 224
28. Lavalley-Fouquet, C.: PhD Thesis, Université J. Fourier, Grenoble (2000) 217, 219, 227, 228, 232, 234, 240
29. Lavalley-Fouquet, C., et al.: *Astron. Astroph. Lett.* **356**, L41 (2000) 234, 235, 239
30. Mendoza, C.: in *Planetary Nebulae*, D. R. Flower (ed.). Reidel, Dordrecht. IAU Symp., 103, 143 (1983) 224
31. Melnikov, S., et al.: *Astron. Astroph.* **483**, 199–208 (2008) 224
32. Nisini, B., et al.: *Astron. Astroph.* **441**, 159–170 (2005) 218, 220, 221, 228, 230, 238
33. Nisini, B., *Jets from Young Stars II, Lecture Notes in Physics*, Vol. 742. Springer-Verlag, Berlin, Heidelberg, p. 79., (2008) 213, 214, 215
34. Nussbaumer, H., Storey, P. J.: *Astron. Astroph.* **193**, 327 (1988) 224
35. Orlando, S., Favata, F.: *Jets from Young Stars II, Lecture Notes in Physics*, Vol. 742. Springer-Verlag, Berlin, Heidelberg, p. 173 (2008) 214
36. Osterbrock D. E.: *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, Mill Valley: University Science Book (1994) 220, 221
37. Pesenti, N., et al.: *Astron. Astroph.* **410**, 155 (2003) 218, 219, 221, 242
38. Pesenti, N., et al.: *Astron. Astroph.* **416**, L9 (2004) 245, 246, 247
39. Podio, L., : *Astron. Astroph.* **456**, 189–204 (2006) 218, 225, 228, 229, 230
40. Porter, J. M., et al.: *Astron. Astroph.* **428**, 327 (2004) 243
41. Ray, T., et al.: In *Protostars and Planets V*, B. Reipurth, D. Jewitt, and K. Keil (eds.), University of Arizona Press, Tucson, p. 231–244 (2007) 226
42. Raga, A. C., Böhm, K. H.: *Astrophys. J.* **308**, 829 (1986) 222
43. Raymond, J. C., et al.: *Astrophys. J.* **326**, 323 (1988) 231
44. Safier, P.: *Astrophys. J.* **392**, 492 (1992) 247, 248
45. Woitas, J., et al.: *Astron. Astrophys.* **432**, 149–160 (2005) 242, 243

Index

A

Active regions, 113, 114, 115, 117–118, 119, 120, 121, 128
Active region transient brightening, 117–118
Alfvén speed, 42, 64, 106, 123, 125, 127, 128, 129, 130, 131, 133, 136, 143, 144
Alfvén wave, 83, 106, 125, 127, 129–130, 133, 153

C

Coronal heating problem, 111, 115–116, 156
Coronal hole, 113, 114, 115, 119, 124
Coronal loop, 114, 120, 124, 125, 126–128, 129, 149, 151, 156, 157
Coronal mass ejection (CME), 117
Current sheet, 135–138, 142, 143, 144, 147–149, 153, 154, 155

F

Fan reconnection, 147
See also Magnetic fan plane
Fast body waves, 127, 128
Fast magnetoacoustic waves, 125–126
Flare, 117, 118, 119, 129, 134, 135, 143
Force-free magnetic field, 123–124
Frozen-flux theorem, 43, 136, 139

K

Kink instability, 41, 147, 148
Kink wave, 128

L

Leaky wave, 128

M

Magnetic carpet, 120, 121, 151
Magnetic fan plane, 145, 146, 154, 158
See also Fan reconnection
Magnetic field recycle, 120, 151

Magnetic flux tube, 126, 147, 148, 150
 tectonics, 151–154
Magnetic neutral point, 134
 O-type, 136, 141, 142
 X-type, 136, 141, 142
Magnetic Reynolds number, 10, 135–136, 143, 144
Magnetic separator, 146, 158
 See also Separator reconnection
Magnetic separatrix surface, 145, 146, 149, 151, 152, 154, 155, 158
Magnetic skeleton, 146–147, 154, 158
Magnetic spine, 145, 146, 154, 158
 See also Spine reconnection
Magnetic topology, 133, 140, 146
Magnetogram, 120–121, 151
MHD equations, 9, 10, 34, 51, 57, 73, 82, 93, 97, 104, 105, 122–124, 139, 140, 141, 196
Microflare, 114, 118–119

N

Nanoflare, 118–119, 151, 156–157
 heating, 151, 156–157

P

Parker's theory of coronal heating, 151
Petschek reconnection, 135, 144
Plasma β , 33, 36, 43, 123, 125
Prominence, 111, 112, 117

Q

Quiet sun region, 114, 115, 117, 118, 119, 120, 121, 158

R

Reconnection, 115, 116, 120, 133–149, 150, 151, 154, 155, 158
 3D, 146, 147–149

jet, 148
plasmoid, 148
rate, 143, 144

S

Sausage wave, 126–127
Separator reconnection, 147
 See also Magnetic separator
Slow body waves, 127
Slow magnetoacoustic waves, 125
Slow MHD shocks, 135, 143
Solar cycle, 111, 113, 117
Solar interior, 109, 111, 115, 129, 150
Sound speed, 36, 64, 73, 88, 104, 106, 125,
 127, 128, 166, 172, 174, 183, 186
Spine reconnection, 147

See also Magnetic spine

Sunspot, 111, 112, 113, 121, 128
Sweet–Parker reconnection, 135, 142–143

T

Tearing mode instability, 135, 139–142

W

Wave dissipation, 124–133
 classical, 129–130
 phase mixing, 130–132, 133, 158
 resonance absorption, 132–133, 158

X

X-ray bright points, 113–114, 118, 119