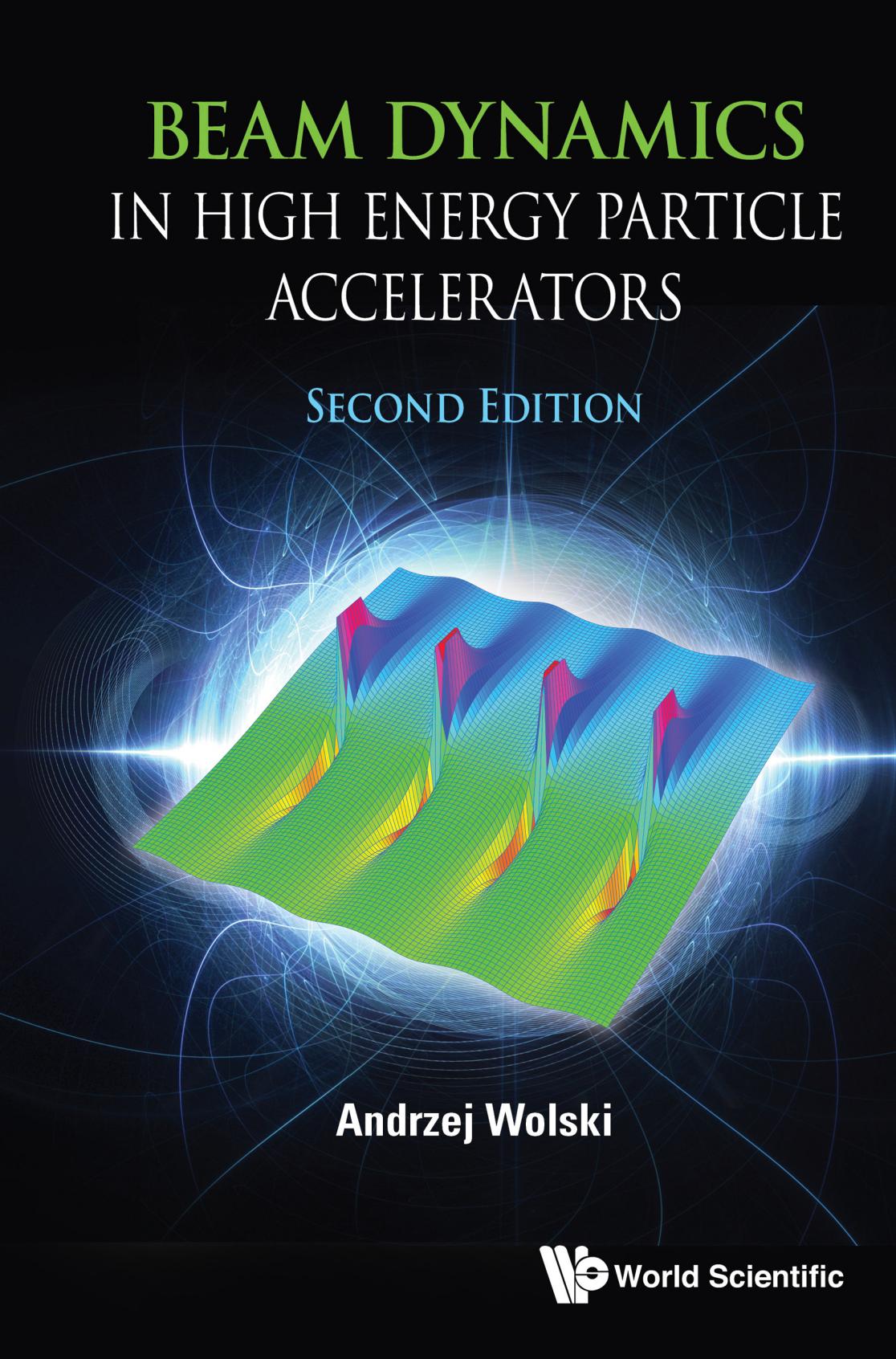


# BEAM DYNAMICS IN HIGH ENERGY PARTICLE ACCELERATORS

SECOND EDITION



Andrzej Wolski

**BEAM DYNAMICS  
IN HIGH ENERGY PARTICLE  
ACCELERATORS**

**SECOND EDITION**

**This page intentionally left blank**

# BEAM DYNAMICS IN HIGH ENERGY PARTICLE ACCELERATORS

SECOND EDITION

**Andrzej Wolski**

University of Liverpool, UK



**World Scientific**

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI • TOKYO

*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

*USA office:* 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

*UK office:* 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Control Number: 2023932604

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

**BEAM DYNAMICS IN HIGH ENERGY PARTICLE ACCELERATORS**

**Second Edition**

Copyright © 2023 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-127-332-2 (hardcover)

ISBN 978-981-127-333-9 (ebook for institutions)

ISBN 978-981-127-334-6 (ebook for individuals)

For any available supplementary material, please visit

<https://www.worldscientific.com/worldscibooks/10.1142/13333#t=suppl>

Desk Editor: Carmen Teo Bin Jie

Printed in Singapore

To my parents, for all the many years of patience and support.

**This page intentionally left blank**

# Contents

<i>Preface to the Second Edition</i>	xiii
<i>Preface to the First Edition</i>	xv
<b>I Electromagnetism and Classical Mechanics</b>	<b>1</b>
1. Electromagnetic Fields in Accelerator Components	3
1.1 Boundary Conditions on Electromagnetic Fields . . . . .	7
1.1.1 Surface of an infinite permeability material . . . . .	8
1.1.2 Surface of an ideal conductor . . . . .	11
1.2 Two-Dimensional Multipole Fields . . . . .	12
1.2.1 Current distribution for a pure multipole . . . . .	18
1.2.2 Geometry of iron-dominated multipole magnets .	22
1.2.3 Multipole decomposition . . . . .	27
1.3 Three-Dimensional Fields . . . . .	33
1.3.1 Cartesian and cylindrical modes . . . . .	34
1.3.2 Generalised gradients . . . . .	42
1.3.3 Analytical expressions for multipole fringe fields .	48
1.4 Fields in Radiofrequency Cavities . . . . .	67
1.4.1 Rectangular cavities . . . . .	70
1.4.2 Cylindrical cavities . . . . .	77
2. Hamiltonian for a Particle in an Accelerator Beam Line	83
2.1 The Hamiltonian for a Straight Beam Line . . . . .	85
2.2 Dynamical Variables for Beam Dynamics . . . . .	91
2.3 The Hamiltonian in a Curved Co-ordinate System . . . . .	96

2.4	Symplectic Transfer Maps and Liouville's Theorem . . . . .	100
<b>II</b>	<b>Single-Particle Linear Dynamics</b>	<b>105</b>
3.	Linear Transfer Maps for Common Components	107
3.1	Drift Space . . . . .	109
3.2	Dipole Magnet . . . . .	112
3.3	Dipole Fringe Fields and Edge Focusing . . . . .	120
3.4	Quadrupole Magnet . . . . .	126
3.5	Solenoid . . . . .	130
3.6	Radiofrequency Cavity . . . . .	134
3.7	Spin Dynamics . . . . .	140
4.	Linear Optics in Uncoupled Beam Lines	151
4.1	A FODO Lattice . . . . .	152
4.2	The Courant–Snyder Parameters . . . . .	156
4.3	Action–Angle Variables . . . . .	160
4.4	Courant–Snyder Parameters in a FODO Beam Line . . . . .	167
4.5	Hill's Equation . . . . .	171
4.6	Courant–Snyder Parameters and Particle Distribution . .	174
5.	Coupled Optics	181
5.1	Transverse–Longitudinal Coupling . . . . .	182
5.1.1	Dispersion . . . . .	182
5.1.2	Momentum compaction and phase slip . . . . .	186
5.1.3	Synchrotron motion . . . . .	192
5.2	Fully Coupled Motion . . . . .	200
5.3	Dispersion Revisited . . . . .	212
5.4	Examples of Coupled Optics . . . . .	214
5.4.1	Uniform solenoid field . . . . .	214
5.4.2	Flat-beam electron source . . . . .	215
5.4.3	Measurement of normal mode emittances . . . . .	219
6.	Linear Imperfections in Storage Rings	227
6.1	The Closed Orbit . . . . .	228
6.2	Dipole Field Errors . . . . .	230
6.3	Quadrupole Alignment Errors . . . . .	235
6.4	Focusing Errors . . . . .	236

6.5	Beam-Based Alignment of Quadrupoles . . . . .	239
6.6	Coupling Errors . . . . .	242
7.	Effects of Synchrotron Radiation	247
7.1	Classical Radiation: Radiation Damping . . . . .	249
7.2	Quantum Radiation: Quantum Excitation . . . . .	262
7.3	Equilibrium Emittance and Lattice Design . . . . .	269
7.3.1	Natural emittance in a FODO storage ring . . . . .	272
7.3.2	Double-bend achromat . . . . .	276
7.3.3	TME lattices and multibend achromats . . . . .	279
7.4	Computation of Equilibrium Emittances . . . . .	285
7.4.1	Effects of betatron coupling . . . . .	286
7.4.2	Effects of vertical dispersion . . . . .	294
7.4.3	Envelope method . . . . .	298
7.5	Synchrotron Radiation and Spin Polarisation . . . . .	305
<b>III Single-Particle Nonlinear Dynamics</b>		<b>311</b>
8.	Examples of Nonlinear Effects in Accelerator Beam Lines	313
8.1	Longitudinal Dynamics in a Bunch Compressor . . . . .	313
8.2	Chromaticity in a Linear FODO Beam Line . . . . .	322
8.3	Chromaticity in Storage Rings . . . . .	327
9.	Representations of Transfer Maps	337
9.1	Lie Transformations . . . . .	339
9.2	Power Series Map for a Sextupole . . . . .	346
9.3	Mixed-Variable Generating Functions . . . . .	351
10.	Particle Tracking by Numerical Integration	359
10.1	The Störmer–Verlet Method . . . . .	360
10.2	Symplectic Integrators . . . . .	366
10.2.1	Splitting methods . . . . .	366
10.2.2	Explicit symplectic integrator for $s$ -dependent fields . . . . .	376
10.2.3	Symplectic Runge–Kutta integrators . . . . .	382

11. Methods for Analysis of Single-Particle Dynamics . . . . .	389
11.1 A Lie Transformation Example: the $-I$ Transformer . . . . .	390
11.2 Canonical Perturbation Theory . . . . .	393
11.2.1 Dipole perturbations: closed orbit distortion . . . . .	401
11.2.2 Quadrupole perturbations: focusing errors . . . . .	404
11.2.3 Skew quadrupole perturbations: coupling . . . . .	406
11.2.4 Sextupole perturbations . . . . .	414
11.3 Resonances and Dynamic Aperture . . . . .	417
11.4 Normal Form Analysis . . . . .	431
11.5 A Numerical Method: Frequency Map Analysis . . . . .	442
<b>IV Collective Effects</b> . . . . .	<b>451</b>
12. Space Charge . . . . .	453
12.1 The Kapchinsky–Vladimirsky Distribution . . . . .	455
12.2 The Envelope Equations for the KV Distribution . . . . .	460
12.3 Elliptically Symmetric Non-KV Distributions . . . . .	465
12.4 Space-Charge Tune Shifts . . . . .	474
12.5 Matching a Continuous Beam to a Solenoid Field . . . . .	483
12.6 Longitudinal Dynamics with Space Charge . . . . .	487
12.7 Beam-Beam Effects . . . . .	498
13. Scattering Effects . . . . .	503
13.1 Touschek Effect . . . . .	504
13.2 Intrabeam Scattering . . . . .	514
13.2.1 Piwinski formulae . . . . .	517
13.2.2 Bjorken–Mtingwa formulae . . . . .	522
13.2.3 High energy approximation . . . . .	524
14. Wake Fields, Wake Functions and Impedance . . . . .	527
14.1 Wake Fields in a Resonant Cavity . . . . .	529
14.2 Resistive-Wall Wake Fields . . . . .	537
14.3 Wake Functions . . . . .	549
14.4 Impedance . . . . .	557
15. Coherent Instabilities . . . . .	569
15.1 Coupled-Bunch Instabilities . . . . .	570

15.1.1	Transverse modes . . . . .	570
15.1.2	Longitudinal modes . . . . .	578
15.2	Potential-Well Distortion . . . . .	583
15.3	Coasting Beams: Microwave Instability . . . . .	587
15.4	Single-Bunch Instabilities . . . . .	598
15.4.1	Head-tail instability . . . . .	599
15.4.2	Sacherer's integral equation . . . . .	613
15.4.3	Discrete modes: Robinson instability . . . . .	622
15.4.4	Mode coupling . . . . .	628
<i>Bibliography</i>		635
<i>Index</i>		653

**This page intentionally left blank**

# Preface to the Second Edition

Accelerator physics is an area of active research, and there have inevitably been many developments in the field since the first edition of this book was published, nearly a decade ago. Continuing improvements in technology and techniques mean that accelerators can now reach parameter regimes that were not accessible at that time: this is exemplified by the rapid progress in x-ray free-electron lasers, which enabled the growing field of ultra-fast science, and with ultra-low emittance storage rings that can now operate as diffraction-limited light sources. There has also been significant progress in recent years with plasma-wakefield acceleration techniques, which offer the prospect of achieving accelerating gradients orders of magnitude higher than those that can be achieved in conventional radio-frequency (RF) accelerating structures. However, control of the beam quality and stability in plasma-wakefield accelerators remains a considerable challenge, so facilities based on this technology for scientific, industrial or medical applications still seem some way off. In terms of beam dynamics, the principles underlying the operation of RF-based accelerators today are the same as they were ten (or more) years ago. Hence, while this second edition has offered the opportunity to add some selected, specific topics, very major changes to the scope have not been needed. Of course, the second edition has also allowed me to correct a number of regrettable errors in the first edition, to revise and improve the text in many places, and to update and add to the references.

New topics in this edition include an analytical treatment of multipole fringe fields (in Chapter 1), the measurement of normal mode emittances (Chapter 5), and the Störmer–Verlet method for numerical integration of conservative systems (Chapter 10). The inclusion of all these topics has been motivated not only by their relevance to my own research and teaching

activities in the past few years, but also by their potentially very widespread practical value in accelerator beam dynamics. A further topic that is of increasing interest and importance in accelerator physics is the use of machine learning for a wide range of tasks: this is an intriguing and rapidly developing area that promises to have a significant impact on the way that studies in beam dynamics are performed. However, machine learning is a very large subject that would require more space than is available in this book to do it justice. Also, mindful of my own very limited experience and still more limited expertise in this area, it seems advisable to leave it to others to contribute to the texts available on the topic. For similar reasons, I decided also not to include discussion of coherent radiation (whether in the form of coherent synchrotron radiation from dipole magnets, or in free-electron lasers). In any case, there are already a number of excellent textbooks covering this important topic, and there will no doubt be more in the future.

The errors in the first edition are, to my great regret, too numerous to list here. Although they largely consist of minor typographical mistakes, I am all too well aware of the frustration that can result even from just an incorrect minus sign or a missing factor in a term in an equation. I am sincerely grateful to the readers who have taken the trouble to contact me, to let me know when they have found an error. Needless to say, I have taken the opportunity in this edition to correct all those errors of which I am aware, and while it seems overly-optimistic to expect that all the errors have been put right, I hope that there are not too many remaining.

Finally, I must express my thanks to the many readers and colleagues who have contributed to this book, either indirectly through thought-provoking discussions on particular topics, or more directly through comments or questions on sections of the first edition. There are too many people to whom I am indebted in various ways for me to name them all, but I feel that I must thank especially Dr. Bruno Muratori of Daresbury Laboratory, and Prof. Moses Chung of UNIST for many enlightening discussions that have influenced this new edition in a number of ways. My thanks also go, of course, to all who have helped me explore the fascinating field of accelerator beam dynamics.

# Preface to the First Edition

Given a beam of charged particles in an accelerator, the challenge for the accelerator physicist is to explain and control the behaviour of that beam. Beam dynamics provides the tools for describing and understanding the particle motion. The electromagnetic fields that determine the dynamics may arise from components such as multipole magnets or radiofrequency cavities, from stray particles in the vacuum chamber, from the beam itself or from other beams of particles in the accelerator. An understanding of the fields generated by these sources and their effect on the beam dynamics is needed for the design, commissioning and operation of an accelerator.

In calculating the motion of particles in an accelerator, the first step is to select the appropriate physical principles. Here, we appear to be in good shape, since nearly all the important features of beam dynamics can be described and explained using physical laws that have been known for more than a century. In particular, the electromagnetic fields in a beam line must satisfy Maxwell's equations, and the motion of a charged particle through those fields is determined by Hamilton's equations, with an appropriate (relativistic) Hamiltonian. There are certainly some aspects of beam behaviour that give glimpses beyond the regime of purely classical physics: these include the quantum excitation of particle oscillations in a storage ring from synchrotron radiation and effects associated with spin polarisation. But from rather few basic ingredients, there results a rather impressive and somewhat daunting diversity of phenomena. Furthermore, many of the problems that occur in practical situations are rather difficult to solve. Much of the discussion in this book concerns techniques for finding approximate solutions, with certain desirable properties, to the classical equations of motion in various situations in accelerator beam lines. Different techniques are appropriate for different cases.

My main motivation for writing this book (apart from an intrinsic interest in the subject) was to bring together in one place the principles and methods that I have found most important, interesting and useful in my work as an accelerator physicist, dealing with situations that range from the relatively simple to the rather complex. The simpler cases can often be modelled adequately using straightforward techniques based on various approximations. However, most accelerator physicists will sooner or later encounter complicated systems that require accurate descriptions of the beam dynamics: and the simplified techniques cannot always be extended to handle such cases. The approach I have taken in this book is to develop the theory of beam dynamics in such a way that it can be applied, where necessary, to some of the more complex situations encountered in high energy particle accelerators. The aim is to provide a solid foundation for a deeper understanding of the various topics than would be provided by a simple discussion at an introductory level.

Although it is assumed that readers will be familiar with electromagnetism and classical mechanics, these topics are covered briefly in introductory chapters because of the way that nearly all of beam dynamics follows naturally from them. Also, these theories are wide enough that it is worth emphasising the aspects that are of particular relevance to the behaviour of charged particles in accelerators. Without the introductory chapters, this book would be incomplete. Special relativity is not discussed explicitly, but (by definition) is intrinsic to just about all aspects of beam dynamics in high energy accelerators.

In addition to some familiarity with electromagnetism, classical mechanics and special relativity, it is assumed that readers have at least a basic knowledge of the general features and principles of operation of high energy particle accelerators. This includes, for example, the purpose and structure of components such as dipole and quadrupole magnets, solenoids and rf cavities. Although these (and other) components are briefly introduced and described in the appropriate places, my intention is simply to provide a context and to define a starting point for the development of the relevant beam dynamics.

When planning this book, I quickly realised that it would be necessary to make some compromise between the range of topics covered and the depth to which they are treated. The contents are guided by my own experience as an accelerator physicist, and fall into one (or more) of three categories. First, I have included those topics that are widely considered, with some justification, to be fundamental to the study of beam dynamics.

In the second category are results and methods that I use (or see used) on a regular basis. Finally, some subjects are covered that may not be thought essential for practical applications, but are included because they provide some insight into certain aspects of beam behaviour, or simply because they involve interesting physics. My hope is that I have not got the balance between range and depth hopelessly wrong, and that this book will be of some use to those relatively new to the subject, as well as to more experienced accelerator physicists who may find it convenient to have the various topics brought together in one place.

Inevitably, there are important topics that ought to have been included but have been omitted for one reason or another. For example, coherent synchrotron radiation and free-electron lasers are mentioned only in passing; beam cooling (apart from synchrotron radiation damping) is not mentioned at all, nor are ion effects, electron cloud and some other instabilities. The discussions here of spin dynamics, beam-beam effects and numerous other phenomena completely fail to do justice to these interesting and important topics in beam dynamics. Furthermore, regarding those topics that are included, I fear that despite my best efforts at accuracy and consistency, there will inevitably be some errors, for which I apologise in advance.

I faced particular challenges at times in finding appropriate notation. Although I have tried to follow convention as far as possible, there are cases where several different notations or definitions are in use by the accelerator community, or where the conventional notation used in one specific topic conflicts with that used in another. I am afraid that I did not always find satisfactory solutions. I may be fortunate enough to have the opportunity to address some of these deficiencies in the future; but for the time being I just hope that this book, as far as it goes, can be of some use for those with an interest in beam dynamics.

Finally, I should like to acknowledge and thank all those who have helped me in my own efforts to learn accelerator physics. I have had the privilege of working with some wonderful people and outstanding physicists from a number of different laboratories and universities. As I still have a very great deal to learn, I look forward to many future interesting and enlightening discussions.

**This page intentionally left blank**

PART I

## **Electromagnetism and Classical Mechanics**

**This page intentionally left blank**

## Chapter 1

# Electromagnetic Fields in Accelerator Components

The properties of the electromagnetic fields in an accelerator are fundamental to the beam dynamics. Electric fields can be used to accelerate charged particles, and to steer and focus beams of particles. Magnetic fields cannot be used for particle acceleration, but can be more effective than electric fields for steering and focusing beams when the particles have high energy (that is, when the velocities of the particles are relativistic). Complicated configurations of electromagnetic fields are used for manipulating beams in special ways; for example, for injecting particles into a storage ring or for the production of synchrotron radiation having particular characteristics. Before discussing accelerator beam dynamics, it is appropriate to consider in some detail the features of a few of the components most widely used in accelerators. In this chapter, we shall pay specific attention to components that will be encountered regularly in later chapters, namely multipole magnets (dipoles, quadrupoles, sextupoles etc.) and radiofrequency (rf) cavities. Our main goal is to describe the fields they produce, and to show how these fields depend on particular features of the component concerned. We shall also discuss mathematical techniques that can be used for constructing a representation of a given magnetic field as a function of the spatial co-ordinates, starting from a set of numerical field data. These techniques will not be needed until later chapters, but their description fits best in the context of a discussion of the electromagnetic fields in accelerator components.

Some of the results that are important for a study of beam physics will be familiar from any general course on electromagnetism, and are well covered by standard texts (for example, [Jackson (1998); Garg (2012); Zangwill (2013)]). However, other results are more specialised for accelerator beam dynamics. It is in any case interesting to see how the results that we shall

need can be developed from fundamental physics. We shall begin, therefore, with Maxwell's equations, and show how they lead to constraints on the behaviour of electric and magnetic fields on boundaries (such as those formed, for example, by the pole faces of an iron-cored electromagnet or the walls of an rf cavity). Multipole magnets are a particularly important class of components in accelerator beam lines, where they can be used for steering (dipoles), focusing (quadrupoles) and for correcting chromatic aberrations (sextupoles). We shall show how multipole fields provide solutions to Maxwell's equations in free space, and, using the boundary conditions, discuss how the required fields can be produced in practice. Our focus is on the physics, rather than on the technology or engineering: the technical aspects are extremely important (and challenging) but the subject is too large and varied for us to cover more than what will be needed for a study of the beam dynamics. However, with the goal of presenting a discussion that is as self-contained as possible, we start from first principles.

Maxwell's equations may be written in differential form as follows:

$$\nabla \cdot \mathbf{D} = \rho, \quad (1.1)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (1.2)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \quad (1.3)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (1.4)$$

The fields **B** (magnetic flux density) and **E** (electric field strength) determine the force on a particle of charge  $q$  travelling with velocity **v** (the Lorentz force equation):

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (1.5)$$

The electric displacement **D** and magnetic intensity **H** are related to the electric field and magnetic flux density by:

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (1.6)$$

$$\mathbf{B} = \mu \mathbf{H}. \quad (1.7)$$

The electric permittivity  $\epsilon$  and magnetic permeability  $\mu$  depend on the medium within which the fields exist. The values of these quantities in vacuum are fundamental physical constants. In SI units:

$$\mu_0 = 4\pi \times 10^{-7} \text{ H m}^{-1},$$

$$\epsilon_0 = \frac{1}{\mu_0 c^2},$$

where  $c$  is the speed of light in vacuum. The permittivity and permeability of a material characterise the response of that material to electric and magnetic fields. In simplified models, they are often regarded as constants for a given material; however, in reality the permittivity and permeability can have a complicated dependence on the fields that are present in. Note that the *relative permittivity*  $\varepsilon_r$  and the *relative permeability*  $\mu_r$  are frequently used. These are dimensionless quantities, defined by:

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0}, \quad (1.8)$$

$$\mu_r = \frac{\mu}{\mu_0}. \quad (1.9)$$

The quantities  $\rho$  and  $\mathbf{J}$  are respectively the electric charge density (charge per unit volume) and electric current density ( $\mathbf{J} \cdot \mathbf{n}$  is the charge crossing unit area perpendicular to unit vector  $\mathbf{n}$  per unit time). The charge density and current density may be regarded as sources of the electromagnetic field. Problems in electromagnetism often take one of two forms: either, to find the electromagnetic fields resulting from given charge and current distributions; or, to find the charge and current distributions needed to produce specified fields. The second kind of problem in particular will be discussed in this chapter. A third kind of problem, which is often encountered in accelerator beam dynamics, and which we shall discuss in later chapters, is to find a ‘self-consistent’ solution for a system of charges: that is, to find the time-dependent electromagnetic fields generated by a system of electric charges (and currents) that move under the influence of those fields.

Two of Maxwell’s equations, (1.2) and (1.4), are independent of the sources  $\rho$  and  $\mathbf{J}$ , and are sometimes referred to as the ‘homogeneous’ equations; the other two of Maxwell’s equations, (1.1) and (1.3), are dependent on  $\rho$  and  $\mathbf{J}$ , and are referred to as the ‘inhomogeneous’ equations. When the charge density and current density are specified (as functions of space and, generally, time), Maxwell’s equations (1.1)–(1.3) can be integrated to find possible electric and magnetic fields in the system. Usually, however, the solution one finds by integration is not unique: for example, the field within an accelerator dipole magnet may be modified by propagating an electromagnetic wave through the magnet. However, by imposing certain constraints (for example, that the fields within a magnet are independent of time) it is possible to obtain a unique solution for the fields in a given system of electric charges and currents.

Most realistic situations are sufficiently complicated that exact solutions

to Maxwell's equations cannot be obtained analytically. A variety of computer codes are available that can provide solutions numerically once the charges, currents and properties of the materials present are all specified. Solving for the fields in realistic (three-dimensional) systems often requires a reasonable amount of computing power; some sophisticated techniques have been developed for solving Maxwell's equations numerically with good efficiency. We do not consider such techniques here, but focus instead on some analytical solutions that may be obtained in idealised situations. Although the solutions in such cases may not be sufficiently accurate to complete the design of a real accelerator magnet, the analytical solutions do provide a useful basis for describing the fields in real magnets, and also provide some important connections with the beam dynamics in an accelerator.

An important feature of Maxwell's equations is that, for systems containing materials with constant permittivity and permeability (i.e. permittivity and permeability that are independent of the fields present), the equations are linear in the fields and sources. That is, each term in the equations involves a field or a source to (at most) the first power, and products of fields or sources do not appear. As a consequence, the *principle of superposition* applies: for example, if  $\mathbf{B}_1$  and  $\mathbf{B}_2$  are solutions of Maxwell's equations with the current densities  $\mathbf{J}_1$  and  $\mathbf{J}_2$ , then the field  $\mathbf{B}_T = \mathbf{B}_1 + \mathbf{B}_2$  will be a solution of Maxwell's equations, with the source given by the total current density  $\mathbf{J}_T = \mathbf{J}_1 + \mathbf{J}_2$ . This means that it is possible to represent complicated fields as superpositions of simpler fields. An important and widely used analysis technique for accelerator magnets is to decompose the field (determined either from a magnetic model or from measurements of the field in an actual magnet) into a set of multipoles. While the ideal goal would often be to produce a field consisting of a single multipole component, this is never perfectly achieved in practice: the multipole decomposition indicates the extent to which components other than the 'desired' multipole are present. Multipole decompositions also produce useful information for modelling the dynamics of beams of particles moving through a given magnet. Although the principle of superposition strictly only applies in systems where the permittivity and permeability are independent of the fields, it is always possible to perform a multipole decomposition of the fields in free space (for example, in the interior of a vacuum chamber), since in that region the permittivity and permeability are constants. However, it should be remembered that for nonlinear materials (where the permeability, for example, depends on the magnetic field strength), the field inside the material from which the magnet is made will

not necessarily be as expected, if one were to simply add together the fields corresponding to the multipole components.

Solutions to Maxwell's equations lead to a rich diversity of phenomena, including the fields around charges and currents in certain simple configurations, and the generation, transmission and absorption of electromagnetic radiation. Many existing texts cover these phenomena in detail; see, for example, the text by Jackson [Jackson (1998)]. In the following sections, we discuss analytical solutions to Maxwell's equations for situations relevant to some types of magnets commonly used in accelerators. These include multipoles (dipoles, quadrupoles, sextupoles and higher-order multipoles), solenoids and insertion devices (undulators and wigglers).

The first step is to derive some important results for the behaviour of electromagnetic fields at boundaries between regions with different electromagnetic properties: these results will be needed when we come to discuss the geometries of magnets required to produce fields with specified characteristics. We shall then consider two-dimensional magnetic fields; that is, magnetic fields that are independent of one co-ordinate (generally, the co-ordinate representing the direction of motion of the beam). We shall show that multipole fields are solutions of Maxwell's equations, and we shall derive the current distributions needed to generate 'pure' multipole fields. We shall then discuss multipole decompositions, and compare techniques for determining the multipole components present in a given field from numerical field data (which may be obtained from a model or from measurements). Finally, we shall discuss how the two-dimensional multipole decomposition may be extended to three-dimensional fields, to include (for example) insertion devices and fringe fields in multipole magnets.

## 1.1 Boundary Conditions on Electromagnetic Fields

Before beginning a detailed discussion of electromagnetic fields in accelerator components, it is useful to obtain some results describing the behaviour of electric and magnetic fields at boundaries between different materials. These results have important implications for the geometries of accelerator magnets and rf cavities: it is the behaviour of the fields on the surfaces of magnetic materials and electrical conductors that determine what these devices look like and how they work. There are two idealised cases that we consider. First, we shall discuss the behaviour of magnetic fields at the surface of a material with infinite magnetic permeability: we will find that the magnetic field at the surface of such a material must be perpendicular

to the surface. Then, we shall discuss electromagnetic fields at the surface of an ideal conductor (that is, a material with infinite electrical conductivity). The results we obtain are standard results in electromagnetism, so we will only give brief outlines of the derivations. Our main purpose is to remind the reader of results that we will use when discussing how to generate electromagnetic fields with the particular properties required to control beams of charged particles in accelerators.

### 1.1.1 *Surface of an infinite permeability material*

Consider a boundary between a non-magnetic material (permeability  $\mu_1 \approx \mu_0$ ) and a material with very large permeability,  $\mu_2 \rightarrow \infty$ . We shall show that the magnetic field at the boundary must be perpendicular to the boundary. As we shall see in Section 1.2.2, this result will allow us to determine the geometry of the pole face needed in an iron-dominated magnet to generate a magnetic field of a specified shape.

Consider a loop  $\mathcal{C}$  that spans the boundary as shown in Fig. 1.1. Integrating  $\nabla \times \mathbf{H}$  over the area  $\mathcal{S}$  bounded by the loop  $\mathcal{C}$  (with area element

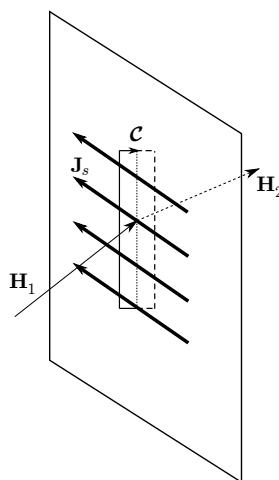


Fig. 1.1 Behaviour of the tangential component of a magnetic field at a plane boundary between different media. A surface current  $\mathbf{J}_s$  flows on the boundary. The change in the component of the magnetic intensity  $\mathbf{H}$  parallel to the boundary is equal to the component of  $\mathbf{J}_s$  perpendicular to the projection of  $\mathbf{H}$  onto the boundary.

$d\mathbf{A}$ ), we obtain from Stokes' theorem:

$$\int_{\mathcal{S}} \nabla \times \mathbf{H} \cdot d\mathbf{A} = \oint_{\mathcal{C}} \mathbf{H} \cdot d\mathbf{l}. \quad (1.10)$$

Now we take the limit where the two sides of the loop parallel to the boundary, of length  $\Delta\ell$ , lie an infinitesimal distance apart, on either side of the boundary. Then, using (1.10) and Maxwell's equation (1.3), we obtain, for small  $\Delta\ell$ :

$$H_{\parallel 1}\Delta\ell - H_{\parallel 2}\Delta\ell = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{A} + \int_{\mathcal{S}} \frac{\partial \mathbf{D}}{\partial t} \cdot d\mathbf{A}, \quad (1.11)$$

where  $H_{\parallel 1}$  and  $H_{\parallel 2}$  are the components of the magnetic intensity  $\mathbf{H}$  parallel to the boundary on either side of the boundary. In the limit that the area  $\mathcal{S}$  bounded by the loop becomes infinitesimally small, the integral of the rate of change of the electric displacement over  $\mathcal{S}$  vanishes, and we are left with:

$$H_{\parallel 1} - H_{\parallel 2} = J_{s\perp}, \quad (1.12)$$

where  $J_{s\perp}$  is the *surface current density* (with SI units  $\text{A m}^{-1}$ ) flowing on the boundary, in a direction perpendicular to the tangential component of the magnetic intensity at the boundary. Equation (1.12) is quite general: so far we have made no assumptions about the properties of the materials on either side of the boundary. If the materials have finite conductivity, then the boundary itself (a layer of zero thickness) must have infinite resistance, and no surface current can flow. Then, the tangential component of the magnetic intensity must be continuous across the boundary. For the magnetic field  $\mathbf{B}$ , this means that:

$$\frac{B_{\parallel 1}}{\mu_1} = \frac{B_{\parallel 2}}{\mu_2}. \quad (1.13)$$

Assuming that the magnetic field  $\mathbf{B}$  remains finite, and if  $\mu_1$  is also finite, we must have:

$$\lim_{\mu_2 \rightarrow \infty} B_{\parallel 1} \rightarrow 0. \quad (1.14)$$

At the surface of a material with infinite permeability, the component of the magnetic field parallel to the surface vanishes; the magnetic field must therefore be perpendicular to the surface.

Now let us consider the component of the magnetic field normal to a boundary between two different media. We shall show that the normal component of the magnetic field must be continuous across a boundary, regardless of the properties of the materials on either side of the boundary.

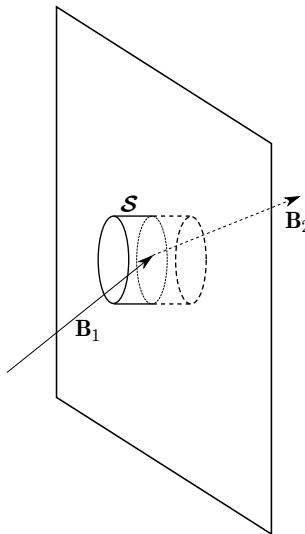


Fig. 1.2 Behaviour of the normal component of a magnetic field at a boundary between different media. By applying Gauss' theorem to Maxwell's equation (1.2) integrated over the volume enclosed by a cylindrical surface  $\mathcal{S}$ , we find that the normal component of a magnetic field  $\mathbf{B}$  is always continuous across a boundary.

First, consider a cylindrical surface  $\mathcal{S}$  as shown in Fig. 1.2, constructed so that the flat ends of the cylinder are an infinitesimal distance apart, on either side of the boundary. Taking Maxwell's equation (1.2), and integrating over the volume  $\mathcal{V}$  enclosed by the surface  $\mathcal{S}$  gives:

$$\int_{\mathcal{V}} \nabla \cdot \mathbf{B} dV = \int_{\mathcal{S}} \mathbf{B} \cdot d\mathbf{A} = 0, \quad (1.15)$$

where we have used Gauss' theorem to convert the integral over the volume  $\mathcal{V}$  into an integral over the surface  $\mathcal{S}$  bounding  $\mathcal{V}$ . In the limit of infinitesimal length of the cylinder, only the field over the ends of the cylinder contributes to the surface integral, and we obtain:

$$B_{\perp 1} = B_{\perp 2}, \quad (1.16)$$

where  $B_{\perp 1}$  and  $B_{\perp 2}$  are the normal components of the magnetic field on either side of the boundary.

Equation (1.16) is a general condition at the boundary between any two materials; there is no assumption about the electrical or magnetic properties of the materials. However, in the case that one of the materials has infinite permeability, it leads to the result that the magnetic intensity  $\mathbf{H}$  vanishes

within a material of infinite permeability. This is because the only field component present at the surface must be normal to the surface. But the normal component of the magnetic field  $\mathbf{B}$  is continuous across the surface; therefore, if the magnetic field is finite, it follows from  $\mathbf{H} = \mathbf{B}/\mu$  that, in the limit  $\mu \rightarrow \infty$ , the magnetic intensity vanishes.

### 1.1.2 Surface of an ideal conductor

The fields within an rf cavity are constrained by the behaviour of the fields at the walls of the cavity. Many of the important properties of rf cavities can be understood from simple models that involve taking the limit in which the conductivity of the walls is infinite, i.e.  $\sigma \rightarrow \infty$ . In that case, the tangential component of the electric field and the normal component of the magnetic field vanish at the walls. We shall show that this behaviour of the fields follows from Maxwell's equations, together with some simple assumptions about the properties of an 'ideal' conductor.

To start, consider a boundary between two different materials, without any assumptions about the properties of the materials. Following the same argument that led to (1.12), but starting from Maxwell's equation, (1.4) leads to the condition that the tangential component of the electric field must be continuous across a boundary:

$$E_{\parallel 1} = E_{\parallel 2}. \quad (1.17)$$

Similarly, following the same argument that led to (1.16), but starting from Maxwell's equation (1.1) leads to:

$$D_{\perp 1} - D_{\perp 2} = \rho_s, \quad (1.18)$$

where  $\rho_s$  is the *surface charge density* (with SI units  $\text{C m}^{-2}$ ), that is, the charge per unit area existing purely on the boundary.

Now consider the case that the material on one side of the boundary is an ideal conductor. In such a material, no electric field can exist: the charges within the material would simply flow (instantaneously, in the ideal case) to cancel any field that is applied. Since  $\mathbf{E} = 0$  within the conductor, it follows immediately from (1.17) that at the surface of the conductor the tangential component of the electric field must vanish;

$$E_{\parallel} = 0. \quad (1.19)$$

We cannot make a similar statement about the component of the electric field perpendicular to the boundary: it is possible for charge to accumulate on the surface, leading to a discontinuity in the electric displacement, and

hence in the electric field. In principle, electric field (of any strength) can exist at the surface of an ideal conductor (on the non-conducting side of the boundary), but the direction of the field must be perpendicular to the surface.

The behaviour of the magnetic field at the surface of a conductor is more complicated, because it is difficult to make any definitive statement about the magnetic field within a conductor, even in the case that the conductivity is infinite. If the conductor is a superconductor, we know that magnetic flux will be excluded from within the material (the Meissner effect), which suggests that  $\mathbf{B} = 0$  within the conductor. However, in reality a magnetic field can penetrate a short distance (the London penetration depth) into a superconductor. In type II superconductors, flux lines can penetrate a considerable distance through the material. We shall put these difficulties aside, and simply assume that the magnetic field is zero within an ideal conductor. In practice, this assumption works quite well: we shall mainly be concerned with the behaviour of the magnetic field at the wall of an rf cavity, where the electric and magnetic fields oscillate at high frequency. The oscillating fields will induce currents in the conducting walls of the cavity, in such a way as to cancel the fields within the walls. Since the normal component of the magnetic field must be continuous across any boundary (1.16), we can say that the normal component of the magnetic field at the wall of an rf cavity must be zero:

$$B_{\perp} = 0. \quad (1.20)$$

We cannot make a similar statement for the tangential component of the magnetic field, because some discontinuity is allowed in the tangential component if there is a surface current existing on the boundary (1.12). Even in a normal conductor (i.e. not an ideal conductor or a superconductor), the currents induced by high frequency fields near the surface can be large enough to lead to an effective cancellation of the magnetic field within the conductor. As a result, it is possible for a magnetic field of any strength to exist at the surface of an ideal conductor, but the direction of the field must be parallel to the surface.

## 1.2 Two-Dimensional Multipole Fields

Magnetic fields provide a convenient way to control beams of high energy charged particles. Magnets that can be described in terms of two-dimensional multipole fields are particularly useful: for example, dipole

fields can be used to control the direction of a beam and quadrupole fields can be used to focus a beam. In this section we define two-dimensional multipole fields and discuss some of their characteristics. Their effects on the motion of particles in an accelerator beam line are discussed in later chapters. More general (three-dimensional) fields are also of importance in accelerators: these are discussed in Section 1.3.

Consider a region of space free of charges and currents: for example, the interior of an accelerator vacuum chamber (at least, in an ideal case, and when the beam is not present). If the fields within this region of space are static, then the magnetic field must satisfy:

$$\nabla \cdot \mathbf{B} = 0, \quad (1.21)$$

$$\nabla \times \mathbf{B} = 0. \quad (1.22)$$

Equation (1.21) is just Maxwell's equation (1.2), and (1.22) follows from Maxwell's equation (1.3) given that  $\mathbf{J} = 0$ ,  $\mathbf{B} = \mu_0 \mathbf{H}$ , and derivatives with respect to time vanish. Equations (1.21) and (1.22) are satisfied by a field  $\mathbf{B} = (B_x, B_y, B_z)$  with  $B_z$  constant, and  $B_x, B_y$  given by:

$$B_y + iB_x = C_n(x + iy)^{n-1}, \quad (1.23)$$

where  $C_n$  is a (complex) constant. Note that the field components  $B_x$  and  $B_y$  are real, and are obtained from the imaginary and real parts of the right-hand side of (1.23). We can show that the field (1.23) satisfies (1.21) and (1.22) as follows. If we apply the differential operator:

$$\frac{\partial}{\partial x} + i \frac{\partial}{\partial y}, \quad (1.24)$$

to the left-hand side of (1.23), we obtain:

$$\left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) (B_y + iB_x) = \left( \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) + i \left( \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} \right). \quad (1.25)$$

Applied to the right-hand side of (1.23), the operator (1.24) gives:

$$\left( \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) C_n(x + iy)^{n-1} = 0. \quad (1.26)$$

Combining (1.23), (1.25) and (1.26), we find:

$$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 0, \quad (1.27)$$

$$\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} = 0. \quad (1.28)$$

Since  $B_z$  is constant, any derivatives of  $B_z$  vanish; furthermore,  $B_x$  and  $B_y$  are independent of  $z$ , so any derivatives of these co-ordinates with respect

to  $z$  vanish. Thus we conclude that the field (1.23) satisfies (1.21) and (1.22).

Fields of the form (1.23) are known as *multipole fields*. The index  $n$  (an integer) indicates the *order* of the multipole:  $n = 1$  is a dipole field,  $n = 2$  is a quadrupole field,  $n = 3$  is a sextupole field, and so on. A solenoid field has  $C_n = 0$  for all  $n$ , and  $B_z$  non-zero; usually, a solenoid field is not considered a multipole field, and we assume (unless stated otherwise) that in a multipole magnet  $B_z = 0$ . We can apply the principle of superposition to construct a field solving Maxwell's equations, by adding together a set of multipole fields:

$$B_y + iB_x = \sum_{n=1}^{\infty} C_n (x + iy)^{n-1}. \quad (1.29)$$

A ‘pure’ multipole field of order  $n$  has  $C_n \neq 0$  for only a single value of  $n$ .

The coefficients  $C_n$  in (1.29) characterise the strength and orientation of each multipole component in a two-dimensional magnetic field. It is sometimes more convenient to express the field using polar co-ordinates, rather than Cartesian co-ordinates. Writing  $x = r \cos(\theta)$  and  $y = r \sin(\theta)$ , (1.29) becomes:

$$B_y + iB_x = \sum_{n=1}^{\infty} C_n r^{n-1} e^{i(n-1)\theta}. \quad (1.30)$$

By writing the multipole expansion in this form, we see immediately that the strength of the field in a pure multipole of order  $n$  varies as  $r^{n-1}$  with distance from the magnetic axis. We can go a stage further, and express the field in terms of polar components:

$$\begin{aligned} B_y + iB_x &= B_r \sin(\theta) + B_\theta \cos(\theta) + iB_r \cos(\theta) - iB_\theta \sin(\theta) \\ &= (B_\theta + iB_r) e^{-i\theta}. \end{aligned} \quad (1.31)$$

Thus:

$$B_\theta + iB_r = \sum_{n=1}^{\infty} C_n r^{n-1} e^{in\theta}. \quad (1.32)$$

From (1.32), we see that, for a pure multipole of order  $n$ , rotation of the magnet through  $\pi/n$  around the  $z$  axis simply changes the sign of the field. We also see that if we write:

$$C_n = |C_n| e^{i\varphi_n}, \quad (1.33)$$

then the value of  $\varphi_n$  (the phase of  $C_n$ ) determines the orientation of the field:

$$B_\theta + iB_r = \sum_{n=1}^{\infty} |C_n| r^{n-1} e^{in(\theta + \varphi_n/n)}. \quad (1.34)$$

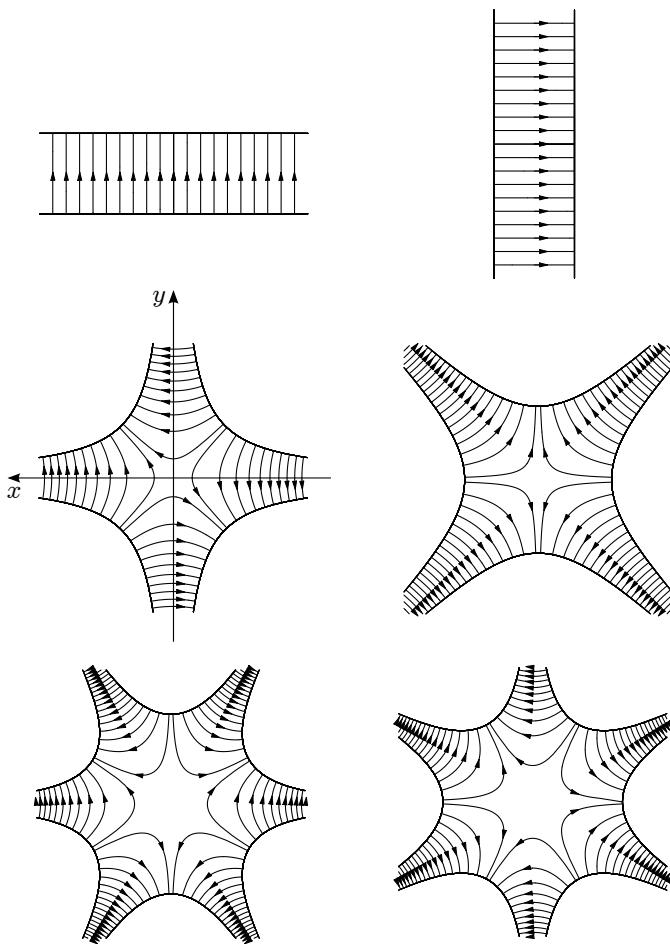


Fig. 1.3 ‘Pure’ multipole fields. Top: dipole. Middle: quadrupole. Bottom: sextupole. Fields on the left are normal ( $a_n = 0$ ); those on the right are skew ( $b_n = 0$ ). Co-ordinate axes are shown only for the normal quadrupole field, but similar axes apply for all fields. The view of each field is that of a beam travelling in the  $+z$  direction. The field lines start and finish on surfaces of constant scalar magnetic potential, which define the pole faces in an iron-dominated electromagnetic as described by equation (1.73).

Conventionally, a pure multipole with  $\varphi_n = 0$  or  $\varphi_n = \pi$  is known as a *normal multipole*, while a pure multipole with  $\varphi_n = \pm\pi/2$  is known as a *skew multipole* (see Fig. 1.3).

The units of  $C_n$  depend on the order of the multipole. In SI units, for a dipole, the units of  $C_1$  are tesla (T); for a quadrupole, the units of  $C_2$  are

$\text{T m}^{-1}$ ; for a sextupole, the units of  $C_3$  are  $\text{T m}^{-2}$ , and so on. It is sometimes preferred to specify multipole components in dimensionless units. In that case, we introduce a reference field  $B_{\text{ref}}$  and a reference radius  $R_{\text{ref}}$ . The multipole expansion is then written, in a standard notation:

$$B_y + iB_x = B_{\text{ref}} \sum_{n=1}^{\infty} (b_n + ia_n) \left( \frac{x+iy}{R_{\text{ref}}} \right)^{n-1}. \quad (1.35)$$

In polar co-ordinates:

$$B_y + iB_x = B_{\text{ref}} \sum_{n=1}^{\infty} (b_n + ia_n) \left( \frac{re^{i\theta}}{R_{\text{ref}}} \right)^{n-1}. \quad (1.36)$$

The reference field and reference radius can be chosen arbitrarily, but must be specified if the coefficients  $a_n$  and  $b_n$  are to be interpreted properly.

We shall use the Hamiltonian formulation of classical mechanics extensively for our study of beam dynamics. In Hamiltonian mechanics, the equations of motion for a charged particle in an electromagnetic field are written in terms of the scalar potential  $\phi$  and vector potential  $\mathbf{A}$ , rather than in terms of the fields. In general, the potentials are related to the fields by:

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (1.37)$$

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \quad (1.38)$$

In the particular case of a multipole field, the magnetic field (1.35) can be derived from the vector potential  $\mathbf{A} = (0, 0, A_z)$ , where:

$$A_z = -B_{\text{ref}} \operatorname{Re} \sum_{n=1}^{\infty} (b_n + ia_n) \frac{(x+iy)^n}{n R_{\text{ref}}^{n-1}}. \quad (1.39)$$

In polar co-ordinates:

$$A_z = -B_{\text{ref}} \operatorname{Re} \sum_{n=1}^{\infty} (b_n + ia_n) \frac{r^n e^{in\theta}}{n R_{\text{ref}}^{n-1}}. \quad (1.40)$$

The scalar potential is zero.

For any given magnetic field, the vector potential is not uniquely defined. In particular, since the curl of the gradient of any function vanishes, it is always possible to make a *gauge transformation*, starting from some vector potential  $\mathbf{A}$ , to a new vector potential  $\bar{\mathbf{A}}$  given by:

$$\bar{\mathbf{A}} = \mathbf{A} + \nabla f, \quad (1.41)$$

where  $f$  is any function of position and, in general, time. The potentials  $\mathbf{A}$  and  $\bar{\mathbf{A}}$  lead to the same magnetic field. If a simultaneous transformation of the scalar potential is made:

$$\bar{\phi} = \phi - \frac{\partial f}{\partial t}, \quad (1.42)$$

then the electric field derived from  $\bar{\phi}$  and  $\bar{\mathbf{A}}$  is the same as the field derived from  $\phi$  and  $\mathbf{A}$ .

The potential (1.39) represents a particular choice of gauge for multipole fields, in which two of the components of the vector potential vanish ( $A_x = A_y = 0$ ). As we shall see in Chapter 3, this leads to significant simplifications for the equations describing the particle motion in multipole fields. However, the ability to choose a gauge in which two components of the vector potential vanish is a particular feature of two-dimensional static magnetic fields; in more complicated cases, it is usually not possible to find a gauge in which more than one component of the vector potential vanishes.

For a pure multipole field of order  $n$ , the coefficients  $a_n$  and  $b_n$  are related to the derivatives of the field components with respect to the  $x$  and  $y$  co-ordinates. Thus, for a normal multipole:

$$\frac{\partial^{n-1} B_x}{\partial y^{n-1}} = \frac{\partial^{n-1} B_y}{\partial x^{n-1}} = (n-1)! \frac{B_{\text{ref}}}{R_{\text{ref}}^{n-1}} b_n, \quad (1.43)$$

and for a skew multipole:

$$\frac{\partial^{n-1} B_x}{\partial x^{n-1}} = -\frac{\partial^{n-1} B_y}{\partial y^{n-1}} = (n-1)! \frac{B_{\text{ref}}}{R_{\text{ref}}^{n-1}} a_n. \quad (1.44)$$

A normal dipole has a uniform vertical field; a normal quadrupole has a vertical field for  $y = 0$  that increases linearly with  $x$ ; a normal sextupole has a vertical field for  $y = 0$  that increases as the square of  $x$ ; and so on.

Multipole fields are of considerable importance in accelerators. Beams of charged particles are steered by dipole fields and focused by quadrupole fields. Higher-order multipoles have applications in controlling effects related to the spread of particle energies and in suppressing some types of beam instability. When discussing the beam dynamics, it is conventional to work with multipole fields multiplied by  $q/P_0$ , where  $q$  is the charge on a single particle in the beam, and  $P_0$  is the *reference momentum* (the particle momentum for which the beam line is designed). Thus, we define for normal multipole fields:

$$k_n = \frac{q}{P_0} \frac{\partial^n B_y}{\partial x^n} = n! \frac{B_{\text{ref}}}{R_{\text{ref}}^n} b_{n+1}, \quad (1.45)$$

and for skew multipole fields:

$$k_n^{(s)} = -\frac{q}{P_0} \frac{\partial^n B_x}{\partial x^n} = -n! \frac{B_{\text{ref}}}{R_{\text{ref}}^n} a_{n+1}. \quad (1.46)$$

As we shall see in later chapters, it is the ‘normalised’ multipole strengths specified by values  $k_n$  and  $k_n^{(s)}$ , rather than the coefficients  $b_n$  and  $a_n$ , that are significant for charged particle optics in an accelerator. For example, when a charged particle beam passes through a quadrupole, it is focused in the same way that a beam of light is focused when it passes through a lens; the focal length of a quadrupole of normalised strength  $k_1$  and length  $L$  is (approximately)  $1/(k_1 L)$ .

It is important to know how to generate multipole fields in practice. It turns out that the magnetic fields from long, straight, parallel wires can be combined to generate pure multipole fields of any given order. It is also possible to generate pure multipole fields using materials with high magnetic permeability, shaped with the appropriate geometry. We shall consider both methods, first focusing on ‘idealised’ geometries in Sections 1.2.1 and 1.2.2 without worrying about practical constraints. However, field imperfections (in the form of unwanted multipole components in a given magnet) can have significant effects on the beam dynamics in an accelerator. Therefore, we discuss in Section 1.2.3 some of the characteristics of field imperfections, and outline techniques for determining the multipole content of a given magnet.

### 1.2.1 Current distribution for a pure multipole

Our goal is to determine a current distribution that will generate a pure multipole field of specified order. As a first step, we derive the multipole components in the field around a long, straight wire carrying a uniform current. The magnetic field in this case has field lines that form circular loops centred around the wire. If we take Maxwell’s equation (1.3) and integrate over a circular disc  $S$  perpendicular to, and centred around the wire, we obtain:

$$\int_S \nabla \times \mathbf{H} \cdot d\mathbf{A} = \oint_C \mathbf{H} \cdot d\mathbf{l} = I, \quad (1.47)$$

where the curve  $C$  bounds the disc  $S$  (with area element  $d\mathbf{A}$ ),  $I$  is the current in the wire, and we have applied Stokes’ theorem to convert the surface integral into a line integral. Since the magnetic field is uniform around the curve  $C$ , we obtain (using  $\mathbf{B} = \mu_0 \mathbf{H}$ ) Ampère’s law:

$$B = \frac{\mu_0 I}{2\pi\ell}, \quad (1.48)$$

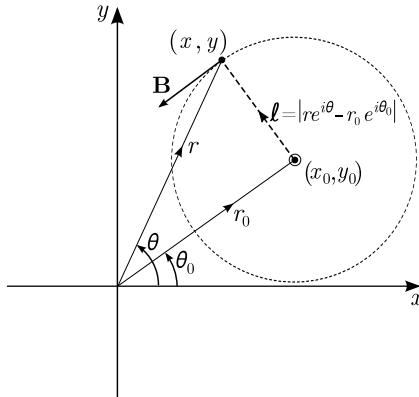


Fig. 1.4 Geometry for calculation of the multipole components in the magnetic field  $\mathbf{B}$  around a long, straight wire carrying a uniform current. The wire passes through the point  $(x_0, y_0)$  (with  $x_0$  and  $y_0$  given by the real and imaginary parts of  $r_0 e^{i\theta_0}$ , respectively) and carries a current in the  $+z$  direction. The distance in the  $x$ - $y$  plane between the wire and an observation point  $(x, y)$  (with  $x$  and  $y$  given by the real and imaginary parts of  $re^{i\theta}$ ) is  $\ell = |re^{i\theta} - r_0 e^{i\theta_0}|$ .

where  $\ell$  is the distance from the wire in a plane perpendicular to the wire. To determine the multipole components in the field, we first derive an expression for the field components at an arbitrary point  $(x, y)$  from a wire carrying current  $I$ , passing through a point  $(x_0, y_0)$  and parallel to the  $z$  axis.

Since we are working in two dimensions, we can represent the components of a vector by the real and imaginary parts of a complex number. Thus, the vector from  $(x_0, y_0)$  to a point  $(x, y)$  is given by  $re^{i\theta} - r_0 e^{i\theta_0}$ , and the magnitude of the field at  $(x, y)$  is:

$$B = \frac{\mu_0 I}{2\pi} \frac{1}{|re^{i\theta} - r_0 e^{i\theta_0}|}. \quad (1.49)$$

The geometry is shown in Fig. 1.4. The direction of the field is perpendicular to the line from  $(x_0, y_0)$  to  $(x, y)$ . Since a rotation through  $90^\circ$  can be represented by a multiplication by  $i$ , we can write:

$$B_x + iB_y = \frac{\mu_0 I}{2\pi} \frac{i(re^{i\theta} - r_0 e^{i\theta_0})}{|re^{i\theta} - r_0 e^{i\theta_0}|^2}, \quad (1.50)$$

and hence:

$$B_y + iB_x = \frac{\mu_0 I}{2\pi} \frac{(re^{-i\theta} - r_0 e^{-i\theta_0})}{|re^{i\theta} - r_0 e^{i\theta_0}|^2} = -\frac{\mu_0 I}{2\pi r_0} \frac{e^{-i\theta_0}}{1 - \frac{r}{r_0} e^{i(\theta-\theta_0)}}. \quad (1.51)$$

Then, using the Taylor series expansion for  $(1 - \zeta)^{-1}$ , where  $\zeta$  is a complex number with  $|\zeta| < 1$ :

$$\frac{1}{1 - \zeta} = \sum_{n=0}^{\infty} \zeta^n, \quad (1.52)$$

we can write:

$$B_y + iB_x = -\frac{\mu_0 I}{2\pi r_0} e^{-i\theta_0} \sum_{n=1}^{\infty} \left(\frac{r}{r_0}\right)^{n-1} e^{i(n-1)(\theta-\theta_0)}. \quad (1.53)$$

Equation (1.53) is valid for  $r < r_0$ . Comparing with the standard multipole expansion (1.36), we see that if we choose the reference field  $B_{\text{ref}}$  and the reference radius  $R_{\text{ref}}$ :

$$B_{\text{ref}} = \frac{\mu_0 I}{2\pi r_0}, \quad (1.54)$$

$$R_{\text{ref}} = r_0, \quad (1.55)$$

then the coefficients for the multipole components in the field are given by:

$$b_n + ia_n = -e^{-in\theta_0}. \quad (1.56)$$

The field around a long, straight wire can be represented as an infinite sum over all multipoles.

Now we consider a current flowing on the surface of a cylinder of radius  $r_0$ . Suppose that the current flowing in a section of the cylinder at angle  $\theta_0$  and subtending angle  $d\theta_0$  at the origin is  $I(\theta_0) d\theta_0$ . By the principle of superposition, we can obtain the total field by summing the contributions from the currents at all values of  $\theta_0$ :

$$B_y + iB_x = -\frac{\mu_0}{2\pi r_0} \sum_{n=1}^{\infty} \left(\frac{r}{r_0}\right)^{n-1} e^{i(n-1)\theta} \int_0^{2\pi} e^{-in\theta_0} I(\theta_0) d\theta_0. \quad (1.57)$$

We see that the multipole components are related to the Fourier components in the current distribution over the cylinder of radius  $r_0$ . In particular, if we consider a current distribution with just a single Fourier component:

$$I(\theta_0) = I_0 \cos(n_0 \theta_0 - \varphi), \quad (1.58)$$

the integral in the right-hand side of (1.57) vanishes except for  $n = n_0$ , and we find:

$$B_y + iB_x = -\frac{\mu_0 I_0}{2\pi r_0} \left(\frac{r}{r_0}\right)^{n_0-1} e^{i(n_0-1)\theta} \pi e^{-i\varphi}. \quad (1.59)$$

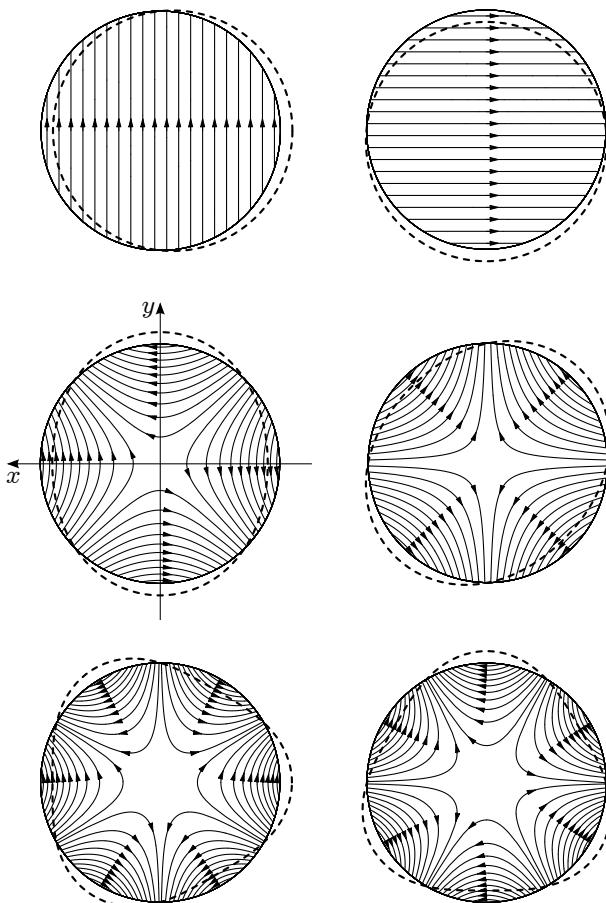


Fig. 1.5 Current distributions for generating pure multipole fields. Top: dipole. Middle: quadrupole. Bottom: sextupole. Fields on the left are normal ( $a_n = 0$ ); those on the right are skew ( $b_n = 0$ ). Co-ordinate axes are shown only for the normal quadrupole field, but similar axes apply for all fields. The view of each field is that of a beam travelling in the  $+z$  direction. The deviation of the dashed line from the circular boundary shows the local current density. Current is flowing in the  $+z$  direction for increased radius, and in the  $-z$  direction for reduced radius.

The current distribution (1.58) generates a pure multipole field of order  $n_0$ . If we choose, as before, the reference field (1.54) and the reference radius (1.55), then the multipole coefficients are:

$$b_n + ia_n = -\pi e^{-i\varphi}. \quad (1.60)$$

The parameter  $\varphi$  gives the ‘angle’ of the current distribution. For  $\varphi = 0$

or  $\varphi = \pi$ , the current generates a normal multipole; for  $\varphi = \pm\pi/2$ , the current generates a skew multipole (see Fig. 1.5).

The fact that a sinusoidal current distribution on a cylinder can generate a pure multipole field is not simply of academic interest. By winding wires in an appropriate pattern on a cylinder, it is possible to approximate a sinusoidal current distribution closely enough to produce a multipole field of acceptable quality for many applications. Usually, several layers of windings are used with a different pattern of wires in each layer, to improve the approximation to a sinusoidal current distribution. Superconducting wires can be used to achieve strong fields: for more information, see [Mess *et al.* (1996)].

### **1.2.2 Geometry of iron-dominated multipole magnets**

Normal-conducting magnets usually use iron cores to increase the flux density achieved by a given current. In such a magnet, the shape of the magnetic field depends mainly on the geometry of the iron. In this section, we shall determine the geometry required to generate a pure multipole of given order. To simplify the problem, we shall make some approximations: in particular, we shall assume that the iron core has uniform cross-section and infinitely extend along  $z$ ; that there are no limits to the iron in the  $x$ - $y$  plane; and that the iron has infinite permeability. The field in a more realistic magnet will, in general, need to be computed numerically; however, the characteristics derived from our idealised model are often a good starting point for the design of an iron-dominated multipole magnet.

We base our analysis on the magnetic scalar potential,  $\phi_{\text{mag}}$ , which is related to the magnetic field  $\mathbf{B}$  by:

$$\mathbf{B} = -\nabla\phi_{\text{mag}}. \quad (1.61)$$

The curl of the field in this case is zero for any function  $\phi_{\text{mag}}$ : this is a consequence of the mathematical properties of the grad and curl operators. Therefore, it follows from Maxwell's equation (1.3) that a magnetic field can only be derived from a scalar potential if there is both zero current density, and zero time-dependent electric displacement at the location where the field is to be calculated. We have already shown that for multipole fields both the divergence and the curl of the field vanish. Since the curl of the grad of any function is identically zero, (1.22) is automatically satisfied for any field  $\mathbf{B}$  derived using (1.61). From (1.21), we find:

$$\nabla^2\phi_{\text{mag}} = 0, \quad (1.62)$$

where  $\nabla^2$  is the Laplacian operator. Equation (1.62) is Laplace's equation: the scalar potential in any specific case is found by solving this equation with given boundary conditions.

The geometry of iron required to generate a pure multipole field can be determined from the scalar potential for a pure multipole field, as follows. Since the magnetic flux density  $\mathbf{B}$  is obtained from the gradient of the scalar potential, the flux density at any point must be perpendicular to a surface of constant scalar potential. However, we know from Section 1.1.1 that the magnetic flux density at the surface of a material with infinite permeability must be perpendicular to that surface. Hence, to generate a pure multipole field in a magnet containing material of infinite permeability, we just need to shape the material so that its surface follows a surface of constant magnetic scalar potential for the required field.

Therefore, we look for a potential  $\phi_{\text{mag}}$  that gives a pure multipole field of order  $n$ , i.e. a potential that satisfies, from (1.61):

$$-\left(\frac{\partial}{\partial y} + i\frac{\partial}{\partial x}\right)\phi_{\text{mag}} = B_y + iB_x = C_n(x + iy)^{n-1}. \quad (1.63)$$

An appropriate solution to (1.63) is:

$$\phi_{\text{mag}} = -|C_n| \frac{r^n}{n} \sin(n\theta - \varphi_n), \quad (1.64)$$

where:

$$x + iy = re^{i\theta}. \quad (1.65)$$

That (1.64) is indeed the potential for a pure multipole of order  $n$  can be shown as follows. In polar co-ordinates, the gradient can be written:

$$\nabla\phi_{\text{mag}} = \hat{r}\frac{\partial\phi_{\text{mag}}}{\partial r} + \frac{\hat{\theta}}{r}\frac{\partial\phi_{\text{mag}}}{\partial\theta}, \quad (1.66)$$

where  $\hat{r}$  and  $\hat{\theta}$  are unit vectors in the directions of increasing  $r$  and  $\theta$ , respectively. Using:

$$\hat{r} = \hat{x}\cos(\theta) + \hat{y}\sin(\theta), \quad (1.67)$$

$$\hat{\theta} = -\hat{x}\sin(\theta) + \hat{y}\cos(\theta), \quad (1.68)$$

it follows from (1.66) that:

$$\begin{aligned} -\nabla\phi_{\text{mag}} &= \hat{x}\sin((n-1)\theta - \varphi_n)|C_n|r^{n-1} \\ &\quad + \hat{y}\cos((n-1)\theta - \varphi_n)|C_n|r^{n-1}. \end{aligned} \quad (1.69)$$

Thus the field derived from the potential (1.64) can be written:

$$B_y + iB_x = |C_n|e^{-i\varphi_n}r^{n-1}e^{i(n-1)\theta}. \quad (1.70)$$

Therefore, if:

$$C_n = |C_n| e^{-i\varphi_n}, \quad (1.71)$$

then:

$$B_y + iB_x = C_n r^{n-1} e^{i(n-1)\theta} = C_n (x + iy)^{n-1}, \quad (1.72)$$

and we see that the potential (1.64) does indeed generate a pure multipole field of order  $n$ .

From the above argument, we can immediately conclude that, to generate a pure multipole field, we can shape a high permeability material such that the surface of the material follows the curve given (in parametric form, with parameter  $\theta$ ) by:

$$r^n \sin(n\theta - \varphi_n) = r_0^n, \quad (1.73)$$

where  $r_0$  is a constant giving the minimum distance between the surface of the material and the origin of the co-ordinate system. The cross-sections of iron-dominated multipole magnets of orders 1, 2 and 3, and the corresponding fields are shown in Fig. 1.3. Note that  $r \rightarrow \infty$  for  $n\theta - \varphi_n \rightarrow \text{integer} \times \pi$ . Treating each region between infinite values of  $r$  as a separate pole, we see that a pure multipole of order  $n$  has  $2n$  poles. We also see that the potential changes sign when moving from one pole to either adjacent pole: that is, poles alternate between ‘north’ and ‘south’. The field must be generated by currents flowing along wires between the poles, parallel to the  $z$  axis: to avoid direct contribution from the field around the wires, these wires should be located a large (in fact, infinite) distance from the origin.

Note that it is possible to determine the shape of the pole face for a magnet containing any specified set of multipoles by summing the potentials for the different multipole components, and then solving for  $r$  as a function of  $\theta$ , for a fixed value of the scalar potential. Accelerator magnets designed to have more than one multipole component are often known as *combined function magnets*. Possibly the most common type of combined function

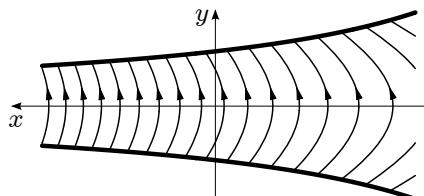


Fig. 1.6 Pole shapes for dipole magnet with quadrupole component.

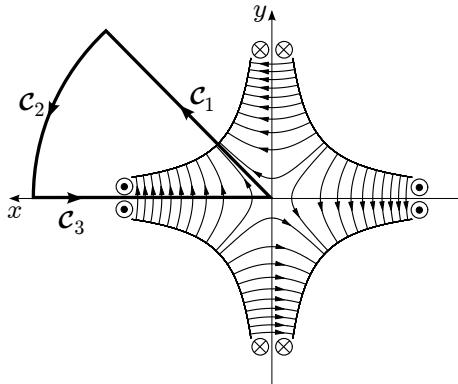


Fig. 1.7 Contour  $\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3$  for line integral used to calculate the field strength in an iron-dominated quadrupole. Current is carried parallel to the  $z$  axis in coils indicated by a dot within a circle (current in the  $-z$  direction) or by a cross within a circle (current in the  $+z$  direction).

magnet is a dipole with a quadrupole component: such magnets can be used to steer and focus a beam simultaneously. The shape of the pole faces and the field lines in a dipole with (strong) quadrupole component are shown in Fig. 1.6.

In practice, some variation from the ideal geometry is needed to account for the fact that any material used in the magnet will have finite permeability, and finitely extend transversely and longitudinally. The wires carrying the current that generates the magnetic flux are arranged in coils around each pole: the geometry of the coils is also likely to lead to some higher-order multipole components in the field.

To complete our discussion of methods for generating multipole fields, we will derive an expression for the field strength in an iron-dominated magnet with a given number of ampere-turns in the coil around each pole. To do this, we consider a line integral as shown in Fig. 1.7. In the figure, we show a quadrupole; however, the generalisation to other orders of multipole is straightforward. In principle, the coils carrying the electric current, and the line segment  $\mathcal{C}_2$ , should be an infinite distance from the origin (the centre of the magnet).

Using Maxwell's equation (1.3), with constant (zero) electric displacement, and integrating over the surface  $\mathcal{S}$  bounded by the curve  $\mathcal{C}_1 + \mathcal{C}_2 + \mathcal{C}_3$  gives:

$$\int_{\mathcal{S}} \nabla \times \mathbf{H} \cdot d\mathbf{A} = \int_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{A} = NI. \quad (1.74)$$

The surface  $\mathcal{S}$  is oriented so that the normal to  $\mathcal{S}$  (in the direction of the area element  $d\mathbf{A}$ ) is parallel to the negative  $z$  axis; and the coil around each pole consists of  $N$  turns of wire carrying current  $I$ . Applying Stokes' theorem gives:

$$\int_{C_1} \mathbf{H} \cdot d\mathbf{l} + \int_{C_2} \mathbf{H} \cdot d\mathbf{l} + \int_{C_3} \mathbf{H} \cdot d\mathbf{l} = NI. \quad (1.75)$$

From Section 1.1.1, we know that the normal component of the magnetic flux density  $\mathbf{B}$  is continuous across any boundary; and the parallel component of the magnetic field vanishes at the surface of a material with infinite permeability. Therefore, since  $\mathbf{B} = \mu\mathbf{H}$ , it follows that for a finite field between the poles, and for permeability  $\mu \rightarrow \infty$ , the magnetic intensity  $\mathbf{H}$  vanishes within the poles, and the integral along  $C_2$  in (1.75) is zero. Also, the field is perpendicular to the line segment  $C_3$ . Thus, the only integral that makes a non-zero contribution to the left-hand side of (1.75) is the integral along  $C_1$  from the origin to the face of the pole. Hence:

$$\int_0^{r_0} \frac{B_r(r)}{\mu_0} dr = NI. \quad (1.76)$$

The contour  $C_1$  is chosen such that along this contour the field only has a radial component, parallel to the contour. From (1.32), we see that, for a multipole of order  $n$ , along this contour we have:

$$B_r = |C_n| r^{n-1}. \quad (1.77)$$

Therefore, by performing the integral in (1.76) we find:

$$|C_n| = \mu_0 NI \frac{n}{r_0^n}. \quad (1.78)$$

For a normal multipole, the field is given by:

$$B_y + iB_x = \frac{\mu_0 n NI}{r_0} \left( \frac{x+iy}{r_0} \right)^{n-1}. \quad (1.79)$$

For example, in a normal quadrupole ( $n = 2$ ), the field gradient is given by:

$$\frac{\partial B_y}{\partial x} = \frac{2\mu_0 NI}{r_0^2}. \quad (1.80)$$

### 1.2.3 Multipole decomposition

In Sections 1.2.1 and 1.2.2, we determined the current density distributions and material geometries needed to generate a pure multipole field of a given order. However, the distributions and geometries required are not perfectly achievable in practice: the currents and materials have infinite longitudinal extent; and we require either a current that exists purely on the surface of a cylinder, or infinite permeability materials with infinite transverse extent.

Real multipole magnets, therefore, will not consist of a single multipole component, but (in general) a superposition of an infinite number of multipole fields. The exact shape of the field can have a significant impact on the beam dynamics in an accelerator. In many simulation codes for accelerator beam dynamics, the magnets are specified by the multipole coefficients: this is because simple techniques exist for approximating the effect of, for example, sextupole, octopole and other higher-order components in the field of a dipole or quadrupole magnet. The question then arises: how does one determine the multipole components in a given magnetic field?

At this point, we should make a distinction between the design field of a magnet, and the field that exists within a fabricated magnet. The design field is one that is still in some sense ‘ideal’; though the design field for a quadrupole magnet (for example) will contain other multipole components, because the design has to respect practical constraints, i.e. the magnet will have finite longitudinal and transverse extent, any currents will flow in wires of non-zero dimension, and any materials present will have finite (and often nonlinear) permeability. Usually, one attempts to optimise the design to minimise the strengths of the multipole components except for the one required: the residual strengths are generally known as *systematic* multipole errors. These errors will be present in any fabricated magnet, although, because of construction tolerances, the errors will vary between any two magnets of the same design. The differences between the multipole components in the design field and the components in a particular magnet are known as *random* multipole errors.

The effects of systematic and random multipole errors on the performance of an accelerator, and hence the specification of upper limits on these quantities, can usually only be properly understood by running beam dynamics simulations. Therefore, accelerator magnet (and lattice) design often progresses iteratively. Some initial estimate of the limits on the errors is often needed to guide the magnet design; but then any design that is developed must be studied by further beam dynamics simulations to determine whether improvements are needed.

It is therefore important to be able to determine the multipole components in a magnetic field from numerical field data: these data may come from either a computer model (based on the design of a particular magnet), or from measurements made on a real device. There are different procedures that can be used to achieve the ‘decomposition’ of a field into its multipole components. In the remainder of this section, we shall consider methods based on Cartesian and polar representations of two-dimensional fields (i.e. fields that are independent of the longitudinal co-ordinate). In Section 1.3 we shall consider decompositions of three-dimensional fields, that is, fields that have explicit dependence on longitudinal as well as transverse co-ordinates. However, we first consider an important concept in the discussion of multipole field errors, namely how the symmetry of a multipole magnet leads to ‘allowed’ and ‘forbidden’ higher-order multipoles.

A pure multipole field of order  $n$  can be written:

$$B_y + iB_x = |C_n| e^{i\varphi_n} r^{n-1} e^{i(n-1)\theta}. \quad (1.81)$$

The parameter  $\varphi_n$  characterises the angular orientation of the magnet around the  $z$  axis. In particular, from (1.73), we see that a change in  $\varphi_n$  by  $n\alpha$  is equivalent to a rotation of the co-ordinates (a change in  $\theta$ ) by  $\alpha$ . Thus, a rotation of a magnet around the  $z$  axis by angle  $\alpha$  may be represented by a change in  $\varphi_n$  by  $-n\alpha$ . In particular, if the magnet is rotated by  $\pm\pi/n$ , then from (1.81), we see that the field at any point simply changes sign:

$$\text{if } \varphi_n \mapsto \varphi_n \pm \pi, \text{ then } \mathbf{B} \mapsto -\mathbf{B}. \quad (1.82)$$

This property of the magnetic field is imposed by the symmetry of the magnet. In a real magnet, it will not be satisfied exactly, because random variations in the geometry will break the symmetry. However, it is possible to maintain the symmetry exactly in the design of the magnet; this means that although higher-order multipoles will in general be present, only those multipoles satisfying the symmetry constraint (1.82) can be present. These are the *allowed multipoles*. Other multipoles, which must be completely absent, are the *forbidden multipoles*.

We can derive a simple expression for the allowed multipoles in a magnet designed with symmetry for a multipole of order  $n$ . Consider an additional multipole (a systematic error) in this field, of order  $m$ . By the principle of superposition, the total field can be written as

$$B_y + iB_x = |C_n| e^{i\varphi_n} r^{n-1} e^{i(n-1)\theta} + |C_m| e^{i\varphi_m} r^{m-1} e^{i(m-1)\theta}. \quad (1.83)$$

By design, the geometry is such that under a rotation about the  $z$  axis through  $\pi/n$ , the magnet looks the same, except that all currents have reversed direction: therefore the field simply changes sign. Under this rotation  $\varphi_n \mapsto \varphi_n - \pi$ ; however,  $\varphi_m \mapsto \varphi_m - m\pi/n$ . This means that we must have:

$$e^{-im\pi/n} = -1. \quad (1.84)$$

Therefore  $m/n$  must be an odd integer. Assuming that  $m \neq n$  (i.e. the multipole error is of a different order than the ‘main’ multipole field), then

$$\frac{m}{n} = 3, 5, 7, \dots \quad (1.85)$$

Thus, for a dipole, the allowed higher-order multipoles are sextupole, decapole, etc.; for a quadrupole, the allowed higher-order multipoles are 12-pole, 20-pole, etc. The fact that the allowed higher-order multipoles have an order given by an odd integer multiplied by the order of the main multipole is a consequence of the fact that magnetic poles always occur in north–south pairs: to maintain the correct rotational symmetry (with the field changing sign under a rotation through  $\pi/2$ ) the first higher-order multipole must be constructed by ‘splitting’ each main pole into three, then into five, and so on. The field in a real magnet will contain all higher-order multipoles, not just the ones allowed by symmetry. However, it is often the case that the strengths of the allowed multipoles dominate over those of the forbidden multipoles.

Suppose we have obtained a set of numerical field data, either from a magnetic model or from measurements on a real magnet. To determine the effect of the field on the beam dynamics in an accelerator, it is helpful to know the multipole components in the field. One way to compute the multipole components is to fit a polynomial to the field data. For example, if we consider a normal multipole (coefficients  $C_n$  are all real), the vertical field along the  $x$  axis (i.e. for  $y = 0$ ) is given by:

$$B_y = \sum_{n=1}^{\infty} C_n x^{n-1}. \quad (1.86)$$

The number of data points determines the highest-order multipole that can be fitted. Fitting may be achieved using, for example, a routine that minimises the squares of the residuals between the data and the fitted function. However, although this procedure can, in principle, produce useful results, it is not very robust. In particular, the presence of multipoles of order higher than those included in the fit can affect the values determined for

the multipoles that are included in the fit. This can lead to significant errors in simulations of the beam dynamics: for example, the effect of a linear focusing error in a beam line (from some variation in the quadrupole strength) can be very different from the effects of higher-order multipole errors. If the multipole coefficients are determined using a polynomial fit, this can lead to inaccurate predictions of the behaviour of the beam in the beam line.

The reason that a polynomial fit sometimes does not work very well is that the basis functions for the fit are simply monomials, i.e. powers of  $x$ , and these functions are not orthogonal. This means that it is possible to construct a set of data using one monomial, and obtain a fit to the data using a different monomial. The coefficients obtained from a fit depend on which monomials are included in the fit. A more robust technique can be developed by using basis functions that are orthogonal, i.e. the coefficients obtained from a fit to a given set of data will be the same, no matter which functions from the set are chosen. In that case, the quadrupole strength (for example) that we find in a particular magnet will not depend on which higher-order terms we include in the fit.

An appropriate set of basis functions for fitting a set of multipoles to a given field is provided by the function  $r^{n-1}e^{in\theta}$ . From (1.32) we know that the field in a multipole magnet can be written in polar co-ordinates as:

$$B_\theta + iB_r = \sum_{n=1}^{\infty} C_n r^{n-1} e^{in\theta}. \quad (1.87)$$

We see that if we make a set of measurements of  $B_r$  and  $B_\theta$  at different values of  $\theta$  and fixed radial distance  $r$ , then we can obtain the coefficients  $C_n$  by a discrete Fourier transform, as follows. Suppose we make  $M$  measurements of the field, at  $\theta = \theta_m$ , where:

$$\theta_m = 2\pi \frac{m}{M}, \quad m = 0, 1, 2 \dots M-1. \quad (1.88)$$

We write the measurement at  $\theta = \theta_m$  as  $B_m$ ; note that  $B_m$  is a complex number, whose real and imaginary parts are given by the azimuthal and radial components of the field at  $\theta = \theta_m$ . Now we construct, for a chosen integer  $n'$ :

$$\sum_{m=0}^{M-1} B_m e^{-2\pi i n' m / M} = \sum_{m=0}^{M-1} \sum_{n=1}^{\infty} C_n r_0^{n-1} e^{2\pi i (n-n') m / M}, \quad (1.89)$$

where  $r_0$  is the radial distance at which the field measurements are made. The summation over  $m$  on the right-hand side vanishes, unless  $n = n'$ .

Thus we can write:

$$\sum_{m=0}^{M-1} B_m e^{-2\pi i n' m / M} = MC_{n'} r_0^{n'-1}. \quad (1.90)$$

If we relabel  $n'$  as  $n$ , then we see that the multipole coefficients  $C_n$  are given by:

$$C_n = \frac{1}{Mr_0^{n-1}} \sum_{m=0}^{M-1} B_m e^{-2\pi i n m / M}. \quad (1.91)$$

As already mentioned, the advantage of this technique over a polynomial fit is that the basis functions used to construct the fit are orthogonal. As a consequence, the value found for any given multipole component using (1.91) is independent of the presence of any other multipoles, of higher or lower order.

A further advantage of using the polar basis instead of the Cartesian basis comes from the dependence of the field on the radial distance. Consider a multipole field with a single component  $C_n$ . Suppose that field data is measured (or obtained from a model) with accuracy  $\Delta B_m$ . Then the accuracy in the multipole coefficients will vary with the reference radius  $r_0$  (at which the measurements are made) as:

$$\Delta C_n \propto \frac{\Delta B_m}{r_0^{n-1}}. \quad (1.92)$$

Better accuracy in the multipole coefficients is obtained by choosing the radius  $r_0$  to be as large as possible. Furthermore, the accuracy in the fitted field will be:

$$\Delta B \propto \Delta C_n r^{n-1}. \quad (1.93)$$

We obtain *improved* accuracy in the field for  $r < r_0$ , although the accuracy deteriorates quickly (particularly for higher-order multipoles) for  $r > r_0$ . It is important to choose the radial distance  $r_0$  large enough to enclose all particles likely to pass through the magnet, otherwise it will be difficult to assess the accuracy of the results from particle tracking simulations.

It is of course possible to represent a magnetic field using a set of numerical field data, giving the three field components on points forming a mesh covering the region of interest. In some ways, this is a very convenient representation, since it is the one usually provided directly by a magnetic modelling code: while further processing is usually required to arrive at other representations. However, while a numerical field map in

two dimensions is often a practical representation, in three dimensions the amount of data in even a relatively simple magnet can become extremely large, especially if a high resolution is required for the mesh. A multipole representation, on the other hand, describes a magnetic field with a relatively small set of coefficients, from which the field components at any point can be reconstructed using the basis functions. In other words, a multipole representation is more ‘portable’ than a numerical field map.

Additionally, a representation based on a multipole expansion lends itself to further manipulation in ways that a numerical field map does not. For example, any noise in the data (from measurement or computational errors) can be ‘smoothed out’ by suppression of higher-order modes. Conversely, random errors can be introduced into data based on a model with perfect symmetry by introducing multipole coefficients corresponding to forbidden harmonics. There will of course be issues surrounding the suppression or enhancement of errors by adjusting the multipole coefficients; however, one benefit of this approach is that for *any* set of multipole coefficients, at least the field is a physical field, in the sense that it satisfies Maxwell’s equations. The same will not usually be true if, for example, a general smoothing algorithm is applied to a numerical field map.

Finally, one of the main motivation for performing a multipole decomposition of a field is to provide data in a format appropriate for a beam dynamics code. Accuracy and efficiency are both important when simulating an accelerator using a modelling code. Characterisation of a storage ring frequently requires tracking thousands of particles over hundreds or thousands of turns, through a beam line that can easily consist of hundreds of magnetic elements. Numerical integration of the equations of motion for a particle in a numerical field map is generally too slow to be considered practical. Instead, transfer maps are used to compute in a small number of steps the position and momentum of a particle at the exit of a magnetic component, given its position and momentum at the entrance of the magnet. For dipoles and quadrupoles, transfer maps are available (as we shall see in Chapter 3) that can be used to accurately compute the change in particle position and momentum in a single step. For higher-order multipoles, other techniques must be used (some of which are described in Chapters 9 and 10); however it is possible to use the multipole decomposition of a given field, to construct a transfer map that provides both accuracy and efficiency for computer simulations.

We should emphasise that in our discussion of multipole decomposition, we have made no clear distinction between field data obtained from a

computational model, or from measurement of a real magnet. Of course, it is much easier to obtain the data required from a computational model: it is then quite straightforward to perform the required decomposition to determine the values of the various multipole coefficients. Unfortunately, modelled data will not, in general, include manufacturing errors, which can be very important. Measurements provide more realistic data: however, many other issues then need to be addressed, including, for example, accuracy of field measurements and alignment of the measurement instruments with respect to the magnet.

### 1.3 Three-Dimensional Fields

In Section 1.2 we discussed two-dimensional magnetostatic fields, i.e. fields dependent on only two spatial co-ordinates. Although the multipole modes that can be used to describe such fields often provide a good description for accelerator multipole magnets, the finite length of a magnet will necessarily place some limitation on any two-dimensional description of the field. In particular, in the ‘fringe field’ regions at the entrance and exit faces of multipole magnets, the field strengths often vary rapidly with longitudinal position, and the fields in these regions cannot be accurately represented by two-dimensional fields. Fortunately, in many accelerators, only the fringe fields of dipole magnets have any significant effect on the particle dynamics, and for dipoles, simplified descriptions of the fringe fields (for example, as will be discussed in Section 3.3) are often sufficient for calculating their impact on particle trajectories and machine optics. However, there are cases where a full three-dimensional description of a magnetic field in an accelerator is desirable, or even necessary. For instance, insertion devices such as wigglers and undulators (see, for example, [Clarke (2004)]) can often be represented as sequences of short dipoles of alternating polarity, but if the lengths of the dipoles are small compared with the gap between the poles, the three-dimensional nature of the field starts to have effects that cannot be ignored. There are even cases where conventional multipoles can have fringe fields that affect the dynamics to a significant extent (for example, where beam properties and space constraints mean that the aperture of the magnet needs to be large in comparison to its length).

It is therefore of practical interest to consider how two-dimensional multipole representations may be generalised to three dimensions. As usual, there are many different ways to approach the problem, and the method that is chosen will depend on the problem to be solved. In the following

sections, we describe three ways of representing three-dimensional fields that, together, provide techniques that will cover many situations arising in accelerators. First, in Section 1.3.1, we consider expressions for the fields based on series expansions in either Cartesian modes (rectangular harmonics), or cylindrical modes. The Cartesian basis is perhaps more straightforward in terms of the mathematical expressions for the different modes, but does have some disadvantages. Although initially appearing a little more complicated, the cylindrical basis is often more practical.

The series expansions presented in Section 1.3.1 describe the longitudinal variation of the field components in terms of Fourier modes. Although this representation for the fields can be useful for a wide variety of different types of magnet, it is perhaps most appropriate for magnets where the field variation in the longitudinal direction is periodic. This is only the case for certain particular kinds of magnet: examples include insertion devices such as undulators or wigglers. For other kinds of magnet, a more general representation is usually preferred. One possibility is to use generalised gradients, as described in Section 1.3.2. Generalised gradients can be useful, but have the drawback that physical fields are represented as summations over an infinite number of terms: truncating the summation at some point results in expressions for the fields that do not satisfy Maxwell's equations. In some situations, this may not be a significant issue, as long as the truncated summations still provide a sufficiently good approximation to the field. However, when exact analytical expressions are available for the fields in accelerator magnets, it may be preferable to use such expressions rather than truncated summations: one case where exact expressions for three-dimensional fields can be written down is that of fringe fields in multipole magnets, which we shall discuss in Section 1.3.3.

### **1.3.1 *Cartesian and cylindrical modes***

Consider the field given by:

$$B_x = -B_0 \frac{k_x}{k_y} \sin(k_x x) \sinh(k_y y) \sin(k_z z), \quad (1.94)$$

$$B_y = B_0 \cos(k_x x) \cosh(k_y y) \sin(k_z z), \quad (1.95)$$

$$B_z = B_0 \frac{k_z}{k_y} \cos(k_x x) \sinh(k_y y) \cos(k_z z). \quad (1.96)$$

This field is often used as a model for the field in an undulator or wiggler magnet. As may easily be verified, the field satisfies:

$$\nabla \times \mathbf{B} = 0. \quad (1.97)$$

Furthermore, the equation:

$$\nabla \cdot \mathbf{B} = 0 \quad (1.98)$$

is satisfied if:

$$k_y^2 = k_x^2 + k_z^2. \quad (1.99)$$

We conclude that, as long as the constraint (1.99) is applied, the fields (1.94)–(1.96) provide solutions to Maxwell's equations in regions with constant permeability and static (or zero) electric fields. It is possible to generalise the expressions in (1.94)–(1.96) by adding constant offsets to each of the co-ordinates. Also, the hyperbolic trigonometric function can appear for the dependence on  $x$  or on  $z$ , rather than on  $y$ . The particular form of the field given in (1.94)–(1.96) is suitable for cases where the main field component is a vertical field generated by a magnet with pole faces at  $y = \pm y_{\text{pf}}$ , for some value  $y_{\text{pf}}$ : in such a magnet, the vertical component is expected to fall off with horizontal transverse position  $x$ , and to increase as the pole faces are approached from  $y = 0$ .

By superposing fields, with appropriate variations on the form given by (1.94)–(1.96), it is possible to construct three-dimensional magnetic fields that are quite general. For example, we can simply superpose fields with different mode numbers and amplitudes:

$$B_x = - \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \frac{k_x}{k_y} \sin(k_x x) \sinh(k_y y) \sin(k_z z), \quad (1.100)$$

$$B_y = \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \cos(k_x x) \cosh(k_y y) \sin(k_z z), \quad (1.101)$$

$$B_z = \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \frac{k_z}{k_y} \cos(k_x x) \sinh(k_y y) \cos(k_z z). \quad (1.102)$$

In this form, we can see how to perform a mode decomposition; that is, how we can determine the coefficients  $\tilde{B}(k_x, k_z)$  as functions of the *mode numbers*  $k_x$  and  $k_z$ . If we consider in particular the vertical field component

on the plane  $y = y_0$  (which should be between the pole faces and the mid-plane  $y = 0$ ), then we have from (1.101):

$$\frac{B_y(x, y_0, z)}{\cosh(k_y y_0)} = \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \cos(k_x x) \sin(k_z z). \quad (1.103)$$

Hence,  $\tilde{B}(k_x, k_z)$  may be obtained from an inverse Fourier transform of  $B_y(x, y_0, z)/\cosh(k_y y_0)$ . Given field data on a grid over  $x$  and  $z$ , we can then perform numerically an inverse discrete Fourier transform, to obtain a set of coefficients  $\tilde{B}(k_x, k_z)$ . Note that once we have obtained these coefficients, we can then reconstruct all field components at all points in space. This is an important consequence of the strong constraints on the fields provided by Maxwell's equations: in general, for a static field, if we know how one field component varies over a two-dimensional plane, then we can deduce how all the field components vary over all space (on and off the plane).

It is worth noting that for fields that are periodic in the longitudinal co-ordinate  $z$ , the field components may be represented by expressions involving summations over harmonics (integer multiples) of a fundamental wave number. In the more general case, where there is no periodicity (for example, in the fringe fields of multipole magnets), the summation is replaced by an integral over a continuous range of longitudinal wave numbers, represented by a variable  $k_z$ . The expressions in Eqs. (1.100)–(1.102) apply to the general case, where there is no assumption of periodicity.

As already mentioned, for beam dynamics studies it can be useful to know a potential from which a given magnetic field may be derived. In the case of the fields (1.100)–(1.102), a convenient form for the potential is:

$$A_x = 0, \quad (1.104)$$

$$A_y = \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \frac{k_z}{k_x k_y} \sin(k_x x) \sinh(k_y y) \cos(k_z z), \quad (1.105)$$

$$A_z = - \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_z \tilde{B}(k_x, k_z) \frac{1}{k_x} \sin(k_x x) \cosh(k_y y) \sin(k_z z). \quad (1.106)$$

Here, we choose a gauge in which the  $x$  component of the vector potential vanishes. It is of course possible to make a gauge transformation, so that either the  $y$  or  $z$  component vanishes instead.

Cartesian modes are often useful for describing fields in insertion devices such as undulators and wigglers, particularly those that have weak variation of the field with  $x$ . The modes reflect the geometry, in the sense that the

poles of an undulator can be used to define the upper and lower faces of a tube of rectangular cross-section, enclosing the region within which particles move through the undulator. Except at the ends of the undulator, the field is often periodic in  $z$ . For simple devices, it is often possible to achieve a good fit to a given field using a small number of modes. To maximise the region over which the fit is reliable, one needs to choose a plane with a value of  $y$  as large as possible, with  $x$  and  $z$  extending out as far as possible on this plane: the fitted coefficients should reproduce the field behaviour accurately for values of the co-ordinates smaller than the values on the plane of the fit, but the residuals of the fit will, in general, increase rapidly for larger values of the co-ordinates. For a planar undulator or wiggler, it is often possible to choose a plane close to the pole faces, and for  $x$  to extend over the entire width of the vacuum chamber.

However, for more general cases, Cartesian modes may not provide a convenient basis. For example, in a magnet with a circular aperture, the plane that provides the largest range in  $x$  will be the mid-plane,  $y = 0$ ; and as  $y$  increases, the available range in  $x$  decreases. To perform a fit using the Cartesian basis then requires some compromise between the range of reliability in the horizontal transverse and vertical directions.

Fortunately, it is possible to choose an alternative basis for magnets with circular aperture, in which the coefficients of the fitted field can be obtained from field data on the surface of a cylinder inscribed through the magnet. In that case, the radius of the cylinder can be close to the aperture limit, maximising the range of reliability of the fit. The appropriate modes in this case are most easily expressed in cylindrical polar co-ordinates.

A field with zero divergence and curl, and hence satisfying Maxwell's equations for static fields in regions with uniform permeability, is given in cylindrical polar co-ordinates by:

$$B_r = \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) I'_n(k_z r) \sin(n\theta) \cos(k_z z), \quad (1.107)$$

$$B_\theta = \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) \frac{n}{k_z r} I_n(k_z r) \cos(n\theta) \cos(k_z z), \quad (1.108)$$

$$B_z = - \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) I_n(k_z r) \sin(n\theta) \sin(k_z z). \quad (1.109)$$

This field may be generalised by including constant offsets in the azimuthal angle  $\theta$  and the longitudinal co-ordinate  $z$ . In (1.107)–(1.109),  $I_n(k_z r)$  is the modified Bessel function of the first kind, of order  $n$ . For small values

of the argument  $\xi$ , the modified Bessel function of order  $n$  has the series expansion:

$$I_n(\xi) = \frac{\xi^{|n|}}{2^{|n|} |n|!} + O(|n| + 1). \quad (1.110)$$

For larger values of the argument, the modified Bessel functions  $I_n(\xi)$  increase exponentially. This is significant: it means that if we fit a field to data on the surface of a cylinder of given radius, then residuals of the fit will decrease exponentially within the cylinder towards  $r = 0$ , and increase exponentially outside the cylinder.

The field (1.107)–(1.109) may be obtained from the vector potential:

$$A_r = - \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) \frac{r}{n} I_n(k_z r) \cos(n\theta) \sin(k_z z), \quad (1.111)$$

$$A_\theta = 0, \quad (1.112)$$

$$A_z = - \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) \frac{r}{n} I'_n(k_z r) \cos(n\theta) \cos(k_z z). \quad (1.113)$$

In the case of multipole fields (in which the field components are independent of  $z$ ) we can find a vector potential in which two of the components are zero. For three-dimensional fields, it is not usually possible to find a gauge in which more than one of the components of the vector potential vanish. For the potential (1.111)–(1.113), we have chosen a gauge in which the azimuthal component of the vector potential vanishes. It is always possible to choose such a gauge: if we have a vector potential  $\mathbf{A}$  with  $A_\theta \neq 0$ , then we can obtain a new vector potential  $\bar{\mathbf{A}}$  in which  $\bar{A}_\theta = 0$  by making the gauge transformation:

$$\bar{\mathbf{A}} = \mathbf{A} + \nabla f, \quad (1.114)$$

where  $f$  satisfies:

$$\frac{1}{r} \frac{\partial f}{\partial \theta} = -A_\theta. \quad (1.115)$$

An attractive feature of the polar basis is that it is possible to draw a direct connection between the three-dimensional modes in this basis and the multipole components in a two-dimensional field. Consider a mode amplitude  $\tilde{B}_n(k_z)$  given (for some particular value of  $n$ ) by

$$\tilde{B}_n(k_z) = 2^n (n-1)! C_n \frac{\delta_D(k_z)}{k_z^{n-1}}, \quad (1.116)$$

where  $\delta_D(k)$  is the Dirac delta function and  $C_n$  is a constant. Substituting this mode amplitude into equations (1.107)–(1.109), using the expansion (1.110), and performing the integral over  $k_z$  gives:

$$B_r = C_n r^{n-1} \sin(n\theta), \quad (1.117)$$

$$B_\theta = C_n r^{n-1} \cos(n\theta), \quad (1.118)$$

$$B_z = 0. \quad (1.119)$$

Comparing with (1.32), we see that this is a multipole field of order  $n$ . Thus a two-dimensional multipole field is a special case of a three-dimensional field (1.107)–(1.109), with a single mode coefficient, given by (1.116).

In general, the mode coefficients  $\tilde{B}_n(k_z)$  may be obtained by a Fourier transform of the field on the surface of a cylinder of given radius,  $r = r_0$ . For example, it follows from (1.107) that

$$\frac{B_r(r_0, \theta, z)}{I'_n(k_z r_0)} = \int_{-\infty}^{\infty} dk_z \sum_{n=-\infty}^{\infty} \tilde{B}_n(k_z) \sin(n\theta) \cos(k_z z). \quad (1.120)$$

Since the mode coefficients  $\tilde{B}_n(k_z)$  are related to the multipole coefficients in a two-dimensional field, we can use these coefficients to extend the idea of a multipole to a three-dimensional field. Strictly speaking, the mode coefficients  $\tilde{B}_n(k_z)$  are related to the field by a two-dimensional Fourier transform; however, we can perform a one-dimensional inverse Fourier transform (in the  $z$  variable) to obtain a set of functions that represent, in some sense, the ‘multipole components’ of a three-dimensional field as a function of  $z$ . Here, we use the term ‘multipole components’ rather loosely, since a multipole field is strictly defined only in the two-dimensional case (i.e. for a field that is independent of the longitudinal co-ordinate). A quantity that is perhaps easier to interpret is the contribution to the field in the plane  $z = z_0$  (for any constant  $z_0$ ) made by the mode coefficients  $\tilde{B}_n(k_z)$  with a given  $n$ . For  $n = 1$ , the field components  $B_r$  and  $B_\theta$  in the plane  $z = z_0$  will behave as for a dipole field; for  $n = 2$ ,  $B_r$  and  $B_\theta$  will behave as for a quadrupole field, and so on.

As an example of the application of the techniques developed in this section for describing three-dimensional fields, it is interesting to consider fringe fields in ‘pure’ multipole magnets. Accelerator magnets are often represented by a hard-edged model, in which the field rises abruptly from zero when crossing the entrance face of the magnet, and falls abruptly to zero when crossing the exit face. However, a finite change in the field over an infinitesimal distance implies that the field has infinite derivative, which will violate Maxwell’s equations. As a more realistic model, the field gradient

is sometimes represented by an Enge function [Enge (1967)], so that at the entrance or exit of a multipole of order  $n$  and multipole gradient  $C_n$ :

$$\frac{1}{(n-1)!} \left. \frac{\partial^{n-1} B_\theta(z)}{\partial r^{n-1}} \right|_{\theta=0} = \frac{C_n}{1 + e^{az}}, \quad (1.121)$$

where  $a$  (the Enge coefficient) is a constant that determines the rate at which the gradient rises from zero ( $a < 0$ ) or falls to zero ( $a > 0$ ) upon entering or exiting the magnet. Here,  $z = 0$  is the position of the face of the magnet (the mid-point of the fringe field). For a pure multipole, we need only a single term in the summation over  $n$  in the expressions for the magnetic field (1.107)–(1.109). An appropriate form for the function  $\tilde{B}_n(k_z)$  can be obtained from a Fourier transform of the Enge function (1.121). The result is that the field components in the fringe field of a multipole of order  $n$ , with Enge-type roll-off of the field, are given by:

$$\begin{aligned} \frac{B_r}{C_n} &= \frac{1}{2} r^{n-1} \sin(n\theta) - \\ &(n-1)! \operatorname{sgn}(a) \frac{\sin(n\theta)}{\pi} \left( \frac{2\pi}{a} \right)^n \int_0^\infty \frac{I'_n(k_z r) \sin(k_z z)}{\sinh^n(\frac{\pi}{a} k_z)} dk_z, \end{aligned} \quad (1.122)$$

$$\begin{aligned} \frac{B_\theta}{C_n} &= \frac{1}{2} r^{n-1} \cos(n\theta) - \\ &(n-1)! \operatorname{sgn}(a) \frac{\cos(n\theta)}{\pi} \left( \frac{2\pi}{a} \right)^n \int_0^\infty \frac{n I_n(k_z r) \sin(k_z z)}{r k_z \sinh^n(\frac{\pi}{a} k_z)} dk_z, \end{aligned} \quad (1.123)$$

$$\frac{B_z}{C_n} = -(n-1)! \operatorname{sgn}(a) \frac{\sin(n\theta)}{\pi} \left( \frac{2\pi}{a} \right)^n \int_0^\infty \frac{I_n(k_z r) \cos(k_z z)}{\sinh^n(\frac{\pi}{a} k_z)} dk_z. \quad (1.124)$$

Fringe fields in multipole magnets are discussed further in Section 1.3.3.

It is also worth noting that the expressions (1.107)–(1.109) can be applied to axially-symmetric fields (for example, in a solenoid magnet), in which case  $B_\theta = 0$ , and the radial and longitudinal field components take the form:

$$B_r = - \int_{-\infty}^{\infty} \tilde{B}(k_z) I_1(k_z r) \cos(k_z z) dk_z, \quad (1.125)$$

$$B_z = \int_{-\infty}^{\infty} \tilde{B}(k_z) I_0(k_z r) \sin(k_z z) dk_z. \quad (1.126)$$

As an example, consider the exit of a solenoid in which the longitudinal field decays along the  $z$  axis, falling from some fixed value for  $z \rightarrow -\infty$  to

zero as  $z \rightarrow \infty$ . The behaviour can be described by an Enge function, so that along the  $z$  axis (i.e. for  $r = 0$ ):

$$B_z = \frac{B_0}{1 + e^{az}}, \quad (1.127)$$

where the Enge coefficient  $a$  (with  $a > 0$  for the exit of the solenoid) determines the rate of roll-off of the field. In this case, the radial and longitudinal components of the field can be written:

$$\frac{B_r}{B_0} = \int_0^\infty \frac{I_1(k_z r) \cos(k_z z)}{|a| \sinh(\frac{\pi}{a} k_z)} dk_z, \quad (1.128)$$

$$\frac{B_z}{B_0} = \frac{1}{2} - \int_0^\infty \frac{I_0(k_z r) \sin(k_z z)}{|a| \sinh(\frac{\pi}{a} k_z)} dk_z. \quad (1.129)$$

The same formulae can be used with  $a < 0$ , to describe the variation of the field at the entrance of the solenoid. The field (in the case  $a > 0$ , i.e. around the exit of the solenoid) is illustrated in Fig. 1.8, and shows the expected behaviour. It should be noted, however, that the field around a solenoid in free space (i.e. in the absence of any magnetic materials) is not described exactly by an Enge function: for further discussion of solenoid fields, and a derivation of exact expressions for the field around a solenoid in free space, see Section 1.3.3.

It is worth making a few final remarks about mode decompositions for three-dimensional fields. First, as already mentioned, in many practical cases a full three-dimensional mode decomposition will not be necessary. While the techniques described in this section do provide a detailed description of the field in a form suitable for beam dynamics studies, three-dimensional decompositions do rely on large sets of accurate and detailed field data. While the appropriate data may be conveniently obtained from a computational model, it may be difficult or impractical to make the necessary measurements on a real magnet. In many cases, a two-dimensional field description in terms of multipoles will be sufficient. Generally, a three-dimensional analysis only needs to be undertaken where there are grounds to believe that the three-dimensional nature of the field is likely to have a significant impact on the beam dynamics.

Second, we have already emphasised that to obtain an accurate description of the field within some region in terms of a mode decomposition, the mode amplitudes should be determined by a fit on a surface enclosing the region of interest. Outside the region bounded by the surface of the fit, the fitted field can be expected to diverge exponentially from the real field.

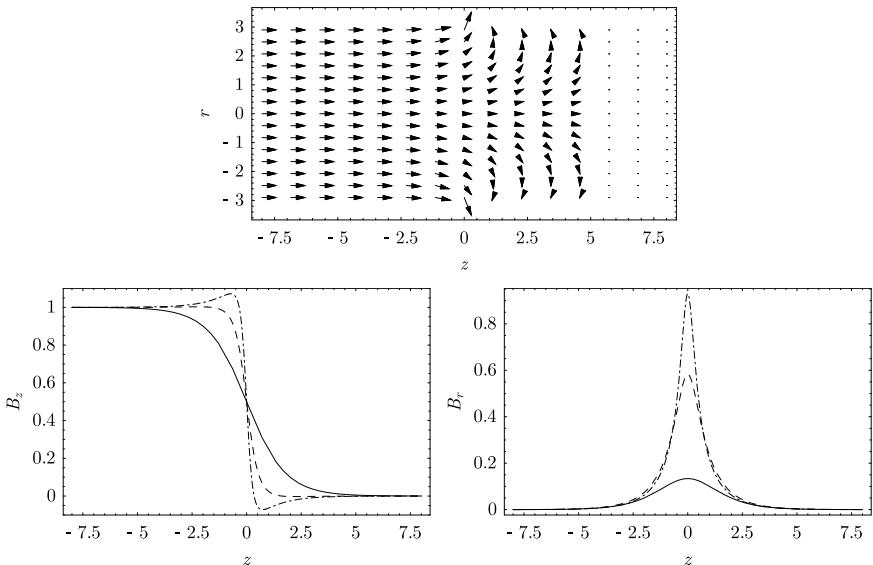


Fig. 1.8 Axially-symmetric magnetic field with Enge-type roll-off along the  $z$  axis. Top: field vectors in a plane of fixed angular co-ordinate  $\theta = \theta_0$ . Bottom: variation in the longitudinal (left) and radial (right) field components along lines parallel to the  $z$  axis, with radial co-ordinate  $r = 1.0$  (solid lines),  $r = 2.5$  (dashed lines) and  $r = 2.8$  (dash-dotted lines). The Enge coefficient of the field is  $a = 1$ , and the field amplitude is  $B_0 = 1$ .

However, in choosing the surface for the fit, the geometry of the magnet will impose some constraints. A magnet with a wide rectangular aperture may lend itself to a description using a Cartesian basis (fitting on the surface of a rectangular box); a circular aperture, however, is more likely to require use of a polar basis (fitting on the surface of a cylinder with circular cross-section). Both cases have been described above. It may be appropriate in other cases to perform a fit on the surface of a cylinder with elliptical cross-section. The basis functions in this case involve Mathieu functions [Mitchell and Dragt (2010); Dragt (2020)].

### 1.3.2 Generalised gradients

The field fitting techniques described in Section 1.3.1 produce representations of magnetic fields in terms of Fourier modes in the longitudinal direction (i.e. along the  $z$  axis). Although this is a reasonable approach in many cases, it can be useful to have alternative techniques, especially for non-periodic fields. One possible alternative is to express the field modes in

terms of a function of  $z$  that is not assumed to be periodic. This leads to the idea of *generalised gradients* [Dragt (2020); Venturini and Dragt (1999)], which we outline in this section.

It is convenient to start with a scalar potential  $\phi_{\text{mag}}$ , from which the magnetic field is derived by:

$$\mathbf{B} = -\nabla\phi_{\text{mag}}. \quad (1.130)$$

Such a representation is possible where both the current density and the rate of change of electric displacement vanish, so that from Maxwell's equation (1.3):

$$\nabla \times \mathbf{B} = 0. \quad (1.131)$$

Equation (1.131) is satisfied for any scalar potential  $\phi_{\text{mag}}$ . Maxwell's equation (1.2) is satisfied if the scalar potential satisfies Laplace's equation:

$$\nabla^2\phi_{\text{mag}} = 0. \quad (1.132)$$

The solution to (1.132) can be expressed as a sum over cylindrical harmonics:

$$\phi_{\text{mag}} = \int_{-\infty}^{\infty} dk_z \sum_{m=-\infty}^{\infty} G_m(k_z) I_m(k_z r) e^{im\theta} e^{ik_z z}, \quad (1.133)$$

where  $I_m(k_z r)$  is a modified Bessel function, and the cylindrical polar coordinates are  $r$ ,  $\theta$  and  $z$ . We define a set of functions  $\Phi_m(r, z)$ , given by:

$$\Phi_m(r, z) = \int_{-\infty}^{\infty} dk_z G_m(k_z) I_m(k_z r) e^{ik_z z}, \quad (1.134)$$

in terms of which the magnetic scalar potential can be written:

$$\phi_{\text{mag}} = \sum_{m=-\infty}^{\infty} \Phi_m(r, z) e^{im\theta}. \quad (1.135)$$

For physical fields, the imaginary part of the magnetic scalar potential must vanish; the condition for this to be the case can be written:

$$\Phi_{-m}(r, z) = \Phi_m^*(r, z), \quad (1.136)$$

where  $\Phi_m^*(r, z)$  is the complex conjugate of  $\Phi_m(r, z)$ . Using the Taylor expansion of the modified Bessel function:

$$I_m(k_z r) = \sum_{\ell=0}^{\infty} \frac{(k_z r)^{2\ell+|m|}}{2^{2\ell+|m|}\ell!(\ell+|m|)!}, \quad (1.137)$$

we can write:

$$\Phi_m(r, z) = \int_{-\infty}^{\infty} dk_z \sum_{\ell=0}^{\infty} G_m(k_z) e^{ik_z z} \frac{(k_z r)^{2\ell+|m|}}{2^{2\ell+|m|} \ell! (\ell + |m|)!}. \quad (1.138)$$

We now define a set of functions  $C_m^{[0]}(z)$ , called the *generalised gradients*:

$$C_m^{[0]}(z) = \frac{1}{2^{|m|} |m|!} \int_{-\infty}^{\infty} dk_z k_z^{|m|} G_m(k_z) e^{ik_z z}. \quad (1.139)$$

The  $n$ th derivative of  $C_m^{[0]}(z)$  is written:

$$C_m^{[n]}(z) = \frac{\partial^n}{\partial z^n} C_m^{[0]}(z) = \frac{i^n}{2^{|m|} |m|!} \int_{-\infty}^{\infty} dk_z k_z^{|m|+n} G_m(k_z) e^{ik_z z}. \quad (1.140)$$

In terms of the generalised gradients, the function  $\Phi_m(r, z)$  can be written:

$$\Phi_m(r, z) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) r^{2\ell+|m|}, \quad (1.141)$$

and the magnetic scalar potential is:

$$\phi_{\text{mag}} = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^\ell |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) r^{2\ell+|m|} e^{im\theta}. \quad (1.142)$$

From (1.136) the condition for the magnetic scalar potential to be a real quantity (i.e. to have imaginary part zero) is:

$$C_{-m}^{[0]}(z) = C_m^{[0]}(z)^*. \quad (1.143)$$

It is possible to determine the generalised gradients  $C_m^{[n]}(z)$  from numerical field data. Before we describe how to do this, we shall give expressions for the vector potential  $\mathbf{A}$  and the magnetic field  $\mathbf{B}$  in terms of the generalised gradients: these expressions will be useful for studies of the effect of a given three-dimensional magnetic field on the beam dynamics in an accelerator. The scalar and vector potentials are related by:

$$\mathbf{B} = \nabla \times \mathbf{A} = -\nabla \phi_{\text{mag}}. \quad (1.144)$$

We shall work in a gauge for the vector potential in which  $A_\theta = 0$  (see the comments preceding equations (1.114) and (1.115)). With this gauge condition, (1.144) leads to the relationships:

$$\frac{\partial A_z}{\partial \theta} = -r \frac{\partial \phi_{\text{mag}}}{\partial r}, \quad (1.145)$$

$$\frac{\partial A_z}{\partial r} - \frac{\partial A_r}{\partial z} = \frac{1}{r} \frac{\partial \phi_{\text{mag}}}{\partial \theta}, \quad (1.146)$$

$$\frac{\partial A_r}{\partial \theta} = r \frac{\partial \phi_{\text{mag}}}{\partial z}. \quad (1.147)$$

Then, from (1.142), the vector potential has components:

$$A_r = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell+1]}(z) r^{2\ell+|m|+1} \frac{i e^{im\theta}}{m}, \quad (1.148)$$

$$A_\theta = 0, \quad (1.149)$$

$$A_z = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^\ell |m|! (2\ell + |m|)}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) r^{2\ell+|m|} \frac{i e^{im\theta}}{m}. \quad (1.150)$$

Note that the  $m = 0$  term must be excluded from the summations. The  $m = 0$  term corresponds to a monopole component in the field. Although magnetic monopoles are excluded from physics by Maxwell's equation (1.2), monopole components in a magnetic field (in the sense of a field component that is independent of angular co-ordinate  $\theta$ ) can occur, for example, in solenoids. The  $m = 0$  term needs special treatment, and we shall return to this case later. For now, however, we assume that the field has no monopole component.

Using (1.144) and (1.142), we find the following expressions for the components of the magnetic field, in terms of the generalised gradients:

$$B_r = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} (2\ell + |m|) |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) r^{2\ell+|m|-1} e^{im\theta}, \quad (1.151)$$

$$B_\theta = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z) r^{2\ell+|m|-1} i m e^{im\theta}, \quad (1.152)$$

$$B_z = \sum_{m=-\infty}^{\infty} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell+1]}(z) r^{2\ell+|m|} e^{im\theta}. \quad (1.153)$$

A field with components given by (1.151)–(1.153) will satisfy Maxwell's equations, for *any* functions  $C_m^{[0]}(z)$ . If the generalised gradients are known for a given field, then the vector potential can be expressed in the form (1.148)–(1.150); the vector potential can then be used for analysis of the beam dynamics, as we shall discuss in Chapter 2 (see also Chapter 10).

When studying the beam dynamics, it is usually more convenient to work in Cartesian co-ordinates rather than cylindrical co-ordinates. The components of the vector potential in a Cartesian basis are readily obtained from (1.148)–(1.150), and can be expressed in terms of Cartesian co-ordinates using:

$$r^{|m|} e^{im\theta} = (x + i \operatorname{sgn}(m)y)^{|m|}, \quad (1.154)$$

$$r = \sqrt{x^2 + y^2}. \quad (1.155)$$

The final question that we need to address at this stage is how to construct the generalised gradients  $C_m^{[0]}(z)$  for a given field. Suppose that we have a set of numerical field data, consisting of the field components  $B_r$ ,  $B_\theta$  and  $B_z$ , on a grid of points for different  $r$ ,  $\theta$  and  $z$ . The first step is to take the data for the radial component  $B_r$  at a particular longitudinal position  $z = z_0$ , and radius  $r = r_0$ . By performing a discrete Fourier transform on the set of values of  $B_r$  as a function of  $\theta$ , a set of coefficients  $\tilde{B}_m(r_0, z_0)$  can be obtained such that:

$$B_r(r_0, \theta, z) = \sum_{m=0}^{M-1} \tilde{B}_m(r_0, z_0) e^{im\theta}. \quad (1.156)$$

By comparison with (1.151), we see that:

$$\tilde{B}_m(r_0, z_0) = \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} (2\ell + |m|)! |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z_0) r_0^{2\ell + |m|-1} e^{im\theta}. \quad (1.157)$$

The next step is to repeat the construction of the set of values  $\tilde{B}_m(r_0, z_0)$  for different values of  $r_0$ , at fixed  $z_0$ . Then, a set of coefficients  $b_{m,\ell}(z_0)$  can be obtained by performing a polynomial fit up to some order  $\ell'_{\max}$ :

$$\tilde{B}_m(r, z_0) = \sum_{\ell=0}^{\ell'_{\max}} b_{m,\ell}(z_0) r^{2\ell + |m|-1}. \quad (1.158)$$

Comparing (1.158) with (1.157), we can say that:

$$b_{m,\ell}(z_0) = \frac{(-1)^{\ell+1} (2\ell + |m|)! |m|!}{2^{2\ell} \ell! (\ell + |m|)!} C_m^{[2\ell]}(z_0). \quad (1.159)$$

Hence, the generalised gradients  $C_m^{[2\ell]}(z_0)$  are obtained at as many different longitudinal positions  $z_0$  as desired (or as data permits). The generalised gradients  $C_m^{[2\ell+1]}(z_0)$  can be obtained by following a similar procedure, but starting with the longitudinal field component  $B_z$  (1.153).

Before concluding our discussion of generalized gradients, let us consider briefly the case of a solenoid field. We noted previously that in the series expansions for the vector potential (1.148)–(1.150), the term  $m = 0$  needs to be excluded from the summations: this term corresponds to a field component that has rotational symmetry about the  $z$  axis. Hence, solenoid magnets (which have axial symmetry) need to be considered as a special case. Series expansions for axially-symmetric fields are given by [Dragt (1990)]:

$$A_r = 0, \quad (1.160)$$

$$A_\theta = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell r^{2\ell+1} b_{2\ell}}{2^{2\ell+1} \ell! (\ell + 1)!}, \quad (1.161)$$

$$A_z = 0, \quad (1.162)$$

where:

$$b_{2\ell} = \frac{\partial^{2\ell}}{\partial z^{2\ell}} B_z(0, 0, z). \quad (1.163)$$

The corresponding magnetic field is:

$$B_r = \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell+1} r^{2\ell+1} b_{2\ell+1}}{2^{2\ell+1} \ell! (\ell+1)!}, \quad (1.164)$$

$$B_\theta = 0, \quad (1.165)$$

$$B_z = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell r^{2\ell} b_{2\ell}}{2^{2\ell} (\ell!)^2}. \quad (1.166)$$

These expressions can be used to describe the field not just in the body of a solenoid (where the field is, at least approximately, independent of the longitudinal co-ordinate  $z$ ), but also in the fringe field region at the entrance or exit of the solenoid, where the field will vary with  $z$ . For further details, see [Dragt (1990, 2020)].

Expressing a magnetic field in the form of a series such as (1.151)–(1.153) (or for the corresponding vector potential, (1.148)–(1.150)) allows for the application of some powerful techniques to study the effects of the field on the beam dynamics in an accelerator (see, for example, Chapter 10). However, achieving an accurate fit to a given set of field data is often a challenging task. In the process described above, for example, the values obtained for the coefficients  $b_{m,\ell}(z_0)$  in (1.158) will depend on the value chosen for  $\ell_{\max}$ , because the monomial functions in the power series are not orthogonal. This is the same issue that was discussed in Section 1.2.3, in the context of fitting multipole components to a set of field data. Fortunately, if sufficient field data is available, the accuracy of a fit can be assessed by comparing the results with data not used in making the fit. This is possible because Maxwell's equations impose strong constraints on how electric and magnetic fields vary in space and time. In principle, for a static magnetic field in free space, we can construct the generalised gradients (including all the derivatives) from two of the field components at a single longitudinal position. Then, from the generalised gradients, we can reconstruct all three field components anywhere in space. It is therefore possible to perform a variety of consistency checks to validate the results of the fit, or to overconstrain the fit to improve accuracy and reliability. The details of the techniques used to fit a set of field data will depend on the specific features of the system concerned.

### **1.3.3 Analytical expressions for multipole fringe fields**

The techniques for describing three-dimensional magnetic fields discussed in Sections 1.3.1 and 1.3.2 are based on series expansions. This approach can have some drawbacks; for example, in some cases a large number of terms may be needed to provide an accurate description of the field, and this will increase the computational cost of numerical evaluation of the field components. Furthermore, in the case of generalised gradients, the series expressions only provide exact solutions of Maxwell's equations if all the terms in the infinite series are included: truncation of the series leads to a field that is not strictly physical. For these reasons, in some applications it may be desirable to have a full description of a three-dimensional field (i.e. a field that varies as a function of all three spatial co-ordinates) in terms of closed-form analytical expressions that provide exact solutions of Maxwell's equations. In this section, we shall present such expressions, in a form suitable for describing fringe fields in multipole magnets, i.e. the fields around the entrance and exit faces of some of the types of magnet commonly used in accelerator beam lines. Under certain conditions, multipole fringe fields can have a significant impact on the motion of particles moving along a beam line: this is often the case, for example, for dipole fringe fields, but fringe fields can also be of concern in quadrupoles, sextupoles etc.

In addition to multipole magnets, accelerator beam lines sometimes also include solenoids (for example, in the interaction region of a collider, or to provide focusing in particle sources). For completeness, as well as discussing analytical expressions for multipole fringe fields, we consider the magnetic field around a finite solenoid. It should be noted, however, that the expressions we give for a solenoid field apply to the case of a solenoid in free space, i.e. in the absence of any magnetic materials. Generally, solenoids in accelerators will include iron cores or clamp plates, which can modify the field significantly: in such cases, the approach described in Section 1.3.1, in which the field at the end of a solenoid can be described (for example) by an Enge function, might be more appropriate. The case of a solenoid in free space may, however, be a useful starting point for studies of the effects of solenoid fields in accelerator beam lines.

Our derivation of expressions for the field around a finite solenoid follows that given in [Callaghan and Maslen (1960)]; analytical expressions for magnetic fields in a variety of systems with axial symmetry, including solenoids, can be found in [Conway (2001)] and [Derby and Olbert (2010)]. Methods for computation of solenoid fields based on a number of different

models are discussed in [Granum *et al.* (2022)]. Our derivation of analytical expressions for fringe fields in multipole magnets follows that of [Muratori *et al.* (2015)]. In this section, we shall consider only the description of the magnetic field, and not the motion of charged particles through the field. The dynamics of particles in dipole fringe fields will be treated in Section 3.3; general techniques for particle tracking in three-dimensional fields will be discussed in Chapter 10.

**Field around a finite solenoid** The magnetic field around a finite solenoid can be derived by first considering the field around a single loop of wire carrying a steady current. The solenoid is then constructed by summing over many coaxial wire loops, distributed uniformly over the length of the solenoid. It is convenient in the present case to start by finding an expression for the vector potential  $\mathbf{A}$ , and then obtain the magnetic field  $\mathbf{B}$  using  $\mathbf{B} = \nabla \times \mathbf{A}$ .

In general, the vector potential  $\mathbf{A}(\mathbf{x}_0)$  at a point  $\mathbf{x}_0$  in the vicinity of a current  $I(\mathbf{x})$  of infinitesimal length at a point  $\mathbf{x}$  in free space is given by:

$$\mathbf{A}(\mathbf{x}_0) = \frac{\mu_0}{4\pi} \frac{I(\mathbf{x}) d\mathbf{l}}{|\mathbf{x} - \mathbf{x}_0|}, \quad (1.167)$$

where  $\mu_0$  is the permeability of free space, and  $d\mathbf{l}$  is an infinitesimal vector in the direction of the current. Consider a single loop of radius  $a$  formed from a wire carrying current  $I$ , with centre at the origin and lying in the plane  $z = 0$  of a cylindrical polar co-ordinate system (with radial co-ordinate  $r$  and angular co-ordinate  $\theta$ ). Equation (1.167) then leads to the following expression for the vector potential around the wire loop:

$$A_z = A_r = 0, \quad A_\theta(\mathbf{x}_0) = \frac{\mu_0 I a}{4\pi} \int_{-\pi}^{\pi} \frac{\cos(\theta')}{\mathcal{D}} d\theta', \quad (1.168)$$

where  $\mathcal{D}$  is the distance from a point on the wire with angular co-ordinate  $\theta'$  to the point at  $\mathbf{x}_0$  where the vector potential is observed. We denote the radial and longitudinal co-ordinates at the observation point  $r$  and  $z$ , respectively; because of the symmetry of the system, we can assume without loss of generality that the observation point has angular co-ordinate  $\theta = 0$ . We then find (see Fig. 1.9):

$$\mathcal{D}^2 = z^2 + a^2 + r^2 - 2ar \cos(\theta'). \quad (1.169)$$

Note that in (1.169),  $r$  and  $z$  are the radial and longitudinal co-ordinates of the observation point (i.e. the point at which the field is to be calculated), while  $\theta'$  is the angular co-ordinate of a point on the wire loop.

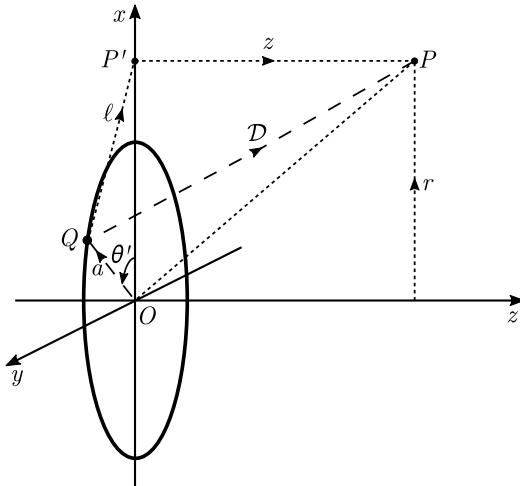


Fig. 1.9 Calculation of the distance  $\mathcal{D}$  between a point  $Q$  on a wire loop of radius  $a$  with centre at the origin  $O$ , and a point  $P$  at which an observer sees the field generated by a current flowing through the loop. The point  $P'$  is the projection of  $P$  onto the  $x-y$  plane (the plane of the loop); hence, the angle between  $QP'$  and  $P'P$  is a right angle, and  $\mathcal{D}^2 = z^2 + \ell^2$ , where  $\ell$  is the length of  $QP'$ . Since the length of  $OP'$  is  $r$  (the radial co-ordinate of  $P$ ), using the cosine rule we find  $\ell^2 = a^2 + r^2 - 2ar \cos(\theta')$ , where  $\theta'$  is the angle between  $OP'$  and  $OQ$ . Hence we have Eq. (1.169):  $\mathcal{D}^2 = z^2 + a^2 + r^2 - 2ar \cos(\theta')$ .

The vector potential around a finite solenoid of length  $L$  (with axis along the  $z$  axis, and centre at the origin) can be found by superposition of the vector potential of a set of coaxial wire loops, with each loop in a plane  $z = z'$  with  $-L/2 \leq z' \leq +L/2$ :

$$A_\theta = \frac{\mu_0 n I a}{2\pi} \int_{-L/2}^{L/2} \int_0^\pi \frac{\cos(\theta')}{\sqrt{(z-z')^2 + a^2 + r^2 - 2ar \cos(\theta')}} d\theta' dz', \quad (1.170)$$

where there are  $n$  loops per unit length along the  $z$  axis. If we define  $\zeta = z - z'$ , and change the variable of integration from  $z'$  to  $\zeta$ , this becomes:

$$A_\theta = \frac{\mu_0 n I a}{2\pi} \int_{\zeta_-}^{\zeta_+} \int_0^\pi \frac{\cos(\theta')}{\sqrt{\zeta^2 + a^2 + r^2 - 2ar \cos(\theta')}} d\theta' d\zeta, \quad (1.171)$$

where  $\zeta_\pm = z \pm \frac{L}{2}$ . Performing the integral over  $\zeta$  gives:

$$A_\theta = \frac{\mu_0 n I a}{2\pi} \int_0^\pi \cos(\theta') \left[ \ln(\zeta + \sqrt{\zeta^2 + a^2 + r^2 - 2ar \cos(\theta')}) \right]_{\zeta_-}^{\zeta_+} d\theta'. \quad (1.172)$$

The field is then obtained from  $\mathbf{B} = \nabla \times \mathbf{A}$ , with the result:

$$B_r = -\frac{\mu_0 n I a}{2\pi} \left[ \int_0^\pi \frac{\cos(\theta') d\theta'}{\sqrt{\zeta^2 + a^2 + r^2 - 2ar \cos(\theta')}} \right]_{\zeta_-}^{\zeta_+}, \quad (1.173)$$

$$B_z = \frac{\mu_0 n I a}{2\pi} \left[ \int_0^\pi \frac{\zeta (a - r \cos(\theta')) d\theta'}{(a^2 + r^2 - 2ar \cos(\theta')) \sqrt{\zeta^2 + a^2 + r^2 - 2ar \cos(\theta')}} \right]_{\zeta_-}^{\zeta_+}. \quad (1.174)$$

The expressions (1.173) and (1.174) can be written in terms of elliptic integrals [Callaghan and Maslen (1960); Abramowitz and Stegun (1964); Gradshteyn and Ryzhik (2014a)]:

$$B_r = \frac{\mu_0 n I}{\pi} \sqrt{\frac{a}{r}} \left[ \left( \frac{\sqrt{\bar{\eta}}}{2} - \frac{1}{\sqrt{\bar{\eta}}} \right) K(\eta) + \frac{E(\eta)}{\sqrt{\bar{\eta}}} \right]_{\zeta_-}^{\zeta_+}, \quad (1.175)$$

$$B_z = \frac{\mu_0 n I}{4\pi} \left[ \zeta \sqrt{\frac{\eta}{ar}} K(\eta) + \pi \operatorname{sgn}((a - r)\zeta) \Lambda_0(\varphi|\eta) \right]_{\zeta_-}^{\zeta_+}, \quad (1.176)$$

where:

$$\eta = \frac{4ar}{\zeta^2 + (a+r)^2}, \quad \text{and} \quad \varphi = \tan^{-1} \left| \frac{\zeta}{a-r} \right|. \quad (1.177)$$

In these expressions,  $K(\eta)$  and  $E(\eta)$  are (respectively) complete elliptic integrals of the first and second kinds. The function  $\Lambda_0(\varphi|\eta)$  is the Heuman lambda function:

$$\Lambda_0(\varphi|\eta) = \frac{F(\varphi|\bar{\eta})}{K(\bar{\eta})} + \frac{2}{\pi} K(\eta) Z(\varphi|\bar{\eta}), \quad (1.178)$$

where  $\bar{\eta} = 1 - \eta$ , and  $Z(\varphi|\bar{\eta})$  is the Jacobi zeta function:

$$Z(\varphi|\bar{\eta}) = E(\varphi|\bar{\eta}) - \frac{E(\bar{\eta}) F(\varphi|\bar{\eta})}{K(\bar{\eta})}. \quad (1.179)$$

$F(\varphi|\bar{\eta})$  and  $E(\varphi|\bar{\eta})$  are incomplete elliptic integrals of the first and second kinds, respectively. Plots illustrating the field around a solenoid are shown in Fig. 1.10.

In some cases, it may be useful to have simplified, approximate expressions for the fields, even if the approximate expressions do not provide exact solutions to Maxwell's equations. For  $r \ll a$ , the fields (1.175) and (1.176) are given approximately by:

$$B_r \approx \frac{\mu_0 n I}{4} \left[ \frac{a^2 r}{(\zeta^2 + a^2)^{\frac{3}{2}}} \right]_{\zeta_-}^{\zeta_+}, \quad (1.180)$$

$$B_z \approx \frac{\mu_0 n I}{2} \left[ \frac{\zeta}{\sqrt{\zeta^2 + a^2}} \right]_{\zeta_-}^{\zeta_+}. \quad (1.181)$$

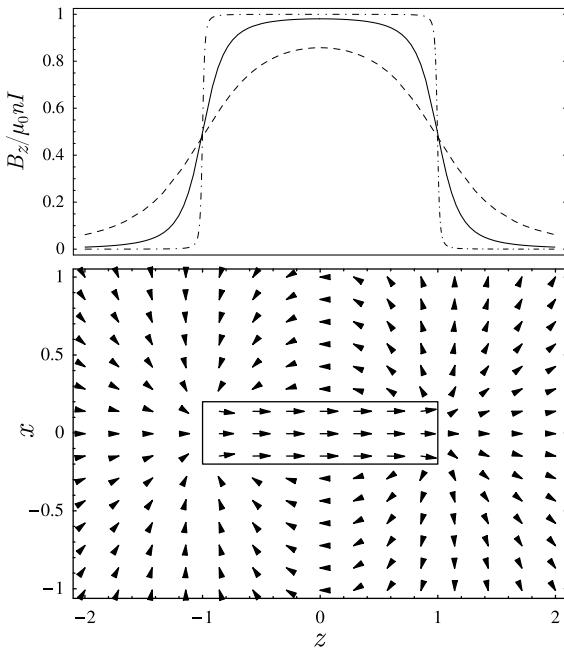


Fig. 1.10 Field around a solenoid in free space. The solenoid extends from  $z = -1$  to  $z = +1$ . The top plot shows the longitudinal field (1.176) along the  $z$  axis for solenoids of different radii,  $a = 0.02$  (dash-dotted line),  $a = 0.2$  (solid line) and  $a = 0.6$  (dashed line). The bottom plot shows the field vectors in the  $x$ - $z$  plane for the case  $a = 0.2$  (with the outline showing the dimensions of the solenoid).

**Dipole fringe fields** Before considering fringe fields in multipole magnets in general, we shall first consider the simplest case, that of a dipole. Within the main part of the magnet, there is a uniform magnetic field: we shall assume that the orientation of the magnet is such that in a Cartesian basis, the field has a non-zero component  $B_y = B_0$ , while the other components are  $B_x = B_z = 0$ . We shall further assume that the entrance and exit faces are parallel to each other, and perpendicular to the  $z$  axis.<sup>1</sup> Approaching (from point of view of a particle in an accelerator) the entrance face of the magnet along the  $z$  axis, the field  $B_y$  increases from zero to the nominal field strength  $B_0$ . Within the magnet, the field remains constant, and finally beyond the exit face the field falls back to zero. The

<sup>1</sup>Note that with this geometry and co-ordinate system, a charged particle entering the magnet along the  $z$  axis will not exit the magnet along this axis, because of the deflection of the trajectory of the particle as it travels through the field within the magnet.

regions around the entrance and exit faces in which the field is changing with position are the ‘fringe field’ regions. In the simplest approximation of the fringe field, the  $B_y$  component rises abruptly from 0 to the nominal value  $B_0$  upon entering the magnet through the entrance face, and then falls abruptly from  $B_0$  to 0 on leaving the magnet through the exit face. However, in this approximation the  $x$  component of the curl of the magnetic field  $\nabla \times \mathbf{B}$  (which involves the derivative of  $B_y$  with respect to  $z$ ) would have an infinite value. To avoid such a non-physical situation, the variation of  $B_y$  along the  $z$  axis needs to be described by a continuous and smoothly-varying function. To construct a better approximation for the fringe field in a dipole, we will consider functions that vary smoothly between the appropriate limits (i.e. zero outside the magnet, and  $B_0$  within the magnet), and that satisfy Maxwell’s equations.

To proceed, we simplify the problem by considering the entrance and exit faces of the dipole separately, and treating the magnet as if it extended an infinite distance beyond the face being considered. At the entrance face, the variation of the field is described by a function that varies smoothly from  $B_y = 0$  in the limit  $z \rightarrow -\infty$  to  $B_y = B_0$  in the limit  $z \rightarrow \infty$ . One possibility in this case is to base the field variation on an Enge function [Enge (1967)]:

$$B_y(0, 0, z) = \frac{B_0}{1 + e^{E(z)}}, \quad \text{where} \quad E(z) = \sum_{k=0}^{N-1} a_k \left(\frac{z}{D}\right)^k. \quad (1.182)$$

The constant  $D$ , with units of length, is introduced for convenience, so that the  $N$  Enge coefficients  $a_k$  are dimensionless.  $D$  is often associated with the aperture of the dipole: a larger aperture leads to a slower variation of the field through the fringe region, which can be described (for fixed  $a_k$ ) by a larger value of  $D$ . The detailed shape of the field variation is determined by the set of  $N$  Enge coefficients  $a_k$ . A reasonably accurate representation of the field may be obtained in some cases with just a single coefficient,  $a_1$  (with  $a_1/D < 0$  at the entrance face, and  $a_1/D > 0$  at the exit face).

The expression (1.182) describes the variation of a single component of the field along a single line in space: for a full description of the field we need to extend this expression so that all three field components are expressed as functions of all three spatial co-ordinates. To achieve this, we use the fact that for any arbitrary function  $f(\eta)$  with  $\eta = z+iy$ , expressions

of the form:

$$B_x = 0, \quad (1.183)$$

$$B_y = f(\eta) + \text{c.c.}, \quad (1.184)$$

$$B_z = -if(\eta) + \text{c.c.}, \quad (1.185)$$

provide solutions to Maxwell's equations for a static field in free space:

$$\nabla \cdot \mathbf{B} = \nabla \times \mathbf{B} = 0. \quad (1.186)$$

In (1.184) and (1.185), 'c.c.' represents the complex conjugate of the preceding term in the expression in which it appears: in that case, the expressions for the fields  $B_y$  and  $B_z$  take real values. Note that the field components in (1.183)–(1.185) have no dependence on the  $x$  co-ordinate: this implies that the system has translational symmetry along the  $x$  axis, which would be the case for dipoles in which the poles have infinite extent in the  $\pm x$  directions.

Equations (1.183)–(1.185) provide a natural way to generalise the Enge function in (1.182) to all field components and all points within a dipole fringe field. For an Enge-type fringe field, we use  $f(\eta) = f_{\text{Enge}}(\eta)$ :

$$f_{\text{Enge}}(\eta) = \frac{B_0}{2(1 + e^{E(\eta)})}, \quad (1.187)$$

which leads to expressions for the field components:

$$B_x = 0, \quad (1.188)$$

$$B_y = \frac{B_0}{2(1 + e^{E(z+iy)})} + \frac{B_0}{2(1 + e^{E(z-iy)})}, \quad (1.189)$$

$$B_z = \frac{iB_0}{2(1 + e^{E(z-iy)})} - \frac{iB_0}{2(1 + e^{E(z+iy)})}. \quad (1.190)$$

The field components can be expressed in terms of (explicitly) real quantities as follows:

$$B_x = 0, \quad (1.191)$$

$$B_y = \frac{B_0(1 + e^\theta \cos(\psi))}{1 + 2e^\theta \cos(\psi) + e^{2\theta}}, \quad (1.192)$$

$$B_z = \frac{-B_0 e^\theta \sin(\psi)}{1 + 2e^\theta \cos(\psi) + e^{2\theta}}, \quad (1.193)$$

where  $\theta$  and  $\psi$  are real variables given by:

$$\theta + i\psi = \sum_{k=0}^{N-1} a_k \left( \frac{z + iy}{D} \right)^k. \quad (1.194)$$

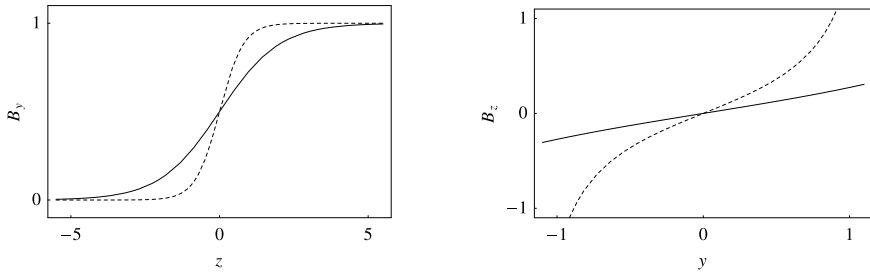


Fig. 1.11 Enge-type fringe field at the entrance of a dipole magnet. The left-hand plot shows the vertical field component ( $B_y$ ) for  $y = 0$ ; the right-hand plot shows the longitudinal field component ( $B_z$ ) for  $z = 0$ . The field has  $B_0 = 1$ , and is described by a single Enge coefficient,  $a_1$ . The solid lines show the case  $a_1/D = -1$ , and the dashed lines show the case  $a_1/D = -2.5$ .

Enge-type fringe fields, each with a single Enge coefficient, at the entrance of a dipole magnet are illustrated in Figs. 1.11 and 1.12. Each figure shows two cases, one with Enge coefficient  $a_1/D = -1$  (solid lines in Fig. 1.11 and upper plots in Fig. 1.12) and another with  $a_1/D = -2.5$  (dashed lines in Fig. 1.11 and lower plots in Fig. 1.12). The field in the body of the dipole magnet is  $B_0 = 1$  in each case. Figure 1.11 shows how the Enge coefficient affects the rate at which the field rises from zero to the full field strength in the body of the dipole. The plots in Fig. 1.12 show the variation of the vertical and longitudinal field components in a vertical plane perpendicular to the  $x$  axis, i.e. the  $y-z$  plane. We observe singularities (in both field components) that are clearly non-physical, but are a consequence of the fact that we have written down the magnetic field as a solution to Maxwell's equations in free space, without regard for the sources of the field. An analogous situation would be the electric field around a charge distribution with spherical symmetry, centred on the origin and contained within a radius  $r = R_0$ , where  $r$  is the distance from the origin. Outside the charge (in free space,  $r > R_0$ ), the field varies in strength as  $1/r^2$ . Although the function describing the field in this region has a singularity at  $r = 0$ , when the spherical charge distribution is taken into account, the field remains finite everywhere: the free-space solution for the field is not valid for  $r < R_0$ , where the charge density is non-zero.

In the case of the magnetic field in Fig. 1.12, the singularities in the field are associated with the sources of the field, as we shall now show. We shall discuss only the case of an Enge-type fringe field, with a single Enge coefficient,  $a_1$ . Consider integrating the field  $\mathbf{B}$  around a closed loop  $\mathcal{C}$  in

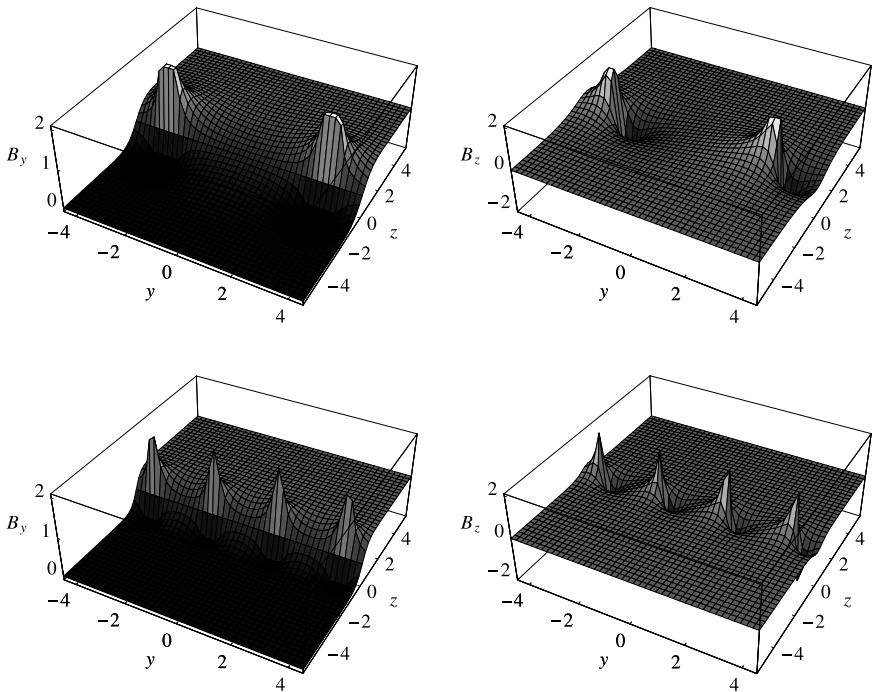


Fig. 1.12 Vertical ( $B_y$ ) and longitudinal ( $B_z$ ) components (left-hand and right-hand plots, respectively) in an Enge-type fringe field at the entrance of a dipole magnet. The field has  $B_0 = 1$ , and is described by a single Enge coefficient,  $a_1$ . The upper plots show the case  $a_1/D = -1$ , and the lower plots show the case  $a_1/D = -2.5$ . Singularities in the vertical component occur at  $z = 0$  and  $y = p\pi D/a_1$  for odd integers  $p$ : the singularities can be understood in terms of currents that act as sources of the field.

the  $y$ - $z$  plane. From Stokes' theorem, we have:

$$\oint_{\mathcal{C}} \mathbf{B} \cdot d\mathbf{l} = \int_{\mathcal{S}} \nabla \times \mathbf{B} \cdot d\mathbf{S}, \quad (1.195)$$

where  $\mathcal{S}$  is the region bounded by the loop  $\mathcal{C}$  in the  $y$ - $z$  plane.  $d\mathbf{l}$  and  $d\mathbf{S}$  are length and area elements of  $\mathcal{C}$  and  $\mathcal{S}$ , respectively. The integral around the loop  $\mathcal{C}$  can also be written:

$$\oint_{\mathcal{C}} \mathbf{B} \cdot d\mathbf{l} = \text{Im} \oint_{\mathcal{C}} (B_y + iB_z)(dz + i dy). \quad (1.196)$$

Here, we treat  $B_y$  and  $B_z$  as the real and imaginary parts of a function of a complex variable  $z + iy$ . From Maxwell's equations (in free space) (1.186), and the fact that  $B_x = 0$ , it follows that  $B_y$  and  $B_z$  satisfy the Cauchy-Riemann conditions for analytic functions:

$$\frac{\partial B_y}{\partial y} = -\frac{\partial B_z}{\partial z}, \quad \frac{\partial B_y}{\partial z} = \frac{\partial B_z}{\partial y}. \quad (1.197)$$

Hence, we can apply the residue theorem:

$$\oint_{\mathcal{C}} (B_y + iB_z)(dz + i dy) = 2\pi i \sum_k \text{Res}(B_y + iB_z, \sigma_k), \quad (1.198)$$

where  $\text{Res}(B_y + iB_z, \sigma_k)$  is the residue of  $B_y + iB_z$  at a point  $\sigma_k$ , and the summation is over all the poles of  $B_y + iB_z$  within the loop  $\mathcal{C}$  (which occur at points  $\sigma_k$ ). The singularities in the field all occur along  $z = 0$ , in which case, for an Enge-type fringe field with a single Enge coefficient:

$$B_y + iB_z = \frac{B_0}{1 + e^{i \frac{a_1}{D} y}}. \quad (1.199)$$

From this, we find that the poles occur at  $y = p\pi D/a_1$ , for  $p$  any odd integer, and that each pole has a residue:

$$\text{Res}(B_y + iB_z, \sigma_k) = -\frac{D}{a_1} B_0. \quad (1.200)$$

Combining (1.195), (1.196), (1.198) and (1.200) we find:

$$\int_{\mathcal{S}} \nabla \times \mathbf{B} \cdot d\mathbf{S} = -2\pi N_{\text{poles}} \frac{D}{a_1} B_0, \quad (1.201)$$

where  $N_{\text{poles}}$  is the number of poles (singularities) in the field within the region  $\mathcal{S}$ . Finally, using Maxwell's equation:

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \quad (1.202)$$

where  $\mathbf{J}$  is the current density,  $\mathbf{E}$  the electric field, and  $\mu_0$ ,  $\varepsilon_0$  are the permeability and permittivity of free space, we find that (by taking the limit of an infinitesimal loop  $\mathcal{C}$  enclosing just one singularity) we can interpret the singularities in the field in terms of currents flowing through infinitesimally thin wires, with the current in each wire given by:

$$I = \frac{2\pi D}{\mu_0 a_1} B_0. \quad (1.203)$$

There are infinitely many wires, each parallel to the  $x$  axis, positioned at regular intervals  $\Delta y = 2\pi D/a_1$  along the  $y$  axis, passing through points  $z = 0$  and  $y = p\pi D/a_1$ , where  $p$  is any odd integer.

In practice, the fringe field in a dipole magnet will have a more complicated form than we have assumed in the present case. The field will be determined by the geometry and material properties of the dipole magnet (including, in an electromagnet, the currents in the coils), and singularities in the field will of course not appear.

As a final observation on the singularities in an Enge-type fringe field in a dipole magnet, it is worth noting that in the limit  $D/a_1 \rightarrow 0$ , the

dipole field becomes ‘hard-edged’: that is, the field changes abruptly over an infinitesimal longitudinal distance, from zero outside the magnet to the full strength  $B_0$  within the magnet. The singularity (this time, in the gradient of the field) can again be understood in terms of electric currents. To see that this is the case, first observe that from Eq. (1.203), using the model in which singularities in the field are associated with currents in infinitesimally thin wires, the current per unit distance along the  $y$  axis is:

$$\frac{I}{\Delta y} = \frac{B_0}{\mu_0}, \quad (1.204)$$

where  $I$  is the current in each wire, and  $\Delta y = 2\pi D/a_1$  is the spacing between the wires. In the hard-edged limit  $D/a_1 \rightarrow 0$ , the wires form a continuous ‘current sheet’, carrying current (in the  $x$  direction) per unit length (along the  $y$  axis) given by (1.204). This is the same result that would be obtained from an elementary application of Maxwell’s equation (1.202), for the change in the magnetic field on crossing an infinitely wide conducting sheet of infinitesimal thickness, carrying current per unit length  $I/\Delta y$  (for example, in a septum magnet). The calculation is similar to that carried out in Section 1.1.1, with surface current density  $J_{s\perp}$  replaced by  $I/\Delta y$ , the current per unit distance.

**Scalar and vector potentials for dipole fringe fields** For some purposes, for example when considering magnet design or for particle tracking, it can be useful to describe the magnetic field in terms of a potential. For the static case, the curl of the magnetic field vanishes ( $\nabla \times \mathbf{B} = 0$ ), and it is then possible to define a scalar potential  $\phi_{\text{mag}}$  (1.61), such that:

$$\mathbf{B} = -\nabla\phi_{\text{mag}}. \quad (1.205)$$

Note that the potential  $\phi_{\text{mag}}$  is distinct from the usual scalar potential, from which the electric field is often derived. Also, the minus sign in (1.205) is a conventional choice, and means that the field lines point in the direction of *decreasing* potential.<sup>2</sup> The magnetic scalar potential (1.205) can be useful in the context of magnet design, since in an iron-core magnet, in the approximation that the iron has infinite magnetic permeability, the field lines outside the core are perpendicular to the surface of the iron, which forms an equipotential surface.

Using the fact that, from (1.184) and (1.185):

$$B_y + iB_z = 2f(\eta), \quad (1.206)$$

---

<sup>2</sup>The opposite choice, with field lines pointing in the direction of increasing potential, is made in [Muratori *et al.* (2015)].

where  $\eta = z + iy$ , it can be shown that the scalar potential can be written:

$$\phi_{\text{mag}}(\eta) = i \int_0^\eta f(\eta') d\eta' + \text{c.c.} \quad (1.207)$$

In the case of an Enge field (1.182) with a single coefficient  $a_1$  ( $a_k = 0$  for  $k \neq 1$ ), we find for the scalar potential in the fringe field of a dipole magnet:

$$\phi_{\text{mag,Enge}} = -yB_0 - \frac{iD}{2a_1} B_0 \ln \left( \frac{1 + e^{\frac{a_1}{D}(z+iy)}}{1 + e^{\frac{a_1}{D}(z-iy)}} \right). \quad (1.208)$$

Note that this expression for the scalar potential is real-valued: there is no need to add the complex conjugate.

A magnetic field can also be derived from a vector potential  $\mathbf{A}$ :

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (1.209)$$

For static fields, it is always possible to choose a gauge in which one component of the vector potential is zero. In the case of the dipole field (1.183)–(1.185), where the field has a translational symmetry along one co-ordinate axis, it is possible to use a gauge in which two of the components of the vector potential vanish. If we choose a gauge such that:

$$\mathbf{A} = (A_x, 0, 0), \quad (1.210)$$

then, again using (1.206), we find that  $A_x$  is given by:

$$A_x = \int_0^\eta f(\eta') d\eta'. \quad (1.211)$$

In the case of the Enge field derived from (1.187), with only a single non-zero Enge coefficient ( $a_k = 0$  for  $k \neq 1$ ) this leads to:

$$A_{x,\text{Enge}} = zB_0 - \frac{D}{2a_1} B_0 \ln \left( 1 + e^{\frac{a_1}{D}(z+iy)} \right) - \frac{D}{2a_1} B_0 \ln \left( 1 + e^{\frac{a_1}{D}(z-iy)} \right). \quad (1.212)$$

Again, this expression is real-valued: there is no need to add the complex conjugate.

### Fringe fields in quadrupole and higher-order multipole magnets

Fringe fields in dipole magnets are important in accelerator beam physics, because the focusing effects of dipole fringe fields can have a significant impact on beam optics. In quadrupole and higher-order multipole magnets, fringe fields are also of importance, but for different reasons. First, the shape of the fringe field can have a significant effect on the effective or ‘magnetic’ length of the component, which may be different from the physical length measured (for example) as the length of the iron core. Second,

in some cases multipole fringe fields can affect the nonlinear dynamics in an accelerator.

In developing analytical expressions for fringe fields in dipole magnets, we assumed that the magnet has translational symmetry along the  $x$ -axis: the corresponding field component is then exactly zero, which simplifies the problem considerably. In quadrupoles and higher-order multipole magnets, we cannot make a similar assumption, and we are therefore forced to take a different approach to the analysis of fringe fields in these cases. As a first step, before discussing the fringe fields in these types of magnet, we first establish an appropriate notation by writing expressions for the fields in the body of a multipole magnet in a suitable form. Once the notation is established, we will write down analytical expressions for the fringe fields. We follow the notation, and the general approach, used in [Muratori *et al.* (2015)].

We begin by defining new co-ordinates  $(u, v, \zeta)$ , related to the Cartesian co-ordinates  $(x, y, z)$  (with  $z$  the co-ordinate along the axis of the magnet):

$$u = \frac{1}{\sqrt{2}}(x + iy), \quad v = \frac{1}{\sqrt{2}}(x - iy), \quad \zeta = \sqrt{2}z. \quad (1.213)$$

We define corresponding components for the magnetic field:

$$B_u = \frac{1}{\sqrt{2}}(B_x + iB_y), \quad B_v = \frac{1}{\sqrt{2}}(B_x - iB_y), \quad B_\zeta = \frac{1}{\sqrt{2}}B_z. \quad (1.214)$$

Using these co-ordinates and field components, Maxwell's equations may be written:

$$\partial_u B_u + \partial_\zeta B_\zeta = 0, \quad (1.215)$$

$$\partial_v B_v + \partial_\zeta B_\zeta = 0, \quad (1.216)$$

$$\partial_\zeta B_u - \partial_v B_\zeta = 0, \quad (1.217)$$

$$\partial_\zeta B_v - \partial_u B_\zeta = 0. \quad (1.218)$$

In the body of a multipole magnet, the field is independent of the distance along the axis, i.e. the field components have no dependence on  $z$  (or  $\zeta$ ). In that case,  $\partial_\zeta B_u = \partial_\zeta B_v = 0$ , and we see that Maxwell's equations have solutions with:

$$B_u = B_u(v), \quad B_v = B_v(u), \quad B_\zeta = \text{constant}. \quad (1.219)$$

For example, we can write a solution:

$$B_u = i\mathcal{G}_n v^n, \quad B_v = -i\mathcal{G}_n u^n, \quad B_\zeta = 0, \quad (1.220)$$

where  $\mathcal{G}_n$  and  $n$  are constants. In this case, the field components  $B_x$  and  $B_y$  are given in terms of the radial co-ordinate  $r$  and polar angle  $\theta$ :

$$B_x = 2^{\frac{1-n}{2}} \mathcal{G}_n r^n \sin(n\theta), \quad B_y = 2^{\frac{1-n}{2}} \mathcal{G}_n r^n \cos(n\theta). \quad (1.221)$$

The field (1.220) represents the field in a multipole of order  $n$  (with  $n = 0$  for a dipole,  $n = 1$  for a quadrupole, etc.).

The next step is to find a solution to Maxwell's equations (1.215)–(1.218) that includes an appropriate dependence on  $\zeta$ , to be able to describe the field in the fringe-field region of a multipole magnet. It is readily verified (by substitution into Maxwell's equations) that a possible solution can be written as a superposition of fields with components:

$$B_{u,j} = i b_j f(\zeta + ih_j) - i b_j g(\zeta - ih_j), \quad (1.222)$$

$$B_{v,j} = \frac{i}{b_j} f(\zeta + ih_j) - \frac{i}{b_j} g(\zeta - ih_j), \quad (1.223)$$

$$B_{\zeta,j} = f(\zeta + ih_j) + g(\zeta - ih_j), \quad (1.224)$$

where  $b_j$  is an arbitrary constant,  $f$  and  $g$  are arbitrary functions, and  $h_j$  is defined by:

$$h_j = \frac{u}{b_j} + b_j v = \frac{1}{\sqrt{2}} \left( \frac{1}{b_j} + b_j \right) x + \frac{i}{\sqrt{2}} \left( \frac{1}{b_j} - b_j \right) y. \quad (1.225)$$

The superposition is expressed as a sum over the index  $j$ , with each term in the summation having a coefficient  $c_j$ :

$$B_u = \sum_j c_j B_{u,j}, \quad B_v = \sum_j c_j B_{v,j}, \quad B_{\zeta} = \sum_j c_j B_{\zeta,j}. \quad (1.226)$$

The fact that  $f$  and  $g$  are arbitrary functions provides sufficient generality in the solution to be able to, eventually, describe multipole fringe fields. For the moment, however, we consider the body of a multipole (where the field has no dependence on  $z$ ), and look for appropriate expressions for the functions  $f$  and  $g$ . In the case of a quadrupole magnet, we simply write:

$$f(\zeta + ih_j) = (\zeta + ih_j) \mathcal{G}_1, \quad g(\zeta - ih_j) = (\zeta - ih_j) \mathcal{G}_1, \quad (1.227)$$

where  $\mathcal{G}_1$  is a constant. From (1.222)–(1.224), we find that:

$$B_{u,j} = -2\mathcal{G}_1 (u + b_j^2 v), \quad B_{v,j} = -2\mathcal{G}_1 \left( \frac{u}{b_j^2} + v \right), \quad B_{\zeta,j} = 2\mathcal{G}_1 \zeta. \quad (1.228)$$

Then, for a quadrupole, we superpose two fields (i.e. fields for different values of the coefficients  $b_j$  and  $c_j$ , with  $j = 1, 2$ ) of the form (1.228) with:

$$c_1 = -\frac{i}{2(b_1^2 - b_2^2)}, \quad c_2 = -\frac{i}{2(b_2^2 - b_1^2)}, \quad (1.229)$$

to obtain:

$$B_u = i\mathcal{G}_1 v, \quad B_v = -i \frac{\mathcal{G}_1}{b_1^2 b_2^2} u, \quad B_\zeta = 0. \quad (1.230)$$

If we choose:

$$b_1 = \pm \frac{1}{b_2}, \quad (1.231)$$

then we have, finally:

$$B_u = i\mathcal{G}_1 v, \quad B_v = -i\mathcal{G}_1 u, \quad B_\zeta = 0, \quad (1.232)$$

which is, as required, the field in a quadrupole magnet. The quadrupole gradient is  $\mathcal{G}_1$ .

The result (1.232) can be generalised to higher-order multipoles. For a multipole of order  $n$ , the functions  $f$  and  $g$  are chosen to be:

$$f(\zeta + ih_j) = (\zeta + ih_j)^n \mathcal{G}_n, \quad g(\zeta - ih_j) = (-1)^{n+1} (\zeta - ih_j)^n \mathcal{G}_n. \quad (1.233)$$

The superposition (1.226) then involves  $n + 1$  terms: the requirement that the resulting expressions take the appropriate form (1.220) leads to a set of algebraic equations for the constants  $c_j$  and  $b_j$  (for  $j = 1 \dots n + 1$ ); details are given in [Muratori *et al.* (2015)]. The solution to the equations for  $c_j$  and  $b_1$  may be expressed:

$$c_j = \frac{(-i)^n b_j^{n-1}}{2 \prod_{k=1, k \neq j}^{n+1} (b_j^2 - b_k^2)}, \quad b_1 = \pm \frac{i^{n+1}}{\prod_{k=2}^{n+1} b_k}. \quad (1.234)$$

Note that the equations do not determine the values of all the parameters  $b_j$ : there is only a single constraint on these quantities (expressed as the second of equations (1.234)), so that in a multipole of order  $n$ , there are  $n$  free parameters. The values chosen for these parameters do not affect the field in the body of the magnet, but do have an effect on the fringe field. For a dipole magnet,  $n = 0$ , there are no free parameters, but the values of  $b_1$  and  $c_1$  are given by:

$$b_1 = \pm i, \quad c_1 = \frac{1}{2b_1}. \quad (1.235)$$

We are now in a position to write down expressions for the fringe fields in multipole magnets. Since fields based on (1.233) with constant  $\mathcal{G}_n$  can be used to construct the field in the body of a multipole magnet (with gradient  $\mathcal{G}_n$ ), we only have to replace  $\mathcal{G}_n$  by a function  $G(\zeta \pm ih_j)$ , where the upper and lower signs are used for terms in the functions  $f$  and  $g$ , respectively, so that the expressions for  $f$  and  $g$  (1.233) become:

$$f(\zeta + ih_j) = (\zeta + ih_j)^n G(\zeta + ih_j), \quad (1.236)$$

$$g(\zeta - ih_j) = (-1)^{n+1} (\zeta - ih_j)^n G(\zeta - ih_j). \quad (1.237)$$

The main constraint on  $G$  is that the function should have the appropriate limiting behaviour for  $z \rightarrow \pm\infty$ . In particular, for the entrance fringe field, we require that:

$$\lim_{z \rightarrow -\infty} G(\zeta \pm ih) = 0, \quad \lim_{z \rightarrow +\infty} G(\zeta \pm ih) = \mathcal{G}_n. \quad (1.238)$$

This ensures that the field vanishes at large distances upstream of the entrance of the magnet, but that the field has the correct behaviour in the body of the magnet. In principle, any function with the appropriate behaviour can be used for  $G$ : the particular choice will depend on the specific case being considered. One possibility is to require that the multipole gradient should vary with distance along the axis of the magnet as an Enge function. For example, for a single Enge coefficient  $a_1$ , and constant  $D$  (with  $a_1/D < 0$  for the entrance fringe field, and  $a_1/D > 0$  for the exit fringe field) we would write:

$$\frac{1}{n!} \left. \frac{\partial^n B_y}{\partial x^n} \right|_{x=y=0} = \frac{\mathcal{G}_n}{1 + e^{\frac{a_1}{D} z}}. \quad (1.239)$$

In this case, (with the appropriate limits, (1.238))  $G$  is given by:

$$\frac{G(\zeta \pm ih_j)}{\mathcal{G}_n} = F\left(n; -\frac{a_1}{\sqrt{2}D}(\zeta \pm ih_j)\right), \quad (1.240)$$

where the function  $F(n; \xi)$  is defined as:

$$F(n; \xi) = -\frac{n!}{\xi^n} \text{Li}_n(-e^\xi) + \sum_{k=1}^n \frac{n!}{(n-k)! \xi^k} \text{Li}_k(-1). \quad (1.241)$$

$\text{Li}_n(\xi)$  is the polylogarithm (or Jonquière function) of order  $n$ .

The above results, in particular (1.213), (1.214), (1.222)–(1.224), (1.226), (1.234), (1.236) and (1.237), provide the ingredients for writing down analytic expressions for the fringe field in a multipole magnet. The exact shape of the fringe field (including, for example, the rate of roll-off of the gradient as a function of distance from the entrance face or exit face of the magnet) is determined by the choice of function  $G$ . For a fringe field described by a single-coefficient Enge function,  $G$  is given by (1.240).

As we observed above, for a multipole of order  $n$ , there are  $n$  free parameters  $b_j$  (for  $j = 1 \dots n+1$ , with one constraint in the form of the second equation of (1.234)). The exact shape of the fringe field will be affected by the choice of values for these parameters, as well as by the choice of function  $G$ . It may be appropriate, in some cases, to superpose solutions with different choices of  $b_j$  and  $G$ , or with particular transformations applied to the

co-ordinates: this allows the construction of solutions with specific properties, such as symmetries associated with the geometry of the magnet. For example, quadrupole magnets are often constructed with pole face geometries such that the magnet has four-fold symmetry under rotations around the axis of the magnet: as a result, the field has symmetry under reflection in the plane  $x = y$  (i.e. under exchange of the  $x$  and  $y$  co-ordinates, the  $B_x$  and  $B_y$  components of the field are also exchanged). A fringe field with this symmetry cannot be produced by a single solution with a particular choice of  $b_j$  and  $G$ . However, the symmetry can be achieved by superposing solutions with the same  $b_j$  and  $G$ , but with the  $x$  and  $y$  co-ordinates exchanged.

**Enge-type fringe field in a quadrupole** In general, expressions for fringe fields in multipole magnets can be rather lengthy and complicated. However, it is possible to write down explicit expressions for the field components in certain cases. As an example, consider the entrance fringe field in a quadrupole magnet with gradient  $\mathcal{G}_1$  in the body of the magnet, and geometry having four-fold rotational symmetry. We assume that the field gradient along the  $z$  axis (i.e. for  $x = y = 0$ ) in the fringe field can be described by an Enge function with a single parameter  $a_1$  and constant  $D$ :

$$\frac{\partial B_x}{\partial y} = \frac{\partial B_y}{\partial x} = \frac{\mathcal{G}_1}{1 + e^{az}}, \quad (1.242)$$

where we define  $a = a_1/D$ . The polylogarithm has the property that  $\text{Li}_1(\chi) = -\ln(1 - \chi)$ ; and hence for  $n = 1$ ,  $F(n; \xi)$  from (1.240) takes the form:

$$F(1; \xi) = \frac{1}{\xi} (\ln(1 + e^\xi) - \ln(2)). \quad (1.243)$$

In the limit  $b_2 \rightarrow 1$ , the field components can be written:

$$B_x = \mathcal{G}_1 \frac{y}{2} \left( \frac{e^{-az} + \cos(ax)}{\cos(ax) + \cosh(az)} \right), \quad (1.244)$$

$$B_y = i \frac{\mathcal{G}_1}{2a} \ln \left( \frac{e^{az} + e^{-i\alpha x}}{e^{az} + e^{i\alpha x}} \right), \quad (1.245)$$

$$B_z = -\mathcal{G}_1 \frac{y}{2} \left( \frac{\sin(ax)}{\cos(ax) + \cosh(az)} \right). \quad (1.246)$$

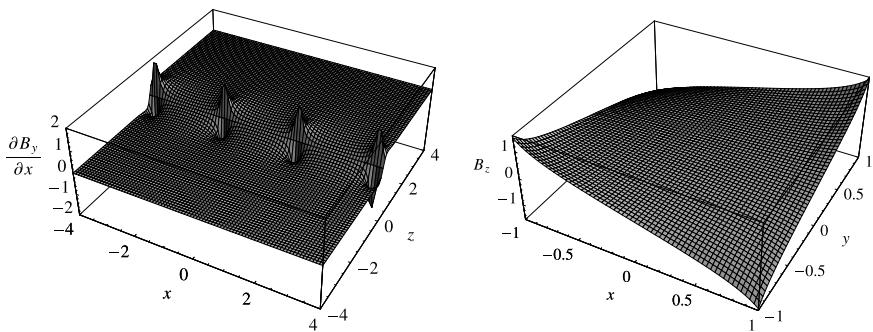


Fig. 1.13 Enge-type fringe field at the entrance of a quadrupole magnet. The left-hand plot shows the quadrupole gradient in the  $x-z$  plane, where  $z$  is the longitudinal co-ordinate (along the axis of the magnet). The field has been symmetrised so that a plot of the gradient in the  $y-z$  plane would be identical. The right-hand plot shows the longitudinal field component in the  $x-y$  plane at  $z = 0$  (halfway through the fringe field region). The fringe field parameters in this case are  $b_2 = 1$ , and  $a_1/D = -2.5$ .

The field may be symmetrised as described above, by constructing:

$$B'_x = \frac{1}{2} (B_x + B_y|_{x \leftrightarrow y}), \quad (1.247)$$

$$B'_y = \frac{1}{2} (B_y + B_x|_{x \leftrightarrow y}), \quad (1.248)$$

$$B'_z = \frac{1}{2} (B_z + B_z|_{x \leftrightarrow y}), \quad (1.249)$$

where  $B'_x$ ,  $B'_y$ , and  $B'_z$  are the components of the symmetrised field, and  $B_x|_{x \leftrightarrow y}$  is the expression for  $B_x$  but with  $x$  and  $y$  interchanged (and similarly for  $B_y|_{x \leftrightarrow y}$  and  $B_z|_{x \leftrightarrow y}$ ). The gradient in an Enge-type fringe field in a quadrupole after symmetrising the field in this way is illustrated in Fig. 1.13.

**Scalar and vector potentials for multipole fringe fields** As mentioned in the discussion of dipole fringe fields, it is useful for some applications to have expressions for potentials from which a given magnetic field can be derived. At locations where there is no electric current and no time-dependent electric field (so that  $\nabla \times \mathbf{B} = 0$ ), the magnetic field can be derived from a scalar potential  $\phi_{\text{mag}}$ , as in (1.61):

$$\mathbf{B} = -\nabla\phi_{\text{mag}}. \quad (1.250)$$

The relationship between the field and the scalar potential leads to:

$$B_\zeta = -\frac{\partial\phi_{\text{mag}}}{\partial\zeta}. \quad (1.251)$$

In a multipole fringe field, the field component  $B_\zeta$  can be constructed from (1.224) and (1.226):

$$B_\zeta = \sum_{j=1}^{n+1} c_j \left( f(\zeta + ih_j) + g(\zeta - ih_j) \right), \quad (1.252)$$

with the coefficients  $c_j$  given by (1.234). It then follows that the scalar potential  $\phi_{\text{mag}}$  can be expressed as:

$$\phi_{\text{mag}} = - \sum_{j=1}^{n+1} c_j \left( \tilde{f}(\zeta + ih_j) + \tilde{g}(\zeta - ih_j) \right), \quad (1.253)$$

where  $\tilde{f}$  and  $\tilde{g}$  are found by integrating the functions  $f$  and  $g$  along the  $z$  axis (i.e. for real values of their respective arguments):

$$\tilde{f}(\zeta) = \int_0^\zeta f(\zeta') d\zeta', \quad \tilde{g}(\zeta) = \int_0^\zeta g(\zeta') d\zeta'. \quad (1.254)$$

In the case of an Enge-type fringe field with parameter  $a_1$  and constant  $D$ , the functions  $f(\zeta)$  and  $g(\zeta)$  take the form:

$$f(\zeta) = (-1)^{n+1} g(\zeta) = \mathcal{G}_n \zeta^n F\left(n; -\frac{a_1}{\sqrt{2}D} \zeta\right). \quad (1.255)$$

The integrals (1.254) then lead to:

$$\tilde{f}(\zeta) = (-1)^{n+1} \tilde{g}(\zeta) = \mathcal{G}_n \frac{\zeta^{n+1}}{n+1} F\left(n+1; -\frac{a_1}{\sqrt{2}D} \zeta\right). \quad (1.256)$$

As usual, to obtain the dependence on the transverse co-ordinates, i.e. a full expression for the scalar potential  $\phi_{\text{mag}}$  (1.253),  $\zeta$  is replaced by  $\zeta + ih_j$  in  $\tilde{f}(\zeta)$ , and by  $\zeta - ih_j$  in  $\tilde{g}(\zeta)$ . For a dipole magnet,  $n = 0$ , and with  $b_1$  and  $c_1$  given by (1.235), the results (1.253) and (1.256) lead to the scalar potential (1.208). In the case of a quadrupole magnet,  $n = 1$ , taking the limit  $b_2 \rightarrow 1$ , we find the scalar potential:

$$\phi_{\text{mag}} = -i \frac{\mathcal{G}_1}{2a} y \ln\left(\frac{e^{az} + e^{-iaz}}{e^{az} + e^{iaz}}\right). \quad (1.257)$$

As usual, this may be symmetrised to match the usual symmetry of a quadrupole magnet:

$$\phi'_{\text{mag}} = \frac{1}{2} \left( \phi_{\text{mag}} + \phi_{\text{mag}}|_{x \leftrightarrow y} \right). \quad (1.258)$$

For some purposes, it can be useful to have an expression for the vector potential from which the magnetic field in the fringe region of a multipole

magnet may be derived. The usual relationship between a magnetic field  $\mathbf{B}$  and a vector potential  $\mathbf{A}$  is:

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (1.259)$$

In terms of the co-ordinates  $u$ ,  $v$  and  $\zeta$ , this relationship can be expressed:

$$B_u = \sqrt{2}i \frac{\partial A_u}{\partial \zeta}, \quad B_v = -\sqrt{2}i \frac{\partial A_v}{\partial \zeta}, \quad B_\zeta = \frac{1}{\sqrt{2}i} \left( \frac{\partial A_u}{\partial u} - \frac{\partial A_v}{\partial v} \right). \quad (1.260)$$

Then, it is readily verified that the field described by (1.222)–(1.224) can be derived from a vector potential:

$$A_u = \frac{b_j}{\sqrt{2}} \left( \tilde{f}(\zeta + ih_j) - \tilde{g}(\zeta - ih_j) \right), \quad (1.261)$$

$$A_v = -\frac{1}{\sqrt{2}b_j} \left( \tilde{f}(\zeta + ih_j) - \tilde{g}(\zeta - ih_j) \right), \quad (1.262)$$

$$A_\zeta = 0, \quad (1.263)$$

where  $\tilde{f}$  and  $\tilde{g}$  are the integrals of  $f$  and  $g$ , (1.254). This potential satisfies the Coulomb gauge condition:

$$\nabla \cdot \mathbf{A} = \frac{\partial A_u}{\partial u} + \frac{\partial A_v}{\partial v} + \frac{\partial A_\zeta}{\partial \zeta} = 0. \quad (1.264)$$

In the particular case of a magnet with an Enge-type fringe field,  $\tilde{f}$  and  $\tilde{g}$  are given by (1.256). Then, for a dipole magnet,  $n = 0$ , and with  $b_1$  and  $c_1$  given by (1.235), the results (1.261)–(1.263) lead to the vector potential (1.212), with an additive constant (which is of no significance for the magnetic field). For an Enge-type fringe field in a quadrupole magnet in the limit  $b_2 \rightarrow 1$  (1.244)–(1.246), the vector potential is found to be:

$$A_x = \frac{\mathcal{G}_1}{a} x \ln(2) + \frac{i\mathcal{G}_1}{2a^2} \left( \text{Li}_2(-e^{-a(z+ix)}) - \text{Li}_2(-e^{-a(z-ix)}) \right), \quad (1.265)$$

$$A_y = \frac{\mathcal{G}_1}{2a} y \left( \ln(1 + e^{-a(z+ix)}) + \ln(1 + e^{-a(z-ix)}) - 2 \ln(2) \right), \quad (1.266)$$

$$A_z = 0. \quad (1.267)$$

## 1.4 Fields in Radiofrequency Cavities

So far, we have considered only (static) magnetic fields. While such fields are important in accelerators for steering and focusing beams of charged particles, to increase the energy of a particle, an electric field is needed. For

the initial acceleration of particles from a source, static electric fields are often used; however, there are practical limitations on the voltages that can be achieved with static electric fields. To accelerate particles to energies in the MeV range or above, oscillating electric fields are required, either in a straight linear accelerator (linac) or in a circular machine (for example, a cyclotron or a synchrotron).

Accelerators can operate over a very wide range of parameters, and we will not attempt a comprehensive review of the different types of component available for providing fields for accelerating particles. Here, we shall limit the discussion to the principles behind radiofrequency (rf) cavities, which are widely used in linacs and circular accelerators. We shall show how the boundary conditions on the electric and magnetic fields in an rf cavity determine the oscillation frequencies of fields that can exist within the cavity. The fields corresponding to each oscillation frequency form particular spatial patterns: understanding the dependence of the fields both in space and in time is important in understanding the effects of rf cavities on the motion of particles passing through the cavities. A complete description of the dynamical effects can be very complicated, but fortunately the dominant effects can often be stated fairly easily. Our goal in this section is to set out the basic principles, allowing the most important features of rf cavities to be understood.

The region inside an rf cavity can be treated as a vacuum, while the walls of an rf cavity can be treated (in a simplifying approximation) as an ideal conductor. Maxwell's equations in free space admit wave-like solutions for the electric and magnetic fields. The boundary conditions imposed by the walls of a cavity can be satisfied by fields that take the form of standing waves within the cavity. In Sections 1.4.1 and 1.4.2, we shall show how the boundary conditions lead to constraints on the oscillation frequencies and spatial variation of fields that can exist in a cavity of specified geometry.

Before discussing fields in cavities, it is worth reviewing briefly the wave equation for electromagnetic fields in free space. Starting with Maxwell's equation (1.4), we take the curl of both sides, and use a vector identity to give:

$$\nabla \times \nabla \times \mathbf{E} \equiv \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial}{\partial t} \nabla \times \mathbf{B}. \quad (1.268)$$

Now we use Maxwell's equations (1.1) and (1.3), together with  $\mathbf{B} = \mu \mathbf{H}$  and  $\mathbf{D} = \epsilon \mathbf{E}$ , to give:

$$\nabla \frac{\rho}{\epsilon} - \nabla^2 \mathbf{E} = -\mu \frac{\partial \mathbf{J}}{\partial t} - \mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad (1.269)$$

In free space, we can say that  $\rho = \mathbf{J} = 0$ ,  $\mu = \mu_0$  and  $\varepsilon = \varepsilon_0$ . Therefore, the electric field satisfies the wave equation:

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad (1.270)$$

where:

$$c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}. \quad (1.271)$$

Starting with Maxwell's equation (1.3), following a very similar procedure leads to the wave equation for the magnetic field:

$$\nabla^2 \mathbf{B} - \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} = 0. \quad (1.272)$$

The wave equations (1.270) and (1.272) have plane wave solutions:

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}, \quad (1.273)$$

$$\mathbf{B} = \mathbf{B}_0 e^{i(\mathbf{k} \cdot \mathbf{x} - \omega t)}. \quad (1.274)$$

Substituting these expressions for the fields into the respective wave equations leads to the dispersion relation between the wave vector  $\mathbf{k}$  and frequency  $\omega$ :

$$\frac{\omega}{|\mathbf{k}|} = c. \quad (1.275)$$

The requirement that the solutions to the wave equations satisfy Maxwell's equations as well as the wave equations leads to constraints on the solutions. For example, from Maxwell's equation (1.4), we find the following relationship between the wave amplitudes and the wave vector:

$$\mathbf{k} \times \mathbf{E}_0 = \omega \mathbf{B}_0. \quad (1.276)$$

Plane electromagnetic waves in free space are transverse waves, with the electric and magnetic fields perpendicular to each other and to the wave vector  $\mathbf{k}$ . With simple plane wave solutions of the form (1.273) and (1.274), we cannot satisfy the boundary conditions on the electromagnetic fields in an rf cavity. In Sections 1.4.1 and 1.4.2, we shall discuss how to construct standing wave solutions (which can be thought of as superpositions of plane wave solutions) that satisfy the wave equations within the interior of a cavity and the boundary conditions at the walls of a cavity.

### 1.4.1 Rectangular cavities

The shapes of rf cavities in accelerators are often highly optimised to maximise performance, and cannot usually be described as simple geometrical objects such as rectangular boxes or cylinders. Nevertheless, some features (of cavities) relevant for beam dynamics can be understood using simplified models. Significant characteristics include the spatial variation of the fields inside the cavity, and the relationship between the spatial variation and the oscillation frequency. In this section, we shall discuss these characteristics using the simplest representation of a cavity, namely, as a rectangular box with perfectly conducting walls. After considering the field patterns and oscillation frequencies of the resonant modes, we shall derive expressions for the stored energy in a rectangular cavity (in a given mode and with given field amplitude), and define the quality factor and shunt impedance. Following from the simple case of a rectangular cavity, we shall discuss in Section 1.4.2 the resonant modes in a cylindrical cavity.

Consider a rectangular cavity with perfectly conducting walls, containing a perfect vacuum (see Fig. 1.14). Within the interior of the cavity, electric and magnetic fields must satisfy the wave equations (1.270) and (1.272). If the walls of the cavity are ideal conductors, then on the walls of the cavity the fields must satisfy the boundary conditions:

$$E_{\parallel} = 0, \quad (1.277)$$

$$B_{\perp} = 0, \quad (1.278)$$

where  $E_{\parallel}$  is the component of the electric field tangential to the wall, and  $B_{\perp}$  is the component of the magnetic field normal to the wall. Free-space

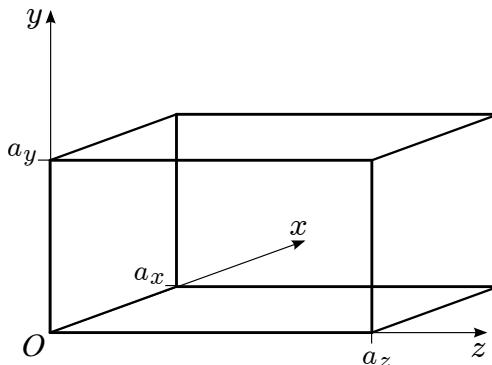


Fig. 1.14 A simple rectangular cavity, with side lengths  $a_x$ ,  $a_y$  and  $a_z$ .

plane wave solutions will not satisfy the boundary conditions. However, we can look for standing wave solutions of the form:

$$E_x = E_{x0} \cos(k_x x) \sin(k_y y) \sin(k_z z) e^{-i\omega t}, \quad (1.279)$$

$$E_y = E_{y0} \sin(k_x x) \cos(k_y y) \sin(k_z z) e^{-i\omega t}, \quad (1.280)$$

$$E_z = E_{z0} \sin(k_x x) \sin(k_y y) \cos(k_z z) e^{-i\omega t}. \quad (1.281)$$

To satisfy the wave equation (1.270), the wave numbers  $k_x$ ,  $k_y$  and  $k_z$  must be related to the frequency  $\omega$  by:

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{c^2}. \quad (1.282)$$

Maxwell's equation  $\nabla \cdot \mathbf{E} = 0$  imposes a constraint on the components of the wave vector and the amplitudes of the field components:

$$k_x E_{x0} + k_y E_{y0} + k_z E_{z0} = 0. \quad (1.283)$$

To satisfy the boundary condition that the tangential component of the electric field vanishes at the walls of the cavity, we require:

$$k_x = \frac{m_x \pi}{a_x}, \quad k_y = \frac{m_y \pi}{a_y}, \quad k_z = \frac{m_z \pi}{a_z}, \quad (1.284)$$

where the *mode numbers*  $m_x$ ,  $m_y$  and  $m_z$  are integers, and  $a_x$ ,  $a_y$  and  $a_z$  are the lengths of the sides of the cavity along the  $x$ ,  $y$  and  $z$  axes, respectively. If the mode numbers are specified, there is still some freedom (in general) in choosing the relative amplitudes of the field components; however, for a given size and shape of cavity, the frequency of oscillation of the field is determined completely by the mode numbers.

The magnetic field can be obtained from the electric field, using Maxwell's equation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (1.285)$$

This gives:

$$B_x = \frac{i}{\omega} (E_{y0} k_z - E_{z0} k_y) \sin(k_x x) \cos(k_y y) \cos(k_z z) e^{-i\omega t}, \quad (1.286)$$

$$B_y = \frac{i}{\omega} (E_{z0} k_x - E_{x0} k_z) \cos(k_x x) \sin(k_y y) \cos(k_z z) e^{-i\omega t}, \quad (1.287)$$

$$B_z = \frac{i}{\omega} (E_{x0} k_y - E_{y0} k_x) \cos(k_x x) \cos(k_y y) \sin(k_z z) e^{-i\omega t}. \quad (1.288)$$

The fields with components (1.279)–(1.281) and (1.286)–(1.288) satisfy Maxwell's equations, and also satisfy the boundary conditions at the walls of the cavity.

To determine the frequency of oscillation of the fields at a given mode in a cavity, we substitute the electric field (1.279)–(1.281) into the wave equation (1.270). This gives:

$$\omega = \pi c \sqrt{\left(\frac{m_x}{a_x}\right)^2 + \left(\frac{m_y}{a_y}\right)^2 + \left(\frac{m_z}{a_z}\right)^2}. \quad (1.289)$$

The same dispersion relation is obtained by substituting the magnetic field (1.286)–(1.288) into the wave equation (1.272). For a cubic cavity ( $a_x = a_y = a_z$ ), there will be some degeneracy, i.e. there will be different sets of mode numbers leading to different fields, but all with the same frequency of oscillation. The degeneracy can be broken by making the lengths of the sides of the cavity different from each other: see Fig. 1.15. Some examples of field patterns in different modes of a rectangular cavity are shown in Fig. 1.16.

It is often useful to know the energy stored in the electromagnetic field in a cavity. Expressions for the energy density and energy flux in an electromagnetic field follow from Poynting's theorem, which may be derived as follows. From Maxwell's equations (1.3) and (1.4), we obtain:

$$\mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H} = -\mathbf{E} \cdot \mathbf{J} - \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t}. \quad (1.290)$$

In a linear medium (i.e. a medium with constant permittivity and permeability), we can write the following:

$$\frac{\partial}{\partial t} \left( \frac{1}{2} \mathbf{D} \cdot \mathbf{E} + \frac{1}{2} \mathbf{B} \cdot \mathbf{H} \right) = -\nabla \cdot (\mathbf{E} \times \mathbf{H}) - \mathbf{E} \cdot \mathbf{J}. \quad (1.291)$$

Integrating over volume  $\mathcal{V}$ , and applying Gauss' theorem gives:

$$\frac{\partial}{\partial t} \int_{\mathcal{V}} U_E + U_B dV = - \int_{\mathcal{S}} \mathbf{S} \cdot d\mathbf{A} - \int_{\mathcal{V}} \mathbf{E} \cdot \mathbf{J} dV, \quad (1.292)$$

where the surface  $\mathcal{S}$  (with area element  $d\mathbf{A}$ ) bounds the volume  $\mathcal{V}$ , and the quantities  $U_E$ ,  $U_B$  and  $\mathbf{S}$  are given by:

$$U_E = \frac{1}{2} \mathbf{D} \cdot \mathbf{E}, \quad (1.293)$$

$$U_B = \frac{1}{2} \mathbf{B} \cdot \mathbf{H}, \quad (1.294)$$

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad (1.295)$$

The integral over  $\mathbf{E} \cdot \mathbf{J}$  gives the total rate at which work is done by the electromagnetic field on any charge within the volume  $\mathcal{V}$ . Conservation of energy then leads to the interpretation of  $U_E$  and  $U_B$  as the energy density

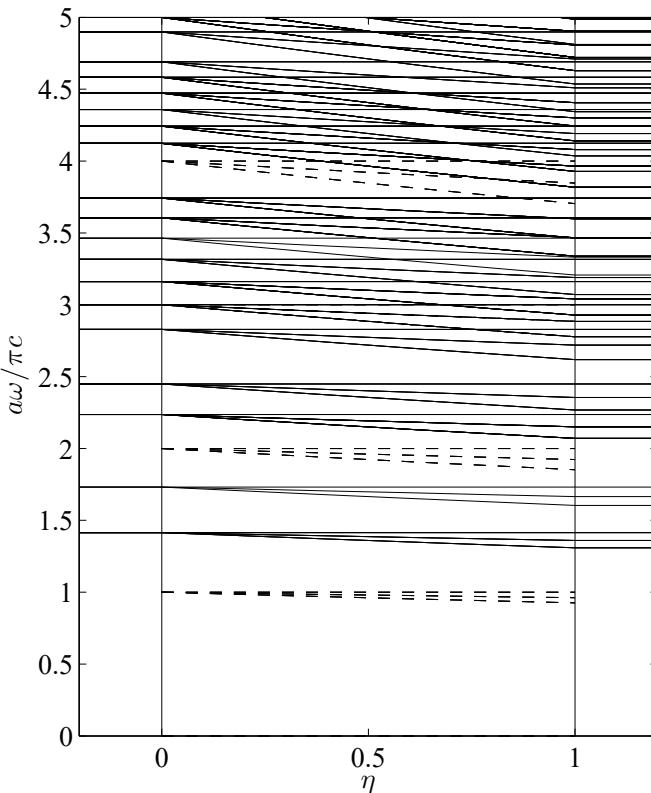


Fig. 1.15 Mode spectra in rectangular cavities. The left-hand side of the plot shows the resonant frequencies in a cavity that has the shape of a cube with side lengths  $a_x = a_y = a_z = a$ . Many of the modes are degenerate (i.e. have the same frequency despite having different mode numbers). The central part of the plot shows the variation in the resonant frequencies with the parameter  $\eta$ , which scales the side lengths so that  $a_x = (1 + 0.04\eta)a$ ,  $a_y = (1 + 0.08\eta)a$  and  $a_z = a$ . Making the side lengths all different breaks the degeneracies in the mode frequencies; the right-hand part of the plot shows the mode spectrum in a rectangular cavity with side lengths  $a_x = 1.04a$ ,  $a_y = 1.08a$  and  $a_z = a$ . Dashed lines in the central part of the plot show modes that necessarily have zero amplitude.

in the electric and magnetic fields, respectively, and  $\mathbf{S}$  as the energy flux (energy crossing unit area per unit time) in the electromagnetic field. The physical significance of these quantities is not without some ambiguity, but for our present purposes can be assumed to be as stated. Equation (1.291) is known as Poynting's theorem, and the vector  $\mathbf{S}$  defined by (1.295) is known as the Poynting vector.

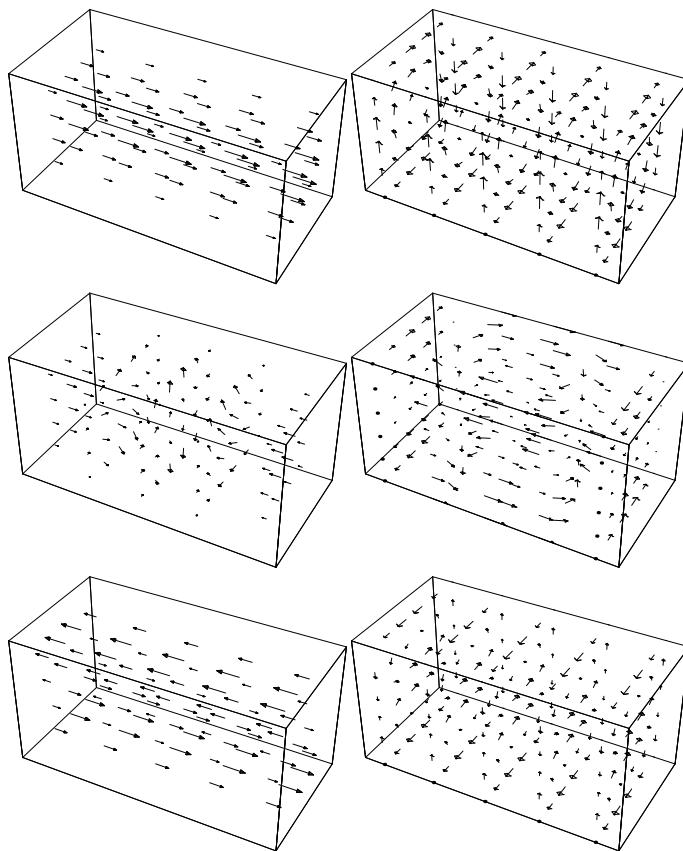


Fig. 1.16 Examples of modes in rectangular cavity. The co-ordinate axes are as shown in Fig. 1.14. From top to bottom, the modes are  $(m_x, m_y, m_z) = (110)$ ,  $(111)$  and  $(120)$ . The size and orientation of the arrows show the strength and direction of the field at different points. Left: electric field. Right: magnetic field.

For a rectangular cavity, it is relatively straightforward to calculate the energy stored in a particular mode with given field amplitudes. If the mode numbers are all non-zero, then using (1.293) we find that the total energy in the electric field is:

$$\mathcal{E}_E = \frac{1}{16} \varepsilon_0 (E_{x0}^2 + E_{y0}^2 + E_{z0}^2) \cos^2(\omega t). \quad (1.296)$$

Similarly, using (1.294) and (1.283), and noting that the magnetic field is  $90^\circ$  out of phase with the electric field, we find that the total energy in the

magnetic field is:

$$\mathcal{E}_B = \frac{1}{16} \frac{1}{\mu_0 c^2} (E_{x0}^2 + E_{y0}^2 + E_{z0}^2) \sin^2(\omega t). \quad (1.297)$$

Finally, using  $1/c^2 = \mu_0 \epsilon_0$ , we find that the total energy in the electromagnetic field (for  $m_x$ ,  $m_y$  and  $m_z$  non-zero integers) is given by:

$$\mathcal{E}_E + \mathcal{E}_B = \frac{1}{16} \epsilon_0 (E_{x0}^2 + E_{y0}^2 + E_{z0}^2). \quad (1.298)$$

The total energy in the cavity is constant over time, although the energy ‘oscillates’ between the electric field and the magnetic field.

The power flux (energy crossing unit area per unit time) in the electromagnetic field is given by the Poynting vector:

$$\mathbf{S} = \mathbf{E} \times \mathbf{H}. \quad (1.299)$$

The electric and magnetic fields in the cavity are  $90^\circ$  out of phase with each other: if the electric field varies as  $\cos(\omega t)$ , then the magnetic field varies as  $\sin(\omega t)$ . Therefore, averaging the Poynting vector over time at any point in the cavity gives zero: this is again consistent with conservation of energy.

Since the tangential component of the electric field vanishes at the surface of an ideal conductor, it follows from the form of the Poynting vector (1.295) that there is no energy flow across the surface. Therefore, in an rf cavity with perfectly conducting walls, all the energy will remain within the cavity, and an oscillating electromagnetic field will persist indefinitely, without any loss of amplitude. This is consistent with the standing wave solutions that we wrote down for the fields; (1.279)–(1.281) and (1.286)–(1.288). In practice, the walls of the cavity will not be perfectly conducting, and the boundary conditions will vary slightly from those we have assumed. The electric and (oscillating) magnetic fields on the walls will induce currents that will dissipate the energy in the fields. The rate of energy dissipation is usually quantified by the *quality factor*,  $Q$ . The equation of motion for a damped harmonic oscillator:

$$\frac{d^2 u}{dt^2} + \frac{\omega}{Q} \frac{du}{dt} + \omega^2 u = 0, \quad (1.300)$$

has the solution:

$$u = u_0 e^{-\frac{\omega t}{2Q}} \cos(\bar{\omega}t - \varphi_0), \quad (1.301)$$

where:

$$\bar{\omega} = \omega \sqrt{\frac{4Q^2 - 1}{4Q^2}}. \quad (1.302)$$

The quality factor  $Q$  (for  $Q \gg 1$ ) is the number of oscillations over which the energy in the oscillator (proportional to the square of the amplitude of  $u$ ) falls by a factor  $1/e$ . More precisely, the rate of energy dissipation (the dissipated power  $P_d$ ) is given by:

$$P_d = -\frac{d\mathcal{E}}{dt} = \frac{\omega}{Q}\mathcal{E}. \quad (1.303)$$

Consider an electromagnetic field inside a cavity that is ‘driven’ at some frequency  $\omega$ . If the frequency corresponds to an allowed mode (i.e. a field described by integer values of  $m_x$ ,  $m_y$  and  $m_z$ ) then it is possible to match the boundary conditions for this mode in a cavity with perfectly conducting walls: the rate of energy dissipation in the walls will be slow, and the fields can be driven to high amplitudes. In other words, the cavity will resonate at the frequency  $\omega$ . However, if an attempt is made to drive oscillations at frequencies not corresponding to integer values of  $m_x$ ,  $m_y$  and  $m_z$ , then the boundary conditions cannot be satisfied, and the energy will be rapidly dissipated.

For a mode with a longitudinal electric field component  $E_{z0} = V_0/L$  (where  $L = a_z$  is the length of the cavity), we define the *shunt impedance*  $R_s$  as follows:

$$R_s = \frac{V_0^2}{2P_d}. \quad (1.304)$$

Note that different definitions of the shunt impedance are used, depending on the context: some definitions omit the factor  $1/2$ . Combining (1.303) and (1.304), we see that:

$$\frac{R_s}{Q} = \frac{V_0^2}{2P_d} \cdot \frac{P_d}{\omega\mathcal{E}} = \frac{V_0^2}{2\omega\mathcal{E}}. \quad (1.305)$$

The quantity  $R_s/Q$  is a measure of the peak field in the cavity, per unit of stored energy. Both the peak field and the stored energy are significant quantities for cavity design. Generally, one of the design goals for an rf cavity in an accelerator is to maximise the peak field (i.e. the accelerating voltage) that can be achieved for a given stored energy: this means maximising the value of  $R_s/Q$ . Although the quality factor characterises the rate of dissipation of the energy in the cavity, we see from (1.305) that  $R_s/Q$  can be expressed purely in terms of quantities that depend on the field amplitude and the mode numbers. Therefore,  $R_s/Q$  must itself depend only on the geometry of the cavity, and not on the material from which the cavity is made. This is a useful result, since it means that when

designing a cavity it is possible to focus initially on optimising the geometry to maximise  $R_s/Q$  for the desired mode (and to minimise  $R_s/Q$  for all other modes), without having to consider from the start the properties of the material from which the cavity will be made.

### 1.4.2 Cylindrical cavities

In Section 1.4.1, we considered the field modes in a rectangular cavity. However, most cavities in accelerators are better approximated with a cylindrical geometry than with a rectangular geometry, and it is therefore worth considering the solutions to Maxwell's equations, subject to the usual boundary conditions, for a cylinder with perfectly conducting walls. In practice, the resonant modes in a realistic accelerator cavity will need to be found by solving Maxwell's equations numerically; however, a simplified model of a cavity as a cylinder can provide some useful results for studies of beam dynamics.

We can find the modes in a cylindrical cavity in the same way as we did for a rectangular cavity: we first look for solutions to the wave equations for the electric and magnetic fields (for example, by using separation of variables). Then, we find the 'allowed' solutions (the resonant modes) by imposing the appropriate boundary conditions. The algebra is more complicated for cylindrical cavities than for rectangular cavities, because we have to work in cylindrical polar co-ordinates. We shall not go through the derivation in detail: the solutions for the fields can be verified by substitution into Maxwell's equations.

One set of modes (not the most general solution) can be expressed as follows:

$$E_r = -iB_0 \frac{n\omega}{k_r^2 r} J_n(k_r r) \sin(n\theta) \sin(k_z z) e^{-i\omega t}, \quad (1.306)$$

$$E_\theta = -iB_0 \frac{\omega}{k_r} J'_n(k_r r) \cos(n\theta) \sin(k_z z) e^{-i\omega t}, \quad (1.307)$$

$$E_z = 0, \quad (1.308)$$

$$B_r = B_0 \frac{k_z}{k_r} J'_n(k_r r) \cos(n\theta) \cos(k_z z) e^{-i\omega t}, \quad (1.309)$$

$$B_\theta = -B_0 \frac{nk_z}{k_r^2 r} J_n(k_r r) \sin(n\theta) \cos(k_z z) e^{-i\omega t}, \quad (1.310)$$

$$B_z = B_0 J_n(k_r r) \cos(n\theta) \sin(k_z z) e^{-i\omega t}. \quad (1.311)$$

$J_n(x)$  is a Bessel function of order  $n$ , and  $J'_n(x)$  is the derivative of  $J_n(x)$ . Bessel functions (some of which are shown in Fig. 1.17) are solutions of the

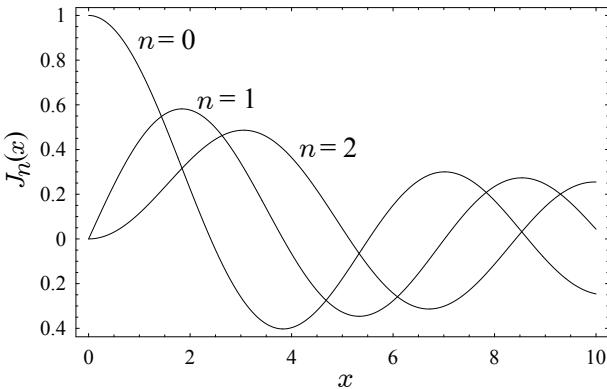


Fig. 1.17 The first three Bessel functions of the first kind,  $J_0(x)$ ,  $J_1(x)$  and  $J_2(x)$ . Bessel functions  $J_n(k_r r)$  describe the radial dependence of the electric and magnetic fields in a cylindrical cavity of radius  $r$ . Possible values of the radial wave number  $k_r$  (which is related by the dispersion relation (1.317) to the oscillation frequency  $\omega$ ) are determined by the boundary conditions on the fields at the cavity walls, and are expressed in terms of zeros of the derivatives of the Bessel functions, through the relationship (1.313).

differential equation:

$$x^2 \frac{d^2 f}{dx^2} + x \frac{df}{dx} + (x^2 - n^2) f = 0, \quad (1.312)$$

which occurs when we separate variables in finding a solution to the wave equation in cylindrical polar co-ordinates.

Because of the dependence of the fields on the azimuthal angle  $\theta$ , we require that  $n$  is an integer:  $n$  provides an azimuthal index in specifying a mode.

From the boundary conditions,  $E_\theta$  and  $B_r$  must both vanish on the curved wall of the cavity, i.e. when  $r = a$ , where  $a$  is the radius of the cylinder. Therefore, we have a constraint on  $k_r$ :

$$J'_n(k_r a) = 0, \quad (1.313)$$

or:

$$k_r = \frac{p'_{nm}}{a} \quad (1.314)$$

where  $p'_{nm}$  is the  $m$ th zero of the derivative of the  $n$ th-order Bessel function. This equation is analogous to the conditions (1.284) that we had for the rectangular cavity. We can use the integer  $m$  as a radial index in specifying a mode.

We also need to have  $B_z = E_r = E_\theta = 0$  on the flat ends of the cavity. Assuming the flat ends of the cavity are at  $z = 0$  and  $z = L$ , these boundary conditions are satisfied if:

$$\sin(k_z L) = 0, \quad (1.315)$$

and hence:

$$k_z = \frac{\ell\pi}{L}, \quad (1.316)$$

where the longitudinal mode number  $\ell$  is an integer.

Substituting the electric field (1.306)–(1.308) into the wave equation (1.270) gives the dispersion relation:

$$k_r^2 + k_z^2 = \frac{\omega^2}{c^2}. \quad (1.317)$$

The same relation is obtained by substituting the magnetic field (1.309)–(1.311) into the wave equation (1.272).

In the field (1.306)–(1.311), the longitudinal component of the electric field vanishes: in other words, the electric field is purely transverse. Modes with this characteristic are known as *TE modes*. A specific mode of this type, with indices  $n$  (azimuthal),  $m$  (radial) and  $\ell$  (longitudinal) is commonly written as  $\text{TE}_{nml}$ . The frequency of a mode  $\text{TE}_{nml}$  depends on the dimensions of the cavity, and is given by:

$$\omega_{nml} = c\sqrt{k_r^2 + k_z^2} = c\sqrt{\left(\frac{p'_{nm}}{a}\right)^2 + \left(\frac{\ell\pi}{L}\right)^2}. \quad (1.318)$$

The fields in the  $\text{TE}_{111}$  mode in a cylindrical cavity are shown in Fig. 1.18.

TE modes are useful for giving a transverse deflection to a beam in an accelerator, but are not much use for providing acceleration. Fortunately, cylindrical cavities allow another set of modes that have non-zero longitudinal electric field:

$$E_r = -E_0 \frac{k_z}{k_r} J'_n(k_r r) \cos(n\theta) \sin(k_z z) e^{-i\omega t}, \quad (1.319)$$

$$E_\theta = E_0 \frac{nk_z}{k_r^2 r} J_n(k_r r) \sin(n\theta) \sin(k_z z) e^{-i\omega t}, \quad (1.320)$$

$$E_z = E_0 J_n(k_r r) \cos(n\theta) \cos(k_z z) e^{-i\omega t}, \quad (1.321)$$

$$B_r = iE_0 \frac{n\omega}{c^2 k_r^2 r} J_n(k_r r) \sin(n\theta) \cos(k_z z) e^{-i\omega t}, \quad (1.322)$$

$$B_\theta = iE_0 \frac{\omega}{c^2 k_r} J'_n(k_r r) \cos(n\theta) \cos(k_z z) e^{-i\omega t}, \quad (1.323)$$

$$B_z = 0. \quad (1.324)$$

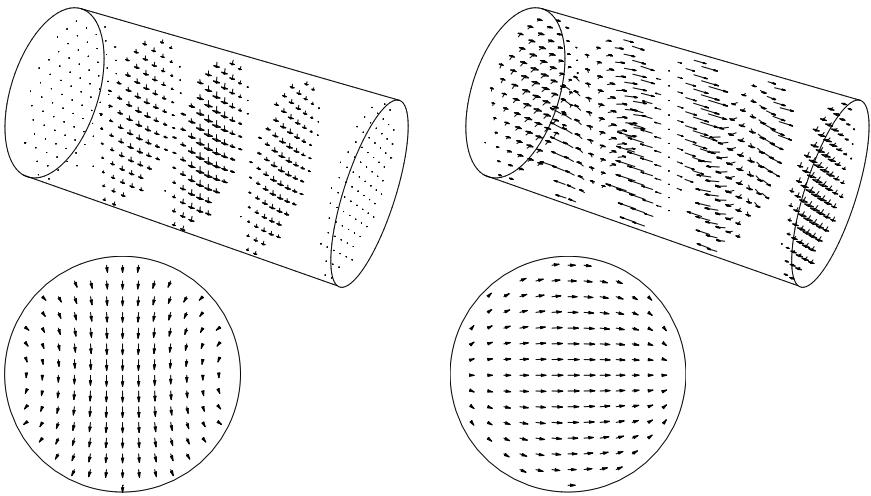


Fig. 1.18 TE<sub>111</sub> mode in a cylindrical cavity. Left: Electric field. Right: Magnetic field. Top: Three-dimensional view. Bottom: Cross-section.

In these modes, the magnetic field is purely transverse (zero longitudinal component); therefore, they are referred to as *TM modes*. As before, for physical fields,  $n$  must be an integer. The boundary conditions on the fields give:

$$k_r = \frac{p_{nm}}{a}, \quad (1.325)$$

$$k_z = \frac{\ell\pi}{L}, \quad (1.326)$$

where  $p_{nm}$  is the  $m$ th zero of the  $n$ th-order Bessel function  $J_n(x)$ . The frequency of a mode TM <sub>$nml$</sub>  is given by:

$$\omega_{nml} = c\sqrt{k_r^2 + k_z^2} = c\sqrt{\left(\frac{p_{nm}}{a}\right)^2 + \left(\frac{\ell\pi}{L}\right)^2}. \quad (1.327)$$

The lowest frequency accelerating mode in a cylindrical cavity is the TM<sub>010</sub> mode ( $n = 0$ ,  $m = 1$ ,  $\ell = 0$ ). The fields in the TM<sub>010</sub> (see Fig. 1.19)

mode are given by:

$$E_r = 0, \quad (1.328)$$

$$E_\theta = 0, \quad (1.329)$$

$$E_z = E_0 J_0\left(p_{01} \frac{r}{a}\right) e^{-i\omega t}, \quad (1.330)$$

$$B_r = 0, \quad (1.331)$$

$$B_\theta = -i \frac{E_0}{c} J_1\left(p_{01} \frac{r}{a}\right) e^{-i\omega t}, \quad (1.332)$$

$$B_z = 0. \quad (1.333)$$

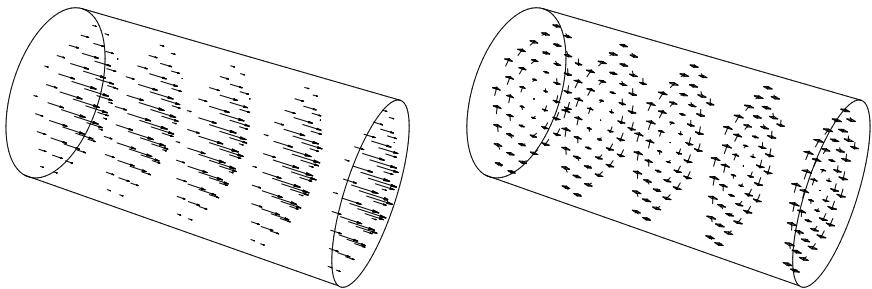


Fig. 1.19 Fields in the  $\text{TM}_{010}$  mode in a cylindrical cavity. Left: Electric field. Right: Magnetic field.

The frequency of the  $\text{TM}_{010}$  mode is determined by the radius of the cavity, and is independent of its length:

$$\omega_{010} = p_{01} \frac{c}{a}, \quad (1.334)$$

where  $p_{01} \approx 2.40483$ . However, for a cylindrical cavity operating in the  $\text{TM}_{010}$  mode in an accelerator, there is a condition on the length necessary to provide the maximum acceleration: the time taken for a particle to pass through the cavity should be one half of the rf period, i.e.  $\pi/\omega$ . If a particle passes through the cavity in a shorter time, it will exit the cavity while there is still a field present in the cavity that could provide additional acceleration. If the field reverses direction before a particle exits the cavity, then the particle will start to be decelerated. Therefore, for maximum efficiency, the length of the cavity should be  $L = \pi v/\omega$ , where  $v$  is the velocity of the particle. In the case of ultra-relativistic particles, for which  $v \approx c$ , the length of the cavity should be  $L = \lambda/2$ , where  $\lambda$  is the wavelength of an electromagnetic wave with angular frequency  $\omega$  in free space.

**This page intentionally left blank**

## Chapter 2

# Hamiltonian for a Particle in an Accelerator Beam Line

The main goal of accelerator beam dynamics is to understand and describe, in as much detail and as accurately as possible, the motion of charged particles in beam lines consisting of sequences of specified electromagnetic fields. In the simplest case, one might consider the motion of just a single particle passing along a beam line where the only fields present are static magnetic fields that take a simple form and are repeated periodically. In more general cases, the fields may take very complex forms and include time-dependent electric and magnetic fields; it may be necessary to account for the interactions between different bunches of particles, or between different particles within a single bunch. Cases of practical importance can range from the very simple to the very complex.

In Chapter 1, we developed some techniques for describing the electromagnetic fields that are used to control charged particle beams in accelerators. We now present some of the basic tools that are used for the description and analysis of the motion of charged particles within electromagnetic fields in accelerator beam lines. For now, we consider only individual particles, and neglect effects arising from the presence of large numbers of particles in a beam line. Our goal in this chapter is to derive some general results that can be applied to describe the dynamics of individual particles in specific types of accelerator component.

In principle, there are only two steps in the analysis of any dynamical system. The first step is to write down the equations of motion; and the second step is to solve them. For particles in accelerators (as in many other systems) there are always several different ways of writing the equations of motion. Often, our ability to solve the equations of motion depends very much on how we choose to write the equations of motion. The approach that we take here is based on Hamiltonian mechanics, and turns out to

have some important advantages when applied to accelerator physics. In particular, it is possible to make use of the conserved quantities that occur in Hamiltonian systems to develop powerful techniques for the analysis of particle dynamics in accelerator beam lines: some of these techniques will be discussed in detail in later chapters (see, for example, Chapters 4, 5 and 9).

Students of physics are commonly introduced to dynamical systems in the framework of Newton's laws of motion. In particular, Newton's second law:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad (2.1)$$

provides an equation of motion for a particle once the force  $\mathbf{F}$  is specified. Note that  $\mathbf{p}$  is the momentum of the particle and  $t$  is the time. For the case of a charged particle in an electromagnetic field, the force is the Lorentz force, given by:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (2.2)$$

where  $q$  is the electric charge on the particle, which is moving with velocity  $\mathbf{v}$ , and  $\mathbf{E}$  and  $\mathbf{B}$  are (respectively) the electric and magnetic fields seen by the particle. In principle, to describe the motion of a particle along an accelerator beam line, we could solve the above equations in the given electric and magnetic fields. However, this is not the approach generally taken in accelerator beam dynamics. The main reason is that solving Newton's second law involves integrating the equation of motion with respect to time, whereas in an accelerator beam line the electric and magnetic fields are usually specified as functions of position along the beam line. An alternative approach, which turns out to provide some very powerful techniques in beam dynamics, is to start instead from Hamilton's equations of motion. Although the time  $t$  may be used as a variable in Hamiltonian mechanics (just as in Newtonian mechanics) to parameterise the trajectory of a particle, Hamiltonian mechanics provides an elegant way to modify the equations to instead use a variable corresponding to the distance along the beam line. This ultimately makes it much simpler to describe the trajectory of a particle moving through a potentially complex sequence of electric and magnetic fields. The techniques of Hamiltonian mechanics are described in numerous works on classical mechanics: see, for example [Goldstein *et al.* (2001a)].

In Hamiltonian mechanics, the dynamics of a given system are determined by a function (the Hamiltonian) that plays a role analogous to the

force in Newtonian mechanics. Using a specific form for the Hamiltonian in a general set of equations (Hamilton's equations) produces the equations of motion for a particular dynamical system. Our goal in this chapter is to derive an expression for the Hamiltonian appropriate for describing the motion of a relativistic charged particle in an accelerator beam line. Later chapters will deal with solving the equations of motion, initially in simple cases (single-particle linear dynamics), but then in progressively more complicated cases (single-particle nonlinear dynamics, and finally multi-particle dynamics).

## 2.1 The Hamiltonian for a Straight Beam Line

Hamilton's equations can be expressed, in general, as:

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}, \quad (2.3)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i}, \quad (2.4)$$

where  $x_i$  are the co-ordinates of the particle ( $i = 1 \cdots N$  in an  $N$ -dimensional co-ordinate space),  $p_i$  are the components of the momentum, and  $H$  is the Hamiltonian. The Hamiltonian defines the dynamics of a particle (in the same way as does the force in Newtonian mechanics), and is expressed as a function of the co-ordinates, momentum and time:

$$H = H(x_i, p_i; t). \quad (2.5)$$

As a simple example, the Hamiltonian:

$$H = \frac{p^2}{2m} + \frac{k}{2}x^2, \quad (2.6)$$

where  $x$  is the co-ordinate and  $p$  the momentum, leads by substituting into Hamilton's equations (2.3) and (2.4) to the equations of motion:

$$\frac{dx}{dt} = \frac{p}{m}, \quad (2.7)$$

$$\frac{dp}{dt} = -kx. \quad (2.8)$$

Equation (2.7) simply gives the momentum of the particle in terms of its velocity and mass. Substituting for  $p$  into (2.8) gives:

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x, \quad (2.9)$$

which we recognise as the equation of motion for a harmonic oscillator. Equation (2.6) therefore gives the Hamiltonian for a particle performing simple harmonic motion in one degree of freedom (i.e. in a co-ordinate space with one dimension). We note that this Hamiltonian takes the form:

$$H = T + V, \quad (2.10)$$

where  $T = p^2/2m$  is the kinetic energy of the particle, and  $V = kx^2/2$  is the potential energy. The Hamiltonian in this case, therefore, represents the total energy of the particle. Note that the Hamiltonian is always expressed as a function of the co-ordinates and the momentum of a particle, and *not* as a function of the velocity of the particle.

In an accelerator beam line, we are interested in charged particles moving through electromagnetic fields. If we are to apply Hamiltonian mechanics to the problem, we need an expression for the Hamiltonian for a charged particle in an electromagnetic field. Since we already know that the equations of motion for such a particle are given by (2.1) and (2.2), all we need to do is find a Hamiltonian that, when substituted into Hamilton's equations (2.3) and (2.4), gives the correct equations of motion, i.e. equations of motion consistent with the Lorentz force (2.2) in Newtonian mechanics. In the non-relativistic case, an appropriate Hamiltonian is:

$$H = \frac{(\mathbf{p} - q\mathbf{A})^2}{2m} + q\phi, \quad (2.11)$$

where  $\phi$  is the scalar potential and  $\mathbf{A}$  is the vector potential, which are related to the electric and magnetic fields by the usual equations:

$$\mathbf{E} = -\nabla\phi - \frac{\partial\mathbf{A}}{\partial t}, \quad (2.12)$$

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (2.13)$$

In general,  $\phi$  and  $\mathbf{A}$  are functions of co-ordinates  $x$ ,  $y$  and  $z$ , and time  $t$ . Note that with the Hamiltonian (2.11), the relationship between the particle velocity and momentum, given by (2.3) is:

$$\mathbf{p} = mv + q\mathbf{A}. \quad (2.14)$$

The momentum  $\mathbf{p}$  defined in this way is known as the *canonical momentum*, to emphasise that it is distinct from the mechanical momentum  $mv$ .

Particles in high energy accelerators tend to be moving with relativistic velocities. We therefore need to generalise the Hamiltonian (2.11) to the relativistic case. In special relativity, the total energy  $E$  and momentum  $\mathbf{p}$

for a particle in free space (i.e. with zero electric and magnetic fields) are related by:

$$E^2 = \mathbf{p}^2 c^2 + m^2 c^4, \quad (2.15)$$

where  $c$  is the speed of light in free space. Equation (2.15) follows from the expressions for the energy and momentum in terms of the mass and velocity of the particle:

$$E = \gamma mc^2, \quad (2.16)$$

$$\mathbf{p} = \beta \gamma mc, \quad (2.17)$$

where  $\beta = \mathbf{v}/c$ , and:

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}. \quad (2.18)$$

In an electromagnetic field, we simply include the contributions of the scalar and vector potentials in the same way as we would for a non-relativistic particle, so that the total energy and canonical momentum become:

$$E = \gamma mc^2 + q\phi, \quad (2.19)$$

$$\mathbf{p} = \beta \gamma mc + q\mathbf{A}. \quad (2.20)$$

It then follows from the relationship (2.18) between  $\gamma$  and  $\beta$  that the relationship between the total energy and the canonical momentum can be written:

$$(E - q\phi)^2 = (\mathbf{p} - q\mathbf{A})^2 c^2 + m^2 c^4. \quad (2.21)$$

We saw above that, for a particle performing simple harmonic motion, the Hamiltonian could be expressed as the total energy of the particle. If we assume that the same is true for a relativistic particle moving in an electromagnetic field, then we propose the following form for the Hamiltonian:

$$H = c \sqrt{(\mathbf{p} - q\mathbf{A})^2 + m^2 c^2} + q\phi. \quad (2.22)$$

Whether or not this is the correct Hamiltonian (that is, whether it gives the correct equations of motion when substituted into Hamilton's equations) must ultimately be tested by experiment. It turns out that (2.22) is indeed the correct Hamiltonian for a relativistic charged particle moving in an electromagnetic field.

So far, there is no apparent benefit to working with Hamilton's equations rather than with the Newtonian equations of motion. However, for accelerator beam dynamics, Hamiltonian mechanics provides a useful technique for expressing the equations of motion using the distance along a

particular path as the independent variable, rather than time. This is convenient when considering the trajectory of a particle along a beam line, where the electric and magnetic fields are given as functions of position along the beam line: in general, it is difficult to work out the time at which a particle enters a particular magnet, but we know immediately how far along the beam line it has travelled when it does so. In what follows, we shall first develop the theory required for changing the independent variable (in our case, from time to path length), considering a particle moving in just one degree of freedom. Then, we shall generalise the results to the case of three degrees of freedom, and finally apply the results to the accelerator Hamiltonian.

Consider a particle moving in one degree of freedom, with co-ordinate  $x$  and canonical momentum  $p$ . Since the co-ordinate and momentum are functions of time  $t$ , we can express the Hamiltonian also as a function of the time  $t$ . Then, we can perform the integral, from time  $t_0$  to time  $t_1$ :

$$S = \int_{t_0}^{t_1} p \dot{x} - H dt. \quad (2.23)$$

The quantity  $S$  is called the *action*. Note that the dot above the co-ordinate  $x$  indicates a time derivative, so that  $\dot{x} = v$  is the velocity of the particle. Hamilton's equations lead to particular expressions for the co-ordinate  $x$  and momentum  $p$  as functions of time: thus, over the period from time  $t_0$  to time  $t_1$ , the particle follows a particular line in a plot of  $p$  versus  $x$ . Such a plot is called a *phase space* plot: phase space is simply the space with axes corresponding to the dynamical variables. Let us consider a different path from  $t_0$  to  $t_1$ , but with the same start and end points in phase space (see Fig. 2.1). At each point along the trajectory, we change the co-ordinate by a (small) amount  $\delta x$  and the momentum by a (small) amount  $\delta p$ . As a result of the changes in  $x$  and  $p$ , the value of the Hamiltonian will change by an amount  $\delta H$ . Finally, the action will change by:

$$\delta S = \int_{t_0}^{t_1} \delta p \dot{x} + p \delta \dot{x} - \delta H dt. \quad (2.24)$$

Now, at any given time, the change in the Hamiltonian can be written:

$$\delta H = \frac{\partial H}{\partial x} \delta x + \frac{\partial H}{\partial p} \delta p. \quad (2.25)$$

Using Hamilton's equations (2.3) and (2.4), for small changes in  $x$  and  $p$  (so that we remain close to the trajectory in phase space defined by Hamilton's equations) the change in the Hamiltonian becomes:

$$\delta H = -\dot{p} \delta x + \dot{x} \delta p. \quad (2.26)$$

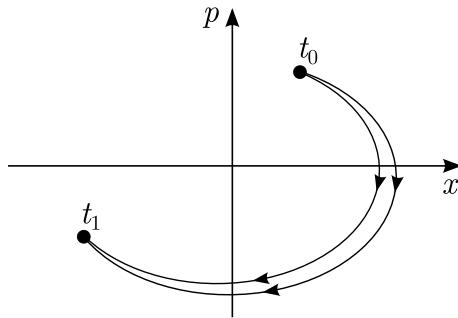


Fig. 2.1 Path integrals in phase space. The dynamical system evolves from some initial condition at time  $t_0$  to a later condition at time  $t_1$ . The actual path through phase space followed by the dynamical system minimises the action (2.23).

Thus, the change in the action is:

$$\delta S = \int_{t_0}^{t_1} p \delta \dot{x} + \dot{p} \delta x dt = \int_{t_0}^{t_1} \frac{d}{dt} (p \delta x) dt = [p \delta x]_{t_0}^{t_1}. \quad (2.27)$$

But, since the start and end points of the path in phase space are fixed, it follows that  $\delta x = 0$  at  $t = t_0$  and  $t = t_1$ ; and hence:

$$\delta S = 0. \quad (2.28)$$

This result expresses the fact that, for a particle obeying Hamilton's equations, the trajectory of a particle in phase space is such that the action is at a minimum (or a maximum) with respect to small changes in the trajectory. We can reverse the argument, and say that Hamilton's equations follow from the fact that the trajectory of a particle in phase space minimises (or maximises) the action (2.23).

Now, instead of expressing the action  $S$  as an integral over the time  $t$ , we can equally well express  $S$  as an integral over the co-ordinate  $x$ :

$$S = \int_{x_0}^{x_1} p \dot{x} \frac{dt}{dx} - H \frac{dt}{dx} dx = \int_{x_0}^{x_1} (-H) \frac{dt}{dx} + p dx, \quad (2.29)$$

where the particle is at  $x = x_0$  at time  $t = t_0$ , and at  $x = x_1$  at time  $t = t_1$ . If we compare (2.29) with (2.23), we conclude that we define the *same path in phase space* if we write equations obtained from Hamilton's equations (2.3) and (2.4) by making the following replacements: the Hamiltonian  $H$  is replaced by  $-p$ , the co-ordinate  $x$  is replaced by  $t$ , the momentum  $p$  is replaced by  $-H$ , and the time  $t$  is replaced by  $x$ . Put another way, in the 'old' variables, the Hamiltonian is  $H$ , the co-ordinate and canonical momentum

are  $(x, p)$ , and the independent variable is  $t$ . In the ‘new’ variables, the Hamiltonian is  $-p$ , the co-ordinate and canonical momentum are  $(t, -H)$ , and the independent variable is  $x$ . Then, Hamilton’s equations (2.3) and (2.4) in the ‘old’ variables become, in the ‘new’ variables:

$$\frac{dt}{dx} = \frac{\partial p}{\partial H}, \quad (2.30)$$

$$\frac{dH}{dx} = -\frac{\partial p}{\partial t}. \quad (2.31)$$

Solving the equations of motion in the ‘new’ variables gives expressions for  $t$  and  $H$  as functions of  $x$ ; the form of the expressions is determined by  $p$ .

As an example, consider again a simple harmonic oscillator, for which the Hamiltonian is:

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2. \quad (2.32)$$

Re-arranging for  $p$ , we obtain the ‘new’ Hamiltonian:

$$p = \sqrt{2mH - kx^2}. \quad (2.33)$$

Substituting into Hamilton’s equations (2.30) and (2.31) expressed in terms of the ‘new’ variables, we obtain the equations of motion:

$$\frac{dt}{dx} = \frac{m}{\sqrt{2mH - kx^2}} = \frac{m}{p}, \quad (2.34)$$

and:

$$\frac{dH}{dx} = 0. \quad (2.35)$$

Equation (2.34) is the familiar relationship between the momentum and the velocity of the particle. Equation (2.35), which follows from the fact that  $p$  has no explicit dependence on the ‘co-ordinate’  $t$ , tells us that the value of  $H$  is conserved during the motion of the particle. Since  $H$  is physically identified with the energy of the particle, we know that this provides a valid solution to the equations of motion.

A particle moving along an accelerator beam line has three degrees of freedom, rather than one. However, we can easily apply the above results to express the equations of motion with the path length along the beam line as the independent variable. In three degrees of freedom, the action is:

$$S = \int_{t_0}^{t_1} p_x \dot{x} + p_y \dot{y} + p_z \dot{z} - H dt. \quad (2.36)$$

For simplicity, consider a beam line with the accelerator components aligned along the  $z$  axis of a Cartesian co-ordinate system. We then replace the

Hamiltonian  $H$  by  $-p_z$ , and the co-ordinate  $z$  by  $t$ . The independent variable is now  $z$ , the distance along the beam line. The other co-ordinate and momentum pairs,  $(x, p_x)$  and  $(y, p_y)$ , are left unchanged. An explicit expression for the new Hamiltonian, which we write  $H_1$ , is obtained by re-arranging (2.22), to give:

$$H_1 = -p_z = -\sqrt{\frac{(E - q\phi)^2}{c^2} - (p_x - qA_x)^2 - (p_y - qA_y)^2 - m^2c^2 - qA_z}, \quad (2.37)$$

where we have replaced the original Hamiltonian  $H$  by the total energy of the particle  $E$ . At this point, we can use Hamilton's equations to write down the equation of motion for a particle moving along a beam line with electric and magnetic fields given by the potentials  $\phi$  and  $\mathbf{A}$  (which are functions of position  $(x, y, z)$ , and, in general, time  $t$ ). Solving the equations of motion will lead to expressions for the co-ordinates and momenta  $(x, p_x)$ ,  $(y, p_y)$  and  $(t, -E)$ , in terms of the independent variable  $z$ , which is the distance along the beam line.

Where the electric and magnetic fields along the beam line are static (that is, independent of time  $t$ ) this approach is particularly convenient. Note that it follows from Hamilton's equation:

$$\frac{d(-E)}{dz} = \frac{\partial H_1}{\partial t}, \quad (2.38)$$

that, if the Hamiltonian is independent of  $t$ , the total energy of the particle is constant along the beam line. Then, it is possible to integrate the equations of motion for the transverse variables  $(x, p_x)$  and  $(y, p_y)$ , knowing only the total energy, and without needing to know the time at which a particle arrives at any particular point along the beam line. Using Newtonian mechanics, however, it would be necessary to integrate the equations of motion for all three co-ordinates (expressing them as functions of time), in order to know the correct values for the electric and magnetic fields at any particular moment. Of course, if the beam line does contain fields that depend on the time (as is the case, for example, for an rf cavity — a common component in accelerators!) then the situation again becomes more complicated; but at least we have localised the problem to one component, or set of components.

## 2.2 Dynamical Variables for Beam Dynamics

Using the Hamiltonian (2.37), and knowing the electromagnetic potentials as functions of position, we can write down the equations of motion for

a particle moving along a straight accelerator beam line. Solving these equations will give the co-ordinates and momentum of the particle (i.e. the dynamical variables) as functions of  $z$ , the distance along the beam line. However, except in very simple cases, the equations of motion are too complex to solve exactly. Usually, we have to make some approximations in order to solve the equations, even using numerical methods. To assist the process, we define new variables to describe the position and momentum of the particle: these variables are designed to remain small as the particle moves along the beam line, and this makes it possible to expand the expressions appearing in the equations of motion (and in their solutions) as power series in these variables. By truncating the power series at some order in the dynamical variables, we can find approximate solutions to the equations of motion. Truncating at low order gives solutions that are less accurate than truncating at higher order; nevertheless, even first-order (i.e. linear) approximations are useful in beam dynamics.

In this section, we shall continue to consider a straight beam line: in practice, this means a beam line without dipole magnets. In Section 2.3, we shall extend our results to include beam lines with dipole magnets.

Our goal is to define a set of new variables for describing the particle position and momentum, suitable for making approximations to the equations of motion. As we shall see in later chapters, Hamiltonian mechanics provides some very important and useful techniques in beam dynamics; we shall therefore be careful to choose new variables that obey Hamilton's equations, for an appropriate Hamiltonian. If the variables used to describe a dynamical system evolve according to Hamilton's equations, then the variables are known as *canonical variables*. A transformation from one set of canonical variables to another set of canonical variables is known as a canonical transformation. We shall take care to ensure that the new variables we define are related to the variables we have used up to now (and that obey Hamilton's equations) by a canonical transformation.

The first step is to apply a scaling to the Hamiltonian itself. That is, we define a new Hamiltonian:

$$\tilde{H} = \frac{H_1}{P_0}, \quad (2.39)$$

where  $P_0$  is a fixed *reference momentum*. Although the reference momentum can, in principle, be chosen arbitrarily, we shall see that a sensible choice is to use the momentum that particles in the beam line are intended to have by design. Then, an ideal particle will have longitudinal momentum  $p_z = P_0$ , and then from (2.37) we can say that, for an ideal particle,

$\tilde{H} = -1$ . Inspecting (2.3) and (2.4), we can see that if we also scale the momenta Hamilton's equations will still be valid for the new Hamiltonian  $\tilde{H}$ :

$$\tilde{p}_x = \frac{p_x}{P_0}, \quad \tilde{p}_y = \frac{p_y}{P_0} \quad \text{and} \quad (-\tilde{E}) = \frac{(-E)}{P_0}. \quad (2.40)$$

In terms of the new momenta, the Hamiltonian is:

$$\tilde{H} = -\sqrt{\frac{(\tilde{E} - q\phi/P_0)^2}{c^2} - (\tilde{p}_x - a_x)^2 - (\tilde{p}_y - a_y)^2 - \frac{m^2 c^2}{P_0^2} - a_z}, \quad (2.41)$$

where we have defined the scaled vector potential:

$$\mathbf{a} = \frac{q}{P_0} \mathbf{A}. \quad (2.42)$$

The next step is to define new longitudinal variables. The problem with the current variables  $(t, -E)$  is that the ‘co-ordinate’ (actually the time  $t$ ) increases with distance down the beam line. In beam dynamics, we are often mostly concerned with the relative distance between two particles in the beam: this distance can be smaller than the length of the beam line by many orders of magnitude. Therefore, if we try to calculate the relative position of two particles after tracking them down a long beam line, we need to take the difference of two large numbers that are almost equal. In practice, it can be very difficult to maintain good accuracy in the calculations using this procedure. Instead, we shall define a new longitudinal co-ordinate that has the physical significance of a distance relative to a nominal particle that moves down the beam line with the reference momentum  $P_0$ . If all particles in the beam have momentum close to  $P_0$ , then we should expect that the new longitudinal co-ordinate should remain conveniently small.

To ensure that our new longitudinal variables are canonical, we define the transformation from the old variables to the new variables in terms of a generating function. There are various types of generating functions; in this case, we shall use a mixed-variable generating function of the second kind, which in general is expressed as a function of the old co-ordinates  $x_i$ , the new momenta  $P_i$ , and the independent variable  $z$ :

$$F_2 = F_2(x_i, P_i; z). \quad (2.43)$$

The generating function defines the relationships between the old and new variables, which are written as:

$$\tilde{p}_i = \frac{\partial F_2}{\partial x_i}, \quad X_i = \frac{\partial F_2}{\partial P_i} \quad \text{and} \quad K = \tilde{H} + \frac{\partial F_2}{\partial z}. \quad (2.44)$$

Defining the new variables in this way ensures that they obey Hamilton's equations, with the Hamiltonian  $K$ . We use the particular choice of generating function:

$$F_2(x, P_X, y, P_Y, t, P_Z; z) = xP_X + yP_Y + \left( \frac{z}{\beta_0} - ct \right) \left( \frac{1}{\beta_0} + P_Z \right), \quad (2.45)$$

where  $\beta_0$  is a constant. The old variables are  $(x, \tilde{p}_x)$ ,  $(y, \tilde{p}_y)$  and  $(t, -\tilde{E})$ ; the new variables are  $(X, P_X)$ ,  $(Y, P_Y)$  and  $(Z, P_Z)$ . Applying the transformations (2.44), we find that the transverse variables are simply related by the identity transformation:

$$X = x, \quad P_X = \tilde{p}_x, \quad Y = y, \quad P_Y = \tilde{p}_y. \quad (2.46)$$

The old and new longitudinal variables are related by:

$$(-\tilde{E}) = \frac{\partial F_2}{\partial t} = -c \left( \frac{1}{\beta_0} + P_Z \right), \quad (2.47)$$

and:

$$Z = \frac{\partial F_2}{\partial P_Z} = \frac{z}{\beta_0} - ct. \quad (2.48)$$

The new Hamiltonian is:

$$K = -\sqrt{\left( P_Z + \frac{1}{\beta_0} - \frac{q\phi}{cP_0} \right)^2 - (P_X - a_x)^2 - (P_Y - a_y)^2 - \frac{m^2 c^2}{P_0^2}} \\ - a_z + \frac{1}{\beta_0^2} + \frac{P_Z}{\beta_0}. \quad (2.49)$$

At this point, it is worth pausing to consider the physical interpretation of the dynamical variables. The transverse variables are straightforward:  $X$  and  $Y$  are simply the horizontal and vertical position of the particle in a plane perpendicular to the  $z$  axis.  $P_X$  and  $P_Y$  are the components of the canonical momentum of the particle, in the  $X$  and  $Y$  directions, respectively. The longitudinal co-ordinate is:

$$Z = \frac{z}{\beta_0} - ct. \quad (2.50)$$

Consider a particle moving along the  $z$  axis, with the reference momentum  $P_0$ ; we call this the *reference particle*. Let us choose  $\beta_0$  so that  $P_0 = \beta_0 \gamma_0 mc$ , where  $\gamma_0 = 1/\sqrt{1 - \beta_0^2}$ ; i.e.  $\beta_0$  is the velocity of the reference particle, scaled by the speed of light. Since  $t$  is the time at which the reference particle arrives at the position  $z$  along the beam line, we see immediately that for the reference particle  $Z = 0$  for all  $z$ . A particle arriving at a particular location at an earlier time than the reference particle

has  $Z > 0$ , and a particle arriving later than the reference particle has  $Z < 0$ . In fact, the  $Z$  co-ordinate of a particular particle is *approximately* the longitudinal distance between that particle and the reference particle.

The longitudinal momentum is:

$$P_Z = \frac{\tilde{E}}{c} - \frac{1}{\beta_0} = \frac{E}{cP_0} - \frac{1}{\beta_0}. \quad (2.51)$$

For the reference particle, the energy (in a region where the scalar potential is zero) is  $E = \gamma_0 mc^2$ , and the momentum is  $P_0 = \beta_0 \gamma_0 mc$ ; hence, for the reference particle,  $P_Z = 0$ . Particles with energy higher than the reference particle have  $P_Z > 0$ ; and particles with energy lower than the reference particle have  $P_Z < 0$ . In fact, the canonical momentum  $P_Z$  of a particle is *approximately* the energy of the particle relative to the reference particle. Conventionally,  $P_Z$  is referred to as the *energy deviation*.

Finally, after the above sequence of canonical transformations, we have ended up with a rather unconventional notation for the dynamical variables used in beam dynamics. We therefore define a new notation for the variables, as follows. The horizontal and vertical transverse co-ordinates are  $x$  and  $y$ , respectively (which are in any case identical to  $X$  and  $Y$ ). The corresponding momenta  $P_X$  and  $P_Y$  we rewrite as  $p_x$  and  $p_y$ , which are expressed in terms of the physical properties (mass, velocity, charge and vector potential) as follows:

$$p_x = \frac{\beta_x \gamma mc + qA_x}{P_0}, \quad (2.52)$$

$$p_y = \frac{\beta_y \gamma mc + qA_y}{P_0}. \quad (2.53)$$

The independent variable  $z$ , which is the distance along the beam line, we rewrite as  $s$ . The longitudinal co-ordinate we rewrite (following the conventional notation) as  $z$ , so that now:

$$z = \frac{s}{\beta_0} - ct. \quad (2.54)$$

The energy deviation is written as  $\delta$ :

$$\delta = \frac{E}{cP_0} - \frac{1}{\beta_0}. \quad (2.55)$$

And, finally, we note that the constant term  $1/\beta_0^2$  in the Hamiltonian (2.49) has no significance for the dynamics; so we simply drop this term, and write the Hamiltonian:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0} - \frac{q\phi}{cP_0}\right)^2 - (p_x - a_x)^2 - (p_y - a_y)^2 - \frac{1}{\beta_0^2 \gamma_0^2} - a_z}. \quad (2.56)$$

Note that we have used  $mc/P_0 = 1/\beta_0\gamma_0$ . The equations of motion for a particle in the beam line are found from Hamilton's equations (2.3) and (2.4), with the variables  $(x, p_x)$ ,  $(y, p_y)$  and  $(z, \delta)$ , and the Hamiltonian (2.56). Understanding the particle dynamics involves solving these equations for given cases of the scalar potential  $\phi$  and vector potential  $\mathbf{A}$ . We shall investigate solutions for a few common cases in Chapter 3. Before we do so, however, we should recall that the independent variable  $s$  represents the distance along a straight line: accelerator beam lines generally include dipole magnets, which lead to a significant curvature of the particle trajectories. A Cartesian co-ordinate system is not appropriate for such cases. Therefore, before proceeding to find solutions for the equations of motion, we discuss, in the following section, how we can generalise the expression for the Hamiltonian to include a curvature of the trajectory along which we measure the distance  $s$ .

### 2.3 The Hamiltonian in a Curved Co-ordinate System

In the previous section, we derived an expression (2.56) for the Hamiltonian describing the dynamics of a particle moving along an accelerator beam line, where the independent variable  $s$  represents the distance along a straight path. To describe accurately the motion in a dipole magnet, it is more appropriate to measure the distance  $s$  along a path that curves with the trajectory of a particle in the dipole: the co-ordinates  $x$  and  $y$  will also be measured with respect to the curved path. In principle, we can define any curved path in space as the origin of a local  $x$ - $y$  co-ordinate system: we refer to the chosen path as the *reference trajectory*. Note that the reference trajectory does not have to be the physical trajectory followed by any particle: however, it is usually convenient to choose the reference trajectory as the ideal ‘design’ trajectory followed by a charged particle as it moves along the beam line. In that case, the beam line will usually have the property that particles remain close to the reference trajectory, even if they do not follow it exactly.

For simplicity, we shall consider only a curvature of the reference trajectory in the horizontal plane. We begin with the Hamiltonian for a relativistic charged particle, using Cartesian variables, with the time  $t$  as the independent variable. The Hamiltonian is given by (2.22):

$$H = c\sqrt{(\mathbf{P} - q\mathbf{A})^2 + m^2c^2} + q\phi. \quad (2.57)$$

We denote the Cartesian co-ordinates  $X$ ,  $Y$  and  $Z$ , and the canonical

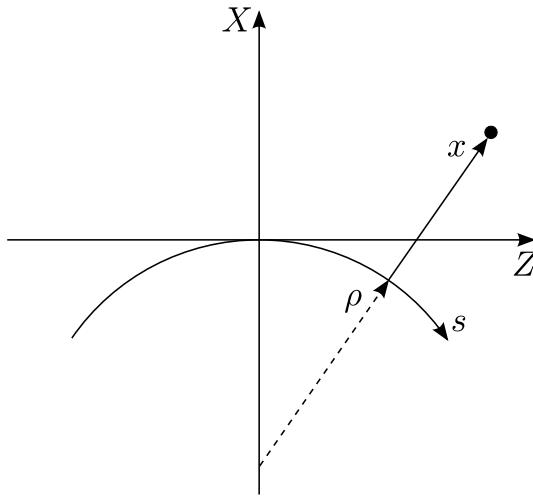


Fig. 2.2 Co-ordinate system with a curved reference trajectory.

momenta  $P_X$ ,  $P_Y$  and  $P_Z$ . Now let us consider a change of variables, so that instead of Cartesian co-ordinates  $(X, Y, Z)$ , we use the set of co-ordinates  $(x, y, s)$ . The co-ordinate  $s$  describes the distance along a path tangent to the  $Z$  axis at  $Z = 0$ , but with radius of curvature  $\rho$  in the  $X-Z$  plane. The co-ordinates  $x$  and  $y$  are, respectively, the horizontal and vertical co-ordinates in a plane perpendicular to the  $s$  axis at any point along this axis: the co-ordinate system is illustrated in Fig. 2.2. The relationship between the Cartesian co-ordinates  $(X, Y, Z)$  and the ‘curved’ co-ordinates  $(x, y, s)$  are:

$$X = (x + \rho) \cos\left(\frac{s}{\rho}\right) - \rho, \quad (2.58)$$

$$Y = y, \quad (2.59)$$

$$Z = (x + \rho) \sin\left(\frac{s}{\rho}\right). \quad (2.60)$$

The new dynamical variables consist of the position of a particle expressed in the curved co-ordinates and the corresponding momenta. To ensure that the new variables are canonical (i.e. that they evolve according to Hamilton’s equations), we look for a canonical transformation from the old variables (in the Cartesian co-ordinate system) to the new variables (in the curved co-ordinate system). We can express an appropriate transformation

in terms of a mixed-variable generating function of the third kind:

$$F_3 = F_3(x_i, P_i; t), \quad (2.61)$$

where  $P_i$  are the old momenta, and  $x_i$  are the new co-ordinates. The equations relating the old and new variables are:

$$X_i = -\frac{\partial F_3}{\partial P_i}, \quad (2.62)$$

$$p_i = -\frac{\partial F_3}{\partial x_i}, \quad (2.63)$$

and the new Hamiltonian is:

$$K = H + \frac{\partial F_3}{\partial t}. \quad (2.64)$$

A generating function giving the transformation of the co-ordinates (2.58), (2.59) and (2.60) is:

$$F_3 = -\left( (x + \rho) \cos\left(\frac{s}{\rho}\right) - \rho \right) P_X - y P_y - (x + \rho) \sin\left(\frac{s}{\rho}\right) P_Z. \quad (2.65)$$

From this, we derive the following expressions for the new momenta:

$$p_x = P_X \cos\left(\frac{s}{\rho}\right) + P_Z \sin\left(\frac{s}{\rho}\right), \quad (2.66)$$

$$p_y = P_Y, \quad (2.67)$$

$$p_s = P_Z \left( 1 + \frac{x}{\rho} \right) \cos\left(\frac{s}{\rho}\right) - P_X \left( 1 + \frac{x}{\rho} \right) \sin\left(\frac{s}{\rho}\right). \quad (2.68)$$

Since  $F_3$  has no explicit dependence on  $t$ , the new Hamiltonian is the same as the old Hamiltonian. However, when changing from the old to the new variables, we also need to express the components of the vector potential as components along the new co-ordinate axes:

$$A_x = A_X \cos\left(\frac{s}{\rho}\right) + A_Z \sin\left(\frac{s}{\rho}\right), \quad (2.69)$$

$$A_y = A_Y, \quad (2.70)$$

$$A_s = A_Z \cos\left(\frac{s}{\rho}\right) - A_X \sin\left(\frac{s}{\rho}\right). \quad (2.71)$$

Substituting the new momenta and the new components of the vector potential for the old momenta and old components of the vector potential in the Hamiltonian (2.57) gives:

---


$$K = c \sqrt{\left( \frac{p_s}{1 + h_x} - q A_s \right)^2 + (p_x - q A_x)^2 + (p_y - q A_y)^2 + m^2 c^2 + q \phi}, \quad (2.72)$$

where we have defined the curvature of the reference trajectory  $h = 1/\rho$ . Perhaps surprisingly, the new Hamiltonian has a very similar form to the old Hamiltonian: the only significant difference is a factor  $(1+hX)$  that divides  $p_s$ . Since the form of the Hamiltonian is so similar, the subsequent steps (changing the independent variable from time  $t$  to path length  $s$ ; scaling by the reference momentum; and changing the longitudinal variables) follow exactly the same lines as in Section 2.2, where we considered only a straight reference trajectory. Therefore, we simply quote the result for the final Hamiltonian:

$$H = \frac{\delta}{\beta_0} - (1 + hx) \sqrt{\left(\delta + \frac{1}{\beta_0} - \frac{q\phi}{cP_0}\right)^2 - (p_x - a_x)^2 - (p_y - a_y)^2 - \frac{1}{\beta_0^2 \gamma_0^2}} \\ - (1 + hx) a_s. \quad (2.73)$$

In the limit that the radius of curvature becomes infinite, the reference trajectory becomes a straight line; in this limit,  $h \rightarrow 0$ , and the Hamiltonian (2.73) then becomes identical to the Hamiltonian (2.56) that we derived in that case.

Some care is needed in the interpretation of the variables in the Hamiltonian (2.73). In particular,  $x$  is the horizontal co-ordinate in a plane perpendicular to the reference trajectory at any point, and  $a_x$  is the component of the vector potential, multiplied by  $q/P_0$ , in the direction of increasing  $x$ . Similarly, the independent variable  $s$  is the distance along the reference trajectory, and  $a_s$  is the component of the vector potential in the direction of increasing  $s$ . The momenta  $p_x$  and  $p_y$  are given by (2.52) and (2.53):

$$p_x = \frac{\gamma m \dot{x} + qA_x}{P_0}, \quad (2.74)$$

$$p_y = \frac{\gamma m \dot{y} + qA_y}{P_0}, \quad (2.75)$$

where  $\dot{x}$  and  $\dot{y}$  are the derivatives of  $x$  and  $y$  with respect to time,  $t$ . The longitudinal co-ordinate  $z$  and its conjugate momentum  $\delta$  (the energy deviation) are given by (2.54) and (2.55).

In practice, accelerator beam lines are often modelled as a sequence of (more or less) standard components, such as solenoids, dipole magnets, quadrupole magnets, etc. Dipole magnets are the main components that have significant impact on the direction in which the beam is travelling; therefore, we only usually need to consider curved reference trajectories inside dipole magnets. The reference trajectory is built up from curved sections (inside dipoles) joined by straight lines (around which are positioned

quadrupoles, solenoids, etc.). In the following chapters, we shall first see how to solve the equations of motion, with some approximations, in individual standard components such as solenoids, dipoles and quadrupoles; and then we shall discuss how to describe the overall dynamics in a beam line consisting of sequences of these components.

## 2.4 Symplectic Transfer Maps and Liouville's Theorem

Dynamical systems that obey Hamilton's equations have an important property that we shall use frequently in our discussion of accelerator beam dynamics: the transfer maps from one point (defined by a given value of the independent variable) to another are *symplectic*. A transfer map is simply a statement of the relationship between the dynamical variables for a particle at different points. In the case of an accelerator beam line, we can write:

$$\vec{x}(s_1) = \vec{M}(\vec{x}(s_0)), \quad (2.76)$$

where  $\vec{x} = (x, p_x, y, p_y, z, \delta)$  is a vector whose components are the dynamical variables, and  $\vec{M}$  (the transfer map) is a vector whose six components are functions of the dynamical variables. Note that we use the notation  $\vec{x}$  for vectors in phase space (vectors with six components in three degrees of freedom), and the notation  $\mathbf{x}$  for vectors in co-ordinate space (vectors with three components in three degrees of freedom).  $\vec{x}(s_0)$  denotes the values of the dynamical variables for a particular particle at  $s = s_0$ . The transfer map  $\vec{M}$  is found by solving Hamilton's equations for a particle moving from  $s_0$  to  $s_1$ . We shall now show an important mathematical property of the transfer map follows from the fact that the system obeys Hamilton's equations.

We denote the components of the vector  $\vec{x}$  by  $x_i$ , where in  $n$  degrees of freedom,  $i = 1 \cdots 2n$ . Then we can write Hamilton's equations in the form:

$$\dot{x}_i = S \frac{\partial H}{\partial x_i}, \quad (2.77)$$

where the dot indicates the derivative with respect to the independent variable, and  $S$  is a block-diagonal  $2n \times 2n$  antisymmetric matrix, with each of the  $n$  block diagonals given by:

$$S_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (2.78)$$

Consider an infinitesimal change in the independent variable from  $s = s_0$  to  $s = s_0 + \delta s$ . The values of the dynamical variables change according to:

$$x_i(s_0 + \delta s) = x_i(s_0) + \dot{x}_i(s_0) \delta s = x_i(s_0) + S_{ik} \frac{\partial H}{\partial x_k} \Big|_{s=s_0} \delta s. \quad (2.79)$$

The Jacobian of the transformation from  $s_0$  to  $s_0 + \delta s$  is given by:

$$J_{ij} = \frac{\partial x_i(s_0 + \delta s)}{\partial x_j(s_0)} = \delta_{ij} + S_{ik} \frac{\partial^2 H}{\partial x_j \partial x_k} \Big|_{s=s_0} \delta s, \quad (2.80)$$

where  $\delta_{ij}$  is the Kronecker delta ( $\delta_{ij} = 1$  if  $i = j$ , and  $\delta_{ij} = 0$  if  $i \neq j$ ). We can write (2.80) in matrix form:

$$J = I + S \tilde{H} \delta s, \quad (2.81)$$

where  $\tilde{H}$  is the symmetric  $2n \times 2n$  matrix:

$$\tilde{H}_{jk} = \frac{\partial^2 H}{\partial x_j \partial x_k} \Big|_{s=s_0}. \quad (2.82)$$

Using the fact that  $S^2 = -I$ , we find that to first order in  $\delta s$  (i.e. for an infinitesimal transformation):

$$J^T S J = S, \quad (2.83)$$

where  $J^T$  is the transpose of the Jacobian  $J$ . A transformation with Jacobian satisfying (2.83) is said to be *symplectic*.

Now consider a second infinitesimal transformation, from  $s_0 + \delta s$  to  $s_0 + 2\delta s$ . The Jacobian for the total transformation from  $s_0$  to  $s_0 + 2\delta s$  can be written:

$$\frac{\partial x_i(s_0 + 2\delta s)}{\partial x_j(s_0)} = \sum_{k=1}^{2n} \frac{\partial x_i(s_0 + 2\delta s)}{\partial x_k(s_0 + \delta s)} \frac{\partial x_k(s_0 + \delta s)}{\partial x_j(s_0)}. \quad (2.84)$$

In matrix form:

$$J_2 = J_1 J_0, \quad (2.85)$$

where  $J_0$  is the Jacobian of the transformation from  $s_0$  to  $s_0 + \delta s$ ,  $J_1$  is the Jacobian of the transformation from  $s_0 + \delta s$  to  $s_0 + 2\delta s$ , and  $J_2$  is the Jacobian of the transformation from  $s_0$  to  $s_0 + 2\delta s$ . We know that if the system obeys Hamilton's equations, then  $J_0$  and  $J_1$  will satisfy (2.83). But it is easy to show that if  $J_0$  and  $J_1$  each satisfy (2.83), then the matrix product  $J_2 = J_0 J_1$  will also satisfy (2.83). Hence, a transformation constructed from a sequence of symplectic transformations will be symplectic. This means that a transformation representing an evolution of the system from any point  $s = s_0$  to any other point  $s = s_1$  will be symplectic, even

if it is not infinitesimal, since it can be constructed from a sequence of infinitesimal transformations.

Note that we have made no assumption about the linearity of the transformations. In general, the Jacobian will be a function of the dynamical variables. However, in the special case of a linear transformation, where the components of the vector  $\vec{M}$  are polynomials in the dynamical variables with zeroth-order and first-order terms, then the Jacobian will be a matrix with constant components. A matrix  $J$  satisfying the symplectic condition (2.83) is said to be a *symplectic matrix*. The first-order part of a linear transfer map can be represented by a symplectic matrix.

One reason that symplecticity is important is that symplectic transformations have conserved quantities associated with them. An example is the density of particles in phase space. To show this, we first note that we can consider a symplectic transformation  $\vec{M}$  as defining a new set of dynamical variables:

$$\vec{X} = \vec{M}(\vec{x}). \quad (2.86)$$

In the original variables  $\vec{x}$ , a volume element  $dv$  of phase space is:

$$dv = \prod_{i=1}^{2n} dx_i. \quad (2.87)$$

In the new variables  $\vec{X}$ , the volume element is:

$$dV = \prod_{i=1}^{2n} dX_i = |\det J| \prod_{i=1}^{2n} dx_i, \quad (2.88)$$

where  $|\det J|$  is the modulus of the determinant of the Jacobian of the transformation from the old to the new variables. Note that  $dV$  is the volume of the region obtained by transforming every point within  $dv$  by the transfer map  $\vec{M}$ . Now, using (2.83), we can write:

$$\det(J^T SJ) = (\det J)^2 \det S = \det S, \quad (2.89)$$

and since  $\det S \neq 0$  (in fact,  $\det S = 1$ ), it follows immediately that:

$$\det J = \pm 1, \quad (2.90)$$

and so:

$$|\det J| = 1. \quad (2.91)$$

We note in passing that although we have shown here that the determinant of a symplectic matrix has to be  $\pm 1$ , it can, in fact, be shown that the

determinant of a symplectic matrix must be equal to 1; but the proof of the stronger statement is not straightforward, and is not required here. It follows from (2.87), (2.88) and (2.91) that:

$$dV = dv. \quad (2.92)$$

In other words, the volume of any given region in phase space is conserved under a symplectic transformation. The density of particles in phase space is simply equal to the number of particles in a given region divided by the volume of the region. Since both the number of particles and the volume occupied by the particles is conserved when the motion of the particles obeys Hamilton's equations, we can see immediately that the density of particles in phase space is conserved in a system obeying Hamilton's equations: this is illustrated in Fig. 2.3. The conservation of the density of particles in phase space, for particles obeying Hamilton's equations, is known as Liouville's theorem, and can be written mathematically as:

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \sum_{i=1}^{2n} \frac{dx_i}{dt} \frac{\partial\rho}{\partial x_i} = 0. \quad (2.93)$$

In accelerator beam dynamics, Liouville's theorem has an important consequences for the behaviour of particles. In particular, the phase space volume of a bunch of particles is conserved under symplectic transport.

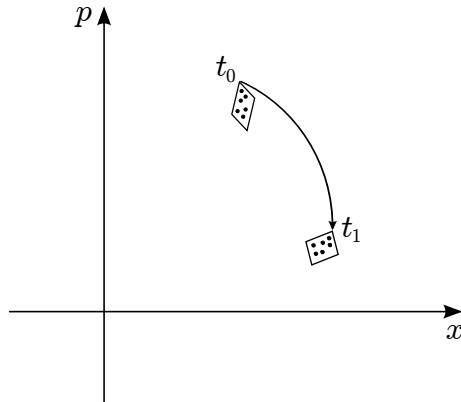


Fig. 2.3 Liouville's theorem: for particles obeying Hamilton's equations, the density of particles in phase space is conserved as the system evolves. The infinitesimal region within the quadrilateral shown at  $t_0$  evolves to the region within the quadrilateral at  $t_1$ ; although the shape of the region may change, its area and the number of particles it contains remain constant.

Consider, for example, a bunch that is focused to a very small spot at some point (the focal point) along a beam line. As the focusing is increased to reduce the size of the spot in co-ordinate space, the divergence will increase so that the particles in the bunch occupy a constant volume in phase space. This means that the smaller the dimensions of the bunch at the focal point, the more rapidly the size of the bunch will increase with distance from the focal point. If we wish to reduce the overall phase space volume of a bunch of particles, then we need to use non-symplectic processes, often known as *beam cooling* methods. Examples include synchrotron radiation damping (discussed in Chapter 7) or stochastic cooling. There are also non-symplectic processes, such as intrabeam scattering, that can lead to an overall increase in the phase space volume of a bunch.

Finally, it is worth noting that although phase space volumes are conserved in all symplectic systems, at least for linear transformations in more than one degree of freedom, symplecticity implies the existence of additional conserved quantities. As we shall see in Chapter 5, linear symplectic transformations in  $n$  degrees of freedom have associated with them  $n$  conserved quantities (the product of which corresponds to the phase space volume). In beam dynamics, these quantities are called the *emittances*, and are of considerable importance in describing the properties of bunches of particles. The significance of the various conserved quantities is one of the main motivations for developing beam dynamics in the framework of Hamiltonian mechanics, and the reason why we place so much emphasis, in the following chapters, on symplectic maps for describing the dynamics of particles in accelerator beam lines.

PART II

## Single-Particle Linear Dynamics

**This page intentionally left blank**

## Chapter 3

# Linear Transfer Maps for Common Components

Substituting the electromagnetic potentials in a particular region of space into the Hamiltonian (2.73) allows us to write down the equations of motion for a charged particle moving through that region. Our goal in this chapter is to find solutions to the equations of motion for particular components that are commonly used in accelerator beam lines, making certain approximations necessary to obtain solutions in a convenient form. Having established the solutions, we can (starting in Chapter 4) investigate the dynamical behaviour of particles in beam lines consisting of sequences of components.

For now, we shall restrict the discussion to accelerator components within which the particle dynamics can be represented, to a good approximation, by linear equations. Such components include drift spaces (i.e. regions with no electromagnetic field), dipole magnets, quadrupole magnets and solenoids. Many important aspects of beam dynamics in accelerators can be understood in terms of the properties of these components, using a linear approximation. However, it should be remembered that the equations of motion in all accelerator components (even drift spaces!) are, in reality, nonlinear; at least when expressed in terms of the conventional dynamical variables using the Hamiltonian (2.73). For many components (including, for example, higher-order multipoles such as sextupoles and octupoles) the dominant effects cannot be represented with a linear approximation. We shall discuss the nonlinear dynamics in accelerators starting in Chapter 8.

Linear approximations to nonlinear dynamical systems can be made in a number of different ways. Our consistent approach here is to make an approximation to the Hamiltonian; specifically, we shall assume that we can expand the Hamiltonian as a power series to second order in the dynamical variables. Substituting a second-order Hamiltonian into Hamilton's

equations (2.3) and (2.4) produces linear equations of motion. The advantage of this approach is that although we have made an approximation to the system, we are still describing the dynamics using a Hamiltonian. The solution therefore retains properties (in particular, the existence of conserved quantities) that prove to be very useful when investigating particle dynamics in accelerator beam lines.

There are various ways in which the approximations can be made to the Hamiltonian in order to derive equations that can be solved exactly. The approximation that is appropriate for any given case will depend on the exact form of the Hamiltonian. Consider, for example, the special case where the scalar potential  $\phi$  vanishes, and the magnetic field can be described by a vector potential with  $A_x = A_y = 0$ . Then, the Hamiltonian (2.73) becomes:

$$H = \frac{\delta}{\beta_0} - (1 + hx) \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}} - (1 + hx) a_s. \quad (3.1)$$

If  $a_s$  is independent of the dynamical variable  $z$ , the Hamiltonian (3.1) has no explicit dependence on  $z$ , and then the energy deviation  $\delta$  will be constant. Expanding to second order in the transverse momenta  $p_x$  and  $p_y$  gives:

$$H = \frac{\delta}{\beta_0} + (1 + hx) \left( \frac{p_x^2}{2D} + \frac{p_y^2}{2D} - D - a_s \right) + O(4), \quad (3.2)$$

where:

$$D = \sqrt{1 + \frac{2\delta}{\beta_0} + \delta^2}, \quad (3.3)$$

and  $O(4)$  represents terms of fourth order and above in the transverse momenta. The approximation is valid for:

$$\frac{p_x^2 + p_y^2}{2D^2} \ll 1. \quad (3.4)$$

This is the case for particles with trajectories nearly parallel to the reference trajectory. Truncation of a power series expansion of the Hamiltonian at low order in  $p_x$  and  $p_y$  is known as the *paraxial approximation*. With further restrictions on some of the parameters, it is sometimes possible to use the paraxial approximation to construct a Hamiltonian for which the equations of motion can be solved analytically. Even if the paraxial approximation on its own is not sufficient to produce equations of motion with analytical

solutions, the paraxial approximation often provides a useful starting point for techniques that lead to analytical (and symplectic) solutions by making some further, carefully chosen approximations. Some of these techniques will be discussed in Chapters 9 and 10, where we consider nonlinear dynamics. In this chapter, we shall apply paraxial approximation to study beam dynamics in accelerator components for which the dominant effects can be described by linear equations of motion.

### 3.1 Drift Space

For the first case, let us consider the simple situation of a region of space where there are no electric or magnetic fields present at all, i.e. a drift space. With an appropriate choice of gauge, the potentials  $\phi$  and  $\mathbf{A}$  can be taken to be both zero. Furthermore, we can use a straight reference trajectory (curvature  $h = 0$ ). Then, the Hamiltonian (2.73) becomes:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}}. \quad (3.5)$$

The equations of motion derived from this Hamiltonian can be solved exactly. Hamilton's equations (2.3) and (2.4) give the equations of motion. Firstly for the horizontal variables:

$$\frac{dp_x}{ds} = -\frac{\partial H}{\partial x} = 0, \quad (3.6)$$

$$\frac{dx}{ds} = \frac{\partial H}{\partial p_x} = \frac{p_x}{d}, \quad (3.7)$$

where we define:

$$d = \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}}. \quad (3.8)$$

The equations of motion for the vertical variables are similar:

$$\frac{dp_y}{ds} = -\frac{\partial H}{\partial y} = 0, \quad (3.9)$$

$$\frac{dy}{ds} = \frac{\partial H}{\partial p_y} = \frac{p_y}{d}. \quad (3.10)$$

Finally, for the longitudinal variables:

$$\frac{d\delta}{ds} = -\frac{\partial H}{\partial z} = 0, \quad (3.11)$$

$$\frac{dz}{ds} = \frac{\partial H}{\partial \delta} = \frac{1}{\beta_0} \left(1 - \frac{1}{d}\right) - \frac{\delta}{d}. \quad (3.12)$$

Since the canonical momenta  $p_x$ ,  $p_y$  and  $\delta$  are all constant,  $d$  is also constant. The equations of motion can be solved exactly without the need for any approximations. The solutions are:

$$x_1 = x_0 + \frac{p_{x0}}{d} L, \quad (3.13)$$

$$p_{x1} = p_{x0}, \quad (3.14)$$

$$y_1 = y_0 + \frac{p_{y0}}{d} L, \quad (3.15)$$

$$p_{y1} = p_{y0}, \quad (3.16)$$

$$z_1 = z_0 + \frac{L}{\beta_0} \left(1 - \frac{1}{d}\right) - \frac{\delta_0}{d} L, \quad (3.17)$$

$$\delta_1 = \delta_0, \quad (3.18)$$

where the subscript 0 denotes the value of the variable at the entrance to the drift ( $s = s_0$ ), and the subscript 1 denotes the value at the exit of the drift ( $s = s_1 = s_0 + L$ ). The equations (3.13)–(3.18), relating the final values of the variables to the initial values of the variables, constitute the *transfer map* for the drift space.

The above transfer map is an exact solution to the equations of motion for a particle in a drift space. In nearly all other cases of accelerator component, we will have to make some approximations in order to obtain a convenient form for the solution to the equations of motion. Note that the transfer map for the drift space is nonlinear: the final values for the co-ordinates have a nonlinear dependence on the values of the canonical momenta. However, in practice, the nonlinear effects are usually rather small, and it is possible and often convenient to make a linear approximation to the transfer map. This can be done in two ways. Perhaps the most obvious approach is to expand the equations (3.13)–(3.18) as power series to first order in the dynamical variables. The result is:

$$x_1 = x_0 + Lp_{x0} + O(2), \quad (3.19)$$

$$p_{x1} = p_{x0}, \quad (3.20)$$

$$y_1 = y_0 + Lp_{y0} + O(2), \quad (3.21)$$

$$p_{y1} = p_{y0}, \quad (3.22)$$

$$z_1 = z_0 + \frac{L}{\beta_0^2 \gamma_0^2} \delta_0 + O(2), \quad (3.23)$$

$$\delta_1 = \delta_0, \quad (3.24)$$

where  $O(2)$  represents terms of second order and higher in the dynamical variables. An alternative approach is to make the paraxial approximation,

expanding the Hamiltonian (3.5) to second order in the dynamical variables:

$$H = -1 + \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2} + O(3). \quad (3.25)$$

Simply dropping the terms of third order and higher in the dynamical variables (and also dropping constant, or zeroth-order, terms, which have no significance for the dynamics) gives the second-order Hamiltonian:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2}. \quad (3.26)$$

We then construct the linear transfer map from Hamilton's equations, using the expression (3.26) for the Hamiltonian. In this case, we obtain the same transfer map, given by (3.19)–(3.24), as we obtained by making a linear approximation directly to the solutions of the equations of motion for the full Hamiltonian. However, there is an important reason for preferring the second approach (approximating the Hamiltonian, rather than the solutions to the equations of motion) in general. Systems that evolve according to Hamilton's equations have useful properties, in particular the existence of conserved quantities, that can be lost by making approximations to the solutions of the equations of motion: the approximate solutions often cannot be derived from any Hamiltonian (except in special cases, such as a linear approximation to the drift space). The conserved quantities associated with Hamiltonian systems are of considerable importance in beam dynamics: we shall discuss them in detail later. For now, we simply note that where it is necessary to make an approximation in order to obtain a convenient solution to the equations of motion, it is preferable to make the approximation to the Hamiltonian, rather than directly to the solutions themselves.

Finally, we observe that it is possible to express the linear transfer map in matrix form:

$$\vec{x}_1 = R_{\text{drift}} \vec{x}_0, \quad (3.27)$$

where  $\vec{x}$  is a phase space vector (with six components) constructed from the dynamical variables:

$$\vec{x} = \begin{pmatrix} x \\ p_x \\ y \\ p_y \\ z \\ \delta \end{pmatrix}, \quad (3.28)$$

and  $R_{\text{drift}}$  is the *transfer matrix* for a drift space:

$$R_{\text{drift}} = \begin{pmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\beta_0^2 \gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.29)$$

Representing the transfer maps for different components in a beam line as matrices means that the maps can easily be combined, by multiplying the matrices, to construct the (linear) map for any section of the beam line.

### 3.2 Dipole Magnet

The field within an accelerator dipole magnet is (ideally) a uniform field, although the field in some types of dipole magnet (combined function dipoles) varies linearly with transverse position. In this section, we shall derive the transfer maps, in linear approximations, for regular dipoles and for combined function dipoles. Although the field in a dipole can be described quite easily, the dynamics of a charged particle is complicated by the fact that we usually wish to use a curved reference trajectory, to ensure that the values of the dynamical variables remain small through the dipole. The reference trajectory is usually chosen so that it is the path followed through the dipole by a particle with the reference momentum.

In a static magnetic field (and zero electric field) the Hamiltonian is, from (2.73):

$$H = \frac{\delta}{\beta_0} - (1 + hx) \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - (p_x - a_x)^2 - (p_y - a_y)^2 - \frac{1}{\beta_0^2 \gamma_0^2}} - (1 + hx) a_s. \quad (3.30)$$

To derive the transfer map for a particle moving through a magnetic field, we need an expression for the vector potential that describes the field. This becomes somewhat complicated when the reference trajectory has some curvature. Consider a simple dipole magnet, in which the field is in the vertical ( $y$ ) direction, and has a constant strength  $B_0$ :

$$\mathbf{B} = (0, B_0, 0). \quad (3.31)$$

There are important effects that occur when a particle moves from a drift region into a dipole field (or from a dipole field into a drift region): we shall

discuss these effects in Section 3.3, but for now we shall only consider the dynamics within the dipole field. We need to find a vector potential  $\mathbf{A}$  from which the magnetic field (3.31) can be derived using the usual relationship:

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (3.32)$$

Our standard co-ordinate system for describing the dynamics of particles in an accelerator beam line is a curvilinear co-ordinate system, with an axis,  $s$ , having some curvature  $h = 1/\rho$  (where  $\rho$  is the radius of curvature), and axes  $x$  and  $y$  perpendicular to the  $s$  axis at any point along the  $s$  axis. To find a vector potential that gives the field (3.31), we need an expression for the curl of a vector in the co-ordinates  $(x, y, s)$ . The general expression for the curl of a vector  $\mathbf{A}$  in orthogonal curvilinear co-ordinates  $(x_1, x_2, x_3)$  is [Gradshteyn and Ryzhik (2014b)]:

$$\nabla \times \mathbf{A} = \frac{1}{u_1 u_2 u_3} \sum_{i,j,k} \epsilon_{ijk} \mathbf{u}_i \frac{\partial(u_k A_j)}{\partial x_j}, \quad (3.33)$$

where  $\epsilon_{ijk}$  is the Levi-Civita symbol (equal to 1 for  $(i, j, k)$  a cyclic permutation of  $(1, 2, 3)$ , equal to  $-1$  for  $(i, j, k)$  an anticyclic permutation of  $(1, 2, 3)$ , and zero otherwise). The vectors  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  and  $\mathbf{u}_3$  are orthogonal basis vectors defined by:

$$\mathbf{u}_i = \frac{\partial \mathbf{r}}{\partial x_i}, \quad (3.34)$$

where:

$$\mathbf{r} = (X, Y, Z) \quad (3.35)$$

is the position vector of a point with Cartesian co-ordinates  $(X, Y, Z)$ . The basis vectors have magnitudes  $u_1$ ,  $u_2$  and  $u_3$ . In our case, the curvilinear co-ordinates  $(x, y, s)$  are related to the Cartesian co-ordinates by (2.58), (2.59) and (2.60). Then, we find from (3.33) that the components of the curl of  $\mathbf{A}$  are:

$$[\nabla \times \mathbf{A}]_x = \frac{\partial A_s}{\partial y} - \frac{1}{1 + hx} \frac{\partial A_y}{\partial s}, \quad (3.36)$$

$$[\nabla \times \mathbf{A}]_y = \frac{1}{1 + hx} \frac{\partial A_x}{\partial s} - \frac{\partial A_s}{\partial x} - \frac{h A_s}{1 + hx}, \quad (3.37)$$

$$[\nabla \times \mathbf{A}]_s = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}. \quad (3.38)$$

For a dipole field, it is possible to use a gauge in which  $A_x = A_y = 0$ : using such a gauge simplifies the form of the Hamiltonian (3.30). By substitution

into (3.36)–(3.38), it is found that the field (3.31) can be derived from the vector potential:

$$\mathbf{A} = \left( 0, 0, -B_0 x + \frac{B_0 h x^2}{2(1 + hx)} \right). \quad (3.39)$$

The scaled vector potential  $\mathbf{a}$  is:

$$\mathbf{a} = \frac{q}{P_0} \mathbf{A} = \left( 0, 0, -k_0 x + \frac{k_0 h x^2}{2(1 + hx)} \right), \quad (3.40)$$

where  $q$  is the particle charge,  $P_0$  is the reference momentum, and we define (for convenience) the normalised field strength  $k_0$ :

$$k_0 = \frac{q}{P_0} B_0. \quad (3.41)$$

Having obtained an expression for the vector potential, the next step is to derive the equations of motion. Substituting the vector potential (3.40) into the expression for the Hamiltonian (3.30) gives the Hamiltonian for a charged particle moving through a dipole field. Since the trajectory of the particle is an arc of a circle, the Hamiltonian admits an exact solution. However, the exact solution has a rather complicated form, and it is sufficient in many cases to make some approximations to obtain a more convenient expression. In particular, we are concerned for the present with the linear dynamics: therefore, we follow the same procedure that we used in the case of a drift space, and make the paraxial approximation. Expanding the Hamiltonian to second order in the dynamical variables (and dropping constant terms, which have no significance for the trajectory of the particle), the Hamiltonian for a dipole field can be approximated by:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2} + \left( k_0 - h \left( 1 + \frac{\delta}{\beta_0} \right) \right) x + \frac{k_0 h x^2}{2}. \quad (3.42)$$

Before writing the solution to the equations of motion for this Hamiltonian, we can already make some observations on the dynamics. First, if  $k_0 = 0$  (i.e. if the field is zero) and if  $h = 0$  (i.e. the reference trajectory has zero curvature), then the Hamiltonian (3.42), unsurprisingly, reduces to the second-order Hamiltonian for a drift space (3.26). In the case where a magnetic field is present and the reference trajectory has some curvature, then there are two additional terms compared to the second-order Hamiltonian for a drift space: the first is proportional to  $x$ , and the second is second order in  $x$ . From Hamilton's equations, we see that the linear term gives a horizontal deflection of the particle, exactly as we would expect for

a charged particle moving through a vertical magnetic field. However, we note that the deflection is zero in the case that:

$$h = \frac{k_0}{1 + \delta/\beta_0}. \quad (3.43)$$

In other words, if the reference trajectory has an appropriate curvature (depending on the magnetic field, the energy deviation and the reference velocity) then a particle initially following the reference trajectory will remain on the reference trajectory through the dipole field. When describing a beam line, however, we need to specify fixed values for the magnetic fields and the curvature of the reference trajectory: these values should not depend on the dynamical variables. A sensible choice for the curvature of the reference trajectory in a dipole magnet would be:

$$h = k_0. \quad (3.44)$$

In this case, a particle with the reference momentum and initially following the reference trajectory will remain on the reference trajectory. However, a particle with non-zero energy deviation  $\delta$  will not follow the reference trajectory, but after some time will deviate from it by a distance depending on the energy deviation and the magnetic field strength. The deviation of a particle trajectory from the reference trajectory induced in this way is known as *dispersion*.

It follows from equation (3.41), and from  $h = 1/\rho$ , that for a particle with momentum equal to the reference momentum:

$$B\rho = \frac{P_0}{q}, \quad (3.45)$$

where  $\rho$  is the radius of curvature of the trajectory of a particle of charge  $q$  and momentum  $P_0$  in a dipole field of strength  $B$ . Equation (3.45) shows that the product  $B\rho$  depends only on the charge and momentum of a particle: this result can also be obtained from Newtonian mechanics, by considering the centripetal force on a particle performing circular motion in a magnetic field. The quantity  $B\rho$  is often written as an alternative to  $P_0q$ , and is referred to as the *beam rigidity*.

The second-order term in  $x$  in the Hamiltonian (3.42) is what we would expect in the Hamiltonian for a simple harmonic oscillator (2.32). Consider a particle with momentum equal to the reference momentum (i.e.  $\delta = 0$ ) in a dipole magnet, with a reference trajectory having curvature  $h = k_0$ . Such a particle will, in general, perform oscillations around the reference trajectory in the horizontal plane, with period:

$$C_0 = \frac{2\pi}{h} = 2\pi\rho, \quad (3.46)$$

where  $C_0$  is the circumference of the reference trajectory in the case that the dipole magnet (and hence the reference trajectory) forms a complete circle. Transverse oscillations around a given reference trajectory are known as *betatron oscillations*, and are a consequence of transverse focusing. In the case of a dipole magnet, the origin of the transverse focusing is not hard to understand: the trajectory of a particle is a circle, which may in general be displaced from the circular reference trajectory. After one complete turn (i.e. distance of  $2\pi\rho$ ), the particle returns to its starting point, as long as  $p_y = 0$ . Of course, in most high energy accelerator beam lines, individual dipole magnets bend the beam through much less than  $360^\circ$ , so particles will only perform a small fraction of a betatron period within the dipole. The focusing provided by dipole magnets is referred to as *weak focusing*. Much stronger focusing is provided by quadrupole magnets.

Note that there is no focusing in the vertical direction. This means that in general the trajectory of a particle will be an arc of a helix, with the particle having constant vertical momentum.

A final observation from the dipole Hamiltonian (3.42) concerns the longitudinal co-ordinate  $z$ . In a drift space, the change in the  $z$  co-ordinate for a particle travelling parallel to the reference trajectory (i.e. with  $p_x = p_y = 0$ ) vanishes in the ultra-relativistic limit  $\gamma_0 \rightarrow 0$ . This reflects the fact that at very high energy the velocity of a particle approaches the fixed limit  $c$ . However, in a dipole, we see that the change in  $z$  co-ordinate of a particle has a dependence on  $x$ : if  $x$  is non-zero, then the change in  $z$  will be non-zero, even in the ultra-relativistic limit for a particle travelling parallel to the reference trajectory. The reason for this, of course, is that the reference trajectory is curved. Consider an interval corresponding to a given change in the independent variable  $s$ . In this interval, a particle with positive  $x$  has to travel a further distance than a particle with zero or negative  $x$  (assuming that the curvature of the reference trajectory is in the direction of negative  $x$ ). The change in the longitudinal co-ordinate for a particle moving through a dipole is a consequence of the variation of velocity with energy (which vanishes in the ultra-relativistic limit), and of the variation of path length with horizontal co-ordinate. The change in longitudinal co-ordinate as a particle moves along a beam line is generally known as *phase slip*, and is a particularly important characteristic of storage rings, where it leads to synchrotron oscillations. Phase slip will be discussed in more detail in Chapter 5.

Hamilton's equations with the Hamiltonian (3.42) have an exact

solution, which can be written as:

$$\vec{x}_1 = R_{\text{dipole}} \vec{x}_0 + \vec{m}_{\text{dipole}}. \quad (3.47)$$

Before giving expressions for  $R_{\text{dipole}}$  and  $\vec{m}_{\text{dipole}}$ , we consider the effect of including a quadrupole gradient in the dipole field: this is commonly done to provide a more compact (i.e. shorter) beam line with the desired beam optics properties for a specific application. A dipole magnet with a quadrupole gradient is commonly known as a *combined function dipole*, since such a magnet is used to provide strong horizontal and vertical focusing, as well as being used to change the direction of the beam.

Ideally, a combined function dipole would provide a field (scaled by  $q/P_0$ ) of the form:

$$\mathbf{b} = (k_1 y, k_0 + k_1 x, 0), \quad (3.48)$$

where  $k_0$  and  $k_1$  are constants, respectively they are the dipole and quadrupole components of the field. Such a magnet would provide a constant vertical field along the reference trajectory (which, if identified with the trajectory of a particle with the reference momentum, would be an arc of a circle), while strong focusing would be provided by a linear dependence of the vertical field on the horizontal offset from the reference trajectory, and of the horizontal field on the vertical offset from the reference trajectory. However, the co-ordinates in (3.48) are not within a Cartesian co-ordinate system, but (in general) within a system with a curved reference trajectory: then, the field given by (3.48) has non-zero divergence and is not a physically realisable magnetic field. To satisfy Maxwell's equations, there must be higher-order terms:

$$\mathbf{b} = (k_1 y + O(2), k_0 + k_1 x + O(2), 0), \quad (3.49)$$

where  $O(2)$  represents terms of second-order and higher in the co-ordinates  $x$  and  $y$ . To write an explicit expression for the Hamiltonian for particle motion in a combined function dipole, we need a vector potential that leads to a field of the form given in (3.49) and that also satisfies:

$$\nabla \times \mathbf{b} = 0. \quad (3.50)$$

(Note that a field  $\mathbf{b}$  obtained from a vector potential  $\mathbf{a}$  by  $\mathbf{b} = \nabla \times \mathbf{a}$  necessarily satisfies  $\nabla \cdot \mathbf{b} = 0$ .) In our curvilinear co-ordinate system, this turns out to be a non-trivial problem; however, a solution for the vector potential can be expressed as  $\mathbf{a} = (0, 0, a_s)$ , where:

$$\begin{aligned} a_s = & -k_0 x + \frac{k_0}{2} \frac{hx^2}{1+hx} - \frac{k_1}{2} \left( x^2 - y^2 \frac{\sinh^{-1}(\alpha)}{\alpha} \right) + \frac{k_1}{6} \frac{hx^3}{1+hx} \\ & + \frac{k_1}{6} y^2 \left( \frac{2}{\alpha^2} + \left( 1 - \frac{2}{\alpha^2} \right) \sqrt{1+\alpha^2} \right), \end{aligned} \quad (3.51)$$

and

$$\alpha = \frac{hy}{1 + hx}. \quad (3.52)$$

The magnetic field may be obtained by taking the curl of this vector potential. The complete expression for the field is not very enlightening, but is given, to first order in the co-ordinates, by (3.49). To find an approximate solution to the equations of motion for a particle moving through the field, we expand the vector potential to second order in the co-ordinates (before substituting into the Hamiltonian):

$$a_s = -k_0 x + \frac{k_0 h}{2} x - \frac{k_1}{2} (x^2 - y^2) + O(3). \quad (3.53)$$

The higher-order terms will have some impact on the dynamics: however, for now we are only concerned with the linear dynamics, so we shall assume that the higher-order effects are negligible (which will be the case for particles that remain sufficiently close to the reference trajectory).

The second-order Hamiltonian for a particle moving through a combined function dipole is the same as for a particle moving through a uniform field (3.42), but with an additional term in  $k_1$  representing the effect of the field gradient:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2} + \left( k_0 - h \left( 1 + \frac{\delta}{\beta_0} \right) \right) x + \frac{k_0 h}{2} x^2 + \frac{k_1}{2} (x^2 - y^2). \quad (3.54)$$

In the transverse variables  $(x, p_x)$  and  $(y, p_y)$ , this is similar to the Hamiltonian for a harmonic oscillator. The term linear in  $x$  leads to a horizontal deflection that grows with distance  $s$  along the reference trajectory; this is superposed on an oscillation arising from the combination of weak focusing from the dipole field (the term in  $k_0 x^2$ ) and strong focusing from the quadrupole component of the field (the term in  $k_1 x^2$ ). The fact that there is a minus sign in front of the  $y^2$  term means that, if  $k_1$  is positive, the solution for the vertical co-ordinate is expressed in terms of hyperbolic trigonometric functions: that is, there is no oscillation in the vertical direction, but an exponential growth in the vertical co-ordinate (and momentum). In the horizontal plane, some additional complexity arises from the dependence on the energy deviation; however, since the energy deviation is constant, when solving the equations of motion  $\delta$  may be treated as a parameter, rather than a variable. Since we have retained only terms to second order in the dynamical variables, the final values of the variables (after evolution of the independent variable  $s$  over a distance  $L$  through the field of the

combined function dipole) have a linear dependence on the initial values of the variables. The linear transfer map for a dipole may be expressed as:

$$\vec{x}_1 = R_{\text{CFD}} \vec{x}_0 + \vec{m}_{\text{CFD}}. \quad (3.55)$$

The (constant) components of the transfer matrix  $R_{\text{CFD}}$  and the vector  $\vec{m}_{\text{CFD}}$  for a combined function dipole (CFD) are expressed in terms of the magnetic field strength, the gradient of the magnetic field, the reference momentum and the evolution distance  $L$ . Explicitly:

$$R_{\text{CFD}} = \begin{pmatrix} c_x & s_x & 0 & 0 & 0 & \bar{h} \frac{1-c_x}{\omega_x^2} \\ -\omega_x^2 s_x & c_x & 0 & 0 & 0 & \bar{h} s_x \\ 0 & 0 & c_y & s_y & 0 & 0 \\ 0 & 0 & \omega_y^2 s_y & c_y & 0 & 0 \\ -\bar{h} s_x & -\bar{h} \frac{1-c_x}{\omega_x^2} & 0 & 0 & 1 & \frac{L}{\beta_0^2 \gamma_0^2} - \bar{h}^2 \frac{L-s_x}{\omega_x^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.56)$$

where:

$$s_x = \frac{\sin(\omega_x L)}{\omega_x}, \quad (3.57)$$

$$c_x = \cos(\omega_x L), \quad (3.58)$$

$$s_y = \frac{\sinh(\omega_y L)}{\omega_y}, \quad (3.59)$$

$$c_y = \cosh(\omega_y L), \quad (3.60)$$

$$\bar{h} = \frac{h}{\beta_0}, \quad (3.61)$$

$$\omega_x = \sqrt{h k_0 + k_1}, \quad (3.62)$$

$$\omega_y = \sqrt{k_1}. \quad (3.63)$$

The vector  $\vec{m}_{\text{CFD}}$  is given by:

$$\vec{m}_{\text{CFD}} = \begin{pmatrix} (h - k_0) \frac{1-c_x}{\omega_x^2} \\ (h - k_0) s_x \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (3.64)$$

Explicit expressions for the matrix  $R_{\text{dipole}}$  and the vector  $\vec{m}_{\text{dipole}}$  for a uniform magnetic field are readily obtained from (3.56) and (3.64) by taking the limit  $k_1 \rightarrow 0$ .

### 3.3 Dipole Fringe Fields and Edge Focusing

The transfer map for a dipole that was derived in Section 3.2 describes the dynamics of a particle moving entirely within the field of the dipole. There are significant effects associated with the fringe fields that a particle sees as it enters and leaves a dipole. Properties of the magnetic field in the vicinity of the ends of a dipole were discussed in Section 1.3.3: in this section, we will look more closely at the effects that fringe fields in dipoles can have on the dynamics of particles entering or leaving a dipole magnet. In order to derive useful expressions for the transfer maps describing the dynamics, we will use simplified models of the fringe fields, which, for dipole magnets, are often sufficient to understand and describe the effects on the particle dynamics. If more detailed or accurate descriptions of the dynamics are needed, then numerical integration of the equations of motion in the fringe fields may be necessary: some of the techniques that may be used in this case will be discussed in Chapter 10.

In a linear approximation, dipole fringe fields can be represented as having horizontal and vertical focusing (or defocusing) effects, with a focusing strength dependent on a variety of factors, including the main field of the dipole, the extent of the fringe field, and the angle of the entrance or exit face with respect to the reference trajectory. In this section, we shall develop a simple model for the fringe fields of a dipole, and derive a linear transfer map to represent the dynamics of a particle passing through a fringe field.

As we have seen in Section 1.3.3, at the entrance of a dipole magnet the field cannot change abruptly from zero to the nominal dipole field, but must rise smoothly (and continuously) over some distance. Similarly, at the exit of the dipole, the field must fall smoothly over some distance. The exact shape of the field at the ends of a dipole will depend on the detailed geometry of the magnet: the gap between the pole faces, for example, will have a significant impact on the extent of the fringe fields. Rather than use a detailed model of the kind that we developed in Section 1.3.3, as a simple approximation we can assume that the vertical field component rises linearly with distance over some distance  $d_f$  measured perpendicularly from the entrance face of the magnet. That is, if the entrance face is at

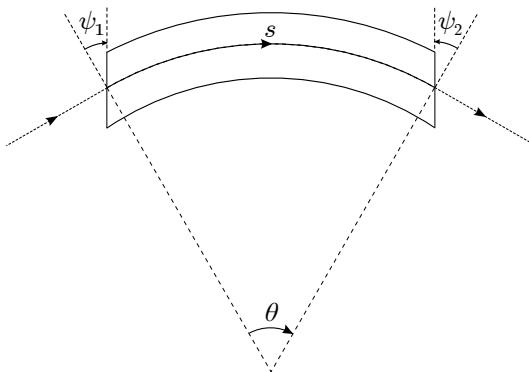


Fig. 3.1 Rotation angles of entrance and exit faces in a dipole magnet. If the pole face rotation angles are  $\psi_1 = \psi_2 = 0$ , then the pole face is perpendicular to the reference trajectory at the entrance and exit of the magnet. With  $\psi_1 = \psi_2 = \theta/2$  (where  $\theta$  is the total bending angle of the reference trajectory in the dipole), the pole faces are parallel.

$s = d_f$ , then the magnetic field in the region  $0 < s < d_f$  is given by:

$$B_x = 0, \quad (3.65)$$

$$B_y = \frac{s}{d_f} B_0, \quad (3.66)$$

$$B_s = \frac{y}{d_f} B_0, \quad (3.67)$$

where  $B_0$  is the magnetic field strength in the main part of the dipole. We can assume that the horizontal component of the magnetic field is zero if the width of the poles are large (so that over some range there is translational symmetry along the  $x$  axis). There is a longitudinal component of the field that depends on the  $y$  co-ordinate: this is required to satisfy Maxwell's equation  $\nabla \times \mathbf{B} = 0$ . Physically, the variation of vertical and longitudinal field components describes the field lines 'bending out' from between the poles of the magnet.

Dipole magnets in accelerators are often built so that the entrance and exit faces are parallel to each other. Since the reference trajectory inside the magnet is curved, this means that there is a rotation of the entrance and exit faces with respect to planes perpendicular to the reference trajectory. The geometry is shown in Fig. 3.1. If the total bending angle of a dipole is  $\theta$ , then the entrance and exit faces will be parallel if the rotation angles are  $\psi_1 = \psi_2 = \theta/2$ . In principle, the entrance and exit faces of a dipole could be rotated by any amount: therefore, we must consider general values of  $\psi_1$  and  $\psi_2$ .

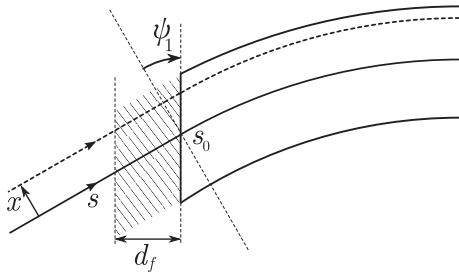


Fig. 3.2 Effect of fringe fields in a dipole magnet. The fringe field extends a perpendicular distance  $d_f$  from the entrance face of the magnet, which is rotated by angle  $\psi_1$ . A particle travelling parallel to the reference trajectory with a horizontal displacement  $x$  from the reference trajectory experiences the fringe field over a distance  $L = d_f / \cos(\psi_1)$ . The particle also has to travel a distance  $x \tan(\psi_1)$  further than a particle on the reference trajectory before entering the main field of the dipole magnet.

The rotation of the entrance (or exit) face of a dipole has three effects. First of all, the fringe field extends over a distance  $L$  of the reference trajectory that is longer than  $d_f$  (the range of the fringe field measured perpendicularly from the face of the magnet): see Fig. 3.2. For the entrance face:

$$L = \frac{d_f}{\cos(\psi_1)}. \quad (3.68)$$

Second, the field itself is rotated with the face of the magnet, so that the component of the field perpendicular to the face of the magnet has components along the  $x$  and  $s$  axes. In the fringe field region leading up to the entrance face,  $0 < s < L$ :

$$B_x = -\frac{y}{d_f} B_0 \sin(\psi_1) = -\frac{y}{L} B_0 \tan(\psi_1), \quad (3.69)$$

$$B_s = \frac{y}{d_f} B_0 \cos(\psi_1) = \frac{y}{L} B_0. \quad (3.70)$$

The third effect of a rotation of the entrance face is that particles with different  $x$  co-ordinates enter the main field of the magnet at different  $s$  positions. This is a complicated effect to take into account: the dynamics of particles in an accelerator are most easily described when the boundaries of each component are defined by planes perpendicular to the reference trajectory, and the fields between those planes are independent of  $s$ . We shall deal with the situation that arises from the rotation of the entrance face of a dipole magnet as follows. Suppose the entrance face of a dipole is at  $s = s_0$ , and the face is rotated through angle  $\psi_1$  about the  $y$  axis at

$s = s_0$ . A particle with co-ordinate  $x$  travelling parallel to the reference trajectory passes the entrance face at  $s = s_0 + x \tan(\psi_1)$ . The particle therefore sees a ‘deficit’ integrated vertical field  $B_0 x \tan(\psi_1)$ . We assume that this can be represented by an additional vertical component of the field in the fringe field region, so that in the region  $0 < s < L$  leading up to the entrance face of a dipole, the vertical field component is given by:

$$B_y = \frac{s}{L} B_0 - \frac{x}{L} B_0 \tan(\psi_1). \quad (3.71)$$

The first term on the right-hand side of (3.71) represents the linear change in vertical field component with distance through the fringe field. The second term represents the change in integrated field seen by the particle over the distance it travels before it actually crosses the rotated entrance face of the magnet.

The field with components (3.69), (3.70) and (3.71) satisfies Maxwell’s equations in free space:

$$\nabla \cdot \mathbf{B} = \nabla \times \mathbf{B} = 0. \quad (3.72)$$

We therefore assume that this field provides a reasonable model for the dynamics in the fringe field of a dipole. The field can be derived from the vector potential:

$$A_x = 0, \quad (3.73)$$

$$A_y = \frac{B_0}{L} xy, \quad (3.74)$$

$$A_s = \frac{B_0}{2L} (x^2 - y^2) \tan(\psi_1) - \frac{B_0}{L} xs. \quad (3.75)$$

It is not possible to find a gauge in which two of the components of the vector potential vanish, because the field has a dependence on all three co-ordinates. We choose a potential in which  $A_x = 0$ . To find the transfer map for a particle moving through the fringe field, we follow the usual procedure: we first expand the Hamiltonian (2.73) to second order in the dynamical variables (and use a straight reference trajectory,  $h = 0$ ), then integrate the equations of motion. In this case, we integrate from  $s = 0$  to

$s = L$ . The result is:

$$\begin{aligned} x(L) &= \left( L - \frac{\sinh(\sqrt{k_f}L)}{\sqrt{k_f}} \right) \cot(\psi_1) + \\ &\quad \cosh(\sqrt{k_f}L)x(0) + \frac{\sinh(\sqrt{k_f}L)}{\sqrt{k_f}} p_x(0), \end{aligned} \quad (3.76)$$

$$\begin{aligned} p_x(L) &= \left( 1 - \cosh(\sqrt{k_f}L) \right) \cot \psi_1 + \\ &\quad \sqrt{k_f} \sinh(\sqrt{k_f}L)x(0) + \cosh(\sqrt{k_f}L)p_x(0), \end{aligned} \quad (3.77)$$

$$y(L) = \cos(\sqrt{k_f}L)y(0) + \frac{\sin(\sqrt{k_f}L)}{\sqrt{k_f}} p_y(0), \quad (3.78)$$

$$p_y(L) = -\sqrt{k_f} \sin(\sqrt{k_f}L)y(0) + \cos(\sqrt{k_f}L)p_y(0), \quad (3.79)$$

$$z(L) = z(0) + \frac{L}{\beta_0^2 \gamma_0^2} \delta(0), \quad (3.80)$$

$$\delta(L) = \delta(0), \quad (3.81)$$

where:

$$k_f = \frac{q}{P_0} \frac{\tan(\psi_1)}{L} B_0. \quad (3.82)$$

Note that in the horizontal part of the transfer map there are zeroth-order terms as well as first-order terms: the zeroth-order terms arise from the fact that in the fringe field there is a non-zero vertical magnetic field that will deflect particles from the straight reference trajectory.

In the limit  $L \rightarrow 0$ , the transfer map for a dipole fringe field can be written in the form of a transfer matrix:

$$R_{\text{fringe}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ k_0 \tan(\psi_1) & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -k_0 \tan(\psi_1) & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.83)$$

where:

$$k_0 = \frac{q}{P_0} B_0. \quad (3.84)$$

If the entrance face of a dipole is rotated so that it is not perpendicular to the reference trajectory, then the fringe field has a focusing effect on the beam even if the fringe field extends an infinitesimal distance from the entrance face. The origin of the vertical focusing is the component

of the fringe field perpendicular to the entrance face (which we refer to as the ‘normal component’ of the fringe field): when the entrance face is rotated, the normal component rotates with it, so a particle moving through the fringe field parallel to the reference trajectory experiences a vertical deflection. The size (and direction) of the deflection depends on the strength of the normal component: this varies linearly with the vertical co-ordinate. The *integrated* strength of the normal component of the fringe field is independent of the distance over which the fringe field extends. The shorter the distance, the more sharply the field rises from zero, and the larger the normal component of the fringe field.

The horizontal (de)focusing effect of the dipole fringe field is essentially a geometric effect. When the entrance face is rotated so that it is not perpendicular to the reference trajectory, a particle moving towards the dipole has to travel a longer or shorter distance to enter the dipole, depending on its horizontal co-ordinate. The horizontal trajectory of the particle is changed as a result.

The transfer map for the exit fringe field is the same as that for the entrance fringe field (3.76)–(3.81), but with  $\psi_1$  replaced by  $\psi_2$ . Since the transfer maps are similar to those for a quadrupole, the focusing effect from the fringe field (when the entrance and exit faces are rotated) is known as *edge focusing*.

In the limit  $\psi_1 \rightarrow 0$ , there is no focusing. The transfer map becomes:

$$x(L) = -\frac{qB_0}{P_0} \frac{L^2}{6} + x(0) + L p_x(0), \quad (3.85)$$

$$p_x(L) = -\frac{qB_0}{P_0} \frac{L}{2} + p_x(0), \quad (3.86)$$

$$y(L) = y(0) + L p_y(0), \quad (3.87)$$

$$p_y(L) = p_y(0), \quad (3.88)$$

$$z(L) = z(0) + \frac{L}{\beta_0^2 \gamma_0^2} \delta(0), \quad (3.89)$$

$$\delta(L) = \delta(0). \quad (3.90)$$

In the vertical and longitudinal variables, the transfer map is the same as for a drift space. In the horizontal variables, there is a deflection arising from the vertical field component in the fringe field.

The simple model of dipole fringe fields used here leads to a convenient form for the transfer map, and is often sufficient for describing the dominant effects on the particle dynamics. It is of course possible (as we saw in Section 1.3.3) to construct more accurate descriptions of fringe fields in

dipoles, as well as in other multipole magnets. In general, modelling the particle dynamics in a fringe field will require numerical integration of the equations of motion, though some approximations may make it possible to write down transfer maps in analytical form in some cases. Where analytical solution of the equations of motion is possible, then different models of a fringe field will lead to different forms for the transfer map across the fringe field [Brown (1982); Sagalovsky (1989); Carey (1992)]. In such cases, the transfer maps can often be parameterised in terms of quantities such as the gap between the poles and integrated strength of the fringe field, as well as the angle of the pole face with respect to the reference trajectory.

### 3.4 Quadrupole Magnet

In an accelerator beam line, the purpose of quadrupole magnets is to control the beam size by providing transverse focusing. Compared to a dipole magnet, deriving the transfer matrix for a quadrupole magnet is straightforward because the reference trajectory can be chosen to be a straight line through the centre of the quadrupole field (the magnetic axis): since, by design, the field is zero along this line, a particle entering along the magnetic axis will continue in a straight line through the length of the magnet. Particles with some transverse offset with respect to the magnetic axis will receive some deflection depending on the size and direction of the offset. This deflection is described by the transfer matrix for the quadrupole.

The magnetic field (scaled by  $q/P_0$ ) in a quadrupole is given by:

$$\mathbf{b} = (k_1 y, k_1 x, 0), \quad (3.91)$$

which can be obtained from the vector potential:

$$\mathbf{a} = \left( 0, 0, -\frac{k_1}{2}(x^2 - y^2) \right). \quad (3.92)$$

Substituting into the general accelerator Hamiltonian (2.73), with  $h = 0$  for a straight reference trajectory, gives the Hamiltonian for particle motion in a quadrupole:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left( \delta + \frac{1}{\beta_0} \right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}} + \frac{k_1}{2}(x^2 - y^2). \quad (3.93)$$

To solve the equations of motion that follow from this Hamiltonian, we need (as usual) to make the paraxial approximation by expanding the square root to second order in the dynamical variables:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2 \gamma_0^2} + \frac{k_1}{2}(x^2 - y^2). \quad (3.94)$$

Note that we have dropped a constant (zeroth-order) term, which has no effect on the dynamics.

Solving the equations of motion that follow from the Hamiltonian  $H_2$  gives the transfer matrix  $R_{\text{quad}}$ :

$$\vec{x}_1 = R_{\text{quad}} \vec{x}_0, \quad (3.95)$$

where, for a distance  $L$  through a quadrupole field:

$$R_{\text{quad}} = \begin{pmatrix} \cos(\omega L) & \frac{\sin(\omega L)}{\omega} & 0 & 0 & 0 & 0 \\ -\omega \sin(\omega L) \cos(\omega L) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cosh(\omega L) & \frac{\sinh(\omega L)}{\omega} & 0 & 0 \\ 0 & 0 & \omega \sinh(\omega L) & \cosh(\omega L) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\beta_0^2 \gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.96)$$

and:

$$\omega = \sqrt{k_1}. \quad (3.97)$$

If  $k_1 > 0$ , then  $\omega$  is real, and the motion in the horizontal plane is that of a harmonic oscillator. In most practical accelerator beam lines, a particle will perform only a small fraction of an oscillation within a quadrupole magnet (the length of most quadrupoles is small compared to the wavelength of the betatron motion). In the vertical plane, however, the motion is very different: a particle entering the quadrupole with a small vertical offset receives a vertical deflection that increases the offset. In other words, the vertical force is *defocusing*, rather than focusing. If  $k_1 < 0$ , the situation is reversed, and the quadrupole focuses in the vertical direction, and defocuses in the horizontal direction.

It appears that although a quadrupole can be used to focus a beam in one transverse direction, it defocuses the beam in the orthogonal transverse direction. This is a consequence of the constraints on the magnetic field imposed by Maxwell's equations. However, as we shall see in Chapter 4, it is possible to combine quadrupoles that focus alternately in the horizontal and vertical transverse directions, to provide overall focusing in both directions.

Inspecting the transfer matrix for a quadrupole given in (3.96), we see that the focusing strength experienced by a particle is independent of the energy deviation  $\delta$ . There appears to be some inconsistency here, since the focusing parameter  $\omega$ , given by (3.97), can be written as:

$$\omega^2 = \frac{q}{P_0} \frac{\partial B_y}{\partial x}, \quad (3.98)$$

where  $P_0$  is the reference momentum,  $q$  the particle charge, and  $B_y$  the vertical component of the magnetic field. The effect of a quadrupole magnet on two particles with different momenta may be described in two different ways: either, with the same reference momentum but different energy deviations; or, with zero energy deviation but different reference momenta. In the latter case, the focusing parameter  $\omega$  will be different for the two particles, which will therefore experience different focusing effects. But in the former case, the focusing parameter will be the same for the two particles; and since the focusing is independent of the energy deviation, the particles experience the same focusing effects. It appears that the same system, described in two different ways, leads to two different results.

The reason for the discrepancy is that the variation of focusing with momentum is a nonlinear effect. A particle entering a quadrupole with  $p_x = 0$  has a value of  $p_x$  at the exit of the quadrupole that is proportional to the initial value of  $x$ : but the constant of proportionality depends on the energy deviation  $\delta$ . When we made a second-order (paraxial) approximation to the Hamiltonian in order to write the transfer map as a matrix, we lost this effect.

The variation of focusing strength with particle energy in an accelerator beam line is known as *chromaticity*. Chromaticity is an important effect, and although it is properly described as a nonlinear phenomenon, it is sometimes useful to be able to include it in a linear description of the dynamics. One way to do this for a quadrupole (and other beam line components with static magnetic fields) is to treat the transverse variables  $(x, p_x)$  and  $(y, p_y)$  separately from the longitudinal variables  $(z, \delta)$ . This is possible, since we know that if we neglect radiation and certain collective effects, the energy of a particle in a static magnetic field is constant. Therefore, the energy deviation  $\delta$  of a particle moving through a quadrupole will be constant. We can then expand the Hamiltonian (3.93) to second order in the transverse variables, while maintaining the exact dependence on  $\delta$ :

$$H'_2 = \frac{p_x^2}{2D} + \frac{p_y^2}{2D} + \frac{\delta}{\beta_0} - D + \frac{k_1}{2}(x^2 - y^2), \quad (3.99)$$

where

$$D = \sqrt{1 + \frac{2\delta}{\beta_0} + \delta^2}. \quad (3.100)$$

The solutions to Hamilton's equations for the transverse variables are then:

$$x(L) = \cos(\omega'L)x(0) + \frac{\sin(\omega'L)}{\omega'D}p_x(0), \quad (3.101)$$

$$p_x(L) = -\omega'D \sin(\omega'L)x(0) + \cos(\omega'L)p_x(0), \quad (3.102)$$

$$y(L) = \cosh(\omega'L)y(0) + \frac{\sinh(\omega'L)}{\omega'D}p_y(0), \quad (3.103)$$

$$p_y(L) = \omega'D \sinh(\omega'L)y(0) + \cosh(\omega'L)p_y(0), \quad (3.104)$$

where

$$\omega' = \sqrt{\frac{k_1}{D}}. \quad (3.105)$$

While this approach can sometimes be useful for including chromaticity in the transverse dynamics, the drawback is that the equations describing the longitudinal dynamics (in particular, the change in the dynamical variable  $z$ ) become very complex. Finally, it is also worth noting that the transverse dynamics in a drift space have a dependence on the energy deviation, as can be seen from (3.13) and (3.15). In other words, even drift spaces have chromaticity.

A skew quadrupole field is obtained from a normal quadrupole field by a rotation of the field through  $\pi/4$  about the reference trajectory. The field (scaled by  $q/P_0$ ) in a skew quadrupole is:

$$\mathbf{b} = (-k_1^{(s)}x, k_1^{(s)}y, 0), \quad (3.106)$$

where the normalised skew quadrupole field strength is:

$$k_1^{(s)} = \frac{q}{P_0} \frac{\partial B_y}{\partial y} = -\frac{q}{P_0} \frac{\partial B_x}{\partial x}. \quad (3.107)$$

The field (3.106) may be obtained from the (scaled) vector potential:

$$\mathbf{a} = (0, 0, -k_1^{(s)}xy). \quad (3.108)$$

The Hamiltonian to second order in the dynamical variables is:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2} + k_1^{(s)}xy. \quad (3.109)$$

The solution to Hamilton's equations can be represented as a transfer matrix:

$$\vec{x}_1 = R_{\text{sq}}\vec{x}_0, \quad (3.110)$$

where  $R_{\text{sq}}$  is related to the transfer matrix for a normal quadrupole (3.96). Specifically:

$$R_{\text{sq}} = R_{\text{rot}}\left(\frac{\pi}{4}\right) R_{\text{quad}} R_{\text{rot}}\left(-\frac{\pi}{4}\right), \quad (3.111)$$

where  $R_{\text{rot}}(\theta)$  is a transfer matrix representing a rotation in co-ordinate space through angle  $\theta$  about the reference trajectory:

$$R_{\text{rot}}(\theta) = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) & 0 & 0 & 0 \\ 0 & \cos(\theta) & 0 & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & 0 & \cos(\theta) & 0 & 0 & 0 \\ 0 & \sin(\theta) & 0 & \cos(\theta) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.112)$$

The form of the transfer matrix given in (3.111) can be interpreted as a rotation of the phase space variables through  $-\pi/4$  about the reference trajectory, followed by transport through a *normal* quadrupole, followed by an inverse rotation of the phase space variables through  $-\pi/4$ . This is the same as transport through a normal quadrupole field rotated (about the reference trajectory) through angle  $\pi/4$ : a normal quadrupole field rotated through this angle is, by definition, the same as a skew quadrupole field.

### 3.5 Solenoid

A solenoid produces a uniform magnetic field that is parallel to the direction in which the beam is travelling. Usually, it is appropriate to choose a straight reference trajectory that is parallel to the field, since a particle travelling along such a trajectory experiences no deflection from the field. Solenoids have various uses in accelerator beam lines: they are commonly used to provide transverse focusing, but can also play a role in capturing particles from sources, or for controlling spin polarisation. Solenoids also occur in the detectors used in high energy physics experiments.

The magnetic field in a solenoid is given by:

$$\mathbf{B} = (0, 0, B_0). \quad (3.113)$$

For dipoles and quadrupoles, we were able to choose a gauge in which only the  $A_s$  component of the vector potential was non-zero. This is not possible for a solenoid. Various gauge choices can be made, but a conventional choice is:

$$\mathbf{A} = \left( -\frac{B_0}{2}y, \frac{B_0}{2}x, 0 \right). \quad (3.114)$$

Some care is needed here, because the dynamical variables  $p_x$  and  $p_y$  are related to the canonical momentum of a particle, to which the vector potential makes a contribution. We therefore need to consider the transformation of the dynamical variables from a field-free region outside the solenoid

(where the vector potential can be taken to be zero) to the region within the solenoid (where the vector potential has non-zero transverse components). This transformation needs to include also the effect of the fringe field of the solenoid: since the field obeys Maxwell's equations, it cannot be abruptly turned on at the entrance to the solenoid and off at the exit. We should use a realistic model of the field at the ends of the solenoid (as discussed, for example, in Sections 1.3.1 and 1.3.3. However, realistic field models can lead to complicated descriptions of the dynamics for particles entering or leaving a solenoid. We therefore use a simplified model in which we consider the field at the end of the solenoid to be confined within a short region between two planes (separated by distance  $\Delta s$ ) perpendicular to the main field of the solenoid. The field lines within the fringe field region are radial, centred on a point on the axis of the solenoid (which we take to be the point where the reference trajectory enters the solenoid). Continuity of the lines of magnetic flux means that, within the fringe field region, the flux density increases linearly with distance from the axis of the solenoid. At the entrance of the solenoid, we can write the fringe field as:

$$\mathbf{B} = \left( -B_0 \frac{x}{2\Delta s}, -B_0 \frac{y}{2\Delta s}, 0 \right). \quad (3.115)$$

Within the body of the solenoid, the total flux passing through a disc of radius  $r$  centred on the axis of the solenoid is  $\pi r^2 B_0$ ; the same flux passes through the curved surface of a cylinder of radius  $r$  and length  $\Delta s$  within the fringe field described by (3.115).

Let us now consider the deflection of a particle crossing the fringe field of the solenoid. If the particle crosses the fringe field on a trajectory that has some angle  $\theta$  to the reference trajectory, then the component of the velocity of the particle in the  $s$  direction is  $v_s = \beta c \cos(\theta)$ . The time taken for the particle to cross the fringe field region is  $\Delta t = \Delta s / \beta c \cos(\theta)$ . Therefore, the change in the horizontal component of the momentum (scaled by the reference momentum) on crossing the fringe field is:

$$\Delta p_x = -\frac{qv_s B_y \Delta t}{P_0} + \frac{q\Delta A_x}{P_0}, \quad (3.116)$$

where  $\Delta A_x$  is the change in the horizontal component of the vector potential on crossing the fringe field. Substituting for  $v_s$ ,  $\Delta t$  and  $B_y$ , we obtain:

$$\Delta p_x = \frac{q}{P_0} \frac{B_0}{2} y + \frac{q\Delta A_x}{P_0}. \quad (3.117)$$

But, from (3.114), we have:

$$\Delta A_x = -\frac{B_0}{2} y. \quad (3.118)$$

Hence, the horizontal kick from the magnetic field in the fringe field is exactly cancelled by the change in the vector potential, and the total change in the horizontal component of the canonical momentum on crossing the fringe field is zero:

$$\Delta p_x = 0. \quad (3.119)$$

Not surprisingly (since the system is rotationally symmetric around the axis of the solenoid), the same is true for the vertical component of the vector potential. Also, there is no change in the energy deviation, and in the limit  $\Delta s \rightarrow 0$  there are no changes in the co-ordinates  $x$ ,  $y$  or  $z$ . Thus, the transfer map for a solenoid fringe field in our simple model is just the identity.

Now we can consider the dynamics of a particle within the main field of the solenoid. The dynamics are described by the accelerator Hamiltonian (2.73), with  $h = 0$  (for a straight reference trajectory), and with the vector potential (3.114). The Hamiltonian for a solenoid is:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - (p_x + k_s y)^2 - (p_y - k_s x)^2 - \frac{1}{\beta_0^2 \gamma_0^2}}, \quad (3.120)$$

where

$$k_s = \frac{q}{P_0} \frac{B_0}{2}. \quad (3.121)$$

As usual, in order to obtain a linear approximation to the solutions to the equations of motion, we make the paraxial approximation, and expand the Hamiltonian to second order in the dynamical variables:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2 \gamma_0^2} + \frac{k_s^2}{2} x^2 + \frac{k_s^2}{2} y^2 - k_s x p_y + k_s y p_x. \quad (3.122)$$

Note that there are focusing terms for both the horizontal and the vertical motion (the fourth and fifth terms respectively in  $H_2$ ): the solenoid provides simultaneous focusing in both transverse degrees of freedom. Also note that the focusing is independent of whether  $k_s$  is positive or negative: it does not matter whether the field lines are parallel or antiparallel to the direction of the beam. The final two terms in the expression (3.122) for the Hamiltonian include factors in horizontal and vertical variables. This means that there is coupling between the horizontal and vertical motion: specifically, there is a horizontal force depending on the vertical momentum, and vice versa. Coupling between the horizontal and vertical motion is not an effect that we found with dipoles or quadrupoles, but is an intrinsic effect in solenoids.

Physically, particles moving through a solenoid follow helical trajectories, performing cyclotron oscillations around the field lines.

Given the second-order Hamiltonian (3.122), we can solve Hamilton's equations to find the transfer matrix for a solenoid of length  $L$ . The result is:

$$\vec{x}_1 = R_{\text{sol}} \vec{x}_0, \quad (3.123)$$

where the transfer matrix  $R_{\text{sol}}$  for a distance  $L$  through a solenoid field is:

$$R_{\text{sol}} = \begin{pmatrix} \cos^2(\omega L) & \frac{1}{2\omega} \sin(2\omega L) & \frac{1}{2} \sin(2\omega L) & \frac{1}{\omega} \sin^2(\omega L) & 0 & 0 \\ -\frac{\omega}{2} \sin(2\omega L) & \cos^2(\omega L) & -\omega \sin^2(\omega L) & \frac{1}{2} \sin(2\omega L) & 0 & 0 \\ -\frac{1}{2} \sin(2\omega L) & -\frac{1}{\omega} \sin^2(\omega L) & \cos^2(\omega L) & \frac{1}{2\omega} \sin(2\omega L) & 0 & 0 \\ \omega \sin^2(\omega L) & -\frac{1}{2} \sin(2\omega L) & -\frac{\omega}{2} \sin(2\omega L) & \cos^2(\omega L) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\beta_0^2 \gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (3.124)$$

and:

$$\omega = k_s = \frac{q}{P_0} \frac{B_0}{2}. \quad (3.125)$$

Note that the transfer matrix for a solenoid field has a different structure from the transfer matrices for a drift, dipole magnet or quadrupole magnet. In particular, the components describing coupling between horizontal and vertical motion (for example, referring to the (row, column) of  $R_{\text{sol}}$ , the (1, 3), (1, 4), (2, 3) and (2, 4) components) have non-zero values in the case of the solenoid. This simply means that the vertical co-ordinate and momentum of a particle at the exit of a solenoid depend (in general) on the horizontal co-ordinate and momentum at the entrance, and vice versa. This feature makes the description of the beam optics in a beam line containing a solenoid somewhat more complicated than in a beam line consisting just of drifts, dipoles and quadrupoles; this will be discussed in more detail in Chapter 5.

A solenoid magnet has chromaticity in much the same way as a quadrupole: the transverse co-ordinates and momenta at the exit of a solenoid depend on the momentum of the particle, as well as on the transverse co-ordinates and momenta at the entrance. However, in making a second-order approximation to the Hamiltonian for a particle in a solenoid field, we 'lost' the chromaticity, in just the same way as we did for a quadrupole. The chromaticity is properly described by including higher-order terms in the transfer map. Alternatively, since the energy deviation

is constant (if we neglect radiation and certain collective effects) the energy deviation can be included as a parameter in a transfer map describing the transverse dynamics. In other words, the same approaches to describing the chromaticity in a solenoid can be used as were discussed for the case of a quadrupole.

### 3.6 Radiofrequency Cavity

Radiofrequency (rf) cavities produce time-dependent electromagnetic fields that are used to control particles in a beam in a variety of different ways. For example, they may be used to accelerate or decelerate the particles within a bunch; or to ensure that particles remain confined within a bunch; or to provide transverse deflections that are dependent on the time that a particle arrives at the cavity. In this section, we shall consider only the simplest (and one of the most common) applications, namely particle acceleration. Even just in this case, the details of the electromagnetic fields and their effect on the beam depend strongly on specific features of the cavity design, and can get very complicated. We shall therefore consider only a simple model, based on a cylindrical cavity operating in the  $\text{TM}_{010}$  mode: transfer maps have been derived for more general and realistic cases [Slater (1948); Chambers (1968); Rosenzweig and Serafini (1994)].

The fields in rf cavities were discussed in Chapter 1. In the case of a cylindrical cavity, the field components in the  $\text{TM}_{010}$  mode are given by (1.328)–(1.333). In cylindrical polar co-ordinates  $(r, \theta, s)$ , and using a notation where the fields are expressed as real numbers, the components of the electric field can be written:

$$E_r = 0, \quad (3.126)$$

$$E_\theta = 0, \quad (3.127)$$

$$E_s = E_0 J_0(kr) \sin(\omega t + \phi_0). \quad (3.128)$$

The components of the magnetic field can be written:

$$B_r = 0, \quad (3.129)$$

$$B_\theta = \frac{E_0}{c} J_1(kr) \cos(\omega t + \phi_0), \quad (3.130)$$

$$B_s = 0. \quad (3.131)$$

We assume that the field amplitude  $E_0$  is constant. At high beam currents the energy transferred from the cavity to the beam can be significant, leading to a drop in amplitude of the fields in the cavity. This effect, known

as *beam loading*, can be compensated over a train of bunches by feedback systems on the rf power sources. The fields generated in the cavities by charged particles in the beam can also be significant: these ‘wake fields’ will be discussed further in Chapter 14. For the present, we assume that the beam intensity is low enough that the effects of beam loading and wake fields can be neglected.

To satisfy Maxwell’s equations, the frequency and wave number are related by  $\omega = kc$ . To satisfy the boundary conditions on the electric and magnetic fields at the walls of the cavity, the wave number  $k$  must be given by:

$$k = \frac{p_{01}}{a}, \quad (3.132)$$

where  $a$  is the radius of the cavity and  $p_{01}$  is the first zero of the Bessel function  $J_0$  ( $p_{01} \approx 2.405$ ).

The  $\text{TM}_{010}$  mode has a longitudinal electric field, and is therefore suitable for accelerating charged particles. Strictly speaking, in the  $\text{TM}_{010}$  mode, there is no constraint on the length of the cavity, since the fields are independent of the longitudinal co-ordinate  $s$ . However, if the cavity is to be used for accelerating particles, it would be appropriate to choose length  $L$  so that:

$$\frac{L}{\beta_0 c} = \frac{\pi}{\omega}, \quad (3.133)$$

where  $\beta_0 c$  is the speed of a particle with the nominal (reference) momentum. If this condition is satisfied, then the time taken for a particle to traverse the cavity is one half of an rf period: if the cavity is shorter than this, then a particle entering an accelerating cavity when  $E_s = 0$  exits the cavity before the field returns to zero, and so does not obtain the maximum possible acceleration from the cavity. If the cavity is longer than specified in (3.133), then the field reverses direction before a particle has crossed the cavity, and the particle will start to be decelerated.

To determine the dynamics of a particle in an rf cavity, we need to write down the vector potential that gives the specified fields, and then substitute this potential into the Hamiltonian. We shall consider only the  $\text{TM}_{010}$  mode. A suitable potential is given (in cylindrical polar co-ordinates) by:

$$\mathbf{A} = \left( 0, 0, \frac{E_0}{\omega} J_0(kr) \cos(\omega t + \phi_0) \right). \quad (3.134)$$

Note that the vector potential depends on the time  $t$ . The time is related to the independent variable  $s$  by (2.54), using which we write:

$$\omega t = \frac{ks}{\beta_0} - kz, \quad (3.135)$$

where  $z$  is a dynamical variable (corresponding to the longitudinal coordinate of a particle). Since the reference trajectory has zero curvature, the Hamiltonian for a particle in an rf cavity in the TM<sub>010</sub> mode is:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}} - \frac{q}{P_0} \frac{E_0}{\omega} J_0(kr) \cos\left(\frac{ks}{\beta_0} - kz + \phi_0\right), \quad (3.136)$$

where  $r = \sqrt{x^2 + y^2}$ . This Hamiltonian has an explicit dependence on the independent variable  $s$ : this is a feature that we have not encountered previously, because we have so far considered only static fields. The explicit dependence on the independent variable (in this case, arising from the time dependence of the fields) makes it difficult to integrate the equations of motion. To simplify the problem, we assume that it is possible to ‘average’ the Hamiltonian over the length of the cavity; that is, we consider the Hamiltonian  $\langle H \rangle$  given by:

$$\langle H \rangle = \frac{1}{L} \int_{-L/2}^{+L/2} H ds. \quad (3.137)$$

We find:

$$\langle H \rangle = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}} - \frac{\alpha}{\pi} J_0(kr) \cos(\phi_0 - kz), \quad (3.138)$$

where

$$\alpha = \frac{qV_0}{P_0 c}. \quad (3.139)$$

The cavity voltage  $V_0$  is defined in terms of the electric field amplitude  $E_0$ , and length  $L$  of the cavity:

$$\frac{V_0}{L} = E_0 T. \quad (3.140)$$

The *transit-time factor*  $T$  takes into account the variation in the electric field over the time taken for a particle to pass through the cavity:

$$T = \frac{2\pi\beta_0}{k^2 L^2} \sin\left(\frac{kL}{2\beta_0}\right). \quad (3.141)$$

As usual, to find approximate solutions to the equations of motion, we make the paraxial approximation for the Hamiltonian; in this case, we expand the averaged Hamiltonian  $\langle H \rangle$  to second order in the dynamical

variables. We also drop constant terms, which makes no contribution to the equations of motion. The resulting Hamiltonian is:

$$H_2 = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{\delta^2}{2\beta_0^2\gamma_0^2} + \frac{\alpha}{4\pi} \cos(\phi_0) k^2(x^2 + y^2) - \frac{\alpha}{\pi} \sin(\phi_0) kz + \frac{\alpha}{2\pi} \cos(\phi_0) k^2 z^2. \quad (3.142)$$

Notice that the fourth term gives (for an appropriate phase  $\phi_0$ ) simultaneous focusing in  $x$  and  $y$ : this is a consequence of the azimuthal magnetic field (3.130). The fifth term in the Hamiltonian (3.142) is proportional to the longitudinal co-ordinate  $z$ . This term leads to a change in the energy deviation  $\delta$  of a particle: this change is proportional to the cavity voltage  $V_0$  and depends on the phase  $\phi_0$ , but is independent of the particle co-ordinates. In other words, the change in  $\delta$  resulting from the fifth term in (3.142) will be the same for every particle in the beam. On the other hand, the sixth term in (3.142) leads to a change in the energy deviation  $\delta$  that depends not just on the cavity voltage and phase, but also on the longitudinal co-ordinate of the particle. In effect, this sixth term provides longitudinal focusing.

Solving the equations of motion derived from the Hamiltonian (3.142) gives the linear transfer map for an rf cavity operated in the  $\text{TM}_{010}$  mode:

$$\vec{x}_1 = R_{\text{rf}} \vec{x}_0 + \vec{m}_{\text{rf}}, \quad (3.143)$$

where:

$$R_{\text{rf}} = \begin{pmatrix} c_\perp & s_\perp & 0 & 0 & 0 & 0 \\ -\omega_\perp^2 s_\perp & c_\perp & 0 & 0 & 0 & 0 \\ 0 & 0 & c_\perp & s_\perp & 0 & 0 \\ 0 & 0 & -\omega_\perp^2 s_\perp & c_\perp & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{||} & \frac{1}{\beta_0^2\gamma_0^2} s_{||} \\ 0 & 0 & 0 & 0 & -\beta_0^2\gamma_0^2\omega_{||}^2 s_{||} & c_{||} \end{pmatrix}, \quad (3.144)$$

and:

$$\vec{m}_{\text{rf}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ (1 - \cos(\omega_{||}L)) \frac{\tan(\phi_0)}{k} \\ \beta_0^2\gamma_0^2\omega_{||} \sin(\omega_{||}L) \frac{\tan(\phi_0)}{k} \end{pmatrix}. \quad (3.145)$$

The various quantities appearing in the transfer map are, in the transverse components:

$$c_{\perp} = \cos(\omega_{\perp} L), \quad (3.146)$$

$$s_{\perp} = \frac{\sin(\omega_{\perp} L)}{\omega_{\perp}}, \quad (3.147)$$

$$\omega_{\perp} = k \sqrt{\frac{\alpha \cos(\phi_0)}{2\pi}}, \quad (3.148)$$

and in the longitudinal components:

$$c_{\parallel} = \cos(\omega_{\parallel} L), \quad (3.149)$$

$$s_{\parallel} = \frac{\sin(\omega_{\parallel} L)}{\omega_{\parallel}}, \quad (3.150)$$

$$\omega_{\parallel} = \frac{k}{\beta_0 \gamma_0} \sqrt{\frac{\alpha \cos(\phi_0)}{\pi}}. \quad (3.151)$$

From the zeroth-order terms in the transfer map (the components of  $\vec{m}_{\text{rf}}$ ), we see that a particle that has initial longitudinal co-ordinate  $z = 0$  experiences a change in energy deviation:

$$\Delta\delta = \beta_0^2 \gamma_0^2 \omega_{\parallel} \sin(\omega_{\parallel} L) \frac{\tan(\phi_0)}{k} \approx \frac{qV_0}{P_0 c} \frac{kL}{\pi} \sin(\phi_0), \quad (3.152)$$

where the approximation is valid for  $\omega_{\parallel} L \ll 1$ . If the cavity length is chosen such that  $kL \approx \pi$  (i.e. the length is one half of the rf wavelength), then:

$$\Delta\delta \approx \frac{qV_0}{P_0 c} \sin(\phi_0). \quad (3.153)$$

The change in the energy deviation is approximately equal to the particle charge multiplied by the cavity voltage (defined by (3.140)) scaled by  $P_0 c$ , and multiplied by the sine of the phase of the rf: this is what we would expect from the physical significance of the cavity voltage and phase.

As usual, by neglecting terms of third order and higher in the Hamiltonian, we have lost some properties of the dynamics: in particular, the transverse focusing in the linear transfer map (3.144) and (3.145) is independent of the energy deviation. That is, there is no chromaticity in the transfer map that we have derived. This is the same situation that occurred when we considered quadrupole and solenoid magnets. In those cases, we derived a linear map for the transverse variables by making use of the fact that the energy deviation was constant, and could therefore be treated as a parameter in the equations of motion, rather than as a variable. For an rf cavity, however, we cannot assume that the energy deviation is constant,

and therefore we cannot use the same approach. Proper treatment of the dynamics in an rf cavity, including higher-order effects, requires more advanced tools than what we have developed in this chapter. Fortunately, in many practical cases, a simple linear approximation to the dynamics in an rf cavity is sufficient to describe the observable effects on the beam.

Finally, we note that the transfer map given in Eqs. (3.144) and (3.145) describes the acceleration of a particle by a change in energy deviation,  $\delta$ . The reference momentum does not change. In a linac, particles can undergo significant change in energy, which means that if the acceleration is described by a change in energy deviation, then the value of the variable  $\delta$  can eventually become very large. This is generally undesirable, since descriptions of the dynamics often depend on approximations that are valid for small values of the dynamical variables. To avoid the energy deviation becoming too large, a change in the reference momentum can be made. For example, if a particle is accelerated from momentum  $10 \text{ MeV}/c$  to  $15 \text{ MeV}/c$ , then if the initial reference momentum is  $10 \text{ MeV}/c$  (so that  $\delta = 0$  initially), without any change in reference momentum, the final value of the energy deviation will be  $\delta \approx 0.5$ . Changing the reference momentum to  $15 \text{ MeV}/c$  will restore  $\delta = 0$ . However, changing the reference momentum will also change the values of the transverse momenta  $p_x$  and  $p_y$ , since these are defined in terms of reference momentum.

A change in reference momentum can be represented by a transfer map:

$$\vec{x}_1 = R_{\Delta P} \vec{x}_0 + \vec{m}_{\Delta P}, \quad (3.154)$$

where:

$$R_{\Delta P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{P_0}{P_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{P_0}{P_1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{P_0}{P_1} \end{pmatrix}, \quad (3.155)$$

and:

$$\vec{m}_{\Delta P} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{\beta_1} \left( \frac{\gamma_0}{\gamma_1} - 1 \right) \end{pmatrix}. \quad (3.156)$$

Here,  $P_0$  and  $P_1$  are the initial and final reference momenta;  $\gamma_0$  and  $\gamma_1$  are the relativistic factors for particles with momenta  $P_0$  and  $P_1$  respectively; and  $\beta_1 c$  is the velocity of a particle with momentum  $P_1$ . The co-ordinates  $x$ ,  $y$  and  $z$  are unchanged.

In contrast to the other transfer maps derived in this chapter, the transfer map given by (3.155) and (3.156) is not symplectic. As we saw in Chapter 2, symplecticity is associated with conserved quantities in a dynamical system. In accelerator beam dynamics, a particularly important conserved quantity is the density of particles in phase space. If we apply a transformation (such as a change in reference momentum) that is not symplectic, then there is no requirement for the density of particles in phase space to be conserved. In fact, an increase in reference momentum results in a decrease in the volume of phase space occupied by a collection of particles. This effect, which is important when particles are accelerated in a linac, is known as *adiabatic damping*, and will be discussed further in Chapter 4.

### 3.7 Spin Dynamics

In addition to their co-ordinates and momenta, fundamental particles have a further degree of freedom that is of importance for many accelerator applications: spin. The spin of a particle describes aspects of its behaviour that would be expected from some internal (or intrinsic) angular momentum. In colliders, for example, there are often specifications on the spin polarisation of the beams at the interaction point, set by the physics goals. In charged particles, spin is associated with each particle acting as the source of a magnetic field, characterised by a magnetic moment that is an intrinsic property of the type of particle concerned. The magnetic moment may be measured to extremely high precision by observation of the precession frequency of the spin of the particle in a uniform magnetic field within a storage ring: this technique forms the basis of tests of the Standard Model of particle physics, by comparison of the measured value for a particular type of particle with the theoretical expected value (perhaps most notably in the case of the muon [Abi *et al.* (2021); Albahri *et al.* (2021)]). Spin is also of significance in synchrotron light sources and other electron storage rings, where it has some (usually small) impact on the beam lifetime, and can be used to precisely determine the beam energy, using resonant depolarisation (see Section 7.5) [Zholentz *et al.* (1980); Melissinos (1995); Klein *et al.* (1997); Minty and Zimmermann (2003a)].

Formally, spin should be treated using the rules of quantum mechanics;

however, the evolution of the spin of a particle in the electromagnetic fields in an accelerator beam line can be described, to a good approximation, by a classical model with certain parameters derived from quantum theory. We do not attempt to give a thorough analysis of spin dynamics in an accelerator: for a summary of the important results and further references see (for example) [Roser *et al.* (1999)]. Spin polarisation in storage rings is treated in more depth by [Barber (1999); Lee (1997)]. Here, we limit ourselves to an outline of the basic concepts. Of particular interest is the *Thomas–Bargmann–Michel–Telegdi* (or *Thomas–BMT*) equation [Thomas (1927); Bargmann *et al.* (1959)] that describes the evolution of the spin of a relativistic particle in an electromagnetic field; but before stating this equation, we first give some basic definitions.

The magnetic moment  $\boldsymbol{\mu}$  generated by a current density  $\mathbf{J}$  is defined as:

$$\boldsymbol{\mu} = \frac{1}{2} \int \mathbf{r} \times \mathbf{J} dV, \quad (3.157)$$

where the integral is taken over all space, and it is assumed that the current density is constant in time. For a circular wire loop of radius  $r$  carrying current  $I$ , the magnetic moment from (3.157) is:

$$\boldsymbol{\mu} = IA, \quad (3.158)$$

where  $\mathbf{A}$  (not to be confused with the magnetic vector potential) is a vector with magnitude equal to the area of the loop, and perpendicular to the plane within which the loop lies. The magnetic field from an infinitesimal magnetic moment located at the origin is:

$$\mathbf{B} = \frac{\mu_0}{4\pi} \left( \frac{3(\boldsymbol{\mu} \cdot \mathbf{r})\mathbf{r}}{|\mathbf{r}|^5} - \frac{\boldsymbol{\mu}}{|\mathbf{r}|^3} \right). \quad (3.159)$$

Now consider a particle with charge  $q$  and mass  $m$  travelling around a circular loop of radius  $r_0$  with speed  $v$ . The current associated with the motion of the particle is:

$$I = \frac{qv}{2\pi r_0}. \quad (3.160)$$

The angular momentum of the particle is:

$$\mathbf{L} = m\mathbf{r}_0 \times \mathbf{v}, \quad (3.161)$$

where the direction of  $\mathbf{L}$  is perpendicular to the plane of the motion of the charge (see Fig. 3.3). Combining (3.158), (3.160) and (3.161), the magnetic moment generated by the motion of the charge can be written in terms of its angular momentum:

$$\boldsymbol{\mu} = \frac{q}{2m} \mathbf{L}. \quad (3.162)$$

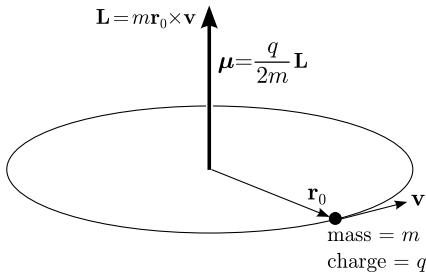


Fig. 3.3 Magnetic moment generated by a point particle with mass  $m$  and charge  $q$  following a circular trajectory of radius  $r_0$  with speed  $v$ . The particle has angular momentum  $\mathbf{L}$  and magnetic moment  $\mu$ .

Quantum mechanics predicts an intrinsic angular momentum (spin) for electrons with magnitude:

$$S = \frac{\hbar}{2}, \quad (3.163)$$

where  $\hbar$  is Planck's constant divided by  $2\pi$ . An electron in free space cannot be modelled as a charged particle following a circular orbit; nevertheless, a relationship very similar to (3.162) exists between the magnetic moment of the electron and its spin angular momentum:

$$\mu = g \frac{q}{2m} S, \quad (3.164)$$

where  $g$  is a constant, known as the *gyromagnetic ratio*. For a Dirac fermion, it is expected (from the Dirac equation) that  $g = 2$ ; however, radiative corrections from quantum electrodynamics lead to some small difference from this value. The difference is measured by the *anomalous magnetic moment*,  $G$ :

$$G = \frac{g - 2}{2}. \quad (3.165)$$

Electrons and muons have small anomalous magnetic moments that can be predicted from quantum field theory. Particles such as protons and deuterons, which are not fundamental but contain constituent particles, also have magnetic moments and spin angular momentum related by (3.164), but the anomalous magnetic moments for such particles are generally much larger than for the leptons. The values of the anomalous magnetic moment for some particles are shown in Table 3.1.

Let us return for a moment to our simple model of a current loop, and consider what happens when the loop is placed in a (uniform) magnetic

Table 3.1 Angular momenta and anomalous magnetic moments of some types of particle [Beringer, J. et al. (Particle Data Group) (2012)].

particle type	spin angular momentum, $ \mathbf{S} $	anomalous magnetic moment, $G$
electron	$\frac{1}{2}\hbar$	$1.159652 \times 10^{-3}$
muon	$\frac{1}{2}\hbar$	$1.165921 \times 10^{-3}$
proton	$\frac{1}{2}\hbar$	1.792847
deuteron	$\hbar$	-0.142987

field. If the magnetic moment of the loop is exactly parallel with the magnetic field, then there will be a force on each section of the loop acting radially outwards from the centre of the loop. If the loop is constructed from rigid wire, then the magnetic forces will be balanced by the mechanical tension in the wire, and nothing will happen. However, if the loop is tilted so that there is some angle between the magnetic moment  $\boldsymbol{\mu}$  and the (external) magnetic field  $\mathbf{B}$ , then there will be a net torque  $\mathbf{T}$  acting on the loop, given by:

$$\mathbf{T} = \boldsymbol{\mu} \times \mathbf{B}. \quad (3.166)$$

Assuming that the current in the loop consists of a single particle of charge  $q$  and mass  $m$ , the equation of motion for the angular momentum of the charge will be:

$$\frac{d\mathbf{L}}{dt} = \mathbf{T} = \boldsymbol{\mu} \times \mathbf{B} = \frac{q}{2m} \mathbf{L} \times \mathbf{B}. \quad (3.167)$$

Since the magnetic moment is parallel to the angular momentum, the net result will be a precession of the angular momentum of the charge around the magnetic field line passing through the (small) current loop.

In the case of a charged particle in an external magnetic field, a similar precession (called the *Larmor precession*) of the angular momentum takes place:

$$\frac{d\mathbf{S}}{dt} = g \frac{q}{2m} \mathbf{S} \times \mathbf{B}. \quad (3.168)$$

In the case of a relativistic particle in an accelerator, there are complicated effects associated with the transformations of the various quantities from the rest frame of the particle to the laboratory frame. We do not discuss the transformations in detail, but simply give the final result:

$$\frac{d\mathbf{S}'}{dt} = \boldsymbol{\Omega}_0 \times \mathbf{S}', \quad (3.169)$$

where  $\mathbf{S}'$  is the spin vector of the particle *in the rest frame of the particle*,  $t$  is the time in the laboratory frame and the vector  $\boldsymbol{\Omega}$  is given by:

$$\boldsymbol{\Omega}_0 = -\frac{q}{\gamma m} \left( (1 + \gamma G) \mathbf{B}_\perp + (1 + G) \mathbf{B}_\parallel + \gamma \left( G + \frac{1}{\gamma + 1} \right) \frac{\mathbf{E} \times \mathbf{v}}{c^2} \right). \quad (3.170)$$

Equation (3.169) is the Thomas–BMT equation. The expression for the precession frequency (3.170) assumes a Cartesian co-ordinate system (i.e. a straight reference trajectory).  $\gamma$  is the relativistic factor of the particle which is moving with velocity  $\mathbf{v}$ ,  $\mathbf{B}_\perp$  and  $\mathbf{B}_\parallel$  are (respectively) the components of the magnetic field perpendicular to and parallel to the velocity, and  $\mathbf{E}$  is the electric field.

Consider a particle following a circular path in a uniform magnetic field  $\mathbf{B}$ . If the particle has momentum  $\mathbf{P}$ , then the equation of motion for the particle can be written:

$$\frac{d\mathbf{P}}{dt} = \mathbf{F} = \boldsymbol{\omega}_0 \times \mathbf{P}, \quad (3.171)$$

where  $\mathbf{F}$  is the Lorentz force (2.2), and:

$$\boldsymbol{\omega}_0 = -\frac{q}{\gamma m} \mathbf{B}. \quad (3.172)$$

The evolution of the spin is given by (3.169), with  $\boldsymbol{\Omega}_0$  in this case being:

$$\boldsymbol{\Omega}_0 = -\frac{q}{\gamma m} (1 + \gamma G) \mathbf{B}. \quad (3.173)$$

Comparing (3.172) and (3.173), we see that if  $G = 0$  (which unfortunately does not correspond to any physical situation) then the momentum and the spin rotate at the same rate. In other words, in the time taken for the particle to complete one orbit, the spin (measured in the rest frame of the particle) precesses through  $360^\circ$ . In beam dynamics, it is usually convenient to use a co-ordinate system based on a curved reference trajectory, with the curvature at any point determined by the local dipole field at that point (and the reference momentum of the beam). In the co-ordinate system where the particle follows the reference trajectory, both the momentum and the spin of the particle (in the case that  $G = 0$ ) are stationary.

In general, we can define a co-ordinate system in which the reference trajectory is defined by the path of a particle with charge  $q$ , mass  $m$  and relativistic factor  $\gamma$  through a region with specified electric and magnetic fields. In this co-ordinate system, the spin evolves according to:

$$\frac{d\mathbf{S}'}{dt} = \boldsymbol{\Omega} \times \mathbf{S}', \quad (3.174)$$

where:

$$\boldsymbol{\Omega} = -\frac{q}{\gamma m} \left( \gamma G \mathbf{B}_\perp + (1+G) \mathbf{B}_\parallel + \gamma \left( G - \frac{1}{\gamma^2 - 1} \right) \frac{\mathbf{E} \times \mathbf{v}}{c^2} \right). \quad (3.175)$$

As usual,  $\mathbf{S}'$  is the spin measured in the rest frame of the particle.

Consider again the case of a particle following a circular path in a uniform magnetic field. We shall choose the co-ordinate system such that the particle moves along the  $s$  axis, and the magnetic field is in the  $y$  direction. Suppose the particle has a spin vector  $\mathbf{S}'$  that has azimuthal angle  $\phi$  about the  $y$  axis. Using (3.174) and (3.175), the angle  $\phi$  evolves as:

$$\frac{d\phi}{dt} = \frac{qB}{m} G. \quad (3.176)$$

Using (3.45):

$$B\rho = \frac{P}{q}, \quad (3.177)$$

where  $\rho$  is the radius of the trajectory of the particle and  $P$  is the momentum of the particle, and:

$$T_0 = \frac{2\pi\rho}{\beta c} = \frac{2\pi\rho\gamma m}{P}, \quad (3.178)$$

where  $T_0$  is the revolution period and  $\beta c$  is the velocity of the particle, we find that the change  $\Delta\phi$  in the azimuthal angle of the spin over one revolution period is:

$$\Delta\phi = 2\pi\nu_s, \quad (3.179)$$

where the *spin tune* is:

$$\nu_s = \gamma G. \quad (3.180)$$

The spin tune multiplied by  $2\pi$  gives the precession angle of the spin vector in one orbital period. Although we have considered the specific case of a particle moving in a uniform magnetic field, the result is the same if we consider a particle in a storage ring. It is interesting that the spin tune is a function of the particle energy (i.e. of the relativistic factor  $\gamma$ ) alone and a fundamental property of the particle (the anomalous magnetic moment  $G$ ). For electrons, with anomalous magnetic moment  $G \approx 1.16 \times 10^{-3}$ , integer spin tunes in a storage ring occur when the beam energy is an integer multiple of 440 MeV.

For a particle in a storage ring, we have to take into account not just the dipole fields, but also the fields of quadrupoles, sextupoles and other magnets that may appear, such as solenoids. In the ideal case, a particle will

follow a trajectory defined by the dipoles, and will pass through quadrupole and higher-order multipole magnets along the magnetic axis of each magnet, where the field is zero. Then, assuming there are no other types of magnet present, the only fields seen by a particle will be the dipole fields from the main bending magnets. The spin precession will be the same as in a uniform magnetic field. In general though, particles will follow paths around the storage ring that do not exactly coincide with the reference trajectory: then, the spin will be affected by the fields of other magnets. On average, we might expect the effects of small deflections of the spin from quadrupoles and other magnets over many turns of the storage ring to cancel out. However, there are important cases where, either by design or by accident, this does not happen. Understanding the evolution of the spin of particles in a bunch is often of particular importance in colliders, where the spin polarisation of the beams at the interaction point can be significant for the physics studies. Various techniques have been developed both for accurate and efficient analysis and modelling of spin dynamics in accelerators, and for the manipulation of the spin polarisation in different situations. We shall first briefly consider some of the general features of spin dynamics and systems that can be used to control the spin polarisation, and then conclude this section with a short discussion of one particular technique for modelling spin dynamics.

Usually, it is not the spin of a single individual particle in an accelerator that is of concern as much as the average spin of a (large) number of particles in a bunch. The *spin polarisation* of a bunch is defined as the average of the spin vectors of all particles within the bunch. The magnitude of the spin polarisation vector measures the extent to which the spin vectors of the particles are aligned. If the spins of the particles in a bunch are all pointing in one particular direction, then the magnitude of the spin polarisation will be equal to 1: the beam has a spin polarisation of 100%. If a bunch with 100% spin polarisation is injected into a storage ring, there are various processes that will lead to a reduction in the polarisation over time: the rate of depolarisation, and the final value of the spin polarisation, depends on the construction of the storage ring and the initial polarisation state.

In the simplest case of a planar storage ring, in which particles on the reference trajectory see only the vertical magnetic field in the main bending magnets, spin depolarisation of a bunch can be minimised if the bunch initially has a spin polarisation vector that is vertical. This can be understood as follows: We have seen that in a magnetic field, the spin vector of a particle precesses about an axis parallel to the field. This means that as a

particle moves around a storage ring, if the spin vector is initially vertical, then the spin vector will remain vertical as long as the only magnetic fields seen by the particle are the fields in the main dipoles that bend the particle trajectory in the horizontal plane. However, if the spin of a particle has some angle to the vertical, then the spin will precess around a vertical axis at a rate determined by the energy of the particle (i.e. by the spin tune). In general, particles within a bunch will have some spread in energies. Then, different particles within a bunch will have different spin tunes. As a result, particles with spin vectors initially aligned at some angle  $\theta$  to the vertical will eventually have spin vectors distributed around a cone with opening angle  $2\theta$ . If the spin vectors are uniformly distributed around this cone, then the magnitude of the spin polarisation vector will be  $\cos(\theta)$ . If  $\theta = 0$ , then no depolarisation will occur; if  $\theta = \pi/2$ , then the beam could become fully depolarised.

In practice, in the situation just described, the spin vectors of the particles in a bunch are unlikely to become uniformly distributed around a cone. This is because, in a synchrotron storage ring, the fields in the rf cavities cause particle energies to oscillate (see the discussion of synchrotron motion in Chapter 5). As a result, particle spins will initially start to spread out around a cone; but after half a synchrotron period, the particle spins will move back until the spins are once more all aligned. The maximum amount of depolarisation depends on the energy spread in the beam.

Things become more complicated if, as will generally be the case, particles see magnetic fields other than the purely vertical fields from the main dipole magnets. For example, a particle moving through a quadrupole with some vertical offset from the magnetic axis will see a horizontal magnetic field. Depending on the orientation of the spin vector of the particle, this field could move the spin vector towards the vertical, away from the vertical, or could rotate the spin vector through some (small) angle around the vertical. When a bunch of particles has some energy spread and some distribution of spin vectors, the evolution of the spin polarisation can be difficult to predict. However, depolarisation effects usually become more pronounced if the spin tune is close to an integer. To understand why this should be the case, consider a storage ring with a solenoid field at one location. Suppose a particle initially enters the solenoid with a vertical spin. The solenoid field will rotate the spin through some angle away from the vertical axis. Then, as the particle makes a turn around the storage ring, the spin precesses around the vertical axis. If the spin tune is a half-integer, when the particle returns to the location of the solenoid, the solenoid field

will rotate the spin vector so that it is aligned more vertically. However, if the spin tune is an integer, then the solenoid field will rotate the spin vector further away from the vertical. In other words, if the spin tune is an integer, then the spin precession ‘resonates’ with the solenoid field, leading to a large variation in the orientation of the spin. Near a half-integer spin tune, on the other hand, the spin vector of a particle will be little affected by solenoid fields in the storage ring. This does not mean, however, that a half-integer spin tune is always the best place to operate a storage ring to avoid depolarisation effects: there can be complicated interactions between particle spins and trajectories that lead to depolarisation from multipole fields (for example, from quadrupoles). Usually, though, depolarisation effects are strongest at integer spin tunes.

Depolarisation effects in storage rings can usually be reduced by orienting the spins of particles so that they are as close as possible to the vertical axis; however, in colliders it is often desirable to have beams with longitudinal spin polarisation, i.e. with spin vectors of the particles aligned parallel (or antiparallel) to the direction in which the particles are travelling. One possible approach to satisfying both criteria is to rotate the spins of the particles close to the interaction point. For example, a solenoid can be used to rotate the polarisation vector from the vertical to the (transverse) horizontal direction; then a dipole (in a final bend before the interaction point) can rotate the polarisation from the horizontal to the longitudinal. On the other side of the interaction point, a dipole followed by a solenoid can be used to rotate the polarisation back to the vertical direction. Various combinations of magnets can be used to manipulate the spin polarisation in different ways. Accelerators requiring control of the spin polarisation often include devices designed to rotate the spins of particles while leaving their trajectories unchanged: a device with this property is known as a *Siberian snake*. Siberian snakes can be constructed from solenoid and/or dipole fields. Some skill is required to design a system that has the desired effects on the spin polarisation, taking into account also the effects on the beam trajectory (and other aspects of the beam dynamics).

One way of investigating the spin dynamics in a given accelerator system is to track the spin of a particle (or a collection of particles) through the beam line in much the same way as can be done for the phase space variables. This involves integrating the Thomas–BMT equation (3.174) with some specified initial conditions, and using appropriate values for the electric and magnetic fields seen by the particle along its trajectory. The integration is not entirely straightforward, since even in a region of uniform

magnetic field (such as in a solenoid or dipole magnet) the direction of the velocity of a particle can change, which will change the components of the field  $\mathbf{B}_\perp$  and  $\mathbf{B}_\parallel$  perpendicular and parallel to the velocity, respectively. However, for high energy beams where particles remain close to the reference trajectory, it is often possible to approximate the motion (as far as the spin dynamics are concerned) by assuming that the direction and magnitude of the velocity are constant. In quadrupole and higher-order multipole magnets, we also need to assume that the magnetic field is approximately constant over the integration step: this will be the case if the step size of the integration is small compared with the betatron wavelength (which will be defined in Chapter 4). Then, the precession vector  $\boldsymbol{\Omega}$  will be constant, and it is possible to apply an analytical solution to (3.174).

For constant  $\boldsymbol{\Omega}$ , the solution to (3.174) can be written:

$$\mathbf{S}'(t) = \mathbf{S}'_0 \cos(\Omega t) + \hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\Omega}} \cdot \mathbf{S}'_0)(1 - \cos(\Omega t)) + \hat{\boldsymbol{\Omega}} \times \mathbf{S}'_0 \sin(\Omega t), \quad (3.181)$$

where  $\mathbf{S}'_0$  is the spin vector  $\mathbf{S}'$  at  $t = 0$ ,  $\Omega = |\boldsymbol{\Omega}|$  is the magnitude of the precession vector  $\boldsymbol{\Omega}$ , and  $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}/\Omega$  is a unit vector in the direction of the precession vector. In tracking particles along an accelerator beam line, it is conventional to use the path length  $s$  along the reference trajectory as the independent variable, rather than the time  $t$ . In (3.181), the time interval  $t$  is related to the path length interval  $\Delta s$  by  $t = \Delta s/\beta_0 c$ , where  $\beta_0 c$  is the velocity of the reference particle (i.e. the velocity of a particle with the reference momentum  $P_0$ ). In a specified magnetic field, the field components  $\mathbf{B}_\perp$  and  $\mathbf{B}_\parallel$  can be found from:

$$\mathbf{B}_\parallel = \frac{(\mathbf{B} \cdot \mathbf{v})\mathbf{v}}{|\mathbf{v}|^2}, \quad (3.182)$$

$$\mathbf{B}_\perp = \mathbf{B} - \mathbf{B}_\parallel, \quad (3.183)$$

where  $\mathbf{v}$  is the velocity of the particle. Equation (3.181) provides a transfer map for the spin of a particle, analogous to the transfer maps we have derived for the phase space variables in various beam line components earlier in this chapter. It should be remembered that the formula in (3.181) depends on approximating the particle velocity and the magnetic field as constant: in some cases, it may be necessary to divide an accelerator component into a number of short lengths in order for the approximations to be valid.

Spin tracking based on the formula (3.181) is rather a ‘brute force’ approach to modelling the spin dynamics, and can be computationally expensive, especially if tracking many particles over many turns in a large

storage ring. More efficient techniques for the analysis of spin dynamics in storage rings have been developed, often based on the concept of the *invariant spin field* [Derbenev and Kondratenko (1973); Yokoya (1986); Barber *et al.* (1999)]. This is defined as follows: Suppose that, at some location  $s = s_0$  in a storage ring, a particle has phase space vector  $\vec{x}_0$ . In general, after a complete turn around the storage ring, the values of the phase space variables will change so that the new phase space vector  $\vec{x}_1$  is different from  $\vec{x}_0$ . If the motion is stable, then after some (possibly large) number of turns  $M$ , the particle will return to a point in phase space close to its starting point, so that:

$$\vec{x}_M \approx \vec{x}_0. \quad (3.184)$$

If we compare the spin vector of the particle after  $M$  turns to the initial spin vector, we shall find that it is rotated about some axis defined by a unit vector  $\hat{\mathbf{n}}(\vec{x}, s_0)$ . The invariant spin field at a particular location  $s_0$  in the storage ring consists of the set of vectors  $\hat{\mathbf{n}}(\vec{x}, s_0)$  for all points  $\vec{x}$  in phase space. It can be shown that, for a collection of particles, the invariant spin field averaged over all the particles  $\langle \hat{\mathbf{n}}(\vec{x}, s_0) \rangle$  places an upper limit on the spin polarisation of the beam over the long term, i.e. after many turns of the ring. In other words, for any initial polarisation of the beam, after a large number of turns the magnitude of the polarisation will be less than the magnitude of  $\langle \hat{\mathbf{n}}(\vec{x}, s_0) \rangle$ . Thus, computing the invariant spin field can provide useful information on the spin dynamics in a storage ring without the need for long-term tracking of particle trajectories and spins.

Finally, we mention that for electron and positron storage rings, synchrotron radiation can have significant effects on the spin polarisation: these effects are discussed further in Chapter 7.

## Chapter 4

# Linear Optics in Uncoupled Beam Lines

In Chapter 3, we derived the linear transfer maps for a number of components commonly used in accelerator beam lines, including drift spaces, dipoles, quadrupoles, solenoids and rf cavities. The linear transfer maps (represented by matrices) provide a description of the first-order dynamics of particles in the individual components within a beam line. While such a description is essential for the effective design and operation of accelerators, it can also be rather cumbersome, especially in beam lines consisting of a large number of components. In practice, it is convenient to have a description of the dynamics at a higher level, which allows important features of the beam dynamics to be readily understood without the need to worry about the detailed motion of particles in individual components. Significant properties of the beam that are routinely required include the trajectory of the beam and the variation of the beam size along a beam line: the analysis of these properties constitutes the study of beam optics and is the subject of the present chapter.

Throughout this chapter, we assume that there is no coupling in the beam line, in other words, that the horizontal, vertical and longitudinal dynamics can all be treated independently. This greatly simplifies the analysis of a beam line, and makes it as straightforward as possible to introduce a number of important new concepts. Linear coupling is discussed in detail in Chapter 5.

To begin our treatment of beam optics, we shall consider the special case of a periodic beam line; i.e. a beam line consisting of a unit cell that is repeated indefinitely. The unit cell itself consists of some specified sequence of components. Periodic beam lines are used frequently in accelerator systems; for example, a storage ring can be viewed as a periodic beam line in which the unit cell consists of the entire ring. The advantage of discussing

beam optics in periodic beam lines first is: the periodicity imposes constraints that lead naturally to the definition of quantities (such as the beta function) that describe the variation of the beam size along the beam line. Of course, not all beam lines are periodic; but having defined the important quantities for the periodic case, it is possible, without great difficulty, to make the generalisation to non-periodic beam lines.

#### 4.1 A FODO Lattice

Consider a beam line consisting of equally spaced quadrupoles, with the quadrupoles alternately focusing (F quadrupoles) and defocusing (D quadrupoles) in the horizontal plane. If all the F quadrupoles have the same gradient and all the D quadrupoles have the same gradient, then we can consider the beam line to be periodic, with a unit cell consisting of a pair of quadrupoles each followed by a drift space. Such a cell is commonly called a *FODO cell* (where the ‘O’ refers to a drift space between two quadrupoles); a beam line consisting of repeated FODO cells is shown in Fig. 4.1.

The motion of individual particles along the beam line can be described using the transfer matrices for drift spaces and quadrupoles. As we shall see, particles can perform transverse (betatron) oscillations about the reference trajectory as they move along a periodic beam line: but the periodicity of the particle oscillations will be different from the periodicity of the magnets in the beam line. Also, the amplitude and wavelength of the betatron oscillations will vary along the beam line. This leads to the fact that under certain conditions the variation of the size of a beam consisting of many particles will be periodic, with the same periodicity as the lattice. Our goal

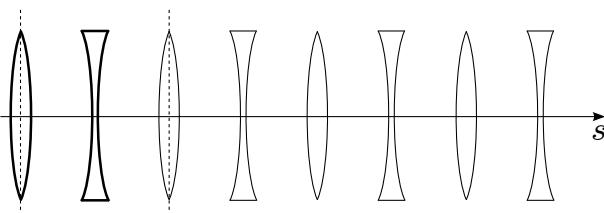


Fig. 4.1 A FODO beam line, consisting of equally spaced, alternating horizontally focusing and horizontally defocusing quadrupole magnets. A single cell (shown between the dashed vertical lines) consists of a horizontally focusing quadrupole and a horizontally defocusing quadrupole together with the drift spaces following each magnet.

in this chapter is to derive the conditions that must be satisfied for the beam size to vary with the same periodicity as the components in the beam line, and to understand how (when the appropriate conditions are met) the beam size depends on the strengths and spacings of the quadrupole magnets. In this section, we shall begin by looking at some of the relevant features of the particle motion in a simple FODO beam line.

To simplify the analysis, we shall use the ‘thin lens’ approximation for the quadrupoles; that is, we shall take the limit that the length  $L_q$  of a quadrupole goes to zero, but with the product of the quadrupole gradient and the length remaining at a fixed value:

$$L_q \rightarrow 0, \quad (4.1)$$

$$k_1 L_q \rightarrow \frac{1}{f}, \quad (4.2)$$

where

$$k_1 = \frac{q}{P_0} \frac{\partial B_y}{\partial x}, \quad (4.3)$$

and  $f$  is a constant. In this limit, the transfer matrix for the horizontally focusing quadrupole (the F quadrupole) becomes, from (3.96):

$$R_F = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.4)$$

A particle entering the quadrupole with  $x = x_0$ ,  $p_x = 0$  is deflected so that at the exit of the quadrupole  $p_x = -x_0/f$ . Since  $p_x$  is (approximately) equal to the angle between the particle trajectory and the reference trajectory, we see that the particle will cross the reference trajectory a distance  $f$  downstream of the quadrupole. This is independent of  $x_0$ . The effect of the quadrupole is therefore to ‘focus’ particles entering the magnet on trajectories parallel to the reference trajectory, to a single point a distance  $f$  from the magnet. In the horizontal plane, the magnet behaves as a focusing lens with focal length  $f$ . In the vertical plane, the magnet acts as a defocusing lens, with the same focal length.

Let us take the horizontally defocusing quadrupole (the D quadrupole) in the FODO cell to have a field gradient of the same magnitude as the F quadrupole, but with opposite sign. The transfer matrix  $R_D$  for the D

quadrupole is then the same as for the F quadrupole (4.4), but with  $f$  replaced by  $-f$ . Finally, we take the drifts following each quadrupole to have the same length  $L$ . The transfer matrix for each drift is, from (3.29):

$$R_{\text{drift}} = \begin{pmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\beta_0^2 \gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (4.5)$$

We can now construct the transfer matrix  $R$  for a full FODO cell simply by multiplying the matrices for the individual components. Before we do so, we note that the result will depend on the choice of starting point for the cell. Since the beam line is periodic, we are free to choose any point in the cell as the starting point. We shall take the starting point to be the point at the entrance to the F quadrupole. If a particle at this point has phase space vector  $\vec{x}_0$ , then after the quadrupole the phase space vector is  $\vec{x}_1$ , given by:

$$\vec{x}_1 = R_F \vec{x}_0. \quad (4.6)$$

After the drift following the F quadrupole, the phase space vector is  $\vec{x}_2$ :

$$\vec{x}_2 = R_{\text{drift}} \vec{x}_1 = R_{\text{drift}} R_F \vec{x}_0. \quad (4.7)$$

Notice that we include the transfer matrix for a given component by *pre-multiplying* the transfer matrix for the beam line up to that component; that is, the transfer matrices appear in the reverse order to the components in the beam line. The full transfer matrix for the FODO cell is:

$$R = R_{\text{drift}} R_D R_{\text{drift}} R_F. \quad (4.8)$$

Since there is no coupling between the horizontal, vertical and longitudinal motion,  $R$  is block diagonal:

$$R = \begin{pmatrix} R_x & 0 & 0 \\ 0 & R_y & 0 \\ 0 & 0 & R_z \end{pmatrix}. \quad (4.9)$$

The  $2 \times 2$  matrices along the block diagonal can be written explicitly in terms of the focal length  $f$  of the quadrupoles and the drift length  $L$ . The horizontal motion is determined by the matrix:

$$R_x = \begin{pmatrix} 1 - \frac{L}{f} \left(1 + \frac{L}{f}\right) \frac{L}{f} (2f + L) \\ -\frac{L}{f^2} & 1 + \frac{L}{f} \end{pmatrix}. \quad (4.10)$$

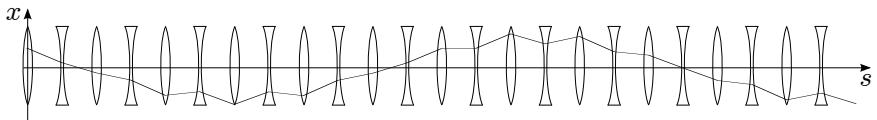


Fig. 4.2 Horizontal trajectory of a particle in a FODO beam line. The quadrupole magnets have focal length  $\pm\sqrt{2}$  m, and the drift spaces have length 1 m. The initial horizontal co-ordinate and momentum are  $x = 1$  mm and  $p_x = 0$ .

The vertical motion is determined by:

$$R_y = \begin{pmatrix} 1 + \frac{L}{f} \left(1 - \frac{L}{f}\right) \frac{L}{f} (2f - L) \\ -\frac{L}{f^2} & 1 - \frac{L}{f} \end{pmatrix}, \quad (4.11)$$

and the transfer matrix for the longitudinal motion is:

$$R_z = \begin{pmatrix} 1 & \frac{2L}{\beta_0^2 \gamma_0^2} \\ 0 & 1 \end{pmatrix}. \quad (4.12)$$

Even though we have made a simplifying approximation by representing the quadrupoles as thin lenses, the transfer matrix for the FODO cell looks fairly complicated. However, the particle dynamics can be described relatively simply. As an example, consider the particular case where  $L = 1$  m, and  $f = \sqrt{2}$  m. For now, we consider only the transverse dynamics: the longitudinal co-ordinate and energy deviation do not affect the horizontal or vertical motion, at least in the linear approximation where we neglect chromaticity. We take a particle with initial co-ordinates  $x = y = 1$  mm, and  $p_x = p_y = 0$ , and track this particle along the beam line by applying the appropriate transfer matrix for each successive component. The horizontal trajectory through 12 cells is shown in Fig. 4.2. We see that, as expected, the particle is deflected towards the reference trajectory by each horizontally focusing magnet, and away from the reference trajectory by each horizontally defocusing magnet; overall, the trajectory resembles a sinusoidal oscillation, though the period of the motion is certainly different from the periodicity of the magnetic lattice. We would find similar behaviour in the vertical plane if we plotted the vertical trajectory (though the horizontally focusing quadrupoles will give vertical defocusing, and vice versa).

We can construct phase space portraits by plotting the horizontal co-ordinate and momentum  $(x, p_x)$  and the vertical co-ordinate and momentum  $(y, p_y)$  of the particle at the exit of each successive FODO cell. The results are shown in Fig. 4.3.

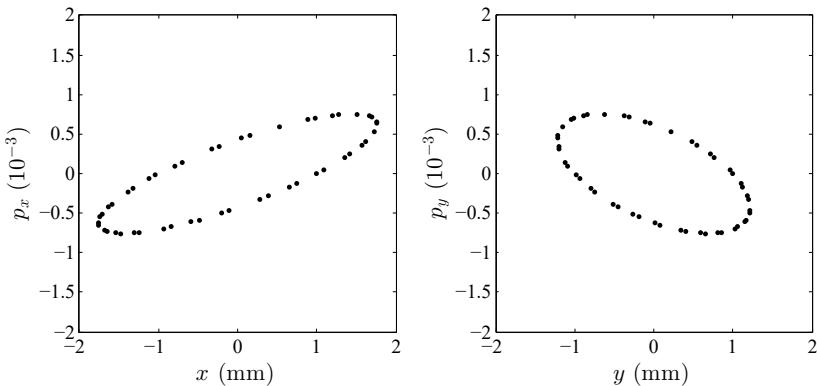


Fig. 4.3 Horizontal (left) and vertical (right) phase space plots for a particle in a FODO beam line, constructed by tracking a particle with initial values  $(x, p_x) = (1 \text{ mm}, 0)$  and  $(y, p_y) = (1 \text{ mm}, 0)$  for the horizontal and vertical dynamical variables, respectively. Each point shows the values of the dynamical variables after tracking through an integer number of FODO cells.

We see that the phase space co-ordinates, plotted at the start of the beam line and after each successive FODO cell, lie on an ellipse in each transverse plane. We obtain an ellipse no matter which starting point we take for the FODO cell, although the size and shape of the ellipse do depend on the starting point. Also, we find that, within a certain range of values, we can vary the focal length of the quadrupoles and length of the drifts and still obtain an ellipse; the size and shape again depend on the values chosen for the quadrupole focal length and the drift length.

## 4.2 The Courant–Snyder Parameters

To understand why we obtain ellipses in phase space when we track a particle repeatedly through a FODO cell, we need to study the properties of the transfer matrix  $R$ ; or, more precisely, we need to look at the properties of the block diagonals  $R_x$  and  $R_y$ . Our first observation is that the transfer matrix is *symplectic*, that is, it satisfies the symplectic condition:

$$R^T S R = S, \quad (4.13)$$

where  $S$  is the antisymmetric matrix:

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \quad (4.14)$$

It is no accident that the transfer matrix  $R$  is symplectic: it follows from the fact that  $R$  is obtained by multiplying matrices that are themselves symplectic; and the transfer matrices for the individual components are symplectic because they were obtained as solutions to Hamilton's equations. The main reason for taking the approach that we did in Chapter 3 to solving the equations of motion in the various beam line components, approximating the Hamiltonian in a form that allowed exact analytical solutions to the equations of motion, was to ensure that the transfer matrices we obtained were symplectic. Since the product of any two symplectic matrices is another symplectic matrix, the transfer map for any section of our FODO beam line can be represented by a symplectic matrix. Furthermore, the  $2 \times 2$  block diagonals of a  $2N \times 2N$  block-diagonal symplectic matrix are all symplectic.

The  $2 \times 2$  transfer matrix  $R_x$  for the horizontal motion can be written quite generally as:

$$R_x = I_2 \cos(\mu_x) + S_2 A_x \sin(\mu_x), \quad (4.15)$$

where  $I_2$  is the  $2 \times 2$  identity matrix,  $A_x$  is a  $2 \times 2$  symmetric matrix (whose components depend on the components of  $R_x$ ),  $\mu_x$  is a parameter (again depending on  $R_x$ ) and  $S_2$  is a  $2 \times 2$  antisymmetric matrix:

$$S_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (4.16)$$

Notice that the right-hand side of (4.15) has four free parameters (given by the three independent components of  $A_x$  and the parameter  $\mu_x$ ) corresponding to the four components of  $R_x$ . However, the symplectic condition imposes a single constraint, so only three of the four parameters are actually independent. To make this explicit, let us write the matrix  $A_x$  as:

$$A_x = \begin{pmatrix} \gamma_x & \alpha_x \\ \alpha_x & \beta_x \end{pmatrix}. \quad (4.17)$$

We shall refer to the quantities  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  as the (horizontal) *Courant–Snyder parameters* [Courant and Snyder (1958)], although they are also

known as the *Twiss parameters* [Twiss and Frank (1949)].  $\beta_x$  is also sometimes called the *betatron function* [Sands (1970)], or simply the *beta function*. The Courant–Snyder parameters play an important role in the linear optics of charged particle beams; we shall explore their properties and physical significance in the following sections. Substituting the expression in (4.15) for  $R_x$  into the symplectic condition:

$$R_x^T S_2 R_x = S_2, \quad (4.18)$$

we obtain:

$$\beta_x \gamma_x - \alpha_x^2 = 1. \quad (4.19)$$

Note that  $R_x$  can be written in terms of the Courant–Snyder parameters  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$ , and the parameter  $\mu_x$  as follows:

$$R_x = \begin{pmatrix} \cos(\mu_x) + \alpha_x \sin(\mu_x) & \beta_x \sin(\mu_x) \\ -\gamma_x \sin(\mu_x) & \cos(\mu_x) - \alpha_x \sin(\mu_x) \end{pmatrix}. \quad (4.20)$$

If the transfer matrix for a single cell is known (for example, by multiplying the transfer matrices for the different components within the cell), then expressing the transfer matrix in the form (4.20) is convenient for calculating the values for the phase advance and the Courant–Snyder parameters. In particular, the phase advance can be obtained from the trace of the transfer matrix:

$$\text{Tr}(R_x) = 2 \cos(\mu_x). \quad (4.21)$$

The values for the Courant–Snyder parameters are then found from the different elements of the transfer matrix.

As we have already mentioned, the elements of the matrix  $A_x$  depend on the elements of  $R_x$ ; the elements of  $R_x$  in turn depend on the starting position we choose within the FODO cell. The question then arises as to how the elements of  $A_x$  vary with position within the cell. To address this question, let us consider how  $A_x$  changes if we choose a different starting point for the cell, namely the exit of the F quadrupole, instead of the entrance of the F quadrupole. Let us denote the entrance of the F quadrupole  $s = s_0$ , and the exit of the F quadrupole  $s = s_1$ . (In the thin lens approximation,  $s_0$  and  $s_1$  will actually be the same value, but we shall treat them as separate for notational convenience.) The transfer matrix for the full FODO cell is then:

$$R(s_1) = R_F R_{\text{drift}} R_D R_{\text{drift}} = R_F R(s_0) R_F^{-1}. \quad (4.22)$$

Assuming that the parameter  $\mu_x$  is independent of the starting position within the cell, it follows from (4.15) that:

$$S_2 A_x(s_1) = R_{Fx} S_2 A_x(s_0) R_{Fx}^{-1}, \quad (4.23)$$

and so:

$$A_x(s_1) = (S_2^{-1} R_{Fx} S_2) A_x(s_0) R_{Fx}^{-1}. \quad (4.24)$$

Here,  $R_{Fx}$  is a  $2 \times 2$  sub-matrix (the first block diagonal) of the transfer matrix for the horizontally focusing quadrupole. Since  $R_F$  is symplectic and block diagonal,  $R_{Fx}$  is symplectic, and we have from (4.18):

$$S_2^{-1} R_{Fx}^T S_2 = R_{Fx}^{-1}, \quad (4.25)$$

and hence, using  $S_2^T = S_2^{-1}$ :

$$S_2^{-1} R_{Fx} S_2 = (R_{Fx}^T)^{-1}. \quad (4.26)$$

Substituting into (4.24), we find:

$$A_x(s_1) = (R_{Fx}^T)^{-1} A_x(s_0) R_{Fx}^{-1}. \quad (4.27)$$

Equation (4.27) gives the transformation of the matrix  $A_x$  (and hence, of the Courant–Snyder parameters  $\alpha_x$ ,  $\beta_x$ , and  $\gamma_x$ ) corresponding to a translation from one point of a periodic beam line to another. We can use this transformation rule to construct an invariant quantity from the phase space co-ordinates of a particle moving along the beam line: we show how to do this in Section 4.3. But as a final remark in the present section, we note that since the Courant–Snyder parameters satisfy the relation (4.19), the inverse of  $A_x$  is given by:

$$A_x^{-1} = \begin{pmatrix} \beta_x & -\alpha_x \\ -\alpha_x & \gamma_x \end{pmatrix}. \quad (4.28)$$

The transformation rule (4.27) for the Courant–Snyder parameters is more conveniently expressed in terms of the inverse of  $A_x$ :

$$A_x(s_1)^{-1} = R_x(s_1, s_0) A_x(s_0)^{-1} R_x(s_1, s_0)^T, \quad (4.29)$$

where  $A_x(s_0)$  and  $A_x(s_1)$  are the matrices constructed from the Courant–Snyder parameters at  $s = s_0$  and  $s = s_1$  respectively, and  $R(s_1, s_0)$  is the transfer matrix from  $s_0$  to  $s_1$ . Equation (4.29) is an important result. Although we have been discussing a periodic FODO beam line in particular, equation (4.29) applies for any uncoupled beam line: if we know the Courant–Snyder parameters at any point  $s_0$  in the beam line, and if we

know the transfer matrix from  $s_0$  to any other point  $s_1$ , then we can immediately calculate the Courant–Snyder parameters at  $s_1$ . In a periodic beam line, the Courant–Snyder parameters can be obtained from the transfer matrix for a single periodic section, using (4.15). If the beam line is not periodic, then we can, in principle, choose any values of the Courant–Snyder parameters that we like at some point in the beam line; the values at all other points are then determined using (4.29). However, as we shall see in Section 4.6, the Courant–Snyder parameters are related to the phase space distribution of particles in a bunch; the appropriate values for the Courant–Snyder parameters in a non-periodic beam line, therefore, are determined by the phase space distribution of particles injected into the beam line.

### 4.3 Action–Angle Variables

Our goal now is to understand the significance of the Courant–Snyder parameters for the beam dynamics. To achieve this, we first introduce the *action variable*  $J_x$ , which is defined in terms of the Courant–Snyder parameters and the phase space co-ordinates. We continue to consider only the horizontal motion in a beam line, which we assume to be independent of the vertical and longitudinal motion. All the results of this chapter can be applied directly to the vertical and longitudinal motion; we shall generalise the results to coupled beam lines in Chapter 5.

We construct the quantity  $J_x$  from the phase space co-ordinates and the  $2 \times 2$  matrix  $A_x$  (defined by (4.15)):

$$J_x = \frac{1}{2} (x \ p_x) A_x \begin{pmatrix} x \\ p_x \end{pmatrix}. \quad (4.30)$$

Under transport through a section of the beam line with  $2 \times 2$  transfer matrix  $R$ :

$$\begin{pmatrix} x \\ p_x \end{pmatrix} \mapsto R \begin{pmatrix} x \\ p_x \end{pmatrix}, \quad (4.31)$$

and from (4.27):

$$A_x \mapsto (R^T)^{-1} A_x R^{-1}. \quad (4.32)$$

Hence the action transforms as:

$$J_x \mapsto \frac{1}{2} (x \ p_x) R^T ((R^T)^{-1} A_x R^{-1}) R \begin{pmatrix} x \\ p_x \end{pmatrix} = \frac{1}{2} (x \ p_x) A_x \begin{pmatrix} x \\ p_x \end{pmatrix} = J_x. \quad (4.33)$$

Equation (4.33) tells us that the quantity  $J_x$  is invariant under transport along the beam line, even though the values of the phase space co-ordinates  $x$  and  $p_x$  and the parameters  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  (which are the components of  $A_x$ ) all vary along the beam line. In terms of the components of  $A_x$ , the action  $J_x$  can be written:

$$J_x = \frac{1}{2} (\gamma_x x^2 + 2\alpha_x x p_x + \beta_x p_x^2). \quad (4.34)$$

Strictly speaking, we do not need the factor of  $1/2$  in the definition of the action for the action to be invariant: however, it is conventional to include this factor, and the reasons for doing so will become apparent shortly.

The fact that the action is an invariant of the particle motion explains the ellipses that we observed when tracking a particle through the lattice that we constructed from repeated FODO cells. The Courant–Snyder parameters  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  are functions of position within a cell. If we observe a particle at corresponding locations within successive cells, the values of  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  will be the same (in other words, these functions have the same periodicity as the lattice). Then, (4.34) for fixed  $J_x$  and with the condition (4.19) defines an ellipse in phase space. The invariance of  $J_x$  constrains the values of  $x$  and  $p_x$  to lie on an ellipse in phase space. The area of the ellipse is  $2\pi J_x$ , and the shape of the ellipse is determined by the values of  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$ : this is illustrated in Fig. 4.4.

Since the action  $J_x$  determines the size of the ellipse in phase space obtained by tracking a particle through successive periodic cells, the location of a particle in phase space can be specified by giving a value for  $J_x$ , together with a value describing the position of the particle around the ellipse. We define the *angle variable*  $\phi_x$  as follows:

$$\tan(\phi_x) = -\beta_x \frac{p_x}{x} - \alpha_x. \quad (4.35)$$

The expressions (4.34) and (4.35) can be inverted to express the phase space co-ordinates  $(x, p_x)$  in terms of  $\phi_x$  and  $J_x$ :

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (4.36)$$

$$p_x = -\sqrt{\frac{2J_x}{\beta_x}} (\sin(\phi_x) + \alpha_x \cos(\phi_x)). \quad (4.37)$$

Equation (4.36) helps to explain the physical significance of the beta function  $\beta_x$ . As a particle moves along a periodic beam line, if the motion is stable it performs transverse oscillations (i.e. betatron oscillations) about the reference trajectory. Since the action  $J_x$  is constant, the beta function

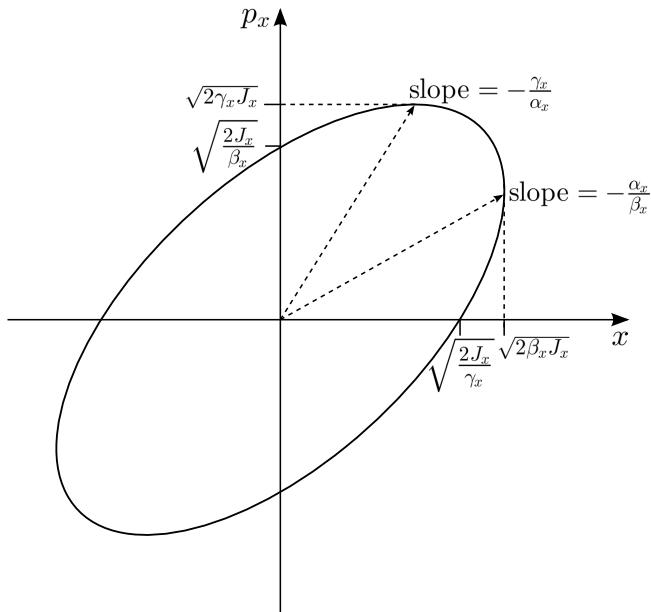


Fig. 4.4 Phase space ellipse in a periodic beam line. As a particle moves along the beam line, the phase space co-ordinates  $(x, p_x)$  at the exit of each cell lie on an ellipse. The area of the ellipse is  $2\pi J_x$ , where  $J_x$  is the action of the particle; the shape of the ellipse is defined by the Courant–Snyder parameters  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$ . The values of the Courant–Snyder parameters depend on how the periodic cell is constructed and which point is chosen as the starting point of the cell.

describes the *local* amplitude of oscillation of the particle. The actual co-ordinate of the particle depends on both the amplitude and phase (or angle  $\phi_x$ ); but the amplitude will vary with the same periodicity as the lattice. In other words, if the amplitude of oscillation of a particle is large at a particular location in one periodic cell, then it will be large at the corresponding locations in all periodic cells. In the case of a beam of many particles, we can expect the values of the action–angle variables of the particles to cover some range; but since the beta function will be the same for all particles, the size of the beam will vary along the beam line with the same periodicity as the beam line itself. The variation in the beam size is described by the beta function: this will be discussed in more detail in Section 4.6.

The variables  $J_x$  and  $\phi_x$  are known as *action–angle variables*: they provide a useful alternative to the Cartesian variables  $x$  and  $p_x$  for describing the motion of a particle in phase space [Goldstein *et al.* (2001b)]. It turns out that, since the action  $J_x$  is constant for linear symplectic transport,

action-angle variables are of particular value in beam dynamics, and we shall make extensive use of these variables in the remainder of this chapter, and in much of the rest of this book. Of course, we can define action-angle variables for the vertical dynamics, and for the longitudinal dynamics, in just the same way as we have for the horizontal dynamics: all we need are the Courant-Snyder parameters, which are obtained from the transfer matrix for a single period of a periodic beam line. For now, we must assume that the motion in each plane (horizontal, vertical or longitudinal) is independent of the motion in the other planes. In Chapter 5, we shall see how to extend the definition of the action-angle variables to the case where the horizontal, vertical and longitudinal dynamics are coupled.

In a simple FODO beam line consisting only of quadrupoles and drift spaces, the longitudinal motion is qualitatively different from the horizontal and the vertical motion: since there is no longitudinal focusing, the longitudinal motion will not be periodic. Instead, the energy deviation will be constant, while the longitudinal co-ordinate  $z$  will increase (or decrease) without limit. If we attempt to calculate the longitudinal Courant-Snyder parameters from the longitudinal part of the transfer matrix for a single FODO cell, we find that the value of  $\beta_z$  must be infinite. More generally, in the case of non-periodic motion, the values of  $\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$  and  $\mu_i$  (where  $i = x, y$  or  $z$ ) will (if  $\beta_i$  is finite) be imaginary. Real, finite values will only be obtained for the Courant-Snyder parameters if the motion of a particle in a given periodic beam line is periodic. We shall discuss the conditions for stable, periodic motion further in Section 4.4.

We know that (by construction) the action of a particle is invariant as the particle moves along a beam line. To complete the picture, we need to know how the angle evolves as the particle moves along a beam line. The first step in providing the answer is to define the vector  $\vec{J}_x$ , whose components are expressed in terms of the action-angle variables:

$$\vec{J}_x = \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix}. \quad (4.38)$$

The vector  $\vec{J}_x$  is related to the Cartesian variables by:

$$\begin{pmatrix} x \\ p_x \end{pmatrix} = N_x^{-1} \vec{J}_x, \quad (4.39)$$

where:

$$N_x^{-1} = \begin{pmatrix} \sqrt{\beta_x} & 0 \\ -\frac{\alpha_x}{\sqrt{\beta_x}} & \frac{1}{\sqrt{\beta_x}} \end{pmatrix}. \quad (4.40)$$

Under a transformation from a point  $s_0$  to a point  $s_1$  along a beam line:

$$\begin{pmatrix} x \\ p_x \end{pmatrix} \mapsto R(s_1, s_0) \begin{pmatrix} x \\ p_x \end{pmatrix}, \quad (4.41)$$

where  $R(s_1, s_0)$  is the  $2 \times 2$  transfer matrix from  $s_0$  to  $s_1$ . It follows from (4.39) that:

$$\vec{J}_x \mapsto N_x(s_1)R(s_1, s_0)N_x(s_0)^{-1}\vec{J}_x, \quad (4.42)$$

where  $N_x(s_0)$  and  $N_x(s_1)$  are the matrices defined by (4.40) using the appropriate Courant–Snyder parameters at  $s_0$  and  $s_1$ , respectively. We know that as a particle travels along a beam line, the action  $J_x$  remains constant; only the angle  $\phi_x$  changes. Therefore, we can write for the transformation of  $\vec{J}_x$ :

$$\begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x + d\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x + d\phi_x) \end{pmatrix}. \quad (4.43)$$

We note that:

$$\begin{pmatrix} \sqrt{2J_x} \cos(\phi_x + d\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x + d\phi_x) \end{pmatrix} = \begin{pmatrix} \cos(d\phi_x) & \sin(d\phi_x) \\ -\sin(d\phi_x) & \cos(d\phi_x) \end{pmatrix} \vec{J}_x. \quad (4.44)$$

Therefore, using the ‘normalised’ phase space vector  $\vec{J}_x$ , the transformation corresponding to particle motion along a beam line appears as a pure rotation. The matrix  $N_x$  (4.40) is called the *normalising matrix*.

Now let us consider an infinitesimal step along the beam line, so that  $s_1 = s_0 + ds$ . In general,  $R(s_1, s_0)$  must be symplectic. Furthermore, we expect that in the limit  $ds \rightarrow 0$ :

$$x \mapsto x + p_x ds. \quad (4.45)$$

Therefore, we consider a matrix  $R(s_1, s_0)$  of the form:

$$R = \begin{pmatrix} \cos(\omega ds) + a \sin(\omega ds) & \frac{1}{\omega} \sin(\omega ds) \\ -\omega(1 + a^2) \sin(\omega ds) & \cos(\omega ds) - a \sin(\omega ds) \end{pmatrix}, \quad (4.46)$$

where  $\omega$  and  $a$  are parameters determined by the values of the components of  $R(s_1, s_0)$ . The values of the Courant–Snyder parameters at  $s_1$  may be obtained from the values at  $s_0$  using (4.29). Then, we find that:

$$N_x(s_1)R(s_1, s_0)N_x(s_0)^{-1} = \begin{pmatrix} 1 & \frac{ds}{\beta_x(s_0)} \\ -\frac{ds}{\beta_x(s_0)} & 1 \end{pmatrix} + O(ds^2), \quad (4.47)$$

where  $\beta_x(s_0)$  is the value of the beta function at  $s = s_0$ . Now, from (4.43) and (4.44), the transformation  $N_x(s_1)R(s_1, s_0)N_x(s_0)^{-1}$  must take the form of a pure rotation. By inspection of (4.47), the change  $d\phi_x$  corresponding

to the advance of a particle by an infinitesimal distance  $ds$  along a beam line is therefore:

$$d\phi_x = \frac{ds}{\beta_x}, \quad (4.48)$$

where  $\beta_x$  is the Courant–Snyder beta function at the location of the particle. We can now write the equations of motion for a particle moving along a beam line in terms of the action–angle variables:

$$\frac{dJ_x}{ds} = 0, \quad (4.49)$$

$$\frac{d\phi_x}{ds} = \frac{1}{\beta_x}. \quad (4.50)$$

From (4.36) we have seen that the beta function describes the variation in the local amplitude of betatron oscillations of particles moving along a beam line. Now we see from (4.50) that the beta function also describes the local wavelength of the betatron oscillations: the larger the beta function, the slower the rate of change of the angle variable, and the slower the phase advance of the betatron oscillation. Note that (4.49) and (4.50) apply separately to the horizontal, vertical and longitudinal motion, and are valid when the three degrees of freedom are independent of each other.

Equation (4.50) can also be derived from equation (4.36), which relates the co-ordinate  $x$  to the action–angle variables. Taking the derivative of (4.36) with respect to  $s$  gives:

$$\frac{dx}{ds} = \frac{1}{2} \sqrt{\frac{2J_x}{\beta_x}} \cos(\phi_x) \frac{d\beta_x}{ds} - \sqrt{2\beta_x J_x} \sin(\phi_x) \frac{d\phi_x}{ds}. \quad (4.51)$$

If the horizontal momentum appears in the Hamiltonian as a term  $p_x^2/2$ , then, from Hamilton's equations:

$$\frac{dx}{ds} = \frac{\partial H}{\partial p_x} = p_x. \quad (4.52)$$

This is the same as the assumption (4.45). Then, comparing (4.51) with (4.37) gives, from the term in  $\sin(\phi_x)$ , the rate of change of the angle variable along the beam line:

$$\frac{d\phi_x}{ds} = \frac{1}{\beta_x}. \quad (4.53)$$

From the term in  $\cos(\phi_x)$ , we obtain a relationship between the Courant–Snyder alpha function and the derivative of the Courant–Snyder beta function:

$$\alpha_x = -\frac{1}{2} \frac{d\beta_x}{ds}. \quad (4.54)$$

Using the Cartesian variables (for example  $(x, p_x)$  in the horizontal plane) the transfer matrices are symplectic: this is a consequence of the fact that the dynamical variables form a canonical pair, and the equations of motion can be derived from an appropriate Hamiltonian. It turns out that the action–angle variables also form a canonical pair: the angle is the co-ordinate of the particle, and the action is the canonically conjugate momentum. This may be shown by constructing an appropriate mixed-variable generating function for a canonical transformation from  $(x, p_x)$  to  $(\phi_x, J_x)$ . For example, using (in Goldstein’s nomenclature [Goldstein *et al.* (2001c)]) a mixed-variable generating function of the first kind:

$$F_1(x, \phi_x) = -\frac{x^2}{2\beta_x}(\tan(\phi_x) + \alpha_x), \quad (4.55)$$

we obtain the following relationship between the Cartesian variables and the action–angle variables:

$$J_x = -\frac{\partial F_1}{\partial \phi_x} = \frac{x^2}{2\beta_x} \sec^2(\phi_x), \quad (4.56)$$

$$p_x = \frac{\partial F_1}{\partial x} = -\frac{x}{\beta_x}(\tan(\phi_x) + \alpha_x). \quad (4.57)$$

Rearranging, these equations give the relationships (4.36) and (4.35). Since the action–angle variables form a canonical pair, we expect that the equations of motion (4.49) and (4.50) can be derived from a Hamiltonian. This is indeed the case, and an appropriate Hamiltonian (in one degree of freedom) is:

$$H = \frac{J_x}{\beta_x}. \quad (4.58)$$

Finally, it is worth noting that since the transformation in (4.42) is a pure rotation (in which the angle variable increases by  $\mu_x$ ), we have:

$$N_x(s_1)R(s_1, s_0)N_x(s_0)^{-1} = \begin{pmatrix} \cos(\mu_x) & \sin(\mu_x) \\ -\sin(\mu_x) & \cos(\mu_x) \end{pmatrix}. \quad (4.59)$$

Using (4.40), it then follows that the transfer matrix  $R(s_1, s_0)$  can be written in terms of the phase advance  $\mu_x$  and the Courant–Snyder parameters ( $\beta_{x0} \equiv \beta_x(s_0)$ , etc.) as follows:

$$\begin{aligned} R(s_1, s_0) = & \\ & \begin{pmatrix} \sqrt{\frac{\beta_{x1}}{\beta_{x0}}}(\cos(\mu_x) + \alpha_{x0} \sin(\mu_x)) & \sqrt{\beta_{x0}\beta_{x1}} \sin(\mu_x) \\ \frac{\alpha_{x0} - \alpha_{x1}}{\sqrt{\beta_{x0}\beta_{x1}}} \cos(\mu_x) - \frac{1 + \alpha_{x0}\alpha_{x1}}{\sqrt{\beta_{x0}\beta_{x1}}} \sin(\mu_x) & \sqrt{\frac{\beta_{x0}}{\beta_{x1}}}(\cos(\mu_x) - \alpha_{x1} \sin(\mu_x)) \end{pmatrix}. \end{aligned} \quad (4.60)$$

#### 4.4 Courant–Snyder Parameters in a FODO Beam Line

We now return to our example of a FODO beam line to calculate the Courant–Snyder parameters and the phase advance (the increase in the value of the angle variable as a particle moves along the beam line). We shall also derive the conditions required for particle motion to be stable.

To begin, recall that the transfer matrix for one complete period of the beam line can be written in the form (4.20). The transformation of the vector  $\vec{J}_x$  (4.38) over one complete period is:

$$\vec{J}_x \mapsto N_x R N_x^{-1} \vec{J}_x, \quad (4.61)$$

where, from (4.40):

$$N_x = \begin{pmatrix} \frac{1}{\sqrt{\beta_x}} & 0 \\ \frac{\alpha_x}{\sqrt{\beta_x}} & \sqrt{\beta_x} \end{pmatrix}. \quad (4.62)$$

Performing the matrix multiplication, we find:

$$N_x R N_x^{-1} = \begin{pmatrix} \cos(\mu_x) & \sin(\mu_x) \\ -\sin(\mu_x) & \cos(\mu_x) \end{pmatrix}. \quad (4.63)$$

The matrix  $N_x$  provides a transformation that puts the transfer matrix for one complete period of the lattice into the form of a pure rotation. Equation (4.44) shows that a rotation matrix applied to the vector  $\vec{J}_x$  leads to an increase in the angle variable by an amount equal to the rotation angle; hence, the parameter  $\mu_x$  in the expression (4.20) for the transfer matrix gives the increase in the angle variable over one periodic section of the beam line: this is called the *phase advance*. In general, there may be a different phase advance over a periodic section of a beam line in each degree of freedom. For the transverse variables  $x$  and  $y$ , we refer to the *betatron phase advance* (which, in general, will be different in the horizontal and the vertical); if there is periodic motion in the longitudinal variables, as in the case of a synchrotron storage ring, then we can refer also to the *synchrotron phase advance*. For a simple FODO beam line, if the F and D quadrupoles have the same magnitude of field gradient (though of course with different signs), the vertical phase advance will be equal to the horizontal phase advance. Without longitudinal focusing, the longitudinal motion is not periodic.

From (4.9), the  $2 \times 2$  transfer matrix for the horizontal motion in one FODO cell, starting from the entrance to the F quadrupole, is:

$$R_x = \begin{pmatrix} 1 - \frac{L}{f} \left(1 + \frac{L}{f}\right) \frac{L}{f} (2f + L) & \\ -\frac{L}{f^2} & 1 + \frac{L}{f} \end{pmatrix}. \quad (4.64)$$

The horizontal Courant–Snyder parameters and phase advance may be obtained by comparing the components of the transfer matrix in the form (4.64) with the components of the transfer matrix in the form (4.20). For example, if we equate the trace of the right-hand side of (4.20) with the trace of the right-hand side of (4.64), we find:

$$\cos(\mu_x) = 1 - \frac{L^2}{2f^2}. \quad (4.65)$$

For the case where the quadrupoles are represented as thin lenses with focal length  $f$  separated by distance  $L$ , (4.65) gives a convenient expression for the phase advance in a FODO cell. We can find a real solution for the phase advance only if  $f$  and  $L$  are such that:

$$-1 \leq 1 - \frac{L^2}{2f^2} \leq 1. \quad (4.66)$$

If the only solutions to (4.65) give imaginary values for  $\mu_x$ , then the motion of a particle in the beam line is unstable, as we shall see shortly. Furthermore, if  $\mu_x = n\pi$  for integer  $n$ , then we are unable to determine values for the Courant–Snyder parameters, and such beam lines are also generally regarded as unstable: although they may allow periodic motion in principle, such beam lines would be extremely difficult to operate in practice. The stability condition may also be expressed in terms of the trace of the transfer matrix:

$$|\text{Tr}(R)| < 2, \quad (4.67)$$

where  $R$  is the  $2 \times 2$  transfer matrix for one periodic section of an uncoupled beam line in one degree of freedom. For the longitudinal motion in a straight FODO beam line  $\text{Tr}(R) = 2$ : this gives an immediate indication that the longitudinal motion is not periodic.

Assuming that the condition (4.67) is satisfied for a FODO cell, then having obtained the value for the phase advance  $\mu_x$  (or  $\mu_y$ ), the Courant–Snyder parameters can be readily found from:

$$\alpha_x \sin(\mu_x) = -\frac{L}{f} \left( 1 + \frac{L}{2f} \right), \quad (4.68)$$

$$\beta_x \sin(\mu_x) = \frac{L}{f} (2f + L), \quad (4.69)$$

$$\gamma_x \sin(\mu_x) = \frac{L}{f^2}. \quad (4.70)$$

These expressions give the values of the Courant–Snyder parameters at the location of the cell corresponding to the point at which the single-cell

transfer matrix (4.64) is calculated, namely the entrance of the horizontally focusing quadrupole. The Courant–Snyder parameters at any other point in the cell may be calculated using (4.29). However, we note that replacing  $f$  by  $-f$  in (4.64) gives the transfer matrix for a single cell starting at the entrance of the horizontally defocusing quadrupole: since the quadrupoles in the FODO cell have the same strength but opposite sign, changing the sign of  $f$  simply corresponds to shifting the entire beam line longitudinally by the distance between two quadrupoles. Then, (4.65) gives the vertical (as well as the horizontal) phase advance across one FODO cell; and (4.68), (4.69) and (4.70) give the values for the *vertical* Courant–Snyder parameters at the entrance of the *vertically focusing* quadrupole. Once the Courant–Snyder parameters have been found at one point in the cell, the Courant–Snyder parameters at any other point can be found by application of (4.29). The horizontal and vertical beta functions in a FODO cell with (thin) quadrupoles of focal length  $\pm\sqrt{2}$  m and drifts of length 1 m are shown in Fig. 4.5.

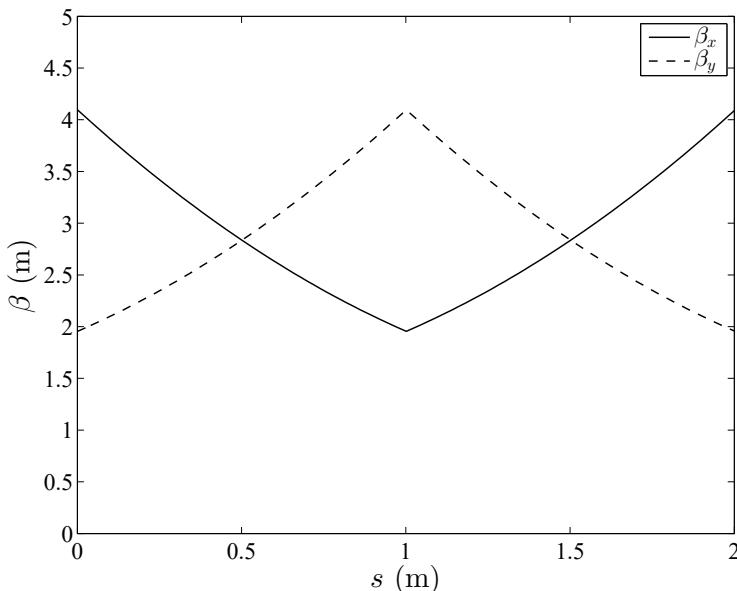


Fig. 4.5 Beta functions in a FODO cell with (thin) quadrupoles of focal length  $\pm\sqrt{2}$  m and drifts of length 1 m. The horizontally focusing quadrupole is at  $s = 0$  m, and the vertically focusing quadrupole is at  $s = 1$  m.

The  $2 \times 2$  transfer matrix for the longitudinal motion in one FODO cell is, from (4.9):

$$R_z = \begin{pmatrix} 1 & \frac{2L}{\beta_0^2 \gamma_0^2} \\ 0 & 1 \end{pmatrix}. \quad (4.71)$$

Therefore, the longitudinal phase advance is given by:

$$\cos(\mu_z) = 1, \quad (4.72)$$

which implies that  $\mu_z = 0$  or  $\mu_z = 2\pi$  (the stability condition (4.67) is not satisfied). If we consider the ultra-relativistic limit  $\gamma_0 \rightarrow \infty$ , for which case the longitudinal transfer matrix reduces to the identity, it seems appropriate to take the solution  $\mu_z = 0$ . In either case, we have  $\sin(\mu_z) = 0$ , and we are unable to determine values for the longitudinal Courant–Snyder parameters. The lack of longitudinal focusing in the FODO cell means that the longitudinal motion cannot be stable. If we plot the longitudinal phase space co-ordinates from one cell to the next, then the points do not lie on a closed ellipse; instead, for non-zero values of the energy deviation  $\delta$ , the magnitude of the longitudinal co-ordinate  $z$  increases indefinitely.

Consider the case of the horizontal dynamics when  $f < L/2$ ; then, from (4.65), the horizontal phase advance is such that  $\cos(\mu_x) < -1$ . This does not admit a real solution for  $\mu_x$ . The dynamics can be described using imaginary values for the phase advance, and for the Courant–Snyder parameters. Alternatively, the matrix decomposition (4.15) can be rewritten as:

$$R_x = I_2 \cosh(\tilde{\mu}_x) + S_2 A_x \sinh(\tilde{\mu}_x), \quad (4.73)$$

where (as before)  $I_2$  is the  $2 \times 2$  identity matrix, the matrix  $S_2$  is given by (4.16), and the matrix  $A_x$  is given by (4.17). Equation (4.73) admits a real solution for  $\tilde{\mu}_x$  when  $|\text{Tr}(R_x)| > 2$ . The symplecticity condition applied to  $R_x$  written in the form (4.73) leads to the relationship between the Courant–Snyder parameters:

$$\beta_x \gamma_x - \alpha_x^2 = -1. \quad (4.74)$$

This should be compared with (4.19), which gives the corresponding relationship between the Courant–Snyder parameters for the case of stable periodic motion. The betatron action  $J_x$  can be defined in the same way as before, using (4.30), and is a constant of the motion whether we write  $R_x$  in the form (4.15) or in the form (4.73). However, when the relationship (4.19) applies, plotting the phase space co-ordinates of a particle after successive

passes through a repeated set of beam line components produces points lying on a closed curve, an ellipse. But in the case that the Courant–Snyder parameters satisfy (4.74), plotting the phase space co-ordinates after successive passes through a repeated set of components produces points lying on a hyperbola. Under these conditions, the values of the phase space co-ordinates eventually increase without limit, and the particle will eventually hit the wall of the vacuum chamber.

## 4.5 Hill's Equation

An alternative approach to developing a description of the linear dynamics of a particle in terms of the Courant–Snyder parameters is provided by Hill's equation. In a beam line in the absence of coupling, the Hamiltonian for each degree of freedom leads to an equation of motion of the form:

$$x'' = -k_1(s)x, \quad (4.75)$$

where the prime indicates a derivative with respect to  $s$  (the distance along the beam line), and  $k_1(s)$  is a function of  $s$  that characterises the focusing at each point along the beam line. In a drift space, for example,  $k_1 = 0$  since the momentum of a particle will be constant. In a quadrupole,  $k_1$  will be related to the field gradient. In the approximation:

$$p_x \approx x', \quad (4.76)$$

we can write:

$$k_1 \approx \frac{q}{P_0} \frac{\partial B_y}{\partial x}, \quad (4.77)$$

where  $q$  is the particle charge,  $P_0$  is the reference momentum and  $B_y = B_y(s)$  is the vertical component of the magnetic field.

In the case that  $k_1(s)$  is a periodic function of  $s$  (4.75) is known as Hill's equation. Based on the similarity to the equation for simple harmonic motion, we try a solution of the form:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (4.78)$$

where  $J_x$  is a constant, and  $\beta_x$  and  $\phi_x$  are functions of  $s$ . Equation (4.78) represents the motion of a particle along a beam line as an oscillation with varying amplitude and wavelength. Substituting (4.78) into (4.75) and equating coefficients of  $\cos(\phi_x)$  and  $\sin(\phi_x)$  on either side of the resulting equation, leads to the following equations for  $\beta_x$  and  $\phi_x$ :

$$\beta'_x \phi'_x + \beta_x \phi''_x = 0, \quad (4.79)$$

$$\frac{\beta''_x}{2\beta_x} - \frac{\beta'^2_x}{4\beta_x^2} - \phi'^2_x = -k_1. \quad (4.80)$$

Equation (4.79) has the solution:

$$\phi'_x = \frac{K}{\beta_x}, \quad (4.81)$$

where  $K$  is a constant. We make the conventional choice  $K = 1$ : choosing a different value for  $K$  leads to equally valid solutions to Hill's equation, but the quantities  $\beta_x$  and  $\phi_x$  will have a different interpretation. Substituting for  $\phi'_x$  from (4.81) into (4.80) gives the differential equation for  $\beta_x$ :

$$\beta''_x - \frac{4 + \beta'^2_x}{2\beta_x} + 2k_1\beta_x = 0. \quad (4.82)$$

If  $k_1$  is known (from the design of the beam line), then for given initial values of  $\beta_x$  and  $\beta'_x$  equation (4.82) may be integrated to give  $\beta_x$  at any position along the beam line. Usually a numerical integration is required.

In the case of a periodic beam line, the focusing strength satisfies the periodicity condition:

$$k_1(s + C_0) = k_1(s), \quad (4.83)$$

where  $s$  is any point along the beam line, which has periodicity  $C_0$ . We can impose the same periodicity condition on the beta function:

$$\beta_x(s + C_0) = \beta_x(s), \quad (4.84)$$

$$\beta'_x(s + C_0) = \beta'_x(s). \quad (4.85)$$

The transformation (4.78) is then known as a *Floquet transformation*. The conditions (4.84) and (4.85) provide constraints that determine a unique (periodic) solution to the differential equation (4.82) for the beta function. Usually, it is simpler to find the Courant–Snyder parameters in an uncoupled periodic beam line using the transfer matrix expressed in the form (4.20) than to solve the differential equation (4.82) subject to the periodicity conditions on the beta function.

Equation (4.82) provides a method for propagating the beta function along a beam line with given initial conditions, if the focusing is known continuously along a beam line. The role of (4.82) is similar to that of (4.29), which allows us to propagate the Courant–Snyder parameters from any point in a beam line to any other point, if we know the transfer matrix between the two points.

Using the approach based on Hill's equation, the Courant–Snyder alpha function may be introduced by taking the derivative of (4.78) with respect to  $s$ :

$$x' = \frac{1}{2} \sqrt{\frac{2J_x}{\beta_x}} \beta'_x \cos(\phi_x) - \sqrt{2\beta_x J_x} \phi'_x \sin(\phi_x). \quad (4.86)$$

If we assume that  $x' \approx p_x$ , then using (4.81) (again with  $K = 1$ ) and comparing (4.86) with (4.37), we conclude that:

$$\alpha_x \approx -\frac{1}{2}\beta'_x. \quad (4.87)$$

The approximation comes from the fact that we defined the Courant–Snyder parameters in terms of the transfer matrices for the canonical variables  $x$  and  $p_x$ . Using Hill's equation, it is more natural to use the variables  $x$  and  $x'$ . Although in many situations the value of  $p_x$  is close to  $x'$ , it should be remembered that these variables represent different physical quantities. In particular:

$$x' = \frac{dx}{ds} = \frac{v_x}{v_s}, \quad (4.88)$$

where  $v_x$  is the component of the velocity of a particle in the  $x$  direction, and  $v_s$  is the component of the velocity along the reference trajectory. For a particle with zero energy deviation ( $\delta = 0$ ), and in a region where the vector potential has  $x$  component  $A_x = 0$ , the momentum variable  $p_x$  is given by:

$$p_x = \frac{v_x}{\beta_0 c}. \quad (4.89)$$

In the limit that the direction of motion of a particle approaches the  $x$  axis,  $x' \rightarrow \infty$ , while  $p_x \rightarrow 1$ . For the approximation  $x' \approx p_x$  to be valid, it is necessary that the direction of motion of the particle make a small angle with the reference trajectory.

As an example, let us consider the evolution of the beta function in a drift space, starting from a point  $s = 0$  at which  $\beta_x(0) = \beta_{x0}$ , and  $\beta'_x(0) = 0$ . Using (4.87) we put  $\alpha(0) = 0$ , and using (4.19) we put  $\gamma_x(0) = 1/\beta_{x0}$ . In a drift space, we can write the transfer matrix over a distance  $s$ :

$$R_x = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}. \quad (4.90)$$

Then, performing the matrix multiplication in (4.29), we find:

$$\beta_x(s) = \beta_{x0} + \frac{s^2}{\beta_{x0}}, \quad (4.91)$$

$$\alpha_x(s) = -\frac{s}{\beta_{x0}}. \quad (4.92)$$

The expression (4.91) is found to solve (4.82) for  $k_1 = 0$  (i.e. no focusing) and with the given initial conditions. Equation (4.91) shows that in a drift space the beta function increases quadratically from a point where the beta function has zero gradient, and that the rate of growth of the beta function increases in inverse proportion to the initial (i.e. the minimum) value of the beta function.

## 4.6 Courant–Snyder Parameters and Particle Distribution

The discussion in this chapter so far has been based on the assumption that the beam line with which we are dealing has a periodic structure. The key result in this case is that if the motion of a particle along the beam line is stable, then the phase space co-ordinates of a particle can be expressed as (4.36) and (4.37):

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (4.93)$$

$$p_x = -\sqrt{\frac{2J_x}{\beta_x}} (\sin(\phi_x) + \alpha_x \cos(\phi_x)). \quad (4.94)$$

Similar equations hold for the vertical and (if there is longitudinal focusing) longitudinal motion. The action  $J_x$  of the particle is constant along the beam line, and between any two points in the beam line the change in the angle variable  $\phi_x$  is given by the phase advance between those points. The functions  $\beta_x$  and  $\alpha_x$  (the Courant–Snyder parameters) vary with position along the beam line, but (in a periodic beam line) have the same periodicity as the components within the beam line. The values of  $\beta_x$  and  $\alpha_x$  at any point are determined by the transfer matrix representing transport of a particle through a complete period starting from that point. The phase advance for a single cell can be determined from the same transfer matrix.

Of course, not all accelerator beam lines are periodic. For a non-periodic beam line, there is apparently no way to determine the values of the Courant–Snyder parameters. In such a case, the Courant–Snyder parameters at any point in the beam line can, in principle, be chosen arbitrarily. However, once the values at some point have been fixed, then the values at all other points in the beam line are determined by propagating the Courant–Snyder parameters using the transfer matrices and (4.29). In fact, we are at liberty to choose arbitrary values for the Courant–Snyder parameters even in a periodic beam line, although the  $\beta_x$  and  $\alpha_x$  functions will have the same periodicity as the beam line only when they are determined from the transfer matrix for a single cell using (4.15). In practice, the values of the Courant–Snyder parameters are often best determined from the beam, rather than from the structure of the beam line. The Courant–Snyder parameters can be determined from the beam because, with certain assumptions, they describe not only the dynamics of individual particles but also the distribution of particles within a bunch. Specifically, if we take the statistical average of  $x^2$  over all particles within a given bunch at a

particular point in a beam line, then from (4.36) we find:

$$\langle x^2 \rangle = 2\beta_x \langle J_x \cos^2(\phi_x) \rangle. \quad (4.95)$$

The brackets  $\langle \cdot \rangle$  indicate the mean of the enclosed quantity over all particles in a bunch. We make two assumptions: first, that the angle variables of the particles within the bunch are uncorrelated with the action variables; and second, that the angle variables are uniformly distributed from 0 to  $2\pi$ . If these assumptions are valid, then:

$$\langle x \rangle = 0, \quad (4.96)$$

and the mean square value of  $x$  becomes, from (4.95):

$$\langle x^2 \rangle = \beta_x \epsilon_x, \quad (4.97)$$

where we have defined the *horizontal emittance* of the bunch:

$$\epsilon_x = \langle J_x \rangle. \quad (4.98)$$

Similarly, using the same assumptions, we find from (4.36) and (4.37):

$$\langle xp_x \rangle = -\alpha_x \epsilon_x, \quad (4.99)$$

$$\langle p_x^2 \rangle = \gamma_x \epsilon_x. \quad (4.100)$$

Combining (4.97), (4.99) and (4.100) with (4.19), the emittance can be expressed in terms of the beam distribution:

$$\epsilon_x = \sqrt{\langle x^2 \rangle \langle p_x^2 \rangle - \langle xp_x \rangle^2}. \quad (4.101)$$

If the parameters of the beam distribution (i.e. the values of  $\langle x^2 \rangle$ ,  $\langle p_x^2 \rangle$  and  $\langle xp_x \rangle$ ) are known at a particular point in a beam line, then the emittance of the beam can be determined from (4.101), and appropriate values for the Courant–Snyder parameters are then found from (4.97), (4.99) and (4.100). If the transformation for each section of the beam line is symplectic, the action for each particle is conserved as the bunch moves along the beam line. It then follows immediately from (4.98) that the emittance of the bunch is also conserved. We see that the Courant–Snyder parameters describe the evolution of the distribution of particles within a bunch as the bunch moves along a beam line. There is no assumption about the periodicity of the beam line in this case. However, if the beam line is periodic, then the Courant–Snyder parameters may be obtained either from analysis of the transfer matrix for a full period of the beam line, or from the particle distribution within the bunch at a particular point in the beam line. If the Courant–Snyder parameters obtained from the two methods are equal, then the bunch is said to be properly *matched* to the beam line, and the beam

distribution will vary along the beam line with the same periodicity as the beam line itself. If the bunch is not properly matched to the beam line, then the Courant–Snyder parameters describing the particle distribution will vary with a different periodicity from the beam line: this phenomenon is sometimes referred to as *beta-beating*.

Note that equations (4.97), (4.99), (4.100) and (4.101), relating the emittance and Courant–Snyder parameters to the particle distribution, are valid only if the horizontal motion can be treated separately from the vertical and the longitudinal; in other words, if the motion is uncoupled. Clearly, in the case of uncoupled motion we can define vertical and longitudinal emittances in exactly the same way as we define the horizontal emittance. A bunch of particles therefore has three conserved emittances, at least in the case of linear symplectic transport. In Chapter 5, we shall see how to extend the definitions of the Courant–Snyder parameters and the emittances to cases where there is coupling between the horizontal, vertical and longitudinal motion.

In a periodic beam line, the action is a measure of the area within the phase space ellipse mapped out by a particle as it moves through successive periods. Since the emittance is the average of the action over all particles in the beam, the emittance is a measure of the area of phase space occupied by particles within the beam. Different definitions of the emittance than (4.98) and (4.101) are sometimes used, and, depending on the distribution of particles, may be more appropriate in particular cases. The emittance that satisfies (4.101) is sometimes called the root mean square (rms) emittance; the emittance may also be defined in terms of the area of phase space within which some specified fraction (for example, 95%) of the particles are to be found [Edwards and Syphers (1999)]. In this book, we shall consistently use the definition of the emittance as the average of the action variables over all particles in a bunch.

Since the emittance is a measure of the area of phase space occupied by a bunch of particles, and the emittance is conserved under linear symplectic transport, this suggests that the density of particles in phase space is also conserved. In fact, we have already seen in Chapter 2 that, when the particle dynamics are governed by Hamilton’s equations, the density of particles in phase space is a constant of the motion: this is Liouville’s theorem. However, Liouville’s theorem identifies only one constant of the motion, irrespective of the number of degrees of freedom. We know from the results in this chapter, that for uncoupled, linear, symplectic motion there is a conserved quantity (the emittance) for each degree of freedom.

We shall see in Chapter 5 that the same holds true even when there is coupling between the degrees of freedom.

In accelerator beam lines, we are often interested in the case where the beam is accelerated, i.e. there is a change in the energy of the particles. If the energy change for a given particle is small, then it may be accounted for by a change of the energy deviation. However, for large energy changes it is preferable to represent the change in energy by a change in the reference momentum: a transfer map for this transformation was given in Chapter 3 (3.154). It was noted at the time that the transformation was not symplectic. In that case, we expect to observe some change in the action (and emittance) when we make a change in the reference momentum.

The change in the emittance and in the Courant–Snyder parameters associated with a change in the reference momentum may be calculated as follows. First, we observe that we can construct a matrix  $\Sigma$  consisting of the second-order moments of the beam distribution, as follows:

$$\Sigma = \begin{pmatrix} \langle x^2 \rangle & \langle xp_x \rangle \\ \langle xp_x \rangle & \langle p_x^2 \rangle \end{pmatrix} = \begin{pmatrix} \beta_x & -\alpha_x \\ -\alpha_x & \gamma_x \end{pmatrix} \epsilon_x, \quad (4.102)$$

where we have used (4.97), (4.99) and (4.100). The transformation properties of  $\Sigma$  follow from the transformation properties of the phase space variables.  $\Sigma$  can be written as:

$$\Sigma = \langle \vec{x} \cdot \vec{x}^T \rangle, \quad (4.103)$$

where  $\vec{x}$  is the phase space vector:

$$\vec{x} = \begin{pmatrix} x \\ p_x \end{pmatrix}. \quad (4.104)$$

Therefore, under the transformation:

$$\vec{x} \mapsto R_x \vec{x}, \quad (4.105)$$

$\Sigma$  transforms as:

$$\Sigma \mapsto R_x \Sigma R_x^T. \quad (4.106)$$

In the case of a change of reference momentum,  $R_x$  is obtained from (3.154):

$$R_x = \begin{pmatrix} 1 & 0 \\ 0 & P_0/P_1 \end{pmatrix}, \quad (4.107)$$

where  $P_0$  and  $P_1$  are the values of the reference momentum before and after the transformation, respectively. The change in the emittance and the Courant–Snyder parameters can then be found from (4.102) and (4.106):

$$\begin{pmatrix} \beta_{x1} & -\alpha_{x1} \\ -\alpha_{x1} & \gamma_{x1} \end{pmatrix} \epsilon_{x1} = \begin{pmatrix} 1 & 0 \\ 0 & P_0/P_1 \end{pmatrix} \begin{pmatrix} \beta_{x0} & -\alpha_{x0} \\ -\alpha_{x0} & \gamma_{x0} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & P_0/P_1 \end{pmatrix} \epsilon_{x0}. \quad (4.108)$$

The initial and final Courant–Snyder parameters satisfy the usual relationship (4.19), so that (respectively):

$$\beta_{x0}\gamma_{x0} - \alpha_{x0}^2 = 1, \quad (4.109)$$

$$\beta_{x1}\gamma_{x1} - \alpha_{x1}^2 = 1. \quad (4.110)$$

Solving (4.108), (4.109) and (4.110) gives:

$$\alpha_{x1} = \alpha_{x0}, \quad (4.111)$$

$$\beta_{x1} = \frac{P_1}{P_0} \beta_{x0}, \quad (4.112)$$

$$\epsilon_{x1} = \frac{P_0}{P_1} \epsilon_{x0}. \quad (4.113)$$

The change in the beta function and the emittance is such that mean square beam size  $\langle x^2 \rangle$  is constant. This is just what we should expect, since the transverse co-ordinates  $x$  and  $y$  are independent of the reference momentum. The emittance is changed in such a way that:

$$\beta_1\gamma_1\epsilon_{x1} = \beta_0\gamma_0\epsilon_{x0}, \quad (4.114)$$

where  $P_0 = \beta_0\gamma_0 mc$ , and  $P_1 = \beta_1\gamma_1 mc$ . The quantities  $\beta_0$ ,  $\gamma_0$ ,  $\beta_1$ ,  $\gamma_1$  are relativistic parameters, and should not be confused with the Courant–Snyder parameters.

Equation (4.114) tells us that when a beam is accelerated (or, strictly speaking, when a change is made in the reference momentum), it is not the emittance  $\epsilon_x$  that is conserved, but the quantity  $\beta_0\gamma_0\epsilon_x$ . This is an important result when dealing with beam dynamics in accelerators. The quantity  $\beta_0\gamma_0\epsilon_x$  is called the *normalised emittance*. To emphasise the distinction from the normalised emittance, the quantity  $\epsilon_x$  is sometimes referred to as the *geometric emittance*. Since the normalised emittance is conserved, an increase in the reference momentum leads to a reduction in the geometric emittance: this effect is referred to as *adiabatic damping*.

Since adiabatic damping is associated with a change in the reference momentum, and the reference momentum can be chosen arbitrarily, it may appear that adiabatic damping is an artificial effect. However, it is important to remember two facts: first, that the canonical momenta are defined in terms of the reference momentum; and second, that changing the reference momentum changes the Courant–Snyder parameters as well as the geometric emittance. Consider the acceleration of a beam in a section of linac. Without a change in the reference momentum, the beam dynamics would be described by certain values of the beta function, and a fixed

geometric emittance. If the reference momentum is changed to match the acceleration, then the beam dynamics are described by a different (larger) beta function, and a different (smaller) geometric emittance. Usually, the choice is made to scale the reference momentum with the nominal momentum of the beam. In that case, for a given set of beta functions, it should be remembered that there is a real reduction in the geometric emittance. The corresponding reduction in the phase space area occupied by particles in the beam can be understood in terms of the change in the canonical momenta associated with a change in the reference momentum.

**This page intentionally left blank**

## Chapter 5

# Coupled Optics

In Chapter 4, we developed a description of the dynamics of a single particle in an accelerator beam line using action–angle variables. Throughout that chapter, we assumed that the motion in the horizontal, vertical, or longitudinal direction are independent of the motion in the other directions; in other words, we assumed that there was no coupling. The analysis of beam dynamics is certainly simplified when we need consider only one degree of freedom at a time; however, coupling will almost always be present in an accelerator in one form or another, either as an intrinsic feature of the system, or by design, or as a result of imperfections in the construction or tuning of the accelerator. In this chapter, we shall first consider coupling between the longitudinal and the transverse planes. This is a particularly important case, since it arises whenever dipole magnets are used in a beam line: the bending angle, and hence the transverse trajectory, depends on the particle energy (specified by the energy deviation, a longitudinal variable). The dependence of the transverse trajectory on the particle energy is described by the *dispersion*, which is a function of position along the beam line. Another aspect of coupling between the transverse and longitudinal degrees of freedom is *chromaticity*, the dependence of the focusing strength of a quadrupole magnet on the energy of a particle. However, as already mentioned in Chapter 3, chromaticity is properly treated as a higher-order effect, and in the present chapter we shall limit the discussion to linear coupling. Chromaticity will be discussed in detail in Chapter 8.

Coupling between horizontal and vertical motion can occur in a beam line either by design (for example, because of the inclusion of skew quadrupole or solenoid magnets), or as a result of alignment errors on the magnets (such as the tilt of a quadrupole around its magnetic axis). It is important to be able to describe coupling and its effects on the beam, and

there are several methods that have been developed to do this in a convenient way. Unfortunately, no single method has been adopted as a universal standard, and it would not be practical to try to cover all (or even several) of the methods that are in use here. Therefore, we restrict our discussion to a single technique that provides a reasonably simple and intuitive way to generalise quantities introduced for uncoupled linear optics, in particular the action-angle variables and Courant-Snyder parameters, to the case of coupled motion with an arbitrary number of degrees of freedom.

## 5.1 Transverse-Longitudinal Coupling

### 5.1.1 Dispersion

In Chapter 4, we considered single-particle dynamics in a FODO beam line, consisting of a sequence of quadrupole magnets arranged in a straight line. In that case, the transfer matrix for each periodic cell (4.9) was made up from  $2 \times 2$  matrices on the block diagonal, so that the horizontal, vertical and longitudinal motions could each be treated independently. In practice, accelerator beam lines generally include dipole magnets to change the direction of the beam. Let us consider a dipole magnet without any field gradient ( $k_1 = 0$ ), and with curvature  $h$  of the reference trajectory equal to the field scaled by the beam rigidity:

$$h = k_0 = \frac{q}{P_0} B_0 = \frac{B_0}{B\rho}, \quad (5.1)$$

where  $q$  is the particle charge,  $P_0$  the reference momentum,  $B_0$  the dipole field and  $B\rho$  is the beam rigidity. Then, the linear transfer map for the dipole is:

$$\vec{x}_1 = R\vec{x}_0, \quad (5.2)$$

where, from (3.56):

$$R = \begin{pmatrix} \cos(\omega L) & \frac{\sin(\omega L)}{\omega} & 0 & 0 & 0 & \frac{1-\cos(\omega L)}{\omega\beta_0} \\ -\omega \sin(\omega L) & \cos(\omega L) & 0 & 0 & 0 & \frac{\sin(\omega L)}{\beta_0} \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -\frac{\sin(\omega L)}{\beta_0} & -\frac{1-\cos(\omega L)}{\omega\beta_0} & 0 & 0 & 1 & \frac{L}{\beta_0^2\gamma_0^2} - \frac{\omega L - \sin(\omega L)}{\omega\beta_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (5.3)$$

Here,  $L$  is the length of the dipole,  $\omega = k_0$ , and  $\beta_0$  and  $\gamma_0$  are respectively the scaled velocity and relativistic factor for a particle with the reference

momentum. Note that if  $h = k_0$ , then there are no zeroth-order terms in the transfer map (5.2).

The transfer matrix for a dipole (5.3) differs from that for a straight FODO cell by having non-zero elements off the block diagonals. In particular, the transfer matrix elements  $R_{16}$  and  $R_{26}$  represent variations in the horizontal co-ordinate and momentum at the exit of the dipole, depending on the energy deviation of the particle. The physical origin of these elements is easily understood: they arise directly from the fact that the trajectory of a charged particle in a dipole field depends on the energy of the particle. Together, the elements  $R_{16}$  and  $R_{26}$  represent dispersion from the dipole. Note that the elements off the block diagonal  $R_{51}$  and  $R_{52}$  are also non-zero: these elements represent variations in the path length depending on the horizontal trajectory. Mathematically, they are related to  $R_{16}$  and  $R_{26}$  by the symplectic condition. Physically, the path length depends on the transverse co-ordinates because of the curvature of the reference trajectory. Because the field in this case is vertical, the bending of the trajectory is purely in the horizontal plane, and there is no vertical dispersion; the elements  $R_{36}$  and  $R_{46}$  are both zero, as are  $R_{53}$  and  $R_{54}$ .

Introducing a dipole magnet into an otherwise straight beam line will modify the transfer matrix across any section of the beam line that includes the dipole magnet. However, the structure of the transfer matrix will always, for a horizontally bending magnet, be the same as (5.3). That is, the horizontal trajectory will depend on the energy deviation, and the path length will depend on the horizontal trajectory. The vertical motion will remain independent of the horizontal and longitudinal motion, and as long as we do not include time-dependent fields (as in rf cavities, for example) the horizontal motion will be independent of the longitudinal co-ordinate  $z$ . In that case, the horizontal motion can be described in terms of a  $2 \times 2$  matrix for the first-order terms, with zeroth-order terms that depend on the energy deviation  $\delta$  (which we treat as a constant parameter). Consider inserting a dipole magnet into a FODO cell, and then repeating the FODO cell so that the beam line closes on itself: in other words, we construct a FODO storage ring. The transfer map for the horizontal motion in one periodic cell, starting at some point  $s_0$  along the beam line, can be written:

$$\begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0+L} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} \begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0} + \begin{pmatrix} R_{16} \\ R_{26} \end{pmatrix} \delta, \quad (5.4)$$

where  $L$  is the length of the cell, and  $R$  is the transfer matrix from  $s = s_0$  to  $s = s_0 + L$ . The energy deviation  $\delta$  is constant. If a particle has zero

energy deviation, and it enters a cell with  $x = p_x = 0$ , then it will exit the cell with  $x = p_x = 0$ ; thus, the particle will simply follow the reference trajectory around the entire ring. This is a rather trivial result, but the dynamics become more interesting if we consider particles with non-zero energy deviation. We can try to find a periodic trajectory by imposing the condition:

$$\begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0+L} = \begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0}. \quad (5.5)$$

In the case of a storage ring, the condition (5.5) determines the horizontal closed orbit, that is the particle trajectory that closes on itself after one complete turn. Properly, the closed orbit in a storage ring is defined in six-dimensional phase space, but here we consider only the horizontal orbit. Combining (5.4) and (5.5) to solve for  $x$  and  $p_x$  at  $s = s_0$  we find:

$$\begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0} = \begin{pmatrix} 1 - R_{11} & -R_{12} \\ -R_{21} & 1 - R_{22} \end{pmatrix}^{-1} \begin{pmatrix} R_{16} \\ R_{26} \end{pmatrix} \delta. \quad (5.6)$$

In a periodic beam line (such as a storage ring), the *dispersion* is defined in terms of the change in the periodic trajectory with respect to the momentum. Explicitly, we write:

$$x(\delta_p) = x|_{\delta_p=0} + \eta_x \delta_p + \eta_x^{(2)} \delta_p^2 + \dots, \quad (5.7)$$

$$p_x(\delta_p) = p_x|_{\delta_p=0} + \eta_{px} \delta_p + \eta_{px}^{(2)} \delta_p^2 + \dots. \quad (5.8)$$

The *momentum deviation*  $\delta_p$  is defined by:

$$\delta_p = \frac{P}{P_0} - 1, \quad (5.9)$$

where  $P$  is the total momentum of the particle, and  $P_0$  is the reference momentum. The coefficient  $\eta_x$  is the dispersion;  $\eta_x^{(2)}$  is the second-order dispersion, and so on. The co-ordinate  $x$  and canonical momentum  $p_x$  in (5.7) and (5.8) are subject to the constraint (5.5).

Note that the momentum deviation  $\delta_p$  is not the same as the energy deviation  $\delta$ . The dispersion  $\eta_x$  is defined in terms of the momentum deviation, while the expression for the change in trajectory in terms of the transfer matrix (5.6) is written in terms of the energy deviation. To proceed, it is useful to have a relationship between the derivative with respect to  $\delta_p$  and the derivative with respect to  $\delta$ . Since  $\delta_p = 0$  when  $\delta = 0$ , we can write:

$$\frac{d}{d\delta_p} \Big|_{\delta_p=0} = \frac{d\delta}{d\delta_p} \Big|_{\delta_p=0} \frac{d}{d\delta} \Big|_{\delta=0}. \quad (5.10)$$

Using (2.55), and:

$$E^2 = P^2 c^2 + m^2 c^4, \quad (5.11)$$

we find:

$$\frac{d\delta}{d\delta_p} \Big|_{\delta_p=0} = P_0 \frac{d\delta}{dP} \Big|_{P=P_0} = \beta_0. \quad (5.12)$$

Hence:

$$\frac{d}{d\delta_p} \Big|_{\delta_p=0} = \beta_0 \frac{d}{d\delta} \Big|_{\delta=0}. \quad (5.13)$$

From (5.7), (5.8) and (5.13), we can write the dispersion as:

$$\begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_0} = \frac{d}{d\delta_p} \Big|_{\delta_p=0} \begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0} = \beta_0 \frac{d}{d\delta} \Big|_{\delta=0} \begin{pmatrix} x \\ p_x \end{pmatrix}_{s=s_0}. \quad (5.14)$$

Then, using (5.6) we find:

$$\begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_0} = \beta_0 \begin{pmatrix} 1 - R_{11} & -R_{12} \\ -R_{21} & 1 - R_{22} \end{pmatrix}^{-1} \begin{pmatrix} R_{16} \\ R_{26} \end{pmatrix}. \quad (5.15)$$

In (5.15), the phase space co-ordinates  $x$  and  $p_x$  are subject to the periodicity condition (5.5). Note that the quantities  $R_{ij}$  are elements of the transfer matrix for one periodic cell of the beam line, starting from the point  $s = s_0$ . The existence of the dispersion function depends on the existence of the matrix inversion appearing in (5.15). Strictly speaking, equation (5.15) is only valid if there are no rf cavities (or other components producing time-dependent fields) in the beam line, and if the energy deviation is constant. We can of course generalise the dispersion to the vertical motion: vertical dispersion can be generated either by vertical bending, or by coupling the horizontal dispersion into the vertical plane (for example, in a skew quadrupole or a solenoid).

Equation (5.15) provides a way to determine the dispersion at a given point in a periodic beam line. Often, we would like to know the dispersion at each point along the entire length of a beam line. Since the dispersion is effectively the trajectory of a particle with non-zero energy deviation, we can evolve the dispersion along a beam line using the transfer maps from one point in the beam line to another. In drift spaces and normal quadrupoles, the dispersion transforms in the same way as the horizontal trajectory. In dipoles, there are additional (zeroth-order) terms to account for the fact that the trajectory of a particle in a dipole depends on the

energy of the particle. In a beam line consisting of drift spaces, normal quadrupoles and horizontally bending dipoles, we can write:

$$\begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_1} = \begin{pmatrix} R_{11}(s_1, s_0) & R_{12}(s_1, s_0) \\ R_{21}(s_1, s_0) & R_{22}(s_1, s_0) \end{pmatrix} \begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_0} + \beta_0 \begin{pmatrix} R_{16}(s_1, s_0) \\ R_{26}(s_1, s_0) \end{pmatrix}, \quad (5.16)$$

where the  $R_{ij}(s_1, s_0)$  are elements of the transfer matrix  $R(s_1, s_0)$  from  $s_0$  to  $s_1$ .

Equation (5.16) provides a way to propagate the dispersion along a beam line even where the beam line is not periodic, provided that the initial values of the dispersion are known. In a periodic beam line, the initial values are determined by the condition (5.5), which leads to (5.15). This situation is analogous to the calculation of the Courant–Snyder parameters: in a periodic beam line, the requirement that the Courant–Snyder parameters have the same periodicity as the beam line leads to constraints that determine unique values for the Courant–Snyder parameters. If the beam line is not periodic, then we need some additional information to determine initial values that can then be propagated along the beam line using the transfer matrices. For the Courant–Snyder parameters, appropriate initial values are determined by the distribution of particles within a bunch, either from design or measurement. In the case of the dispersion, we can use the variation of the trajectory of a particle (or the mean trajectory of particles within a bunch) with respect to the beam energy to determine appropriate values for the dispersion at some point in the beam line. The variation of trajectory with energy can either be given as a design specification, or measured experimentally.

The higher-order dispersion terms ( $\eta_x^{(2)}$  etc.) arise from nonlinearities in the lattice. In practice, they can be calculated numerically using a computer model that includes the relevant nonlinear effects.

### **5.1.2 Momentum compaction and phase slip**

The revolution frequency of a particle in a storage ring depends on the energy of the particle for two reasons. First, as the energy increases, the speed of the particle increases. However, for ultra-relativistic particles (with relativistic factor  $\gamma \gg 1$ ), the increase in speed accompanying any increase in energy is small. The second reason for the dependence of revolution frequency on particle energy is usually significant even at high energy: the presence of dispersion in a storage ring means that the length of the trajectory followed by a particle as it moves around the ring depends on the

energy of the particle. In most cases, the path length increases with energy: this would be the case, for example, for a particle following a circular path in a uniform magnetic field. The result is that for energies below some threshold value (called the *transition energy*) the revolution frequency increases with energy, as the increase in speed dominates over the increase in path length; but above the transition energy, the revolution frequency decreases as the energy increases, as the change in path length dominates over the change in speed.

In designing and operating a storage ring, it is important to know how both the path length and the revolution frequency vary with energy. In particular, since storage rings generally include rf cavities with fields oscillating at fixed frequencies, the dependence of the revolution frequency of a particle on the energy of the particle plays an important role in the longitudinal dynamics. We will discuss the dependence of the revolution frequency on energy in Section 5.1.2; but first, we shall consider the dependence of the path length on energy.

Consider a beam line in which the reference trajectory has total length  $C_0$ . We assume that the reference trajectory is chosen so that it is the physical trajectory for a particle with initial phase space vector  $(x, p_x, y, p_y, z, \delta) = 0$ . Now consider a short section of the beam line of length  $ds$  in which the reference trajectory has radius of curvature  $\rho$  in the horizontal plane. A particle with co-ordinate  $x$  follows a trajectory with radius of curvature  $\rho + x$ ; the path length  $dC$  corresponding to a distance  $ds$  along the reference trajectory is:

$$dC = (\rho + x) d\theta = (\rho + x) \frac{ds}{\rho}, \quad (5.17)$$

where  $d\theta$  is the angle subtended by the distance  $ds$  along the reference trajectory at the centre of curvature: see Fig. 5.1. The total path length  $C$  followed by a particle (neglecting any vertical motion) as it travels from the start to the end of the beam line is:

$$C = \int_0^{C_0} \left( 1 + \frac{x}{\rho} \right) ds = C_0 + \int_0^{C_0} \frac{x}{\rho} ds. \quad (5.18)$$

Now suppose that the value of  $x$  at any point along the beam line is determined entirely by the dispersion and the energy deviation, i.e. that the horizontal action  $J_x$  is zero. Then, from (5.7), we can write:

$$x = \eta_x \delta_p + \eta_x^{(2)} \delta_p^2 + \dots \quad (5.19)$$

We define the *momentum compaction factor*  $\alpha_p$  to be the relative change in path length with respect to momentum for particles with zero momentum

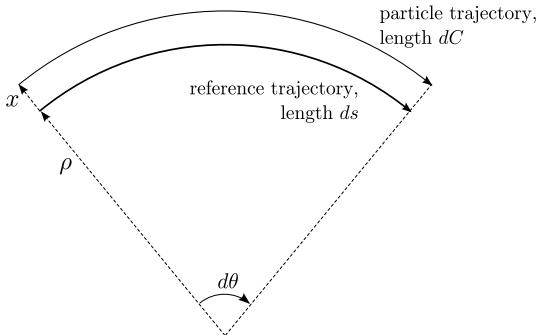


Fig. 5.1 The angle subtended at the centre of curvature by the trajectory of a particle travelling parallel to the reference trajectory is  $d\theta = dC/(x + \rho) = ds/\rho$ , where  $dC$  is the length of the trajectory,  $x$  is the horizontal co-ordinate with respect to the reference trajectory,  $ds$  is the corresponding distance along the reference trajectory, and  $\rho$  is the radius of curvature of the reference trajectory. Thus, the path length  $dC$  is given in terms of  $ds$  by  $dC = (1 + x/\rho) ds$ .

deviation. Combining (5.18) and (5.19) we have:

$$\alpha_p = \frac{1}{C_0} \frac{dC}{d\delta_p} \Big|_{\delta_p=0} = \frac{1}{C_0} \int_0^{C_0} \frac{\eta_x}{\rho} ds. \quad (5.20)$$

A change in path length with respect to energy arises from the presence of dispersion in sections of the beam line that have non-zero curvature (where the curvature is defined as  $1/\rho$ ). In sections of the beam line where the reference trajectory is a straight line, we have  $\rho \rightarrow \infty$ : such sections make no contribution to the integral in (5.20).

In a storage ring, the *first synchrotron radiation integral*  $I_1$  is defined:

$$I_1 = \oint \frac{\eta_x}{\rho} ds, \quad (5.21)$$

where the integral is taken around the entire circumference of the ring. The momentum compaction factor in a storage ring can then be written:

$$\alpha_p = \frac{I_1}{C_0}. \quad (5.22)$$

Just as we defined higher-order dispersion terms in (5.7), we can define higher-order momentum compaction factors by writing the path length as a power series in the energy deviation:

$$\frac{C}{C_0} = 1 + \alpha_p \delta_p + \alpha_p^{(2)} \delta_p^2 + \alpha_p^{(3)} \delta_p^3 + \dots \quad (5.23)$$

The fact that dispersion is required to generate momentum compaction indicates that momentum compaction is a consequence of coupling between

the transverse and longitudinal degrees of freedom. Momentum compaction is actually an indirect effect of coupling: the direct coupling effect is the dependence of the transverse co-ordinates on the momentum deviation. However, the dependence of the transverse co-ordinates on momentum couples back into the longitudinal direction because of the curvature of the reference trajectory, and results in a variation of path length with respect to momentum deviation.

The momentum compaction factor describes the change in the length of a particle trajectory along a beam line resulting from a change in momentum. It is often useful to know also how the time of flight along a beam line depends on the momentum: this is described by the *phase slip factor*  $\eta_p$  (not to be confused with the dispersion  $\eta_x$ ). The phase slip factor is defined by analogy with the momentum compaction factor (5.20):

$$\eta_p = \frac{1}{T_0} \frac{dT}{d\delta_p} \Big|_{\delta_p=0}, \quad (5.24)$$

where  $T$  is the time taken for a particle to travel between two specified points in a beam line. Just as we defined higher-order momentum compaction factors (5.23), we can define higher-order phase slip factors by writing the time of flight as a power series in the momentum deviation:

$$\frac{T}{T_0} = 1 + \eta_p \delta_p + \eta_p^{(2)} \delta_p^2 + \eta_p^{(3)} \delta_p^3 + \dots \quad (5.25)$$

Let us consider a particle moving with constant momentum, and hence with constant speed. The time of flight  $T$  along a beam line can be written in terms of the length of the trajectory  $C$  along the beam line:

$$T = \frac{C}{\beta c}. \quad (5.26)$$

Hence:

$$\frac{1}{T} \frac{dT}{d\delta_p} = \frac{1}{C} \frac{dC}{d\delta_p} - \frac{1}{\beta c} \frac{d\beta}{d\delta_p}. \quad (5.27)$$

Using the definition of the momentum deviation (5.9), we have:

$$\beta\gamma = (1 + \delta_p) \beta_0 \gamma_0, \quad (5.28)$$

where  $\beta_0 c$  and  $\gamma_0$  are (respectively) the speed and relativistic factor for a particle with the reference momentum  $P_0$ . From (5.28) it follows that:

$$\frac{d(\beta\gamma)}{d\delta_p} = \gamma^3 \frac{d\beta}{d\delta_p} = \beta_0 \gamma_0. \quad (5.29)$$

Using the result (5.29) in (5.27), and evaluating the terms at  $\delta = 0$  gives:

$$\eta_p = \alpha_p - \frac{1}{\gamma_0^2}. \quad (5.30)$$

The momentum compaction factor is determined principally by the design of the beam line (in particular, by the dispersion and the curvature of the reference trajectory), whereas the value of  $\gamma_0$  depends on the beam energy. Therefore, depending on the design of the beam line and the beam energy, the phase slip factor may be negative, zero or positive. In a storage ring, the momentum compaction factor is usually positive: this may be understood from (5.20), remembering that the circumference of a particle trajectory generally increases with the momentum of the particle. Therefore, the phase slip factor in a storage ring is usually positive for beams of sufficiently high energy; such a ring is said to be operating ‘above transition’. The relativistic factor at the transition energy  $\gamma_T$  is defined by the condition  $\eta_p = 0$ , in which case, from (5.30):

$$\gamma_T = \frac{1}{\sqrt{\alpha_p}}. \quad (5.31)$$

Above transition, the revolution period increases with increasing momentum, because the increase in speed of a particle is insufficient to compensate the increase in path length. At low energy  $\gamma < \gamma_T$ , the phase slip factor is negative: a storage ring in this regime is said to be operating ‘below transition’, in which case the revolution period decreases with increasing momentum. The speed of a particle in a storage ring below transition is sufficiently slow compared to the speed of light, that increasing the momentum leads to an increase in speed that more than compensates the increase in path length around the ring. The phenomenon of transition has implications for the operation of synchrotrons used for accelerating particles: we shall discuss this in more detail in Chapter 7.

Since the phase slip factor characterises the variation of revolution period with respect to the momentum deviation, we can express  $\eta_p$  in terms of the transfer matrix for a complete turn of a storage ring. Recall that the longitudinal co-ordinate  $z$  is defined by (2.54):

$$z = \frac{s}{\beta_0} - ct, \quad (5.32)$$

where the particle crosses the plane perpendicular to the reference trajectory at  $s$  at time  $t$ . Suppose a particle is at  $s = s_0$  at  $t = t_0$ , and completes one revolution in time  $T$  (so that the particle is at  $s = s_0 + C_0$  at time

$t = t_0 + T$ ). The change in the longitudinal co-ordinate over one turn is:

$$\Delta z = \frac{C_0}{\beta_0} - cT. \quad (5.33)$$

Now suppose that there is zero dispersion at the point  $s_0$ . If we construct the transfer matrix  $R$  for a complete turn of the ring starting at  $s_0$ , then the revolution period for a particle will be completely determined by the particle energy and the  $R_{56}$  element of the transfer matrix. Note that the dispersion must be zero for this to be the case. If there is dispersion at  $s_0$ , then a change in the energy deviation would change the co-ordinates at  $s_0$  for a particle following a closed orbit; and the change in the co-ordinates would contribute (through the  $R_{51}$  and  $R_{52}$  matrix elements) to the change in revolution period. If there is zero dispersion at  $s_0$ , then  $R_{16} = R_{26} = 0$ , and, as a consequence of the symplectic condition for the transfer matrix, we also have  $R_{51} = R_{52} = 0$ . We can then write the following relationship between the  $R_{56}$  matrix element and the revolution period:

$$R_{56} = \left. \frac{\partial}{\partial \delta} \right|_{\delta=0} \Delta z = -c \left. \frac{dT}{d\delta} \right|_{\delta=0}. \quad (5.34)$$

We use (5.13) to change the derivative with respect to energy deviation  $\delta$  into a derivative with respect to momentum deviation  $\delta_p$ . Then, using (5.24), we obtain:

$$R_{56} = -c \frac{T_0}{\beta_0} \eta_p = -\frac{C_0}{\beta_0^2} \eta_p. \quad (5.35)$$

Equation (5.35) generalises to the case where  $R$  is the transfer matrix between any two points  $s_0$  and  $s_1$  in a beam line, where  $C_0 = s_1 - s_0$ , as long as the dispersion at  $s_0$  is zero. The phase slip factor can be calculated from the momentum compaction factor (5.20) and the reference momentum using (5.30). If there is dispersion at  $s_0$ , then the contribution of the transfer matrix elements  $R_{51}$  and  $R_{52}$  (in the case of horizontal dispersion) to the change in the longitudinal co-ordinate  $z$  need to be taken into account.

It is desirable in some situations for the time of flight of a particle between two specified points in a beam line to be independent of the energy deviation, at least to first order. The transfer matrix between the specified points has the property:

$$R_{56} = 0. \quad (5.36)$$

In this case, the section of beam line between the specified points are said to be *isochronous*.

### **5.1.3 Synchrotron motion**

In a synchrotron storage ring, there is coupling between the transverse and longitudinal motion of particles, because of the presence of dispersion generated by the dipole magnets. The coupling between the transverse and longitudinal motion plays an important role in the longitudinal (synchrotron) oscillations performed by particles as they move around the ring. In this section, we shall discuss the longitudinal motion, with the goal of understanding the main characteristics, such as the oscillation frequency and the shape of the ellipse mapped out by a particle in longitudinal phase space. Knowing how the longitudinal dynamics depends on such parameters as the phase slip factor and rf voltage is important for the design and operation of storage rings.

Although the longitudinal and transverse dynamics in a storage ring are intrinsically coupled, this coupling appears implicitly rather than explicitly in our analysis, essentially being contained within the phase slip factor,  $\eta_p$ . For simplicity, for the moment we shall neglect transverse (betatron) oscillations, and only consider the synchrotron oscillations. A method for the study of fully coupled motion, in which particles perform betatron and synchrotron oscillations simultaneously, will be discussed in Section 5.2.

Outlined, synchrotron oscillations are easily understood. Consider a reference particle in a storage ring operating above transition. The reference particle has zero energy deviation, and arrives at the rf cavities at exactly the right time each turn for the electric field in the cavities to restore the energy lost by synchrotron radiation. Now consider a particle moving close to the reference particle, but with positive energy deviation. Since the ring is above transition, this second particle will tend to slip back, relative to the reference particle, on each turn. If the rf phase is such that particles arriving at the cavities late receive less energy than those arriving early, the particle with initially positive energy deviation will eventually lose enough energy (through synchrotron radiation, or through deceleration in the rf cavities) that its energy deviation becomes negative. At that point, its revolution frequency increases, and it starts to catch up with the reference particle. Eventually, its energy deviation becomes positive once again, and the whole cycle repeats.

Having outlined the processes involved in synchrotron oscillations, let us consider the dynamics in more detail. Our goal will be to derive expressions for the synchrotron oscillation frequency, and for the ratio of energy spread to bunch length in an invariant beam distribution. We shall also

find that, as a result of the sinusoidal variation of the electric field in the rf cavities, there is a maximum energy deviation that a particle can have while performing stable synchrotron oscillations.

Consider a single particle moving through a synchrotron storage ring. From the definition of the longitudinal co-ordinate  $z$  (2.54), the change in  $z$  for a particle travelling with speed  $\beta c$  over one turn of the storage ring is:

$$\Delta z = \frac{C_0}{\beta_0} - \frac{C}{\beta}, \quad (5.37)$$

where  $C_0$  is the circumference along the reference trajectory, and  $C = \beta cT$  is the actual path length in time  $T$  that it takes the particle to complete one revolution. If we assume that we can average the change in  $z$  over the entire circumference, we have:

$$\frac{dz}{ds} \approx \frac{\Delta z}{C_0} = \frac{1}{\beta_0} \left( 1 - \frac{T}{T_0} \right), \quad (5.38)$$

where  $T_0$  is the revolution period for the reference particle. The revolution periods  $T$  and  $T_0$  are related by the phase slip factor (5.25) and the momentum deviation  $\delta_p$ ; thus, we can write the equation of motion for  $z$ :

$$\frac{dz}{ds} \approx -\frac{\eta_p}{\beta_0} \delta_p. \quad (5.39)$$

To express the equation of motion in terms of canonical variables, we need to replace the momentum deviation  $\delta_p$  by the energy deviation  $\delta$  (given by (2.55)). For particles with small momentum deviation, we can say that:

$$\delta_p \approx \frac{\delta}{\beta_0}. \quad (5.40)$$

Then, the equation of motion for the longitudinal co-ordinate  $z$  becomes:

$$\frac{dz}{ds} = -\frac{\eta_p}{\beta_0^2} \delta. \quad (5.41)$$

Now consider the change in the energy deviation  $\Delta\delta$  over one revolution of the ring. This is given by the difference between the energy gained from the rf cavities and the energy lost through synchrotron radiation:

$$\Delta\delta = \frac{qV_{\text{rf}}}{cP_0} \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}}z}{c}\right) - \frac{U_0}{cP_0}, \quad (5.42)$$

where  $q$  is the electric charge on the particle,  $V_{\text{rf}}$  is the amplitude of the rf voltage,  $\omega_{\text{rf}}$  is the rf frequency,  $\phi_{\text{rf}}$  is a fixed phase, and  $E_0$  is the reference energy.  $U_0$  is the energy lost per turn by the particle, through synchrotron radiation (7.20) or other effects. Synchrotron radiation will be discussed in

Chapter 7: there, we shall find that the energy radiated by a charged particle depends on the energy of the particle, and that this leads to damping of the synchrotron oscillations. We shall ignore the damping effect for now, and just assume that  $U_0$  is a constant, determined by the beam energy and the lattice design (the magnetic field strengths).

In general, a synchrotron storage ring will contain a number of rf cavities. Here, we assume that the total rf voltage from all the cavities can be represented by a single amplitude  $V_{\text{rf}}$ , frequency  $\omega_{\text{rf}}$  and phase  $\phi_{\text{rf}}$ .

Again assuming that we can average the change in the energy of a particle over the entire circumference, the equation of motion for the energy deviation is:

$$\frac{d\delta}{ds} = \frac{qV_{\text{rf}}}{cP_0C_0} \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}}z}{c}\right) - \frac{U_0}{cP_0C_0}. \quad (5.43)$$

Equations (5.41) and (5.43) describe the longitudinal dynamics of a particle in a synchrotron storage ring. Taking the derivative of (5.41) with respect to  $s$  and substituting for  $d\delta/ds$  from (5.43) gives:

$$\frac{d^2z}{ds^2} = -\frac{qV_{\text{rf}}}{cP_0C_0} \frac{\eta_p}{\beta_0^2} \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}}z}{c}\right) + \frac{\eta_p}{\beta_0^2} \frac{U_0}{cP_0C_0}. \quad (5.44)$$

If the rf phase is set so that  $\phi_{\text{rf}} = \phi_s$ , where:

$$\sin(\phi_s) = \frac{U_0}{qV_{\text{rf}}}, \quad (5.45)$$

then expanding (5.44) to first order in the dynamical variable  $z$  gives:

$$\frac{d^2z}{ds^2} = -k_z^2 z, \quad (5.46)$$

where:

$$k_z^2 = -\frac{qV_{\text{rf}}}{cP_0} \frac{\omega_{\text{rf}}}{cC_0} \frac{\eta_p}{\beta_0^2} \cos(\phi_s). \quad (5.47)$$

We see that if:

$$qV_{\text{rf}}\eta_p \cos(\phi_s) < 0, \quad (5.48)$$

then  $k_z^2 > 0$  and a particle with motion described by (5.46) will perform stable oscillations with frequency  $\omega_z$  about the reference particle; the reference particle will see a fixed rf phase of  $\phi_s$ , which is known as the *synchronous phase*. The *synchrotron tune*  $\nu_z$  is the number of synchrotron oscillations completed in one turn of the storage ring (i.e. over a distance  $C_0$ ):

$$\nu_s = \frac{k_z C_0}{2\pi} = \frac{1}{2\pi} \sqrt{-\frac{qV_{\text{rf}}}{cP_0} \frac{\omega_{\text{rf}} C_0}{c} \frac{\eta_p}{\beta_0^2} \cos(\phi_s)}. \quad (5.49)$$

The phase slip factor  $\eta_p$  describes the change in revolution period of a particle with respect to the energy of the particle. If  $\eta_p < 0$ , i.e. if the storage ring is operating below transition, then the revolution period decreases as the energy of the particle increases. The increase in velocity with energy more than compensates the change in path length that also accompanies a change in energy of a particle in a storage ring. In a storage ring operating above transition, any increase in velocity with increasing energy is negligible; then assuming that the momentum compaction factor  $\alpha_p > 0$ , the revolution period will increase with increasing energy, as a result of increasing path length. Assuming that  $qV_{\text{rf}} > 0$ , then for stable synchrotron oscillations in a ring below transition ( $\eta_p < 0$ ), the synchronous phase must satisfy  $\sin(\phi_s) > 0$  and  $\cos(\phi_s) > 0$ ; hence:

$$0 < \phi_s < \frac{\pi}{2}. \quad (5.50)$$

For a ring above transition ( $\eta_p > 0$ ), the synchronous phase must satisfy  $\sin(\phi_s) > 0$  and  $\cos(\phi_s) < 0$ , and hence the synchronous phase must lie in the range:

$$\frac{\pi}{2} < \phi_s < \pi. \quad (5.51)$$

The simple harmonic motion implied by (5.46) is based on two approximations: first, that the synchrotron motion is slow compared to the revolution period, so that phase slip effects in the dipoles and energy changes in the dipoles and rf cavities can be smoothed out over the entire circumference; and second, that the oscillation amplitude is small compared to the rf wavelength. The first approximation is valid in the large majority of cases: typically, the synchrotron tune in a storage ring will be of order 0.01. The second approximation is often valid for particles in bunches with the equilibrium distribution determined (in electron storage rings) by synchrotron radiation effects, but cannot be assumed to be valid in general. For example, particles injected into a storage ring may initially perform synchrotron oscillations with very large amplitude; also, scattering effects (which will be discussed in Chapter 13) can lead to the excitation of large amplitude synchrotron oscillations of individual particles. At very large amplitudes, synchrotron oscillations can become unstable, leading to the loss of particles from the ring. The maximum amplitude of the energy deviation for a particle performing stable synchrotron oscillations in a storage ring is characterised by the *energy acceptance*. To understand the longitudinal dynamics at large synchrotron amplitude, it is helpful to develop a description of the system using Hamiltonian mechanics.

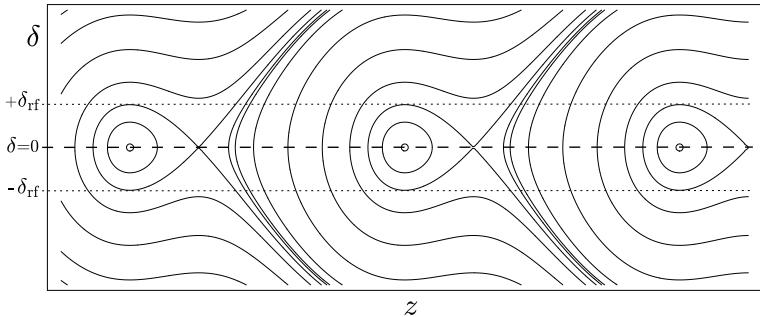


Fig. 5.2 Longitudinal phase space portrait for a synchrotron storage ring. The lines show contours of constant value of the Hamiltonian  $H$ , given by (5.52), and correspond to the phase space trajectories of particles in the storage ring. The positions of the stable fixed points (centres of the closed loops) and the unstable fixed points (where contours intersect) are given by (5.55) and (5.56) respectively: the fixed points here occur along the line  $\delta = 0$ , shown by a dashed line in the figure. The rf bucket height (or rf acceptance)  $\delta_{rf}$  is the maximum value of the energy deviation  $\delta$  within a closed loop, and is given by (5.59) (dotted lines in the figure). The distance between adjacent rf buckets corresponds to one rf wavelength.

The equations of motion (5.41) and (5.43) may be obtained from the Hamiltonian:

$$H = \frac{qV_{rf}}{\omega_{rf}P_0C_0} \left( \sin(\phi_s) \frac{\omega_{rf}z}{c} - \cos\left(\phi_s - \frac{\omega_{rf}z}{c}\right) \right) - \frac{1}{2} \frac{\eta_p}{\beta_0^2} \delta^2, \quad (5.52)$$

where we have used (5.45) to eliminate  $U_0$ . Particles performing synchrotron oscillations will follow contours of constant value of the Hamiltonian in longitudinal phase space. A longitudinal phase space portrait is shown in Fig. 5.2. Note the existence of a series of stable fixed points at intervals along the line  $\delta = 0$ : the stable fixed points occur at the centres of the closed loops visible in the phase space portrait. A particle at a stable fixed point has zero energy deviation, and arrives at the rf cavities at the synchronous phase, i.e. so that the energy gain from the rf cavities exactly compensates the energy lost by synchrotron radiation. Particles performing stable synchrotron oscillations follow closed loops around the stable fixed points.

Also visible in the longitudinal phase space portrait in Fig. 5.2 is a sequence of unstable fixed points: these occur where contour lines cross. At an unstable fixed point, particles again arrive at the rf cavities at a phase such that the energy gain exactly cancels the energy lost by synchrotron radiation; however, the rate of change of the field in the cavity has the opposite sign to that at the stable fixed points. Thus, a particle with a

positive energy deviation arriving initially at a cavity with the correct phase to compensate the synchrotron radiation energy loss will, over successive turns, gain *more* energy, leading to an increase in the energy deviation. The result is that the particle will not perform oscillations in longitudinal phase space, but will continually gain (or lose) energy, until it is lost from the storage ring.

The presence of unstable fixed points as well as stable fixed points is a consequence of the sinusoidal variation of the fields in the rf cavities. For stable oscillations, particles must arrive at the rf cavities not only when the rf voltage has (approximately) the correct value, but also when the rate of change of the voltage has the correct sign. The fact that stable synchrotron oscillations can occur when the rf voltage meets the appropriate conditions is known as the principle of *phase stability*.

The contours passing through the unstable fixed points in the longitudinal phase space diagram are known as *separatrices*: a separatrix forms a boundary between a region of stable motion and a region of unstable motion. The regions of stable motion are known as *rf buckets*. The *rf acceptance* of a storage ring is the maximum energy deviation within an rf bucket: if a particle has an energy deviation larger than the rf acceptance, it will be lost from the storage ring. There may be other, smaller limits on the maximum energy deviation for which stable motion is possible; however, the rf acceptance is still a significant quantity for making estimates of properties such as the beam lifetime. We can find an expression for the rf acceptance in a storage ring with given parameters, as follows.

The equations of motion from the Hamiltonian (5.52) are:

$$\frac{dz}{ds} = -\frac{\eta_p}{\beta_0^2} \delta, \quad (5.53)$$

$$\frac{d\delta}{ds} = -\frac{qV_{\text{rf}}}{cP_0C_0} \left( \sin(\phi_s) - \sin\left(\phi_s - \frac{\omega_{\text{rf}}z}{c}\right) \right). \quad (5.54)$$

We see that a stable fixed point occurs at:

$$\begin{pmatrix} z \\ \delta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad (5.55)$$

and an unstable fixed point occurs at:

$$\begin{pmatrix} z \\ \delta \end{pmatrix} = \begin{pmatrix} (2\phi_s - \pi)c/\omega_{\text{rf}} \\ 0 \end{pmatrix}. \quad (5.56)$$

Substituting for  $z$  and  $\delta$  from (5.56) into (5.52), we find that the value of the Hamiltonian at the unstable fixed point is:

$$H_{\text{ufp}} = \frac{qV_{\text{rf}}}{\omega_{\text{rf}}P_0C_0} (\cos(\phi_s) + (2\phi_s - \pi)\sin(\phi_s)). \quad (5.57)$$

The value of the Hamiltonian is equal to  $H_{\text{ufp}}$  along the separatrix that bounds the rf bucket. The maximum value of the energy deviation within the rf bucket occurs at  $z = 0$  (when  $d\delta/ds = 0$ ). Let this maximum value of the energy deviation be  $\delta_{\text{rf}}$ . Substituting  $z = 0$  and  $\delta = \delta_{\text{rf}}$  into the Hamiltonian (5.52), we find:

$$H_{\text{ufp}} = -\frac{qV_{\text{rf}}}{\omega_{\text{rf}}P_0C_0} \cos(\phi_s) - \frac{1}{2} \frac{\eta_p}{\beta_0^2} \delta_{\text{rf}}^2. \quad (5.58)$$

Equating (5.57) and (5.58) gives:

$$\delta_{\text{rf}} = \frac{2\nu_z\beta_0^2}{h|\eta_p|} \sqrt{1 + \left(\phi_s - \frac{\pi}{2}\right) \tan(\phi_s)}, \quad (5.59)$$

where  $\nu_z$  is the synchrotron tune, given by (5.49), and  $h$  is the *harmonic number*, defined as the ratio of the rf frequency to the beam revolution frequency:

$$h = \frac{\omega_{\text{rf}}}{\omega_0}. \quad (5.60)$$

The harmonic number gives the total number of rf buckets in longitudinal phase space, and is therefore equal to the maximum number of bunches that can be stored in a synchrotron storage ring. Note that as  $\phi_s \rightarrow \pi/2$ ,  $\delta_{\text{rf}} \rightarrow 0$ : this limit corresponds to the situation where the rf voltage is only just large enough to replace the energy lost by synchrotron radiation. As this limit is approached, the distance between the stable and unstable fixed points in longitudinal phase space becomes infinitesimally small.

Finally, we note that it is possible, from the Hamiltonian, to determine the ratio of the rms bunch length  $\sigma_z$  to rms energy spread  $\sigma_\delta$  in an invariant distribution, i.e. in a distribution that remains unchanged as particles perform synchrotron oscillations. The ratio  $\sigma_z/\sigma_\delta$  is simply the ratio of the width to the height of a contour (within an rf bucket) in longitudinal phase space. Consider a particle that has values of the phase space variables  $z = \sigma_z$  and  $\delta = 0$ . After one quarter of a synchrotron oscillation, the phase space variables will be  $z = 0$  and  $\delta = \sigma_\delta$ . Since the Hamiltonian is a constant of the motion, we can say that for this particle:

$$\begin{aligned} H &= \frac{qV_{\text{rf}}}{\omega_{\text{rf}}P_0C_0} \left( \sin(\phi_s) \frac{\omega_{\text{rf}}\sigma_z}{c} - \cos\left(\phi_s - \frac{\omega_{\text{rf}}\sigma_z}{c}\right) \right), \\ &= -\frac{qV_{\text{rf}}}{\omega_{\text{rf}}P_0C_0} \cos(\phi_s) - \frac{1}{2} \frac{\eta_p}{\beta_0^2} \sigma_\delta^2. \end{aligned} \quad (5.61)$$

If the bunch length is short compared to the rf wavelength, i.e.:

$$\frac{\omega_{\text{rf}}\sigma_z}{c} \ll 1, \quad (5.62)$$

then we find from (5.61) and (5.62) that:

$$\frac{\sigma_z}{\sigma_\delta} = \frac{C_0}{2\pi} \frac{|\eta_p|}{\beta_0^2 \nu_z}. \quad (5.63)$$

The synchrotron motion can be described using action-angle variables in the same way as the betatron motion. The longitudinal action  $J_z$  and angle  $\phi_z$  are defined so that:

$$z = \sqrt{2\beta_z J_z} \cos(\phi_z), \quad (5.64)$$

$$\delta = -\sqrt{\frac{2J_z}{\beta_z}} (\sin(\phi_z) + \alpha_z \cos(\phi_z)), \quad (5.65)$$

where  $\beta_z$  and  $\alpha_z$  are longitudinal Courant-Snyder functions. By making the smooth focusing approximation in writing the equations of motion (5.41) and (5.43), we effectively assume that  $\alpha_z = 0$ , and  $\beta_z$  is constant. This is a good assumption if the longitudinal focusing is weak, i.e. if the synchrotron tune is small. Then, averaging over all particles in a bunch at any point in a storage ring, we can say that:

$$\beta_z = \sqrt{\frac{\langle z^2 \rangle}{\langle \delta^2 \rangle}} = \frac{\sigma_z}{\sigma_\delta} = \frac{C_0}{2\pi} \frac{|\eta_p|}{\beta_0^2 \nu_z}. \quad (5.66)$$

We define the synchrotron frequency  $\omega_z$  as follows:

$$\omega_z = \frac{c}{\beta_0} \frac{d\phi_z}{ds}. \quad (5.67)$$

Note that for a storage ring below transition ( $\eta_p < 0$ ),  $\omega_z > 0$ , and for a storage ring above transition ( $\eta_p > 0$ ),  $\omega_z < 0$ . The synchrotron frequency is related to the synchrotron tune (the number of synchrotron oscillations per orbit of the ring) by:

$$\nu_z = \frac{|\omega_z|}{\omega_0}, \quad (5.68)$$

where the revolution frequency is:

$$\omega_0 = \frac{c}{\beta_0} \frac{2\pi}{C_0}. \quad (5.69)$$

Combining (5.66), (5.68) and (5.69) gives:

$$\beta_z = \frac{c}{\beta_0} \frac{|\eta_p|}{|\omega_z|}. \quad (5.70)$$

Since the phase slip factor and synchrotron frequency always have opposite signs, we can write:

$$\beta_z = -\frac{c}{\beta_0} \frac{\eta_p}{\omega_z}. \quad (5.71)$$

Substituting for  $\omega_z$  from (5.67) gives:

$$\frac{d\phi_z}{ds} = -\frac{\eta_p}{\beta_0^2 \beta_z}. \quad (5.72)$$

It is interesting to compare (5.72) with (4.81), which relates the rate of betatron phase advance to the transverse beta function. In both cases, the rate of the phase advance is inversely proportional to the beta function. However, for the horizontal motion, the beta function is conventionally defined so that:

$$\frac{d\phi_x}{ds} = \frac{1}{\beta_x}, \quad (5.73)$$

and similarly for the vertical motion. But for the longitudinal motion, the phase advance also depends on the phase slip factor: if there is no phase slip, there is no synchrotron phase advance. For ultra-relativistic particles, the dominant contribution to the phase slip comes from the dipoles, and therefore most of the synchrotron phase advance takes place in the dipoles. Also, the direction of the phase advance depends on the sign of the phase slip. In a storage ring below transition, particles move clockwise in longitudinal phase space; above transition, particles move anticlockwise.

Description of the synchrotron motion in terms of longitudinal action-angle variables and Courant–Snyder parameters assumes that the synchrotron motion can be treated independently of the betatron motion. At locations in a storage ring where the dispersion is zero (and the single-turn transfer matrix takes a block-diagonal form), it is possible to treat the betatron and synchrotron motion separately; in many situations, however, there will be coupling between the longitudinal and transverse motion. In the most general case, there will be coupling between the longitudinal, horizontal and the vertical motion. A fully consistent description of the dynamics in an accelerator should take this into account. There are various methods that can be applied for the analysis of coupled motion in three degrees of freedom: we discuss one of them in Section 5.2.

## 5.2 Fully Coupled Motion

In the preceding sections, we discussed specific aspects of the coupling between the transverse and longitudinal motion of a particle in an accelerator beam line. In particular, we considered the phenomena of dispersion, momentum compaction and phase slip. These are convenient concepts in the design, analysis and operation of accelerators; however, they give special

status to the energy deviation, which is treated as a fixed (or slowly varying) parameter on which the transverse trajectory, path length and time of flight all depend on. For the description and analysis of coupling in more general cases (including coupling between the horizontal and vertical degrees of freedom), more powerful techniques are needed. There are various methods available for describing coupled optics, and the one that is chosen is often ultimately a question of convenience for a specific application. One method that is widely used is that of Edwards and Teng [Edwards and Teng (1973)], which is based on constructing a transformation that puts the transfer matrix into block-diagonal form; optical functions analogous to the Courant–Snyder parameters are then obtained in the usual way from each block diagonal. This technique does have some advantages, particularly in practical application for coupling measurements [Sagan and Rubin (1999); Sagan *et al.* (2000)]. However, it also has some drawbacks. For example, the transformation required to put the transfer matrix into block-diagonal form is not uniquely defined, which means that the coupled Courant–Snyder parameters are also not uniquely defined. Also, the technique is most readily applied to matrices describing the motion in two degrees of freedom, and does not generalise easily to the full three degrees of freedom. Here, we shall discuss an alternative approach in which we construct a transformation that puts the transfer matrix into the form of a pure rotation; the generalised Courant–Snyder parameters are then obtained from the matrix representing the transformation, rather than from the transformed transfer matrix. This method has the advantage of providing a relatively direct and straightforward generalisation of the concepts of uncoupled motion to an arbitrary number of degrees of freedom [Wolski (2006)]. For example, the relationships (4.97), (4.99) and (4.100) between the second-order moments of the beam distribution and the Courant–Snyder parameters generalise to  $n$  degrees of freedom as follows:

$$\Sigma_{ij} = \sum_k \beta_{ij}^k \epsilon_k, \quad (5.74)$$

where  $\Sigma_{ij} = \langle x_i x_j \rangle$  are elements of a  $2n \times 2n$  matrix  $\Sigma$  consisting of the second-order moments of the beam distribution,  $\beta_{ij}^k$  are elements of a set of  $2n \times 2n$  matrices  $B^k$ , and  $\epsilon_k$  are a set of invariant emittances given as the averages over the action variables of all particles in the bunch. The summation in (5.74) ranges over all values of the index  $k$ , which is associated with the degrees of freedom in the system: in three degrees of freedom, we write  $k = \text{I}, \text{II}$  or  $\text{III}$ . The elements of the matrices  $B^k$  are the generalisation of the Courant–Snyder parameters: in one degree of freedom,  $B = A^{-1}$ ,

where the matrix  $A^{-1}$  is given by (4.28). Given  $B^k(s_0)$  at some point in a beam line  $s = s_0$ , the matrices  $B^k(s_1)$  at some other point  $s = s_1$  can be obtained using the transfer matrix  $R$  from  $s_0$  to  $s_1$ :

$$B^k(s_1) = RB^k(s_0)R^T. \quad (5.75)$$

This approach to coupled optics emphasises the significance of the Courant–Snyder parameters for describing the variation of the beam distribution along a beam line. In addition, it provides an interesting insight into the beam optics, namely that a description of the optics in terms of the Courant–Snyder parameters and phase advances is essentially a description in terms of the eigenvectors and eigenvalues of the transfer matrices.

We begin the development of the theory, as we did for uncoupled motion, by considering a periodic beam line. Let  $R$  be the transfer matrix for one period of this beam line. We can choose the reference trajectory so that it is the physical trajectory of a particle starting with phase space vector  $(x, p_x, y, p_y, z, \delta) = 0$ ; in particular, we choose the curvature  $h$  of the reference trajectory in a dipole magnet so that:

$$h = \frac{q}{P_0} B, \quad (5.76)$$

where  $q$  is the particle charge,  $B$  is the dipole field, and  $P_0$  is the reference momentum. In this case, the transfer map has no zeroth-order terms and (in the linear approximation) is completely described by the transfer matrix  $R$ . Since we are using canonical variables, the transfer matrix is symplectic, i.e. it satisfies:

$$R^T S R = S, \quad (5.77)$$

where  $S$  is the antisymmetric matrix (4.14). Symplectic matrices with distinct eigenvalues have two important mathematical properties on which the results developed in this section are based. First, the eigenvalues occur in reciprocal pairs, so that:

$$\lambda_{-k} \lambda_k = 1, \quad (5.78)$$

where  $\lambda_{\pm k}$  are the eigenvalues of the symplectic matrix. Second, subject to certain conditions, the eigenvectors can be normalised and arranged into a matrix  $E$  such that:

$$E^T S E = iS. \quad (5.79)$$

We can prove the properties (5.78) and (5.79) as follows. Let  $\vec{e}_m$  be the eigenvector of the symplectic matrix  $R$  with eigenvalue  $\lambda_m$ :

$$R\vec{e}_m = \lambda_m \vec{e}_m. \quad (5.80)$$

It follows that:

$$\vec{e}_\ell^T R^T S R \vec{e}_m = \lambda_\ell \lambda_m \vec{e}_\ell^T S \vec{e}_m. \quad (5.81)$$

Applying the symplectic condition (5.77), we find:

$$\vec{e}_\ell^T S \vec{e}_m = \lambda_\ell \lambda_m \vec{e}_\ell^T S \vec{e}_m, \quad (5.82)$$

which means that if:

$$\vec{e}_\ell^T S \vec{e}_m \neq 0, \quad (5.83)$$

then:

$$\lambda_\ell \lambda_m = 1. \quad (5.84)$$

From (5.84) we can sort the eigenvalues into reciprocal pairs, as stated in (5.78); if the eigenvalues are distinct, it then follows that for  $\ell \neq m$ :

$$\vec{e}_{-\ell}^T S \vec{e}_m = 0, \quad (5.85)$$

where, in  $n$  degrees of freedom,  $\ell, m = \pm 1, \pm 2, \dots, \pm n$ . If, for all  $m$ :

$$\vec{e}_{-m}^T S \vec{e}_m \neq 0, \quad (5.86)$$

then the eigenvectors  $\vec{e}_m$  can be normalised so that:

$$\vec{e}_{-m}^T S \vec{e}_m = i, \quad (5.87)$$

and the result (5.79) then follows. Not every symplectic matrix has eigenvalues that are distinct and eigenvectors that satisfy the condition (5.86); however, in the case of a transfer matrix  $R$  for a linearly stable, periodic beam line not on a resonance (i.e. with distinct, non-integer tunes), we can assume that the conditions required for (5.78) and (5.79) are satisfied.

Since  $E$  is constructed from the eigenvectors of  $R$  with appropriate ordering, it follows that  $E^{-1}RE$  is a diagonal matrix, with elements along the diagonal given by the eigenvalues  $\lambda_{\pm k} = e^{\pm i\mu_k}$ . Now let us define the matrix  $Q$  with block diagonals  $Q_2$  given by:

$$Q_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}. \quad (5.88)$$

In  $n$  degrees of freedom,  $Q$  will be a  $2n \times 2n$  block-diagonal matrix. Using  $Q$ , we can construct the *normalising matrix*  $N$ :

$$N = QE^{-1}, \quad (5.89)$$

where  $E$  is the matrix of eigenvectors of  $R$  sorted and normalised to satisfy (5.79). Using the fact that  $E^{-1}RE$  is a diagonal matrix with elements

$e^{\pm i\mu_k}$ , it is straightforward to show (by constructing  $QE^{-1}REQ^{-1}$ ) that  $N$  has the property:

$$NRN^{-1} = \bar{R}(\mu_k), \quad (5.90)$$

where  $\bar{R}(\mu_k)$  is a block-diagonal rotation matrix with rotation angles  $\mu_k$ . Explicitly, in three degrees of freedom:

$$\bar{R}(\mu_k) = \begin{pmatrix} \cos(\mu_I) & \sin(\mu_I) & 0 & 0 & 0 & 0 \\ -\sin(\mu_I) \cos(\mu_I) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos(\mu_{II}) & \sin(\mu_{II}) & 0 & 0 \\ 0 & 0 & -\sin(\mu_{II}) \cos(\mu_{II}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos(\mu_{III}) & \sin(\mu_{III}) \\ 0 & 0 & 0 & 0 & -\sin(\mu_{III}) \cos(\mu_{III}) & 0 \end{pmatrix}. \quad (5.91)$$

We can show that  $N$  is symplectic. Since  $N = QE^{-1}$ , we can write:

$$N^T S N = (E^{-1})^T Q^T S Q E^{-1} = i(E^{-1})^T S E^{-1}, \quad (5.92)$$

where the final step follows from (5.88). Then, using (5.79), we find:

$$N^T S N = S, \quad (5.93)$$

which shows that  $N$  satisfies the symplectic condition. We use the matrix  $N$  to define a vector  $\vec{J}$  of new dynamical variables  $(X_k, P_k)$ . In  $n$  degrees of freedom:

$$\vec{J} = \begin{pmatrix} X_1 \\ P_1 \\ \vdots \\ X_n \\ P_n \end{pmatrix} = N \vec{x}. \quad (5.94)$$

Since the variables  $(X_k, P_k)$  are derived from the original canonical variables by a symplectic transformation, the new variables are also canonical. Define a mixed-variable generating function of the first kind, as follows:

$$F_1(X_k, \phi_k) = -\frac{1}{2} \sum_k X_k^2 \tan(\phi_k), \quad (5.95)$$

where the summation extends over all the degrees of freedom. From the standard formulae:

$$P_k = \frac{\partial F}{\partial X_k}, \quad (5.96)$$

$$J_k = -\frac{\partial F}{\partial \phi_k}, \quad (5.97)$$

we derive the following relationships between the old variables  $(X_k, P_k)$  and the new variables  $(\phi_k, J_k)$ :

$$J_k = \frac{1}{2} (X_k^2 + P_k^2), \quad (5.98)$$

$$\tan(\phi_k) = -\frac{P_k}{X_k}. \quad (5.99)$$

Note that  $J_k$  and  $\phi_k$  are the action-angle variables for motion in an arbitrary number of degrees of freedom. By constructing the variables in this way, we ensure that the motion described by the action-angle variables for one value of  $k$  is independent of the values of the variables for the other values of  $k$ . That is, the motion in any one of the degrees of freedom is uncoupled from the motion in the other degrees of freedom. This is the case even when there is arbitrary coupling between the different degrees of freedom described using original variables (the components of the phase space vector  $\vec{x}$ ). The variables  $(X_k, P_k)$  defined by (5.94) and the action-angle variables  $(\phi_k, J_k)$  defined by (5.98) and (5.99) describe the *normal modes* of the system.

In three degrees of freedom, the components of the phase space vector  $\vec{J}$  can be written:

$$\vec{J} = \begin{pmatrix} \sqrt{2J_I} \cos(\phi_I) \\ -\sqrt{2J_I} \sin(\phi_I) \\ \sqrt{2J_{II}} \cos(\phi_{II}) \\ -\sqrt{2J_{II}} \sin(\phi_{II}) \\ \sqrt{2J_{III}} \cos(\phi_{III}) \\ -\sqrt{2J_{III}} \sin(\phi_{III}) \end{pmatrix}. \quad (5.100)$$

The transformation of  $\vec{J}$  under transport over one periodic cell of the beam line can be found as follows. First note that:

$$\begin{aligned} \vec{J}(s_1) &= N\vec{x}(s_1) \\ &= NR\vec{x}(s_0) \\ &= NRN^{-1}N\vec{x}(s_0) \\ &= NRN^{-1}\vec{J}(s_0). \end{aligned} \quad (5.101)$$

Then, using (5.90), we see that:

$$\vec{J}(s_1) = \bar{R}(\mu_k)\vec{J}(s_0). \quad (5.102)$$

In other words,  $\vec{J}$  transforms from one point in the beam line to another according to a pure rotation. Since  $\bar{R}(\mu_k)$  is block diagonal, and the elements of  $\vec{J}$  are of the form shown in (5.100), the rotation is easily applied. It is found that the action variables  $J_k$  are conserved under  $\vec{x} \mapsto R\vec{x}$ :

$$J_k \mapsto J_k, \quad (5.103)$$

and the angle variables increase by  $\mu_k$ :

$$\phi_k \mapsto \phi_k + \mu_k. \quad (5.104)$$

The rotation angles  $\mu_k$  give the phase advance in each degree of freedom across one periodic cell of the beam line.

Continuing the discussion of periodic beam lines, the next step is to introduce the concept of a *matched distribution*. A matched distribution at any point in a periodic beam line is a phase space distribution of particles that is unchanged after the bunch is transported along one periodic section of the beam line. For the present purposes, we only need to consider the second-order moments of the distribution: we do not need to specify whether the distribution is uniform, parabolic, Gaussian or some more exotic function. For convenience, we define the  $2n \times 2n$  matrix  $\Sigma$  (in  $n$  degrees of freedom) with elements  $\Sigma_{ij}$  defined by:

$$\Sigma_{ij} = \langle x_i x_j \rangle, \quad (5.105)$$

and  $x_i$  are the elements of the phase space vector (in Cartesian co-ordinates)  $\vec{x}$ . In matrix form, (5.105) may be written as the average over the outer product of the phase space vectors for each particle in the bunch:

$$\Sigma = \langle \vec{x} \vec{x}^T \rangle. \quad (5.106)$$

By writing  $\Sigma$  in this way, it is clear that under transport through one periodic cell, for which  $\vec{x} \mapsto R\vec{x}$ ,  $\Sigma$  transforms as:

$$\Sigma \mapsto R\Sigma R^T. \quad (5.107)$$

A matched distribution satisfies the condition:

$$R\Sigma R^T = \Sigma, \quad (5.108)$$

in other words, the second-order moments of the beam distribution are unchanged after linear transport through one periodic cell of the beam line.

In order to proceed, it is useful to express  $\Sigma$  in terms of the action-angle variables. Using the definition of the phase space vector (in action-angle co-ordinates)  $\vec{J}$  (5.94), we can write:

$$\Sigma = \langle \vec{x} \vec{x}^T \rangle = N^{-1} \langle \vec{J} \vec{J}^T \rangle (N^{-1})^T. \quad (5.109)$$

Let us assume that the angle variables of particles in the bunch are uncorrelated, in the sense that:

$$\langle \sin(\phi_i) \sin(\phi_j) \rangle = \langle \cos(\phi_i) \cos(\phi_j) \rangle = \frac{1}{2} \delta_{ij}, \quad (5.110)$$

$$\langle \sin(\phi_i) \cos(\phi_j) \rangle = 0. \quad (5.111)$$

Then, using the explicit expression for  $\vec{J}$  (5.100) in  $n$  degrees of freedom, we find that:

$$\Sigma = \sum_k N^{-1} T^k (N^{-1})^T \epsilon_k. \quad (5.112)$$

In three degrees of freedom, the matrices  $T^k$  are  $6 \times 6$  matrices defined by:

$$T^I = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad T^{II} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad T^{III} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (5.113)$$

with an obvious generalisation to any number of degrees of freedom. The *emittance* of each degree of freedom is defined by:

$$\epsilon_k = \langle J_k \rangle. \quad (5.114)$$

In  $n$  degrees of freedom, we define a set of  $n$  matrices  $B^k$  (with each matrix having  $2n \times 2n$  elements) such that:

$$T^k = NB^kN^T. \quad (5.115)$$

Then, (5.112) may be written:

$$\Sigma = \sum_k B^k \epsilon_k. \quad (5.116)$$

Now, we observe that the matrices  $B^k$  are determined by the transfer matrix over one periodic cell of the beam line. Therefore, to construct a matrix  $\Sigma$  at a given point in a periodic beam line using (5.116) we only have  $n$  parameters, corresponding to the emittances in  $n$  degrees of freedom. This is much fewer than the number of independent elements of a symmetric  $2n \times 2n$  matrix. This means that the distributions represented by (5.116) cannot be a completely general distribution: in fact, a distribution constructed using this equation must be a *matched distribution*, as we shall now show. Using (5.112) we find:

$$R\Sigma R^T = \sum_k RN^{-1} T^k (N^{-1})^T R^T \epsilon_k. \quad (5.117)$$

But from (5.90), we have:

$$RN^{-1} = N^{-1}\bar{R}, \quad (5.118)$$

where  $\bar{R}$  is a pure rotation. Hence:

$$R\Sigma R^T = \sum_k N^{-1} \bar{R} T^k \bar{R}^T (N^{-1})^T \epsilon_k. \quad (5.119)$$

But since  $\bar{R}$  is just a rotation matrix, it satisfies:

$$\bar{R}^T = \bar{R}^{-1}, \quad (5.120)$$

and so, using the explicit form for the matrices  $T^k$ :

$$\bar{R} T^k \bar{R}^T = T^k. \quad (5.121)$$

Then, using this result in (5.119):

$$R\Sigma R^T = \sum_k N^{-1} T^k (N^{-1})^T \epsilon_k = \Sigma. \quad (5.122)$$

Therefore, a matrix  $\Sigma$  constructed using (5.116) is unchanged under transport through one periodic section of the beam line. In other words, such a matrix represents a matched distribution. Note that this is true for any value of the emittances  $\epsilon_k$ : there are infinitely many matched distributions for a given beam line, although the number of degrees of freedom in choosing a matched distribution is only equal to the number of degrees of freedom in the particle motion.

We have now shown that, in a periodic beam line, a matched distribution at a given point can be expressed as (5.116) in terms of a set of matrices  $B^k$  that are determined by the transfer matrix over a single periodic cell, and a set of emittances  $\epsilon_k$  that can (mathematically) take any values. The emittances represent the average of the action variables over all particles in the bunch. The action for a particle gives the amplitude of the oscillation of a particle as it moves down a beam line, and is a positive number: this implies that the emittances must be positive in order to produce a physical distribution. Equation (5.115) provides a recipe for constructing the matrices  $B^k$  from the transfer matrix for one periodic section, via (5.89). The final step is to find how the matrices  $B^k$  evolve along the beam line. To do this, we first show that the eigenvalues of  $\Sigma S$  are  $\pm i\epsilon_k$ , where  $S$  is the antisymmetric matrix (4.14) and  $\epsilon_k$  are the emittances.

To find the relationship between the eigenvalues of  $\Sigma S$  and the beam emittances, we proceed as follows. From (5.116), it follows that:

$$\Sigma S = \sum_k B^k S \epsilon_k. \quad (5.123)$$

Using the definition of the matrices  $B^k$  (5.115), we can write:

$$B^k S = N^{-1} T^k (N^{-1})^T S = N^{-1} T^k S N, \quad (5.124)$$

where the final step follows from the fact that  $N$  is symplectic. Substituting from the definition of  $N$  (5.89), we find that:

$$B^k S = EQ^{-1}T^k SQE^{-1}, \quad (5.125)$$

from which it follows that:

$$E^{-1}B^k SE = Q^{-1}T^k SQ. \quad (5.126)$$

The matrices on the right-hand side of (5.126) are all constant. Evaluating these matrices, we find:

$$E^{-1}B^k SE = -i\bar{T}^k, \quad (5.127)$$

where, in three degrees of freedom:

$$\bar{T}^I = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{T}^{II} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \bar{T}^{III} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \quad (5.128)$$

Finally, from (5.123) and (5.127), we have (again, in three degrees of freedom):

$$E^{-1}\Sigma SE = -i \begin{pmatrix} \epsilon_I & 0 & 0 & 0 & 0 & 0 \\ 0 & -\epsilon_I & 0 & 0 & 0 & 0 \\ 0 & 0 & \epsilon_{II} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\epsilon_{II} & 0 & 0 \\ 0 & 0 & 0 & 0 & \epsilon_{III} & 0 \\ 0 & 0 & 0 & 0 & 0 & -\epsilon_{III} \end{pmatrix}. \quad (5.129)$$

Since  $E^{-1}\Sigma SE$  is diagonal, the matrix  $E$  must be constructed from the eigenvectors of  $\Sigma S$ . Recall that originally we constructed  $E$  from the eigenvectors of the transfer matrix  $R$  for one periodic section of the beam line: the eigenvectors of  $R$  are the same as the eigenvectors of  $\Sigma S$ , where the second-order beam moment matrix  $\Sigma$  is invariant under transport through one periodic section,  $\Sigma = R\Sigma R^T$ . The elements of the diagonalisation of  $\Sigma S$  are the eigenvalues of  $\Sigma S$ . It follows that:

$$\text{eigenvalues}(\Sigma S) = \pm i\epsilon_k, \quad (5.130)$$

where  $\epsilon_k$  are the emittances.

Now let  $R(s_1, s_0)$  be the transfer matrix from  $s = s_0$  to  $s = s_1$ . This is not necessarily the transfer matrix for one periodic section:  $s_0$  and  $s_1$  can be chosen arbitrarily. However, we assume that the matrices  $B^k$  and the matched distribution  $\Sigma$  at  $s_0$  are known from the periodic constraint. Under transport from  $s_0$  to  $s_1$  the Cartesian phase space vector transforms as:

$$\vec{x}(s_1) = R(s_1, s_0)\vec{x}(s_0). \quad (5.131)$$

It follows that the beam distribution matrix transforms as:

$$\Sigma(s_1) = R(s_1, s_0)\Sigma(s_0)R(s_1, s_0)^T. \quad (5.132)$$

Hence, we have:

$$\Sigma(s_1)S = R(s_1, s_0)\Sigma(s_0)R(s_1, s_0)^TS = R(s_1, s_0)\Sigma(s_0)SR(s_1, s_0)^{-1}, \quad (5.133)$$

where the last step follows from the fact that the transfer matrix  $R$  is symplectic. But for any matrices  $U$  and  $V$ , the eigenvalues of  $V$  are the same as the eigenvalues of  $UVU^{-1}$ . Hence, the eigenvalues of  $\Sigma(s_1)S$  are the same as the eigenvalues of  $\Sigma(s_0)S$ . The eigenvalues of  $\Sigma(s_0)S$  are  $\pm i\epsilon_k$ , where  $\epsilon_k$  are the emittances; hence, the emittances of a bunch are conserved as a bunch is transported along a beam line, as long as the transport is linear and symplectic.

Since the bunch emittances are conserved, and the matrix  $\Sigma$  is related to matrices  $B^k$  through (5.116), the matrices  $B^k$  must transform the same way as  $\Sigma$  under transport along the beam line. In other words:

$$B^k(s_1) = R(s_1, s_0)B^k(s_0)R(s_1, s_0)^T. \quad (5.134)$$

The situation is exactly analogous to that for motion in one degree of freedom. For a non-periodic beam line in one degree of freedom, we can determine appropriate values for the Courant–Snyder parameters (and emittance) from the second-order moments of the beam distribution specified either by design or measurement. The normalising matrix  $N$  can be expressed in terms of the Courant–Snyder parameters in the form (4.40). The Courant–Snyder parameters can be propagated along the beam line using the transfer matrices, as expressed by (4.29). The transfer matrix between any two points in a beam line can be expressed in terms of the Courant–Snyder parameters and the phase advance in the form (4.60). All these results generalise to the case of many degrees of freedom: from the second-order moments of the beam distribution  $\Sigma$  at some point in the beam line, we can determine the matrices  $B^k$  (the elements of which generalise the

Courant–Snyder parameters), the bunch emittances  $\epsilon_k$ , and the normalising transformation  $N$  (for which we impose the condition  $N_{12} = 0$ ). The emittances are conserved under symplectic linear transport, and the matrices  $B^k$  can be evolved along the beam line using the transfer matrices. By generalising (5.101) to any two points along a beam line (which does not have to be periodic), we see that the generalisation of the expression (4.60) for the transfer matrix in terms of the Courant–Snyder parameters and the phase advance is:

$$R(s_1, s_0) = N(s_1)^{-1} \bar{R}(\mu_k) N(s_0), \quad (5.135)$$

where  $\bar{R}(\mu_k)$  is a rotation matrix through angles given by the phase advances  $\mu_k$ .

In one degree of freedom, the Courant–Snyder parameters can be used to express the action variable in terms of the co-ordinate and canonical momentum:

$$2J_x = \gamma_x x^2 + 2\alpha_x x p_x + \beta_x p_x^2. \quad (5.136)$$

This can be written:

$$2J_x = (x \ p_x) S^T \begin{pmatrix} \beta_x & -\alpha_x \\ -\alpha_x & \gamma_x \end{pmatrix} S \begin{pmatrix} x \\ p_x \end{pmatrix}. \quad (5.137)$$

Equation (5.137) generalises to coupled motion in an arbitrary number of degrees of freedom, with the Courant–Snyder parameters given by elements of the matrices  $B^k$ :

$$2J_k = \vec{x}^T S^T B^k S \vec{x}, \quad (5.138)$$

where  $\vec{x}$  is the phase space vector in Cartesian variables. Equation (5.138) is readily shown, as follows. First, we substitute for  $B^k$  from (5.115):

$$\vec{x}^T S^T B^k S \vec{x} = \vec{x}^T S^T N^{-1} T^k (N^{-1})^T S \vec{x}. \quad (5.139)$$

Since the normalising matrix  $N$  is symplectic, its inverse is also symplectic:

$$(N^{-1})^T S N^{-1} = S. \quad (5.140)$$

It then follows that:

$$\vec{x}^T S^T B^k S \vec{x} = \vec{x}^T N^T S^T T^k S N \vec{x}. \quad (5.141)$$

Finally, using (5.94), (5.100) and the definitions of the matrices  $T^k$ :

$$\vec{x}^T S^T B^k S \vec{x} = \vec{J}^T S^T T^k S \vec{J} = \vec{J}^T T^k \vec{J} = 2J_k. \quad (5.142)$$

The fact that the actions  $J_k$  are invariant under linear symplectic transformations can be obtained from (5.138) using the transformation of the (Cartesian) phase space vector (5.131) and the transformation of the Courant–Snyder parameter matrix  $B^k$  given in (5.134).

### 5.3 Dispersion Revisited

In Section 5.1.1 we defined the dispersion as the change in the horizontal (or vertical) trajectory of a particle with respect to a change in the momentum of the particle. Dispersion is therefore a specific case of coupling between the longitudinal variables (in particular, the energy deviation) and the transverse variables. In Section 5.2 we developed a formalism for describing optics in beam lines where there are general coupling effects. Since dispersion is a specific case of coupling, we expect to be able to express the dispersion function within the formalism of Section 5.2. In this section, we shall show how to do this.

Suppose that we know the lattice functions in a beam line. The beam line may be periodic, in which case the lattice functions are determined by the periodicity condition, or it may be non-periodic, in which case the lattice functions are determined from some initial conditions at a given point in the beam line. If we also know the emittances of a bunch of particles, then we can write for the mean square horizontal co-ordinate:

$$\langle x^2 \rangle = \beta_{11}^I \epsilon_I + \beta_{11}^{II} \epsilon_{II} + \beta_{11}^{III} \epsilon_{III}. \quad (5.143)$$

This equation is exact. However, by making some assumptions, we can derive an alternative expression for  $\langle x^2 \rangle$  in terms of the uncoupled Courant–Snyder parameters and the dispersion.

First, we assume that there is no coupling between the horizontal and vertical motion. In that case,  $\beta_{11}^{II} = 0$ . Also, if the coupling between the horizontal and longitudinal planes is weak, then we can make the approximation:

$$\beta_{11}^I \approx \beta_x, \quad (5.144)$$

where  $\beta_x$  is the horizontal beta function that we would calculate in the beam line considering only the horizontal motion. A particle with zero energy deviation follows a horizontal trajectory given by:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (5.145)$$

where  $J_x$  is the horizontal action (a constant of the motion) and  $\phi_x$  is the horizontal angle variable. If the particle has non-zero energy deviation, then, by definition of the dispersion  $\eta_x$ , the position of the particle at any point changes by  $\eta_x \delta_p$ , where  $\delta_p$  is the momentum deviation (5.9). We shall consider the ultra-relativistic case,  $\beta_0 \rightarrow 1$ , so that the momentum deviation and energy deviation are approximately equal,  $\delta_p \approx \delta$ . Then,

taking into account the change in trajectory resulting from dispersion, a more general expression than (5.145) for the horizontal trajectory is:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x) + \eta_x \delta. \quad (5.146)$$

Now we assume that the angle variable  $\phi_x$  of all particles in the beam are uncorrelated, and that the energy deviation  $\delta$  is uncorrelated with the horizontal variables  $J_x$  or  $\phi_x$ . Then, from (5.146), we find that:

$$\langle x^2 \rangle = \beta_x \epsilon_x + \eta_x^2 \langle \delta^2 \rangle, \quad (5.147)$$

where  $\epsilon_x = \langle J_x \rangle$  is the horizontal emittance. For beam distributions that are Gaussian and that also have  $\langle x \rangle = \langle \delta \rangle = 0$ , (5.147) is often written in the form:

$$\sigma_x = \sqrt{\beta_x \epsilon_x + \eta_x^2 \sigma_\delta^2}, \quad (5.148)$$

where  $\sigma_x$  and  $\sigma_\delta$  are the horizontal rms beam size and the rms energy spread (i.e. the standard deviations of the horizontal co-ordinate and the energy deviation) respectively.

Using the assumption that the energy deviation is uncorrelated with the horizontal action or the horizontal angle variable, we also find from (5.146):

$$\langle x\delta \rangle = \eta_x \langle \delta^2 \rangle. \quad (5.149)$$

But we can also write, assuming that there is no coupling between the vertical motion and either the horizontal or longitudinal motion:

$$\langle x\delta \rangle = \beta_{16}^I \epsilon_I + \beta_{16}^{III} \epsilon_{III}, \quad (5.150)$$

and:

$$\langle \delta^2 \rangle = \beta_{66}^I \epsilon_I + \beta_{66}^{III} \epsilon_{III}. \quad (5.151)$$

It is usually the case (except for beam lines that include rather special accelerator components) that, although the energy deviation can have a significant impact on the horizontal trajectory of a particle, the horizontal trajectory has little effect on the energy deviation. Then, we can make the approximations that  $\beta_{16}^I \approx 0$  and  $\beta_{66}^I \approx 0$ . Finally, combining (5.149), (5.150) and (5.151), we find:

$$\eta_x \approx \frac{\beta_{16}^{III}}{\beta_{66}^{III}}. \quad (5.152)$$

Equation (5.152) provides a convenient way to calculate the horizontal dispersion, if the coupled beta functions are known. By similar arguments, we can show that:

$$\eta_{px} \approx \frac{\beta_{26}^{III}}{\beta_{66}^{III}}. \quad (5.153)$$

Also, we can find the vertical dispersion using:

$$\eta_y \approx \frac{\beta_{36}^{\text{III}}}{\beta_{66}^{\text{III}}}, \quad (5.154)$$

and:

$$\eta_{py} \approx \frac{\beta_{46}^{\text{III}}}{\beta_{66}^{\text{III}}}. \quad (5.155)$$

It should be remembered that equations (5.152)–(5.155) are approximations for the dispersion, although in most cases they are good approximations.

## 5.4 Examples of Coupled Optics

### 5.4.1 Uniform solenoid field

A uniform solenoid field provides constant transverse focusing. Considering only the transverse motion, we can apply the techniques developed in Section 5.2 to find the matched distribution in a solenoid, i.e. the transverse distribution that remains invariant as the beam is transported along the solenoid. Since the solenoid provides no longitudinal focusing, it is not possible to find an invariant distribution in three degrees of freedom. However, the transverse case is still of interest (for example, when transporting a continuous beam).

The transfer matrix for a solenoid is given by (3.124). Considering only the transverse degrees of freedom, the transfer matrix is written:

$$R_{\text{solenoid}} = \begin{pmatrix} \cos^2(\omega L) & \frac{1}{2\omega} \sin(2\omega L) & \frac{1}{2} \sin(2\omega L) & \frac{1}{\omega} \sin^2(\omega L) \\ -\frac{\omega}{2} \sin(2\omega L) & \cos^2(\omega L) & -\omega \sin^2(\omega L) & \frac{1}{2} \sin(2\omega L) \\ -\frac{1}{2} \sin(2\omega L) & -\frac{1}{\omega} \sin^2(\omega L) & \cos^2(\omega L) & \frac{1}{2\omega} \sin(2\omega L) \\ \omega \sin^2(\omega L) & -\frac{1}{2} \sin(2\omega L) & -\frac{\omega}{2} \sin(2\omega L) & \cos^2(\omega L) \end{pmatrix}. \quad (5.156)$$

Working through the analysis leading to (5.115) for a beam line consisting of a single solenoid of any given length and field, we find for the matched lattice functions:

$$B^I = \begin{pmatrix} \frac{1}{2\omega} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{\omega}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2\omega} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{\omega}{2} \end{pmatrix}, \quad (5.157)$$

and:

$$B^{II} = \begin{pmatrix} \frac{1}{2\omega} & 0 & 0 & -\frac{1}{2} \\ 0 & \frac{\omega}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2\omega} & 0 \\ -\frac{1}{2} & 0 & 0 & \frac{\omega}{2} \end{pmatrix}, \quad (5.158)$$

where:

$$\omega = \frac{q}{P_0} \frac{B_0}{2}. \quad (5.159)$$

As usual,  $q$  is the particle charge,  $P_0$  is the reference momentum (which must be equal to the particle momentum in this case, since we are neglecting the longitudinal variables), and  $B_0$  is the solenoid field. Since  $B^I$  and  $B^{II}$  are independent of the solenoid length  $L$ , the distribution is matched (and remains invariant) along the entire length of the solenoid.

A matched distribution can be constructed using (5.116):

$$\Sigma = B^I \epsilon_I + B^{II} \epsilon_{II}. \quad (5.160)$$

In this case, we find:

$$\Sigma = \begin{pmatrix} \frac{1}{2\omega}(\epsilon_I + \epsilon_{II}) & 0 & 0 & \frac{1}{2}(\epsilon_I - \epsilon_{II}) \\ 0 & \frac{\omega}{2}(\epsilon_I + \epsilon_{II}) & -\frac{1}{2}(\epsilon_I - \epsilon_{II}) & 0 \\ 0 & -\frac{1}{2}(\epsilon_I - \epsilon_{II}) & \frac{1}{2\omega}(\epsilon_I + \epsilon_{II}) & 0 \\ \frac{1}{2}(\epsilon_I - \epsilon_{II}) & 0 & 0 & \frac{\omega}{2}(\epsilon_I + \epsilon_{II}) \end{pmatrix}. \quad (5.161)$$

As expected, the eigenvalues of  $\Sigma S$  are  $\pm i\epsilon_I$  and  $\pm i\epsilon_{II}$ . Any distribution not of the form (5.161) (for  $\omega$  defined by the solenoid field, and  $\epsilon_I$  and  $\epsilon_{II}$  taking any desired values) will change as it moves along the solenoid.

#### 5.4.2 Flat-beam electron source

As a second example of coupled optics, we shall consider a system capable of producing a ‘flat’ electron beam; that is, a beam in which the ratio of the horizontal emittance to the vertical emittance is much larger than unity [Brinkmann *et al.* (1999); Edwards *et al.* (2000)]. The source itself consists of a photocathode that produces a bunch of electrons when hit by a laser pulse. Commonly, such a system has (approximate) rotational symmetry around an axis perpendicular to the photocathode, with the result that the horizontal and vertical emittances of the beam produced from the source are approximately equal. However, an asymmetry between the emittances of the electron beam can be achieved by immersing the photocathode in a solenoid field. The electron beam emerging from the solenoid is coupled, in the sense that there are correlations between the horizontal and vertical dynamical variables. As we shall show, the correlation can be removed (thus producing an uncoupled beam) using a set of three skew quadrupoles with appropriate strengths and spacings [Derbenev (1998)].

In the absence of a solenoid field, and assuming rotational symmetry of the photocathode and the laser spot, a bunch of electrons produced

by a photocathode will have equal horizontal and vertical emittances, and no correlations between variables in different planes. The bunch from the photocathode is accelerated towards an anode by the electric field between the photocathode and the anode. If the acceleration is sufficiently rapid, and the distance between the photocathode and the anode is small, then the transverse part of the beam correlation matrix as the bunch passes through the anode can be written:

$$\Sigma = \begin{pmatrix} \beta\epsilon & 0 & 0 & 0 \\ 0 & \epsilon/\beta & 0 & 0 \\ 0 & 0 & \beta\epsilon & 0 \\ 0 & 0 & 0 & \epsilon/\beta \end{pmatrix}, \quad (5.162)$$

where the emittance  $\epsilon$  and beta function  $\beta$  are dependent on the geometry and other properties of the photocathode and laser.

Now suppose we insert the photocathode inside a solenoid field, with the magnetic field perpendicular to the surface of the photocathode. We can assume the mechanical part of the momentum of a particle will be unchanged by the magnetic field; but the canonical momentum will pick up a contribution from the vector potential. If the magnetic field strength is  $B_0$ , then the vector potential can be written (3.114):

$$\mathbf{A} = \left( -\frac{B_0}{2}y, \frac{B_0}{2}x, 0 \right). \quad (5.163)$$

In terms of action-angle variables, the phase space co-ordinates of a particle after acceleration to the anode will be:

$$x = \sqrt{2\beta J_x} \cos(\phi_x), \quad (5.164)$$

$$p_x = -\sqrt{\frac{2J_x}{\beta}} \sin(\phi_x) - \frac{qB_0}{2P_0} \sqrt{2\beta J_y} \cos(\phi_y), \quad (5.165)$$

$$y = \sqrt{2\beta J_y} \cos(\phi_y), \quad (5.166)$$

$$p_y = -\sqrt{\frac{2J_y}{\beta}} \sin(\phi_y) + \frac{qB_0}{2P_0} \sqrt{2\beta J_x} \cos(\phi_x), \quad (5.167)$$

where  $P_0$  is the reference momentum at the anode. By averaging over products of the phase space variables, and using  $\langle J_x \rangle = \langle J_y \rangle = \epsilon$ , we find that the beam distribution matrix is now:

$$\Sigma = \begin{pmatrix} \beta\epsilon & 0 & 0 & \psi\epsilon \\ 0 & (1 + \psi^2)\epsilon/\beta & -\psi\epsilon & 0 \\ 0 & -\psi\epsilon & \beta\epsilon & 0 \\ \psi\epsilon & 0 & 0 & (1 + \psi^2)\epsilon/\beta \end{pmatrix}, \quad (5.168)$$

where:

$$\psi = \frac{qB_0}{2P_0}\beta. \quad (5.169)$$

From the eigenvalues of  $\Sigma S$ , we find that the transverse emittances of the beam are:

$$\epsilon_I = \sqrt{1 + 2\psi\sqrt{1 + \psi^2} + 2\psi^2}\epsilon, \quad (5.170)$$

$$\epsilon_{II} = \sqrt{1 - 2\psi\sqrt{1 + \psi^2} + 2\psi^2}\epsilon. \quad (5.171)$$

Note that for any value of  $\psi$ :

$$\epsilon_I\epsilon_{II} = \epsilon^2. \quad (5.172)$$

Hence, one of the emittances can be made small only at the expense of making the other emittance large. The product of the beam emittances cannot be reduced by immersing the photocathode in a solenoid field. For  $\psi \gg 1$ , the emittances are approximately  $\epsilon_I \approx 2\psi\epsilon$  and  $\epsilon_{II} \approx \epsilon/2\psi$ .

Now suppose that we wish to decouple the beam, i.e. apply a transformation so that there are no correlations between horizontal and vertical phase space variables. An appropriate transformation can be achieved using a set of three skew quadrupoles. The transfer matrix for a skew quadrupole is given by (3.111); in the thin lens approximation, the transfer matrix is:

$$R_{sq} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -k_1^{(s)}L & 0 \\ 0 & 0 & 1 & 0 \\ -k_1^{(s)}L & 0 & 0 & 1 \end{pmatrix}, \quad (5.173)$$

where  $k_1^{(s)}L$  is the integrated gradient of the skew quadrupole, divided by the beam rigidity. Decoupling of a beam with the distribution matrix (5.168) is achieved using two skew quadrupoles with integrated strengths:

$$(k_1^{(s)}L)_A = (1 + \sqrt{2})\frac{\sqrt{1 + \psi^2}}{\beta}, \quad (5.174)$$

with a third skew quadrupole placed mid-way between them; the middle skew quadrupole should have integrated strength:

$$(k_1^{(s)}L)_B = -2\sqrt{2}\frac{\sqrt{1 + \psi^2}}{\beta}. \quad (5.175)$$

The drift spaces between the skew quadrupoles should have equal length:

$$L = \frac{1}{2\sqrt{1 + \sqrt{2}}}\frac{\beta}{\sqrt{1 + \psi^2}}. \quad (5.176)$$

The set of three skew quadrupoles should be located so that the beam enters the first skew quadrupole immediately after leaving the solenoid field (recall from Section 3.5 that the transfer matrix for a solenoid fringe field is the identity). After the third skew quadrupole the beam is decoupled and has distribution matrix:

$$\Sigma = \begin{pmatrix} u_+ \beta \epsilon & 0 & 0 & 0 \\ 0 & v_+ \epsilon / \beta & 0 & 0 \\ 0 & 0 & u_- \beta \epsilon & 0 \\ 0 & 0 & 0 & v_- \epsilon / \beta \end{pmatrix}, \quad (5.177)$$

where:

$$u_{\pm} = 1 \pm \frac{\psi}{\sqrt{1 + \psi^2}}, \quad (5.178)$$

$$v_{\pm} = 1 \pm \psi \sqrt{1 + \psi^2} + \psi^2. \quad (5.179)$$

The eigenvalues of  $\Sigma S$ , with  $\Sigma$  given by (5.177), are (5.170) and (5.171). Examples of the lattice functions in a flat-beam electron source are shown in Fig. 5.3.

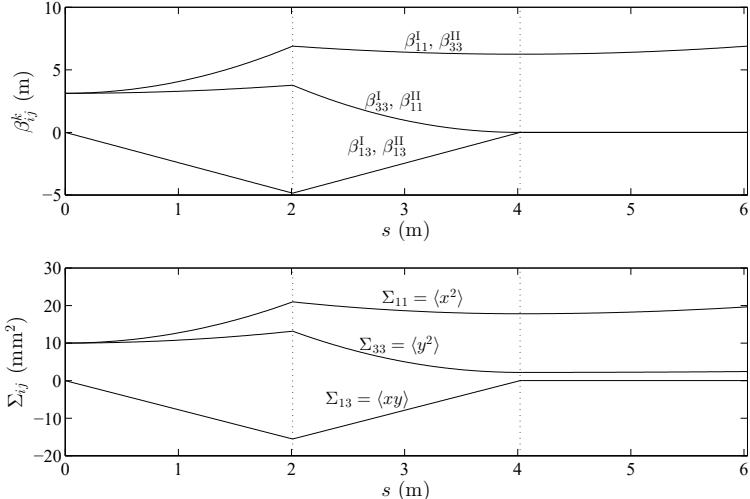


Fig. 5.3 Coupled lattice functions (top) and mean square beam size (bottom) as a function of longitudinal position in a flat-beam electron source. The beam exits the solenoid field at  $s = 0$ , where the first skew quadrupole is located. The positions of the other two skew quadrupoles are indicated by vertical dotted lines. In this example,  $\beta = 10 \text{ m}$ ,  $\psi = 1.25$  and  $\epsilon = 1 \mu\text{m}$ .

### 5.4.3 Measurement of normal mode emittances

Optimisation of the performance of an accelerator often requires measurement of the beam emittance. In beam lines where coupling is present (either by design, or through the effects of imperfections) it can be useful to measure the normal mode emittances rather than the projected emittances, since it is the normal mode emittances that are conserved under linear, symplectic transport. In this section, we discuss a technique for measuring the  $4 \times 4$  covariance (beam distribution) matrix  $\Sigma$  describing the second-order moments of the distribution of particles in transverse phase space, i.e. in the plane perpendicular to the beam trajectory. The normal mode transverse emittances are readily found from the covariance matrix as the eigenvalues of  $\Sigma S$ , where  $S$  is the antisymmetric matrix (4.14). We do not consider technical aspects of the diagnostics needed to perform the measurements, but rather focus on the principles of the beam dynamics. We shall limit the discussion to emittance measurements in linear beam lines, and will not consider the case of storage rings; we also assume that the longitudinal beam distribution is uncorrelated with the transverse distribution.

The technique that we discuss is based on observation of the transverse beam profile in  $x$ - $y$  co-ordinate space at a given point in an accelerator beam line. The emittances and optics functions describing the beam distribution in transverse phase space are found by observing how the co-ordinate space distribution varies with changes in the strengths of upstream quadrupole magnets. Typically, the beam profile is found by imaging light from a fluorescent (or scintillating) screen placed in the path of the beam, or by recording an intensity signal from a slit or a wire scanned across the beam. The choice of diagnostic technique will depend on the machine and beam parameters (particle type, beam energy, intensity, beam size regime etc.). Information on diagnostics techniques and devices for transverse beam profile measurements can be found elsewhere (for example, [Koziel (2005); Bravin (2009, 2020)]); here, we simply assume that it is possible to make measurements of the second-order moments of the beam distribution in co-ordinate space,  $\langle x^2 \rangle$ ,  $\langle y^2 \rangle$ , and  $\langle xy \rangle$ , at an appropriate location in the beam line. In some cases, (depending on the parameter regime of the accelerator) it may also be possible to measure the second-order moments characterising the transverse momenta, i.e.  $\langle p_x^2 \rangle$ ,  $\langle p_x p_y \rangle$ ,  $\langle p_y^2 \rangle$ , and the correlations between transverse momenta and co-ordinates,  $\langle x p_x \rangle$ ,  $\langle x p_y \rangle$ ,  $\langle y p_x \rangle$ , and  $\langle y p_y \rangle$ . If all these quantities can be measured directly, then all elements of the  $4 \times 4$  beam distribution matrix  $\Sigma$  are known, and it is possible

immediately to determine the normal mode emittances. However, suitable diagnostics for measurement of the transverse momenta distributions are not always available, and we shall assume that direct measurements are possible only of the co-ordinate space distributions.

If the beam distribution is close to a Gaussian, then the second-order moments can provide a good description of the distribution. However, the transverse phase space distribution often contains artifacts unrelated to the beam, and the beam itself may exhibit structures that are not very well described by a Gaussian profile. In such cases, although it is usually still possible to determine values for the second-order moments of the beam distribution, these quantities may be of limited value, and a more detailed description may be needed, such as one that can be provided by phase space tomography (see, for example, [McKee *et al.* (1995); Wolski *et al.* (2020)]). For the purposes of the present discussion, however, we shall assume that the second-order moments of the charge distribution in themselves provide a good description of the beam.

The first step in the procedure is to define a ‘reconstruction point’ and an ‘observation point’. The reconstruction point is the location in the beam line at which the transverse distribution will be determined. The observation point is a location downstream from the reconstruction point, at which appropriate diagnostics are positioned to allow measurement of the second-order moments of the charge distribution in transverse co-ordinate space. Focusing elements (such as quadrupoles) with strengths that can be varied over some range are needed between the reconstruction point and the observation point, to allow the beam size at the observation point to be varied while the distribution at the reconstruction point remains fixed. The focusing elements are calibrated so that for any given focusing strengths, the  $4 \times 4$  (transverse) transfer matrix  $M$  from the reconstruction point to the observation point is known with good accuracy.

Let us define two vectors, a 10-component vector  $\vec{\xi}$  consisting of the second-order moments of the transverse phase space distribution at the reconstruction point, and a 3-component vector  $\vec{\eta}$  consisting of the second-order moments of the transverse co-ordinate space distribution at the

observation point:

$$\vec{\xi} = \begin{pmatrix} \langle x^2 \rangle \\ \langle xp_x \rangle \\ \langle xy \rangle \\ \langle xp_y \rangle \\ \langle p_x^2 \rangle \\ \langle p_xy \rangle \\ \langle p_xp_y \rangle \\ \langle y^2 \rangle \\ \langle yp_y \rangle \\ \langle p_y^2 \rangle \end{pmatrix}, \quad \text{and} \quad \vec{\eta} = \begin{pmatrix} \langle x^2 \rangle_{\text{obs}} \\ \langle xy \rangle_{\text{obs}} \\ \langle y^2 \rangle_{\text{obs}} \end{pmatrix}. \quad (5.180)$$

The subscript on the elements of  $\vec{\eta}$  emphasises that these quantities refer to the observation point. For simplicity of notation, we omit any subscript on the second-order moments of the charge distribution at the reconstruction point (i.e. on the elements of  $\vec{\xi}$ ). The order in which elements appear within each vector is arbitrary. All the elements of the vector  $\vec{\eta}$  can be found by direct observation. However, only three of the elements of  $\vec{\xi}$  are directly observable; namely,  $\langle x^2 \rangle$ ,  $\langle xy \rangle$  and  $\langle y^2 \rangle$ , assuming appropriate diagnostics at the reconstruction point. Our goal is to use measurements of the elements of  $\vec{\eta}$  to determine all the elements of  $\vec{\xi}$ . This can be done by relating the elements of  $\vec{\eta}$  (for different strengths of the focusing elements between the reconstruction point and the observation point) to the elements of  $\vec{\xi}$  through the transfer matrix  $M$ , which will vary with changes in the focusing between reconstruction point and observation point.

For a given transfer matrix  $M$ , the covariance matrices  $\Sigma$  at the reconstruction point and  $\Sigma_{\text{obs}}$  at the observation point are related by:

$$\Sigma_{\text{obs}} = M\Sigma M^T. \quad (5.181)$$

The elements of the covariance matrices  $\Sigma$  and  $\Sigma_{\text{obs}}$  are the second-order moments of the beam distribution in phase space at the reconstruction and observation points, respectively. Hence, we can restructure (5.181) to write a relationship between elements of the covariance matrices in terms of the vectors  $\vec{\eta}$  and  $\vec{\xi}$ :

$$\vec{\eta} = D\vec{\xi}, \quad (5.182)$$

where the  $3 \times 10$  matrix  $D$  is constructed from the elements  $m_{ij}$  of the

transfer matrix  $M$ :

$$D^T = \begin{pmatrix} m_{11}^2 & m_{11}m_{31} & m_{31}^2 \\ 2m_{11}m_{12} & m_{12}m_{31} + m_{11}m_{32} & 2m_{31}m_{32} \\ 2m_{11}m_{13} & m_{13}m_{31} + m_{11}m_{33} & 2m_{31}m_{33} \\ 2m_{11}m_{14} & m_{14}m_{31} + m_{11}m_{34} & 2m_{31}m_{34} \\ m_{12}^2 & m_{12}m_{32} & m_{32}^2 \\ 2m_{12}m_{13} & m_{13}m_{32} + m_{12}m_{33} & 2m_{32}m_{33} \\ 2m_{12}m_{14} & m_{14}m_{32} + m_{12}m_{34} & 2m_{32}m_{34} \\ m_{13}^2 & m_{13}m_{33} & m_{33}^2 \\ 2m_{13}m_{14} & m_{14}m_{33} + m_{13}m_{34} & 2m_{33}m_{34} \\ m_{14}^2 & m_{14}m_{34} & m_{34}^2 \end{pmatrix}. \quad (5.183)$$

For a single set of focusing strengths between the reconstruction point and the observation point (i.e. for a single transfer matrix), it is not possible to solve (5.182) for all the elements of  $\vec{\xi}$ , since the number of constraints is insufficient: there are ten unknown values that we wish to determine, and we have only three constraints. However, by measuring the elements of  $\vec{\eta}$  for a number of different transfer matrices, it is possible to provide additional constraints, allowing (5.182) to be solved for  $\vec{\xi}$ . For each set of focusing strengths, we extend the vector  $\vec{\eta}$  and the matrix  $D$ , so (for  $N$  sets of focusing strengths, corresponding to  $N$  different transfer matrices):

$$\vec{\eta}_{\text{ext}} = \begin{pmatrix} \vec{\eta}^{(1)} \\ \vec{\eta}^{(2)} \\ \vec{\eta}^{(3)} \\ \vdots \\ \vec{\eta}^{(N)} \end{pmatrix}, \quad \text{and} \quad D_{\text{ext}} = \begin{pmatrix} D^{(1)} \\ D^{(2)} \\ D^{(3)} \\ \vdots \\ D^{(N)} \end{pmatrix}. \quad (5.184)$$

The vector  $\vec{\eta}_{\text{ext}}$  has  $3N$  elements, and  $D_{\text{ext}}$  is a  $3N \times 10$  matrix. Since  $\vec{\xi}$  is independent of the focusing strengths of beam line components downstream of the reconstruction point, with a sufficient number of observations at different focusing strengths, Eq. (5.182) can be inverted:

$$\vec{\xi} = D_{\text{ext}}^{-1}\vec{\eta}. \quad (5.185)$$

It is often desirable to ‘over-constrain’ the problem by making a large number  $N$  of observations, so that the matrix  $D_{\text{ext}}$  has many more rows than columns. The matrix  $D_{\text{ext}}^{-1}$  can then be constructed as the pseudoinverse of  $D_{\text{ext}}$  [Wang *et al.* (2018)], or constructed using singular value decomposition [Press *et al.* (2007a)]. The elements of  $\vec{\xi}$  can be reassembled to construct the covariance matrix  $\Sigma$ : the normal mode emittances and lattice functions

at the reconstruction point can then be found using the methods described in Section 5.2.

To develop a better understanding of the principles behind the technique described above for measuring the transverse normal mode emittances, it may be helpful to consider the special case where there is no coupling present: we can then treat the horizontal and vertical dynamics separately. We shall consider just the horizontal dynamics. Let us further suppose, for simplicity, that the beam line between the reconstruction point and the observation point contains only a thin quadrupole of focusing strength  $K$ , immediately followed by a field-free region (drift space) of length  $L$ . The transfer matrix from the reconstruction point to the observation point is then:

$$M = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -K & 1 \end{pmatrix} = \begin{pmatrix} 1 - KL & L \\ -K & 1 \end{pmatrix}. \quad (5.186)$$

The matrix  $D$  in this case is constructed from (5.183), omitting all elements that involve  $m_{ij}$  for  $i > 2$  or  $j > 2$  (i.e. any element referring to the vertical dynamics). We are left with a  $1 \times 3$  matrix, relating the beam size  $\langle x^2 \rangle_{\text{obs}}$  at the observation point to the elements  $\langle x^2 \rangle$ ,  $\langle xp_x \rangle$  and  $\langle p_x^2 \rangle$  of the covariance matrix at the reconstruction point:

$$D = (m_{11}^2 \ 2m_{11}m_{12} \ m_{12}^2) = ((1 - KL)^2 \ 2L(1 - KL) \ L^2). \quad (5.187)$$

If we measure the beam size at the observation point for three different quadrupole focusing strengths ( $K^{(1)}$ ,  $K^{(2)}$  and  $K^{(3)}$ ), then the extended matrix  $D_{\text{ext}}$  is given by:

$$D_{\text{ext}} = \begin{pmatrix} (1 - K^{(1)}L)^2 & 2L(1 - K^{(1)}L) & L^2 \\ (1 - K^{(2)}L)^2 & 2L(1 - K^{(2)}L) & L^2 \\ (1 - K^{(3)}L)^2 & 2L(1 - K^{(3)}L) & L^2 \end{pmatrix}. \quad (5.188)$$

Then, from (5.185) we have:

$$\begin{pmatrix} \langle x^2 \rangle \\ \langle xp_x \rangle \\ \langle p_x^2 \rangle \end{pmatrix} = D_{\text{ext}}^{-1} \begin{pmatrix} \langle x^2 \rangle_{\text{obs}}^{(1)} \\ \langle x^2 \rangle_{\text{obs}}^{(2)} \\ \langle x^2 \rangle_{\text{obs}}^{(3)} \end{pmatrix}. \quad (5.189)$$

With three observations of the beam size at three different quadrupole strengths, the  $3 \times 3$  matrix  $D_{\text{ext}}$  can be inverted to solve (5.189). In practice, because of experimental errors, it can be advantageous even in this relatively simple situation to make more than three observations of the beam size (so that  $D_{\text{ext}}$  is  $N \times 3$ , with  $N > 3$ ), and use singular value

decomposition or a similar pseudoinverse algorithm to solve (5.189): the values then found for  $\langle x^2 \rangle$ ,  $\langle xp_x \rangle$  and  $\langle p_x^2 \rangle$  represent a best fit to the data.

For uncoupled motion (i.e. motion in one degree of freedom), we can express the  $(2 \times 2)$  covariance matrix in terms of the emittance  $\epsilon_x$  and Courant–Snyder parameters  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  in the usual way (4.102):

$$\begin{pmatrix} \langle x^2 \rangle & \langle xp_x \rangle \\ \langle xp_x \rangle & \langle p_x^2 \rangle \end{pmatrix} = \begin{pmatrix} \beta_x & -\alpha_x \\ -\alpha_x & \gamma_x \end{pmatrix} \epsilon_x, \quad (5.190)$$

where the condition  $\beta_x \gamma_x - \alpha_x^2$  leads to:

$$\epsilon_x = \sqrt{\langle x^2 \rangle \langle p_x^2 \rangle - \langle xp_x \rangle^2}. \quad (5.191)$$

For uncoupled motion, and with a beam line containing only a thin quadrupole and drift between the reconstruction point and the observation point, a variation on the technique discussed so far may be developed by noting how the beam size at the observation point  $\langle x^2 \rangle_{\text{obs}}$  depends on the focusing strength  $K$  of the quadrupole. From the relationship:

$$\langle x^2 \rangle_{\text{obs}} = D \begin{pmatrix} \langle x^2 \rangle \\ \langle xp_x \rangle \\ \langle p_x^2 \rangle \end{pmatrix}, \quad (5.192)$$

we find that for any given value of  $K$ :

$$\langle x^2 \rangle_{\text{obs}} = aK^2L^2 - 2bKL + c, \quad (5.193)$$

where the constants  $a$ ,  $b$  and  $c$  depend only on the elements of the covariance matrix at the reconstruction point, and the drift length  $L$ :

$$a = \langle x^2 \rangle, \quad (5.194)$$

$$b = \langle x^2 \rangle + L\langle xp_x \rangle, \quad (5.195)$$

$$c = L^2\langle p_x^2 \rangle + 2L\langle xp_x \rangle + \langle x^2 \rangle. \quad (5.196)$$

We see from (5.193) that the square of the beam size at the observation point is given by a second-order polynomial in  $KL$ , with coefficients  $a$ ,  $-2b$  and  $c$ . To determine the values of these coefficients, the experimental procedure (as before) involves recording the beam size at the observation point for a number of discrete settings of the quadrupole strength  $K$ : the values of  $a$ ,  $b$  and  $c$  may then be obtained by fitting a quadratic curve to a set of points on a plot of  $\langle x^2 \rangle_{\text{obs}}$  as a function of  $KL$ . Inverting equations (5.194), (5.195) and (5.196), we have:

$$\langle x^2 \rangle = a, \quad (5.197)$$

$$\langle xp_x \rangle = \frac{b - a}{L}, \quad (5.198)$$

$$\langle p_x^2 \rangle = \frac{a - 2b + c}{L^2}. \quad (5.199)$$

Using (5.190) and (5.191), we then find:

$$\epsilon_x = \frac{\sqrt{ac - b^2}}{L}, \quad (5.200)$$

$$\beta_x = \frac{aL}{\sqrt{ac - b^2}}, \quad (5.201)$$

$$\alpha_x = \frac{a - b}{\sqrt{ac - b^2}}. \quad (5.202)$$

$\gamma_x$  can be found from the usual relationship  $\beta_x \gamma_x - \alpha_x^2 = 1$ .

Although the data analysis takes a different form in this latter approach (which avoids the need to construct and invert the matrix  $D_{\text{ext}}$ ) the two techniques are fundamentally similar: equations (5.197)–(5.199) essentially provide a change of basis from the quantities  $a$ ,  $b$  and  $c$  (determined by fitting a curve to a set of measured data points) to the quantities  $\langle x^2 \rangle$ ,  $\langle xp_x \rangle$  and  $\langle p_x^2 \rangle$  (determined by inverting a matrix relating these quantities to a set of measured values). To make this explicit, we can rewrite (5.193) in the form:

$$\langle x^2 \rangle_{\text{obs}} = (K^2 L^2 \quad KL \ 1) \begin{pmatrix} a \\ -2b \\ c \end{pmatrix}. \quad (5.203)$$

Substituting from (5.194), (5.195) and (5.196) then gives:

$$\langle x^2 \rangle_{\text{obs}} = (K^2 L^2 \quad KL \ 1) \begin{pmatrix} \langle x^2 \rangle \\ -2\langle x^2 \rangle - 2L\langle xp_x \rangle \\ L^2\langle p_x^2 \rangle + 2L\langle xp_x \rangle + \langle x^2 \rangle \end{pmatrix}, \quad (5.204)$$

which can be written in the form:

$$\langle x^2 \rangle_{\text{obs}} = (K^2 L^2 \quad KL \ 1) \begin{pmatrix} 1 & 0 & 0 \\ -2 & -2L & 0 \\ 1 & 2L & L^2 \end{pmatrix} \begin{pmatrix} \langle x^2 \rangle \\ \langle xp_x \rangle \\ \langle p_x^2 \rangle \end{pmatrix}. \quad (5.205)$$

Finally, the first two factors in the matrix multiplication on the right-hand side of (5.205) can be combined to give:

$$\langle x^2 \rangle_{\text{obs}} = ((1 - K^2 L^2) \quad 2L(1 - KL) \quad L^2) \begin{pmatrix} \langle x^2 \rangle \\ \langle xp_x \rangle \\ \langle p_x^2 \rangle \end{pmatrix}, \quad (5.206)$$

which is in agreement with (5.187). In the first of the two approaches, the unknowns are determined by inverting a matrix which is then applied to a set of measured data points; in the second approach, the unknowns

(expressed in a different basis) are determined by fitting a curve to a set of measured data points.

As we mentioned at the start of this section, measurement of the transverse emittances is an important and routine procedure in many types of accelerator. There are numerous techniques that can be applied, but the quadrupole scan method, using a single quadrupole as described here, is widely used. For further discussion and some examples, see (for example) [Minty and Zimmermann (2003b); Dijkstal *et al.* (2015); Green and Shin (2015); Keil *et al.* (2017)].

## Chapter 6

# Linear Imperfections in Storage Rings

Achieving the performance specifications in an accelerator often places demanding requirements on the field quality and alignment precision of the magnets and other components that make up the beam line. The movement of a quadrupole magnet (that may weigh tens of kilograms and have dimensions of a few metres) by just a few micrometres can have a measurable effect on the beam. Usually, identifying and correcting the significant sources of machine errors is a major task when commissioning a new accelerator; and efforts to minimise errors and improve machine performance often continue throughout the lifetime of a facility. Achieving a good level of performance in an accelerator requires a thorough understanding of the effects of different types of error that can occur. Furthermore, modelling the sensitivity to various errors is an essential part of the design process for a new accelerator: the design study must include not just realistic models of machine errors and their effects, but should also consider in detail the diagnostics needed to identify the sources of errors, and the correction systems needed to compensate them.

The relative importance of different kinds of error depends very much on the type and design of the accelerator concerned. In this chapter, we shall not attempt an exhaustive treatment of machine errors and their effects; instead, we shall introduce and explain the important concepts and techniques by focusing on errors in synchrotron storage rings. We shall discuss the errors that are most commonly of importance in the operation of storage rings, including steering, focusing and coupling errors. A comprehensive review of practical techniques for characterising and optimising accelerator beam lines can be found, for example, in [Minty and Zimmermann (2003c)].

## 6.1 The Closed Orbit

In a storage ring, the reference trajectory is usually defined in such a way that a particle initially travelling on the reference trajectory will remain on the reference trajectory and return to its initial position (with its initial momentum) after one complete turn. In other words, the phase space vector initially and after one revolution is  $\vec{x} = (0, 0, 0, 0, 0, 0)$ . This will be the case if the reference trajectory passes along the magnetic axis of all quadrupole (and higher-order multipole) magnets, and if the curvature  $h$  of the reference trajectory in all dipoles is equal to the curvature of the physical trajectory of particles with the reference momentum:

$$h = \frac{1}{\rho} = \frac{q}{P_0} B, \quad (6.1)$$

where  $\rho$  is the radius of curvature of the reference trajectory,  $q$  is the particle charge,  $P_0$  is the reference momentum, and  $B$  is the dipole field. We also require that the energy of the particle remains constant, so we ignore effects of synchrotron radiation, and we assume that the fields in rf cavities are phased so that a particle sees no accelerating (or decelerating) field when passing through them.

The *closed orbit* in a storage ring is defined as the trajectory that has periodicity equal to one turn. That is, if the reference trajectory has circumference  $C_0$ , then the phase space vector  $\vec{x}(s_0)$  is on the closed orbit at  $s = s_0$  if:

$$\vec{x}(s_0 + C_0) = \vec{x}(s_0). \quad (6.2)$$

For simplicity, we shall assume that in the absence of any errors in the storage ring under consideration, the closed orbit is the same as the reference trajectory. That is, the ring has been designed so that all the assumptions in the previous paragraph are satisfied. When the ring is constructed, errors in the fields and alignment of the magnets will mean that the closed orbit does not follow the reference trajectory. For example, an error in the strength of a dipole field results in the trajectory of a particle deviating from the reference trajectory as it passes through the dipole, even if it enters the dipole exactly on the reference trajectory. When modelling the effect of errors in a storage ring, the first task is to construct a model of the storage ring including the errors of interest. Assuming this has been done, the next task is often to find the closed orbit.

If the single-turn map for the storage ring is known, then the closed orbit can be found as follows. Let the zeroth-order and first-order terms in

the transfer map be represented by a vector  $\vec{m}$  and matrix  $R$  respectively, so that the *linear* transfer map can be written as:

$$\vec{x}(s_0 + C_0) = R\vec{x}(s_0) + \vec{m}. \quad (6.3)$$

If  $\vec{x}_{\text{co}}$  is the phase space vector for the closed orbit at  $s = s_0$ , then by definition of the closed orbit:

$$\vec{x}_{\text{co}} = R\vec{x}_{\text{co}} + \vec{m}. \quad (6.4)$$

From this, it follows that:

$$\vec{x}_{\text{co}} = (I - R)^{-1}\vec{m}, \quad (6.5)$$

where  $I$  is the identity matrix. The phase space vector for the closed orbit at other points around the ring can be found by applying the transfer maps from  $s_0$  to different locations. There is no guarantee that the inverse of  $I - R$  exists: if it does not, then it is likely to be difficult to store a beam in the storage ring. Even if the inverse of  $I - R$  does exist, then even though the closed orbit can be found mathematically, there may be points where it hits some physical obstruction within the vacuum chamber, or even lies outside the vacuum chamber altogether.

Higher-order terms in the map are not always negligible, and this means that it is often necessary to apply (6.5) iteratively. This can be done as follows. Suppose that  $\vec{x}_n$  is an approximation to the closed orbit, so that  $\vec{x}_{\text{co}} \approx \vec{x}_n$ . Then, to obtain an improved approximation, the linear part of the map around  $\vec{x}_n$  is constructed; that is, we find the map for which the first-order and zeroth-order terms are represented by a matrix  $R_n$  and vector  $\vec{m}_n$  respectively, and where the effect of the linear map can be written:

$$\vec{x}(s_0 + C_0) - \vec{x}_n = R_n(\vec{x}(s_0) - \vec{x}_n) + \vec{m}_n. \quad (6.6)$$

An improved estimate for the closed orbit (assuming that it exists) is then found from:

$$\vec{x}_{n+1} = \vec{x}_n + (I - R_n)^{-1}\vec{m}_n. \quad (6.7)$$

Note that as  $\vec{x}_n \rightarrow \vec{x}_{\text{co}}$ ,  $\vec{m}_n \rightarrow 0$ . For each iteration step, the matrix  $R_n$  and vector  $\vec{m}_n$  can be constructed numerically. A possible method is to track a set of particles with initial phase space vectors  $\vec{x}^{(j)}$  close to  $\vec{x}_n$ :

$$\frac{[\vec{x}^{(j)} - \vec{x}_n]_i}{d} = \delta_{ij}, \quad (6.8)$$

where the notation  $[\vec{a}]_i$  indicates the  $i$ th component of the vector  $\vec{a}$ ,  $\delta_{ij}$  is the Kronecker delta, and  $d$  is a small parameter. Then the matrix  $R_n$  is given by:

$$[R_n]_{ij} = \lim_{d \rightarrow 0} \frac{[\vec{x}^{(j)}(s_0 + C_0) - \vec{x}_n(s_0 + C_0)]_i}{d}, \quad (6.9)$$

where  $[R_n]_{ij}$  is the element in the  $i$ th row and  $j$ th column of  $R$ . The vector  $\vec{m}_n$  is given by:

$$\vec{m}_n = \vec{x}_n(s_0 + C_0) - \vec{x}_n(s_0). \quad (6.10)$$

As an initial approximation to the closed orbit, it is often sufficient simply to choose the reference trajectory, so that  $\vec{x}_0 = (0, 0, 0, 0, 0, 0)$ .

## 6.2 Dipole Field Errors

The methods discussed in Section 6.1 are useful for computing the closed orbit in a model of a storage ring with given errors. However, they do not give much insight into the relevant features of the storage ring design and how they affect the sensitivity of the storage ring to different types of error. In this section, we shall take a different approach, and develop a formula for computing the closed orbit distortion arising from a dipole field error, given the Courant–Snyder parameters and other optical properties of the storage ring. For simplicity, we shall consider a field error that has only a horizontal field component, i.e. a field component in the plane of the ring. We shall also assume that there is no coupling between the vertical motion and either the horizontal or longitudinal motion, so that the vertical motion can be considered independently of motion in the other degrees of freedom. Our goal is to find an expression for the vertical closed orbit (with respect to the reference trajectory) in terms of the field error and the lattice functions in the storage ring. After discussing the vertical closed orbit, we shall consider at the end of this section some of the complications that arise when dealing with horizontal orbit distortion in synchrotron storage rings.

To derive the formula for the vertical closed orbit, we begin by writing the trajectory of a single particle in terms of action–angle variables:

$$y = \sqrt{2\beta_y J_y} \cos(\phi_y), \quad (6.11)$$

$$p_y = -\sqrt{\frac{2J_y}{\beta_y}} (\sin(\phi_y) + \alpha_y \cos(\phi_y)), \quad (6.12)$$

where  $y$  is the vertical co-ordinate,  $p_y$  the scaled canonical momentum,  $J_y$  and  $\phi_y$  the vertical action and angle variables respectively, and  $\alpha_y$  and  $\beta_y$  are the vertical Courant–Snyder alpha and beta functions. Note that all quantities are (at least in principle) functions of the independent variable  $s$  that represents the distance along the reference trajectory. For linear symplectic transport, the action  $J_y$  is constant. In the absence of any vertical steering, the closed orbit is given by  $J_y = 0$ , i.e.  $y = p_y = 0$ , which

is the same as the reference trajectory. A horizontal magnetic dipole field at some point in the ring will cause a change in  $p_y$  as the particle passes that point: with such a field, the closed orbit is no longer the reference trajectory. We shall look for a particle trajectory that closes on itself after one turn in the presence of a horizontal dipole field at one point in the ring. This trajectory will be the closed orbit.

Suppose that the dipole field is located at a single point  $s = s_0$  in the storage ring, and that it causes a change  $\Delta p_y$  in  $p_y$  as a particle passes that point:

$$\Delta p_y \approx \frac{q}{P_0} \int B_x ds, \quad (6.13)$$

where  $P_0$  is the reference momentum, and  $\int B_x ds$  is the integrated strength of the horizontal dipole field along the reference trajectory. The conditions for the particle trajectory to close on itself can be written in terms of the Cartesian variables:

$$y(s_0 + C_0) = y(s_0), \quad (6.14)$$

$$p_y(s_0 + C_0) + \Delta p_y = p_y(s_0), \quad (6.15)$$

where we treat  $s_0$  as being the point immediately after the dipole field error. Let  $J_{y0}$  and  $\phi_{y0}$  be the action and angle of a particle on the closed orbit immediately after the horizontal dipole field. If  $\mu_y$  is the total phase advance over one complete turn of the ring, then the closed orbit conditions (6.14) and (6.15) can be written in terms of the action-angle variables:

$$\sqrt{2\beta_y J_{y0}} \cos(\phi_{y0} + \mu_y) = \sqrt{2\beta_y J_{y0}} \cos(\phi_{y0}), \quad (6.16)$$

and:

$$\begin{aligned} -\sqrt{\frac{2J_{y0}}{\beta_y}} (\sin(\phi_{y0} + \mu_y) + \alpha_y \cos(\phi_{y0} + \mu_y)) + \Delta p_y = \\ -\sqrt{\frac{2J_{y0}}{\beta_y}} (\sin(\phi_{y0} + \mu_y) + \alpha_y \cos(\phi_{y0} + \mu_y)), \end{aligned} \quad (6.17)$$

where the Courant-Snyder parameters are to be evaluated at  $s = s_0$ . Equations (6.16) and (6.17) may be solved to give:

$$J_{y0} = \frac{\beta_y \Delta p_y^2}{8 \sin^2(\pi \nu_y)}, \quad (6.18)$$

$$\phi_{y0} = -\pi \nu_y, \quad (6.19)$$

where  $\nu_y = \mu_y/2\pi$  is the vertical tune.

From (6.18), we immediately see that if the tune  $\nu_y$  is an integer, then the action is infinite: no closed orbit will exist. If the tune is close to an integer, then  $\sin(\pi\nu_y)$  will be small, and even a small deflection  $\Delta p_y$  can lead to a large distortion of the closed orbit. Notice also that the size of the closed orbit distortion depends on the value of the beta function at the location of the field error: if the beta function is large, then a small deflection can again lead to a large closed orbit distortion. To reduce sensitivity of the orbit to field errors, the storage ring must be designed with a tune far from an integer, and with beta functions kept reasonably low (typically, less than 100 m). In some situations (for example, interaction regions in colliders) it may not be possible to avoid beta functions reaching large values, even at many kilometres: in such situations, careful attention must be paid to the design, fabrication and installation of magnets in the large-beta regions, to make sure that the fields that may cause distortion of the closed orbit are kept as small as possible.

Using (6.18) and (6.19) we can write an expression for the closed orbit at a location  $s$ , resulting from a single dipole field error at  $s = s_0$ :

$$y_{\text{co}}(s) = \frac{\sqrt{\beta_y(s)\beta_y(s_0)}}{2\sin(\pi\nu_y)} \Delta p_y \cos(\mu_y(s, s_0) - \pi\nu_y), \quad (6.20)$$

where  $\mu_y(s, s_0)$  is the vertical phase advance from  $s_0$  to  $s$ . The closed orbit at any point depends on the beta function at that point, as well as on the beta function at the location of the dipole field error. We also observe that the closed orbit ‘oscillates’ about the reference trajectory as we move around the ring. The rate of oscillation is the same as the rate at which the angle variable increases: a horizontal dipole field error at a single point in the ring will lead to  $\nu_y$  oscillations of the closed orbit, where  $\nu_y$  is the vertical tune.

Of course, there is usually not just one error in a storage ring, but a large number of errors distributed around the ring. If the orbit distortion is small enough that we can neglect nonlinear effects, then we may simply use linear superposition to add together the effects from all the different errors. In that case, for a horizontal dipole field error  $B_x(s)$ , which is a function of position around the ring, the closed orbit is:

$$y_{\text{co}}(s) = \int_0^{C_0} \frac{\sqrt{\beta_y(s)\beta_y(s')}}{2\sin(\pi\nu_y)} \frac{q}{P_0} B_x(s') \cos(\mu_y(s, s') - \pi\nu_y) ds', \quad (6.21)$$

where  $C_0$  is the length of the reference trajectory (i.e. the ring circumference). The shape of the closed orbit is clearly more complicated when dipole field errors are located at many points around the ring, compared to

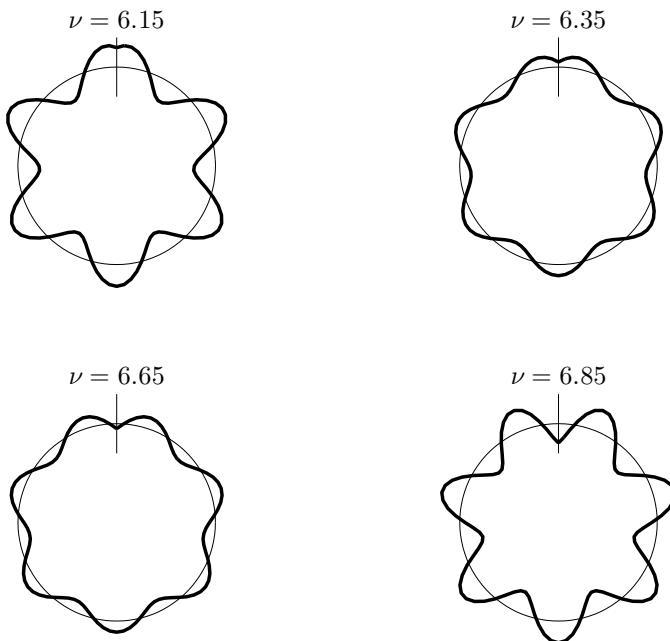


Fig. 6.1 Distortion of the closed orbit in a storage ring with different tune values. In each case, the reference trajectory is represented by a circle, with a straight line cutting the circumference at the location of a dipole ‘kick’. The size of the dipole kick is the same for all cases. The closed orbit is represented by the distance of the heavy line from the reference trajectory, and is obtained from (6.20) with constant beta functions. The amplitude of the closed orbit is largest for tune values close to an integer, and smallest for tune values close to a half integer.

when a field error is present at only a single location; but in practice, it is often found that a small number of errors dominate, so that a closed orbit distortion frequently appears with an oscillation period roughly equal to the betatron period. The effect of different tune values on the closed orbit in a storage ring is shown schematically in Fig. 6.1.

Equation (6.21) provides a useful formula for calculating the closed orbit. It should be remembered that it is based on several assumptions: in particular, we have assumed that we are dealing with the vertical orbit in a planar storage ring, where the closed orbit in the absence of field errors is the same as the reference trajectory, and that there is no coupling between the vertical and either the horizontal or longitudinal motion. Nevertheless, even where these assumptions are not valid, some of the general properties

we observe in (6.21) still hold true. For example, the closed orbit is very sensitive to field errors when the tune is an integer, or if the field error occurs at a location where the beta function is large.

When coupling is present, complicated effects can occur. This is true for both coupling between horizontal and vertical motion, and coupling between either horizontal or vertical motion and longitudinal motion. For example, consider a vertical field error that leads to a horizontal deflection and a distortion of the horizontal closed orbit. An orbit distortion in a region where the reference trajectory is curved (for example, in a main dipole magnet) results in a change in path length. In a synchrotron storage ring, the overall path length in the machine must remain fixed, in order for the revolution period of particles to remain in phase with the rf cavities. In order to compensate a change in the path length resulting from a closed orbit distortion, the beam will change energy, to follow a dispersive orbit. The final orbit will therefore be a combination of a ‘direct’ orbit distortion resulting from the field error, and an ‘indirect’ orbit distortion (actually, just some amount of dispersion). Equation (6.21) will not be valid in such a case, and a more general approach, such as that described in Section 6.1, must be used to find the closed orbit.

It is apparent from (6.20) that the closed orbit at any particular location in a storage ring depends not only on the steering errors in the ring, but also on the optics. As a consequence, by measuring the change in closed orbit in response to a known change in a dipole field at a particular location, it is possible to characterise the storage ring optics. This is the basis of a technique for identifying tuning errors in a storage ring, known as *orbit response matrix analysis* (ORM analysis) [Safranek (1997); Minty and Zimmerman (2003d)]. A storage ring will usually have some number of small dipole (orbit corrector) magnets distributed around the circumference of the ring for making small adjustments to the closed orbit. The orbit response matrix is constructed by measuring the change in the closed orbit at each beam position monitor in response to a change in the current in each orbit corrector magnet. By comparing the measured orbit response matrix with an orbit response matrix constructed from a computer model of the storage ring, it is possible to determine with good accuracy the source of various kinds of optics errors, including focusing and coupling errors.

### 6.3 Quadrupole Alignment Errors

Quadrupole alignment errors are usually a significant source of closed orbit distortion in a storage ring. Suppose we have a storage ring in which all the magnets are perfectly aligned, so that a particle on the reference trajectory passes through the magnetic centre of every quadrupole. Such a particle sees no magnetic field in any of the quadrupoles, and so passes through each quadrupole in a straight line. The only steering comes from the dipole magnets, which we assume bend the beam horizontally, so that the reference trajectory (and, for a perfectly aligned machine, the closed orbit) lies in a horizontal plane. If one quadrupole is then moved vertically, then a particle moving along the reference trajectory at that point now sees a horizontal magnetic field. This will result in the particle receiving a vertical deflection; the closed orbit is no longer exactly along the reference trajectory.

It is useful to be able to calculate the amount of closed orbit distortion arising from quadrupole alignment errors in a storage ring. We shall restrict the discussion to vertical alignment errors in a ring where, in the absence of errors, the closed orbit is the same as the reference trajectory, which lies in a horizontal plane. Suppose that a single quadrupole, of integrated gradient  $k_1 L$ , is moved a vertical distance  $Y_{\text{quad}}$  with respect to the reference trajectory. The normalised quadrupole gradient is given by:

$$k_1 = \frac{q}{P_0} \frac{\partial B_x}{\partial y}, \quad (6.22)$$

and  $k_1 > 0$  for a *horizontally focusing* (i.e. vertical defocusing) quadrupole. Now let  $y_{\text{co}}$  be the change in the closed orbit at the location of the quadrupole, resulting from the movement of the quadrupole. Taking into account the new vertical position of the quadrupole, a particle moving along the closed orbit passes through the quadrupole a vertical distance  $y_{\text{co}} - Y_{\text{quad}}$  above the magnetic axis. Using the thin lens approximation for the quadrupole, the vertical deflection (i.e. the change in  $p_y$ ) resulting from the quadrupole field is:

$$\Delta p_y = k_1 L (y_{\text{co}} - Y_{\text{quad}}). \quad (6.23)$$

Using (6.20), the closed orbit at the location of the quadrupole is:

$$y_{\text{co}} = \frac{\beta_y}{2} \cot(\pi\nu_y) \Delta p_y = \frac{\beta_y}{2} \cot(\pi\nu_y) k_1 L (y_{\text{co}} - Y_{\text{quad}}), \quad (6.24)$$

where  $\beta_y$  is the vertical beta function at the location of the quadrupole. Solving for  $y_{\text{co}}$ , we find:

$$y_{\text{co}} = \frac{a}{a-1} Y_{\text{quad}}, \quad (6.25)$$

where:

$$a = \frac{\beta_y}{2} \cot(\pi\nu_y) k_1 L. \quad (6.26)$$

The change in  $p_y$  at the location of the quadrupole is:

$$\Delta p_y = k_1 L (y_{\text{co}} - Y_{\text{quad}}) = \frac{k_1 L}{a - 1} Y_{\text{quad}}. \quad (6.27)$$

Using (6.20) and (6.27), the closed orbit at any point around the ring can be calculated as a function of the vertical alignment of a single quadrupole.

The beta function and the tune play a role in determining the sensitivity of the closed orbit to quadrupole alignment errors, just as they did in the case of horizontal dipole field errors. However, in the case of a quadrupole alignment error, the behaviour is a little more complicated. If the beta function at a quadrupole is small, the integrated strength of the quadrupole is not too large, and the tune of the ring is not too close to an integer, then we may have  $a \ll 1$ , and the closed orbit will not be very sensitive to an alignment error on the quadrupole. In the other extreme, if  $a$  is large ( $a \gg 1$ ), then as the quadrupole is moved vertically, the closed orbit will change so that it always passes through the quadrupole close to the magnetic axis of the quadrupole: equation (6.25) shows that in this case  $y_{\text{co}} \approx Y_{\text{quad}}$ , and from (6.27) we have  $\Delta p_y \approx 0$ . However, if  $a \approx 1$  then the closed orbit is very sensitive to the quadrupole alignment: a small alignment error on the quadrupole can lead to a very large orbit distortion.

## 6.4 Focusing Errors

In Section 6.3, we saw that an alignment error on a quadrupole in a storage ring leads to a distortion of the closed orbit. In this section, we shall consider the effect of a focusing error arising from a variation of the field strength in a quadrupole. If the quadrupole is aligned so that a particle on the closed orbit passes through the centre of the quadrupole, then varying the strength of the quadrupole field has no effect on the closed orbit, since the field at the centre of a quadrupole is zero. However, particles performing betatron oscillations about the closed orbit will see a focusing effect that depends on the strength of the quadrupole.

To understand the effect of varying the strength of a quadrupole at one location in a storage ring, let us begin by writing the transfer matrix for one complete turn of the storage ring starting from a point immediately after the quadrupole. We shall assume that there is no coupling in the storage ring, so that the horizontal motion can be treated independently

of the vertical and longitudinal motion: then, we can write the transfer matrix for the horizontal motion as a  $2 \times 2$  matrix. Although we consider specifically the horizontal motion, if there is no coupling then the same analysis can also be applied to the vertical motion (though bearing in mind that a horizontally focusing field is vertically defocusing, and vice versa). We initially consider the situation when the quadrupole is set to its nominal strength, i.e. there is no focusing error. In terms of the Courant–Snyder parameters at the location of the quadrupole and the betatron tune, the single-turn transfer matrix is (from (4.15)):

$$R = \begin{pmatrix} \cos(2\pi\nu_x) + \alpha_x \sin(2\pi\nu_x) & \beta_x \sin(2\pi\nu_x) \\ -\gamma_x \sin(2\pi\nu_x) & \cos(2\pi\nu_x) - \alpha_x \sin(2\pi\nu_x) \end{pmatrix}, \quad (6.28)$$

where  $\alpha_x$ ,  $\beta_x$ ,  $\gamma_x$  are the horizontal Courant–Snyder parameters, and  $\nu_x$  is the horizontal tune.

Now we introduce a focusing error arising from an increase  $\Delta K$  in the focusing strength of the quadrupole. Using thin lens approximation, the focusing error can be represented by a transfer matrix  $R_{\text{err}}$ :

$$R_{\text{err}} = \begin{pmatrix} 1 & 0 \\ -\Delta K & 1 \end{pmatrix}. \quad (6.29)$$

The transfer matrix for one complete turn of the storage ring including the focusing error is given by:

$$\bar{R} = R R_{\text{err}}. \quad (6.30)$$

We can express the new transfer matrix  $\bar{R}$  in terms of modified Courant–Snyder parameters  $\bar{\alpha}_x$ ,  $\bar{\beta}_x$ ,  $\bar{\gamma}_x$  and a modified betatron tune  $\bar{\nu}_x$ :

$$\bar{R} = \begin{pmatrix} \cos(2\pi\bar{\nu}_x) + \bar{\alpha}_x \sin(2\pi\bar{\nu}_x) & \bar{\beta}_x \sin(2\pi\bar{\nu}_x) \\ -\bar{\gamma}_x \sin(2\pi\bar{\nu}_x) & \cos(2\pi\bar{\nu}_x) - \bar{\alpha}_x \sin(2\pi\bar{\nu}_x) \end{pmatrix}. \quad (6.31)$$

We can perform the matrix multiplication in (6.30) to find an expression for  $\bar{R}$  in terms of the original Courant–Snyder parameters, the original betatron tune and the focusing error. Comparing the elements of the matrix product  $R R_{\text{err}}$  with the elements of  $\bar{R}$  in the form (6.31) gives a set of equations that can be solved to obtain expressions for the new Courant–Snyder parameters and betatron tune. In particular, we find for the new betatron tune:

$$\bar{\nu}_x = \nu_x + \Delta\nu_x, \quad (6.32)$$

where:

$$\Delta\nu_x \approx \frac{\Delta K \beta_x}{4\pi}. \quad (6.33)$$

The change in betatron tune  $\Delta\nu_x$  depends on the change in focusing strength, and on the value of the beta function at the location of the error. The new beta function is given by:

$$\bar{\beta}_x \approx \frac{\beta_x}{1 + \frac{1}{2}\Delta K \beta_x \cot(2\pi\nu_x)}. \quad (6.34)$$

The fact that the tune change resulting from a focusing error depends on the beta function at the location of the error provides a convenient way to measure the beta function in a storage ring. The betatron tune can be measured by exciting transverse oscillations of the beam using an oscillating electric or magnetic field: the amplitude of the induced beam oscillations is measured as a function of the field oscillation frequency. At the betatron frequency, the beam oscillations resonate with the field oscillations, and the oscillations of the beam are driven to large amplitude. The betatron tune is found from the resonant frequency. The beta function at the location of a given quadrupole can be found by measuring the betatron tune as a function of the focusing strength of the quadrupole.

Now let us look in more detail at the change in the beta function resulting from a change in focusing strength of a quadrupole. From (6.34) it appears that the beta function in the presence of a focusing error vanishes for integer or half-integer values of the tune  $\nu_x$ , even for infinitesimal focusing errors (since  $\cot(2\pi\nu_x)$  becomes infinite for integer or half-integer values of  $\nu_x$ ). In reality, the beta function cannot completely vanish or change sign: note that (6.34) gives an approximate expression for the beta function in the presence of a focusing error. However, the beta function does become extremely sensitive to focusing errors at integer or half-integer values of the tune. This phenomenon is also apparent from the transfer matrix (6.28): at integer values of the tune,  $R = I$ , and at half-integer values,  $R = -I$ , where  $I$  is the identity matrix. In either case, it is not possible to determine the values of the Courant–Snyder parameters. Recall from Section 6.2 that the closed orbit in a storage ring is extremely sensitive to dipole field errors at integer values of the tune: now we find that the focusing effects in a storage ring (which are characterised by the beta function) are extremely sensitive to quadrupole field errors at integer and half-integer values of the tune. The sensitivity of the closed orbit and the beta function to particular types of error at certain values of the tune can be understood in terms of *betatron resonances*: resonance effects are discussed in more detail in Chapter 11.

In certain cases, we may wish to operate a storage ring close to a betatron resonance in order to exploit the effects that occur at particular values

of the tunes. For example, consider a storage ring used in an electron–positron collider. At the collision point, the electromagnetic fields around one beam provide focusing forces acting on the other beam. If the tune is just above a half-integer, then the focusing force acts to reduce the beta function at the collision point: this in turn reduces the beam size and, if carefully controlled, leads to an increase in luminosity. A similar effect could be achieved by operating with a tune just above an integer value; however, the closed orbit then becomes very sensitive to steering errors, and it could be difficult to keep the beams in collision. From the lens of the closed orbit, a half-integer tune minimises sensitivity to steering errors. Of course, the drawback with operating close to the half-integer to improve luminosity is that the focusing becomes sensitive to focusing errors at all points in the ring (arising, for example, from errors on the currents in the quadrupoles), and it can be difficult to control the beam size. Optimal operation of a collider for maximum luminosity will depend on precise correction of focusing errors, to allow controlled operation as close as possible to a half-integer value of the tune. Generally, the greatest benefit is obtained by adjusting the vertical tune as close as possible to a half-integer: some separation between the horizontal and vertical tunes is needed to limit sensitivity to coupling errors. Adverse effects of beam collisions (such as the generation of backgrounds in the detector) are mitigated by arranging for the horizontal beam size to be large compared to the vertical beam size. Finally, note that if the tune is adjusted to just *below* a half-integer, then the focusing effects from the beam collisions would lead to an increase in beta function at the interaction point and a reduction in luminosity.

## 6.5 Beam-Based Alignment of Quadrupoles

When constructing a storage ring, positioning of the magnets is usually performed using conventional surveying techniques. However, using these techniques it is often difficult to achieve the alignment precision necessary to meet the performance specifications. Typically, the beam quality is sensitive to changes in position of quadrupole magnets at the level of around 10 µm. To reach the highest levels of performance, the positions of the magnets need to be known as accurately and precisely as possible, preferably to within a few micrometres. A useful step in the process of determining the magnet alignment can be to use *beam-based alignment* (BBA), where the beam itself is used as a surveying tool. BBA refers to a variety of different techniques [Minty and Zimmermann (2003e)]: one commonly used

technique is to measure the response of the closed orbit to variations in quadrupole strength. For a beam that passes through the exact centre of a quadrupole, the closed orbit does not change significantly when the strength of the quadrupole is changed. If the goal is to steer the beam (or position the magnets) so that the closed orbit passes through the centre of every quadrupole, then it is possible to use an empirical approach, in which the closed orbit position in a given quadrupole is adjusted (using small dipole steering magnets) until the closed orbit is unaffected by a change in the quadrupole strength. This is repeated for each and every quadrupole around the storage ring.

More generally, it is useful to be able to estimate the orbit offset in a quadrupole from measurements of the closed orbit as a function of quadrupole strength. In the remainder of this section, we shall derive formulae for the change in closed orbit resulting from a known change in focusing strength of a quadrupole, assuming that the beam has some offset from the centre of the quadrupole [Wolski and Zimmermann (2004)]. We shall only consider motion in the horizontal plane, making the assumption that this motion is uncoupled from both the vertical and the longitudinal motion. In particular, the assumption that there is no coupling between the horizontal and longitudinal motion requires that there is no dispersion at the location of the quadrupole.

Consider a thin quadrupole of focusing strength  $K = 1/f$  (where  $f$  is the focal length). The transfer matrix for this quadrupole, in the thin lens approximation is:

$$R_q(K) = \begin{pmatrix} 1 & 0 \\ -K & 1 \end{pmatrix}. \quad (6.35)$$

Suppose that a particle on the closed orbit passes through the quadrupole a distance  $x_{bq}$  from the centre of the quadrupole, i.e.  $x_{bq}$  is the horizontal offset of the beam (centred on the closed orbit) with respect to the centre of the quadrupole. The effect of the quadrupole on such a particle is to transform the phase space co-ordinates as follows:

$$\begin{pmatrix} x_{co} \\ p_{x,co} \end{pmatrix} \mapsto \begin{pmatrix} x_{co} \\ p_{x,co} - Kx_{bq} \end{pmatrix}. \quad (6.36)$$

Now let  $\hat{R}$  be the transfer matrix from the exit of the quadrupole, around the ring, and back to the entrance of the quadrupole (note that the effect of the quadrupole is not included in  $\hat{R}$ ). Applying  $\hat{R}$  to the phase space co-ordinates of the closed orbit at the exit of the quadrupole must give

the phase space co-ordinates of the closed orbit at the entrance of the quadrupole. Therefore:

$$\hat{R} \begin{pmatrix} x_{\text{co}} \\ p_{x,\text{co}} - K x_{\text{bq}} \end{pmatrix} = \begin{pmatrix} x_{\text{co}} \\ p_{x,\text{co}} \end{pmatrix}. \quad (6.37)$$

Since  $\hat{R}$  is independent of the quadrupole strength, we can write (6.37) for a different quadrupole strength  $K^{(1)} = K + \Delta K$ :

$$\hat{R} \begin{pmatrix} x_{\text{co}}^{(1)} \\ p_{x,\text{co}}^{(1)} - K^{(1)} x_{\text{bq}}^{(1)} \end{pmatrix} = \begin{pmatrix} x_{\text{co}}^{(1)} \\ p_{x,\text{co}}^{(1)} \end{pmatrix}. \quad (6.38)$$

If we subtract (6.37) from (6.38), we find:

$$\begin{pmatrix} \Delta x_{\text{co}} \\ \Delta p_{x,\text{co}} \end{pmatrix} = \hat{R} \begin{pmatrix} \Delta x_{\text{co}} \\ \Delta p_{x,\text{co}} - \Delta K x_{\text{bq}} - K^{(1)} \Delta x_{\text{co}} \end{pmatrix}, \quad (6.39)$$

where  $\Delta x_{\text{co}} = x_{\text{co}}^{(1)} - x_{\text{co}}$ , and  $\Delta p_{x,\text{co}} = p_{x,\text{co}}^{(1)} - p_{x,\text{co}}$ .

The single-turn matrix from the entrance of the quadrupole, through the quadrupole, around the ring and back to the entrance of the quadrupole is:

$$R = \hat{R} R_q(K), \quad (6.40)$$

from which we write:

$$\hat{R} = R R_q(K)^{-1} = R R_q(-K). \quad (6.41)$$

We can write the single-turn transfer matrix  $R$ , in terms of the tune and the Courant–Snyder parameters at the entrance to the quadrupole, in the form (6.28). Then, from (6.35) and (6.41), we can construct explicit expressions for the elements of  $\hat{R}$ . Finally, solving (6.39) for  $\Delta x_{\text{co}}$  and  $\Delta p_{x,\text{co}}$  gives:

$$\Delta x_{\text{co}} = -\frac{x_{\text{bq}} \beta_x \Delta K \cot(\pi \nu_x)}{2 + \beta_x \Delta K \cot(\pi \nu_x)}, \quad (6.42)$$

$$\Delta p_{x,\text{co}} = \frac{x_{\text{bq}} \Delta K (1 + \alpha_x \cot(\pi \nu_x))}{2 + \beta_x \Delta K \cot(\pi \nu_x)}. \quad (6.43)$$

Equations (6.42) and (6.43) provide the formulae we wanted, relating the change in closed orbit in a storage ring to a known change in focusing strength of a quadrupole, in the case that the beam has some offset from the centre of the quadrupole. The values to be used for the Courant–Snyder parameters  $\alpha_x$  and  $\beta_x$  are those at the entrance of the quadrupole when the quadrupole has initial strength  $K$ . Similarly, the value to be used for the tune  $\nu_x$  is the value corresponding to the initial strength of the quadrupole. Although (6.42) and (6.43) give the change in closed orbit only at one location (namely, the entrance of the quadrupole), the change in closed orbit at other locations around the ring can be found by propagating the phase space vector  $(\Delta x_{\text{co}}, \Delta p_{x,\text{co}})$  using the appropriate transfer matrices.

## 6.6 Coupling Errors

Coupling errors result from unwanted skew quadrupole or solenoid fields. The magnetic field in a skew quadrupole, for example, can be written:

$$B_x = -(B\rho)k_1^{(s)}x, \quad (6.44)$$

$$B_y = (B\rho)k_1^{(s)}y, \quad (6.45)$$

$$B_z = 0, \quad (6.46)$$

where  $B\rho = P_0/q$  is the beam rigidity ( $P_0$  is the reference momentum and  $q$  is the charge of a particle), and  $k_1^{(s)}$  is a constant (the normalised skew quadrupole gradient). In the thin lens approximation, a skew quadrupole has the transfer matrix:

$$R_{\text{sq}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -K & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -K & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad (6.47)$$

where  $K = k_1^{(s)}L$ , and  $L$  is the length of the skew quadrupole. The thin lens approximation represents the transfer matrix of the magnet in the limit  $L \rightarrow 0$ , with constant  $K$ . A particle passing through a skew quadrupole experiences a horizontal deflection proportional to its vertical position (with respect to the magnetic axis) and a vertical deflection proportional to its horizontal position. The analysis of Section 6.4 can be repeated to find the effect on the tunes and beta functions of skew quadrupole errors. We do not go through the analysis in detail, but simply note the result that the expression for the tune shift (6.33) generalises if we use the definition of the coupled lattice functions (5.115). In particular, if we consider a transfer matrix of a slightly more general form than (6.47):

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -K_{11} & 1 & -K_{13} & 0 & -K_{15} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -K_{31} & 0 & -K_{33} & 1 & -K_{35} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ -K_{51} & 0 & -K_{53} & 0 & -K_{55} & 1 \end{pmatrix}, \quad (6.48)$$

then the tune shifts resulting from a component with this transfer matrix inserted into the ring at a given location are approximately:

$$\Delta\nu_k \approx \frac{1}{4\pi} \sum_{i,j=1,3,5} \beta_{ij}^k K_{ij}, \quad (6.49)$$

where  $\beta_{ij}^k$  are the elements of the matrices  $B^k$  defined by (5.115), and the approximation is valid for small tune shifts.

A more accurate expression than (6.49) for the tune shifts resulting from a skew quadrupole field in an otherwise uncoupled storage ring can be obtained as follows. We consider the case where the horizontal and vertical motions are both uncoupled from the longitudinal. Then we can write the transverse part of the transfer matrix for one complete turn around the ring, starting from the entrance to a (thin lens) skew quadrupole of integrated strength  $K = k_1^{(s)} L$ , as the matrix product:

$$R = \begin{pmatrix} R_x & 0 \\ 0 & R_y \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -K & 0 \\ 0 & 0 & 1 & 0 \\ -K & 0 & 0 & 1 \end{pmatrix}, \quad (6.50)$$

where:

$$R_x = \begin{pmatrix} \cos(2\pi\nu_x) + \alpha_x \sin(2\pi\nu_x) & \beta_x \sin(2\pi\nu_x) \\ -\gamma_x \sin(2\pi\nu_x) & \cos(2\pi\nu_x) - \alpha_x \sin(2\pi\nu_x) \end{pmatrix}, \quad (6.51)$$

$$R_y = \begin{pmatrix} \cos(2\pi\nu_y) + \alpha_y \sin(2\pi\nu_y) & \beta_y \sin(2\pi\nu_y) \\ -\gamma_y \sin(2\pi\nu_y) & \cos(2\pi\nu_y) - \alpha_y \sin(2\pi\nu_y) \end{pmatrix}. \quad (6.52)$$

The normal mode tunes are obtained from the eigenvalues of  $R$ , which we write as  $e^{\pm 2\pi i\nu_I}$  and  $e^{\pm 2\pi i\nu_{II}}$ . To find explicit expressions for the normal mode tunes  $\nu_I$  and  $\nu_{II}$  involves some complicated algebra; the results can be expressed as:

$$2 \cos(2\pi\nu_I) = \cos(2\pi\nu_x) + \cos(2\pi\nu_y) - \sqrt{(\cos(2\pi\nu_x) - \cos(2\pi\nu_y))^2 + \beta_x \beta_y K^2 \sin(2\pi\nu_x) \sin(2\pi\nu_y)}, \quad (6.53)$$

$$2 \cos(2\pi\nu_{II}) = \cos(2\pi\nu_x) + \cos(2\pi\nu_y) + \sqrt{(\cos(2\pi\nu_x) - \cos(2\pi\nu_y))^2 + \beta_x \beta_y K^2 \sin(2\pi\nu_x) \sin(2\pi\nu_y)}. \quad (6.54)$$

Note that in the limit  $K \rightarrow 0$ ,  $\nu_I \rightarrow \nu_x$  and  $\nu_{II} \rightarrow \nu_y$ . If the ring is tuned so that  $\nu_x = \nu_y = \nu$ , then we see that:

$$\cos(2\pi\nu_I) = \cos(2\pi\nu) + \frac{1}{2} \sqrt{\beta_x \beta_y} |K \sin(2\pi\nu)|, \quad (6.55)$$

$$\cos(2\pi\nu_{II}) = \cos(2\pi\nu) - \frac{1}{2} \sqrt{\beta_x \beta_y} |K \sin(2\pi\nu)|. \quad (6.56)$$

Although the ring in the absence of coupling is tuned so that the betatron tunes are the same, the presence of a skew quadrupole field leads to a split in the normal mode tunes. For weak coupling, the difference between the normal mode tunes is small, and we find:

$$\Delta\nu_I \approx \frac{1}{4\pi} \sqrt{\beta_x \beta_y} |K|, \quad (6.57)$$

$$\Delta\nu_{II} \approx -\frac{1}{4\pi} \sqrt{\beta_x \beta_y} |K|. \quad (6.58)$$

By making some approximations to (6.53) and (6.54), we can see more clearly how a skew quadrupole error affects the tunes. First, we write the tunes in the absence of coupling in terms of a small deviation  $\Delta\nu$  from the mean  $\nu_0$  of the horizontal and vertical tunes:

$$\nu_x = \nu_0 + \frac{1}{2} \Delta\nu, \quad (6.59)$$

$$\nu_y = \nu_0 - \frac{1}{2} \Delta\nu. \quad (6.60)$$

Similarly, we write the tunes in the presence of the skew quadrupole:

$$\nu_I = \bar{\nu}_0 + \frac{1}{2} \Delta\bar{\nu}, \quad (6.61)$$

$$\nu_{II} = \bar{\nu}_0 - \frac{1}{2} \Delta\bar{\nu}. \quad (6.62)$$

Substituting in (6.53) and using some standard trigonometric identities yields:

$$\begin{aligned} \cos(2\pi\bar{\nu}_0) \cos(\pi\Delta\bar{\nu}) - \sin(2\pi\bar{\nu}_0) \sin(\pi\Delta\bar{\nu}) &= \cos(2\pi\nu_0) \cos(\pi\Delta\nu) - \\ \sqrt{\sin^2(2\pi\nu_0) \sin^2(\pi\Delta\nu) + \frac{\beta_x \beta_y}{4} K^2} \sin(2\pi\nu_0 + \pi\Delta\nu) \sin(2\pi\nu_0 - \pi\Delta\nu). \end{aligned} \quad (6.63)$$

Assuming that  $K$  and  $\Delta\nu$  are small, then:

$$K^2 \sin(2\pi\nu_0 + \pi\Delta\nu) \sin(2\pi\nu_0 - \pi\Delta\nu) = K^2 \sin^2(2\pi\nu_0) + O(4), \quad (6.64)$$

where  $O(4)$  represents terms of fourth order and higher in small quantities. Assuming we can drop the higher-order terms, (6.63) becomes:

$$\begin{aligned} \cos(2\pi\bar{\nu}_0) \cos(\pi\Delta\bar{\nu}) - \sin(2\pi\bar{\nu}_0) \sin(\pi\Delta\bar{\nu}) &\approx \\ \cos(2\pi\nu_0) \cos(\pi\Delta\nu) - \sin(2\pi\nu_0) \sqrt{\sin^2(\pi\Delta\nu) + \frac{\beta_x \beta_y}{4} K^2}. \end{aligned} \quad (6.65)$$

Finally, comparing terms on each side of (6.65), it appears that:

$$\bar{\nu}_0 \approx \nu_0, \quad (6.66)$$

and:

$$\Delta\bar{\nu} \approx \sqrt{\Delta\nu^2 + \frac{\beta_x\beta_y}{4\pi^2}K^2}. \quad (6.67)$$

Hence, for  $|\nu_x - \nu_y| \ll 1$  and  $|K| \ll 1$ , (6.53) can be written in the simpler (approximate) form:

$$\nu_I \approx \frac{1}{2}(\nu_x + \nu_y) + \frac{1}{2}\sqrt{(\nu_x - \nu_y)^2 + \frac{\beta_x\beta_y}{4\pi^2}K^2}. \quad (6.68)$$

Similarly:

$$\nu_{II} \approx \frac{1}{2}(\nu_x + \nu_y) - \frac{1}{2}\sqrt{(\nu_x - \nu_y)^2 + \frac{\beta_x\beta_y}{4\pi^2}K^2}. \quad (6.69)$$

From these expressions, it is clear that if the normal mode tunes are plotted as a function of the uncoupled tune difference  $\nu_x - \nu_y$ , the curves appear as hyperbolae, with the normal mode tunes approaching the uncoupled tunes for  $|\nu_x - \nu_y| \gg \sqrt{\beta_x\beta_y}|K|/2\pi$ : see Fig. 6.2. The minimum separation of the normal mode tunes occurs for  $\nu_x = \nu_y$ , and is given by:

$$\nu_I - \nu_{II} = \frac{\sqrt{\beta_x\beta_y}}{2\pi}|K|. \quad (6.70)$$

The tunes in a storage ring are easily controlled by adjusting quadrupole strengths, and the normal mode tunes can be measured by resonant excitation. Thus, experimental data can readily be collected for constructing plots such as those shown in Fig. 6.2, from which a measure of the coupling strength in the storage ring can be obtained [Minty and Zimmermann (2003d)]. Although we have dealt here with the case of a single skew quadrupole error, in general there will be many such errors present in a machine. The generalisation of the above formulae, in particular (6.68) and (6.69), to the case of multiple skew quadrupole errors distributed around a storage ring is discussed in Chapter 11 (Section 11.2.3). The impact of skew quadrupoles on the equilibrium emittances in an electron storage ring (taking synchrotron radiation effects into account) is considered in Chapter 7 (Section 7.4.1).

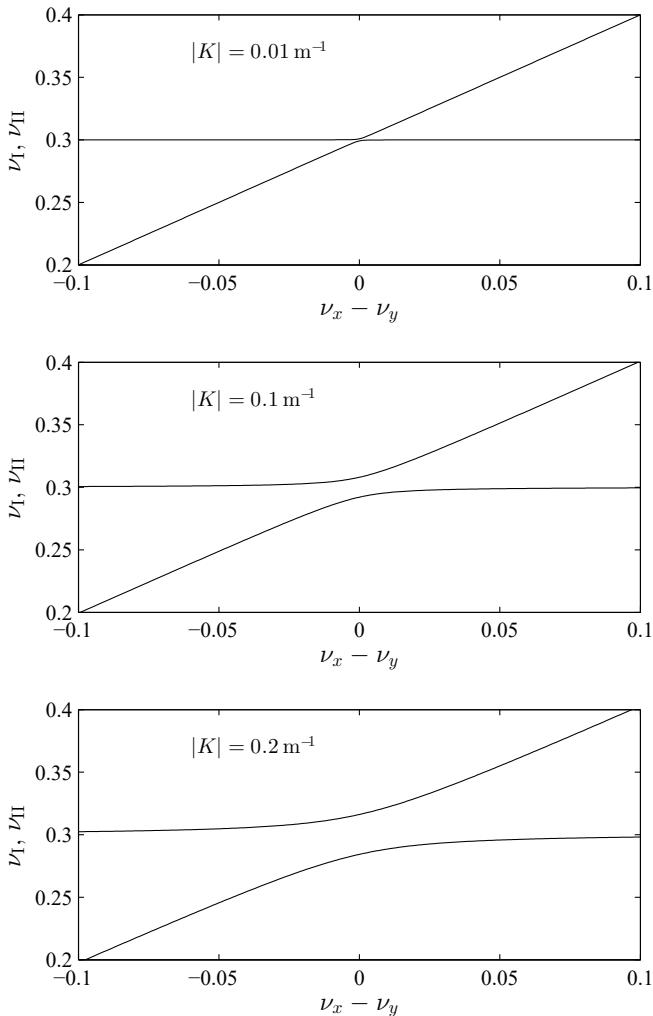


Fig. 6.2 Normal mode betatron tunes in a storage ring in the presence of a skew quadrupole, as a function of the tune difference  $\nu_x - \nu_y$  in the storage ring without the skew quadrupole. The vertical tune is fixed at  $\nu_y = 0.3$  while the horizontal tune is varied over the range  $0.2 < \nu_x < 0.4$ . The beta functions at the skew quadrupole location (in the absence of the skew quadrupole) are  $\beta_x = \beta_y = 1 \text{ m}$ . The three plots show the variation in the normal mode tunes with different values of the integrated normalised skew quadrupole strength,  $0.01 \text{ m}^{-1}$  (top),  $0.1 \text{ m}^{-1}$  (middle) and  $0.2 \text{ m}^{-1}$  (bottom).

## Chapter 7

# Effects of Synchrotron Radiation

The development of the first large accelerators was motivated mainly by the aim of performing experiments in particle and nuclear physics, and little thought was given to possible uses for the electromagnetic radiation generated by beams of charged particles moving through magnetic fields. However, even as the first machines were being built, it was soon realised that synchrotron radiation was a potential limitation on the energies that could be achieved by electrons or positrons in storage rings, and this prompted detailed studies on the production of synchrotron radiation [Schwinger (1949)]. It was soon found that synchrotron radiation had useful properties in a wide range of scientific fields. This led to the development of accelerators optimised for the production of synchrotron radiation, and there are now far more synchrotron light sources in the world than high-energy colliders.

‘Synchrotron radiation’ is a general term for electromagnetic radiation produced by a relativistic charged particle. Classical electromagnetic theory shows that any charged particle will generate radiation when it undergoes acceleration: this could be in a linac, for example, or in a dipole magnet. However, the rate at which electromagnetic energy is radiated scales with the square of the rate of acceleration. In the case of a point-like particle of charge  $q$  moving with relativistic velocity  $\mathbf{v}$ , the total radiation power is given by Liénard’s formula (a relativistic generalisation of Larmor’s formula):

$$P_\gamma = \frac{q^2}{6\pi\varepsilon_0 c} \gamma_0^6 (\dot{\beta}^2 - (\boldsymbol{\beta} \times \dot{\boldsymbol{\beta}})^2), \quad (7.1)$$

where  $\boldsymbol{\beta} = \mathbf{v}/c$ ,  $\gamma_0 = (1 - \beta^2)^{-1/2}$  is the relativistic factor, and  $\varepsilon_0$  is the permittivity of free space. The dot indicates a derivative with respect to time, so  $\dot{\boldsymbol{\beta}}$  is the acceleration of the particle divided by the speed of light.

A relativistic particle in a linac can experience only a small amount of acceleration because of the limit on the velocity; the effects of synchrotron radiation in a linac can usually be ignored. On the other hand, in a dipole magnet, the acceleration corresponding to a change in the direction of the velocity can be very large. Synchrotron radiation from dipole magnets in electron storage rings has significant effects on the beam properties and behaviour.

There are two aspects to studies of radiation produced by particles in accelerators: the study of the properties of the radiation itself, and the study of the impact of the production of the radiation on the particles in the accelerator. In this chapter, we will focus on the effects of synchrotron radiation on the particles generating the radiation; detailed discussion of the properties of the radiation can be found elsewhere, for example [Hofmann (2004); Clarke (2004)]. Also, we shall consider only *incoherent synchrotron radiation*, where each particle radiates independently so that the total radiation from the beam has the same properties as the radiation from a single particle, but with an intensity that scales linearly with the number of particles in the beam. This is the usual regime for radiation from dipoles and many insertion devices. However, the radiation power from a point-like particle, expressed in equation (7.1), varies with the square of the charge of the particle. In the case of a bunch of many particles, if the bunch is small compared to the wavelength of the emitted radiation, then the bunch behaves (from the point of view of the radiation) as a single particle: the radiation is emitted coherently [Schiff (1946); Wiedemann (1999)]. One of the characteristic features of coherent radiation is that the intensity is proportional to the square of the bunch charge, rather than simply being proportional to the charge. In an accelerator, the number of particles within a bunch can be very large, so the intensity of coherent synchrotron radiation is potentially many orders of magnitude larger than the intensity of the incoherent radiation, even if only a fraction of the total number of particles within a given bunch radiate coherently. Coherent synchrotron radiation may appear as an undesirable effect having an impact on beam quality [Stupakov and Heifets (2002); Byrd *et al.* (2002)]; or it may be generated intentionally, as in a free-electron laser [Madey (1971); Deacon *et al.* (1977); Freund and Antonsen, Jr (1996); Saldin *et al.* (2008); Schmüser *et al.* (2008)].

Our discussion in this chapter will focus on the effects of the radiation on particles in storage rings, although many of the results can be applied in other kinds of accelerator. We shall assume that the radiation has relatively

weak effects on the particle motion: that is, the effects of the radiation can be treated as a perturbation to the motion. This is often a valid assumption, but there are parameter regimes (particularly, when there is a very high peak current in the beam) where the radiation has strong effects on the beam; this includes the cases of coherent synchrotron radiation and free-electron lasers.

We shall begin the discussion by treating radiation as a classical phenomenon, ignoring the fact that electromagnetic energy is emitted from the beam in discrete quanta (photons). This will lead to an understanding of radiation damping, by which the beam emittances gradually reduce over some number of turns in a storage ring. However, quantum effects cannot be ignored, when accounting for them, we find that the emission of photons leads to an excitation of the beam emittances, which ultimately results in equilibrium emittances being achieved through a balance between the damping and excitation. The values of the emittances at equilibrium in a storage ring are determined by a number of factors, including in particular the beam energy and lattice design. Magnet alignment and focusing errors can be important as well, especially for the vertical emittance. The calculation of the equilibrium emittances in a given lattice can be quite complicated, but some simple cases provide useful approximations, and gives some insight into features of the lattice that are important for determining the emittances.

## 7.1 Classical Radiation: Radiation Damping

To keep the discussion reasonably concise, we shall quote some results for the properties of radiation from relativistic particles. The first result is that the power radiated by a relativistic particle is generated predominantly in the direction of the velocity of the particle. The radiation flux as a function of angle with respect to the particle velocity has a width of order  $1/\gamma$ , where  $\gamma$  is the relativistic factor of the particle [Jackson (1998)]. Although this ‘opening angle’ of the radiation does have some consequences, we shall for the moment assume that we are dealing with ultra-relativistic particles with large  $\gamma$ , so we may neglect the opening angle altogether, and treat the radiation as being emitted in the direction of the instantaneous velocity of the particle. We shall show that the radiation leads to the reduction in the amplitude of betatron and synchrotron oscillations of particles in a storage ring, and shall calculate the damping times for these oscillations.

Consider a particle with initial momentum  $P$ . Suppose that the

particle emits radiation (for example, by travelling a short distance through a magnetic field) carrying momentum  $dP$ , so that the particle has final momentum  $P_1$ . Note that here we use an upper case  $P$  to indicate the canonical momentum *not* scaled by a reference momentum; i.e.  $P = |\beta\gamma mc + q\mathbf{A}|$ . If the momentum carried by the radiation is in the same direction as the momentum of the particle, then:

$$P_1 = P - dP \approx P \left(1 - \frac{dP}{P_0}\right). \quad (7.2)$$

$P_0$  is the reference momentum, and the approximation is valid for  $P \approx P_0$  (i.e. the energy deviation has magnitude  $|\delta| \ll 1$ ). By assuming that the momentum of the radiation is in the same direction as the momentum of the particle, we are in effect assuming that the momentum of the particle is in the same direction as its velocity. For a particle in a magnetic field, the contribution of the vector potential to the momentum will mean that, in general, the momentum will not be parallel to the velocity of the particle. However, in dipoles or other multipole magnets, we can work in a gauge in which the only non-zero component of the vector potential is parallel to the reference trajectory; and then, for a relativistic particle with small transverse momentum, the canonical momentum will be very close to the mechanical momentum.

Let us now consider specifically the vertical motion of the particle. Since the momentum of the radiation and the path in which the particle is travelling are parallel, there will be no change in the direction of the particle resulting from the radiation emission. Therefore, the vertical momentum is reduced by the same factor as the total momentum:

$$p_{y1} \approx p_y \left(1 - \frac{dP}{P_0}\right), \quad (7.3)$$

where  $p_y$  is the initial vertical momentum scaled by the reference momentum (i.e.  $p_y = P_y/P_0$ ), and  $p_{y1}$  is the final vertical momentum (again scaled by  $P_0$ ). Substituting (7.3) into the expression for the action (see (4.34)):

$$J_y = \frac{1}{2}(\gamma_y y^2 + 2\alpha_y y p_y + \beta_y p_y^2), \quad (7.4)$$

we find the change in the action resulting from the emission of the radiation:

$$J_{y1} = J_y + dJ_y, \quad (7.5)$$

where:

$$dJ_y = -(\alpha_y y p_y + \beta_y p_y^2) \frac{dP}{P_0}. \quad (7.6)$$

In writing (7.4) for the vertical action, we have assumed that the vertical motion is not coupled either to the horizontal motion or to the longitudinal motion: that is, that there is no betatron coupling or vertical dispersion. Effects of betatron coupling and vertical dispersion will be considered later, in Sections 7.4.1 and 7.4.2.

The vertical emittance is the average of the vertical action over all the particles in the beam, so we can find the change in vertical emittance resulting from the radiation from all particles in the beam by averaging the change in the vertical action (7.6) over all particles in the beam. Using versions of equations (4.99), (4.100) and (4.19) for the vertical motion, we find from (7.6):

$$d\epsilon_y = \langle dJ_y \rangle = -\epsilon_y \frac{dP}{P_0}. \quad (7.7)$$

For a particle moving around a storage ring, we can integrate the loss in momentum around the ring to find the total change in momentum over one turn. The emittance is conserved under symplectic transport, and radiation is the only non-symplectic effect we are considering. If the radiation effects are weak so that the change in emittance of the beam in one revolution is small, then from (7.7) we can write:

$$\frac{d\epsilon_y}{dt} = -\frac{\epsilon_y}{T_0} \oint \frac{dP}{P_0}, \quad (7.8)$$

where the integral is taken over one complete revolution, and  $T_0$  is the revolution period.

For convenience, the rate of change of emittance is usually written in terms of the rate of energy loss, rather than the rate of momentum loss. For a particle with velocity  $\beta_0 c$ ; the energy  $E_0$  and momentum  $P_0$  are related by:

$$P_0 = \frac{\beta_0}{c} E_0. \quad (7.9)$$

If the particle is ultra-relativistic, we can neglect the change in velocity associated with a change in momentum, so that:

$$\frac{dP}{P_0} \approx \frac{dE}{E_0}. \quad (7.10)$$

Then, the rate of change of the emittance (7.8) becomes:

$$\frac{d\epsilon_y}{dt} \approx -\frac{U_0}{E_0 T_0} \epsilon_y, \quad (7.11)$$

where  $U_0$  is the energy loss per turn of a particle with the reference momentum following the reference trajectory. From (7.11) we see that the emittance decays (or *damps*) exponentially:

$$\epsilon_y(t) = \epsilon_y(0)e^{-2t/\tau_y}, \quad (7.12)$$

where the damping time  $\tau_y$  is defined by:

$$\tau_y = 2 \frac{E_0}{U_0} T_0. \quad (7.13)$$

Strictly speaking, if we choose to define the damping time so that (7.13) is an exact equality, then (7.12) is an approximation. However, for ultrarelativistic beams the approximations involved are often very good in practice, and any deviations from the expected decay rate are difficult to observe (or at least, to distinguish from other effects that may affect the emittance).

The factor 2 appearing in the definition of the damping time (7.13) is a matter of convention. Since the action variable is proportional to the square of the amplitude of a betatron oscillation, the damping time defined by (7.13) actually gives the exponential decay rate of the betatron amplitude of a particle; the exponential decay rate of the emittance is  $\tau_y/2$ .

From (7.13), we see that the damping time  $\tau_y$  is twice the length of time it would take a particle to lose all its energy (starting from the reference energy  $E_0$ ), if it continued to lose energy at a constant rate. However, while this observation may provide a convenient way to remember the formula (7.13), it actually has limited physical significance. We assumed in deriving the formula (7.13) that the particle momentum  $P$  was close to the reference momentum  $P_0$ : if a particle continues to radiate energy without the energy being restored, eventually this assumption will no longer be valid.

Particle trajectories within storage rings are usually stable for only a small range of particle energies, because of the limitation of the rf acceptance, and because of chromaticity and other nonlinear effects. A particle that produces significant amounts of synchrotron radiation, without the energy being restored, will therefore soon be lost from the storage ring. In electron storage rings, the energy lost through synchrotron radiation is restored by rf cavities. We have seen that the energy lost through synchrotron radiation leads to damping of the amplitude of betatron oscillations: we should therefore consider the effect of the energy gain from rf cavities on betatron oscillations. In most cases, the rf cavities are designed so that a particle sees a longitudinal electric field as it passes through the cavity. There is then no change in the transverse (and in particular, the vertical) momentum of the particle, and therefore no change in the betatron action.

This is consistent with the analysis in Chapter 3 of the dynamics of a particle in an rf cavity: there, we found that the dynamics could be represented by a symplectic transfer matrix. We saw in Chapter 5 that the emittances of a bunch of particles are conserved under linear symplectic transport.

To complete the calculation of the vertical damping time  $\tau_y$ , we need to find an expression for  $U_0$ , the energy lost by a particle through synchrotron radiation on each complete turn of the storage ring. The total power radiated by a relativistic, accelerated particle of charge  $q$  is given by Liénard's formula (7.1). In the case of a particle with constant speed  $\beta_0 c$  following a circular trajectory of radius  $\rho$ , the radiation power becomes [Jackson (1998)]:

$$P_\gamma = \frac{q^2 c}{6\pi\varepsilon_0} \frac{\beta_0^4 \gamma_0^4}{\rho^2}. \quad (7.14)$$

Equation (7.14) can be written in the form:

$$P_\gamma = \frac{C_\gamma c}{2\pi} \frac{\beta_0^4 E_0^4}{\rho^2}, \quad (7.15)$$

where  $C_\gamma$  [Sands (1970)] is a constant given by:

$$C_\gamma = \frac{q^2}{3\varepsilon_0(mc^2)^4}, \quad (7.16)$$

with  $m$  the mass of the particle. We can find the energy lost in one revolution by integrating the synchrotron radiation power over one period:

$$U_0 = \oint P_\gamma dt = \oint P_\gamma \frac{ds}{\beta_0 c}, \quad (7.17)$$

where we assume that the particle (with reference momentum  $P_0$ ) follows the reference trajectory. Using the expression (7.15) for  $P_\gamma$ , we find:

$$U_0 = \frac{C_\gamma}{2\pi} \beta_0^3 E_0^4 \oint \frac{1}{\rho^2} ds. \quad (7.18)$$

Following convention, we define the *second synchrotron radiation integral*  $I_2$ :

$$I_2 = \oint \frac{1}{\rho^2} ds, \quad (7.19)$$

where the integral is taken along the reference trajectory over one complete turn of the ring. In terms of  $I_2$ , the energy loss per turn for a particle with energy  $E_0$  travelling along the reference trajectory can be written:

$$U_0 = \frac{C_\gamma}{2\pi} \beta_0^3 E_0^4 I_2. \quad (7.20)$$

$I_2$  is conveniently given in units of metres<sup>-1</sup>; therefore, if  $E_0$  is given in GeV, using a value of  $C_\gamma$  in m GeV<sup>-3</sup> gives the energy loss per turn  $U_0$  in GeV. For electrons,  $C_\gamma \approx 8.846 \times 10^{-5}$  m GeV<sup>-3</sup>.

$I_2$  is a property only of the reference trajectory, and does not depend on any beam properties. Using this fact, we see from (7.20) that for ultra-relativistic particles ( $\beta_0 \approx 1$ ) the energy loss per turn  $U_0$  scales as the fourth power of the energy of the particles. In an isomagnetic ring, where the radius of curvature of the reference trajectory is constant around the entire circumference, we can write:

$$I_2 = \frac{1}{\rho} \oint \frac{ds}{\rho} = \frac{2\pi}{\rho}, \quad (7.21)$$

and then in this particular case:

$$U_0 = C_\gamma \beta_0^3 \frac{E_0^4}{\rho}. \quad (7.22)$$

The energy loss per turn is inversely proportional to the radius of the ring; this general scaling holds true even when the ring is not perfectly isomagnetic. For high energy storage rings, the synchrotron radiation power can become a significant technical issue, because of the power load on components in the vacuum chamber, and because of the cost of the rf system needed to replace the power lost through synchrotron radiation. The power losses can be reduced by increasing the size of the ring; however, the radiation power scales much more strongly with energy than with the radius of the ring. If the energy is doubled, in order to maintain the energy lost per turn at a fixed value, the circumference of the ring must be increased by a factor of 16. Ultimately, this places a practical limit on the beam energy that can be achieved in an electron storage ring. For protons, which have a much larger mass than electrons (and hence a much smaller value for  $C_\gamma$ ), synchrotron radiation energy losses are rarely of much significance.

Let us now consider the effects of synchrotron radiation on the horizontal emittance. We can proceed along much the same lines as we followed for the analysis in the case of the vertical emittance, except there are three additional complications that need to be dealt with now. First, the horizontal motion is often strongly coupled to the longitudinal motion: this coupling is characterised by the dispersion, and must be taken into account when considering the change in horizontal motion arising from a change in energy of a particle. Second, where the reference trajectory is curved (as is usually the case in a dipole magnet in a storage ring) the path length taken by a particle depends on the horizontal co-ordinate of the particle with respect

to the reference trajectory. Third, dipole magnets are sometimes built with a gradient in the magnetic field, which means that the vertical field seen by a particle in a dipole depends on the horizontal co-ordinate of the particle. Fortunately, it is not too difficult to take all these effects into account.

To begin the analysis, we write the horizontal action in the presence of dispersion as follows:

$$J_x = \frac{1}{2} (\gamma_x \bar{x}^2 + 2\alpha_x \bar{x} \bar{p}_x + \beta_x \bar{p}_x^2), \quad (7.23)$$

where:

$$\bar{x} = x - \eta_x \delta_p, \quad (7.24)$$

$$\bar{p}_x = p_x - \eta_{px} \delta_p. \quad (7.25)$$

$\eta_x$  and  $\eta_{px}$  are the dispersion, describing the change in closed orbit in the ring with respect to changes in the momentum deviation  $\delta_p$ , defined by (5.9).  $\bar{x}$  and  $\bar{p}_x$  are the horizontal co-ordinate and horizontal momentum (scaled by the reference momentum) with respect to the off-energy closed orbit: a particle with  $\bar{x} = \bar{p}_x = 0$  is following the closed orbit for its particular energy deviation, and the horizontal action  $J_x$  is zero. Suppose that the particle emits synchrotron radiation carrying momentum  $dP$ . Assuming that the momentum of the radiation is in the same direction as the momentum  $P$  of the particle, the horizontal co-ordinate and horizontal momentum after the emission of the synchrotron radiation are:

$$\bar{x}_1 \approx \bar{x} + \eta_x \frac{dP}{P_0}, \quad (7.26)$$

$$\bar{p}_{x1} \approx \bar{p}_x \left( 1 - \frac{dP}{P_0} \right) + \eta_{px} \frac{dP}{P_0}. \quad (7.27)$$

We can write the resulting change in the action as:

$$J_{x1} = J_x + dJ_x, \quad (7.28)$$

where:

$$dJ_x = -w_1 \frac{dP}{P_0} + w_2 \left( \frac{dP}{P_0} \right)^2, \quad (7.29)$$

and  $w_1$  and  $w_2$  are given by:

$$w_1 = \alpha_x \bar{x} \bar{p}_x + \beta_x \bar{p}_x^2 - \eta_x (\gamma_x \bar{x} + \alpha_x \bar{p}_x) - \eta_{px} (\alpha_x \bar{x} + \beta_x \bar{p}_x), \quad (7.30)$$

$$w_2 = \frac{1}{2} (\gamma_x \eta_x^2 + 2\alpha_x \eta_x \eta_{px} + \beta_x \eta_{px}^2) - (\alpha_x \eta_x + \beta_x \eta_{px}) \bar{p}_x + \frac{1}{2} \beta_x \bar{p}_x^2. \quad (7.31)$$

From (7.29), the rate of change of the horizontal action is:

$$\frac{dJ_x}{dt} = -w_1 \frac{1}{P_0} \frac{dP}{dt} + w_2 \frac{dP}{P_0^2} \frac{dP}{dt}. \quad (7.32)$$

If we treat radiation as a classical phenomenon, then in the limit  $dt \rightarrow 0$ , the radiation emitted in the time interval  $dt$  also tends towards zero, i.e.  $dP \rightarrow 0$ . This means that in (7.32) for the rate of change of the action, the second term on the right-hand side vanishes. In a classical model, we do not need the expression (7.31) for  $w_2$ ; however, in a quantum model, where radiation is emitted in discrete units (photons), we cannot take the limit  $dP \rightarrow 0$ , and the term in  $w_2$  for the rate of change of the action must be retained. We shall return to this point in Section 7.2; however, for now we shall use a classical model, in which case:

$$\frac{dJ_x}{dt} = -w_1 \frac{1}{P_0} \frac{dP}{dt} \approx -w_1 \frac{P_\gamma}{E_0}. \quad (7.33)$$

Since  $w_1$  is a function of the Courant–Snyder parameters, the rate of change of the action varies around the ring. To find the time-average rate of change of the action, we integrate over one revolution:

$$\left\langle \frac{dJ_x}{dt} \right\rangle_t = -\frac{1}{T_0} \oint w_1 \frac{P_\gamma}{E_0} dt. \quad (7.34)$$

In a time interval  $dt$ , a particle travels distance  $dC = \beta_0 c dt$ . When we considered the vertical motion, we assumed that the reference trajectory had no vertical curvature, and the particle had no horizontal offset from the reference trajectory; hence, we wrote  $dC = ds$ , where  $ds$  is the distance travelled by a reference particle along the reference trajectory in time  $dt$ . However, if the particle has a horizontal offset from the reference trajectory, then (as shown in Fig. 5.1):

$$dC = \left(1 + \frac{x}{\rho}\right) ds, \quad (7.35)$$

where  $x$  is the horizontal offset (i.e. the horizontal co-ordinate) and  $\rho$  is the curvature of the reference trajectory. Therefore, the average rate of change of the horizontal action becomes:

$$\left\langle \frac{dJ_x}{dt} \right\rangle_t = -\frac{1}{T_0 E_0} \oint w_1 P_\gamma \left(1 + \frac{x}{\rho}\right) \frac{ds}{\beta_0 c}. \quad (7.36)$$

To proceed, we substitute the expression (7.14) for the synchrotron radiation power  $P_\gamma$  into (7.36). The presence of a gradient in the field of a dipole magnet can be taken into account by writing:

$$B = B_0 + x \frac{\partial B_y}{\partial x}. \quad (7.37)$$

When we considered the vertical motion, we neglected any change in the field with respect to the vertical co-ordinate, even though the presence of a gradient  $\partial B_y / \partial x$  implies (by Maxwell's equations) the presence of a variation  $\partial B_x / \partial y$ . A *normal* quadrupole component leads to the introduction of a horizontal field component for particles with non-zero vertical co-ordinates; if the main dipole field is vertical (to give horizontal bending), then the change in the field magnitude with vertical position is second order in the vertical co-ordinate, and we assume that (for vertical betatron oscillations with small amplitude) this change in field magnitude can be neglected. The situation is different if there is a *skew* quadrupole component present in the dipole: however, we do not consider this case. A normal quadrupole component gives a variation in field magnitude that is first order in the horizontal co-ordinate, and this cannot be neglected when considering the horizontal motion. Using (7.14), (7.37) and (7.30), and averaging over all particles in the beam, (after some algebra) we find that:

$$\oint \left\langle w_1 P_\gamma \left( 1 + \frac{x}{\rho} \right) \right\rangle \frac{ds}{\beta_0 c} \approx U_0 \left( 1 - \frac{I_4}{I_2} \right) \epsilon_x, \quad (7.38)$$

where  $U_0$  is the energy loss per turn given by (7.20),  $\epsilon_x = \langle J_x \rangle$  is the horizontal emittance,  $I_2$  is the second synchrotron radiation integral (given by (7.19)), and  $I_4$  is the fourth synchrotron radiation integral, given by:

$$I_4 = \oint \frac{\eta_x}{\rho} \left( \frac{1}{\rho^2} + 2k_1 \right) ds. \quad (7.39)$$

$k_1$  is the quadrupole gradient in the dipole field, scaled by the reference momentum:

$$k_1 = \frac{q}{P_0} \frac{\partial B_y}{\partial x}. \quad (7.40)$$

Finally, from (7.36) and (7.38), and assuming that the radiation effects are weak so that the change in action of any particle over one turn is small, for the rate of change of the emittance we can write:

$$\frac{d\epsilon_x}{dt} \approx -\frac{\tau_x}{2} \epsilon_x, \quad (7.41)$$

where the *horizontal damping time*  $\tau_x$  is defined:

$$\tau_x = \frac{2}{j_x} \frac{E_0}{U_0} T_0, \quad (7.42)$$

and the *horizontal damping partition number*  $j_x$  is defined:

$$j_x = 1 - \frac{I_4}{I_2}. \quad (7.43)$$

Note that the horizontal damping partition number is written in lower case ( $j_x$ ), to distinguish it from the horizontal action, which is written in upper case ( $J_x$ ).

We see from (7.41) that the horizontal emittance decays exponentially with time in the same way as the vertical emittance. The expression for the horizontal damping time (7.42) is also similar to the expression for the vertical damping time (7.13); the main difference is the presence of the horizontal damping partition number  $j_x$ . In the absence of dispersion,  $j_x = 1$ . In a storage ring, of course, dispersion is an inevitable consequence of steering resulting from the dipoles: the presence of dispersion has an effect on the horizontal synchrotron radiation damping time. The value of  $j_x$  also depends on the quadrupole gradient in the dipoles. In many cases, if there is no quadrupole gradient in the dipoles, then  $I_4 \ll I_2$ , and  $j_x \approx 1$ ; but introducing a quadrupole gradient in the dipoles can lead to a significant change in the damping partition number.

Having considered the effects of (classical) synchrotron radiation on the horizontal and vertical motion of particles in a storage ring, we now turn to the effects on the longitudinal motion. The longitudinal dynamics in a storage ring have been discussed in Section 5.1.3: there, we included synchrotron radiation in the approximation that all particles lost energy at the same rate. Our goal now is to include a more accurate description of the synchrotron radiation, taking account of the dependence of the energy loss on the energy of the particle. We shall find that this leads to damping of the amplitude of synchrotron oscillations.

In principle, we could take a similar approach to studying the longitudinal motion as we did for the vertical and horizontal motion. However, the longitudinal motion in a storage ring differs significantly from the transverse motion, in that the synchrotron oscillations are relatively slow compared to the revolution period (i.e. the synchrotron tune is much less than one), whereas the transverse oscillations are relatively fast (the betatron tunes are much larger than one). It is convenient, therefore, to take a rather different approach to the longitudinal motion than to the transverse. Following along the same lines as the arguments we used in Section 5.1.3, we shall first write down the equations of motion for the longitudinal variables  $z$  and  $\delta$  for a particle performing synchrotron oscillations, including the radiation energy loss. Then, we shall express the energy loss per turn as a function of the energy deviation  $\delta$ : this was the effect that we neglected previously, but now we shall find that it leads to a damping term in the equations of motion. Finally, solving the equations of motion gives synchrotron oscillations with an amplitude that decays exponentially.

Assuming that the synchrotron motion is slow compared to the revolution frequency, the equations of motion for the longitudinal dynamics in a synchrotron storage ring can be written as (5.41) and (5.43):

$$\frac{dz}{ds} = -\eta_p \delta, \quad (7.44)$$

$$\frac{d\delta}{ds} = \frac{qV_{\text{rf}}}{cP_0C_0} \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}}z}{c}\right) - \frac{U}{cP_0C_0}. \quad (7.45)$$

Here,  $z$  is the longitudinal co-ordinate and  $\delta$  the energy deviation;  $\eta_p$  is the phase slip factor (5.24);  $P_0$  is the reference momentum;  $C_0$  is the length of the reference trajectory over one complete turn around the ring;  $V_{\text{rf}}$  is the amplitude of the rf voltage;  $\omega_{\text{rf}}$  is the (angular) rf frequency;  $\phi_{\text{rf}}$  is a constant phase; and  $U$  is the energy lost by the particle on each turn, through synchrotron radiation. In Section 5.1.3 we assumed that the energy loss per turn was independent of the energy deviation: we shall now use a more detailed model, in which we take into account the variation of the synchrotron radiation power with the particle energy.

Now we make three assumptions. First, we assume that the longitudinal co-ordinate  $z$  is small compared to the rf wavelength, that is  $\omega_{\text{rf}}z/c \ll 1$ . This means we can make the approximation:

$$\sin\left(\phi_s - \frac{\omega_{\text{rf}}z}{c}\right) \approx \sin(\phi_s) - \cos(\phi_s) \frac{\omega_{\text{rf}}z}{c}. \quad (7.46)$$

Second, we assume that the energy deviation  $\delta$  is small, so that we can write the energy loss per turn to first order in  $\delta$ :

$$U \approx U_0 + \Delta E \left. \frac{dU}{dE} \right|_{E=E_0} \approx U_0 + E_0 \delta \left. \frac{dU}{dE} \right|_{E=E_0}. \quad (7.47)$$

Finally, we assume that the phase  $\phi_{\text{rf}}$  is such that for a particle with  $z = \delta = 0$ , the energy gain over one revolution from the rf cavities exactly balances the energy lost by synchrotron radiation, that is:

$$\phi_{\text{rf}} = \phi_s, \quad (7.48)$$

where the *synchronous phase*  $\phi_s$  satisfies:

$$qV_{\text{rf}} \sin(\phi_s) = U_0. \quad (7.49)$$

Using (7.46), (7.47) and (7.49) in (7.44) and (7.45), the equation of motion (7.45) for the energy deviation becomes:

$$\frac{d\delta}{ds} = -\frac{qV_{\text{rf}} \cos(\phi_s)}{cP_0C_0} \frac{\omega_{\text{rf}}z}{c} - \frac{E_0}{cP_0C_0} \delta \left. \frac{dU}{dE} \right|_{E=E_0}. \quad (7.50)$$

Taking the derivative of (7.50) with respect to  $s$ , and substituting for  $dz/ds$  from (7.44) gives a second-order equation of motion for  $\delta$ :

$$\frac{d^2\delta}{ds^2} + \left( \frac{E_0}{cP_0C_0} \frac{dU}{dE} \Big|_{E=E_0} \right) \frac{d\delta}{ds} - \frac{qV_{\text{rf}} \cos(\phi_s)}{cP_0C_0} \frac{\omega_{\text{rf}}}{c} \eta_p \delta = 0. \quad (7.51)$$

Multiplying by  $\beta_0^2 c^2$ , and changing the derivatives from derivatives with respect to path length  $s$  to derivatives with respect to time  $t$  gives:

$$\frac{d^2\delta}{dt^2} + 2\alpha_E \frac{d\delta}{dt} + \omega_z^2 \delta = 0, \quad (7.52)$$

where:

$$\alpha_E = \frac{\beta_0 E_0}{2P_0 C_0} \frac{dU}{dE} \Big|_{E=E_0} = \frac{c}{2C_0} \frac{dU}{dE} \Big|_{E=E_0} = \frac{1}{2\beta_0 T_0} \frac{dU}{dE} \Big|_{E=E_0}, \quad (7.53)$$

and:

$$\frac{\omega_z^2}{\omega_0^2} = \nu_z^2. \quad (7.54)$$

$\omega_0$  is the angular revolution frequency  $2\pi\beta_0 c/C_0$ , and  $\nu_z$  is the synchrotron tune given by (5.49).

Equation (7.52) is the equation of motion for a damped harmonic oscillator, with frequency  $\omega_z$  and damping constant  $\alpha_E$ . If  $\alpha_E \ll |\omega_z|$ , then the longitudinal co-ordinate and energy deviation varies as:

$$z(t) = \sqrt{2\beta_z J_{z0}} e^{-\alpha_E t} \cos(\phi_z) = \sqrt{2\beta_z J_{z0}} e^{-\alpha_E t} \cos(\omega_z t + \phi_{z0}), \quad (7.55)$$

$$\delta(t) = -\sqrt{\frac{2J_{z0}}{\beta_z}} e^{-\alpha_E t} \sin(\phi_z) = -\sqrt{\frac{2J_{z0}}{\beta_z}} e^{-\alpha_E t} \sin(\omega_z t + \phi_{z0}), \quad (7.56)$$

where  $J_{z0}$  and  $\phi_{z0}$  are the synchrotron action and angle (respectively) at  $t = 0$ , and  $\beta_z$  is the longitudinal beta function, given by (5.66).

The longitudinal motion is the same as we found in Section 5.1.3, except now there is an additional term that describes damping of the oscillation amplitude, with damping constant  $\alpha_E$ . To complete the analysis of the longitudinal motion, we need to find an explicit expression for the damping constant in terms of known properties of the storage ring: in particular, we need an expression for the energy lost per turn by a particle emitting synchrotron radiation, as a function of the energy of the particle. To find the energy lost per turn, we can repeat our previous calculation and integrate the synchrotron radiation power  $P_\gamma$  over one revolution period. However,

we now need to take into account the fact that the revolution period depends on the energy deviation, i.e. there is some phase slip. For a particle with zero betatron amplitude,  $x = \eta_x \delta_p \approx \beta_0 \eta_x \delta$ , the path length element is given by:

$$dC = \left(1 + \frac{\eta_x}{\rho} \beta_0 \delta\right) ds, \quad (7.57)$$

where  $ds$  is an element of the reference trajectory, and  $\rho$  the radius of curvature of the reference trajectory. Hence, we can write, for the energy lost per turn:

$$U = \oint P_\gamma dt = \oint P_\gamma \frac{dC}{\beta_0 c} = \oint P_\gamma \left(1 + \frac{\eta_x}{\rho} \beta_0 \delta\right) \frac{ds}{\beta_0 c}. \quad (7.58)$$

Using (7.14) for  $P_\gamma$  and (7.37) for the magnetic field, we find (after some algebra) that:

$$\frac{dU}{dE} \Big|_{E=E_0} = \beta_0 j_z \frac{U_0}{E_0}, \quad (7.59)$$

where the energy loss per turn  $U_0$  for a reference particle is given by (7.20). The *longitudinal damping partition number*  $j_z$  is given by:

$$j_z = 2 + \frac{I_4}{I_2}. \quad (7.60)$$

$I_2$  and  $I_4$  are the synchrotron radiation integrals defined by (7.19) and (7.39) respectively. Finally, we can write the longitudinal damping time:

$$\tau_z = \frac{1}{\alpha_E} = \frac{2}{j_z} \frac{E_0}{U_0} T_0. \quad (7.61)$$

Note that  $\tau_z$  is the exponential decay time for the *amplitude* of a synchrotron oscillation. However, the longitudinal emittance is related to the mean square synchrotron amplitude; if the longitudinal motion is not coupled to the transverse motion, the longitudinal emittance can be written as:

$$\epsilon_z = \sqrt{\langle z^2 \rangle \langle \delta^2 \rangle - \langle z\delta \rangle^2}. \quad (7.62)$$

Hence, the longitudinal emittance is damped exponentially, with a decay time  $\tau_z/2$ :

$$\epsilon_z(t) = \epsilon_z(0) e^{-2t/\tau_z}. \quad (7.63)$$

Comparing (7.13), (7.42) and (7.61), we see that the radiation damping time for the horizontal, vertical and longitudinal motion can be written in the general form:

$$\tau_i = \frac{2}{j_i} \frac{E_0}{U_0} T_0, \quad (7.64)$$

where  $i = x, y$  or  $z$ . The horizontal and longitudinal damping partition numbers  $j_x$  and  $j_z$  are defined by (7.43) and (7.60), and we can, for completeness, define a vertical damping partition number  $j_y$ . For a planar storage ring without vertical dispersion, such as we have considered here, the value of  $j_y$  is simply:

$$j_y = 1, \quad (7.65)$$

but other values of  $j_y$  are possible in more general cases (for example, when there are vertical bends). From (7.43), (7.60) and (7.65), we see that for a planar storage ring:

$$j_x + j_y + j_z = 4. \quad (7.66)$$

The sum of the damping partition numbers is independent of the lattice design. This result, which also applies to more general cases than those we have considered here, is known as the *Robinson damping theorem* [Robinson (1958)]. In effect, the Robinson damping theorem states that although it is possible to shift the balance of the radiation damping between the different degrees of freedom in a storage ring, the overall rate of damping is fixed by the beam energy and the rate of energy loss through synchrotron radiation.

## 7.2 Quantum Radiation: Quantum Excitation

In Section 7.1, we considered the effects of synchrotron radiation on a beam in a storage ring, using a classical model for the radiation. We saw that the emission of the radiation led to an exponential decay of the horizontal, vertical and longitudinal emittances, with the emittances approaching zero as  $t \rightarrow \infty$ . In reality, it is observed that the emittances in an electron (or positron) storage ring reach non-zero equilibrium values. There are two reasons for this. First, radiation is not emitted continuously by a particle in a magnetic field, but in discrete quanta (photons). The random emission of photons acts as a ‘noise’ term in the equations of motion for a particle, which leads to some excitation of betatron and synchrotron oscillations. We shall discuss this *quantum excitation* in this section: taking this effect into account leads to equilibrium horizontal and longitudinal emittances with values determined by a balance between excitation and damping. In a planar storage ring, with zero vertical dispersion and no betatron coupling, the results of the analysis suggest that quantum effects lead to excitation only of the horizontal and longitudinal emittances, and that the vertical emittance will damp to zero. However, this is based on the assumption

that photons are emitted exactly in the instantaneous direction of motion of the radiating particle; this is only an approximation, and there is actually a small opening angle (approximately  $1/\gamma$ ) in the spatial distribution of synchrotron radiation. Taking this into account, it turns out that there is a small, but non-zero lower limit on the vertical emittance. In practice, though, the vertical emittance in storage rings is dominated by machine errors (such as magnet alignment errors) that introduce small amounts of vertical dispersion and betatron coupling. We shall discuss how to compute the equilibrium emittances in a storage ring in the presence of such errors in Section 7.4.

Let us first consider the effect of quantum excitation on the horizontal motion of a particle. To simplify the analysis, we shall take the ultra-relativistic limit,  $\beta_0 \rightarrow 1$ .  $dJ_x$ , the change in the horizontal action resulting from emission of radiation carrying momentum  $dP$  is given by (7.29):

$$dJ_x = -w_1 \frac{dP}{P_0} + w_2 \left( \frac{dP}{P_0} \right)^2. \quad (7.67)$$

In this equation,  $w_1$  and  $w_2$  are functions (of the particle co-ordinates, Courant–Snyder parameters and dispersion) given by (7.30) and (7.31), and  $P_0$  is the reference momentum. Dividing each term in (7.67) by a short time interval  $dt$  and taking the limit  $dt \rightarrow 0$  gives an expression for the rate of change of the horizontal action. Using a classical model for the radiation, we neglected the second term on the right-hand side of (7.67), since we assumed that:

$$\lim_{dt \rightarrow 0} \frac{(dP)^2}{dt} = 0. \quad (7.68)$$

However, using a quantum model, we can no longer make this assumption. Instead, we write:

$$\lim_{dt \rightarrow 0} \frac{dP}{dt} = \frac{1}{c} \int_0^\infty \dot{N}(u) u \, du = \frac{\langle \dot{N}(u) u \rangle}{c}, \quad (7.69)$$

and:

$$\lim_{dt \rightarrow 0} \frac{(dP)^2}{dt} = \frac{1}{c} \int_0^\infty \dot{N}(u) u^2 \, du = \frac{\langle \dot{N}(u) u^2 \rangle}{c}, \quad (7.70)$$

where  $\dot{N}(u) \, du$  is the rate of photon emission in the energy range  $u$  to  $u+du$ . The rate of change of the horizontal action is then given by:

$$\frac{dJ_x}{dt} = -w_1 \frac{\langle \dot{N}(u) u \rangle}{P_0 c} + w_2 \frac{\langle \dot{N}(u) u^2 \rangle}{P_0^2 c^2}. \quad (7.71)$$

The first term on the right-hand side of (7.71) gives radiation damping, as in the classical model. The second term on the right-hand side of (7.71) gives quantum excitation. Assuming that radiation effects are slow compared to the revolution frequency, we can average the excitation term around the circumference of the ring:

$$w_2 \frac{\langle \dot{N}(u)u^2 \rangle}{P_0^2 c^2} \approx \frac{1}{C_0} \oint w_2 \frac{\langle \dot{N}(u)u^2 \rangle}{P_0^2 c^2} ds. \quad (7.72)$$

Using (7.31) for  $w_2$ , we find that for  $x \ll \eta_x$  and  $p_x \ll \eta_{px}$ :

$$w_2 \frac{\langle \dot{N}(u)u^2 \rangle}{P_0^2 c^2} \approx \frac{1}{2E_0^2 C_0} \oint \mathcal{H}_x \langle \dot{N}(u)u^2 \rangle ds, \quad (7.73)$$

where the *dispersion H-function* (sometimes called the ‘curly-H function’)  $\mathcal{H}_x$  is defined by:

$$\mathcal{H}_x = \gamma_x \eta_x^2 + 2\alpha_x \eta_x \eta_{px} + \beta_x \eta_{px}^2. \quad (7.74)$$

The number of photons emitted within a given energy range by a charged particle in a magnetic field can be obtained from the intensity spectrum of the radiation. This is given by [Jackson (1998)]:

$$\frac{d\mathcal{P}}{d\vartheta} = \frac{9\sqrt{3}}{8\pi} P_\gamma \vartheta \int_\vartheta^\infty K_{5/3}(x) dx, \quad (7.75)$$

where  $d\mathcal{P}/d\vartheta$  is the energy radiated per unit time per unit frequency range, and  $\vartheta = \omega/\omega_c$  is the radiation (angular) frequency  $\omega$  divided by the *critical frequency*:

$$\omega_c = \frac{3}{2} \frac{\gamma_0^3 c}{\rho}. \quad (7.76)$$

$P_\gamma$  is the total energy radiated per unit time, given by (7.15), and  $K_{5/3}(x)$  is a modified Bessel function. Since the energy of a photon of frequency  $\omega$  is  $u = \hbar\omega$ , it follows that:

$$\dot{N}(u) du = \frac{1}{\hbar\omega} \frac{d\mathcal{P}}{d\vartheta} d\vartheta, \quad (7.77)$$

and hence, from (7.75):

$$\langle \dot{N}(u)u^2 \rangle = \hbar\omega_c \int_0^\infty \left( \int_\vartheta^\infty K_{5/3}(x) dx \right) \frac{9\sqrt{3}}{8\pi} P_\gamma \vartheta^2 d\vartheta. \quad (7.78)$$

It is possible to perform the integrals. Using (7.15) and (7.75), the result may be expressed as:

$$\langle \dot{N}(u)u^2 \rangle = 2C_q \gamma_0^2 E_0 \frac{P_\gamma}{\rho}, \quad (7.79)$$

where  $\gamma_0$  is the relativistic factor for a particle with energy  $E_0$ , and the quantum radiation constant  $C_q$  is [Sands (1970)]:

$$C_q = \frac{55}{32\sqrt{3}} \frac{\hbar}{mc}. \quad (7.80)$$

For electrons,  $C_q \approx 3.832 \times 10^{-13}$  m. Using (7.73) and (7.79) in (7.71), and averaging over all particles in the beam gives the rate of change of the horizontal emittance:

$$\frac{d\epsilon_x}{dt} = -\frac{2}{\tau_x} \epsilon_x + \frac{2}{j_x \tau_x} C_q \gamma_0^2 \frac{I_5}{I_2}, \quad (7.81)$$

where  $\tau_x$  and  $j_x$  are the horizontal damping time and the horizontal damping partition number, given by (7.42) and (7.43) respectively.  $I_5$  is the fifth synchrotron radiation integral, defined by:

$$I_5 = \oint \frac{\mathcal{H}_x}{|\rho^3|} ds. \quad (7.82)$$

We see in (7.81) that using a quantum model for the synchrotron radiation adds an excitation term to the damping of the horizontal emittance. However, whereas the damping term is proportional to the emittance (leading to an exponential decay), the excitation term is independent of the emittance. This means that an equilibrium emittance  $\epsilon_0$  can exist, where the excitation term exactly balances the damping term:

$$\epsilon_0 = C_q \gamma_0^2 \frac{I_5}{j_x I_2}. \quad (7.83)$$

The emittance given by (7.83) is called the *natural emittance* of a storage ring. In a given storage ring, we can change the beam energy by changing the fields in the dipoles. Assuming that if we change the dipole fields, we simultaneously change all other magnets by the same factor, the lattice functions (Courant–Snyder parameters and dispersion) and the synchrotron radiation integrals will remain constant. In that case, the equilibrium horizontal emittance will scale with the square of the beam energy. The exact value of the equilibrium emittance will depend on the lattice functions: we shall discuss this dependence in Section 7.3.

The solution to (7.81), taking into account radiation damping and quantum excitation, is:

$$\epsilon_x(t) = \epsilon_x(0) e^{-2t/\tau_x} + \epsilon_0 (1 - e^{-2t/\tau_x}). \quad (7.84)$$

$\epsilon_x(0)$  is the horizontal emittance at time  $t = 0$ ; in the limit  $t \rightarrow \infty$ , the emittance approaches the equilibrium emittance  $\epsilon_0$ .

The quantum excitation of the horizontal emittance depends on the horizontal dispersion. Similarly, the quantum excitation of the vertical emittance depends on the vertical dispersion. If there is no vertical dispersion (which may be the design intention in a planar storage ring), it may be expected that there will be no quantum excitation of the vertical emittance, and that the vertical emittance would therefore damp to zero. However, quantum excitation of the horizontal emittance may affect the vertical motion if synchrotron radiation emission occurs at a location where there is betatron coupling. This may be understood in physical terms as follows: Consider a particle following the appropriate closed orbit for the energy of the particle, i.e. the particle has zero betatron amplitude. If the particle emits a photon, then it experiences a sudden loss of energy. If the photon emission occurs at a location where the dispersion is non-zero, then the particle will not be on the closed orbit for the new energy. If the vertical dispersion is zero, then the particle will have a horizontal offset with respect to the closed orbit for the new energy. But if there is betatron coupling at the location where the photon emission occurs, then a horizontal offset corresponds to some linear combination of motion in the two transverse normal modes. Thus, both transverse normal modes are excited, and we expect to see some vertical betatron motion of the particle as it continues to travel around the storage ring.

In the absence of either vertical dispersion or betatron coupling, it may then be expected that the vertical emittance will damp to zero. However, the above analysis of quantum excitation does not take into account the non-zero opening angle of the synchrotron radiation. In reality, photons are not emitted exactly along the instantaneous direction of motion of a particle, but with some distribution around the direction of motion. Hence, a particle moving in a horizontal plane may emit a photon with some vertical momentum; to conserve momentum, the particle must itself then have some vertical momentum after emitting the photon, and hence it will have non-zero vertical betatron amplitude. This effect leads to a fundamental lower limit on the vertical emittance in a given storage ring. In practice, the vertical emittance in storage rings is generally dominated by alignment and tuning errors that generate vertical dispersion and betatron coupling. We do not carry out here an analysis of the fundamental lower limit on the vertical emittance; we simply quote, for reference, the result for the lower limit [Rauenheimer (1991)]:

$$\epsilon_{y,\min} = \frac{13}{55} \frac{C_q}{I_2} \oint \frac{\beta_y}{|\rho^3|} ds. \quad (7.85)$$

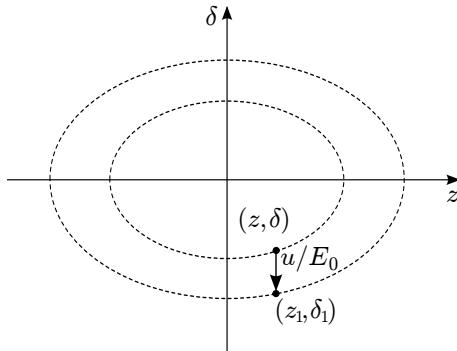


Fig. 7.1 Change in longitudinal phase space co-ordinates when a particle emits a photon of energy  $u$ . The reference energy is  $E_0$ . Each ellipse represents the trajectory of a particle in phase space with fixed longitudinal action  $J_z$ : the area of the ellipse is  $2\pi J_z$ . The longitudinal action changes as a result of the photon emission.

Finally, we consider the quantum excitation of longitudinal emittance. We take a different approach to longitudinal emittance compared to the one we took for the horizontal emittance, because synchrotron oscillations are slow compared to revolution frequency: to a good approximation, we can treat the longitudinal co-ordinate  $z$  and energy deviation  $\delta$  of any particle as being constant over one revolution.

Consider a particle with energy deviation:

$$\delta = -\sqrt{\frac{2J_z}{\beta_z}} \sin(\phi_z), \quad (7.86)$$

where  $J_z$  is the synchrotron action,  $\phi_z$  is the synchrotron angle, and  $\beta_z$  is the longitudinal beta function given by (5.66). (In the case that the longitudinal focusing is not too strong, we can assume that the longitudinal alpha function is zero.) If the particle emits a photon of energy  $u$ , then the energy deviation following the photon emission is:

$$\delta_1 = -\sqrt{\frac{2J_{z1}}{\beta_z}} \sin(\phi_{z1}) = -\sqrt{\frac{2J_z}{\beta_z}} \sin(\phi_z) - \frac{u}{E_0}, \quad (7.87)$$

where  $J_{z1}$  and  $\phi_{z1}$  are the new synchrotron action and angle following the photon emission (we assume that the longitudinal beta function  $\beta_z$  is independent of the particle energy). The longitudinal co-ordinate  $z$  does not change following the emission of a photon (see Fig. 7.1), and hence the ‘new’ co-ordinate is:

$$z_1 = \sqrt{2\beta_z J_z} \cos(\phi_z) = \sqrt{2\beta_z J_{z1}} \cos(\phi_{z1}). \quad (7.88)$$

Combining (7.87) and (7.88) and solving for  $J_{z1}$ , we find:

$$J_{z1} = J_z + \sqrt{2\beta_z J_z} \sin(\phi_z) \frac{u}{E_0} + \frac{\beta_z}{2} \frac{u^2}{E_0^2}. \quad (7.89)$$

Averaging over the bunch gives the change in the rms energy deviation:

$$\Delta\sigma_\delta^2 = \Delta\langle\delta^2\rangle = \frac{\Delta\langle J_z\rangle}{\beta_z} = \frac{\langle u^2 \rangle}{2E_0^2}. \quad (7.90)$$

Including the effects of radiation damping, the equation of motion for the energy spread can be written:

$$\frac{d\sigma_\delta^2}{dt} = -\frac{2}{\tau_z} \sigma_\delta^2 + \frac{1}{2E_0^2 C_0} \oint \langle \dot{N}(u) u^2 \rangle ds. \quad (7.91)$$

Using (7.79), we find:

$$\frac{d\sigma_\delta^2}{dt} = -\frac{2}{\tau_z} \sigma_\delta^2 + C_q \gamma_0^2 \frac{2}{j_z \tau_z} \frac{I_3}{I_2}, \quad (7.92)$$

where the third synchrotron radiation integral,  $I_3$ , is given by:

$$I_3 = \oint \frac{1}{|\rho^3|} ds. \quad (7.93)$$

We see that the rate of change of the energy spread includes a damping term proportional to the energy spread (from the radiation damping, leading to an exponential decay of the energy spread at large values), and an excitation term that is constant. The situation is exactly analogous to that for the horizontal emittance. The energy spread reaches equilibrium when the radiation damping is exactly balanced by the quantum excitation. The energy spread at equilibrium is called the *natural energy spread*,  $\sigma_{\delta0}$ , and is given by:

$$\sigma_{\delta0}^2 = C_q \gamma_0^2 \frac{I_3}{j_z I_2}. \quad (7.94)$$

The natural energy spread in a storage ring is determined by the beam energy and the bending radii of the dipoles. Although the rf system plays an important role in the longitudinal motion (by providing the longitudinal focusing necessary for synchrotron oscillations to occur), the natural energy spread is independent of the rf parameters: neither the rf frequency nor the rf voltage affect the natural energy spread.

A particle performing synchrotron oscillations follows an ellipse in longitudinal phase space, in which the maximum value of the longitudinal co-ordinate  $\hat{z}$  is related to the maximum value of the energy deviation  $\hat{\delta}$  by:

$$\hat{z} = \beta_z \hat{\delta} = -\beta_0 c \frac{\eta_p}{\omega_z} \hat{\delta}. \quad (7.95)$$

Radiation effects (damping and excitation) will lead to the beam reaching an equilibrium distribution in which the shape of the distribution is invariant over one complete revolution. The rms bunch length  $\sigma_z = \sqrt{\langle z^2 \rangle}$  is then related to the rms energy spread in the same way that the longitudinal oscillation amplitude  $\hat{z}$  for a single particle is related to the energy oscillation amplitude. Hence, at equilibrium with energy spread  $\sigma_{\delta 0}$  the rms bunch length is:

$$\sigma_{z0} = -\beta_0 c \frac{\eta_p}{\omega_z} \sigma_{\delta 0}. \quad (7.96)$$

The *natural bunch length*  $\sigma_{z0}$  depends not just on the beam energy and dipole bending radii (through the natural energy spread), but also on the phase slip factor and the synchrotron frequency. Reducing the phase slip factor leads to a reduction in the natural bunch length; increasing the rf voltage or rf frequency increases the synchrotron frequency, and also leads to a reduction in the natural bunch length.

### 7.3 Equilibrium Emittance and Lattice Design

The natural emittance of an electron or positron synchrotron storage ring often has a significant impact on the performance of the storage ring. In a synchrotron light source, for example, an important figure of merit for synchrotron radiation users is the brightness  $\mathcal{B}$ , which is a measure of the photon flux per unit phase space area of the source:

$$\mathcal{B} = \frac{\Phi}{4\pi^2 \Sigma_x \Sigma'_x \Sigma_y \Sigma'_y}. \quad (7.97)$$

Here,  $\Phi$  is the photon flux (photons per second) within a specified spectral range (typically within 0.1% of a specified frequency).  $\Sigma_x$  is given by the sum in quadrature of the electron horizontal beam size  $\sigma_x$  and  $\sigma_R$ , the size of the radiation beam produced by a single electron:

$$\Sigma_x = \sqrt{\sigma_x^2 + \sigma_R^2}. \quad (7.98)$$

Similarly,  $\Sigma'_x$  is determined by the particle horizontal beam divergence, and the divergence of the radiation beam from a single electron:

$$\Sigma'_x = \sqrt{\sigma'_x{}^2 + \sigma'_R{}^2}. \quad (7.99)$$

Corresponding formulae apply in the vertical direction for  $\Sigma_y$  and  $\Sigma'_y$ . For radiation at wavelength  $\lambda$  generated in an undulator of length  $L$ ,  $\sigma_R \approx \sqrt{\lambda/2L}$  and  $\sigma'_R \approx \sqrt{\lambda L/8\pi^2}$ . If the electron beam size and divergence

are comparable to the beam size and divergence of the radiation from a single electron, then the source is said to be ‘diffraction-limited’. Above the diffraction limit, where the electron beam size and divergence are large compared to the size and divergence of the radiation beam from a single electron, the brightness can be approximated by:

$$\mathcal{B} \approx \frac{\Phi}{4\pi^2 \sigma_x \sigma'_x \sigma_y \sigma'_y} \approx \frac{\Phi}{4\pi^2 \epsilon_x \epsilon_y}, \quad (7.100)$$

where  $\epsilon_x$  and  $\epsilon_y$  are the horizontal and vertical emittances of the electron beam. Hence, for a synchrotron light source operating above the diffraction limit, the smaller the natural emittance of the storage ring, the higher the brightness of the synchrotron radiation produced.

In a collider, the figure of merit corresponding to the brightness of a light source is the luminosity, which is a measure of the rate of particle collisions per unit area of the particle beams (see equation (12.250)). The lower the emittance of the beams, the higher the luminosity, and the higher the rate of events of given cross-section. It is therefore important for synchrotron light sources and for lepton colliders to understand how the natural emittance depends on the design of a storage ring.

We saw in Section 7.2 that the natural emittance in a storage ring is given by (7.83):

$$\epsilon_0 = C_q \gamma_0^2 \frac{I_5}{j_x I_2}, \quad (7.101)$$

where  $C_q$  is the quantum constant,  $\gamma_0$  is the relativistic factor of the beam,  $j_x$  is the horizontal damping partition number, and  $I_2$  and  $I_5$  are the second and fifth synchrotron radiation integrals respectively.  $I_2$  is determined by the radius of curvature of the reference trajectory (which we assume, for now, to be the same as the closed orbit) around the ring, and is obtained by a relatively straightforward integral.  $j_x$  is a more complicated function of the lattice, but can often be approximated, at least when the dipoles have no quadrupole component, by  $j_x \approx 1$ .  $I_5$  is also a complicated function of the lattice. For some important cases, there are useful approximations that can be made: we shall discuss a few of these cases in this section. We shall see that the natural emittance depends strongly on the design of the lattice. Our goal in this section is to derive relatively simple expressions for the natural emittance in electron storage rings of specified lattice designs. Although we must consider particular lattices, the examples we consider will be general enough to be relevant to many storage rings that are in operation, particularly in third-generation synchrotron light sources.

To begin, let us consider the second synchrotron radiation integral, given by:

$$I_2 = \oint \frac{ds}{\rho^2}. \quad (7.102)$$

Suppose that all the dipoles in the storage ring have the same field  $B$  (and no quadrupole component). Then the beam rigidity is:

$$B\rho = \frac{P_0}{q} \approx \frac{E_0}{qc}, \quad (7.103)$$

where  $q$  is the charge on a particle,  $P_0$  is the reference momentum,  $E_0$  is the energy of a particle with momentum  $P_0$ , and we have made the approximation for ultra-relativistic beams  $P_0 \approx E_0/c$ . Since  $\rho \rightarrow \infty$  outside of the dipoles, we can write:

$$I_2 = \int_{\text{dipoles}} \frac{ds}{\rho^2} = \int_{\text{dipoles}} \frac{B}{B\rho} \frac{ds}{\rho} \approx \frac{cB}{E_0/q} \int_{\text{dipoles}} \frac{ds}{\rho}, \quad (7.104)$$

where in the final step we have made use of the assumption that all dipoles have the same field, and the fact that the beam rigidity is constant around the ring. The integral in the final expression in (7.104) is the total bending angle of the dipoles: in a storage ring where all dipoles bend in the same direction, this must be  $2\pi$ . Hence, for a storage ring where all dipoles have the same field  $B$ :

$$I_2 \approx 2\pi \frac{cB}{E_0/q}. \quad (7.105)$$

Our main goal is to find an expression for the natural emittance in a storage ring of specified design; but we note in passing that using (7.105) in (7.20) gives, for the energy loss per turn:

$$U_0 \approx C_\gamma q c E_0^3 B, \quad (7.106)$$

and then from (7.42) the horizontal damping time is given by:

$$\tau_x \approx \frac{2T_0}{C_\gamma j_x q c E_0^2 B}. \quad (7.107)$$

The energy loss per turn varies with the third power of the beam energy, and is proportional to the dipole field strength; and the horizontal damping time is inversely proportional to the square of the beam energy, and inversely proportional to the dipole field strength.

### 7.3.1 Natural emittance in a FODO storage ring

Having obtained an estimate for  $I_2$  (7.105) in a storage ring, and if we assume that  $I_4 \approx 1$ , then to calculate the natural emittance for a given lattice we need to evaluate  $I_5$  (which is given, in general, by (7.82)). The exact evaluation of  $I_5$  requires detailed knowledge of the lattice functions (Courant–Snyder parameters and dispersion) through the dipoles. Figure 7.2 shows an example of the lattice functions in a FODO cell. In a realistic case, the formulae for the lattice functions become very complicated; it is often convenient to be able to make estimates based on a simplified description of the lattice. Consider, for example, a storage ring that consists entirely of repeated FODO cells, where the space between the quadrupole magnets in each cell is filled entirely by dipole magnets. To simplify things as far as possible, we represent the quadrupoles as thin lenses with focal lengths (in the horizontal plane):  $+f$  for the horizontally focusing quadrupole, and  $-f$  for the vertically focusing quadrupole. We assume that all dipoles have the same length  $L$  and the same bending radius  $\rho$ .

The lattice functions can be calculated using the techniques described in Sections 4.4 and 5.1.1. Using the transfer matrices for thin quadrupoles and long dipoles (of length  $L$ ), we find that the horizontal beta function at the horizontally focusing quadrupole is given by:

$$\beta_x = \frac{4f\rho \sin(\theta) (2f \cos(\theta) + \rho \sin(\theta))}{\sqrt{16f^4 - (\rho^2 - (4f^2 + \rho^2) \cos(2\theta))^2}}, \quad (7.108)$$

where  $\theta = L/\rho$  is the bending angle of a single dipole. The dispersion at a horizontally focusing quadrupole is:

$$\eta_x = \frac{2f\rho(2f + \rho \tan(\frac{\theta}{2}))}{4f^2 + \rho^2}. \quad (7.109)$$

By symmetry, at the centre of a quadrupole,  $\alpha_x = \eta_{px} = 0$ . We also know from Sections 4.2 and 5.1.1 how to evolve the Courant–Snyder parameters along a beam line using the transfer matrices. For the Courant–Snyder parameters:

$$A(s_1)^{-1} = R(s_1, s_0) A(s_0)^{-1} R(s_1, s_0)^T, \quad (7.110)$$

where  $R(s_1, s_0)$  is the (horizontal part of the) transfer matrix from  $s_0$  to  $s_1$ , and:

$$A^{-1} = \begin{pmatrix} \beta_x & -\alpha_x \\ -\alpha_x & \gamma_x \end{pmatrix}. \quad (7.111)$$

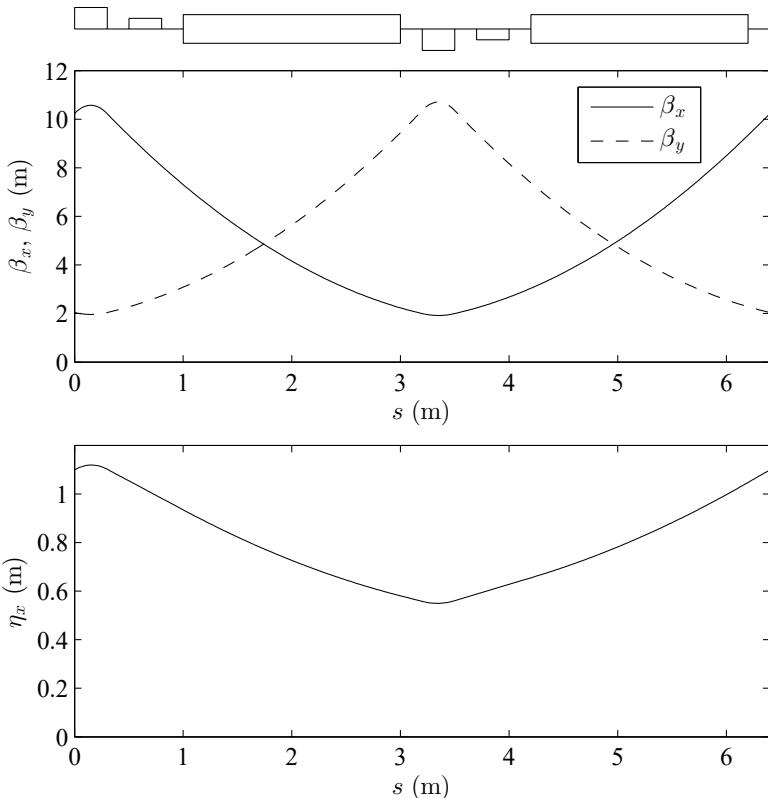


Fig. 7.2 Lattice functions in a FODO cell. The diagram at the top shows the positions of the dipole magnets (boxes extending above and below the line), quadrupole magnets (taller boxes extending above the line for horizontally focusing and below the line for horizontally defocusing quadrupoles) and sextupole magnets (shorter boxes). The plots show the horizontal and vertical beta functions (upper plot) and the horizontal dispersion (lower plot). For the FODO cell shown, the horizontal and vertical phase advances are each close to  $90^\circ$ .

The dispersion evolves through a dipole of bending radius  $\rho$  according to:

$$\begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_0+\Delta s} = R_{\text{dip}}(\Delta\theta) \begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix}_{s=s_0} + \begin{pmatrix} \rho(1 - \cos(\Delta\theta)) \\ \sin(\Delta\theta) \end{pmatrix}, \quad (7.112)$$

where  $\Delta\theta = \Delta s / \rho$ , and the horizontal part of the transfer matrix through a dipole of bending angle  $\Delta\theta$  is:

$$R_{\text{dip}}(\Delta\theta) = \begin{pmatrix} \cos(\Delta\theta) & \rho \sin(\Delta\theta) \\ -\frac{1}{\rho} \sin(\Delta\theta) & \cos(\Delta\theta) \end{pmatrix}. \quad (7.113)$$

The transfer matrix for a thin quadrupole of focal length  $f$  is:

$$R_{\text{quad}} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}. \quad (7.114)$$

Using the results (7.108)–(7.114), we can calculate the Courant–Snyder parameters and the dispersion through the FODO cell. We can then perform the integral (7.82) to find an expression for  $I_5$  in terms of the quadrupole focal length  $f$ , the dipole length  $L$  and the dipole bending radius  $\rho$ . The algebra is rather formidable, and not especially enlightening. The result, for the ratio  $I_5/I_2$ , can be expressed as a power series in the dipole bending angle  $\theta = L/\rho$ :

$$\frac{I_5}{I_2} = \left(4 + \frac{\rho^2}{f^2}\right)^{-\frac{3}{2}} \left(8 - \frac{\rho^2}{2f^2}\theta^2 + O(\theta^4)\right). \quad (7.115)$$

For small  $\theta$  (i.e. for storage rings consisting of many FODO cells), the expression for  $I_5/I_2$  can be written:

$$\frac{I_5}{I_2} \approx 8 \left(1 - \frac{L^2}{16f^2}\right) \left(4 + \frac{\rho^2}{f^2}\right)^{-\frac{3}{2}}. \quad (7.116)$$

If  $\rho \gg 2f$  (which is often the case in practice) this can be further simplified:

$$\frac{I_5}{I_2} \approx \left(1 - \frac{L^2}{16f^2}\right) \frac{8f^3}{\rho^3}. \quad (7.117)$$

If  $4f \gg L$  (which is not always the case in practice) this can be simplified still further:

$$\frac{I_5}{I_2} \approx \frac{8f^3}{\rho^3}. \quad (7.118)$$

With the ratio  $I_5/I_2$  given by the approximation (7.118), and using the approximation  $j_x \approx 1$ , the natural emittance in a FODO storage ring is:

$$\epsilon_0 \approx C_q \gamma^2 \left(\frac{2f}{L}\right)^3 \theta^3. \quad (7.119)$$

Some care must be taken when applying this equation, since the approximations required are not always valid. However, it does give some indication of how the emittance varies with the parameters of the storage ring. In particular, we see that the natural emittance increases with the square of the beam energy, and with the cube of the dipole bending angle. We shall see that these scalings are the same for other types of lattice (not just FODO lattices). Although increasing the energy of the beam in a storage ring

often has benefits for the users (for example, by increasing the energy of photons in synchrotron radiation), the drawback is an increase in natural emittance, which can be detrimental for the users. But for a given energy, it is possible to reduce the natural emittance by increasing the number of cells in the lattice; in effect, by increasing the size of the storage ring. This simple scaling is one reason why a storage ring for a light source designed to operate at, for example, 7 GeV will have a larger circumference than a storage ring for a light source designed to operate at 3 GeV.

We also see from (7.119) that the emittance is proportional to the cube of the quadrupole focal length: increasing the quadrupole strength reduces the focal length, and hence leads to a lower natural emittance. However, the quadrupole strength in a FODO lattice cannot be increased indefinitely. Neglecting the focusing provided by the dipoles, the phase advance in a FODO cell is given by (4.65):

$$\cos(\mu_x) = 1 - \frac{L^2}{2f^2}. \quad (7.120)$$

For stable motion, we must have:

$$\frac{f}{L} \geq \frac{1}{2}. \quad (7.121)$$

In the limiting case,  $\mu_x = 180^\circ$ , and the quadrupole focal length has the minimum value  $f = L/2$ . Using the approximation (7.119) the minimum emittance in a FODO lattice is expected to be:

$$\epsilon_0 \approx C_q \gamma^2 \theta^3. \quad (7.122)$$

However, as we increase the focusing strength, the approximations leading to (7.119) start to break down: for example, we can no longer assume that  $4f \gg L$ . If we use the exact formula for  $I_5/I_2$  (given to second order in  $\theta$  by (7.115)), then we find there is a minimum in the natural emittance:

$$\epsilon_0 \approx 1.2C_q \gamma^2 \theta^3, \quad (7.123)$$

which occurs for  $\mu_x \approx 137^\circ$ .

As an illustrative example, consider a 2 GeV storage ring constructed from 16 FODO cells (i.e. with 32 dipoles in total), with  $90^\circ$  phase advance per cell ( $f = L/\sqrt{2}$ ). Using the approximation (7.119), we expect that the natural emittance will be around 125 nm. This turns out to be a reasonably accurate figure, despite the approximations involved. Many modern applications, including synchrotron light sources, require emittances one or two orders of magnitude smaller in order to achieve the beam brightness

demanded by their users. The options to reduce the natural emittance are: first, to reduce the beam energy; second, to increase the number of cells; third, to increase the focusing; and fourth, to consider a different lattice design altogether. Reducing the energy of the electron beam also reduces the energies of photons in the synchrotron radiation, which is undesirable for some light source users. To reduce the emittance by an order of magnitude by increasing the number of cells would require the size of the ring to be more than doubled, which would more than double the cost of the storage ring. Because of the stability limits on FODO lattices, there is little that can be gained by increasing the focusing strengths of the quadrupoles. Consequently, we are left to consider changing the lattice design as to reduce the natural emittance. Fortunately, there are lattice designs, based on cells only slightly more complicated than FODO cells, that can reduce the natural emittance (for given beam energy and ring circumference) very effectively. We shall consider some such designs in the following sections.

### **7.3.2 Double-bend achromat**

The natural emittance in a storage ring is determined by the balance between quantum excitation (characterised by the fifth synchrotron radiation integral  $I_5$ ) and radiation damping (characterised by the second synchrotron radiation integral  $I_2$ ). Basing a lattice design on FODO cells makes no attempt to optimise the ratio of quantum excitation to radiation damping: a FODO cell is just the simplest kind of lattice design that one could consider for a synchrotron storage ring. As an initial attempt to reduce the natural emittance, we might consider a design in which the dipoles occur in pairs within each cell, with zero dispersion ( $\eta_x = \eta_{px} = 0$ ) at the entrance to the first dipole and at the exit of the second dipole. The dispersion  $\mathcal{H}$ -function (7.74) that appears in the integral for  $I_5$  is then zero at the entrance and exit of each cell. Although each dipole generates dispersion (and hence the value of  $\mathcal{H}_x$  cannot remain zero through the dipoles) we might hope that this approach leads to an overall reduction in the integral of  $\mathcal{H}_x$  in each dipole, compared to the dipoles in a FODO cell, and hence to a reduction in the quantum excitation, and finally to a reduction in the natural emittance.

It turns out to be reasonably straightforward to design a cell where the dispersion at the entrance and the exit are zero. Suppose the dipoles are separated by a distance  $2L_{\text{drift}}$ , with a quadrupole exactly half-way between them (i.e. a distance  $L_{\text{drift}}$  from each dipole). We shall assume

that the dipoles have no quadrupole component. If the dispersion at the entrance of the cell is zero, then by symmetry, the dispersion at the exit of the cell will be zero if the quadrupole reverses the gradient of the dispersion. In the thin lens approximation for the quadrupole, this condition can be written:

$$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} \eta_x \\ \eta_{px} \end{pmatrix} = \begin{pmatrix} \eta_x \\ \eta_{px} - \frac{\eta_x}{f} \end{pmatrix} = \begin{pmatrix} \eta_x \\ -\eta_{px} \end{pmatrix}, \quad (7.124)$$

where  $\eta_x$  and  $\eta_{px}$  are (respectively) the dispersion and the gradient of the dispersion at the entrance of the quadrupole. Hence, the quadrupole must have focal length:

$$f = \frac{\eta_x}{2\eta_{px}}. \quad (7.125)$$

The actual values of the dispersion and the gradient of the dispersion are determined by the dipole bending angle  $\theta$ , the dipole bending radius  $\rho$  and the drift length  $L_{\text{drift}}$ :

$$\eta_x = \rho (1 - \cos(\theta)) + L_{\text{drift}} \sin(\theta), \quad (7.126)$$

$$\eta_{px} = \sin(\theta). \quad (7.127)$$

A cell containing two dipoles, with zero dispersion at the entrance and exit of the cell, is known as a *double-bend achromat* (DBA), or as a *Chasman–Green* lattice [Chasman *et al.* (1975)]. As we have seen, the zero-dispersion condition can be achieved by a quadrupole of suitable strength placed mid-way between the two dipoles. To complete the cell, we need some quadrupoles to control the Courant–Snyder parameters. Ideally, the Courant–Snyder parameters at the entrance to the first dipole will be set so that the integral of  $\mathcal{H}_x$  over the dipole is minimised. If the DBA cell has a symmetry point mid-way between the dipoles, then minimising the integral of  $\mathcal{H}_x$  over the first dipole will also minimise  $I_5$ , and hence (for given dipole parameters) minimise the natural emittance. The full analysis of the optics in a DBA cell involves complicated algebra; here, we simply state the result that the emittance in a DBA lattice is minimised if the Courant–Snyder parameters at the entrance to the first dipole in each cell take the values:

$$\beta_x = \sqrt{\frac{12}{5}}L + O(\theta^3), \quad (7.128)$$

$$\alpha_x = \sqrt{15} + O(\theta^2), \quad (7.129)$$

where  $\theta$  is the dipole bending angle, and  $L = \rho\theta$  is the length of each dipole. These values for the Courant–Snyder parameters can be achieved using a

pair of quadrupole magnets (with opposite polarity and some drift length between them) upstream of the first dipole magnet, with a similar pair in the reverse order (to maintain the symmetry of the cell) downstream of the second dipole. Of course, it is necessary to consider the vertical focusing as well as the horizontal, but fortunately it is usually not too difficult to find solutions that give stable horizontal and vertical motion while providing minimum  $I_5$  in the cell. An example of the lattice functions in a DBA cell is shown in Fig. 7.3.

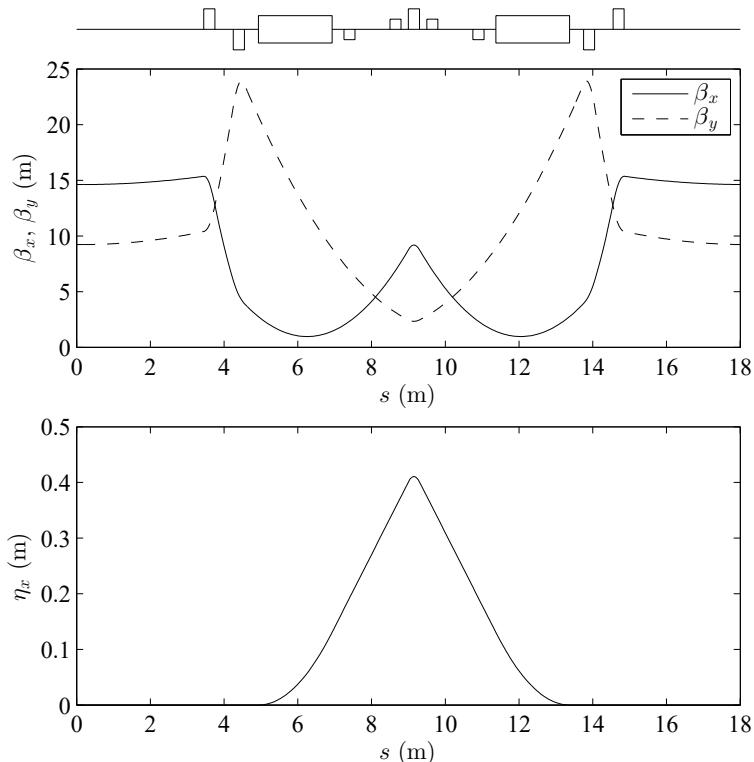


Fig. 7.3 Lattice functions in a double-bend achromat cell. The diagram at the top shows the positions of the dipole magnets (boxes extending above and below the line), quadrupole magnets (taller boxes extending above the line for horizontally focusing and below the line for horizontally defocusing quadrupoles) and sextupole magnets (shorter boxes). The plots show the horizontal and vertical beta functions (upper plot) and the horizontal dispersion (lower plot). The dispersion is zero entering the first dipole, and leaving the second dipole.

The minimum value of  $I_5$  in a DBA lattice (when the values of the Courant–Snyder parameters at the entrance of the first dipole are given by (7.128) and (7.129)) is:

$$I_5 = \frac{1}{4\sqrt{15}} \frac{\theta^4}{\rho} + O(\theta^6). \quad (7.130)$$

Since we know that the value of  $I_2$  in a single dipole is:

$$I_2 = \int \frac{ds}{\rho^2} = \frac{\theta}{\rho}, \quad (7.131)$$

we can now write an expression for the minimum natural emittance in a DBA lattice, using (7.83):

$$\epsilon_0 \approx \frac{1}{4\sqrt{15}} C_q \gamma^2 \theta^3. \quad (7.132)$$

The approximation is valid for small  $\theta$ , and we have used the fact that there is no quadrupole component in the dipole fields to make the approximation  $j_x \approx 1$ .

If we compare the expression (7.132) for the minimum natural emittance in a DBA cell with the corresponding expression (7.123) for the minimum natural emittance in a FODO cell, we see that in both cases the minimum emittance scales with the square of the beam energy, and the cube of the dipole bending angle. However, for a given beam energy and dipole bending angle, the emittance in a DBA lattice is smaller than that in a FODO lattice by roughly a factor  $4\sqrt{15} \approx 15.5$ . This is a significant improvement; however, it is possible to reduce the natural emittance even further.

### 7.3.3 TME lattices and multibend achromats

A DBA lattice has the constraint that the dispersion is zero at the entrance and at the exit of each pair of dipoles. By relaxing this constraint, we can minimise the emittance in a storage ring with given energy and dipole parameters. We continue to assume that the dipoles have no quadrupole component. The resulting lattice is known as a *theoretical minimum emittance* (TME) lattice.

A TME cell consists of a single dipole, with quadrupoles to control the Courant–Snyder parameters and dispersion through the cell. We assume a point of symmetry for the cell at the centre of the dipole. To minimise the natural emittance, we simply minimise  $I_5$ :

$$I_5 = \int_{\text{dipole}} \frac{\mathcal{H}_x}{\rho} ds, \quad (7.133)$$

with respect to variations in the Courant–Snyder parameters and dispersion. The integral in (7.133) is taken over the full length of the dipole. As is usually the case for these calculations, the algebra is complicated: so we simply state the result. The minimum value for  $I_5$  in a TME lattice is:

$$I_5 = \frac{1}{12\sqrt{15}} \frac{\theta^4}{\rho} + O(\theta^6). \quad (7.134)$$

Using (7.131) for  $I_2$ , and (7.83) we find that the minimum natural emittance in a storage ring with a TME lattice is:

$$\epsilon_0 \approx \frac{1}{12\sqrt{15}} C_q \gamma^2 \theta^3. \quad (7.135)$$

The minimum natural emittance is achieved when the dispersion and the horizontal beta function reach minimum values at the centre of the dipole, given respectively by:

$$\eta_x = \left(1 - \frac{2}{\theta} \sin\left(\frac{\theta}{2}\right)\right) \rho = \frac{L\theta}{24} + O(\theta^4), \quad (7.136)$$

and:

$$\beta_x = \frac{L}{2\sqrt{15}} + O(\theta^2). \quad (7.137)$$

It is sometimes useful to know the values of the lattice functions at the entrance to a dipole that minimise the emittance in a TME cell. The Courant–Snyder parameters are:

$$\beta_x = \frac{8}{\sqrt{15}} L + O(\theta^2), \quad (7.138)$$

$$\alpha_x = \sqrt{15} + O(\theta^2), \quad (7.139)$$

and the dispersion is:

$$\eta_x = \frac{1}{6} L\theta + O(\theta^3), \quad (7.140)$$

$$\eta_{px} = -\frac{\theta}{2} + O(\theta^3). \quad (7.141)$$

Comparing (7.135) and (7.132), we see that for a given beam energy and dipole bending angle, a TME lattice can achieve a natural emittance that is three times smaller than the natural emittance that can be achieved in a DBA lattice. This is potentially a significant benefit; unfortunately, TME lattices also have some significant drawbacks. First, in practice it is difficult to achieve the exact conditions for the dispersion and Courant–Snyder parameters in a TME lattice in a reasonably compact cell. Even

where the conditions are achieved, correction of chromaticity can be very difficult, and the dynamic aperture tends to be very small, which leads to a reduction in beam lifetime. Second, there is naturally non-zero dispersion throughout the lattice. If insertion devices are placed between the dipoles (to enhance the production of synchrotron radiation for light source users, for example) then this can lead to an increase in the natural emittance, despite the optimum conditions being achieved in the dipoles. On the other hand, in a DBA lattice, there is naturally zero dispersion in the regions outside of each pair of dipoles: this provides good locations for insertion devices where the only dispersion is the small amount generated by the insertion devices themselves, and any blow-up of the natural emittance is therefore limited.

To optimise a storage ring lattice for a light source, several approaches are possible. For example, if the lattice is based on a DBA structure, then by detuning the quadrupole(s) between each pair of dipoles and thereby allowing some dispersion to ‘leak’ into the region outside each dipole pair, the lattice functions can be adjusted to reduce the emittance below that of a strict DBA. In effect, the lattice functions in each dipole are somewhere between the conditions for a DBA lattice and the conditions for a TME lattice. Although there is then some dispersion at locations where insertion devices may be placed, if the dispersion is carefully controlled then the resulting increase of emittance from quantum excitation in the insertion devices can be limited so that there is still overall a benefit from detuning the DBA. This is a technique that is widely used in storage rings for light sources.

It should be remembered that for each of the lattice types we have considered (FODO, DBA and TME) our analysis has been based on the assumption that the dipole fields have no quadrupole component. In practice, including a gradient in the dipole field provides an extra degree of freedom that can be used to reduce the natural emittance, either by reducing the quantum excitation (i.e. reducing  $I_5$ ) or by increasing the horizontal damping partition number  $j_x$  (in effect, shifting the balance between the equilibrium emittances away from the horizontal plane and towards the longitudinal). The drawback of this technique is that the design and construction of the dipole magnets becomes more complicated when some field gradient is required; and there is also some loss in operational flexibility in having some of the transverse focusing provided by a quadrupole field that can only be adjusted in proportion to the main dipole field.

To complete our discussion of natural emittance and lattice design,

we shall consider multibend achromats. A multibend achromat (MBA) is formed by a group of dipoles with zero dispersion at the entrance and exit of the group. A DBA cell is a special case of an MBA cell, with two dipoles. In principle, the dipoles within a cell may have different parameters: for simplicity, we shall consider the special case where all dipoles have the same bending radius (i.e. the same field strength) but vary in length. In particular, let the outer two dipoles have bending angle  $a\theta$  and the remaining dipoles have bending angle  $b\theta$ , where  $\theta = 2\pi/MN$  is the average bending angle per dipole, and the lattice consists of  $N$  cells with  $M$  dipoles per cell. Then:

$$2a + (M - 2)b = M. \quad (7.142)$$

To minimise the natural emittance, we need to control the dispersion and the Courant–Snyder parameters to minimise the contribution of each dipole to  $I_5$ . Since the contributions of the dipoles to the synchrotron radiation integrals are additive, the total value of  $I_5$  is minimised (subject to achromat condition) if lattice functions in the outer two dipoles match the lattice functions in a DBA, and the lattice functions in the remaining dipoles match the lattice functions in a TME. Using (7.130) and (7.134) we can immediately write down an expression for  $I_5$  in an MBA lattice:

$$I_5 \approx \frac{2}{4\sqrt{15}} \frac{(a\theta)^4}{\rho} + \frac{(M - 2)}{12\sqrt{15}} \frac{(b\theta)^4}{\rho} = \frac{(6a^4 + (M - 2)b^4)}{12\sqrt{15}} \frac{\theta^4}{\rho}. \quad (7.143)$$

From (7.131),  $I_2$  is approximated by:

$$I_2 \approx 2 \frac{a\theta}{\rho} + (M - 2) \frac{b\theta}{\rho}. \quad (7.144)$$

Combining (7.143) and (7.144) gives:

$$\frac{I_5}{I_2} \approx \frac{1}{12\sqrt{15}} \left( \frac{6a^4 + (M - 2)b^4}{2a + (M - 2)b} \right) \theta^3. \quad (7.145)$$

For a given value of  $M$ , we can eliminate  $b$  using (7.142). Then, minimising  $I_5/I_2$  with respect to  $a$ , we find that the minimum value of  $I_5/I_2$  occurs when:

$$\frac{a}{b} = \frac{1}{\sqrt[3]{3}}, \quad (7.146)$$

in which case:

$$\frac{6a^4 + (M - 2)b^4}{2a + (M - 2)b} \approx \frac{M + 1}{M - 1}. \quad (7.147)$$

Equation (7.146) tells us that to minimise the emittance in an  $M$ -bend achromat, the inner dipoles in each achromat should be longer than the outer two dipoles by a factor of  $\sqrt[3]{3}$ . Combining (7.145) and (7.147), and using (7.83), the minimum natural emittance in a multibend achromat with  $M$  dipoles per cell is:

$$\epsilon_0 \approx \frac{C_q \gamma^2}{12\sqrt{15}} \left( \frac{M+1}{M-1} \right) \theta^3. \quad (7.148)$$

Although our assumptions suggest that (7.148) should be valid for  $2 < M < \infty$ , we see that we obtain the correct results; (7.132) for the case  $M = 2$  (DBA lattice), and (7.135) for the case  $M \rightarrow \infty$  (TME lattice).

Historically, synchrotron light sources are often classified as first, second or third-generation [Winick (1994)]. Starting in the early 1950's, first-generation light sources typically exploited synchrotron radiation produced from dipoles in storage rings built for other purposes, such as experimental high-energy physics: since the brightness of the radiation was not of any real concern in the design of the storage ring, first-generation light sources often used simple (e.g. FODO-style) lattices leading to high values of the natural emittance. As the community of synchrotron radiation users began to grow, it was realised that higher brightness radiation would greatly enhance the range and quality of the science that could be performed using synchrotron radiation. In the early 1980's this led to the development of second-generation light sources, designed specifically for the production of synchrotron radiation, and typically use lower-emittance lattice designs, such as the double-bend achromat. The dipole magnets were still the primary source of synchrotron radiation in second-generation light sources, but starting from the mid-1980's, as experimental techniques advanced, more demanding requirements drove the development of third-generation light sources in which the storage rings were optimised for production of synchrotron radiation from insertion devices placed in the straight sections between the dipoles. The possibility to increase the number of source points in a storage ring allowed greater capacity in terms of the number of available photon beamlines. In addition, the ability to control the properties of the synchrotron radiation (spectral range, polarisation etc.) by modifying the design and parameters of an insertion device, without significantly affecting the storage ring operation, led to light sources with much greater flexibility compared to first and second-generation machines.

Modern light sources now aim to operate close to the diffraction limit, where the emittance of the electron beam is equal to the intrinsic emittance of the photon beam: any reduction in electron beam emittance below

the diffraction limit does not increase the brightness of the synchrotron radiation (7.97), and is not observable by light-source users. The intrinsic emittance of a photon beam at a wavelength  $\lambda$  is  $\lambda/4\pi$ , so for an x-ray light source this implies reaching an electron beam emittance of a few tens of picometres. Multi-bend achromats offer means of achieving emittances in this regime in a relatively compact lattice with electron beam energies of a few GeV (suitable for the production of x-rays), and despite some challenging beam dynamics issues [Takao (2013)], lattice designs based on multi-bend achromats have been developed for a number of existing or proposed facilities; see, for example [Eriksson *et al.* (2008); Leemann *et al.* (2009); Shimosaki *et al.* (2011); Cai *et al.* (2012); Farvacque *et al.* (2013); Schroer *et al.* (2018); Singh *et al.* (2018); Hellert *et al.* (2022)]. As an alternative to multi-bend achromats, or as a route to even lower emittances, it is possible to consider more exotic designs, for example using dipole magnets in which the field varies along the length of the dipole, or using structures in which some dipoles act as ‘reverse’ bends (see, for example, [Nagaoka and Wrulich (2007); Wang (2009); Riemann and Streun (2019); Yang *et al.* (2021)]). Introducing additional degrees of freedom in the lattice design in this way allows greater control over the lattice functions in order to reduce the natural emittance (minimising the quantum excitation while maximising the radiation damping).

The continuous advances in synchrotron light sources, particularly with techniques to reduce the natural emittance, means that these machines are likely to remain of great importance for scientific research for some time. Applications cover a wide range of fields, including life sciences, materials science, technology and engineering. The capacity offered by large numbers of insertion devices providing photon beams to dozens of experimental stations operating simultaneously is an attractive feature of a storage-ring based facility, as is the intrinsic stability provided by an electron beam maintained (usually with the assistance of numerous feedback systems) in equilibrium. However, the development of free-electron lasers (FELs) [Madey (1971); Deacon *et al.* (1977)] driven by linear accelerators has provided the ability to generate synchrotron radiation with higher brightness and shorter pulse lengths than can be achieved from a storage ring [Seddon *et al.* (2017); Maroju *et al.* (2020)]. Light sources based on free-electron lasers are often referred to as fourth-generation facilities, though the difference between an FEL and a synchrotron light source is more significant than the differences between machines in the first three generations of light sources. The properties of the light produced by an FEL depend on

coherent emission, in which large numbers of particles act as a single source of radiation. This leads to an increase in brightness of many orders of magnitude over the usual radiation generated from a dipole or insertion device in a storage ring, where particles radiate individually. We do not discuss the physics of coherent radiation here, but refer the reader to other texts (for example, [Schmüser *et al.* (2008, 2014); Pellegrini *et al.* (2016); Kim *et al.* (2017)]).

## 7.4 Computation of Equilibrium Emittances

Although the formulae in Section 7.3 are useful for making estimates of the natural emittance in certain cases, for detailed studies of particular lattices we need techniques for accurate calculations of the emittances, including the vertical and longitudinal emittances as well as the horizontal (i.e. including the effects of coupling between the three degrees of freedom). Generally, the calculation of the equilibrium emittances needs to be done by computer, based on a model of the lattice including magnet alignment and tuning errors. There are various methods for computing the equilibrium emittances in a general storage ring: see, for example, Chao's method [Chao (1979)], and Sagan's method that generalises the usual synchrotron radiation integrals to the coupled case, using the normal modes [Sagan (2022)]. In Section 7.4.3, we describe the envelope method [Ohmi *et al.* (1994)], which is based on computing first the equilibrium beam distribution in the form of the matrix  $\Sigma$  (5.74) consisting of the second-order moments of the phase space variables. The emittances are then obtained from the matrix  $\Sigma$  using the result from Section 5.2 that the eigenvalues of  $\Sigma S$  are  $\pm i\epsilon_k$ , where  $S$  is the antisymmetric matrix (4.14), and  $\epsilon_k$  are the beam emittances (with  $k = \text{I}, \text{II}, \text{III}$ ; an index used to label the degrees of freedom).

To begin, we consider two special cases for which, with some assumptions, we can derive approximate formulae to describe the effect of magnet misalignments on the vertical emittance in a storage ring. In Section 7.4.1 we discuss the generation of vertical emittance from betatron coupling arising (for example) from a skew quadrupole in a storage ring lattice. Then, in Section 7.4.2, we consider the effect of vertical dispersion, which may arise from steering or coupling errors. We show that for betatron coupling and vertical dispersion considered separately, it is possible to derive relatively simple formulae for calculating the vertical emittance generated in each case. However, in general, calculation of the equilibrium emittances in a storage ring must be found computationally, from a model including

all the relevant lattice errors: the envelope method provides one possible technique for this task, and is described in Section 7.4.3.

### 7.4.1 Effects of betatron coupling

Our aim in this section is to derive a formula for the equilibrium vertical emittance in a storage ring (determined by synchrotron radiation effects) in the presence of betatron coupling. We shall assume that the coupling arises from skew quadrupole fields distributed around the ring. To derive the required formula for vertical emittance, we shall first consider the motion of a single particle moving round the storage ring in the absence of synchrotron radiation. We shall find that the skew quadrupole fields drive some regular, periodic variation in the horizontal and vertical betatron actions, but under certain conditions (and with some simplifications) the sum of the two betatron actions remains constant. If we identify the total betatron action with the natural emittance determined by synchrotron radiation, we obtain convenient expressions for the equilibrium horizontal and vertical emittances of a beam of particles in a storage ring, determined by synchrotron radiation and betatron coupling.

To begin, we represent the skew quadrupole fields as a function of position around the storage ring:

$$k_1^{(s)} = -\frac{q}{P_0} \frac{\partial B_x}{\partial x}, \quad (7.149)$$

where  $q$  is the charge on a particle in the beam,  $P_0$  is the reference momentum,  $B_x$  is the horizontal field component as a function of distance along the reference trajectory, and  $x$  is the horizontal co-ordinate. Assuming zero momentum deviation ( $\delta = 0$ ) and neglecting synchroton motion, the motion of a particle in a skew quadrupole field may be described (approximately) by the Hamiltonian (3.109):

$$H_{\text{sq}} = \frac{1}{2} p_x^2 + \frac{1}{2} p_y^2 + k_1^{(s)} xy. \quad (7.150)$$

It is convenient to work in action-angle variables  $(\phi_x, J_x)$  and  $(\phi_y, J_y)$  for the horizontal and vertical motion respectively. In terms of these variables the horizontal and vertical co-ordinates can be written:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (7.151)$$

$$y = \sqrt{2\beta_y J_y} \cos(\phi_y). \quad (7.152)$$

Following (4.58), the Hamiltonian (7.150) then takes the form:

$$H = \frac{J_x}{\beta_x} + \frac{J_y}{\beta_y} + 2k_1^{(s)} \sqrt{\beta_x \beta_y} \sqrt{J_x J_y} \cos(\phi_x) \cos(\phi_y). \quad (7.153)$$

where  $\beta_x$  and  $\beta_y$  are the horizontal and vertical beta functions (which, in general, vary with position around the ring). Note that since we added the skew quadrupole term explicitly to the Hamiltonian, it is appropriate to use the beta functions in the storage ring lattice in the absence of any skew quadrupole fields here. We will ignore any synchrotron motion of the particle.

The next step is to derive the equations of motion for a particle moving around the storage ring. The motion is complicated because beta functions and skew quadrupole fields vary around the ring. In order to make progress, we therefore assume that we can replace these quantities by ‘averaged’ values, so that the Hamiltonian becomes:

$$H = \omega_x J_x + \omega_y J_y + K_\Delta \sqrt{J_x J_y} \cos(\phi_x - \phi_y) + K_\Sigma \sqrt{J_x J_y} \cos(\phi_x + \phi_y), \quad (7.154)$$

where:

$$\omega_{x,y} = \frac{1}{C_0} \int_0^{C_0} \frac{1}{\beta_{x,y}} ds. \quad (7.155)$$

$C_0$  is the circumference of the storage ring. The quantities  $K_\Delta$  and  $K_\Sigma$  characterise the effects of the skew quadrupole fields around the storage ring: expressions for these quantities are given below.

Note that we have written the product  $\cos(\phi_x) \cos(\phi_y)$  as a sum of two terms: this allows us to take into account the betatron phase advance of a particle moving between successive skew quadrupoles. The importance of accounting for the phase advance in this way can be illustrated as follows. Consider two skew quadrupoles of opposite strength at locations in the storage ring such that the beta functions at the second skew quadrupole have the same values as at the first. Suppose that the phase advance from the first to the second skew quadrupole is  $\pi$  in the horizontal plane, and  $2\pi$  in the vertical plane. In that case, in moving from the first to the second skew quadrupole, co-ordinates of a particle will transform such that  $\cos(\phi_x) \mapsto -\cos(\phi_x)$  and  $\cos(\phi_y) \mapsto \cos(\phi_y)$ . As a result, taking into account the opposite strengths of the skew quadrupoles, the kicks from the skew quadrupole fields (represented by the final term in the Hamiltonian (7.153)) will combine to give an effect roughly twice as large as either of the single skew quadrupoles alone. However, if we simply average the

skew quadrupole strengths, without taking into account the phase advance between them, then the net effect would be zero.

To allow for the effect of the phase advance, the skew quadrupole fields around the ring are represented in the averaged Hamiltonian (7.154) by quantities  $K_\Delta$  and  $K_\Sigma$ , defined by:

$$K_\Delta e^{i\chi_\Delta} = \frac{1}{C_0} \int_0^{C_0} e^{i(\psi_x - \psi_y)} \sqrt{\beta_x \beta_y} k_1^{(s)} ds, \quad (7.156)$$

and:

$$K_\Sigma e^{i\chi_\Sigma} = \frac{1}{C_0} \int_0^{C_0} e^{i(\psi_x + \psi_y)} \sqrt{\beta_x \beta_y} k_1^{(s)} ds, \quad (7.157)$$

where  $\psi_x(s)$  and  $\psi_y(s)$  represent the phase advances in the horizontal and vertical planes respectively, from  $s = 0$  to a given point around the storage ring. The complex phases of the integrals in (7.156) and (7.157) are accounted in the phase angles  $\chi_\Delta$  and  $\chi_\Sigma$ , so that  $K_\Delta$  and  $K_\Sigma$  are purely real.

To make a further simplification, let us consider the case that the fractional parts of the betatron tunes are approximately equal, i.e. the storage ring is operated close to a *difference resonance*, where the betatron tunes are such that  $\nu_x - \nu_y$  is an integer. As a result, the value of  $\cos(\phi_x - \phi_y)$  for a particle will remain roughly constant at any given point in the storage ring over successive turns. The value of  $\cos(\phi_x + \phi_y)$ , on the other hand, will oscillate within the range  $-1$  to  $+1$ . Consequentially, over many turns we can expect the ‘difference’ term (with coefficient  $K_\Delta$ ) in the Hamiltonian (7.154) to dominate over the ‘sum’ term (with coefficient  $K_\Sigma$ ). We therefore assume that we can drop the sum term, and consider the Hamiltonian:

$$H = \omega_x J_x + \omega_y J_y + K_\Delta \sqrt{J_x J_y} \cos(\phi_x - \phi_y). \quad (7.158)$$

Although we have made a number of assumptions and approximations, the benefit is some simplification of the equations of motion that follow from the Hamiltonian (7.158):

$$\begin{aligned} \frac{d\phi_x}{ds} &= \omega_x + \frac{K_\Delta}{2} \sqrt{\frac{J_y}{J_x}} \cos(\phi_x - \phi_y), & \frac{dJ_x}{ds} &= K_\Delta \sqrt{J_x J_y} \sin(\phi_x - \phi_y), \\ \frac{d\phi_y}{ds} &= \omega_y + \frac{K_\Delta}{2} \sqrt{\frac{J_x}{J_y}} \cos(\phi_x - \phi_y), & \frac{dJ_y}{ds} &= -K_\Delta \sqrt{J_x J_y} \sin(\phi_x - \phi_y). \end{aligned} \quad (7.159)$$

Despite the approximations we have made, it is still not easy to find a general solution to these equations. However, even without writing down a complete solution, we can see that:

$$\frac{d}{ds}(J_x + J_y) = 0. \quad (7.160)$$

Hence, the sum of the betatron actions of a particle remains constant as the particle moves around the storage ring. If a difference resonance dominates over a sum resonance (where  $\nu_x + \nu_y$  is an integer, and the term with coefficient  $K_\Sigma$  dominates over that with coefficient  $K_\Delta$  in the Hamiltonian (7.154)), the presence of skew quadrupole fields will lead to transfer of betatron amplitude between the horizontal and vertical motion, but the overall total betatron action will stay the same. The rate of change of each of the betatron actions depends on  $\sin(\phi_x - \phi_y)$ : this means that the closer the betatron tunes are to a difference resonance, the slower the variation in the betatron actions. The betatron actions will vary periodically and in antiphase, at a rate determined by the proximity of the tunes to the difference resonance.

A further result that can be derived from the equations of motion is that for particular values of the betatron actions, the phase difference  $\Delta\phi = \phi_x - \phi_y$  remains constant. The values of  $J_x$  and  $J_y$  required to satisfy this condition can be found from:

$$\frac{d}{ds}\Delta\phi = \Delta\omega + \left( \sqrt{\frac{J_y}{J_x}} - \sqrt{\frac{J_x}{J_y}} \right) \frac{K_\Delta}{2} \cos(\Delta\phi) = 0, \quad (7.161)$$

and:

$$J_x + J_y = J_0, \quad (7.162)$$

where  $J_0$  is a constant. The solution to these equations is:

$$J_x = \left( 1 + \frac{1}{\sqrt{1 + \kappa^2}} \right) \frac{J_0}{2}, \quad (7.163)$$

$$J_y = \left( 1 - \frac{1}{\sqrt{1 + \kappa^2}} \right) \frac{J_0}{2}, \quad (7.164)$$

where:

$$\kappa = \frac{K_\Delta}{\Delta\omega} \cos(\Delta\phi). \quad (7.165)$$

Equations (7.163) and (7.164) provide the formulae we need to write expressions for the equilibrium horizontal and vertical emittances determined by synchrotron radiation in a storage ring in the presence of betatron coupling. Recall that the emittance is given by the average of the betatron

action. We assume that we can identify the natural emittance of the storage ring (in the absence of betatron coupling) with the sum of the horizontal and vertical betatron action:

$$\langle J_0 \rangle = \epsilon_0. \quad (7.166)$$

Although the formulae (7.163) and (7.164) apply to a special case in which the difference between the betatron phases is constant, the fact that this leads to constant values for the horizontal and vertical betatron actions means that this represents a ‘fixed point’ in terms of the dynamics of a particle. On the principle that radiation effects will tend to dampen the motion of particles towards an equilibrium, we assume that the equilibrium horizontal and vertical emittances for a beam of particles are given by the betatron actions (7.163) and (7.164) with the maximum value of  $\kappa$ , i.e. with  $\cos(\Delta\phi) \approx 1$  so that  $\kappa = K_\Delta / \Delta\omega$ .

However, before writing explicit expressions for the equilibrium emittances, we need to consider some of the approximations that we have made in deriving the formulae (7.163) and (7.164), in particular, the fact that these results were derived on the assumption that we could ‘average’ the Hamiltonian around the storage ring. In practice, skew quadrupole fields will be localised, so even with the coupling strength  $K_\Delta$  calculated using (7.156) (and thus taking into account local variation in beta functions and phase advances) some modification to the final results will be needed. Since the momentum kick given to a particle by a skew quadrupole field depends on the betatron phases at the location of the skew quadrupole, the significant quantity in estimating the impact of a skew quadrupole over multiple turns is the fractional part of the phase advance (divided by  $2\pi$ ) over one turn, rather than the total phase advance.

We therefore define the coupling strength:

$$\hat{\kappa} = \frac{C_0}{2\pi} \frac{K_\Delta}{\text{frac}(\nu_x - \nu_y)}, \quad (7.167)$$

where  $\text{frac}(\nu_x - \nu_y)$  represents the fractional part of the difference in the betatron tunes. Finally, we then have, for the equilibrium horizontal and vertical emittances determined by synchrotron radiation in a storage ring, in the presence of betatron coupling:

$$\epsilon_x = \left(1 + \frac{1}{\sqrt{1 + \hat{\kappa}^2}}\right) \frac{\epsilon_0}{2}, \quad (7.168)$$

$$\epsilon_y = \left(1 - \frac{1}{\sqrt{1 + \hat{\kappa}^2}}\right) \frac{\epsilon_0}{2}. \quad (7.169)$$

Note that in the limit of weak coupling, these formulae give the expected behaviour:

$$\lim_{\hat{\kappa} \rightarrow 0} \epsilon_x = \epsilon_0, \quad \text{and} \quad \lim_{\hat{\kappa} \rightarrow 0} \epsilon_y = 0. \quad (7.170)$$

In the limit of strong coupling, we find:

$$\lim_{\hat{\kappa} \rightarrow \infty} \epsilon_x = \lim_{\hat{\kappa} \rightarrow \infty} \epsilon_y = \frac{1}{2} \epsilon_0. \quad (7.171)$$

In other words, with strong coupling the natural emittance is ‘shared’ equally between the horizontal and the vertical motion. An example of the variation in equilibrium emittances in an electron storage ring as functions of the tune split (the difference in the fractional betatron tunes) is shown in Fig. 7.4.

It should be noted that in general, the horizontal and vertical emittances defined by (7.168) and (7.169) are *not* the same as the normal mode emittances defined by the normal mode analysis discussed in Section 5.2. This is because in the present section, we have defined the emittances  $\epsilon_x$  and  $\epsilon_y$  in terms of the horizontal and vertical betatron actions  $J_x$  and  $J_y$ , which describe the motion of a particle in the absence of coupling; the skew quadrupole fields are treated as a perturbation to the uncoupled motion. This is apparent in the fact that the equations of motion (7.159) show the betatron actions varying with position around the storage ring, as a result of the skew quadrupole fields; since particle transport through a skew quadrupole field is symplectic, the normal mode action variables will be constants of motion, even including the effects of coupling from the skew quadrupole fields. The normal mode emittances  $\epsilon_I$ ,  $\epsilon_{II}$  (and  $\epsilon_{III}$ , considering the synchrotron motion as well as the betatron motion) are related to the normal mode action variables through (5.114), and should take coupling between the horizontal and vertical (and longitudinal) motion fully into account. Unfortunately, the possibility to make numerous different definitions of the emittances can lead to considerable confusion, particularly since the behaviour of the emittances depends on the definition. Before performing any beam dynamics measurements, calculations or simulations involving the emittances, it is advisable to make a clear statement of the definitions being used for the emittances.

As a final observation, it is worth noting that the behaviour of the betatron tunes described by the equations of motion (7.159) is consistent with the behaviour that we found in Section 6.6, where we used an approach based on transfer matrices to study the impact of a skew quadrupole in a storage ring. In deriving the equations of motion (7.159) we made a number

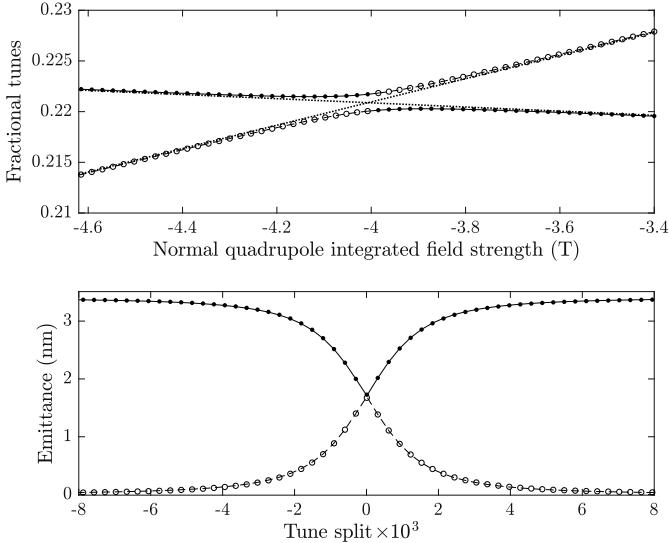


Fig. 7.4 Effect of skew quadrupoles on betatron tunes and equilibrium emittances in an electron storage ring. The storage ring lattice has a 24-cell double-bend achromat structure, and the beam energy is 3 GeV: the natural emittance is 3.4 nm. Two skew quadrupoles are placed in the lattice at locations of zero dispersion, to create betatron coupling without generating vertical dispersion. Normal quadrupole fields superposed on the (fixed) skew quadrupole fields are varied to change the betatron tunes. Top plot: horizontal (solid points, solid line) and vertical (open points, dashed line) fractional tunes as functions of normal quadrupole strength. Lines are calculated using (6.68) and (6.69). Dotted lines show the tunes with skew quadrupoles turned off. Bottom plot: horizontal (solid points, solid line) and vertical (open points, dashed line) emittances as functions of the difference in the fractional betatron tunes with skew quadrupole fields turned off. Points are calculated using (7.168) and (7.169). The lines are calculated using the envelope method, described in Section 7.4.3.

of assumptions and approximations; to show consistency with the results in Section 6.6, in particular equations (6.68) and (6.69), we need to make some further approximations and broad associations between the quantities involved. First, we note that it follows from (7.163) and (7.164) that for  $|\kappa| \ll 1$ :

$$\frac{J_y}{J_x} = \frac{\sqrt{1 + \kappa^2} - 1}{\sqrt{1 + \kappa^2} + 1} \approx \frac{\frac{1}{2}\kappa^2}{\frac{1}{2}\kappa^2 + 2} \approx \frac{\kappa^2}{4}. \quad (7.172)$$

Substituting into the first of equations (7.159), and using (7.165) we have:

$$\frac{d\phi_x}{ds} \approx \omega_x + \frac{K_\Delta^2}{4\Delta\omega} \cos^2(\Delta\phi). \quad (7.173)$$

If the ring contains a single skew quadrupole of integrated strength  $k_1^{(s)}L$ , then it follows from (7.156) that:

$$K_\Delta = \frac{\sqrt{\beta_x \beta_y} k_1^{(s)} L}{C_0}, \quad (7.174)$$

and hence:

$$\frac{d\phi_x}{ds} \approx \omega_x + \frac{\beta_x \beta_y \left(k_1^{(s)} L\right)^2}{4C_0^2 \Delta\omega} \cos^2(\Delta\phi). \quad (7.175)$$

Then, if we assume that the betatron frequency is roughly constant around the storage ring, we can write:

$$\frac{d\phi_x}{ds} \approx \frac{2\pi\nu'_x}{C_0}, \quad \omega_x \approx \frac{2\pi\nu_x}{C_0}, \quad \text{and} \quad \Delta\omega \approx \frac{2\pi\Delta\nu}{C_0}, \quad (7.176)$$

where  $\nu_x$  is the horizontal betatron tune with zero skew quadrupole strength,  $\nu'_x$  is the horizontal betatron tune when the effect of the skew quadrupole is included, and  $\Delta\nu = \nu_x - \nu_y$  is the difference in the betatron tunes with zero skew quadrupole strength. If we take the case  $\cos(\Delta\phi) = \pm 1$ , then we have finally, for the case  $|k_1^{(s)} L / \Delta\nu| \ll 1$ :

$$\nu'_x \approx \nu_x + \frac{\beta_x \beta_y \left(k_1^{(s)} L\right)^2}{16\pi^2 \Delta\nu}. \quad (7.177)$$

To compare this result with (6.68) (derived using an approach based on transfer matrices) we note that for  $|K/\Delta\nu| = |k_1^{(s)} L / \Delta\nu| \ll 1$ , we find:

$$\nu_I \approx \nu_x + \frac{\beta_x \beta_y \left(k_1^{(s)} L\right)^2}{16\pi^2 \Delta\nu}. \quad (7.178)$$

We can also consider the case close to the difference resonance,  $|\nu_x - \nu_y| \ll 1$ , for which we find:

$$\nu'_x \approx \nu_I \approx \nu_x + \frac{1}{4\pi} \sqrt{\beta_x \beta_y} k_1^{(s)} L. \quad (7.179)$$

Similar results to (7.177), (7.178) and (7.179) (though with opposite signs for the second term in each equation) are obtained for the vertical tune  $\nu'_y$  and normal mode II tune  $\nu_{II}$ . Although some caution is needed in identifying the horizontal and vertical betatron tunes  $\nu'_x$  and  $\nu'_y$  with the normal mode tunes  $\nu_I$  and  $\nu_{II}$  it is reasonable to assume that these results show that the tune shift estimated from the equations of motion (7.159) is consistent with the betatron tunes found from an analysis based on transfer matrices (Section 6.6), at least for the case of a single (weak) skew quadrupole in a storage ring without other sources of coupling.

### **7.4.2 Effects of vertical dispersion**

As we have seen in Section 7.2, quantum excitation leads to an increase in the amplitude of betatron and synchrotron oscillations of particles in an electron storage ring. The equilibrium emittance in each plane is determined by the balance between quantum excitation and radiation damping. Quantum excitation occurs when a particle at a location of non-zero dispersion in the storage ring emits a photon, leading to a change in energy of the particle. Because of the dispersion, the change in energy is accompanied by a corresponding change in the closed orbit for the particle. Even though the position and direction of motion of the particle do not change, the change in the closed orbit means that there will be a change in the betatron amplitude of the particle.

In a planar storage ring, the dispersion will be in the same plane as the orbit, which is generally the horizontal plane. This means that, in the absence of coupling, the vertical motion of particles will not be affected to any great extent by quantum excitation. The amplitude of vertical betatron oscillations will still be damped by synchrotron radiation, however, because the radiation from a particle is emitted predominantly along the instantaneous direction of motion of the particle, and it is the loss of momentum in the direction of motion of the particle that leads to the reduction of the amplitude of betatron and synchrotron oscillations. We therefore expect the equilibrium vertical emittance in a planar, uncoupled electron storage ring to be zero; however, the fact that synchrotron radiation is not emitted purely in the direction of motion of a particle, but with some (narrow) distribution around the direction of motion means that there will be quantum excitation of vertical betatron oscillations.

The equilibrium vertical emittance determined by the distribution of synchrotron radiation around the trajectory of a particle (given by Eq. (7.85)) represents a fundamental lower limit. In practice, however, the vertical emittance in an operational storage ring is determined by effects associated with magnet alignment and focusing errors, which lead to betatron coupling (for example, through the generation of skew quadrupole fields) and vertical dispersion. The vertical emittance generated by betatron coupling was discussed in Section 7.4.1; in the present section, we shall consider the vertical emittance generated by vertical dispersion.

Since dispersion describes the variation in the closed orbit in a storage ring with respect to the energy deviation, any effect that impacts the vertical closed orbit is likely also to contribute to vertical dispersion. For

example, vertical steering from the tilt of a dipole field or a vertical alignment error on a quadrupole magnet will generate vertical dispersion. If a skew quadrupole field is present at a location with non-zero horizontal dispersion, then vertical dispersion can be generated by coupling from the horizontal plane. In contrast to the horizontal dispersion, which is usually dominated by the layout of the main dipole magnets and the optics in the lattice and therefore varies in a regular fashion with the same periodicity as the lattice itself, the vertical dispersion in a planar storage ring tends to be irregular, since it arises from errors that occur randomly around the machine. However, since the physics in both cases are fundamentally the same, the vertical dispersion can be calculated using the same techniques as the horizontal dispersion. For example, we can generalise Eq. (5.15) as follows:

$$\begin{pmatrix} \eta_x \\ \eta_{px} \\ \eta_y \\ \eta_{py} \end{pmatrix} = \beta_0 (I - R_{[1:4,1:4]})^{-1} \begin{pmatrix} R_{16} \\ R_{26} \\ R_{36} \\ R_{46} \end{pmatrix}, \quad (7.180)$$

where  $\eta_x$ ,  $\eta_{px}$ ,  $\eta_y$  and  $\eta_{py}$  are the horizontal and vertical dispersion functions,  $\beta_0$  is the velocity (scaled by the speed of light) of a particle with the reference momentum,  $I$  is the  $4 \times 4$  identity matrix, and  $R$  is the  $6 \times 6$  single-turn transfer matrix, with  $4 \times 4$  submatrix  $R_{[1:4,1:4]}$  constructed from the elements in the first four rows and columns of  $R$ . Equation (7.180) gives the dispersion functions at the same point in the lattice to which the single-turn transfer matrix  $R$  applies. Once calculated at a particular location in a storage ring, the horizontal and vertical dispersion functions can be propagated around the ring using a generalisation of (5.16). The vertical dispersion may also be calculated using the techniques described in Section 5.3: see in particular Eqs. (5.154) and (5.155).

Once the vertical dispersion has been calculated (for a given set of magnet errors) in a storage ring, it is not too difficult to estimate the contribution of the dispersion to the vertical emittance. We start from Eq. (7.83) for the natural emittance in a planar storage ring, and generalise this formula to the vertical plane:

$$\epsilon_y = C_q \gamma_0^2 \frac{I_{5y}}{j_y I_2}. \quad (7.181)$$

Here,  $C_q$  is the quantum constant (7.80),  $\gamma_0$  is the relativistic factor for a particle with the reference momentum, and  $I_2$  is the second synchrotron radiation integral (7.19). The vertical damping partition number  $j_y$  is given

by a generalisation of the formula for the horizontal damping partition number (7.43):

$$j_y = 1 - \frac{I_{4y}}{I_2}, \quad (7.182)$$

where, from generalisation of (7.39):

$$I_{4y} = \oint \frac{\eta_y}{\rho} \left( \frac{1}{\rho^2} - 2k_1 \right) ds. \quad (7.183)$$

Note that  $\rho$  is the radius of curvature of the reference trajectory taking into account horizontal as well as vertical bending. Finally, the generalisation of the fifth synchrotron radiation integral (7.82) to the vertical plane is:

$$I_{5y} = \oint \frac{\mathcal{H}_y}{|\rho^3|} ds, \quad (7.184)$$

where, by generalisation of (7.74):

$$\mathcal{H}_y = \gamma_y \eta_y^2 + 2\alpha_y \eta_y \eta_{py} + \beta_y \eta_{py}^2. \quad (7.185)$$

Hence, if we know the vertical dispersion and Courant–Snyder parameters as functions of position around the storage ring, we can calculate the vertical emittance in the same way as we calculate the natural emittance. It should be noted, however, that this approach neglects betatron coupling (except where it enters through the vertical dispersion): the contribution of betatron coupling to the vertical emittance must be added separately, as discussed in Section 7.4.1.

If, as is often the case, the vertical dispersion is generated by random errors around the storage ring, then it is possible to make some approximations to arrive at a convenient formula to estimate the vertical emittance in terms of the mean square vertical dispersion (weighted by the reciprocal of the vertical beta function), and the natural energy spread. First, we can assume that the vertical dispersion is not correlated with the curvature of the reference trajectory (in contrast to the horizontal case, where it is generally the curvature of the reference trajectory that determines the horizontal dispersion). With this assumption, we can replace  $\mathcal{H}_y$  in (7.184) by its mean value around the storage ring:

$$I_{5y} \approx \langle \mathcal{H}_y \rangle \oint \frac{1}{|\rho^3|} ds = \langle \mathcal{H}_y \rangle I_3, \quad (7.186)$$

where we use the definition (7.93) to write the integral as the third synchrotron radiation integral. Hence, we have from (7.181):

$$\epsilon_y \approx C_q \gamma_0^2 \langle \mathcal{H}_y \rangle \frac{I_3}{j_y I_2}. \quad (7.187)$$

Using (7.94), this can be written in terms of the natural energy spread  $\sigma_{\delta 0}$ :

$$\epsilon_y \approx \frac{j_z}{j_y} \langle \mathcal{H}_y \rangle \sigma_{\delta 0}^2, \quad (7.188)$$

where  $j_z$  is the longitudinal damping partition number (7.60). To simplify further the calculation of the vertical emittance, we note the similarity in the forms of the expressions for  $\mathcal{H}_y$  (7.185) and the vertical betatron action:

$$2J_y = \gamma_y y^2 + 2\alpha_y y p_y + \beta_y p_y^2. \quad (7.189)$$

Since we can write the vertical co-ordinate  $y$  in terms of the action-angle variables  $J_y$  and  $\phi_y$ :

$$y = \sqrt{2\beta_y J_y} \cos(\phi_y), \quad (7.190)$$

the similarity between (7.185) and (7.189) implies that we can write the dispersion in terms of a phase angle  $\phi_{\eta y}$ :

$$\eta_y = \sqrt{\beta_y \mathcal{H}_y} \cos(\phi_{\eta y}). \quad (7.191)$$

Even without knowing how the phase angle  $\phi_{\eta y}$  varies around the ring, Eq. (7.191) provides us with an expression for the average of  $\eta_y^2/\beta_y$ :

$$\left\langle \frac{\eta_y^2}{\beta_y} \right\rangle = \frac{1}{2} \langle \mathcal{H}_y \rangle. \quad (7.192)$$

Finally, substituting for  $\langle \mathcal{H}_y \rangle$  into (7.188) gives:

$$\epsilon_y \approx 2 \frac{j_z}{j_y} \left\langle \frac{\eta_y^2}{\beta_y} \right\rangle \sigma_{\delta 0}^2. \quad (7.193)$$

Although Eq. (7.193) is not an exact expression for the vertical emittance, and neglects the contribution of betatron coupling to the vertical emittance, it does provide a convenient way to estimate the vertical emittance given a certain level of vertical dispersion in an electron storage ring. This can be useful for gauging the level to which the vertical dispersion in a machine must be corrected, to keep the vertical emittance within a specified limit. For more detailed studies, however, an exact calculation of the equilibrium emittances in an electron storage ring (including the effects of both betatron coupling and vertical dispersion) may be needed: we discuss one possible technique for accomplishing this task in the following section.

### 7.4.3 Envelope method

Although the results from Sections 7.4.1 and 7.4.2 provide some insight into the effects of betatron coupling and vertical dispersion, and can be useful for calculating the vertical emittance in a storage ring in particular cases, the complexity of modern accelerators and the range of magnet alignment errors and other imperfections that are usually present in real machines mean that a more general computational technique is needed in practice. A range of methods have been developed, including (for example) those of Chao [Chao (1979)] and Sagan [Sagan (2022)]. In this section, however, we present a method for computing the equilibrium emittances based on the ‘envelope method’ described by Ohmi, Hirata and Oide [Ohmi *et al.* (1994)].

Recall from Section 5.2 that, in the absence of synchrotron radiation effects, a distribution is said to be matched to a periodic lattice if it is invariant over transport of the particles through one period:

$$\Sigma = R\Sigma R^T, \quad (7.194)$$

where  $R$  is the transport matrix. We assume that there are no zeroth-order components in the map. In the present case, the reference trajectory is the same as the closed orbit, and  $R$  is the single-turn transfer matrix, which will be symplectic in the absence of synchrotron radiation (and certain other effects). Equation (7.194) is sufficient to determine the *shape* of the matched distribution in a storage ring (i.e. to determine the lattice functions) but is not sufficient to determine the *size* of the distribution, i.e. the beam emittances. This is clear from the fact that for any matrix  $\Sigma$  that satisfies (7.194) a matrix  $\tau\Sigma$  for any scalar  $\tau$  also satisfies (7.194).

In the presence of synchrotron radiation, we need to take account of radiation damping and quantum excitation. Radiation damping means that the single-turn transfer matrix is no longer symplectic. In that case, (7.194) has, in general, only the solution  $\Sigma = 0$ , i.e. the only matched distribution is one in which the beam occupies zero volume in phase space: this represents the limit of radiation damping without quantum excitation. However, quantum excitation can be represented as a zeroth-order term to be added to the right-hand side of (7.194), thus:

$$\Sigma = R\Sigma R^T + D, \quad (7.195)$$

where  $D$  is a constant matrix. In three degrees of freedom,  $D$  is a  $6 \times 6$  matrix. The matrices  $R$  and  $D$  are determined by the effects of the magnets

and the synchrotron radiation in the lattice, and (7.195) has (except in pathological cases) a unique solution for  $\Sigma$ .

In this section, we describe a method for computing the equilibrium emittances in a storage ring in the general case. The method consists of three steps. Firstly, the matrices  $R$  and  $D$  are constructed from a model of the lattice, taking into account the fields from electromagnetic components in the lattice, and radiation effects. Secondly, (7.195) is solved to find the equilibrium distribution  $\Sigma$ . We make the implicit assumption that radiation effects cause any arbitrary initial distribution to damp towards a unique matched distribution. Finally, the equilibrium emittances  $\epsilon_k$  are obtained from the eigenvalues  $\pm i\epsilon_k$  of  $\Sigma S$ .

Beginning with the construction of the matrices  $R$  and  $D$ , we note that the most important case is that of a short section (or ‘thin slice’) of a dipole. We consider this case in particular, because in most storage rings radiation effects are significant only in dipoles. Furthermore, long dipoles can be constructed by composition of the maps for a number of thin slices. The reason for considering a thin slice rather than a complete dipole will become apparent shortly. Once we have the map (including radiation effects) for a thin slice of dipole, we shall be able to construct the map (in the form of the matrices  $R$  and  $D$ ) for a complete revolution of the storage ring, starting from any given point. The construction of the single-turn map is achieved by composition of the maps for individual components.

To simplify the analysis, we shall work in the ultra-relativistic limit,  $\beta_0 \rightarrow 1$ . When a particle emits radiation carrying momentum  $dP$ , the phase space variables transform as:

$$x \mapsto x, \quad (7.196)$$

$$p_x \mapsto \left(1 - \frac{dP}{P_0}\right) p_x, \quad (7.197)$$

$$y \mapsto y, \quad (7.198)$$

$$p_y \mapsto \left(1 - \frac{dP}{P_0}\right) p_y, \quad (7.199)$$

$$z \mapsto z, \quad (7.200)$$

$$\delta \mapsto \delta - \frac{dP}{P_0}, \quad (7.201)$$

where  $P_0$  is the reference momentum. In general,  $dP$  is a function of the co-ordinates. To find how the beam distribution matrix  $\Sigma$  transforms through a thin slice of dipole, we find an explicit expression for  $dP/P_0$ , and then write down the above transformations to first order.

For an ultra-relativistic particle, the momentum lost through radiation in a short time interval  $dt$  can be expressed in terms of the synchrotron radiation power  $P_\gamma$  (the energy loss per unit time) as follows:

$$\frac{dP}{P_0} \approx \frac{P_\gamma}{E_0} dt \approx \frac{P_\gamma}{E_0} \left(1 + \frac{x}{\rho}\right) \frac{ds}{c}, \quad (7.202)$$

where  $\rho$  is the radius of curvature of the reference trajectory. The radiation power from a particle of charge  $q$  and energy  $E$  in a magnetic field  $B$  is given by (7.14):

$$P_\gamma = \frac{C_\gamma}{2\pi} c^3 q^2 E^2 B^2. \quad (7.203)$$

In general, the dipole may have a quadrupole component in the magnetic field, so we can write for the field:

$$B = B_0 + \frac{\partial B_y}{\partial x} x = B_0 + \frac{P_0}{q} k_1 x. \quad (7.204)$$

Also, the particle may have some energy deviation, so the total energy is:

$$E = E_0(1 + \delta). \quad (7.205)$$

Using (7.203), (7.204) and (7.205) in (7.202), we find that to first order in the dynamical variables:

$$\frac{dP}{P_0} \approx \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds + \frac{C_\gamma}{2\pi} \left( \frac{1}{\rho^2} + 2k_1 \right) E_0^3 \frac{x}{\rho} ds + \frac{C_\gamma}{2\pi} E_0^3 \frac{2\delta}{\rho^2} ds. \quad (7.206)$$

Then, the transformations of the dynamical variables are, to first order:

$$x \mapsto x, \quad (7.207)$$

$$p_x \mapsto \left(1 - \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds\right) p_x, \quad (7.208)$$

$$y \mapsto y, \quad (7.209)$$

$$p_y \mapsto \left(1 - \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds\right) p_y, \quad (7.210)$$

$$z \mapsto z, \quad (7.211)$$

$$\delta \mapsto \delta - \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds - \frac{C_\gamma}{2\pi} \left( \frac{1}{\rho^2} + 2k_1 \right) E_0^3 \frac{x}{\rho} ds - \frac{C_\gamma}{2\pi} E_0^3 \frac{2\delta}{\rho^2} ds. \quad (7.212)$$

This map represents the effect of radiation in a short length  $ds$  of dipole: note that it does not include the effect of transport of the beam through the dipole itself.

Let us write the map (7.207)–(7.212) for the phase space vector  $\vec{x}$  in vector notation:

$$\vec{x} \mapsto R_{\text{rad}} \vec{x} + \vec{d}, \quad (7.213)$$

where the matrix  $R_{\text{rad}}$  contains the first-order terms in the map, and the vector  $\vec{d}$  contains the zeroth-order terms. The beam distribution matrix  $\Sigma$  can be written as the outer product of the phase space vector averaged over all particles in the beam:

$$\Sigma = \langle \vec{x} \vec{x}^T \rangle. \quad (7.214)$$

It follows that, as a result of emission of radiation in a thin slice of dipole (of length  $ds$ ), the beam distribution matrix transforms as:

$$\Sigma \mapsto R_{\text{rad}} \Sigma R_{\text{rad}}^T + D_{\text{rad}}, \quad (7.215)$$

where:

$$D_{\text{rad}} = \vec{d} \vec{d}^T, \quad (7.216)$$

and we have assumed that:

$$\langle \vec{x} \rangle = 0, \quad (7.217)$$

i.e. the bunch centroid is on the reference trajectory. Note that the map (7.213) for  $\vec{x}$  has a non-zero zeroth-order term in the sixth component of  $\vec{d}$ : this term is of order  $ds$ , and leads to a non-zero component of order  $ds^2$  in the sixth row, sixth column of  $D_{\text{rad}}$ . In the limit  $ds \rightarrow 0$ , i.e. for radiation from an infinitesimal slice of dipole, the zeroth-order term vanishes compared to the first-order terms. However, this is based on a classical model for the radiation, in which the radiation field can carry arbitrarily small quantities of energy and momentum. In reality, the radiation field is quantised, with the energy carried in discrete units (photons). Working to zeroth order in the dynamical variables, we can write:

$$\frac{d}{ds} \langle \delta^2 \rangle \approx \frac{\langle \dot{N}(u) u^2 \rangle}{E_0^2 c}, \quad (7.218)$$

where  $\dot{N}(u) du$  is the number of photons emitted by a charged particle per unit time in the energy range  $u$  to  $u + du$ . Using the results already quoted, (7.79) for  $\langle \dot{N}(u) u^2 \rangle$  and (7.15) for  $P_\gamma$ , we find that to zeroth order in the dynamical variables:

$$\frac{d}{ds} \langle \delta^2 \rangle \approx 2C_q \gamma^2 \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^3}. \quad (7.219)$$

Hence, for radiation from a thin slice of dipole of (short) length  $ds$ :

$$d \langle \delta^2 \rangle \approx 2C_q \gamma^2 \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^3} ds + O(1), \quad (7.220)$$

where  $O(1)$  represents terms that are first order and higher in the dynamical variables. The first-order terms must be the same as those given by the classical model for the radiation, and can be obtained from (7.212). The significant feature of the quantum model for the radiation is that there is now a zeroth-order (in the dynamical variables) term in the map for  $\langle \delta^2 \rangle$  that is first order in  $ds$ . This term, given by (7.220), does not vanish in the limit  $ds \rightarrow 0$  compared to the terms that are first order in the dynamical variables.

Collecting the results together, we can write explicit expressions for the matrices  $R_{\text{rad}}$  and  $D_{\text{rad}}$ :

$$R_{\text{rad}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_{22} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & R_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ R_{61} & 0 & 0 & 0 & 0 & R_{66} \end{pmatrix}, \quad (7.221)$$

and:

$$D_{\text{rad}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & D_{66} \end{pmatrix}, \quad (7.222)$$

where the matrix elements are:

$$R_{22} = R_{44} = 1 - \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds, \quad (7.223)$$

$$R_{61} = -\frac{C_\gamma}{2\pi} \left( \frac{1}{\rho^2} + 2k_1 \right) \frac{E_0^3}{\rho} ds, \quad (7.224)$$

$$R_{66} = 1 - 2 \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^2} ds, \quad (7.225)$$

$$D_{66} = 2C_q \gamma^2 \frac{C_\gamma}{2\pi} \frac{E_0^3}{\rho^3} ds. \quad (7.226)$$

Having obtained the linear map for the distribution matrix  $\Sigma$  for emission of synchrotron radiation from a thin slice of dipole, we can proceed to construct the full linear map for  $\Sigma$  over a complete revolution of the storage ring. This involves composition of the maps, and can be done in a straightforward fashion numerically, using a computer. However, some care is needed in handling the zeroth-order part of the map. For example,

given the distribution matrix  $\Sigma(s_0)$  at a location  $s = s_0$  along the reference trajectory, we find the distribution matrix at a location  $s_1 = s_0 + ds$  using:

$$\Sigma(s_1) = R(s_1, s_0)\Sigma(s_0)R(s_1, s_0)^T + D(s_1, s_0). \quad (7.227)$$

Then the distribution matrix  $\Sigma(s_2)$  at  $s = s_2$  is given by:

$$\begin{aligned} \Sigma(s_2) &= R(s_2, s_1)\Sigma(s_1)R(s_2, s_1)^T + D(s_2, s_1) \\ &= R(s_2, s_1)R(s_1, s_0)\Sigma(s_0)R(s_1, s_0)^T R(s_2, s_1)^T \\ &\quad + R(s_2, s_1)D(s_1, s_0)R(s_2, s_1)^T + D(s_2, s_1). \end{aligned} \quad (7.228)$$

But we can also write:

$$\Sigma(s_2) = R(s_2, s_0)\Sigma(s_0)R(s_2, s_0)^T + D(s_2, s_0). \quad (7.229)$$

Hence:

$$R(s_2, s_0) = R(s_2, s_1)R(s_1, s_0), \quad (7.230)$$

and:

$$D(s_2, s_0) = R(s_2, s_1)D(s_1, s_0)R(s_2, s_1)^T + D(s_2, s_1). \quad (7.231)$$

The map for a short distance  $ds$  through a dipole can be constructed by composition of a map representing synchrotron radiation over  $ds$ , with a map representing symplectic transport through the dipole field over distance  $ds$ . The map for a full dipole is constructed by composition of maps for short lengths  $ds$ .

The second step in finding the equilibrium distribution in a storage ring is to solve the equation representing the matching condition (7.195):

$$\Sigma = R\Sigma R^T + D, \quad (7.232)$$

where the matrices  $R$  and  $D$  (including the effects both of radiation and of symplectic transport around the ring) can be obtained as described above. The solution for  $\Sigma$  proceeds as follows. Let  $U$  be a matrix constructed of the eigenvectors of  $R$ , and  $\Lambda$  a matrix of the eigenvalues of  $R$ , so that:

$$RU = U\Lambda. \quad (7.233)$$

Now define the matrices  $\tilde{\Sigma}$  and  $\tilde{D}$  by:

$$\Sigma = U\tilde{\Sigma}U^T, \quad (7.234)$$

$$D = U\tilde{D}U^T. \quad (7.235)$$

The equilibrium distribution matrix  $\Sigma$  can be found from:

$$\tilde{\Sigma}_{ij} = \frac{\tilde{D}_{ij}}{1 - \Lambda_i \Lambda_j}, \quad (7.236)$$

where  $\tilde{\Sigma}_{ij}$  are the elements of  $\tilde{\Sigma}$ ,  $\tilde{D}_{ij}$  are the elements of  $\tilde{D}$ , and  $\Lambda_i$  are the diagonal elements of  $\Lambda$ .

The final step in obtaining the equilibrium emittances is simply to compute the eigenvalues of  $\Sigma S$ , where  $S$  is the antisymmetric matrix defined by (4.14): the eigenvalues are  $\pm i\epsilon_k$ , where  $\epsilon_k$  are the emittances.

It is worth making a few remarks concerning the mechanism by which the emittances are generated. The envelope method for calculating the equilibrium emittances in a general storage ring lattice (including the effects, for example, of vertical dispersion and betatron coupling) emphasises the fact that whereas radiation damping acts on the horizontal, vertical and longitudinal degrees of freedom, quantum excitation acts *directly* only on the longitudinal motion. This is emphasised by the fact that the only non-zero component in the zeroth-order part of the map for the beam distribution  $\Sigma$  corresponds to excitation of the energy spread. Fundamentally, when a particle emits radiation, the statistical ‘noise’ arising from the random emission of the radiation in discrete quanta leads to a change in the mean square energy deviation. To put it in another way, if the radiation emission were completely deterministic (as in a classical model) so that all particles emitted radiation in the same way, then if all the particles in a beam had exactly the same energy at the entrance of a dipole, all the particles would have exactly the same energy at the exit of the dipole. The fact that the radiation emission is different for different particles (even though the particles may have the exact same energy, and see the exact same magnetic field) can lead to a change in the mean square energy deviation.

However, the excitation of the energy spread from random emission of photons has no direct impact on the horizontal or vertical motion. If the vertical motion is completely decoupled from the horizontal and the longitudinal motion then, in the approximation that the radiation emission is directly along the instantaneous direction of motion of a particle, there is no excitation of the vertical betatron action for any particle, and the vertical emittance will damp to zero. The same would be true of the horizontal motion if it could be completely decoupled from the longitudinal. However, in a storage ring, the dipole magnets inevitably introduce coupling between the horizontal and longitudinal motion, since the deflection of a particle in a dipole field depends on the energy of the particle. Thus, any excitation of the energy spread will also have an impact on the horizontal distribution. In other words, quantum excitation at a location where there is horizontal dispersion acts (through the dispersion) to excite horizontal betatron motion as well as the energy spread.

Where coupling is present, it is often preferable to describe the motion in terms of the normal modes, where each normal mode corresponds to oscillations at a single well-defined frequency. If the coupling is not too strong, then one can associate each of the normal modes with predominantly horizontal (mode I), vertical (mode II) or longitudinal (mode III) motion. Motion in a normal mode will be excited by emission of radiation if the phase space axes for that normal mode have a component parallel to the energy deviation. Broadly speaking, if there is horizontal dispersion present, then one expects the mode I motion to have a longitudinal component, and radiation emission at a location where there is horizontal dispersion will then excite mode I emittance. Similarly, radiation emission where vertical dispersion is present will excite mode II emittance. If there is betatron coupling present, then horizontal motion corresponds to a combination of mode I and mode II motion; in that case, if the horizontal dispersion is non-zero, both mode I and mode II will be coupled to the longitudinal motion even in the absence of vertical dispersion. So radiation emission at a location where there is horizontal dispersion and betatron coupling will, in general, excite mode II emittance. The important point is that the generation of mode II emittance depends on the coupling in the dipole magnets, where radiation is emitted: it is quite possible to have strong betatron coupling in a storage ring and not generate any mode II emittance, as long as the betatron coupling in the dipoles is zero (and there is also no vertical dispersion in the dipoles).

## 7.5 Synchrotron Radiation and Spin Polarisation

In Chapter 3, we briefly discussed the spin dynamics of particles in accelerators. The magnetic field generated by a charged particle at rest is described by the magnetic moment of the particle. The relationship between the magnetic moment and the spin angular momentum of a particle is given by the gyromagnetic ratio. If a charged particle is placed in an external magnetic field, then the particle experiences a torque that results in a precession (the Larmor precession) of the angular momentum about an axis parallel to the external field. Transforming the spin precession from the rest frame of a particle to a frame in which the particle is moving with some velocity through a given electromagnetic field leads to the Thomas–BMT equation (3.169), which can be used to describe the evolution of the spin of a particle moving along an accelerator beam line.

Considering only the Larmor precession, the angle between the magnetic moment of the particle and the axis defined by the external field will remain constant. To change this angle requires a change in the energy of the particle,  $dU$ , given by:

$$dU = T d\theta = \mu B \sin(\theta) d\theta, \quad (7.237)$$

where  $T$  is the magnitude of the torque (3.166),  $\mu$  is the magnitude of the magnetic moment of the particle,  $B$  is the magnetic field strength, and  $\theta$  is the angle between the field and the magnetic moment. The lowest energy of the system occurs when the magnetic moment is parallel to the field: for positrons, this happens when the spin of the particle is parallel to the field; for electrons, it happens when the spin is antiparallel to the field. The difference in energy between the two spin states leads, in the case of a beam of electrons or positrons in a storage ring, to an equilibrium spin polarisation of the beam, in which the majority of the particles are aligned parallel (for positrons) or antiparallel (for electrons) to the field in the main dipoles. The equilibrium spin polarisation is reached through the emission of synchrotron radiation, as we now describe.

When an electron or positron emits synchrotron radiation, its spin state will usually remain the same, although since photons carry spin angular momentum, there is some non-zero probability for the spin state to change. A particle with magnetic moment antiparallel to the magnetic field is more likely to reverse its spin than a particle with magnetic moment parallel to the field, since the state with magnetic moment parallel to the field has lower energy. This means that, over time, an electron beam in a storage ring will become polarised with particle spins predominantly antiparallel to the field in the dipoles, while a positron beam will become polarised with particle spins predominantly parallel to the field in the dipoles. The polarisation of an electron or positron beam by the emission of synchrotron radiation is known as the *Sokolov–Ternov effect* [Sokolov and Ternov (1964)]. We shall not discuss the Sokolov–Ternov effect in detail, but give here, for reference, some of the principal results. A more complete overview is given in [Barber and Ripken (1999)].

For a bunch of electrons in a storage ring with the spins of the particles aligned (randomly) parallel or antiparallel to the field in the dipoles, the transition rate for spins changing from parallel to antiparallel to the field by photon emission is:

$$\Gamma_{\uparrow\downarrow} = W_{\uparrow\downarrow} N_{\uparrow}, \quad (7.238)$$

where  $N_{\uparrow}$  is the number of particles in the bunch with spin parallel to the dipole field, and:

$$W_{\uparrow\downarrow} = \frac{5\sqrt{3}}{16} \frac{e^2 \gamma^5 \hbar}{4\pi \varepsilon_0 m^2 c^2 |\rho|^3} \left(1 + \frac{8}{5\sqrt{3}}\right). \quad (7.239)$$

In (7.239),  $e$  is the magnitude of the charge on the electron,  $\gamma$  is the relativistic factor of the electrons,  $m$  is the mass of the electron,  $\rho$  is the bending radius of the trajectory of an electron in the field of a dipole, and  $\varepsilon_0$  is the permittivity of free space. The transition rate for spins changing from antiparallel to parallel is:

$$\Gamma_{\downarrow\uparrow} = W_{\downarrow\uparrow} N_{\downarrow}, \quad (7.240)$$

where  $N_{\downarrow}$  is the number of electrons in the bunch with spin antiparallel to the dipole field, and:

$$W_{\downarrow\uparrow} = \frac{5\sqrt{3}}{16} \frac{e^2 \gamma^5 \hbar}{4\pi \varepsilon_0 m^2 c^2 |\rho|^3} \left(1 - \frac{8}{5\sqrt{3}}\right). \quad (7.241)$$

If there are no other effects changing the spins of the particles, then the beam reaches an equilibrium polarisation state when:

$$\Gamma_{\uparrow\downarrow} = \Gamma_{\downarrow\uparrow}. \quad (7.242)$$

The polarisation of the beam is defined as:

$$P_{\uparrow} = \frac{N_{\uparrow} - N_{\downarrow}}{N_{\uparrow} + N_{\downarrow}}. \quad (7.243)$$

The evolution of the polarisation from any given initial state can be obtained as follows. The rate of change of  $N_{\uparrow}$  is given by the rate at which particles change spin from antiparallel (to the field in the dipoles) to parallel, minus the rate at which particles change spin from parallel to antiparallel:

$$\frac{dN_{\uparrow}}{dt} = W_{\downarrow\uparrow} N_{\downarrow} - W_{\uparrow\downarrow} N_{\uparrow}. \quad (7.244)$$

Similarly:

$$\frac{dN_{\downarrow}}{dt} = W_{\uparrow\downarrow} N_{\uparrow} - W_{\downarrow\uparrow} N_{\downarrow}. \quad (7.245)$$

Hence, we can write the rate of change of the polarisation:

$$\frac{dP_{\uparrow}}{dt} = \frac{2}{N_b} (W_{\downarrow\uparrow} N_{\downarrow} - W_{\uparrow\downarrow} N_{\uparrow}), \quad (7.246)$$

where  $N_b$  is the total number of electrons in the bunch. Since:

$$N_{\uparrow} + N_{\downarrow} = N_b, \quad (7.247)$$

it follows from (7.243) that:

$$N_{\uparrow} = \frac{N_b}{2}(1 + P_{\uparrow}), \quad (7.248)$$

$$N_{\downarrow} = \frac{N_b}{2}(1 - P_{\uparrow}). \quad (7.249)$$

Substituting for  $N_{\uparrow}$  and  $N_{\downarrow}$  in (7.246) gives:

$$\frac{dP_{\uparrow}}{dt} = \frac{P_{\text{ST}} - P_{\uparrow}}{\tau_{\text{ST}}}, \quad (7.250)$$

where:

$$P_{\text{ST}} = \frac{W_{\downarrow\uparrow} - W_{\uparrow\downarrow}}{W_{\downarrow\uparrow} + W_{\uparrow\downarrow}} = -\frac{8}{5\sqrt{3}}, \quad (7.251)$$

and (averaging the dipole fields around the circumference  $C_0$  of the storage ring):

$$\frac{1}{\tau_{\text{ST}}} = W_{\downarrow\uparrow} + W_{\uparrow\downarrow} = \frac{5\sqrt{3}}{8} \frac{e^2 \gamma^5 \hbar}{4\pi \varepsilon_0 m^2 c^2} \frac{1}{C_0} \int_0^{C_0} \frac{ds}{|\rho|^3}. \quad (7.252)$$

The solution to (7.250) is:

$$P_{\uparrow}(t) = P_{\uparrow}(0)e^{-t/\tau_{\text{ST}}} + P_{\text{ST}} \left(1 - e^{-t/\tau_{\text{ST}}}\right), \quad (7.253)$$

where  $P_{\uparrow}(0)$  is the spin polarisation of the beam at time  $t = 0$ . Starting from any initial spin polarisation, the beam approaches an equilibrium polarisation  $P_{\text{ST}}$ , at an exponential rate  $\tau_{\text{ST}}^{-1}$ . The same result for the equilibrium polarisation can be obtained from (7.242), using (7.238), (7.240) and (7.247). The results (7.251) and (7.252) assume that changes to the particle spins occur only as a result of synchrotron radiation. In practice, there are likely to be other effects that will have some impact on the polarisation rate and the equilibrium polarisation, especially if the spin tune is close to an integer (as discussed in Chapter 3).

The equilibrium beam polarisation  $P_{\text{ST}}$  is independent of the beam energy and the dipole field: if the spin polarisation is dominated by the Sokolov–Ternov effect, then the beam in an electron storage ring will always reach an equilibrium polarisation of about 92%, antiparallel to the field in the main dipoles. In a positron storage ring, the beam will reach the same level of polarisation, but with the direction of the polarisation parallel to the field in the main dipoles (the transition rates for positrons are given by the same expressions as in (7.239) and (7.241), but with the signs reversed). The rate of change of the spin polarisation, however, does depend on the beam energy and on the dipole field. The value of  $\tau_{\text{ST}}$  can

vary widely depending on the storage ring parameters, but typical values range from some tens of minutes to several hours. In proton storage rings, as might be expected from the very much smaller amounts of synchrotron radiation generated compared to electron storage rings, spin polarisation times are extremely long, and the Sokolov–Ternov effect is not significant.

The spin polarisation state of a beam has some impact on the rate of scattering between particles within a bunch, and hence affects the Touschek lifetime. This provides the basis for a technique based on resonant depolarisation for making very precise measurements of the beam energy [Zholentz *et al.* (1980); Melissinos (1995); Klein *et al.* (1997); Minty and Zimmermann (2003c)]. The fact that a beam of particles in an electron or positron storage ring acquires a degree of spin polarisation naturally through the Sokolov–Ternov effect means that the resonant depolarisation technique can be applied even in storage rings where polarisation is not specifically required for other purposes. The Touschek effect will be discussed further in Chapter 13.

**This page intentionally left blank**

### PART III

## Single-Particle Nonlinear Dynamics

**This page intentionally left blank**

## Chapter 8

# Examples of Nonlinear Effects in Accelerator Beam Lines

The analysis of beam dynamics in previous chapters was based on linear approximations to the equations of motion for particles in an accelerator. While this provides an important foundation for understanding beam behaviour, many significant effects can only be described using nonlinear equations. It is the goal of this and the following chapters to develop the tools needed to analyse and understand nonlinear aspects of beam dynamics.

We begin by discussing, in some detail, two examples of accelerator systems in which nonlinear dynamics play an important role. We shall then, in later chapters, proceed to develop some useful mathematical tools for analysis of nonlinear dynamics: the approach we have taken to linear dynamics, based on Hamilton's equations, provides a good foundation for some powerful mathematical techniques, in particular the Lie algebra methods. We shall show how these methods can be applied to construct transfer maps for nonlinear beam line components. We shall also discuss some numerical methods that are complementary to the algebraic techniques, and that can be of great value in the characterisation and optimisation of accelerator performance.

### 8.1 Longitudinal Dynamics in a Bunch Compressor

As a first example, we shall derive the equations for the longitudinal beam dynamics in a bunch compressor. The purpose of a bunch compressor is to manipulate the longitudinal phase space distribution of a bunch; usually, the goal is to increase the peak current by reducing the bunch length. Some kinds of accelerator systems (for example, free-electron lasers) require a high peak current in order to operate, but in particle sources at low energy, a

high peak current can make the beam sensitive to collective effects that degrade beam quality. The solution is to generate long bunches at the source, and use a bunch compressor to increase the peak current once the beam is at high energy and less susceptible to instabilities.

The two main subsystems in a bunch compressor are: first, a set of radiofrequency (rf) cavities that introduce a correlation between particle energy and longitudinal position within the bunch; and second, a set of dipole magnets that (by means of dispersion) lead to a variation in the length of the path followed by a particle, depending on the energy of the particle. Suppose that the particles in a bunch are travelling at relativistic velocities, so that their speed is essentially unchanged if their energy changes by a small amount. The phase of the rf cavities is set so that, after passing through the cavities, particles near the tail of the bunch have a higher energy than those near the head. Then, by arranging the dipoles in a chicane, particles with a higher energy (at the tail of the bunch) take a shorter path than those with a lower energy (at the head of the bunch). The result is that, after passing through the chicane, particles at the tail of the bunch have moved towards the head, and particles at the head have moved towards the tail: there is a reduction in the total length of the bunch.

Let us consider the system in more detail, starting with the rf cavities. We shall assume that the particles in the beam are travelling at relativistic velocities, so that all particles travel at effectively the same speed  $c$ , as long as their energy deviation  $\delta$  is not too large. Therefore, assuming that a particle stays close to the reference trajectory when passing along a section of beam line with a straight reference trajectory, the final value of the longitudinal co-ordinate  $z(s_1)$  is equal to the initial value  $z(s_0)$ :

$$z(s_1) = z(s_0), \quad (8.1)$$

where  $s = s_0$  is at the start of the section of beam line, and  $s = s_1$  is at the exit of the section. Now suppose that this straight section of beam line contains a number of rf cavities. On passing through the cavities the energy deviation changes so that the final energy deviation  $\delta(s_1)$  is related to the initial energy deviation  $\delta(s_0)$  by:

$$\delta(s_1) = \delta(s_0) + \frac{qV_{\text{rf}}}{E_0} \sin\left(\frac{\omega_{\text{rf}}z(s_0)}{c} + \phi_{\text{rf}}\right), \quad (8.2)$$

where  $q$  is the charge of the particle,  $V_{\text{rf}}$  the total voltage amplitude of the cavities,  $\omega_{\text{rf}}$  is the rf frequency,  $\phi_{\text{rf}}$  a fixed phase offset (independent of the particle co-ordinates), and  $E_0$  is the reference energy.

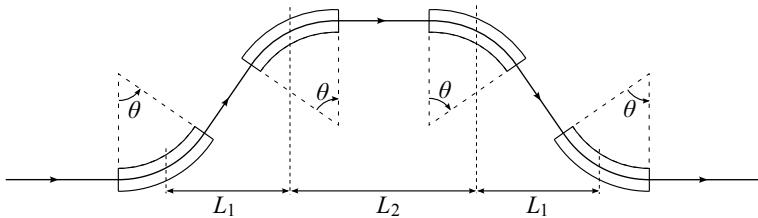


Fig. 8.1 Layout of dipole magnets in a chicane for a bunch compressor. The bending angle within each dipole is  $\theta$ . The distances along the reference trajectory between the centres of the magnets are (for small bending angle and short magnets)  $L_1/\cos(\theta)$  (from the first to the second, and from the third to the fourth dipoles) and  $L_2$  (from the second to the third dipole).

Now let us consider the motion of particles through the dipoles. We shall look only at the case where there are four dipoles with equal bending angles arranged in a chicane, as shown in Fig. 8.1. The total path length of a particle through the chicane is:

$$L = \frac{2L_1}{\cos(\theta)} + L_2, \quad (8.3)$$

where we neglect the lengths of the dipoles themselves;  $L_1$  is the distance from the first to the second dipole and from the third to the fourth dipole; and  $L_2$  is the distance from the second to the third dipole. The bending angle  $\theta$  is a function of the energy deviation:

$$\theta = \frac{\theta_0}{1 + \delta}, \quad (8.4)$$

where  $\theta_0$  is the bending angle for a particle with the reference energy. Neglecting effects such as synchrotron radiation, the energy deviation of the particles stays constant in the dipoles:

$$\delta(s_2) = \delta(s_1), \quad (8.5)$$

where  $s_2$  is a point on the reference trajectory at the exit of the chicane. However, the longitudinal co-ordinate changes from  $z(s_1)$  to  $z(s_2)$ , where:

$$z(s_2) = z(s_1) + 2L_1 \left( \frac{1}{\cos(\theta_0)} - \frac{1}{\cos(\theta)} \right). \quad (8.6)$$

Note that  $\theta$  depends on  $\delta(s_1)$ , which in turn depends on  $z(s_0)$ :  $z(s_2)$  therefore has an implicit dependence on  $z(s_0)$ .

Equations (8.1), (8.2), (8.5) and (8.6) constitute the transfer map for the bunch compressor. Clearly, the map is nonlinear; however, the impact of the nonlinearities depends on the parameter regime. To proceed, we shall

first consider only the linear terms: this will guide us in an appropriate choice of parameters, from which we can then determine the impact of the nonlinear effects.

Making linear approximations to equations (8.1), (8.2), (8.5) and (8.6), and considering only the longitudinal motion, the complete transfer map for the bunch compressor can be written in matrix form as:

$$\begin{pmatrix} z \\ \delta \end{pmatrix}_{s=s_2} = \begin{pmatrix} 1 & R_{56} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ R_{65} & 1 \end{pmatrix} \begin{pmatrix} z \\ \delta \end{pmatrix}_{s=s_0}, \quad (8.7)$$

where:

$$R_{65} = -\frac{qV_{\text{rf}} \cos(\phi_{\text{rf}})}{E_0} \frac{\omega_{\text{rf}}}{c}, \quad (8.8)$$

and:

$$R_{56} = 2L_1 \frac{\theta_0 \tan(\theta_0)}{\cos(\theta_0)}. \quad (8.9)$$

The transfer matrix  $R$  has longitudinal part:

$$R_{\parallel} = \begin{pmatrix} 1 & R_{56} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ R_{65} & 1 \end{pmatrix} = \begin{pmatrix} 1 + R_{65}R_{56} & R_{56} \\ R_{65} & 1 \end{pmatrix}. \quad (8.10)$$

Note that  $R$ , being the product of two symplectic matrices, is itself symplectic, and satisfies:

$$R^T S R = S. \quad (8.11)$$

The beam distribution matrix  $\Sigma = \langle \vec{x} \vec{x}^T \rangle$  has longitudinal part:

$$\Sigma_{\parallel} = \begin{pmatrix} \langle z^2 \rangle & \langle z\delta \rangle \\ \langle z\delta \rangle & \langle \delta^2 \rangle \end{pmatrix}. \quad (8.12)$$

Recall that the beam distribution matrix transforms under a transfer matrix  $R$  as:

$$\Sigma(s_2) = R\Sigma(s_0)R^T. \quad (8.13)$$

We shall assume that we can neglect the effects of coupling between the transverse and longitudinal degrees of freedom, so that when transforming the beam distribution we need consider only the longitudinal parts of the transfer matrix and the beam distribution matrix.

Usually, a bunch compressor is designed so that there is zero correlation between the longitudinal co-ordinates of the particles and their energy deviations, both at the start of the bunch compressor and at the exit:

$$\langle z(s_2)\delta(s_2) \rangle = \langle z(s_0)\delta(s_0) \rangle = 0. \quad (8.14)$$

In that case, using equations (8.10) and (8.13) we find that the parameters  $R_{65}$  and  $R_{56}$  must be related by:

$$\frac{R_{65}}{R_{56}}(1 + R_{65}R_{56}) = -\frac{\langle \delta(s_0)^2 \rangle}{\langle z(s_0)^2 \rangle}. \quad (8.15)$$

Given this constraint, the compression factor  $f$  is given by:

$$f \equiv \sqrt{\frac{\langle z(s_2)^2 \rangle}{\langle z(s_0)^2 \rangle}} = \sqrt{1 + R_{65}R_{56}}. \quad (8.16)$$

As an example, consider a bunch compressor designed to reduce the rms bunch length  $\sigma_z$  from 5 mm to 0.5 mm (so that  $f = 1/10$ ), with initial rms energy spread  $\sigma_\delta = 0.5\%$ . Using equations (8.15) and (8.16), we find that appropriate values for the parameters  $R_{56}$  and  $R_{65}$  are:

$$R_{56} \approx 0.09950 \text{ m}, \quad (8.17)$$

$$R_{65} \approx -9.950 \text{ m}^{-1}. \quad (8.18)$$

To illustrate the longitudinal dynamics, we show an artificial distribution in longitudinal phase space, in which particles are arranged in three lines, with each line having a different energy deviation ( $-\sigma_\delta$ , 0 and  $\sigma_\delta$ ), and extending from  $-\sigma_z$  to  $+\sigma_z$ . Figure 8.2 shows the effect of the transformation, using the full nonlinear transfer map given by equations (8.1), (8.2), (8.5) and (8.6). For this example, we have chosen a beam energy  $E_0 = 100$  MeV, an rf frequency  $\omega_{\text{rf}}/2\pi = 1.3$  GHz, rf phase  $\phi_{\text{rf}} = 0$ , and a dipole bending angle  $\theta_0 = 200$  mrad. From (8.8) and (8.17), we find  $qV_{\text{rf}} = -36.5$  MeV; and from equations (8.9) and (8.18) we have  $L_1 = 1.20$  m.

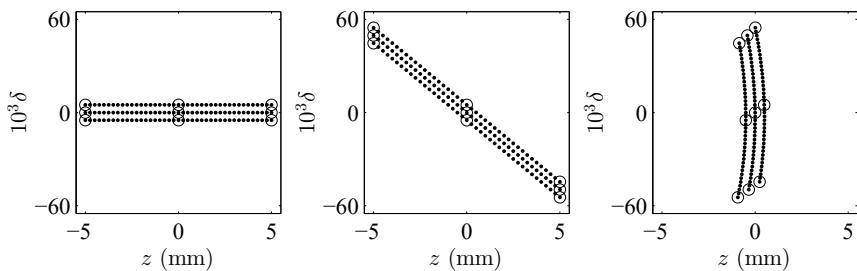


Fig. 8.2 Transformations in longitudinal phase space in a bunch compressor. The phase space is shown at the entrance to the bunch compressor (left), after the rf cavities (centre) and after the chicane (right). Particles with initial longitudinal positions  $z = \pm 5$  mm or  $z = 0$  are circled to emphasise the longitudinal phase advance (rotation in phase space). The parameters of the bunch compressor are designed to reduce the bunch length by a factor of 10.

The rf cavities give a change in energy deviation depending on longitudinal co-ordinate; and the chicane gives a change in longitudinal co-ordinate depending on energy deviation. The resulting transformation has the effect of rotating the distribution in phase space through an angle close to  $90^\circ$ , while simultaneously reducing the length of the bunch by a factor of 10. As well as the desired effect of the reduction in bunch length, we observe that the energy spread is increased by a factor of 10. This is an inevitable consequence of the fact that the transformation is symplectic: the area of phase space covered by the distribution remains the same. In other words, the longitudinal emittance is conserved.

Closer inspection of the final phase space distribution in Fig. 8.2 reveals a further undesirable effect: there is a distortion in the final shape of the phase space distribution, which will lead to a final bunch length somewhat longer than the intended 0.5 mm. This is more evident if we reproduce the final phase space plot from Fig. 8.2 on a different scale: see Fig. 8.3. The distortion results from a quadratic dependence of the final value of the coordinate  $z(s_2)$  on the initial value  $z(s_0)$ : this is an effect of nonlinearities in the transfer map, and by modifying the map (adding additional components to the beam line) there are ways to mitigate the distortion. To understand how to do this, instead of making a linear approximation to the transfer map we expand the map to second order in the dynamical variables. In

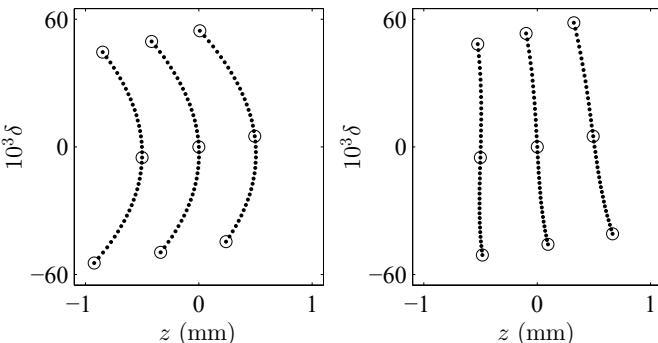


Fig. 8.3 Longitudinal phase space at the exit of a bunch compressor, without nonlinear correction (left), and with nonlinear correction using third-harmonic rf cavities (right). The particle positions in the left-hand and right-hand plots are the same as in the final plots in Figs. 8.2 and 8.4 (respectively), but the scale on the  $z$  axis is changed to show the nonlinear distortion of the distribution more clearly. The third-harmonic cavities reduce the quadratic terms in the transfer map significantly, but there are still some effects from terms of third order and higher.

principle, we can expand the map to whatever order we like; but since the dominant nonlinearity in this case appears to be second order, expanding to second order should be sufficient. In a general form, the expansion can be written:

$$z(s_2) = R_{55}z(s_0) + R_{56}\delta_0 + T_{555}z(s_0)^2 + 2T_{556}z(s_0)\delta(s_0) + T_{566}\delta(s_0)^2, \quad (8.19)$$

$$\delta(s_2) = R_{65}z(s_0) + R_{66}\delta_0 + T_{655}z(s_0)^2 + 2T_{656}z(s_0)\delta(s_0) + T_{666}\delta(s_0)^2, \quad (8.20)$$

where the coefficients  $R_{ij}$  and  $T_{ijk}$  (following the TRANSPORT notation [Brown *et al.* (1977)]) are functions of the parameters of the system, such as the voltage and frequency of the rf cavities, the dipole bending angle, and the spacing between dipoles in the chicane. In particular,  $R_{56}$  and  $R_{65}$  are given by equations (8.9) and (8.8), respectively. The notation is such that the transformation of the  $i$ th component of the phase space vector  $\vec{x}$  can be written:

$$x_i(s_1) = q_i + \sum_{j=1}^6 R_{ij}x_j(s_0) + \sum_{j,k=1}^6 T_{ijk}x_j(s_0)x_k(s_0) \quad (8.21)$$

$$+ \sum_{j,k,l=1}^6 U_{ijkl}x_j(s_0)x_k(s_0)x_l(s_0) + \dots \quad (8.22)$$

Coefficients of the fourth-order terms are denoted  $V_{ijklm}$ , and so on. By symmetry, coefficients that are related by interchanging any except the first index are equal to each other, so that (for example)  $T_{ijk} = T_{ikj}$ .

In our example of a bunch compressor, the nonlinear distortion in the phase space appears to be dominated by the term with coefficient  $T_{555}$ : this gives the quadratic dependence of the final co-ordinate  $z(s_2)$  on the initial co-ordinate  $z(s_0)$ . Carrying out the Taylor expansion of the transfer map, we find:

$$T_{555} = -\frac{1}{2}R_{56}R_{65}^2(2 + \theta_0 \cot(\theta_0) + 2\theta_0 \tan(\theta_0)). \quad (8.23)$$

Since the values of  $R_{65}$  and  $R_{56}$  are fixed by the required compression factor, it appears that the only way to reduce the value of  $T_{555}$ , and hence reduce the nonlinear distortion of the final phase space, is to reduce the dipole bending angle  $\theta_0$ . However, for fixed  $R_{56}$ , reducing  $\theta_0$  leads to an increase in  $L_1$ : generally, it is desirable (for economic reasons) to keep the length of the bunch compressor as short as possible. We therefore need to look at alternative ways to mitigate the nonlinear effects. One possibility is

to introduce additional rf cavities that operate at a higher frequency (and lower voltage) than the main rf cavities. Usually, the higher frequency is a harmonic (integer multiple) of the main rf frequency. For example, if we use third-harmonic cavities, the energy change of a particle passing through the rf section will be:

$$\delta(s_1) = \delta(s_0) + \frac{qV_{\text{rf}}}{E_0} \sin\left(\frac{\omega_{\text{rf}} z(s_0)}{c}\right) + \frac{qV_{\text{rf}}^{(3)}}{E_0} \cos\left(\frac{3\omega_{\text{rf}} z(s_0)}{c}\right), \quad (8.24)$$

where we have made a particular (and, as we shall see, appropriate) choice of phase for the main and third-harmonic rf cavities. With the modified map, the coefficient  $T_{555}$  becomes:

$$T_{555} = -\frac{1}{2} R_{56} R_{65}^2 \left( 2 + \theta_0 \cot(\theta_0) + 2\theta_0 \tan(\theta_0) + \frac{9}{R_{65}^2} \frac{qV_{\text{rf}}^{(3)}}{E_0} \frac{\omega_{\text{rf}}^2}{c^2} \right). \quad (8.25)$$

We see that if we choose the total voltage  $V_{\text{rf}}^{(3)}$  of the third-harmonic rf cavities such that:

$$\begin{aligned} qV_{\text{rf}}^{(3)} &= -E_0 \frac{R_{65}^2}{9} \frac{c^2}{\omega_{\text{rf}}^2} (2 + \theta_0 \cot(\theta_0) + 2\theta_0 \tan(\theta_0)) \\ &= -\frac{(qV_{\text{rf}})^2}{9E_0} (2 + \theta_0 \cot(\theta_0) + 2\theta_0 \tan(\theta_0)), \end{aligned} \quad (8.26)$$

then  $T_{555} = 0$ ; that is, we cancel completely the second-order term that makes a dominant contribution to the nonlinear distortion of the longitudinal phase space. With the parameters given above (in particular,  $E_0 = 100$  MeV,  $qV = -36.5$  MeV and  $\theta_0 = 200$  mrad), then we find that the required voltage for the third-harmonic cavities is  $qV_{\text{rf}}^{(3)} = -4.55$  MeV. In this example, we have chosen the phases of the rf cavities so that a reference particle arrives at the main cavities at the zero-crossing of the voltage, and at the third-harmonic cavities at the minimum voltage. It is possible to operate the cavities at different phases, but for operation of the system just for bunch compression, there is no advantage in doing so.

Figure 8.4 shows the transformations in longitudinal phase space when the third-harmonic rf cavities are included with the appropriate voltage needed to cancel the  $T_{555}$  term in the transfer map. The curvature apparent in the final phase space plot in Fig. 8.2 is no longer visible. Close inspection of Fig. 8.4 shows that there is still some distortion from nonlinear terms in the transfer map; however, we have made no attempt to cancel any coefficients other than  $T_{555}$ . The residual nonlinear effects appear relatively small, and should not significantly affect the performance of the bunch compressor.

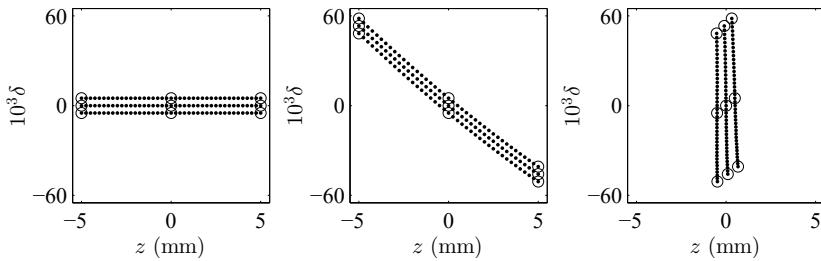


Fig. 8.4 Transformations in longitudinal phase space in a bunch compressor, with nonlinear correction using third-harmonic cavities. The phase space is shown at the entrance to the bunch compressor (left), after the rf (centre) and after the chicane (right). Particles with initial longitudinal positions  $z = \pm 5$  mm or  $z = 0$  are circled. Apart from the presence of the third-harmonic cavities, the parameters of the bunch compressor are the same as in Fig. 8.2.

It is worth making some remarks as to the physical origin of the damaging  $T_{555}$  term in the transfer map. We see from equation (8.23) that  $T_{555}$  has a strong dependence on  $\theta_0$ , the bending angle of the dipoles in the chicane. This suggests that the origin of this term is in the chicane; and this is indeed the case. However, the ultimate origin of the problem is really the  $T_{566}$  term in the transfer map for the chicane: the large  $R_{65}$  term *for the transfer map across the rf cavities* gives a strong dependence of the energy deviation on the longitudinal co-ordinate, and the  $T_{566}$  term *for the chicane* describes the dependence of the final longitudinal co-ordinate on the energy deviation of a particle entering the chicane. In other words, the  $T_{555}$  term for the full bunch compressor results from a combination of the  $R_{65}$  term for the rf cavities and the  $T_{566}$  term for the chicane. Introducing third-harmonic cavities leads to a contribution to the  $T_{655}$  term in the rf section of the bunch compressor: combined with the  $R_{56}$  in the chicane, this gives an independent contribution to the  $T_{555}$  term for the full bunch compressor. By choosing an appropriate voltage for the third-harmonic cavities, the two contributions to the  $T_{555}$  term can be made to cancel.

Techniques other than the use of harmonic rf cavities are possible to correct the  $T_{555}$  term in a bunch compressor. In some situations, it is possible to operate the rf cavities at a phase so that the reference particle arrives off the zero-crossing of the rf voltage. However, this can lead to significant beam *deceleration*, and is not always a practicable technique. The best technique to use (and even whether any correction at all is really necessary) will depend on the specification and particular design features of the bunch compressor being considered.

We close this section with a few general remarks. We have used the example of a bunch compressor to show how nonlinear effects can have an impact on the particle dynamics in an accelerator beam line. The severity of nonlinear effects can be strongly dependent on the operating regime and parameter values for a given system. Strong nonlinearities can impose performance limitations. In the example of a bunch compressor, if we had aimed for a compression factor much greater than 10, then the nonlinear effects would quickly have become more significant and more difficult to mitigate. Attempts to cancel damaging nonlinear terms can, in some cases, drive the scale and cost of accelerator beam lines beyond practical limits.

In the example we have discussed, analysis of the nonlinear effects in the transfer map was fairly straightforward. With some approximations, we were able to write down expressions for the transformations in the rf cavities and the dipoles, which could be readily expressed as power series expansions to any desired order in the dynamical variables. Inspection of the nonlinear transformations in phase space indicated the dominant nonlinear term, and there was a relatively simple way (at least in principle) to cancel the dominant term by introducing additional components into the system. Although power series expansions of the transfer map are a useful tool in the analysis of nonlinear dynamics in accelerators, it is not often the case that the power series can be calculated as easily, and the nonlinearities mitigated in so straightforward a fashion as in this example. A more serious drawback in some situations is that by truncating the power series representation of a transfer map at some arbitrary (and necessarily finite) order, the symplecticity of the map is lost. For single-pass systems such as bunch compressors, this is not always a significant problem; however, for long-term tracking in storage rings it can be important to retain symplecticity, and then more powerful mathematical techniques are required to provide accurate representations of the nonlinear dynamics. In the following chapters, we shall describe some techniques that can often be useful in such cases.

## **8.2 Chromaticity in a Linear FODO Beam Line**

Consider a straight periodic beam line consisting of repeated identical FODO cells. The deflection of a particle by a quadrupole magnet in the beam line depends on the trajectory of the particle as it enters the quadrupole, the length of the quadrupole, the magnetic field in the quadrupole, and on the energy of the particle. The dependence of the

particle motion on the energy of the particle will lead to a variation of the lattice functions with energy. In particular, the betatron phase advance per cell will vary with energy: this variation is known as *chromaticity*.

If we describe the dynamics using linear transfer maps, then it appears that the betatron phase advance is independent of energy. In other words, it seems that there is no chromaticity: this is because chromaticity is, strictly speaking, a higher-order effect. However, there are different ways of treating the effects of an energy variation, and although they must all describe the same physical behaviour, they lead to somewhat different interpretations that can sometimes cause confusion. In this section, we shall analyse the chromaticity in a simple FODO beam line in two different ways: first, by using linear transfer maps and considering a variation in the reference momentum; and second, by writing the maps so that they include the transverse variables to first order, but maintain an exact dependence on the energy deviation. The second technique is often useful where the energy of the particles is constant, and does not need to be treated as a dynamical variable.

To simplify the analysis, we shall consider explicitly only the horizontal motion. In a FODO beam line, the vertical motion is essentially the same as the horizontal. We shall also take the ultra-relativistic limit  $\beta_0 \rightarrow 1$ , so that we can neglect terms in  $1/\gamma_0$ . With these simplifications, the transfer matrix for a drift space of length  $L$  can be written:

$$R_{\text{drift}}(L) = \begin{pmatrix} 1 & L & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (8.27)$$

where the dynamical variables are  $(x, p_x, z, \delta)$ . The transfer matrix for a thin (length  $\rightarrow 0$ ) quadrupole is:

$$R_{\text{quad}}(K) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -K & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (8.28)$$

where  $K$  is the integrated quadrupole gradient scaled by the particle charge and reference momentum:

$$K = \frac{q}{P_0} \int_0^{L_{\text{quad}}} \frac{\partial B_y}{\partial x} ds. \quad (8.29)$$

The transfer matrix for a periodic (FODO) cell in the beam line can be obtained by multiplying the matrices:

$$\begin{aligned} R_{\text{FODO}} &= R_{\text{drift}}(L) R_{\text{quad}}(-K) R_{\text{drift}}(L) R_{\text{quad}}(K) \\ &= \begin{pmatrix} 1 - KL - K^2 L^2 & (2 + KL)L & 0 & 0 \\ -K^2 L & 1 + KL & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned} \quad (8.30)$$

Since the transfer matrices are all block diagonal, there is no coupling between the horizontal and longitudinal motion, and we could in principle work only with the  $2 \times 2$  matrices for the horizontal motion. However, we write the transfer maps as  $4 \times 4$  matrices to emphasise the fact that we are treating the longitudinal variables  $z$  and  $\delta$  as dynamical variables.

We can use the fact that the transfer matrix  $R_{\text{FODO}}$  for one FODO cell is block diagonal to calculate the horizontal phase advance  $\mu_x$  across one cell. Using (4.20), we can write:

$$\cos(\mu_x) = 1 - \frac{1}{2} K^2 L^2. \quad (8.31)$$

Note that  $\mu_x$  has no dependence on the energy deviation  $\delta$ : strictly speaking, the value of  $\mu_x$  that we obtain in this way is valid for particles with  $\delta = 0$ , i.e. for particles with momentum exactly equal to the reference momentum. However, it appears that the phase advance is the same for all particles, whatever their energy. But we can also consider changing the reference momentum  $P_0$ . This will change the value of the quadrupole strength  $K$ , and will therefore change the phase advance. Using (8.29), we find that:

$$\frac{dK}{dP_0} = -\frac{K}{P_0}. \quad (8.32)$$

Then, using (8.31):

$$\frac{d}{dP_0} \cos(\mu_x) = -\sin(\mu_x) \frac{d\mu_x}{dP_0} = -KL^2 \frac{dK}{dP_0} = \frac{K^2 L^2}{P_0} = \frac{2(1 - \cos(\mu_x))}{P_0}. \quad (8.33)$$

Finally, we can write:

$$P_0 \frac{d\mu_x}{dP_0} = -\frac{2(1 - \cos(\mu_x))}{\sin(\mu_x)} = -2 \tan\left(\frac{\mu_x}{2}\right). \quad (8.34)$$

We define the horizontal chromaticity  $\xi_x$  in a beam line as:

$$\xi_x = \frac{P_0}{2\pi} \frac{d\mu_x}{dP_0}. \quad (8.35)$$

The vertical chromaticity  $\xi_y$  can be similarly defined in terms of the variation of vertical phase advance  $\mu_y$  with respect to momentum. From (8.34) the horizontal chromaticity in a FODO cell with horizontal phase advance  $\mu_x$  is:

$$\xi_x = -\frac{1}{\pi} \tan\left(\frac{\mu_x}{2}\right). \quad (8.36)$$

Since the phase advance is always positive, the chromaticity is always negative: an increase in momentum always leads to a reduction in the phase advance per cell. This is easily understood from the fact that for a fixed field gradient the focusing strength of a quadrupole gets smaller as the momentum of particles in a beam increases. This is a general property of accelerator beam lines: any beam line constructed only from quadrupoles, drift spaces and dipoles always has negative chromaticity. In the case of a FODO cell, we see that the chromaticity increases roughly linearly with the phase advance if the phase advance is not too large; but as the phase advance approaches  $180^\circ$ , the chromaticity starts to increase more and more rapidly, approaching an infinite value as  $\mu_x \rightarrow \pi$ .

Now let us consider the same beam line; but instead of treating the energy deviation  $\delta$  as a dynamical variable, we shall treat it as a parameter in the transfer maps. This is possible because the energy deviation remains constant as a particle moves along the beam line. Considering only the horizontal motion, and taking the ultra-relativistic limit, the Hamiltonian for motion in a quadrupole is:

$$H = \delta - \sqrt{(1 + \delta)^2 - p_x^2} + \frac{1}{2} k_1 x^2. \quad (8.37)$$

Expanding to second order in the dynamical variables ( $x$  and  $p_x$ ), and taking the limit that the length  $L_{\text{quad}}$  of the quadrupole approaches zero while  $k_1 L_{\text{quad}}$  remains fixed ( $k_1 L_{\text{quad}} = K$ ), we find that the transfer map can be written in matrix form:

$$R_{\text{quad}}(K) = \begin{pmatrix} 1 & 0 \\ -K & 1 \end{pmatrix}, \quad (8.38)$$

which is exactly the same as the horizontal part of the transfer matrix (8.28). In other words, it looks as though the dynamics still have no dependence on the momentum of the particles. However, let us now consider a drift space. The exact transfer map for a drift space is given in equations (3.13), (3.15) and (3.17). Expanding to first order in the dynamical variables  $x$  and  $p_x$ , making the approximation  $\beta_0 \approx 1$ , and neglecting terms in  $1/\gamma_0$ , we can write the transfer matrix for a drift space of length  $L$  as:

$$R_{\text{drift}}(L) = \begin{pmatrix} 1 & \frac{L}{1+\delta} \\ 0 & 1 \end{pmatrix}. \quad (8.39)$$

We can construct the transfer matrix for a FODO cell by multiplying the matrices for the quadrupoles and the drift spaces, as we did before:

$$R_{\text{FODO}} = \begin{pmatrix} 1 - \frac{KL}{1+\delta} - \frac{K^2 L^2}{(1+\delta)^2} & \left(2 + \frac{KL}{1+\delta}\right) \frac{L}{1+\delta} \\ -\frac{K^2 L}{1+\delta} & 1 + \frac{KL}{1+\delta} \end{pmatrix}. \quad (8.40)$$

We then find:

$$\cos(\mu_x) = 1 - \frac{K^2 L^2}{2(1+\delta)^2}, \quad (8.41)$$

and:

$$\lim_{\delta \rightarrow 0} \frac{d\mu_x}{d\delta} = -2 \tan\left(\frac{\mu_x}{2}\right). \quad (8.42)$$

Using the definition (2.55) of the energy deviation  $\delta$ :

$$\delta = \frac{E}{P_0 c} - \frac{1}{\beta_0}, \quad (8.43)$$

we can say that in the ultra-relativistic limit:

$$\Delta\delta = \frac{\Delta E}{P_0 c} \approx \frac{\Delta P_0}{P_0}. \quad (8.44)$$

Hence, we find for the horizontal chromaticity:

$$\xi_x = \frac{P_0}{2\pi} \frac{d\mu_x}{dP_0} \approx \frac{1}{2\pi} \frac{d\mu_x}{d\delta} = -\frac{1}{\pi} \tan\left(\frac{\mu_x}{2}\right). \quad (8.45)$$

We obtain the same expression for the chromaticity in a FODO beam line using either of the two approaches we have considered: first, treating  $\delta$  as a dynamical variable, and finding the change in the phase advance with respect to the reference momentum; and second, treating the energy deviation  $\delta$  as a parameter in the transfer map, and considering the change in the phase advance with respect to  $\delta$ .

A curious feature of this analysis is that in one case the source of the chromaticity appears to be the quadrupoles; and in the other case the source appears to be the drift spaces. This rather surprising result can be understood in terms of the physical significance of the dynamical variables. In particular, the variable  $p_x$  is defined (2.52) so that in a drift space (with zero vector potential):

$$p_x = \frac{\gamma m v_x}{P_0}, \quad (8.46)$$

where  $v_x$  is the horizontal component of the velocity. Using the Hamiltonian (with  $\delta$  as a parameter, in the ultra-relativistic limit) for a drift space:

$$H = \delta - \sqrt{(1+\delta)^2 - p_x^2}, \quad (8.47)$$

we find:

$$\frac{dx}{ds} = \frac{\partial H}{\partial p_x} = \frac{p_x}{\sqrt{(1+\delta)^2 - p_x^2}} \approx \frac{p_x}{1+\delta}. \quad (8.48)$$

Consider two particles with different momenta entering a thin quadrupole, with both particles having the same value for the  $x$  co-ordinate, and both having  $p_x = 0$ . Despite their different momenta, both particles are deflected such that they have the same value of  $p_x$  as they leave the quadrupole. But from (8.48) we see that the particle with the higher momentum follows a trajectory at a shallower angle than the particle with the lower momentum: in other words, the deflection angle from the quadrupole is smaller for the higher momentum particle, despite the fact that both particles have the same change in  $p_x$ .

### 8.3 Chromaticity in Storage Rings

In a storage ring, chromaticity leads to a variation of the betatron tunes with particle energy. This can have a number of adverse consequences. If the chromaticity is large enough, then the tunes of particles with even quite small energy deviations can reach integer or half-integer values: as we saw in Chapter 6, the dynamics of particles with tunes close to an integer or half-integer are very sensitive to steering or focusing errors. Certain types of beam instability are also dependent on the chromaticity: an example is the head-tail effect, which we shall discuss in Chapter 15. Beam stability, especially at high current, can depend on having a chromaticity close to zero (or even slightly positive). It is therefore important to be able to predict and control the chromaticity in a storage ring.

An expression for the chromaticity in a linear FODO beam line was derived in Section 8.2. In this section, a more general expression for the chromaticity in a storage ring, in terms of the lattice functions and magnet strengths, will be derived. We shall then discuss the use of sextupole magnets for controlling the chromaticity. Unfortunately, introducing sextupoles for chromatic correction in a storage ring has some undesirable consequences for other aspects of the dynamics, so we conclude this section by briefly considering some of the potential adverse impacts on the beam dynamics in a storage ring from nonlinear magnets such as sextupoles.

Consider a storage ring consisting of dipoles, quadrupoles and drift spaces. In the thin lens approximation, the transfer map for the horizontal

variables in a quadrupole can be represented to first order by a matrix:

$$R_{\text{quad}}(K, P_0) = \begin{pmatrix} 1 & 0 \\ -K & 1 \end{pmatrix}, \quad (8.49)$$

where  $K$  is the integrated strength of the quadrupole (8.29), which depends on the reference momentum  $P_0$ . Since we are working with a linear approximation to the dynamics in which the Hamiltonian is expanded to second-order in the dynamical variables, we drop terms in the Hamiltonian that describe a variation in focusing with energy deviation (i.e. terms including products  $x^2\delta$  and  $y^2\delta$ ). Therefore, to calculate the chromaticity in a storage ring, we model the effect of a change in the energy of a particle by considering a change in the reference momentum, while keeping a fixed energy deviation  $\delta = 0$ . Using the result:

$$\frac{dK}{dP_0} = -\frac{K}{P_0}, \quad (8.50)$$

we can write for a small variation  $\Delta P_0$  in the reference momentum:

$$R_{\text{quad}}(K, P_0 + \Delta P_0) = \begin{pmatrix} 1 & 0 \\ -K \left(1 - \frac{\Delta P_0}{P_0}\right) & 1 \end{pmatrix}. \quad (8.51)$$

In other words, the effect of a small increase in the reference momentum can be represented as a focusing error  $\Delta K = -K\Delta P_0/P_0$ . Using equation (6.33), we can immediately write down the contribution of the quadrupoles to the horizontal chromaticity in a storage ring:

$$\xi_x = P_0 \frac{d\nu_x}{dP_0} = -\frac{1}{4\pi} \oint \beta_x k_1 ds, \quad (8.52)$$

where  $\nu_x$  is the horizontal tune. The integral is taken around the entire circumference of the ring, and  $k_1$  is the usual field gradient scaled by the particle charge and reference momentum:

$$k_1 = \frac{q}{P_0} \frac{\partial B_y}{\partial x}. \quad (8.53)$$

We similarly find for the vertical chromaticity:

$$\xi_y = P_0 \frac{d\nu_y}{dP_0} = \frac{1}{4\pi} \oint \beta_y k_1 ds. \quad (8.54)$$

Note that  $k_1$  is positive in horizontally focusing quadrupoles, and negative in vertically focusing quadrupoles. Also, the beta function in each plane will be larger (on average) in quadrupoles that focus, rather than defocus, in that plane. Thus, the total contribution of the quadrupoles to both the horizontal and vertical chromaticity will be negative.

The transfer matrix for a drift space has no dependence on the reference momentum, and so the drift spaces in a storage ring make no contribution to the chromaticity. However, there is a contribution from the dipoles, which can be estimated by comparing the Hamiltonian for a dipole with the Hamiltonian for a quadrupole. In the ultra-relativistic limit, the Hamiltonian for a quadrupole expanded to second order in the dynamical variables is (3.94):

$$H_{\text{quad}} = \frac{p_x^2}{2} + \frac{p_y^2}{2} + \frac{k_1}{2}(x^2 - y^2). \quad (8.55)$$

With the same approximations, the Hamiltonian for a dipole is (3.54):

$$H_{\text{dipole}} = \frac{p_x^2}{2} + \frac{p_y^2}{2} + (k_0 - h(1 + \delta))x + \frac{hk_0}{2}x^2 + \frac{k_1}{2}(x^2 - y^2), \quad (8.56)$$

where  $h = 1/\rho$  is the curvature of the reference trajectory (with radius of curvature  $\rho$ ). The Hamiltonian for a dipole contains horizontal focusing terms (i.e. terms in  $x^2$ ) that will lead to chromaticity in a similar way to the focusing terms in the Hamiltonian for a quadrupole. In a dipole, however, the focusing depends on the field  $k_0$  and the curvature  $h$  of the reference trajectory (which lead to weak focusing), as well as on any field gradient (i.e. strong focusing, given by  $k_1$ ). The dipole Hamiltonian has an additional term that is linear in the horizontal co-ordinate  $x$ : however, this term leads to a variation in the trajectory of a particle with energy, and does not affect the focusing directly (in other words, it is associated with dispersion rather than chromaticity).

To find the contribution of weak focusing in a dipole to the chromaticity in a storage ring, we can consider the effect of a small change in the reference momentum, in the same way as we did for a quadrupole. For any strong focusing, represented by the term in  $k_1$  in the Hamiltonian, this leads to the same result for a dipole as for a quadrupole. Regarding the weak focusing, the scaled field strength  $k_0$  in a dipole, defined by:

$$k_0 = \frac{q}{P_0}B_y, \quad (8.57)$$

varies with the reference momentum in the same way as  $k_1$ :

$$\frac{dk_0}{dP_0} = -\frac{k_0}{P_0}. \quad (8.58)$$

Although the curvature of the reference trajectory  $h$  is, in principle, independent of the reference momentum, when making a change in the reference momentum in the present case (while keeping the energy deviation fixed at

$\delta = 0$ ) for consistency, we need to make a corresponding change in  $h$ . This is because a particle with zero energy deviation should follow a non-dispersive trajectory: if the dispersion term in the Hamiltonian (8.56) vanishes at the original reference momentum for a particle with  $\delta = 0$ , it should still vanish if we make a small change in the reference momentum  $P_0$ , but keep  $\delta = 0$ . Hence, the change in  $k_0$  (8.58) should be accompanied by a change in the curvature:

$$\frac{dh}{dP_0} = -\frac{h}{P_0}. \quad (8.59)$$

Thus, the change in the weak focusing term in the Hamiltonian (8.56) resulting from a change in the reference momentum is given by:

$$\frac{d}{dP_0} hk_0 = -\frac{2hk_0}{P_0}. \quad (8.60)$$

Then, to take account of the contribution of weak focusing in the dipoles to the chromaticity, we simply need to add another term to the integral in (8.52):

$$\xi_x = -\frac{1}{4\pi} \oint \beta_x (2hk_0 + k_1) ds. \quad (8.61)$$

In a strong focusing lattice, the contribution to the chromaticity from weak focusing in the dipoles is usually much smaller than that from strong focusing in the quadrupoles (and from any field gradient in the dipoles), and a good approximation for the chromaticity of a storage ring can be found just from (8.52). No change is necessary to the expression for the vertical chromaticity (8.54) to take account of the dipoles, since weak focusing in the dipoles (neglecting fringe field effects) does not contribute to the vertical focusing.

As mentioned at the start of this section, the chromaticity in a storage ring can have undesirable consequences. When designing a storage ring, it is important to keep the chromaticity generated by the dipole and quadrupole fields as small as possible: this can be achieved, for example, by keeping the beta functions and quadrupole strengths as small as possible. However, there are always design constraints that impose limitations: for example, final focus systems in a collider invariably require strong quadrupoles placed at locations where the beta functions are large, in order to focus the beams as tightly as possible at the interaction point. Therefore, additional methods are required to control the chromaticity; usually, this control must be achieved without affecting the lattice functions (the closed orbit, beta functions and dispersion). Fortunately, it is possible to use sextupole magnets

to control the chromaticity without having a significant impact on the linear dynamics.

The magnetic field in a sextupole has components:

$$b_x = k_2 xy, \quad (8.62)$$

$$b_y = \frac{1}{2} k_2 (x^2 - y^2), \quad (8.63)$$

$$b_z = 0, \quad (8.64)$$

where  $(b_x, b_y, b_z) = \mathbf{b} = q\mathbf{B}/P_0$  is the magnetic field scaled by the particle charge and the reference momentum. The sextupole strength  $k_2$  is related to the field in the sextupole by:

$$k_2 = \frac{q}{P_0} \frac{\partial^2 B_y}{\partial x^2}. \quad (8.65)$$

The field (8.62)–(8.64) can be obtained from the vector potential:

$$\mathbf{a} = \frac{q}{P_0} \mathbf{A} = \left( 0, 0, -\frac{1}{6} k_2 (x^3 - 3xy^2) \right). \quad (8.66)$$

The magnetic axis of the sextupole is defined by the line  $x = y = 0$ . Along the magnetic axis, the field is zero. However, suppose that the closed orbit in a storage ring passes through a sextupole magnet with some offset from the magnetic axis. If  $\bar{x}$  and  $\bar{y}$  are the co-ordinates with respect to the closed orbit, which is at  $x_0$  and  $y_0$  with respect to the magnetic axis, then the transverse components of the field can be written:

$$b_x = k_2 (x_0 + \bar{x})(y_0 + \bar{y}), \quad (8.67)$$

$$b_y = \frac{1}{2} k_2 ((x_0 + \bar{x})^2 - (y_0 + \bar{y})^2). \quad (8.68)$$

We shall be interested in the case that  $y_0 = 0$ . Then, the field is:

$$b_x = k_2 x_0 \bar{y} + k_2 \bar{x} \bar{y}, \quad (8.69)$$

$$b_y = \frac{1}{2} k_2 x_0^2 + k_2 x_0 \bar{x} + \frac{1}{2} k_2 (\bar{x}^2 - \bar{y}^2). \quad (8.70)$$

With respect to the closed orbit, the field has a dipole component of strength  $\frac{1}{2} k_2 x_0^2$ , a quadrupole component with strength  $k_2 x_0$ , and a sextupole component with strength  $k_2$ . The quadrupole component is of particular interest: this will provide focusing of the beam, which contributes to the chromaticity.

Now let us suppose that the sextupole is positioned in the storage ring so that the magnetic axis lies along the reference trajectory, and that the closed orbit coincides with the reference trajectory. Suppose further that

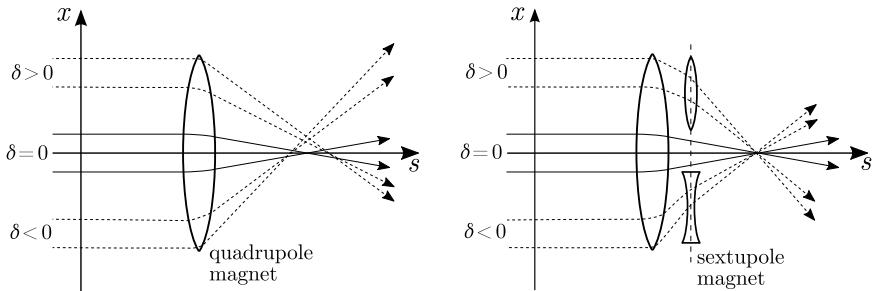


Fig. 8.5 Correction of chromaticity in a quadrupole using a sextupole magnet. Without a sextupole (left), particles with higher energy (positive energy deviation  $\delta$ ) are focused less strongly than particles with lower energy ( $\delta < 0$ ), leading to a variation of focal point position with particle energy. A sextupole magnet (right) has focusing strength that varies with transverse position: thus, if the particle energy is correlated with transverse position (e.g. through dispersion), the sextupole can provide additional focusing for particles with  $\delta > 0$ , and introduce defocusing for particles with  $\delta < 0$ , effectively correcting the variation in focal strength of the quadrupole with particle energy.

there is some horizontal dispersion  $\eta_x$  at the location of the sextupole: the system is illustrated schematically in Fig. 8.5. Then, if we change the reference energy from  $P_0$  to  $P_0 + \Delta P_0$ , the closed orbit will move so that:

$$x_0 = \eta_x \frac{\Delta P_0}{P_0}. \quad (8.71)$$

The beam will now see a quadrupole field at the location of the sextupole, with strength:

$$\Delta k_1 = \left. \frac{\partial B_y}{\partial \bar{x}} \right|_{\bar{x}=\bar{y}=0} = \eta_x k_2 \frac{\Delta P_0}{P_0}. \quad (8.72)$$

The quadrupole field  $\Delta k_1$  can be regarded as a focusing error that occurs when the beam momentum is changed from  $P_0$  to  $P_0 + \Delta P_0$ . The change in focusing with respect to momentum is:

$$P_0 \frac{dk_1}{dP_0} = \eta_x k_2. \quad (8.73)$$

Hence, the sextupole will make a contribution  $\beta_x \eta_x k_2 L_{\text{sext}}$  to the chromaticity, where  $\beta_x$  is the horizontal beta function at the location of the sextupole, and  $L_{\text{sext}}$  is the length of the sextupole. Including the dipole, quadrupole and sextupole contributions, the total horizontal chromaticity can be written:

$$\xi_x = -\frac{1}{4\pi} \oint \beta_x (2hk_0 + k_1 - \eta_x k_2) ds, \quad (8.74)$$

and the total vertical chromaticity can be written:

$$\xi_y = \frac{1}{4\pi} \oint \beta_y(k_1 - \eta_x k_2) ds. \quad (8.75)$$

A sextupole with  $k_2 > 0$  at a location with  $\eta_x > 0$  makes the horizontal chromaticity more positive; but the vertical chromaticity is made more negative. However, by using sextupoles with  $k_2 > 0$  at locations where  $\beta_x > \beta_y$ , and sextupoles with  $k_2 < 0$  where  $\beta_y > \beta_x$  (and assuming that  $\eta_x > 0$  in both cases), it is possible to achieve overall zero chromaticity in a storage ring in both the horizontal and vertical planes simultaneously.

Unfortunately, sextupoles have effects on the beam dynamics other than just an effect on the chromaticity. The full dynamics in a sextupole can be described by the Hamiltonian:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\delta + \frac{1}{\beta_0}\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2} + \frac{1}{6} k_2(x^3 - 3xy^2)}. \quad (8.76)$$

This is obtained from (2.73) with  $h = \phi = 0$ , and scaled vector potential  $\mathbf{a}$  given by (8.66). Unfortunately, this Hamiltonian is not integrable: it is not possible to write down solutions to the resulting equations of motion in closed form. We encountered the same situation when dealing with dipoles, quadrupoles and solenoids. In those cases, we addressed the problem by expanding the Hamiltonian to second order in the dynamical variables. Unfortunately, this will not work for sextupoles, since if we expand the Hamiltonian to second order in the dynamical variables, we lose the sextupole field altogether. Much of the discussion in the following chapters is concerned with how to find useful (approximate) solutions to the equations of motion in cases such as this. For now, we simply note that if we consider a sextupole of infinitesimal length, then (since the co-ordinates  $x$ ,  $y$  and  $z$  of a particle do not change when the particle passes through the sextupole) we can describe the dynamics using the much simpler Hamiltonian:

$$H = \frac{1}{6} k_2(x^3 - 3xy^2). \quad (8.77)$$

The equations of motion are, from Hamilton's equations (2.3) and (2.4):

$$\frac{dp_x}{ds} = -\frac{\partial H}{\partial x} = -\frac{1}{2} k_2(x^2 - y^2), \quad (8.78)$$

$$\frac{dp_y}{ds} = -\frac{\partial H}{\partial y} = k_2 xy. \quad (8.79)$$

Hence, the transfer map is:

$$\Delta p_x = -\frac{1}{2} k_2 L_{\text{sext}}(x^2 - y^2), \quad (8.80)$$

$$\Delta p_y = k_2 L_{\text{sext}} xy, \quad (8.81)$$

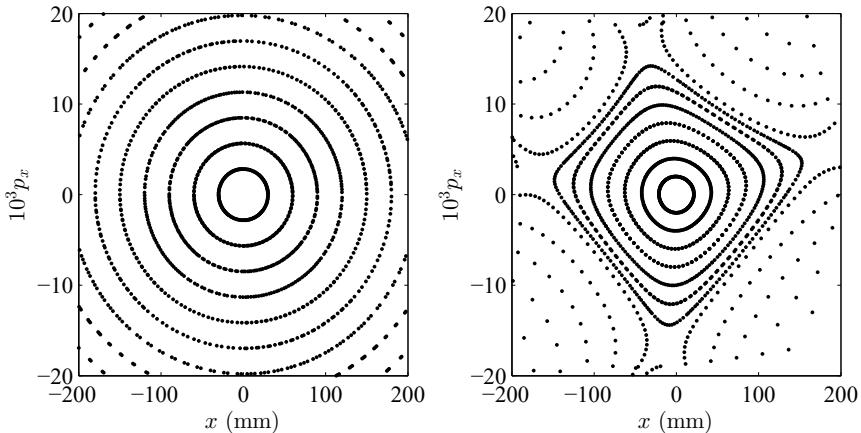


Fig. 8.6 Phase space portraits of a FODO storage ring without (left) and with (right) sextupoles for correction of chromaticity.

where  $k_2 L_{\text{sext}}$  is the integrated strength of the sextupole:

$$k_2 L_{\text{sext}} = \frac{q}{P_0} \int_0^{L_{\text{sext}}} \frac{\partial^2 B_y}{\partial x^2} ds. \quad (8.82)$$

All the other dynamical variables are unchanged when the particle passes through the sextupole. Note that the changes in the transverse momenta have a nonlinear dependence on the co-ordinates  $x$  and  $y$ .

Using the transfer map for a (thin) sextupole, we can track particles in a storage ring consisting of drift spaces, dipoles, quadrupoles and sextupoles. One way of visualising the dynamics is by constructing a *phase space portrait* (a kind of Poincaré section). This is done by tracking particles over many turns through the storage ring, plotting the phase space co-ordinates for all the particles after each turn on a single plot. The initial co-ordinates of the particles are usually selected to cover a range of betatron amplitudes.

Figure 8.6 shows horizontal phase space portraits for a FODO storage ring under different conditions. Without sextupoles, we see that the trajectory of each particle in phase space lies on an ellipse: the shape of the ellipse is described by the Courant–Snyder parameters, and the area is  $\pi J_x$ , where  $J_x$  is the betatron action. When sextupoles are included in the lattice to set the chromaticity to zero, the phase space portrait changes dramatically. At small betatron amplitudes, the particles still follow ellipses in phase space. But at larger amplitudes the ellipses start to become distorted, and eventually break up altogether. Under certain conditions (depending on details of

the lattice; see Section 11.3) smaller ellipses not centred on the origin appear. The introduction of nonlinear components (in this case, sextupoles) in the lattice makes the dynamics much more complicated.

In the following chapters, we shall investigate nonlinear effects in more detail. We shall have two main goals. The first goal is to develop techniques for constructing accurate transfer maps (which should also, ideally, be symplectic) for given components, such as sextupole or octupole magnets. The second goal is to understand the impact that nonlinear components have on the dynamics in an accelerator beam line.

**This page intentionally left blank**

## Chapter 9

# Representations of Transfer Maps

In Chapter 2 we derived the equations of motion for a charged particle in an accelerator beam line using the Hamiltonian formalism. Use of Hamiltonian mechanics (instead of, for example, Newtonian mechanics) has the advantage that it provides a convenient and systematic framework for constructing symplectic transfer maps for various types of beam line component. Neglecting certain effects such as radiation, the motion of a charged particle in a magnetic field is symplectic: as a consequence, there are a number of invariants associated with the motion. However, for many systems the Hamiltonian equations of motion do not admit exact solutions, in which case a numerical or approximate analytical technique must be applied to determine how the system evolves. Unless the integration technique is carefully constructed, the symplectic nature of the solution is lost. Methods for symplectic integration have long been known (see, for example, [Hairer *et al.* (2006a)]), but the particular challenges arising in modelling the motion of particles in accelerators has led to the development of somewhat specialised techniques for accelerator beam lines [Forest (2006)].

Although it is not always essential, for particular studies in beam dynamics it is important that the solutions to the equations of motion preserve the symplectic nature of the dynamics [Ruth (1983); Berg *et al.* (1994)]. This is the case, for example, when tracking particles over many turns in a storage ring: use of non-symplectic transfer maps can lead to artificial growth or damping of the betatron and synchrotron oscillations, resulting in inaccurate information on the stability of the motion [Kleiss *et al.* (1992)]. Even in electron storage rings, where the motion is non-symplectic because of synchrotron radiation effects, it can be useful to apply symplectic integrators for certain aspects of the beam transport, so that evaluation of the radiation effects is not influenced by artifacts of the integration

algorithms. As has already been mentioned, use of the Hamiltonian formalism leads naturally to techniques for constructing approximate solutions to the equations of motion that are exactly symplectic. As an example, in Chapter 3 we constructed the transfer maps for components including dipoles, quadrupoles and solenoids, by first expanding the Hamiltonian to second order in the dynamical variables, and then finding exact solutions to the resulting equations of motion. Although the transfer maps obtained in this way are approximate (since they are based on an approximate Hamiltonian), the fact that they are exact solutions to Hamilton's equations means that they are symplectic [Goldstein *et al.* (2001c)].

Unfortunately, this approach fails when we start to consider nonlinear components such as sextupoles. The important properties of these components are described by terms of third order and higher in the Hamiltonian, making a second-order approximation to the Hamiltonian insufficient. However, the equations of motion for a Hamiltonian involving third-order terms do not, except in very special (and usually artificial) cases, admit of exact solutions. In this chapter, we shall start to develop techniques for constructing transfer maps for nonlinear components in accelerator beam lines. A particularly powerful technique involves the use of Lie transformations [Dragt and Finn (1976); Dragt and Forest (1983); Forest (1998); Dragt (2020); Chao (2022a)], which are described in Section 9.1. A Lie transformation in itself can be used to represent a transfer map in an exact (and exactly symplectic) way, for any given time-independent Hamiltonian. Unfortunately, although Lie transformations are very useful for algebraic manipulation and analysis, they do not allow for convenient evaluation of transfer maps. If we wish to calculate the change in the phase space co-ordinates of a particle when it moves through a sextupole, for example, we first need to convert the map from a Lie transformation to some alternative representation.

By contrast, a transfer map expressed as a power series in the dynamical variables can be evaluated very easily. Furthermore, it is reasonably straightforward to convert a transfer map in the form of a Lie transformation to a power series. The drawback is that an exact representation of a transfer map for a real accelerator component (such as a sextupole) will generally involve a power series with an infinite number of terms. To evaluate the map, the power series must be truncated at some order, and the resulting approximate map will, in general, not be symplectic. Nevertheless, some powerful computational tools have been developed for performing accurate and efficient calculations involving power series: the *truncated power*

*series algebra* (TPSA) is a notable example [Berz (1987, 1989, 1999); Berz *et al.* (2015); Chao (2022b)].

If symplectic tracking is required, then, starting from a power series or Lie algebra representation, there are three possible solutions. The first is to modify the truncated power series in such a way that it is made to be symplectic while still containing a finite number of terms. A symplectic power series map with a finite number of terms is known as a *Cremona map* [Abell (1995)]. Although techniques have been developed to convert a given truncated power series into a Cremona map, they are not particularly easy to apply. The second solution is to modify the Lie transformation so that it has an exact explicit representation in the form of a power series with a finite number of terms [Forest and Ruth (1990); Yoshida (1990)]. This is often a useful technique, and is discussed in Section 9.2. Finally, one can convert the power series into yet another representation, such as a mixed-variable generating function [Berz (1991); Yan *et al.* (1992); Berg *et al.* (1994)], that provides a compromise between accuracy (in particular, the symplectic nature of the dynamics is preserved) and ease of evaluation. The use of mixed-variable generating functions for representing transfer maps is discussed in Section 9.3.

## 9.1 Lie Transformations

In this section, we introduce the concepts of *Lie operators* and *Lie transformations*, and show how Lie transformations can be used to represent transfer maps for accelerator components. Lie transformations provide a powerful tool for the construction of transfer maps, and for the analysis of nonlinear dynamics in accelerators. They are developed from the Hamiltonian formalism for particle dynamics; one of the motivations for basing our discussion in previous chapters on Hamiltonian dynamics was to provide a useful foundation for the development of Lie transformations.

Suppose we have a set of dynamical variables, consisting of co-ordinates  $x_i$  and conjugate momenta  $p_i$  (in  $n$  degrees of freedom,  $i = 1 \cdots n$ ). The dynamics of a particle in a given field are described by a Hamiltonian; solution of Hamilton's equations with given initial conditions leads to expressions for the co-ordinates and momenta as functions of the independent variable  $s$ . For particles moving along an accelerator beam line, the independent variable represents the distance along the reference trajectory. Now suppose that we have any function  $f$  of the phase space variables:

$$f = f(x_i, p_i). \quad (9.1)$$

We impose the condition on  $f$  that it has no explicit dependence on the independent variable. However, if we evaluate  $f$  for a particle moving along a beam line, the value of  $f$  will evolve with  $s$ , because the values of the dynamical variables evolve with  $s$ . The rate of change of  $f$  with distance along the beam line can be written generally as:

$$\frac{df}{ds} = \sum_{i=1}^n \frac{dx_i}{ds} \frac{\partial f}{\partial x_i} + \frac{dp_i}{ds} \frac{\partial f}{\partial p_i}. \quad (9.2)$$

Using Hamilton's equations (2.3) and (2.4), this becomes:

$$\frac{df}{ds} = \sum_{i=1}^n \frac{\partial H}{\partial p_i} \frac{\partial f}{\partial x_i} - \frac{\partial H}{\partial x_i} \frac{\partial f}{\partial p_i}. \quad (9.3)$$

We define the Lie operator  $:g:$  for any function  $g(x_i, p_i)$  by:

$$:g: = \sum_{i=1}^n \frac{\partial g}{\partial x_i} \frac{\partial}{\partial p_i} - \frac{\partial g}{\partial p_i} \frac{\partial}{\partial x_i}. \quad (9.4)$$

The function  $g$  used to construct the Lie operator has (like  $f$ ) no explicit dependence on the independent variable  $s$ . In the case that the Hamiltonian  $H$  has no explicit dependence on  $s$ , we can construct a Lie operator  $:H:$ . In terms of this operator, equation (9.3) can be written:

$$\frac{df}{ds} = -:H:f. \quad (9.5)$$

Equation (9.5) is an equation of motion for the function  $f(x_i, p_i)$ . The form of this equation suggests that we can write the solution:

$$f|_{s=s_0+\Delta s} = e^{-\Delta s:H:} f|_{s=s_0}, \quad (9.6)$$

where the operator  $e^{-\Delta s:H:}$  is defined by the usual power series for an exponential:

$$e^{-\Delta s:H:} = \sum_{m=0}^{\infty} \frac{(-\Delta s)^m}{m!} :H:^m. \quad (9.7)$$

Equation (9.6) may require a little further justification. We observe that we can express  $f$  at  $s = s_0 + \Delta s$  in terms of  $f$  at  $s = s_0$ , in terms of a Taylor series:

$$\begin{aligned} f|_{s=s_0+\Delta s} &= f|_{s=s_0} + \Delta s \frac{df}{ds} \Big|_{s=s_0} + \frac{\Delta s^2}{2} \frac{d^2 f}{ds^2} \Big|_{s=s_0} + \dots \\ &= \sum_{m=0}^{\infty} \frac{\Delta s^m}{m!} \frac{d^m f}{ds^m} \Big|_{s=s_0} \\ &= e^{\Delta s \frac{d}{ds}} f \Big|_{s=s_0}. \end{aligned} \quad (9.8)$$

Equation (9.5) tells us that we can replace the differential operator  $d/ds$  with the Lie operator  $-:H:$ . Making this replacement in equation (9.8) gives (9.6).

The operator  $e^{-\Delta s:g}$  is known as a *Lie transformation*, with generator  $g$ . In the context of accelerator beam dynamics, applying a Lie transformation with the Hamiltonian as the generator to a function  $f$  produces a transfer map for  $f$ . The function  $f$  can be any function of the dynamical variables; if we set  $f$  equal to each of the dynamical variables in turn, we construct the full transfer map for the system. Any Lie transformation represents the evolution of a conservative dynamical system, with Hamiltonian corresponding to the generator of the Lie transformation. Therefore, the map represented by a Lie transformation must be symplectic: this is an important property of Lie transformations in connection with their use for accelerator beam dynamics.

Before we consider the application of Lie transformations to the dynamics of a particle in an accelerator, let us illustrate the ideas with a simple example. Consider the Hamiltonian in one degree of freedom:

$$H = \frac{1}{2}p_x^2 + \frac{1}{2}\omega^2x^2. \quad (9.9)$$

This is of course just the Hamiltonian for a simple harmonic oscillator. In this case, the equations of motion for  $x$  and  $p_x$  are integrable, and the solutions are well known. However, as an illustration, let us integrate the system using a Lie transformation. To apply the Lie transformation, we first need to know the results of applying the Lie operator  $:H:$  to the dynamical variables. Using the definition (9.4), we find:

$$:H:x = \frac{\partial H}{\partial x} \frac{\partial x}{\partial p_x} - \frac{\partial H}{\partial p_x} \frac{\partial x}{\partial x} = -\frac{\partial H}{\partial p_x} = -p_x. \quad (9.10)$$

Similarly, we find:

$$:H:p_x = \omega^2x. \quad (9.11)$$

Hence:

$$:H:^2x = :H:(-p_x) = -\omega^2x, \quad (9.12)$$

$$:H:^3x = :H:(-\omega^2x) = \omega^2p_x, \quad (9.13)$$

$$:H:^4x = :H:(\omega^2p_x) = \omega^4x, \quad (9.14)$$

and so on. Therefore, we can write:

$$\begin{aligned}
 e^{-\Delta s:H}x &= x - \Delta s:H:x + \frac{\Delta s^2}{2}:H.^2x - \frac{\Delta s^3}{6}:H.^3x + \frac{\Delta s^4}{24}:H.^4x - \dots \\
 &= x + \Delta s p_x - \frac{\Delta s^2}{2}\omega^2 x - \frac{\Delta s^3}{6}\omega^2 p_x + \frac{\Delta s^4}{24}\omega^4 x + \dots \\
 &= \left(1 - \frac{\Delta s^2}{2}\omega^2 + \frac{\Delta s^4}{24}\omega^4 - \dots\right)x + \left(\Delta s - \frac{\Delta s^3}{6}\omega^2 + \dots\right)p_x \\
 &= \cos(\omega \Delta s)x + \frac{\sin(\omega \Delta s)}{\omega}p_x.
 \end{aligned} \tag{9.15}$$

The Lie transformation of  $x$ , written as  $e^{-\Delta s:H}x$ , is a function of the dynamical variables  $x$  and  $p_x$ . Evaluating this function at  $s = s_0$  gives the value of  $x$  at  $s = s_0 + \Delta s$ . That is:

$$x(s_0 + \Delta s) = e^{-\Delta s:H}x|_{s=s_0+\Delta s} = \cos(\omega \Delta s)x(s_0) + \frac{\sin(\omega \Delta s)}{\omega}p_x(s_0). \tag{9.16}$$

Similarly (we leave this as an exercise for the reader!) it is found that:

$$p_x(s_0 + \Delta s) = e^{-\Delta s:H}p_x|_{s=s_0+\Delta s} = \cos(\omega \Delta s)p_x(s_0) - \omega \sin(\omega \Delta s)x(s_0). \tag{9.17}$$

Equations (9.16) and (9.17) constitute the transfer map for the simple harmonic oscillator in one degree of freedom. In this case, applying Lie transformations to construct the transfer map (essentially, to solve the equations of motion) probably appears unnecessarily complicated. But the beauty of the method is that, since it only requires differentiation, it can be applied to construct a transfer map even in cases where the equations of motion are not integrable. The transfer map is given in the form of a power series. In the case of the simple harmonic oscillator, we can collect terms in the power series and express them in terms of sine and cosine functions. This allows us to express the infinite series (obtained by applying the Lie transformation) in closed form. In general, it is not always possible to do this. However, the power series maps can still be useful, as long as they converge. It is of course necessary to halt the calculation of the terms at some point: this amounts to a truncation of the power series, and usually results in the loss of the symplecticity of the transfer map. If symplecticity is required, then some modification of the technique is necessary to produce a symplectic map in an explicit form that can be applied directly, without the need to solve additional equations. A map in this form is sometimes called a *symplectic integrator*.

Lie transformations have some mathematical properties that make them useful for the modelling and analysis of dynamics in conservative systems.

We shall complete the present section by deriving two results that we will need in Chapter 10, where we discuss the construction of symplectic integrators using Lie transformations. The first result can be stated as follows: if  $f$  and  $h$  are functions of the dynamical variables, and  $g = g(h)$  is a function of  $h$ , then:

$$e^{:f:}g(h) = g(e^{:f:}h). \quad (9.18)$$

That is, if we express  $g$  as a function of the dynamical variables through the function  $h$  and then apply a Lie transformation, the result is the same as expressing  $g$  as a function of the dynamical variables through the transformed function  $e^{:f:}h$ . The second result is the expression for a similarity transformation. If  $f$  and  $g$  are any functions of the dynamical variables, then:

$$e^{:f:} \exp(:g:) e^{-:f:} = \exp(:e^{:f:}g:). \quad (9.19)$$

We shall not present rigorous proofs of these results, nevertheless we hope that our arguments illustrate some properties of Lie operators and Lie transformations. More comprehensive and rigorous treatments can be found in [Forest (1998); Dragt (2020)].

We turn our attention first to the result (9.18). Consider the effect of applying  $:f:$  to  $g(h)$ . From (9.4), it follows that:

$$:f:g(h) = \frac{dg}{dh} :f:h. \quad (9.20)$$

It also follows from (9.4) that Lie operators obey the usual product rule; that is, for any two functions  $a$  and  $b$  of the dynamical variables:

$$:f:ab = a:f:b + b:f:a. \quad (9.21)$$

Using (9.21), we can write the result of applying a Lie operator to the function  $g$  multiple times, thus:

$$:f:^2g(h) = \frac{d^2g}{dh^2} (:f:h)^2 + \frac{dg}{dh} (:f:^2h), \quad (9.22)$$

$$:f:^3g(h) = \frac{d^3g}{dh^3} (:f:h)^3 + 3\frac{d^2g}{dh^2} (:f:h)(:f:^2h) + \frac{dg}{dh} (:f:^3h), \quad (9.23)$$

$$\begin{aligned} :f:^4g(h) = & \frac{d^4g}{dh^4} (:f:h)^4 + 6\frac{d^3g}{dh^3} (:f:h)^2(:f:^2h) + 3\frac{d^2g}{dh^2} (:f:^2h)^2 \\ & + 4\frac{d^2g}{dh^2} (:f:h)(:f:^3h) + \frac{dg}{dh} (:f:^4h), \end{aligned} \quad (9.24)$$

and so on. Using the definition of a Lie transformation (9.7), we can write:

$$e^{:f:}g(h) = g(h) + :f:g(h) + \frac{1}{2}:f:^2g(h) + \frac{1}{6}:f:^3g(h) + \frac{1}{24}:f:^4g(h) + \dots \quad (9.25)$$

Substituting for the various terms in (9.25) from (9.20) and from (9.22)–(9.24), and collecting terms in derivatives of  $g$  gives:

$$\begin{aligned}
 e^{f:h} g(h) &= g(h) \\
 &+ \frac{dg}{dh} \left( :f:h + \frac{1}{2} :f:^2 h + \frac{1}{6} :f:^3 h + \frac{1}{24} :f:^4 h + \dots \right) \\
 &+ \frac{1}{2} \frac{d^2 g}{dh^2} \left( (:f:h)^2 + (:f:h)(:f:^2 h) + \frac{1}{4} (:f:^2 h)^2 \right. \\
 &\quad \left. + \frac{1}{3} (:f:h)(:f:^3 h) + \dots \right) \\
 &+ \frac{1}{6} \frac{d^3 g}{dh^3} \left( (:f:h)^3 + \frac{3}{2} (:f:h)^2 (:f:^2 h) + \dots \right) \\
 &+ \frac{1}{24} \frac{d^4 g}{dh^4} ((:f:h)^4 + \dots) + \dots. \tag{9.26}
 \end{aligned}$$

This can be written as:

$$\begin{aligned}
 e^{f:h} g(h) &= g(h) \\
 &+ \frac{dg}{dh} \left( :f:h + \frac{1}{2} :f:^2 h + \frac{1}{6} :f:^3 h + \frac{1}{24} :f:^4 h + \dots \right) \\
 &+ \frac{1}{2} \frac{d^2 g}{dh^2} \left( :f:h + \frac{1}{2} :f:^2 h + \frac{1}{6} :f:^3 h + \frac{1}{24} :f:^4 h + \dots \right)^2 \\
 &+ \frac{1}{6} \frac{d^3 g}{dh^3} \left( :f:h + \frac{1}{2} :f:^2 h + \frac{1}{6} :f:^3 h + \frac{1}{24} :f:^4 h + \dots \right)^3 \\
 &+ \frac{1}{24} \frac{d^4 g}{dh^4} \left( :f:h + \frac{1}{2} :f:^2 h + \frac{1}{6} :f:^3 h + \frac{1}{24} :f:^4 h + \dots \right)^4 + \dots \\
 &= g(h) + \frac{dg}{dh} ((e^{f:h} - 1)h) + \frac{1}{2} \frac{d^2 g}{dh^2} ((e^{f:h} - 1)h)^2 \\
 &\quad + \frac{1}{6} \frac{d^3 g}{dh^3} ((e^{f:h} - 1)h)^3 + \frac{1}{24} \frac{d^4 g}{dh^4} ((e^{f:h} - 1)h)^4 + \dots. \tag{9.27}
 \end{aligned}$$

In the final lines, we recognise a Taylor series expansion, from which we infer the result (9.18):

$$e^{f:h} g(h) = g(h + (e^{f:h} - 1)h) = g(e^{f:h} h). \tag{9.28}$$

Let us now consider the expression for a similarity transformation (9.19). This result can be shown as follows. We begin by introducing the *Poisson bracket*  $[g, h]$ , defined for any two functions  $g$  and  $h$  of the dynamical variables in  $n$  degrees of freedom by:

$$[g, h] = \sum_{i=1}^n \frac{\partial g}{\partial x_i} \frac{\partial h}{\partial p_i} - \frac{\partial h}{\partial x_i} \frac{\partial g}{\partial p_i}. \tag{9.29}$$

From (9.4), we see that a Lie operator can be written in terms of the Poisson bracket, thus:

$$:g:h = [g, h]. \quad (9.30)$$

The result of applying a Lie operator to a function is another function. In terms of the Poisson bracket in one degree of freedom, we can write:

$$[g(x, p_x), h(x, p_x)]_{x, p_x} = F(x, p_x), \quad (9.31)$$

where we write the Poisson bracket in (9.31) as  $[\cdot, \cdot]_{x, p_x}$  to emphasise that it is expressed in terms of derivatives of  $x$  and  $p_x$ .  $F(x, p_x)$  is some function of the dynamical variables. Using (9.18):

$$e^{:f:}[g(x, p_x), h(x, p_x)]_{x, p_x} = e^{:f:}F(x, p_x) = F(X, P_X), \quad (9.32)$$

where:

$$X = e^{:f:}x, \quad (9.33)$$

$$P_X = e^{:f:}p_x. \quad (9.34)$$

Since a Lie transformation is symplectic,  $X$  and  $P_X$  can be treated as canonical variables. Then, in (9.31) we can simply substitute  $X$  for  $x$  and  $P_X$  for  $p_x$ :

$$[g(X, P_X), h(X, P_X)]_{X, P_X} = F(X, P_X). \quad (9.35)$$

The Poisson bracket in (9.35) is expressed in terms of derivatives with respect to the new variables  $X$  and  $P_X$ . However, the Poisson bracket is invariant under a canonical transformation. To explain and prove this statement, observe that the Poisson bracket can be written as:

$$[g, h]_{X, P_X} = (\nabla_{\vec{X}} g)^T S_2 \nabla_{\vec{X}} h, \quad (9.36)$$

where:

$$\nabla_{\vec{X}} = \left( \begin{array}{c} \frac{\partial}{\partial X} \\ \frac{\partial}{\partial P_X} \end{array} \right), \quad (9.37)$$

and  $S_2$  is given by (2.78). The operator  $\nabla_{\vec{x}}$  is related to the operator  $\nabla_{\vec{X}}$  by:

$$\nabla_{\vec{X}} = J \nabla_{\vec{x}}, \quad (9.38)$$

where the matrix  $J$  is the Jacobian of the transformation:

$$J = \frac{\partial(X, P_X)}{\partial(x, p_x)}. \quad (9.39)$$

Substituting for  $\nabla_{\vec{x}}$  from (9.38), the Poisson bracket (9.36) becomes:

$$[g, h]_{X, P_X} = (\nabla_{\vec{x}} g)^T J^T S_2 J \nabla_{\vec{x}} h. \quad (9.40)$$

For a canonical transformation, the Jacobian is a symplectic matrix:

$$J^T S_2 J = S_2, \quad (9.41)$$

and so:

$$[g, h]_{X, P_X} = (\nabla_{\vec{x}} g)^T S_2 \nabla_{\vec{x}} h = [g, h]_{x, p_x}. \quad (9.42)$$

Equation (9.42) expresses the invariance of the Poisson bracket under a canonical transformation. Although we have considered just one degree of freedom, the result generalises to any number of degrees of freedom. Applying (9.42) to (9.35) gives:

$$\begin{aligned} e^{:f:}[g(x, p_x), h(x, p_x)]_{x, p_x} &= F(X, P_X) \\ &= [g(X, P_X), h(X, P_X)]_{X, P_X} \\ &= [g(X, P_X), h(X, P_X)]_{x, p_x} \\ &= [e^{:f:}g(x, p_x), e^{:f:}h(x, p_x)]_{x, p_x}. \end{aligned} \quad (9.43)$$

In the final step, we have used (9.18). Then, using (9.30), it follows that:

$$e^{:f:}g:e^{-:f:}h = :e^{:f:}g:e^{:f:}e^{-:f:}h = :e^{:f:}g:h, \quad (9.44)$$

where in the last step we have used the fact that  $e^{:f:}e^{-:f:}$  is the identity transformation. The result (9.44) can be generalised as follows:

$$e^{:f:}g:2e^{-:f:}h = e^{:f:}g:e^{:f:}e^{-:f:}e^{:f:}g:e^{-:f:}h = :e^{:f:}g:2h, \quad (9.45)$$

where we have again made use of the fact that  $e^{:f:}e^{-:f:}$  is the identity transformation. Finally, using the definition (9.7) for a Lie transformation in terms of a power series in Lie operators, we obtain the result (9.19):

$$e^{:f:}\exp(:g:)e^{-:f:}h = \sum_{m=0}^{\infty} e^{:f:}\frac{g:m}{m!}e^{-:f:}h = \sum_{m=0}^{\infty} \frac{:e^{:f:}g:m}{m!}h = \exp(:e^{:f:}g:)h. \quad (9.46)$$

## 9.2 Power Series Map for a Sextupole

A power series in the dynamical variables provides a very convenient form for the transfer map when performing numerical tracking using a computer. In this section, we shall use the example of a sextupole magnet to show how Lie transformations can be used to construct a transfer map for an accelerator component in the form of a power series, if an analytical expression

for the vector potential is known. The method that we discuss has two significant limitations. The first is that it applies only to cases where the vector potential has no dependence on the independent variable  $s$ . This makes the technique appropriate for multipole magnets, but not for more complicated components such as undulators, where it is often necessary to include the full three-dimensional nature of the magnetic field. The second limitation is that, in general, the power series produced by application of a Lie transformation is only symplectic if all the terms of the power series are retained. Generally, application of a Lie transformation produces a power series with an infinite number of terms. If the series is truncated at some finite order then the resulting power series will not be symplectic. This may not be too serious for modelling single-pass systems, although if the aim is to estimate the impact of fairly weak non-symplectic effects (such as wake fields or synchrotron radiation) then it is desirable that any deviation from symplecticity in the basic transport of particles through the magnetic fields in the beam line be as small as possible. For long-term tracking in storage rings, particularly in proton storage rings where radiation effects are usually negligible, it is often essential to use transfer maps that are symplectic to numerical precision. In Chapter 10, we shall discuss how to modify and extend the techniques presented in this section, so as to be able to construct an explicit representation of a symplectic transfer map (i.e. a representation that can be applied directly, without the need for further solution of differential or algebraic equations), even in the case where the magnetic field depends on the independent variable  $s$ .

The full Hamiltonian for a sextupole can be written:

$$H = \frac{\delta}{\beta_0} - \sqrt{\left(\frac{1}{\beta_0} + \delta\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}} + \frac{1}{6} k_2 (x^3 - 3xy^2). \quad (9.47)$$

The transfer map for a sextupole can be written in terms of a Lie transformation with generator given by the Hamiltonian (9.47). For a sextupole of length  $L$ :

$$\vec{x}(s = L) = e^{-L:H:\vec{x}}|_{s=0}. \quad (9.48)$$

The Lie transformation is applied to each component of the phase space vector  $\vec{x} = (x, p_x, y, p_y, z, \delta)$  separately, and the resulting functions are evaluated at  $s = 0$  to give the values of the dynamical variables (i.e. the components of the phase space vector) at  $s = L$ . Although (9.48) is a concise, accurate and exactly symplectic representation of the transfer map, it is not in a form that can easily be implemented in a tracking code on a computer,

for example. To produce an explicit form of the map, we must apply the Lie transformation to each of the dynamical variables in turn. This involves differentiation, and summation of the different terms in the power series expansion of the exponential. This is a rather laborious process, and the resulting expressions rapidly become rather lengthy. Up to second order in  $L$ , the transfer map is:

$$x(L) = x_0 + \frac{p_{x0}}{C}L + \frac{(2x_0p_{x0}y_0p_{y0} - (x_0^2 - y_0^2)(C^2 + p_{x0}^2))}{4C^3}k_2L^2 + O(L^3), \quad (9.49)$$

$$p_x(L) = p_{x0} - \frac{1}{2}(x_0^2 - y_0^2)k_2L - \frac{(x_0p_{x0} - y_0p_{y0})}{2C}k_2L^2 + O(L^3), \quad (9.50)$$

$$y(L) = y_0 + \frac{p_{y0}}{C}L - \frac{((x_0^2 - y_0^2)p_{x0}p_{y0} - 2x_0y_0(C^2 + p_{y0}^2))}{4C^3}k_2L^2 + O(L^3), \quad (9.51)$$

$$p_y(L) = p_{y0} + x_0y_0k_2L + \frac{(x_0p_{y0} + y_0p_{x0})}{2C}k_2L^2 + O(L^3), \quad (9.52)$$

$$\begin{aligned} z(L) &= z_0 + \frac{(1 + \beta_0\delta_0 - C)L}{\beta_0C} \\ &+ \frac{((x_0^2 - y_0^2)p_{x0} - 2x_0y_0p_{y0})(1 + \beta_0\delta_0)}{4\beta_0C^3}k_2L^2 + O(L^3), \end{aligned} \quad (9.53)$$

$$\delta(L) = \delta_0, \quad (9.54)$$

where:

$$C = \sqrt{\left(\frac{1}{\beta_0} + \delta_0\right)^2 - p_{x0}^2 - p_{y0}^2 - \frac{1}{\beta_0^2\gamma_0^2}}, \quad (9.55)$$

and  $x_0 = x(0)$ ,  $p_{x0} = p_x(0)$ ,  $y_0 = y(0)$ ,  $p_{y0} = p_y(0)$  and  $\delta_0 = \delta(0)$  are the values of the dynamical variables at the entrance of the sextupole.

The convergence of the power series should be considered. Since the power series expression for  $e^\alpha$  converges for all  $\alpha$ , we expect the power series obtained by applying a Lie transformation to each of the dynamical variables also to converge. The rate of convergence depends very much on the parameter regime (i.e. the focusing strength  $k_2$ , the length  $L$  of the magnet, and the range of values of the dynamical variables that is of interest), and may be quite slow. The complexity of the expressions involved means that it is not easy to write down an expression for the rate of convergence. However, given the form of the Lie transformation, it is expected in the case of the sextupole that the ratio of successive terms in each power series will be of order  $k_2Lx_i^2$ , where  $x_i^2$  represents second-order products of the dynamical variables.

Of equal importance to the rate of convergence is the symplectic error (i.e. the amount by which the map deviates from the symplectic condition). A symplectic map will satisfy:

$$J^T SJ = S, \quad (9.56)$$

where  $J$  is the Jacobian of the map; that is, a matrix with elements:

$$J_{ij} = \frac{\partial x_i(L)}{\partial x_j(0)}, \quad (9.57)$$

and  $S$  is the antisymmetric matrix (4.14). For a power series map, the algebra involved in calculating the Jacobian and the matrix product  $J^T SJ$  is formidable; but it is found that for a power series map calculated to order  $L^m$  using the Lie transformation for a sextupole, that:

$$J^T SJ = S + O(k_2^m L^{m+1} x_i^{2m-1}). \quad (9.58)$$

To ensure a small symplectic error, the power series map needs to be calculated to order  $m$  such that:

$$k_2^m L^{m+1} x_i^{2m-1} \ll 1. \quad (9.59)$$

The size of the symplectic error that is acceptable will depend on the situation being considered. If it is important that the transfer maps used in a tracking simulation be symplectic (for example, for long-term tracking in a proton storage ring), then usually it is preferable not to use a truncated power series produced from a Lie transformation, but to use instead a technique for constructing an explicit transfer map that consists of a finite number of terms and is exactly symplectic. Some possible techniques are described in Chapter 10.

As an illustration of the use of truncated power series maps, we generate the phase space portrait for a particle in a storage ring with a single sextupole. For simplicity, we consider only horizontal motion. If vertical motion is included, then the dynamics become significantly more complicated, because (as can be seen from equations (9.49)–(9.52)) a sextupole couples horizontal and vertical motion. We also assume the ultra-relativistic limit  $\gamma_0 \rightarrow \infty$  (so that  $\beta_0 \rightarrow 1$ ), and take  $\delta = 0$ . Then, the map for the sextupole becomes (to second order in  $L$ ):

$$x(L) = x_0 + \frac{p_{x0}}{\sqrt{1 - p_{x0}^2}} L - \frac{x_0^2}{4(1 - p_{x0}^2)^{\frac{3}{2}}} k_2 L^2 + O(L^3), \quad (9.60)$$

$$p_x(L) = p_{x0} - \frac{1}{2} x_0^2 k_2 L - \frac{x_0 p_{x0}}{2\sqrt{1 - p_{x0}^2}} k_2 L^2 + O(L^3). \quad (9.61)$$

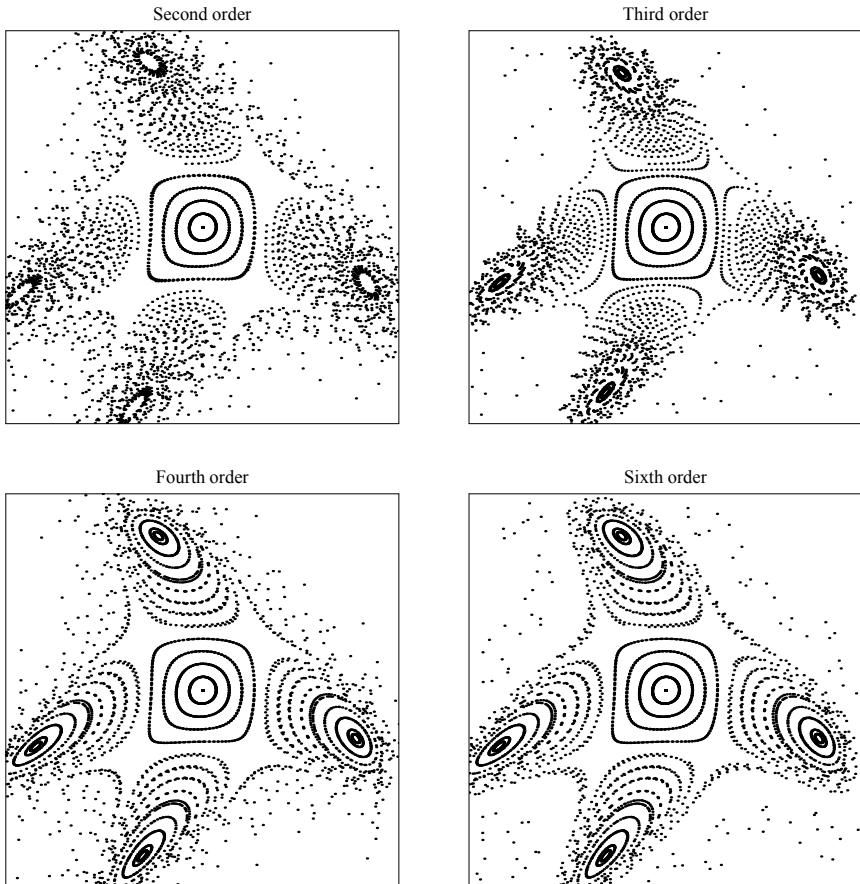


Fig. 9.1 Horizontal phase space portraits of a simple storage ring with a single sextupole. The storage ring lattice excluding the sextupole is represented by a linear transfer map with phase advance  $0.246 \times 2\pi$ . The transfer map for the sextupole is a power series obtained from a Lie transformation applied to the dynamical variables. In each case, the power series is truncated at a different order (as shown above each plot) in the length of the sextupole. The horizontal axis in each plot is the co-ordinate  $x$ , and the vertical axis is the canonical momentum  $p_x$ .

Rather than include a detailed model of the rest of the lattice in the storage ring (which is not important for our illustration), we represent the ring by a simple rotation in horizontal phase space, with rotation angle  $2\pi\nu_x$ , where  $\nu_x$  is the horizontal betatron tune.

Figure 9.1 shows the phase space portraits obtained by tracking 1000 turns in the storage ring, using power series maps for the sextupole magnet.

The power series maps are obtained by applying a Lie transformation to the phase space variables, and truncating the power series at a specified order in  $L$ . In all the cases shown, the betatron tune has fractional part 0.246 (the integer part has no effect on the phase space portrait), and the sextupole magnet has focusing strength  $k_2 = 4000 \text{ m}^{-3}$  and length 0.1 m. The phase space portraits show a range of co-ordinate  $x$  from  $-4 \text{ mm}$  to  $+4 \text{ mm}$ , and a range of momentum  $p_x$  from  $-4 \times 10^{-3}$  to  $+4 \times 10^{-3}$ . With these parameters, we expect to see fairly rapid convergence in the power series map, since  $k_2 L x_i^2 = 0.0064$ . However, over many turns, the effects of even small differences between maps truncated at different order can build up to become visible. The symplectic error at order  $m$  is of order  $k_2^m L^{m+1} x_i^{2m-1} = 0.16^m$ .

### 9.3 Mixed-Variable Generating Functions

Mixed-variable generating functions can be used to define the relationship between two sets of canonical variables. We used mixed-variable generating functions in Chapter 2 when we derived a convenient form for the Hamiltonian for particle dynamics in an accelerator beam line: a mixed-variable generating function allowed us to change from one set of dynamical variables to another (for example, when changing from Cartesian co-ordinates to a co-ordinate system with a curved reference trajectory). In the context of nonlinear dynamics, mixed-variable generating functions play a rather different role, in allowing us to write down an exactly symplectic transfer map in closed form.

Consider a conservative dynamical system (i.e. a system obeying Hamilton's equations) in  $n$  degrees of freedom. The dynamics may be described using various sets of dynamical variables. Suppose that we have an ‘old’ set of variables  $\vec{x}$ ; we can define a ‘new’ set of variables  $\vec{X}$  using a relationship of the form:

$$\vec{X} = \vec{g}(\vec{x}), \quad (9.62)$$

where the function  $\vec{g}$  is a vector with  $2n$  components. If the ‘old’ variables  $\vec{x}$  are canonical, then we can ensure that the ‘new’ variables  $\vec{X}$  are also canonical by deriving  $\vec{g}$  from a mixed-variable generating function.

The idea of a canonical transformation between two sets of dynamical variables can be applied to transfer maps, since a transfer map is a relationship between the dynamical variables at two different points in a beam line,  $s = s_0$  and  $s = s_1$ :

$$\vec{x}(s_1) = \vec{f}(\vec{x}(s_0)). \quad (9.63)$$

The function  $\vec{f}$  is a function of the values of the dynamical variables at  $s = s_0$ . Comparing (9.62) and (9.63), we see that if we identify the initial values  $\vec{x}(s_0)$  with the ‘old’ variables  $\vec{x}$ , and the final values  $\vec{x}(s_1)$  with the ‘new’ variables  $\vec{X}$ , then a mixed-variable generating function can be used to represent a transfer map. Representing a transfer map in terms of a mixed-variable generating function has two advantages over a power series representation. First, a generating function map is always exactly symplectic, even if some approximation was made in constructing the generating function. Second, the representation of the map is more compact (i.e. requires the specification of fewer coefficients for a given order) than an explicit power series map. The disadvantage is that a mixed-variable generating function is not an explicit representation of a transfer map, and requires the solution of algebraic equations whenever it needs to be applied. A power series map is explicit, since all that is required is the evaluation of a set of expressions into which specific values of the dynamical variables are substituted.

It should be remembered that (9.62) is a relationship between two sets of *variables*, whereas (9.63) is a relationship between two sets of *values*. However, in both cases the left-hand side is expressed as a function of quantities on the right-hand side, and it is the form of the function that is of principal interest. In using a mixed-variable generating function to represent a transfer map, we are really interpreting a relationship between two sets of variables in terms of a function that relates two sets of variables and a function that relates two sets of values.

There are different types of mixed-variable generating function. Consider the case of a mixed-variable generating function  $F_3(X_i, p_j)$ . This is a function of the ‘new’ co-ordinates and the ‘old’ momenta: in Goldstein’s nomenclature [Goldstein *et al.* (2001c)],  $F_3(X_i, p_j)$  is a mixed-variable generating function of the third kind. The general relationship between the ‘old’ and ‘new’ variables in terms of a mixed-variable generating function of the third kind is as follows:

$$x_i = -\frac{\partial F_3}{\partial p_i}, \quad (9.64)$$

$$P_i = -\frac{\partial F_3}{\partial X_i}, \quad (9.65)$$

where, in  $n$  degrees of freedom,  $i = 1 \dots n$  and  $j = 1 \dots n$ . As an example, consider the specific case:

$$F_3(X, Y, Z, p_x, p_y, \delta) = -Xp_x + \frac{1}{2}KX^2 - Yp_y - \frac{1}{2}KY^2 + Z\delta. \quad (9.66)$$

Using (9.64), we find:

$$x = X, \quad (9.67)$$

$$y = Y, \quad (9.68)$$

$$z = Z. \quad (9.69)$$

Using (9.65):

$$P_X = p_x - KX, \quad (9.70)$$

$$P_Y = p_y + KY, \quad (9.71)$$

$$P_Z = \delta. \quad (9.72)$$

Note that (9.70) and (9.71) are equations involving the new co-ordinates and new momenta; in this case, it is simple to eliminate the new co-ordinates (since they are just equal to the old co-ordinates), giving expressions for the new co-ordinates and momenta purely in terms of the old co-ordinates and momenta. Making the appropriate substitutions, and then writing the old variables  $x$  as  $x(s_0)$ ,  $p_x$  as  $p_x(s_0)$  etc., and writing the new variables  $X$  as  $x(s_1)$ ,  $P_X$  as  $p_x(s_1)$  etc., equations (9.67)–(9.72) become:

$$x(s_1) = x(s_0), \quad (9.73)$$

$$y(s_1) = y(s_0), \quad (9.74)$$

$$z(s_1) = z(s_0), \quad (9.75)$$

$$p_x(s_1) = p_x(s_0) - Kx(s_0), \quad (9.76)$$

$$p_y(s_1) = p_y(s_0) + Ky(s_0), \quad (9.77)$$

$$\delta(s_1) = \delta(s_0). \quad (9.78)$$

We recognise the transfer map for a thin quadrupole of focusing strength  $K$ . Hence, the mixed-variable generating function (9.66) represents the transfer map for a thin quadrupole.

A general (not necessarily symplectic) first-order transformation from the variables at  $s = s_0$  to the variables at  $s = s_1$  can be defined by 36 coefficients: these are essentially the components of the  $6 \times 6$  transfer matrix. A mixed-variable generating function for a first-order transformation is written as a sum of second-order monomials; in three degrees of freedom (i.e. with six dynamical variables), a generating function for a first-order transformation can therefore be defined by just 21 coefficients, corresponding to the coefficients of all second-order monomials. Thus, the mixed-variable generating function can be represented more concisely than a transfer matrix. The fact that a mixed-variable generating function for

a linear map can be specified using fewer coefficients than a linear transfer matrix is related to the implicit symplecticity of a transformation derived from a mixed-variable generating function. The symplectic condition:

$$M^T S M = S, \quad (9.79)$$

imposes  $n(2n-1)$  constraints on the  $2n \times 2n$  matrix  $M$ : this follows from the fact that the left- and right-hand sides of (9.79) are antisymmetric matrices. Therefore, a symplectic matrix can be described by  $n(2n+1)$  parameters: in the case of three degrees of freedom,  $n = 3$ , and a symplectic transfer matrix can be determined by specifying 21 independent parameters. This corresponds to the number of coefficients in a mixed-variable generating function for a first-order symplectic transformation in three degrees of freedom.

It should be noted that there is no single kind of generating function that is capable of representing all symplectic transformations [Erdélyi and Berz (2001, 2004); Erdélyi (2006)]. For example, the identity transformation (which is certainly symplectic) cannot be represented by mixed-variable generating functions of the first or fourth kind (in Goldstein's nomenclature). For individual accelerator components, where the transformation is often close to an identity, it is more appropriate to represent the transformations using mixed-variable generating functions of either the second or third kind.

Although mixed-variable generating functions are sometimes a useful tool for symplectic tracking, it turns out not to be an easy task to construct the mixed-variable generating function for a given accelerator component, even if the form of the Hamiltonian is known (as in the case, for example, of a multipole magnet). One way to construct the mixed-variable generating function is first to express the map as a power series up to some order in the dynamical variables, and then use an iterative procedure to convert the power series map into a mixed-variable generating function [Yan (1999)]. To illustrate the process, we shall consider the case of a sextupole. To simplify the working, we shall assume only one degree of freedom, so that the dynamical variables are the co-ordinate  $x$  and the canonical momentum  $p_x$ . We shall also consider the ultra-relativistic limit,  $\gamma_0 \rightarrow \infty$  (so that  $\beta_0 \rightarrow 1$ ).

The transfer map to a given order can be found by applying a Lie transformation to each of the dynamical variables, and expanding the resulting expression to a given order in the dynamical variables. Using the

Hamiltonian for a sextupole in one degree of freedom:

$$H = -\sqrt{1 - p_x^2} + \frac{1}{6}k_2x^3, \quad (9.80)$$

we find the map for a sextupole of length  $L$ :

$$\begin{aligned} x(L) &= x_0 + Lp_{x0} - \frac{1}{4}k_2L^2x_0^2 - \frac{1}{6}k_2L^3x_0p_{x0} - \frac{1}{24}k_2L^4p_{x0}^2 + O(3), \\ &\quad (9.81) \end{aligned}$$

$$p_x(L) = p_{x0} - \frac{1}{2}k_2Lx_0^2 - \frac{1}{2}k_2L^2x_0p_{x0} - \frac{1}{6}k_2L^3p_{x0}^2 + O(3), \quad (9.82)$$

where  $O(3)$  represents terms of third order (and higher) in the dynamical variables. We expect the full map to be symplectic, but after truncating to second order (i.e. neglecting terms of third order and higher) we find:

$$J^T SJ = \begin{pmatrix} 0 & 1 + \Delta \\ -1 - \Delta & 0 \end{pmatrix}, \quad (9.83)$$

where  $J$  is the Jacobian of the map, and:

$$\Delta = \frac{1}{72}k_2^2L^4(6x_0^2 + 6Lx_0p_{x0} + L^2p_{x0}^2). \quad (9.84)$$

To construct a mixed-variable generating function that represents the map, we write a function  $F_3(X, p_x)$  in the form:

$$F_3 = -\left(p_x \ X\right) \sum_{m=1}^{\infty} \frac{\vec{U}^{(m)}}{m+1}, \quad (9.85)$$

where  $\vec{U}^{(m)}$  is a vector (with  $2n$  components in  $n$  degrees of freedom) whose components are power series of order  $m$  in the dynamical variables. Then, the relationship between the ‘old’ variables  $(x, p_x)$  and the ‘new’ variables  $(X, P_X)$  is given by:

$$x = -\frac{\partial F_3}{p_x} = \sum_{m=1}^{\infty} U_1^{(m)}, \quad (9.86)$$

$$P_X = -\frac{\partial F_3}{X} = \sum_{m=1}^{\infty} U_2^{(m)}. \quad (9.87)$$

To find the vectors  $\vec{U}^{(m)}$ , we first re-write (9.81) and (9.82) in the form:

$$x = X - Lp_x + \frac{1}{4}k_2L^2x^2 + \frac{1}{6}k_2L^3xp_x + \frac{1}{24}k_2L^4p_x^2 + O(3), \quad (9.88)$$

$$P_X = p_x - \frac{1}{2}k_2Lx^2 - \frac{1}{2}k_2L^2xp_x - \frac{1}{6}k_2L^3p_x^2 + O(3). \quad (9.89)$$

We find  $\vec{U}^{(1)}$  simply by taking (9.88) and (9.89) to first order in  $x$  and  $p_x$ :

$$\vec{U}^{(1)} = \begin{pmatrix} x \\ P_X \end{pmatrix} = \begin{pmatrix} X - Lp_x \\ p_x \end{pmatrix}. \quad (9.90)$$

The vector  $\vec{U}^{(2)}$  is found by taking the second-order terms in (9.88) and

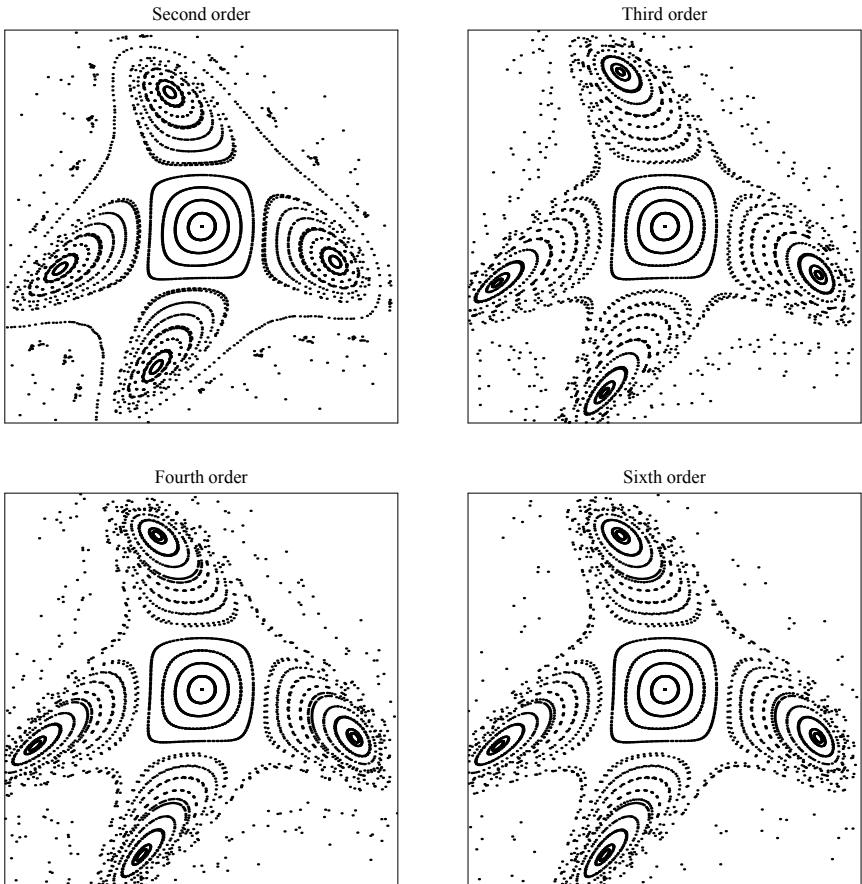


Fig. 9.2 Horizontal phase space portraits of a simple storage ring with a single sextupole. The storage ring lattice excluding the sextupole is represented by a linear transfer map with phase advance  $0.246 \times 2\pi$ . The transformations of the dynamical variables in the sextupole are calculated by numerical solution of the algebraic equations obtained from a mixed-variable generating function of the third kind (given to third order by (9.92)). Each plot shows the results of using a mixed-variable generating function calculated to some order  $n+1$  in the dynamical variables, resulting in a transfer map that is effectively of order  $n$  (shown above the plot). The horizontal axis in each plot is the co-ordinate  $x$ , and the vertical axis is the canonical momentum  $p_x$ .

(9.89), and substituting for  $x$  from  $U_1^{(1)}$ . The result is:

$$\vec{U}^{(2)} = \begin{pmatrix} \frac{1}{4}k_2L^2X^2 - \frac{1}{3}k_2L^3Xp_x + \frac{1}{8}k_2L^4p_x^2 \\ -\frac{1}{2}k_2LX^2 + \frac{1}{2}k_2L^2Xp_x - \frac{1}{6}k_2L^3p_x^2 \end{pmatrix}. \quad (9.91)$$

To continue to the next order, i.e. to calculate  $\vec{U}^{(3)}$ , we would have to take the third-order terms in (9.88) and (9.89), and substitute for  $x$  from  $U_1^{(1)} + U_1^{(2)}$ . The process can be continued to whatever order is desired. Since the transformation derived from any mixed-variable generating function is symplectic, we can truncate the summation in (9.85) at any value of  $m$ , without loss of symplecticity. However, the accuracy of the map will be affected by a truncation. To compute the generating function  $F_3(X, p_x)$  to third order, it is sufficient to calculate the vectors  $\vec{U}^{(m)}$  up to  $m = 2$ . Using (9.90) and (9.91) in (9.85) gives:

$$\begin{aligned} F_3(X, p_x) = & -Xp_x + \frac{1}{2}Lp_x^2 + \frac{1}{6}k_2LX^3 - \frac{1}{4}k_2L^2X^2p_x \\ & + \frac{1}{6}k_2L^3Xp_x^2 - \frac{1}{24}k_2L^4p_x^3 + O(4). \end{aligned} \quad (9.92)$$

Figure 9.2 shows the horizontal phase space in a FODO storage ring with a single sextupole. The storage ring, with the exception of the sextupole is represented by a linear transfer map with phase advance  $0.246 \times 2\pi$ . The transfer map in the sextupole is represented by a mixed-variable generating function; the transformations of the variables are computed by solving the implicit equations (9.64) and (9.65). The parameters used are the same as for Fig. 9.1.

The parameters used are the same as for Fig. 9.1, and it is interesting to compare the tracking results shown in the two figures. The loss of symplecticity from truncating the power series representation of the Lie transformation results in an increased scatter of points in Fig. 9.1, even where the motion should be regular. This is particularly evident in the low-order cases, where the motion found from the mixed-variable generating function map (Fig. 9.2) clearly appears more regular, even at second-order. Both approaches, however, give similar results at higher-order, as we should expect: in principle, increasing the order of the map should give an increasingly accurate approximation to the ‘true’ dynamics. It should be remembered, however, that the preservation of the regularity of the dynamics from use of a symplectic map does not in itself guarantee accuracy: the results of tracking using a symplectic approximation may be less accurate than results of tracking using a non-symplectic method, even if some desirable characteristics of the tracking appear to be better preserved by the symplectic map.

**This page intentionally left blank**

## Chapter 10

# Particle Tracking by Numerical Integration

In Chapter 9 we saw how Lie transformations can be used to construct the transfer map for an accelerator component in the form of a power series, as long as the Hamiltonian has no explicit dependence on the independent variable  $s$ . Although the exact transfer map is symplectic, in most cases an exact representation of the map in the form of a power series would require an infinite number of terms. Truncating the power series at some finite order will generally mean that the resulting map is not symplectic. We also saw that, starting from a truncated power series, a symplectic representation of the map could be reconstructed in the form of a mixed-variable generating function. The drawback of this approach is that a mixed-variable generating function provides an implicit representation of a map, consisting of a set of algebraic equations that must be solved at each step of the integration. Solution of the equations usually needs to be attained numerically, often using an iterative technique, which will impact computational efficiency.

There are of course many numerical integration techniques that could, in principle, be applied to solve the equations of motion for a particle moving through the electromagnetic fields in an accelerator. Ideally, the integration method that is used will satisfy three criteria: the solution will be symplectic, the method will provide the particle momentum as well as the co-ordinates along the trajectory, and it will be computationally efficient. Unfortunately, few techniques meet all these requirements. For example, the symplectic Euler method [Hairer *et al.* (2006b)] is implicit and therefore requires some iteration at each step to solve the equations, which will impact computational efficiency. The explicit Euler method is straightforward to implement, and computationally efficient; however, it is not symplectic. The Störmer–Verlet method [Hairer *et al.* (2006b); Press *et al.* (2007b)] is both symplectic and explicit; but provides a solution only for the co-ordinates, and not for the momentum of a particle.

Fortunately, not all studies of accelerator systems require tracking techniques that have all the features desired of an ‘ideal’ method. It is often the case, for example, that as long as the integration is reasonably accurate, it does not need to be exactly symplectic. This significantly increases the range of possible options: as one example, in Section 10.1 we discuss the extension of the Störmer–Verlet method to provide the velocity (and hence the momentum) as well as the co-ordinates of a particle moving through a three-dimensional ( $s$ -dependent) magnetic field. Although the modification needed to obtain the velocity as well as the co-ordinates means that the integration is no longer exactly symplectic, the symplectic error does remain within some bounds, rather than growing throughout the integration as can be the case with some other methods.

For some accelerator studies, however, such as long-term tracking in storage rings, it is important to have transfer maps that are both symplectic and explicit (i.e. in the form of expressions that just require substitution of particular values for the dynamical variables for their application). A method that produces a symplectic map in explicit form is known as an *explicit symplectic integrator*. Fortunately, methods do exist for constructing explicit symplectic integrators (albeit with some approximations), and in Sections 10.2.1 and 10.2.2, we shall consider techniques based on Lie transformations. The splitting methods discussed in Section 10.2.1 are generally applied where the magnetic field is independent of the distance along the reference trajectory. Where the field does vary with longitudinal position, then the method of Wu, Forest and Robin can be used, as will be discussed in Section 10.2.2.

For completeness, in Section 10.2.3 we consider symplectic Runge–Kutta integrators. Although these methods are implicit, and can be computationally expensive as a result, they can be useful where the approximations required for the explicit symplectic integrators presented in Sections 10.2.1 and 10.2.2 are not valid.

## 10.1 The Störmer–Verlet Method

The Störmer–Verlet method (also known as Verlet integration, the Störmer method, or the leap-frog method) is a technique for numerical solution of second-order ordinary differential equations that describe the motion of conservative systems. There are a number of variations of the technique: in its most basic form, Störmer–Verlet is a symplectic method. However, it only provides the co-ordinates of a system at specified time intervals

(following given initial conditions), and not the momenta. The algorithm can be extended to provide the velocities as well as the co-ordinates, but then the integration is not symplectic. Despite the loss of symplecticity, the *velocity Verlet* method can be an attractive choice in some situations since it provides good accuracy and stability, is relatively fast, and can readily be implemented in a tracking code. The stability of the technique is reflected in the fact that nominal constants of the motion (such as the energy) tend to oscillate about fixed values, rather than drifting away from the initial values as the integration proceeds.

Most implementations of the Störmer–Verlet method have strong restrictions on the systems to which they can be applied. In particular, it is usually necessary that the forces acting on the system depend only on the co-ordinates, and not on the velocity. This would seem to exclude application of the technique to particles moving through a magnetic field; however, the specific properties of magnetic forces make it possible to extend the basic form of the Störmer–Verlet method to apply to particles in magnetic fields. In particular, we can exploit the fact that the acceleration of a particle in a magnetic field is always perpendicular to its velocity, so that the magnitude of the velocity is constant. Unfortunately, the technique cannot be applied to motion in combined electric and magnetic fields. This is certainly a drawback for particle tracking in accelerators, but the Störmer–Verlet method can still be useful in components where only magnetic fields are present, e.g. multipole magnets.

Despite several limitations, the Störmer–Verlet method has some attractive features for numerical tracking of particles through accelerators. In addition to the properties mentioned above, the method applies not just to fields that are independent of the longitudinal co-ordinate (as in, for example, the body of a multipole magnet) but also to cases where the field varies with longitudinal position (for example, in fringe fields). In this section, we shall derive the formulae for Verlet integration of a particle moving through a magnetic field, providing the velocity of the particle along the trajectory as well as the co-ordinates. Our derivation will follow that given by Chambliss and Franklin [Chambliss and Franklin (2020)]. We shall also present an example of the application of the technique in the context of particle tracking in an accelerator beam line.

To begin, let us write the equation of motion for a particle of charge  $q$ , and mass  $m$  in a magnetic field in the form:

$$\frac{d^2\mathbf{x}}{dt^2} = \frac{q}{\gamma m} \mathbf{v} \times \mathbf{B}(\mathbf{x}), \quad (10.1)$$

where  $\mathbf{x}$  is the position vector of the particle at time  $t$ ,  $\gamma$  is the relativistic factor,  $\mathbf{v}$  is the velocity of the particle, and  $\mathbf{B}(\mathbf{x})$  is the magnetic field at the position of the particle. Since the acceleration is perpendicular to the velocity, the magnitude of the velocity  $|\mathbf{v}|$  and the relativistic factor  $\gamma$  are both constant.

Let us suppose that we are given the position  $\mathbf{x}_0$  and velocity  $\mathbf{v}_0$  of a particle at a time  $t = t_0$ . Our aim is to calculate the (approximate) position  $\mathbf{x}_1$  and velocity  $\mathbf{v}_1$  at a later time,  $t = t_0 + \Delta t = t_1$  (i.e. after one time step). Explicit formulae for  $\mathbf{x}_1$  and  $\mathbf{v}_1$  can be derived using the velocity Verlet method as follows. First, we estimate the velocity  $\mathbf{v}_{\frac{1}{2}}$  at time  $t_0 + \frac{1}{2}\Delta t$ :

$$\mathbf{v}_{\frac{1}{2}} = \mathbf{v}_0 + \frac{1}{2}\mathbf{a}_0\Delta t, \quad (10.2)$$

where:

$$\mathbf{a}_0 = \frac{q}{\gamma m}\mathbf{v}_0 \times \mathbf{B}_0, \quad (10.3)$$

is the acceleration of the particle (resulting from the magnetic force) at time  $t_0$ , with  $\mathbf{B}_0 = \mathbf{B}(\mathbf{x}_0)$  the magnetic field at the initial position of the particle,  $\mathbf{x}_0$ . We then estimate the position of the particle at time  $t_1$ :

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{v}_{\frac{1}{2}}\Delta t. \quad (10.4)$$

From this, we can estimate the magnetic field  $\mathbf{B}_1$  at the new position  $\mathbf{x}_1$ ; and the acceleration at time  $t_1$  will then be given by:

$$\mathbf{a}_1 = \frac{q}{\gamma m}\mathbf{v}_1 \times \mathbf{B}_1. \quad (10.5)$$

To complete the algorithm, we therefore need to find the velocity  $\mathbf{v}_1$  at time  $t_1$  (as well as the position  $\mathbf{x}_1$ ). The velocity  $\mathbf{v}_1$  will be given (approximately) by:

$$\mathbf{v}_1 = \mathbf{v}_{\frac{1}{2}} + \frac{1}{2}\mathbf{a}_1\Delta t = \mathbf{v}_{\frac{1}{2}} + \alpha\mathbf{v}_1 \times \mathbf{B}_1, \quad (10.6)$$

where we have substituted for  $\mathbf{a}_1$  from (10.5), and for convenience we define:

$$\alpha \equiv \frac{q\Delta t}{2\gamma m}. \quad (10.7)$$

To find an explicit expression for  $\mathbf{v}_1$  in terms of known quantities, we need to rearrange (10.6). This can be achieved in two steps. First, taking the vector product of each side of (10.6) with  $\mathbf{B}_1$  we have:

$$\mathbf{v}_1 \times \mathbf{B}_1 = \mathbf{v}_{\frac{1}{2}} \times \mathbf{B}_1 + \alpha (\mathbf{v}_1 \times \mathbf{B}_1) \times \mathbf{B}_1, \quad (10.8)$$

$$= \mathbf{v}_{\frac{1}{2}} \times \mathbf{B}_1 + \alpha ((\mathbf{v}_1 \cdot \mathbf{B}_1)\mathbf{B}_1 - \mathbf{B}_1^2 \mathbf{v}_1), \quad (10.9)$$

Second, taking the scalar product of each side of (10.6) with  $\mathbf{B}_1$ , and making use of the fact that  $\mathbf{v}_1 \times \mathbf{B}_1$  is perpendicular to  $\mathbf{B}_1$  (so that  $(\mathbf{v}_1 \times \mathbf{B}_1) \cdot \mathbf{B}_1 = 0$ ) we have:

$$\mathbf{v}_1 \cdot \mathbf{B}_1 = \mathbf{v}_{\frac{1}{2}} \cdot \mathbf{B}_1. \quad (10.10)$$

Finally, substituting for  $\mathbf{v}_1 \cdot \mathbf{B}_1$  from (10.10) into (10.9), and substituting the resulting expression for  $\mathbf{v}_1 \times \mathbf{B}_1$  into (10.6) we obtain, after a little further algebra:

$$\mathbf{v}_1 = \frac{\mathbf{v}_{\frac{1}{2}} + \alpha \mathbf{v}_{\frac{1}{2}} \times \mathbf{B}_1 + \alpha^2 (\mathbf{v}_{\frac{1}{2}} \cdot \mathbf{B}_1) \mathbf{B}_1}{1 + \alpha^2 \mathbf{B}_1^2}. \quad (10.11)$$

The formulae in equations (10.2), (10.4), (10.11) constitute the *magnetic velocity Verlet* method: using these formulae, we can integrate numerically the equations of motion for a (relativistic) charged particle in a magnetic field. Note that the new values of the position and velocity at each step of the integration are expressed purely in terms of known quantities. This means that there is no need to solve (numerically) any algebraic equations when applying each step; in other words, this is an *explicit* numerical integration technique.

There are, of course, many other algorithms available for numerical integration of differential equations. In Section 10.2.3 we shall discuss Runge–Kutta integrators: these provide techniques that are more general, in that they can be applied (for example) to particles moving through regions with electric as well as magnetic fields. A further advantage with Runge–Kutta integration is that versions exist that are symplectic. However, since symplectic Runge–Kutta integrators are implicit, they are more complicated to implement than explicit methods such as Verlet integration, and are comparatively slow in performing the integration. Explicit Runge–Kutta integrators are not symplectic, though for sufficiently small step size they can achieve good accuracy. However, for situations where the magnetic velocity Verlet method can be applied, it has an advantage in providing a more stable integration algorithm than an explicit Runge–Kutta integrator with similar step size working to the same order.

As an example, we shall compare the magnetic velocity Verlet method with an explicit (non-symplectic) Runge–Kutta integrator, applying each method to a particle moving through a combined quadrupole and sextupole field  $\mathbf{B} = (B_x, B_y, B_z)$ , with components given by:

$$\frac{q}{P_0} B_x = k_1 y + k_2 x y, \quad \frac{q}{P_0} B_y = k_1 x + \frac{1}{2} k_2 (x^2 - y^2), \quad B_z = 0. \quad (10.12)$$

Here,  $q$  is the charge of the particle, and  $P_0$  is the reference momentum. For simplicity, we shall consider initial conditions with  $y = v_y = 0$ : the motion of the particle is then entirely in the  $x$ – $z$  plane. Assuming that the particle has zero energy deviation (so that the total momentum is equal to  $P_0$ ), the Hamiltonian for a particle moving in this field can be written:

$$H = -\sqrt{1 - p_x^2} + \frac{1}{2} k_1 x^2 + \frac{1}{6} k_2 x^3. \quad (10.13)$$

We can test the accuracy of the integration by plotting the trajectory of the particle in phase space: since the Hamiltonian is a constant of the motion, as the particle moves through the field the real (physical) trajectory will follow a contour of the Hamiltonian in phase space. Some results with particular initial conditions are shown in Fig. 10.1: the left-hand plot shows the horizontal phase space obtained using the magnetic velocity Verlet method, and the right-hand plot shows the horizontal phase space from a Runge–Kutta integrator, specifically, the explicit midpoint method [Press *et al.* (2007c)]. In both cases, the same initial conditions and step sizes are used.

We choose the explicit midpoint method for the comparison because like the magnetic velocity Verlet method, it is second order in terms of accuracy (i.e. has errors that scale with the square of the step size), and the two methods are both explicit, and require a similar number of calculations at each step. The midpoint method can be applied as follows. Given initial position  $\mathbf{x}_0$ , velocity  $\mathbf{v}_0$  and acceleration  $\mathbf{a}_0$  at time  $t = t_0$ , the position and velocity at time  $t = t_0 + \frac{1}{2}\Delta t$  are first estimated from:

$$\mathbf{x}_{\frac{1}{2}} = \mathbf{x}_0 + \frac{1}{2} \mathbf{v}_0 \Delta t, \quad (10.14)$$

$$\mathbf{v}_{\frac{1}{2}} = \mathbf{v}_0 + \frac{1}{2} \mathbf{a}_0 \Delta t. \quad (10.15)$$

The acceleration  $\mathbf{a}_{\frac{1}{2}}$  is then calculated for a particle at position  $\mathbf{x}_{\frac{1}{2}}$  with velocity  $\mathbf{v}_{\frac{1}{2}}$ . For a particle in a magnetic field:

$$\mathbf{a}_{\frac{1}{2}} = \frac{q}{\gamma m} \mathbf{v}_{\frac{1}{2}} \times \mathbf{B}(\mathbf{x}_{\frac{1}{2}}). \quad (10.16)$$

Then, the position  $\mathbf{x}_1$  and velocity  $\mathbf{v}_1$  at time  $t = t_0 + \Delta t = t_1$  are estimated from:

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{v}_0 \Delta t + \frac{1}{2} \mathbf{a}_0 \Delta t^2, \quad (10.17)$$

$$\mathbf{v}_1 = \mathbf{v}_0 + \mathbf{a}_{\frac{1}{2}} \Delta t. \quad (10.18)$$

We see in Fig. 10.1 that the two methods produce rather different results for our example of a particle moving through a quadrupole magnet with a superposed sextupole field component. Using the magnetic velocity Verlet method, the particle trajectories for a range of different initial conditions follow closely contours of the Hamiltonian, as expected from the symplecticity of the system (reflecting the fact that the Hamiltonian is a conserved quantity). However, when using the explicit midpoint method, each phase space trajectory diverges from the contour on which it starts, and the motion is eventually unstable, even where it should, in reality, be periodic.

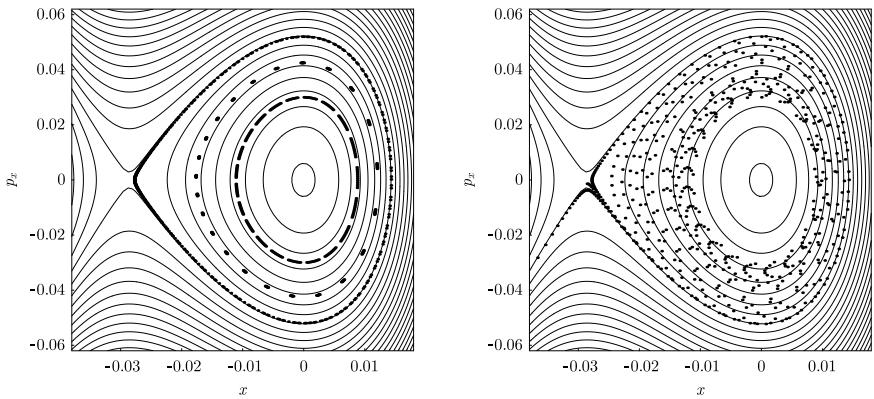


Fig. 10.1 Phase space trajectories of particles in a quadrupole magnet with superposed sextupole component (10.12), calculated using Verlet integration (left) and using the midpoint method (a Runge–Kutta algorithm, right). In each plot, solid lines show contours of constant value of the Hamiltonian (10.13), with  $k_1 = 10 \text{ m}^{-2}$  and  $k_2 = 450 \text{ m}^{-3}$ . Three sets of initial conditions are used:  $x_0 = 0$  in every case, while the values of  $p_{x0}$  are 0.03, 0.0425 and 0.052. The step size is  $c\Delta t = 0.1 \text{ m}$  for  $p_{x0} = 0.03$  and  $p_{x0} = 0.0425$ , and  $c\Delta t = 0.05 \text{ m}$  for  $p_{x0} = 0.052$ . The quadrupole component of the field results in oscillatory motion in the horizontal plane; the sextupole component leads to some non-linear distortion of the phase space. For each set of initial conditions, the particle is tracked for a total distance of approximately 25 m, or approximately 12 periods of oscillation, with the points in the phase space plots showing the phase space co-ordinates after each step. In the left-hand plot (Verlet method) the points follow closely contours of the Hamiltonian. In the right-hand plot (midpoint method) the points move further from the initial contour as tracking proceeds, and the motion is eventually unstable.

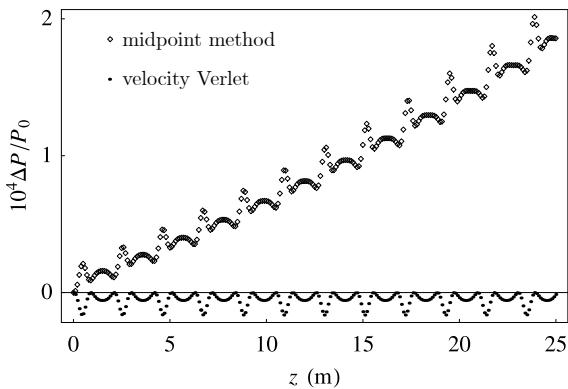


Fig. 10.2 Variation of momentum over the course of tracking a particle through a quadrupole magnet with superposed sextupole component (10.12). Tracking is performed using the midpoint method (open diamonds, positive values) or with the velocity Verlet method (dots, negative values). The parameters are the same as for the results shown in Fig. 10.1; the initial conditions in the case shown here are  $x_0 = 0$ ,  $p_{x0} = 0.03$ , and the step size is  $c\Delta t = 0.1$  m.

The difference between the two integration methods is further illustrated in Fig. 10.2, which shows the variation in the total momentum of the particle over the course of tracking through the combined quadrupole and sextupole field. Since the force on the particle is perpendicular to its direction of motion, the total momentum should be conserved. Figure 10.2 shows that neither the magnetic velocity Verlet method nor the explicit midpoint method exactly conserves the momentum. However, in the case of the magnetic velocity Verlet method, the momentum remains close to its initial value over the course of the tracking, though there is some small periodic variation; by contrast, in the case of the explicit midpoint method, while there is again some small periodic variation of the momentum, this is superposed on a steady drift of the momentum away from the initial value. The behaviour of the momentum in each case is reflected in the phase space plots in Fig. 10.1, where a steady drift in the momentum in the case of the explicit midpoint method is associated with the divergence of the phase space trajectory from a contour of fixed value of the Hamiltonian. Even though neither method is symplectic, the magnetic velocity Verlet method may (in certain circumstances) have some advantages over alternative methods, from point of view of the preservation of conserved quantities.

## 10.2 Symplectic Integrators

### 10.2.1 *Splitting methods*

Our goal in this section is to develop a technique, based on Lie transformations, for constructing symplectic transfer maps that can be represented as power series with a finite number of terms. In general, even for ‘linear’ components such as quadrupoles, it is not possible to find an exact, closed-form solution to Hamilton’s equations. For that reason, in Chapter 3 we constructed the transfer maps for linear components by making an approximation to the Hamiltonian (specifically, expanding the Hamiltonian to second order in the dynamical variables), so that the equations of motion derived from the approximate Hamiltonian did admit exact, closed-form solutions. In other words, we approximated the Hamiltonian rather than the solution to Hamilton’s equations: this ensured the solution we obtained was symplectic. Unfortunately, for higher-order components such as sextupoles, this technique fails, because making a second-order approximation to the Hamiltonian for these components means that we lose their significant effects. However, it turns out that for higher-order multipoles we can follow an approach that, while not exactly the same as the one we used for dipoles and quadrupoles, is broadly similar in principle. In certain cases, it is possible to make an approximation to the generator of a Lie transformation in such a way that, with the approximate generator, the Lie transformation has an exact representation as a power series map with a finite number of terms. One way to do this is to ‘split’ the original generator into a number of terms, each of which provides a generator for a transformation that has an exact (and finite) power series representation. An approximation for the original transformation is then constructed from the composition of the new transformations.

As a simple example, consider the Hamiltonian for a sextupole magnet in one degree of freedom (with dynamical variables  $x$  and  $p_x$ ) and in the ultra-relativistic limit  $\gamma_0 \rightarrow \infty$ :

$$H = -\sqrt{1 - p_x^2} + \frac{1}{6}k_2x^3. \quad (10.19)$$

Applying the Lie transformation  $e^{-L:H}$ : (representing a sextupole with length  $L$ ) to the dynamical variables produces power series with infinitely many terms. However, we note that we can write the Hamiltonian as the sum of two terms:

$$H = H_d + H_k, \quad (10.20)$$

where:

$$H_d = -\sqrt{1 - p_x^2}, \quad (10.21)$$

and:

$$H_k = \frac{1}{6}k_2x^3. \quad (10.22)$$

If the Lie transformation  $e^{-L:H_k:}$  is applied to the dynamical variables, the result can be written exactly:

$$e^{-L:H_k:} \begin{pmatrix} x \\ p_x \end{pmatrix} = \begin{pmatrix} x \\ p_x - \frac{1}{2}k_2Lx^2 \end{pmatrix}. \quad (10.23)$$

Then, applying the Lie transformation  $e^{-L:H_d:}$  to the right-hand side of (10.23), we obtain:

$$\begin{aligned} e^{-L:H_d:} e^{-L:H_k:} \begin{pmatrix} x \\ p_x \end{pmatrix} &= e^{-L:H_d:} \begin{pmatrix} x \\ p_x - \frac{1}{2}k_2Lx^2 \end{pmatrix} \\ &= \begin{pmatrix} x + \frac{Lp_x}{\sqrt{1-p_x^2}} \\ p_x - \frac{1}{2}k_2Lx^2 - \frac{k_2L^2xp_x}{\sqrt{1-p_x^2}} - \frac{k_2L^3p_x^2}{2(1-p_x^2)} \end{pmatrix}. \end{aligned} \quad (10.24)$$

If it can be shown that:

$$e^{-L:H_d:} e^{-L:H_k:} \approx e^{-L:H_d+H_k:} = e^{-L:H:}, \quad (10.25)$$

then (10.24) provides an approximation to the transfer map for a sextupole. The approximate map is explicit, since it is given in the form of power series, which only require substitution of the values of the dynamical variables for their evaluation. The approximate map is also symplectic, since it is obtained from a Lie transformation without further approximation (in particular, without truncating terms from the power series).

Our task now is to put (10.25) on a more rigorous footing; that is, to find an expression for the composition of two Lie transformations. In doing so, we shall see that, while it is possible to make the approximation in (10.25), there are better approximations that can be made. However, we first make some remarks about the result obtained in (10.24). In particular, note that the result was obtained by first applying the Lie transformation  $e^{-L:H_k:}$  to the dynamical variables, and then applying the Lie transformation  $e^{-L:H_k:}$  to the resulting expressions. However, the same result can be obtained in a slightly different way. Specifically, it is possible to apply each of the Lie transformations  $e^{-L:H_d:}$  and  $e^{-L:H_k:}$  separately to the dynamical variables;

then substitute the result obtained from the first transformation into the result of the second. In other words, if:

$$e^{:f:}\vec{x} = \vec{F}(\vec{x}), \quad (10.26)$$

$$e^{:g:}\vec{x} = \vec{G}(\vec{x}), \quad (10.27)$$

then:

$$e^{:f:}e^{:g:}\vec{x} = \vec{G}\left(\vec{F}(\vec{x})\right). \quad (10.28)$$

This is perhaps most easily understood by thinking of a Lie transformation as a canonical transformation from one set of dynamical variables to another: the Lie transformation with generator  $g$  gives expressions for the new variables  $\vec{X} = \vec{G}(\vec{x})$  in terms of the original variables  $\vec{x}$ . The Lie transformation with generator  $f$  again acts on the original variables, regardless of the fact that they now appear in expressions for a new set of variables. Hence, the order in which Lie transformations appear when applied in sequence is the same order in which the transformations are applied. This is in contrast to the situation with transfer matrices, which appear in an expression in the *reverse* order to which the corresponding transformations are applied. The difference arises because transfer matrices act on *values*, whereas Lie transformations act on *variables*.

To find an expression for the composition of two Lie transformations it is helpful to write down some of their properties. The most important properties of Lie transformations for our purpose follow from the properties of Lie operators. We define the commutator  $\{:f:, :g:\}$  in the usual way:

$$\{ :f:, :g: \} = :f::g: - :g::f:. \quad (10.29)$$

Using the explicit form of the Lie operator (9.4), it is possible to derive three significant results. First, for all functions  $f$  of the dynamical variables:

$$\{ :f:, :f: \} = 0. \quad (10.30)$$

Second:

$$\{ a:f: + b:g:, :h: \} = a\{ :f:, :h: \} + b\{ :g:, :h: \}, \quad (10.31)$$

where  $a$  and  $b$  are constants. And third, it can be shown that Lie operators satisfy the Jacobi identity:

$$\{ :f:, \{ :g:, :h: \} \} + \{ :g:, \{ :h:, :f: \} \} + \{ :h:, \{ :f:, :g: \} \} = 0. \quad (10.32)$$

The properties (10.30), (10.31) and (10.32) are the defining properties of a *Lie algebra*. For any operators that form a Lie algebra, the product of exponentials of the operators is given by the Baker–Campbell–Hausdorff (BCH) formula; thus, for Lie transformations:

$$e^{a:f:} e^{b:g:} = e^{:h:}, \quad (10.33)$$

where  $:h:$  is given by:

$$\begin{aligned} :h: &= a:f: + b:g: + \frac{ab}{2}\{f:,g:\} \\ &\quad + \frac{a^2b}{12}\{f:\{f:,g:\}\} + \frac{ab^2}{12}\{g:\{g:,f:\}\} + O(4). \end{aligned} \quad (10.34)$$

$O(4)$  represents terms of fourth order and higher in products of the constants  $a$  and  $b$ . Unfortunately, the BCH formula (10.34) contains an infinite number of terms, and cannot be written down completely; however, it does give us the expression we were seeking for the composition of two Lie transformations.

For Lie transformations, the BCH formula can be written in a slightly more elegant way using the Poisson bracket  $[f, g]$  defined (in  $n$  degrees of freedom) by:

$$[f, g] = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial x_i} \frac{\partial f}{\partial p_i}, \quad (10.35)$$

where  $x_i$  are the co-ordinates and  $p_i$  the canonical momenta. Using the explicit form for the Lie operator (9.4), it can be shown that:

$$\{f:,g:\} = :[f, g]:. \quad (10.36)$$

Hence, the BCH formula (10.34) can be written:

$$h = af + bg + \frac{ab}{2}[f, g] + \frac{a^2b}{12}[f, [f, g]] + \frac{ab^2}{12}[g, [g, f]] + O(4). \quad (10.37)$$

Associated with the BCH formula is the Zassenhaus formula, which gives an expression for factorising an exponential of the sum of two elements of a Lie algebra. In the case of Lie transformations (where the commutator can be replaced by the Poisson bracket) the Zassenhaus formula can be written:

$$e^{f+g:} = e^{f:} e^{g:} e^{-\frac{1}{2}:[f,g]:} e^{\frac{1}{3}:[g,[f,g]]:+\frac{1}{6}:[f,[f,g]]:} \dots \quad (10.38)$$

As may be expected, the Zassenhaus formula involves an infinite product of factors.

We shall now show how to use the BCH and Zassenhaus formulae to construct explicit symplectic integrators for accelerator components. Suppose that the Hamiltonian for a particular component can be written as a sum of two terms:

$$H = H_d + H_k, \quad (10.39)$$

where each of the terms  $H_d$  or  $H_k$  on its own corresponds to an integrable Hamiltonian, i.e. a Hamiltonian for which Hamilton's equations admit of an exact closed-form solution. In the case of a sextupole magnet in one degree of freedom,  $H_d$  and  $H_k$  would be given by (10.21) and (10.22) respectively. Using the Zassenhaus formula, we find:

$$e^{-\epsilon L:H_d+H_k:} = e^{-\epsilon L:H_d:} e^{-\epsilon L:H_k:} e^{-\frac{1}{2}\epsilon^2 L^2:[H_d,H_k]:} \dots, \quad (10.40)$$

where  $\epsilon$  is a dimensionless parameter, the purpose of which is to help us to keep track of the order of various terms. The factors not written explicitly in (10.40) involve exponentials of terms of order  $\epsilon^3$ . The approximation (10.25) for the transfer map for a sextupole is obtained by neglecting terms of order  $\epsilon^2$  and higher in (10.40): the Zassenhaus formula provides a foundation for the approximation (10.25).

Suppose that the Hamiltonian is expressed not as the sum of two terms, but as the sum of three terms:

$$H = \frac{1}{2}H_d + H_k + \frac{1}{2}H_d. \quad (10.41)$$

This is clearly equivalent to (10.39), but the reason for writing the Hamiltonian in this particular way will become clear in a moment. Applying the Zassenhaus formula gives:

$$e^{-\epsilon L:\frac{1}{2}H_d+H_k+\frac{1}{2}H_d:} = e^{-\frac{1}{2}\epsilon L:H_d:} e^{-\epsilon L:H_k+\frac{1}{2}H_d:} e^{-\frac{1}{2}\epsilon^2 L^2:[\frac{1}{2}H_d,H_k+\frac{1}{2}H_d]:} \dots \quad (10.42)$$

Using the bilinearity of the Poisson bracket:

$$\left[ \frac{1}{2}H_d, H_k + \frac{1}{2}H_d \right] = \left[ \frac{1}{2}H_d, H_k \right] + \left[ \frac{1}{2}H_d, \frac{1}{2}H_d \right] = \left[ \frac{1}{2}H_d, H_k \right], \quad (10.43)$$

and a further application of the Zassenhaus formula, we obtain:

$$\begin{aligned} e^{-\epsilon L:\frac{1}{2}H_d+H_k+\frac{1}{2}H_d:} &= e^{-\frac{1}{2}\epsilon L:H_d:} \left( e^{-\epsilon L:H_k:} e^{-\frac{1}{2}\epsilon L:H_d:} e^{-\frac{1}{2}\epsilon^2 L^2:[H_k,\frac{1}{2}H_d]:} \dots \right) \\ &\quad \times e^{-\frac{1}{2}\epsilon^2 L^2:[\frac{1}{2}H_d,H_k]:} \dots \end{aligned} \quad (10.44)$$

Neglecting terms of order  $\epsilon^3$  and higher then gives:

$$\begin{aligned} e^{-\epsilon L:\frac{1}{2}H_d+H_k+\frac{1}{2}H_d:} &\approx e^{-\frac{1}{2}\epsilon L:H_d:} e^{-\epsilon L:H_k:} e^{-\frac{1}{2}\epsilon L:H_d:} \\ &\quad \times e^{-\frac{1}{2}\epsilon^2 L^2:[H_k,\frac{1}{2}H_d]:} e^{-\frac{1}{2}\epsilon^2 L^2:[\frac{1}{2}H_d,H_k]:}. \end{aligned} \quad (10.45)$$

The factors with generator of order  $\epsilon^2$  have the same generator to within a minus sign. Therefore, using the fact (as can be seen from the BCH formula, for example) that:

$$e^{f_1} e^{-f_2} = 1, \quad (10.46)$$

the final result is:

$$e^{-\epsilon L: \frac{1}{2}H_d + H_k + \frac{1}{2}H_d:} \approx e^{-\frac{1}{2}\epsilon L:H_d:} e^{-\epsilon L:H_k:} e^{-\frac{1}{2}\epsilon L:H_d:}, \quad (10.47)$$

where the approximation comes from neglecting terms with generators of order  $\epsilon^3$  and higher. Notice that terms of order  $\epsilon^2$  have cancelled: this is a consequence of writing the Hamiltonian in a symmetric form, and provides an improvement (that is, a higher-order approximation) compared with the result obtained by writing the Hamiltonian simply as a sum of two terms as in (10.39). If the terms  $H_d$  and  $H_k$  correspond to integrable Hamiltonians, then the right-hand side of (10.47) can be applied to each of the dynamical variables to produce a transfer map in the form of power series in the dynamical variables, with each power series having a finite number of terms. Since application of the map requires only substitution of specific values for the dynamical variables, the map is explicit. The map is obtained by applying Lie transformations to the dynamical variables without truncating the resulting power series: therefore, the map will be symplectic. Equation (10.47) provides a second-order explicit symplectic integrator.

In the case of a sextupole (for example), where  $H_d$  and  $H_k$  are given by (10.21) and (10.22), the integrator (10.47) has a simple physical interpretation: it represents three successive transformations. The first transformation corresponds to a drift through half the length of the sextupole; the second transformation to a ‘kick’ (i.e. an instantaneous change in the momenta, without any change in the co-ordinates); and the third transformation to another drift, through the second half of the sextupole. This is sometimes known as a *drift-kick-drift approximation*. It provides a better (higher-order) approximation than a drift-kick approximation, in which the kick is applied after a drift through the entire length of the sextupole. The cost is that three transformations are required, rather than two.

Rather than splitting the Hamiltonian as (10.41) it could be split instead as follows:

$$H = \frac{1}{2}H_k + H_d + \frac{1}{2}H_k. \quad (10.48)$$

This represents a kick-drift-kick approximation, and also leads to a second-order explicit symplectic integrator. Depending on the exact form of the

Hamiltonian involved, there may be some advantage in computational efficiency in using one approximation (either (10.41) or (10.48)) over the other.

The process leading to (10.47) is known as *symmetric factorisation*, or *Yoshida factorisation*. Explicit symplectic integrators constructed using *splitting methods* (in which the Hamiltonian is expressed as a sum of integrable terms) have been known for some time and are widely used [Ruth (1983); Forest and Ruth (1990); Yoshida (1990); Forest (2006); Chao (2022c)]. The goal is to look for the optimum way of splitting the Hamiltonian into different terms, to produce an approximation of the desired order, with the final map in as simple a form as possible.

As an example of applying symmetric factorisation to produce a fourth-order explicit symplectic integrator, consider the Hamiltonian (10.39) split into terms as follows:

$$H = d_1 H_d + c_1 H_k + d_2 H_d + c_2 H_k + d_3 H_d + c_4 H_k + d_4 H_d. \quad (10.49)$$

Substituting this into the Zassenhaus formula, one can look for values of the constants  $d_1$ ,  $d_2$ ,  $c_1$  and  $c_2$  that lead to cancellation of terms of order  $\epsilon^2$  and  $\epsilon^3$  when the Lie transformation  $e^{-\epsilon L:H}$  is factorised. The solution is:

$$d_1 = \frac{1}{12} \left( 4 + 2\sqrt[3]{2} + \sqrt[3]{4} \right), \quad (10.50)$$

$$d_2 = \frac{1}{2} - d_1, \quad (10.51)$$

$$c_1 = 2d_1, \quad (10.52)$$

$$c_2 = 1 - 4d_1. \quad (10.53)$$

In the case of a sextupole (for example), with  $H_d$  and  $H_k$  given by (10.21) and (10.22), the Hamiltonian (10.49) corresponds to a sequence of drifts and kicks of different lengths and amplitudes. The technique of symmetric factorisation, using Lie transformations and the Zassenhaus formula, shows that there is a way of choosing the relative sizes of the drifts and kicks to optimise the accuracy of the integrator. In the case of the fourth-order integrator where the Hamiltonian is split into terms with coefficients (10.50)–(10.53), the relative lengths of the drift transformations are  $d_1 \approx 0.676$ , and  $d_2 \approx -0.176$ : so some of the transformations actually correspond to drifting *backwards* through the accelerator component (though  $2d_1 + 2d_2 = 1$ , so the total drift distance is equal to the length of the component). Although it is not intuitively obvious, this does provide a better approximation to the exact transfer map than simply splitting the component into a number of drifts and kicks of equal length and amplitude.

Although it can be useful in many circumstances, the fourth-order integrator (10.49) does have two drawbacks. First, it requires the evaluation of seven Lie transformations, which (depending on the form of the Hamiltonian) may be computationally expensive. Second, the appearance of negative drifts can lead to numerical instability if the step size of the integration is large. It turns out that it is possible to develop splitting methods that are computationally more efficient than a Yoshida factorisation with comparable accuracy and achieve better numerical stability by avoiding negative drifts. Such methods were first proposed by McLachlan [McLachlan (1995)], and developed more systematically by Laskar and Robutel [Laskar and Robutel (2001)]. The basic technique is as follows. Consider a Hamiltonian of the form:

$$H = H_d + \eta H_k, \quad (10.54)$$

where  $\eta$  is a small parameter. Rather than split this Hamiltonian into seven terms, as in (10.49), we split it into five as follows:

$$H = \bar{d}_1 H_d + \bar{c}_1 \eta H_k + \bar{d}_2 H_d + \bar{c}_1 \eta H_k + \bar{d}_1 H_d. \quad (10.55)$$

We construct Lie transformations from each term in (10.55), and write:

$$e^{-\epsilon L:K:} = e^{-\epsilon L \bar{d}_1 H_d} e^{-\epsilon L \bar{c}_1 \eta H_k} e^{-\epsilon L \bar{d}_2 H_d} e^{-\epsilon L \bar{c}_1 \eta H_k} e^{-\epsilon L \bar{d}_1 H_d}, \quad (10.56)$$

where  $K$  can be expressed as a series in  $\epsilon$  using the BCH formula. We can find values for the coefficients  $\bar{c}_1$ ,  $\bar{c}_2$  and  $\bar{d}_1$  such that:

$$K = H + O(\epsilon^2). \quad (10.57)$$

In a fourth-order integrator, the difference  $K - H$  would be of order  $\epsilon^3$ . We do not have enough coefficients in (10.56) to be able to achieve this; however, if  $\eta$  is small we can do the next best thing, which is to arrange for terms of order  $\epsilon^2 \eta$  to vanish, leaving terms in  $\epsilon^2 \eta^2$  as the lowest-order residuals in  $K - H$ . For small  $\eta$  we obtain an integration method that is more accurate than the second-order integrator (10.47). Furthermore, it turns out to be possible to find a solution in which  $\bar{c}_1$ ,  $\bar{c}_2$  and  $\bar{d}_1$  are all positive, thus avoiding negative drifts. This means that the integrator will have better numerical stability than the fourth-order integrator based on the splitting in (10.49). Also, since we have split the Hamiltonian into five terms rather than seven, we improved computational efficiency compared to the fourth-order integrator. The resulting integrator of the form (10.56)

is known as SABA<sub>2</sub>, and has values for the coefficients as follows:

$$\bar{c}_1 = \frac{1}{2} \left( 1 - \frac{1}{\sqrt{3}} \right), \quad (10.58)$$

$$\bar{c}_2 = \frac{1}{\sqrt{3}}, \quad (10.59)$$

$$\bar{d}_1 = \frac{1}{2}. \quad (10.60)$$

To improve the accuracy of the SABA<sub>2</sub> scheme, it is possible in some cases to add corrections [Laskar and Robutel (2001)] so that the integrator takes the form:

$$\begin{aligned} e^{-\epsilon L:H:} &\approx \\ e^{-\epsilon^3 \eta^2 L \bar{c}_2 :H_c:} e^{-\epsilon L \bar{d}_1 :H_d:} e^{-\epsilon L \bar{c}_1 \eta :H_k:} e^{-\epsilon L \bar{d}_2 :H_d:} e^{-\epsilon L \bar{c}_1 \eta :H_k:} e^{-\epsilon L \bar{d}_1 :H_d:} e^{-\epsilon^3 \eta^2 L \bar{c}_2 :H_c:}, \end{aligned} \quad (10.61)$$

where:

$$H_c = [[H_d, H_k], H_k], \quad (10.62)$$

and:

$$\bar{c}_2 = \frac{2 - \sqrt{3}}{48}. \quad (10.63)$$

The generator  $H_c$  gives an integrable transformation if  $H_d$  is quadratic in the momenta, and  $H_k$  depends only on the co-ordinates: this will be the case, for example, in a multipole magnet. The integrator represented by (10.61) is known as SABA<sub>2</sub>C. By introducing the correction terms, we lose some of the computational efficiency of the SABA<sub>2</sub> integrator, but we retain the numerical stability that comes from avoiding negative drifts. Other variations on the SABA<sub>2</sub> integrator are possible, and the technique can be extended up to a higher order. However, the SABA<sub>2</sub>C integrator provides a robust, accurate and reasonably efficient method for symplectic integration through multipole magnets in accelerators [Nadolski and Laskar (2002); Skokos and Papaphilippou (2008)].

It is worth mentioning that a significant amount of work has been done in developing techniques for manipulating Lie transformations, including methods to provide different factorisations of a given Lie transformation, and for constructing Lie transformations that represent given power series up to some order. Lie transformations provide a set of powerful tools not just for the construction of explicit symplectic integrators, but also for the more general analysis of nonlinear dynamics in accelerator beam lines: see, for example, [Forest (1998); Dragt (2020)].

### 10.2.2 Explicit symplectic integrator for $s$ -dependent fields

The technique presented in Section 10.2.1 for constructing explicit symplectic integrators is valid only when the Hamiltonian has no explicit dependence on the independent variable  $s$ . Despite this limitation, symmetric factorisation is still a useful technique in many situations (for example, where the field in a magnet can be represented as a sum over multipole components). However, there are cases in accelerator beam dynamics when we wish to construct the transfer map for a component in which the field is a continuously varying function of the longitudinal position. An example is an undulator or wiggler magnet, in which the field might be approximated by (1.94)–(1.96):

$$B_x = -B_0 \frac{k_x}{k_y} \sin(k_x x) \sinh(k_y y) \cos(k_s s), \quad (10.64)$$

$$B_y = B_0 \cos(k_x x) \cosh(k_y y) \cos(k_s s), \quad (10.65)$$

$$B_z = -B_0 \frac{k_s}{k_y} \cos(k_x x) \sinh(k_y y) \sin(k_s s), \quad (10.66)$$

where  $B_0$  is the field amplitude, and  $k_s$  is the period of the oscillation of the field along the axis of the magnet (which we take to define the reference trajectory), and  $k_x$  is a constant that describes the rate at which the field strength ‘rolls off’ with increasing distance along the  $x$  axis. Maxwell’s equations are satisfied by the above field components if:

$$k_y^2 = k_x^2 + k_s^2. \quad (10.67)$$

In practice, real undulators will have more complicated fields, but it is often possible to achieve a good representation of the field using a sum over modes with the form (10.64)–(10.66) with various values of  $B_0$ ,  $k_s$  and  $k_x$ .

To write down the Hamiltonian, we need expressions for the components of the vector potential. As discussed in Chapter 1, it is always possible to choose a gauge in which one component of the vector potential vanishes. With the choice of gauge  $A_x = 0$ , a potential leading to the field (10.64)–(10.66) is:

$$A_x = 0, \quad (10.68)$$

$$A_y = -B_0 \frac{k_s}{k_x k_y} \sin(k_x x) \sinh(k_y y) \sin(k_s s), \quad (10.69)$$

$$A_z = -B_0 \frac{1}{k_y} \sin(k_x x) \cosh(k_y y) \cos(k_s s). \quad (10.70)$$

The Hamiltonian with the vector potential (10.68)–(10.70) will have an explicit dependence on the independent variable  $s$ ; hence, it will not be

possible to construct an explicit symplectic transfer map for the undulator using the integrator described in Section 10.2.1. However, there is an alternative method, developed by Wu, Forest and Robin [Wu *et al.* (2003)], that can be applied to cases such as this. The basis of the technique is to extend phase space by making  $s$  a dynamical variable, introducing a new canonical momentum  $p_s$  conjugate to  $s$ , and introducing also a new independent variable  $\sigma$ . To justify this step, consider a general Hamiltonian:

$$H = H(x_i, p_i; s), \quad (10.71)$$

where  $x_i$  are the co-ordinates,  $p_i$  are the conjugate momenta, and  $s$  is the independent variable. Hamilton's equations give the equations of motion for the dynamical variables:

$$\frac{dx_i}{ds} = \frac{\partial H}{\partial p_i}, \quad (10.72)$$

$$\frac{dp_i}{ds} = -\frac{\partial H}{\partial x_i}. \quad (10.73)$$

Let the Hamiltonian in the extended phase space be  $\bar{H}$ , where:

$$\bar{H} = H(x_i, p_i; s) + p_s. \quad (10.74)$$

Note that  $\bar{H}$  has no explicit dependence on the new independent variable  $\sigma$ . The equations of motion for the original dynamical variables are the same as before:

$$\frac{dx_i}{d\sigma} = \frac{\partial \bar{H}}{\partial p_i} = \frac{\partial H}{\partial p_i}, \quad (10.75)$$

$$\frac{dp_i}{d\sigma} = -\frac{\partial \bar{H}}{\partial x_i} = -\frac{\partial H}{\partial x_i}. \quad (10.76)$$

The equations of motion for the new dynamical variables  $(s, p_s)$  are:

$$\frac{ds}{d\sigma} = \frac{\partial \bar{H}}{\partial p_s} = 1, \quad (10.77)$$

$$\frac{dp_s}{d\sigma} = -\frac{\partial \bar{H}}{\partial s} = -\frac{\partial H}{\partial s}. \quad (10.78)$$

With an appropriate choice for the constant of integration, (10.77) gives simply:

$$s = \sigma. \quad (10.79)$$

The equation of motion for  $p_s$  could be quite complicated; however,  $p_s$  is a rather abstract variable, and the solution to (10.78) is not always required.

Given the relationship between the old and new independent variables expressed by (10.79), it may appear that little has been achieved by extending the phase space, and that any change is conceptual rather than practical. However, the fact that the independent variable  $s$  can now be treated like any of the other co-ordinates is crucial, as we shall see. The particle motion is now described by four co-ordinates,  $x$ ,  $y$ ,  $z$  (which is really the time at which a particle arrives at a particular position along the reference trajectory) and  $s$ , and their corresponding canonical momenta. Since the Hamiltonian  $\bar{H}$  has no explicit dependence on the independent variable  $\sigma$ , we can express the evolution of any function of the phase space variables in terms of a Lie transformation:

$$f|_{\sigma=\sigma_0+\Delta\sigma} = e^{-\Delta\sigma:\bar{H}:} f\Big|_{\sigma=\sigma_0}. \quad (10.80)$$

Now consider the Hamiltonian for an accelerator component. For simplicity, we shall consider a straight reference trajectory, although it is possible to apply the method we now describe even if the reference trajectory is curved. We shall assume that a gauge is chosen for the vector potential in which  $A_x = 0$ . In the extended phase space, the Hamiltonian is:

$$\bar{H} = \frac{\delta}{\beta_0} - \sqrt{\left(\frac{1}{\beta_0} + \delta\right)^2 - p_x^2 - (p_y - a_y)^2} - \frac{1}{\beta_0^2 \gamma_0^2} - a_s + p_s. \quad (10.81)$$

In general,  $a_y$  and  $a_s$  will both depend on the co-ordinates  $x$ ,  $y$  and  $s$ . This means that, as it stands, this Hamiltonian cannot be split into integrable terms. To proceed, it is necessary to make the paraxial approximation; in this case, the square root is expanded as a power series to second order in  $p_x$ ,  $p_y$  and  $a_y$ . This approximation is valid if these variables are small. The Hamiltonian is then:

$$\bar{H} \approx H_1 + H_2 + H_3, \quad (10.82)$$

where:

$$H_1 = \frac{\delta}{\beta_0} - D + \frac{p_x^2}{2D} + p_s, \quad (10.83)$$

$$H_2 = \frac{(p_y - a_y)^2}{2D}, \quad (10.84)$$

$$H_3 = -a_s, \quad (10.85)$$

and:

$$D = \sqrt{1 + \frac{2\delta}{\beta_0} + \delta^2}. \quad (10.86)$$

$H_3$  depends only on the co-ordinates: in a system with Hamiltonian given just by  $H_3$ , the co-ordinates will be constant, and the system will be integrable.  $H_1$  depends only on the momenta: again, this corresponds to an integrable system. The remaining term,  $H_2$ , depends on the momentum  $p_y$  as well as on the co-ordinates. As we now show, a Lie transformation with generator  $H_2$  can be written as a composition of integrable transformations (i.e. a composition of Lie transformations with integrable generators).

To express the Lie transformation  $e^{-\Delta\sigma:H_2:}$  as a composition of integrable transformations, we make use of equations (9.19) and (9.20). Using these formulae, it can be seen that if:

$$e^{:I_y:} p_y = p_y - a_y, \quad (10.87)$$

for some function  $I_y$ , then:

$$e^{-\Delta\sigma:H_2:} = e^{:I_y:} \exp\left(-\Delta\sigma: \frac{p_y^2}{2D}: \right) e^{-:I_y:}. \quad (10.88)$$

A function  $I_y$  with the desired property (10.87) is given by:

$$I_y = - \int_0^y a_y(x, y', s) dy'. \quad (10.89)$$

Using the second-order explicit symplectic integrator (10.47), the Lie transformation  $e^{-\Delta\sigma:\bar{H}:}$  can now be written as a composition of integrable Lie transformations. There are various ways to do this, depending on exactly how one chooses to split the Hamiltonian. For example:

$$\begin{aligned} e^{-\Delta\sigma:\bar{H}:} &\approx e^{-\Delta\sigma:H_1+H_2+H_3:} \\ &\approx e^{-\frac{\Delta\sigma}{2}:H_1+H_3:} e^{-\Delta\sigma:H_2:} e^{-\frac{\Delta\sigma}{2}:H_1+H_3:} \\ &\approx e^{-\frac{\Delta\sigma}{4}:H_1:} e^{-\frac{\Delta\sigma}{2}:H_3:} e^{-\frac{\Delta\sigma}{4}:H_1:} e^{-\Delta\sigma:H_2:} e^{-\frac{\Delta\sigma}{4}:H_1:} e^{-\frac{\Delta\sigma}{2}:H_3:} e^{-\frac{\Delta\sigma}{4}:H_1:}, \end{aligned} \quad (10.90)$$

where  $e^{-\Delta\sigma:H_2:}$  should be written as in (10.88). To complete the process, it is only necessary to write down explicit formulae for the transformations

associated with each of the Lie transformations in (10.90). In particular:

$$e^{-\frac{\Delta\sigma}{4}:H_1:x} = x + \frac{p_x}{D} \frac{\Delta\sigma}{4}, \quad (10.91)$$

$$e^{-\frac{\Delta\sigma}{4}:H_1:z} = z + \left( \frac{1}{\beta_0} - \left( \frac{p_x^2}{2D^3} + \frac{1}{D} \right) \left( \frac{1}{\beta_0} + \delta \right) \right) \frac{\Delta\sigma}{4}, \quad (10.92)$$

$$e^{-\frac{\Delta\sigma}{4}:H_1:s} = s + \frac{\Delta\sigma}{4}, \quad (10.93)$$

$$e^{\pm:I_y:p_x} = p_x \mp \int_0^y \frac{\partial a_y(x, y', s)}{\partial x} dy', \quad (10.94)$$

$$e^{\pm:I_y:p_y} = p_y \mp a_y(x, y, s), \quad (10.95)$$

$$e^{\pm:I_y:p_s} = p_s \mp \int_0^y \frac{\partial a_y(x, y', s)}{\partial s} dy', \quad (10.96)$$

$$e^{-\Delta\sigma:H'_2:y} = y + \frac{p_y}{D} \Delta\sigma, \quad (10.97)$$

$$e^{-\Delta\sigma:H'_2:z} = z - \frac{p_y^2}{2D^3} \left( \frac{1}{\beta_0} + \delta \right) \Delta\sigma, \quad (10.98)$$

$$e^{-\frac{\Delta\sigma}{2}:H_3:p_x} = p_x + \frac{\partial a_s}{\partial x} \frac{\Delta\sigma}{2}, \quad (10.99)$$

$$e^{-\frac{\Delta\sigma}{2}:H_3:p_y} = p_y + \frac{\partial a_s}{\partial y} \frac{\Delta\sigma}{2}, \quad (10.100)$$

$$e^{-\frac{\Delta\sigma}{2}:H_3:p_s} = p_s + \frac{\partial a_s}{\partial s} \frac{\Delta\sigma}{2}, \quad (10.101)$$

where:

$$H'_2 = \frac{p_y^2}{2D}. \quad (10.102)$$

Transformations of variables not shown explicitly are equal to the identity transformation. Strictly speaking, the transformations of  $p_s$  are not of much interest; but they are included here for completeness.

Each step of the integration (corresponding to a change in the independent variable from  $\sigma_0$  to  $\sigma_0 + \Delta\sigma$ ) involves applying a sequence of transformations to the dynamical variables. In some of the transformations, there is a change in the original independent variable  $s$ , which in the extended phase space has become a dynamical variable. This is the key aspect of extending the phase space by introducing a new independent variable: without doing this, the co-ordinate  $s$  would have a fixed value at each step of the integration. By extending the phase space, at each step of the integration the vector potential can be ‘sampled’ at different points along the reference trajectory. This is of course what is expected (and necessary)

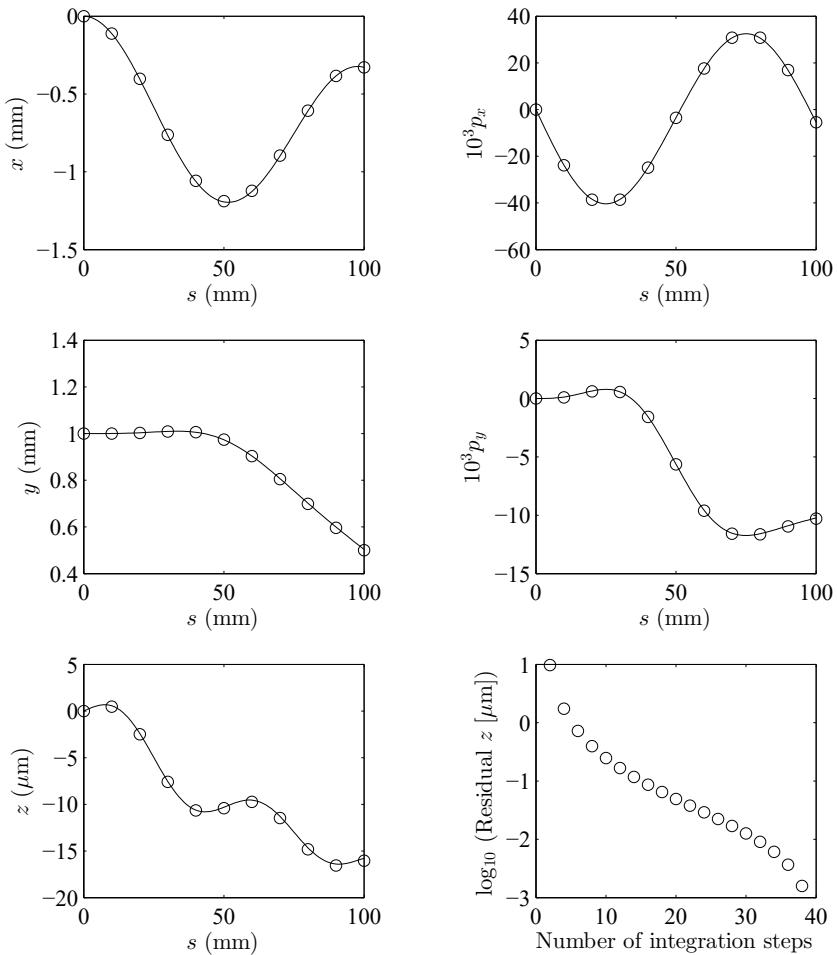


Fig. 10.3 Particle trajectory in an undulator magnet. The field in the undulator is described by (10.64)–(10.66) with period  $2\pi/k_s = 100$  mm, peak field  $B_0 = 0.1$  T, and horizontal ‘roll-off parameter’  $k_x = 400$  m $^{-1}$ . The reference particle has relativistic factor  $\gamma_0 = 25$ . The plots show the trajectory of a particle through one period of the undulator, starting at  $x = p_x = p_y = z = 0$ , vertical offset  $y = 1$  mm, and with energy deviation  $\delta = 0.1$ . The line shows the results of a numerical integration through the field using a high-order adaptive Runge–Kutta method. The circles show the results of using the Wu–Forest–Robin integrator (10.90) with step size  $\Delta\sigma = 10$  mm. The bottom right-hand plot shows the residual in the final value of  $z$  between the two methods, as a function of the number of steps used in the Wu–Forest–Robin integrator.

when the magnetic field, and hence the vector potential, depends on the position along the reference trajectory.

As an illustration of the use of the Wu–Forest–Robin integrator, we apply the integrator to track particles through an undulator magnet, with field described by (10.64)–(10.66). The results are shown in Fig. 10.3. In this case, the undulator has period  $2\pi/k_s = 100$  mm, peak field  $B_0 = 0.1$  T, and parameter (describing the horizontal roll-off of the field)  $k_x = 400$  m $^{-1}$ . We choose the reference energy such that the reference particle has relativistic factor  $\gamma_0 = 25$ . The plots in Fig. 10.3 show the variation in the dynamical variables for a particle moving through one period of the undulator, starting at  $s = 0$  with initial values  $x = p_x = p_y = z = 0$ , and vertical offset  $y = 1$  mm; the particle has fixed energy deviation  $\delta = 0.1$ . The solid lines show the results of integrating the equations of motion obtained from the Hamiltonian (10.81) using a high-order adaptive Runge–Kutta algorithm; the circles show the results of applying the Wu–Forest–Robin integrator (10.90) with step size  $\Delta\sigma = 10$  mm (i.e. with ten steps per period of the undulator). The bottom right-hand plot in Fig. 10.3 shows the residual in the final value of  $z$  calculated from the two different methods. With ten integration steps, the difference in the final value of  $z$  is less than a quarter of a micrometre, or about 1.5% of the overall change in  $z$ .

### **10.2.3 Symplectic Runge–Kutta integrators**

The explicit symplectic integrator described in Section 10.2.2 required the Hamiltonian to be expanded in the paraxial approximation. This is quite often a valid approximation for accelerator beam lines; however, there are cases where it may fail to give an accurate description of the dynamics (for example, where the transverse momenta  $p_y$  and  $p_x$  reach large values). In such cases, alternative integration techniques must be considered. Runge–Kutta methods provide a class of widely used schemes for integration of ordinary differential equations, and if certain conditions are satisfied then the integration will be symplectic. Unfortunately, the required conditions restrict the available methods to implicit Runge–Kutta schemes, which means that at each step of the integration a set of algebraic equations must be solved. This certainly increases the computational cost of performing the integration, but is not usually an insurmountable obstacle. In this section, Runge–Kutta methods are briefly reviewed, and then their particular application to symplectic integration is discussed.

Suppose that a variable  $x(s)$  satisfies the ordinary differential equation:

$$\frac{dx}{ds} = f(x, s), \quad (10.103)$$

for some function  $f$ , and with a given initial condition:

$$x(s_0) = x_0, \quad (10.104)$$

for particular values of  $s_0$  and  $x_0$ . Runge–Kutta methods provide a way to integrate the equation (10.103) by calculating the value of  $x$  at intervals  $\Delta s$ :

$$x(s_n + \Delta s) = x(s_n) + \sum_{i=1}^m b_i \kappa_i, \quad (10.105)$$

where the  $b_i$  constitute a set of  $m$  constant coefficients, and the quantities  $\kappa_i$  are given by:

$$\kappa_i = \Delta s f(x_n^{(i)}, s_n^{(i)}). \quad (10.106)$$

The ‘intermediate’ values  $x_n^{(i)}$  and  $s_n^{(i)}$  are given by:

$$x_n^{(i)} = x_n + \sum_{j=1}^m a_{ij} \kappa_j, \quad (10.107)$$

$$s_n^{(i)} = s_n + c_i \Delta s. \quad (10.108)$$

The quantities  $a_{ij}$  and  $c_i$  are constants. A particular Runge–Kutta scheme is given by specifying the values of the constants  $a_{ij}$ ,  $b_i$  and  $c_i$ . Conventionally, the values are displayed in a *Butcher tableau*:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1m} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_m & a_{m1} & a_{m2} & \cdots & a_{mm} \\ \hline b_1 & b_1 & b_2 & \cdots & b_m \end{array} \quad (10.109)$$

In general, it is required that:

$$c_i = \sum_{j=1}^m a_{ij}. \quad (10.110)$$

An explicit Runge–Kutta scheme has values  $a_{ij}$  such that:

$$a_{ij} = 0 \quad \text{if } j \geq i. \quad (10.111)$$

The Butcher tableau for such a scheme will appear diagonal. For example, a simple second-order explicit Runge–Kutta scheme has:

$$\begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array} \quad (10.112)$$

This is known as the *midpoint method*. The order of the scheme indicates the accuracy of the integration for a given integration step size: at each step, an  $n$ th-order scheme is accurate to order  $\Delta s^n$ . A widely used fourth-order explicit Runge–Kutta scheme (often known simply as RK4) is defined by:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ \hline 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \quad (10.113)$$

The advantage of an explicit scheme, as can be seen from (10.106) and (10.107), is that at each step each of the  $\kappa_i$  can be calculated directly from the  $\kappa_j$  with  $j < i$ . Thus:

$$\kappa_1 = \Delta s f(x_n, s_n), \quad (10.114)$$

$$\kappa_2 = \Delta s f(x_n + a_{21}\kappa_1, s_n + c_1\Delta s), \quad (10.115)$$

$$\kappa_3 = \Delta s f(x_n + a_{31}\kappa_1 + a_{32}\kappa_2, s_n + c_2\Delta s), \quad (10.116)$$

and so on. In the case of an implicit scheme, where in general all  $a_{ij}$  are non-zero, equations (10.106) and (10.107) lead to a set of algebraic equations that must be solved (perhaps iteratively) at each step of the integration. Unfortunately, explicit Runge–Kutta schemes are not symplectic; however, they may still be useful in beam dynamics if exact symplecticity (or at least symplecticity to numerical precision) is not an important requirement.

Runge–Kutta schemes can readily be applied to Hamiltonian systems: it is only necessary to generalise (10.103) to several variables. The function  $f$  is replaced by derivatives of the Hamiltonian. Writing the phase space vector  $\vec{x} = (x, p_x, y, p_y, z, \delta)$ , Hamilton's equations are:

$$\frac{d\vec{x}}{ds} = S \nabla_{\vec{x}} H(\vec{x}, s), \quad (10.117)$$

where  $S$  is the usual antisymmetric matrix (4.14), and  $\nabla_{\vec{x}}$  is the vector differential operator:

$$\nabla_{\vec{x}} = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial p_x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial p_y}, \frac{\partial}{\partial z}, \frac{\partial}{\partial \delta} \right). \quad (10.118)$$

The formula for an integration step in a Runge–Kutta scheme is then:

$$\vec{x}(s_n + \Delta s) = \vec{x}(s_n) + \sum_{i=1}^m b_i \vec{\kappa}_i, \quad (10.119)$$

where:

$$\vec{\kappa}_i = \Delta s S \nabla_{\vec{x}} H(\vec{x}_n^{(i)}, s_n^{(i)}), \quad (10.120)$$

and the intermediate values are:

$$\vec{x}_n^{(i)} = \vec{x}(s_n) + \sum_{j=1}^m a_{ij} \vec{\kappa}_j, \quad (10.121)$$

$$s_n^{(i)} = s_n + c_i \Delta s. \quad (10.122)$$

In principle, any Runge–Kutta method (explicit or implicit) can be used to integrate the Hamiltonian equations of motion; however, the integration will only be symplectic for specific Butcher tableaux. It can be shown (see, for example, [Hairer *et al.* (2006c)]) that if the equations of motion derive from a Hamiltonian, and the Butcher tableau is such that:

$$b_i a_{ij} + b_j a_{ji} = b_i b_j, \quad (10.123)$$

for all  $i$  and  $j$ , then the integration will be symplectic.

As an example, consider the Hamiltonian:

$$H = \frac{1}{2} p_x^2 + \frac{1}{6} k_2 x^3. \quad (10.124)$$

This is, of course, the Hamiltonian for an ultra-relativistic particle in a sextupole magnet in the paraxial approximation. A second-order Runge–Kutta integrator is provided by the Butcher tableau:

$$\begin{array}{c|cc} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array} \quad (10.125)$$

This is called the *implicit midpoint rule*. Since the parameters  $a_{ij}$  and  $b_i$  satisfy (10.123), the integration is symplectic. Applying (10.119) and (10.120) gives:

$$\begin{aligned} x(s_n + \Delta s) &= x(s_n) + \Delta s \left. \frac{\partial H}{\partial p_x} \right|_{x=x_n^{(1)}, p=p_{xn}^{(1)}} \\ &= x(s_n) + \Delta s p_{xn}^{(1)}, \end{aligned} \quad (10.126)$$

and:

$$\begin{aligned} p_x(s_n + \Delta s) &= p_x(s_n) - \Delta s \left. \frac{\partial H}{\partial x} \right|_{x=x_n^{(1)}, p=p_{xn}^{(1)}} \\ &= p_x(s_n) - \frac{1}{2} k_2 \Delta s (x_n^{(1)})^2. \end{aligned} \quad (10.127)$$

Applying (10.121), the intermediate values  $x_n^{(1)}$  and  $p_{xn}^{(1)}$  are found from:

$$\begin{aligned} x_n^{(1)} &= x(s_n) + \frac{1}{2} \Delta s \left. \frac{\partial H}{\partial p_x} \right|_{x=x_n^{(1)}, p=p_{xn}^{(1)}} \\ &= x(s_n) + \frac{1}{2} \Delta s p_{xn}^{(1)}, \end{aligned} \quad (10.128)$$

and:

$$\begin{aligned} p_{xn}^{(1)} &= p_x(s_n) - \frac{1}{2} \Delta s \left. \frac{\partial H}{\partial x} \right|_{x=x_n^{(1)}, p=p_{xn}^{(1)}} \\ &= p_x(s_n) - \frac{1}{4} k_2 \Delta s (x_n^{(1)})^2. \end{aligned} \quad (10.129)$$

Note that (10.128) and (10.129) are algebraic equations for  $x_n^{(1)}$  and  $p_{xn}^{(1)}$ : this integrator is an implicit integrator. However, in this case we can write the exact solutions:

$$x_n^{(1)} = \tau, \quad (10.130)$$

$$p_{xn}^{(1)} = \frac{2\tau}{\Delta s} - \frac{x(s_n)}{2\Delta s}, \quad (10.131)$$

where:

$$\tau = \frac{4}{k_2 \Delta s^2} \left( \sqrt{1 + \frac{1}{2} k_2 \Delta s^2 \left( x(s_n) + \frac{1}{2} \Delta s p_x(s_n) \right)} - 1 \right). \quad (10.132)$$

Finally, substituting these expressions for  $x_n^{(1)}$  and  $p_{xn}^{(1)}$  into (10.126) and (10.127) gives:

$$x(s_n + \Delta s) = x(s_n) + \Delta s p_x(s_n) - \frac{1}{4} k_2 \Delta s^2 \tau^2, \quad (10.133)$$

$$p_x(s_n + \Delta s) = p_x(s_n) - \frac{1}{2} k_2 \Delta s \tau^2. \quad (10.134)$$

Although it is not immediately obvious from (10.133) and (10.134), the Jacobian  $J$  of the transformation satisfies:

$$J^T S_2 J = S_2, \quad (10.135)$$

where  $S$  is the  $2 \times 2$  antisymmetric matrix (2.78). Hence, the transformation for each step of the integration is symplectic, and the overall integration will be symplectic.

The implicit midpoint rule (10.125) is a member of a class of implicit Runge–Kutta integrators known as *Gauss methods* or *Gauss–Legendre*

methods that provide symplectic integration schemes to arbitrarily high order. For example, the fourth-order Gauss method has Butcher tableau [Hairer *et al.* (1993, 2006d); Butcher (2008)]:

$$\begin{array}{c|cc} \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \hline \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \quad (10.136)$$

As can easily be imagined, it becomes rapidly more difficult to implement symplectic Runge–Kutta integrators as the order of the integrator increases. The main difficulty is that symplectic Runge–Kutta methods are implicit methods, and the algebraic equations that must be solved at each step add significantly to the computational cost. It is unlikely that a symplectic Runge–Kutta integrator of higher than fourth order would be used for particle tracking in an accelerator; but for reference, the Butcher tableau for the sixth-order Gauss method is:

$$\begin{array}{c|cccc} \frac{1}{2} - \frac{\sqrt{15}}{10} & \frac{5}{36} & \frac{2}{9} - \frac{\sqrt{15}}{15} & \frac{5}{36} - \frac{\sqrt{15}}{30} \\ \hline \frac{1}{2} & \frac{5}{36} + \frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36} - \frac{\sqrt{15}}{24} \\ \hline \frac{1}{2} + \frac{\sqrt{15}}{10} & \frac{5}{36} + \frac{\sqrt{15}}{30} & \frac{2}{9} + \frac{\sqrt{15}}{15} & \frac{5}{36} \\ \hline & \frac{5}{18} & \frac{4}{9} & \frac{5}{18} \end{array} \quad (10.137)$$

**This page intentionally left blank**

## Chapter 11

# Methods for Analysis of Single-Particle Dynamics

In Chapter 10 Lie transformations were applied to the construction of symplectic integrators. The generator for a Lie transformation corresponding to particle transport through a given accelerator beam line component is obtained from the Hamiltonian describing the dynamics in that component. In general, it is not possible (except in certain special cases) to write down an exact solution to the equations of motion in closed form, but a solution in the form of a power series in the initial values of the dynamical variables can be constructed by applying a Lie transformation (with generator obtained from the Hamiltonian) to the dynamical variables. Usually, the power series representing the solution will contain an infinite number of terms. However, by making approximations to the generator of a Lie transformation for a beam line component, it is possible to obtain a symplectic transfer map that can be expressed as a power series with a finite number of terms.

The construction of symplectic transfer maps (using Lie transformations or other methods) is an important part of the study of nonlinear dynamics in particle accelerators. Accurate transfer maps are necessary to be able to model particle motion in an accelerator beam line. However, to understand the nonlinear dynamics in an accelerator, there is a need for techniques that yield information about the behaviour of particles in complex beam lines consisting of perhaps dozens or hundreds of components. Transfer maps for individual components are important building blocks; but to design and operate accelerators that must achieve demanding performance specifications, some insight is needed into the dynamics that result from putting the building blocks together.

In this chapter, we shall give some examples of techniques for the analysis of particle dynamics in beam lines in which nonlinear components

have significant effects. Often, one of the goals of beam line design is to control the nonlinear effects so as to achieve particular beam properties. In many cases, components with strong nonlinear effects are introduced in a beam line for particular purposes: for example, sextupoles are widely used to correct chromatic effects, and octupoles are sometimes used to provide Landau damping to suppress beam instabilities. However, the desired effects are often accompanied by undesirable side-effects, for example a reduction in the dynamic aperture in a storage ring. The goal in the analysis of nonlinear dynamics in an accelerator is often simply to find ways to minimise the undesirable effects, while maintaining or maximising the desired effects.

### 11.1 A Lie Transformation Example: the $-I$ Transformer

In Section 8.3, it was shown how sextupoles could be used to control the chromaticity in a storage ring. A lattice constructed just from dipoles and quadrupoles will have large negative chromaticity; to achieve a good energy acceptance in a beam line, it is usually necessary to reduce the chromaticity. In a storage ring, for example, bringing the first-order chromaticity close to zero means that the energy deviation of a particle can be quite large before the tunes reach integer or half-integer values. Unfortunately, sextupoles have geometric as well as chromatic effects: that is, they impact the dynamics of particles that have zero energy deviation in addition to their effects on particles with non-zero energy deviation. However, by arranging sextupoles in pairs with an appropriate linear transformation between them, the geometric effects can be made (almost) to cancel. This can be readily understood by using Lie transformations to represent the dynamics, as we now show.

The transfer map for a sextupole of length  $L_s$  can be represented by a Lie transformation:

$$\mathcal{S} = e^{-L_s:H_s:}, \quad (11.1)$$

where  $H_s$  is the Hamiltonian for a sextupole:

$$H_s = H_d + H_k. \quad (11.2)$$

We have split the Hamiltonian into a term representing a drift:

$$H_d = \frac{\delta}{\beta_0} - \sqrt{\left(\frac{1}{\beta_0} + \delta\right)^2 - p_x^2 - p_y^2 - \frac{1}{\beta_0^2 \gamma_0^2}}, \quad (11.3)$$

and a term representing the field in the sextupole:

$$H_k = \frac{1}{6} k_2 (x^3 - 3xy^2). \quad (11.4)$$

The sextupole gradient  $k_2$  is defined by:

$$k_2 = \frac{q}{P_0} \frac{\partial^2 B_y}{\partial x^2}. \quad (11.5)$$

It was shown in Section 10.2.1 that a second-order symplectic approximation to the exact transfer map can be written:

$$\mathcal{S}_2 = e^{-\frac{1}{2}L_s:H_d:} e^{-L_s:H_k:} e^{-\frac{1}{2}L_s:H_d:}. \quad (11.6)$$

In this approximation, the geometric aberrations introduced by the sextupole arise from the Lie transformation with generator  $H_k$ .

Now consider two identical sextupoles (that is, with the same length and gradient) separated by a section of accelerator beam line that can be represented by some Lie transformation  $\mathcal{R}$ . The total transfer map from the entrance of the first sextupole to the exit of the second sextupole is:

$$\mathcal{M} = \mathcal{S}\mathcal{R}\mathcal{S} \approx \mathcal{M}_2 = \mathcal{S}_2\mathcal{R}\mathcal{S}_2. \quad (11.7)$$

Writing  $\mathcal{S}_2$  explicitly, the map  $\mathcal{M}_2$  (that is a second-order approximation to the exact transfer map  $\mathcal{M}$ ) is:

$$\mathcal{M}_2 = e^{-\frac{1}{2}L_s:H_d:} e^{-L_s:H_k:} \bar{\mathcal{R}} e^{-L_s:H_k:} e^{-\frac{1}{2}L_s:H_d:}, \quad (11.8)$$

where:

$$\bar{\mathcal{R}} = e^{-\frac{1}{2}L_s:H_d:} \mathcal{R} e^{-\frac{1}{2}L_s:H_d:}. \quad (11.9)$$

Using the fact that:

$$\bar{\mathcal{R}}\bar{\mathcal{R}}^{-1} = \mathcal{I}, \quad (11.10)$$

where  $\bar{\mathcal{R}}^{-1}$  is the inverse of  $\bar{\mathcal{R}}$  and  $\mathcal{I}$  is the identity transformation, we can write  $\mathcal{M}_2$  as:

$$\mathcal{M}_2 = e^{-\frac{1}{2}L_s:H_d:} \bar{\mathcal{R}} (\bar{\mathcal{R}}^{-1} e^{-L_s:H_k:} \bar{\mathcal{R}}) e^{-L_s:H_k:} e^{-\frac{1}{2}L_s:H_d:}. \quad (11.11)$$

We have inserted the identity transformation  $\mathcal{I} = \bar{\mathcal{R}}\bar{\mathcal{R}}^{-1}$  into the expression for  $\mathcal{M}_2$  so that we can apply the similarity transformation (9.19):

$$\bar{\mathcal{R}}^{-1} e^{-L_s:H_k:} \bar{\mathcal{R}} = \exp(-L_s: \bar{\mathcal{R}}^{-1} H_k :). \quad (11.12)$$

If the transformation  $\bar{\mathcal{R}}^{-1}$  is such that:

$$\bar{\mathcal{R}}^{-1} H_k = -H_k, \quad (11.13)$$

then:

$$\mathcal{M}_2 = e^{-\frac{1}{2}L_s:H_d:} \bar{\mathcal{R}} e^{L_s:H_k:} e^{-L_s:H_k:} e^{-\frac{1}{2}L_s:H_d:}. \quad (11.14)$$

Using:

$$e^{L_s:H_k:} e^{-L_s:H_k:} = \mathcal{I}, \quad (11.15)$$

$\mathcal{M}_2$  becomes:

$$\mathcal{M}_2 = e^{-\frac{1}{2}L_s:H_d:} \bar{\mathcal{R}} e^{-\frac{1}{2}L_s:H_d:} = e^{-L_s:H_d:} \bar{\mathcal{R}} e^{-L_s:H_d:}. \quad (11.16)$$

The nonlinear effects of the sextupoles, arising from the Lie transformation  $e^{-L_s:H_k:}$ , have disappeared from  $\mathcal{M}_2$ . The transfer maps for the sextupoles (in the second-order approximation) in (11.7) can simply be replaced by transfer maps for drift spaces. The cancellation of the nonlinear kicks from the sextupoles depends on the beam line between the sextupoles having the property (11.13). Given the form of  $H_k$  (11.4), this can be achieved if:

$$\bar{\mathcal{R}} x = -x, \quad (11.17)$$

$$\bar{\mathcal{R}} y = \pm y. \quad (11.18)$$

This is a linear transformation, so we expect that the Lie transformation  $\bar{\mathcal{R}}$  can be expressed in terms of a generator that is a second-order polynomial in the dynamical variables. We therefore consider a Lie transformation of the form  $e^{-\mu_x:\bar{J}_x:}$ , where  $\mu_x$  is a constant, and:

$$\bar{J}_x = \frac{1}{2} (c_x x^2 + 2a_x x p_x + b_x p_x^2). \quad (11.19)$$

Since  $\mu_x$  applies an overall scaling to the generator, we can impose the condition on the coefficients  $a_x$ ,  $b_x$  and  $c_x$ :

$$b_x c_x - a_x^2 = 1. \quad (11.20)$$

Using similar working to that leading to (9.16) and (9.17), we find that:

$$e^{-\mu_x:\bar{J}_x:} \begin{pmatrix} x \\ p_x \end{pmatrix} = \begin{pmatrix} \cos(\mu_x) + a_x \sin(\mu_x) & b_x \sin(\mu_x) \\ -c_x \sin(\mu_x) & \cos(\mu_x) - a_x \sin(\mu_x) \end{pmatrix} \begin{pmatrix} x \\ p_x \end{pmatrix}. \quad (11.21)$$

A transformation satisfying (11.17) is obtained by setting  $\mu_x = (2n_x + 1)\pi$ , where  $n_x$  is any integer. The coefficients  $a_x$ ,  $b_x$  and  $c_x$  can take any values satisfying (11.20). Applying the same arguments to the vertical part of the transfer map, it is found that a possible solution for  $\bar{\mathcal{R}}$  is:

$$\bar{\mathcal{R}} = \exp(-\mu_x: \bar{J}_x : - \mu_y: \bar{J}_y :), \quad (11.22)$$

where, for any integers  $n_x$  and  $n_y$ :

$$\mu_x = (2n_x + 1)\pi, \quad (11.23)$$

$$\mu_y = n_y\pi. \quad (11.24)$$

The transformation (11.22) can be represented by a transfer matrix, with transverse part:

$$\bar{\mathcal{R}} \begin{pmatrix} x \\ p_x \\ y \\ p_y \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & \pm 1 \end{pmatrix} \begin{pmatrix} x \\ p_x \\ y \\ p_y \end{pmatrix}. \quad (11.25)$$

The horizontal part of the transfer matrix is  $-I_2$ , where  $I_2$  is the  $2 \times 2$  identity matrix. The vertical part of the transfer matrix in (11.25) is  $\pm I_2$ .

For particles with zero energy deviation, it is possible to construct a beam line with transfer matrix (11.25) using combinations of quadrupoles: for example, two successive FODO cells with phase advance  $90^\circ$  could be used. A beam line with phase advance  $(2n + 1)\pi$  for integer  $n$  is known as a  $-I$  transformer, since the transfer matrix is simply  $-1$  times the identity. Generally, because of chromatic effects, a  $-I$  transformer only has the required phase advance for particles with zero energy deviation. In a beam line consisting of two sextupoles separated by a  $-I$  transformer, the effects of the sextupoles on particles with zero energy deviation will cancel to second order. The significance of higher-order effects will depend on the accuracy of the second-order approximation (11.7): in general, this approximation is better for short magnets than for long magnets. The energy deviation at which the geometric effects of the sextupoles start to become significant depends on the chromaticity of the beam line between the sextupoles.

When sextupoles are introduced in a storage ring lattice to control the chromaticity, arranging them in pairs separated by a  $-I$  transformer is one strategy to improve the dynamic aperture of the storage ring (dynamic aperture is discussed in Section 11.3). However, it is not always possible to implement this technique because of other constraints; for example, on the size of the beam line and the number of components within it. Where a  $-I$  transformer cannot be used, other methods are needed to analyse and optimise the dynamics.

## 11.2 Canonical Perturbation Theory

The  $-I$  transformer discussed in Section 11.1 provides an example of the use of Lie transformations for analysis of the beam dynamics in a beam

line consisting of a number of different components. The main result — that the geometric effects of a pair of identical sextupoles approximately cancel when separated by a section of beam line over which the phase advance is  $180^\circ$  — can of course be obtained by other methods. But Lie transformations offer an elegant approach that has some advantages when it is necessary to carry out detailed investigations of complicated nonlinear systems. We shall return to analysis techniques based on Lie transformations in Section 11.4, but in this section we discuss a technique based on canonical perturbation theory [Goldstein *et al.* (2001d); Ruth (1986a,b)] that is useful for calculating the impact not just of nonlinear effects, but also the effects of linear focusing errors or coupling errors in an accelerator beam line.

The principle behind the perturbation theory technique is to construct a generating function for a canonical transformation to new dynamical variables, in which the Hamiltonian takes a simpler form to some order in a perturbation parameter. Dropping terms above a given order in the perturbation parameter allows an approximate solution to be found for the equations of motion in the new variables. A solution in the original variables is then obtained by using the generating function that relates the new variables to the original variables. As well as providing approximate solutions for systems that are not integrable because of perturbative terms in the Hamiltonian, this technique provides a way of characterising significant features of the dynamics, such as the variation of betatron frequency with the amplitude of betatron oscillations.

For clarity in developing the technique, it is helpful to work in one degree of freedom, and to use action-angle variables  $(\phi_x, J_x)$ . Consider the Hamiltonian for a periodic beam line:

$$H = H_0(J_x; s) + \epsilon V(\phi_x, J_x; s), \quad (11.26)$$

where  $s$  is the independent variable.  $\epsilon$  is a (small) parameter that is introduced to help keep track of the order to which the perturbation appears.  $V$  is a function of the dynamical variables  $\phi_x$  and  $J_x$ , and the independent variable  $s$ . If the beam line has periodicity  $C_0$  (for example,  $C_0$  may be the circumference of a storage ring) then:

$$V(\phi_x, J_x; s + C_0) = V(\phi_x, J_x; s). \quad (11.27)$$

$H_0(J_x; s)$  is a term describing the dynamics in the absence of the perturbation  $V$ , i.e. in the limit  $\epsilon \rightarrow 0$ . It is assumed that this term is independent of the angle variable  $\phi_x$ ; this means that the Hamiltonian  $H_0$  is integrable.

The equations of motion in the limit  $\epsilon \rightarrow 0$  are:

$$\frac{d\phi_x}{ds} = \frac{\partial H_0}{\partial J_x}, \quad (11.28)$$

$$\frac{dJ_x}{ds} = -\frac{\partial H_0}{\partial \phi_x} = 0. \quad (11.29)$$

Hence,  $J_x$  is constant; and the equations of motion have solution:

$$\phi_x(s_1) = \phi_x(s_0) + \int_{s_0}^{s_1} \frac{\partial H_0}{\partial J_x} ds, \quad (11.30)$$

$$J_x = J_x(s_0), \quad (11.31)$$

where  $\phi_x(s_0)$  and  $J_x(s_0)$  are constants of integration, corresponding to the initial values of  $\phi_x$  and  $J_x$  at a specified point  $s = s_0$  along the reference trajectory. In the case of a beam line consisting only of linear components without coupling (for example, drifts, dipoles and quadrupoles, with zero energy deviation),  $H_0$  can be written as (4.58):

$$H_0 = \frac{J_x}{\beta_x(s)}, \quad (11.32)$$

where  $\beta_x(s)$  is the Courant–Snyder beta function. However, more general forms for  $H_0$  (including higher-order terms in  $J_x$ ) are allowed in (11.26). For example, for a beam line containing octupole magnets with strength:

$$k_3(s) = \frac{q}{P_0} \frac{\partial^3 B_y(s)}{\partial x^3}, \quad (11.33)$$

the Hamiltonian in one (horizontal) degree of freedom would be written, using (4.36):

$$\begin{aligned} H &= \frac{J_x}{\beta_x(s)} + \frac{1}{24} k_3(s) x^4 \\ &= \frac{J_x}{\beta_x(s)} + \frac{1}{24} k_3(s) (2\beta_x(s) J_x)^2 \cos^4(\phi_x) \\ &= \frac{J_x}{\beta_x(s)} + \frac{1}{48} k_3(s) \beta_x(s)^2 J_x^2 (3 + 4 \cos(2\phi_x) + \cos(4\phi_x)). \end{aligned} \quad (11.34)$$

Therefore, in this case we would write:

$$H_0 = \frac{J_x}{\beta_x(s)} + \frac{1}{16} k_3(s) \beta_x(s)^2 J_x^2, \quad (11.35)$$

and:

$$\epsilon V(\phi_x, J_x; s) = \frac{1}{48} k_3(s) \beta_x(s)^2 J_x^2 (4 \cos(2\phi_x) + \cos(4\phi_x)). \quad (11.36)$$

The goal in perturbation theory is to find a canonical transformation to new variables in which the Hamiltonian has no dependence on the angle variable, up to some order. In the following, we show how to construct such a transformation; but the transformation has no effect on  $H_0$ , so the higher-order terms in  $J_x$  remain. This is not a problem, however, since a Hamiltonian that can be expressed as a power series in  $J_x$  (with no dependence on  $\phi_x$ ) is integrable. The higher-order terms in  $J_x$  describe a tune shift with amplitude. In the case of an octupole perturbation, for example, the phase advance through an octupole of length  $L$  located at  $s = s_0$ , averaged over all particles in the beam is:

$$\langle \Delta\phi_x \rangle = \int_{s_0}^{s_0+L} \left\langle \frac{\partial H}{\partial J_x} \right\rangle ds = \frac{L}{\beta_x(s_0)} + \frac{1}{8} k_3 L \beta_x(s_0)^2 \langle J_x \rangle. \quad (11.37)$$

An octupole gives a *tune spread* that is proportional to the integrated strength of the octupole multiplied by the square of the beta function, and is proportional also to the beam emittance (which is equal to average betatron action  $\langle J_x \rangle$ ).

To understand the effects of perturbations in more detail, we shall follow the approach of perturbation theory, and look for a canonical transformation that removes the terms dependent on  $\phi_x$  from the Hamiltonian. Consider a mixed-variable generating function of the second kind:

$$F_2(\phi_x, J_{1x}; s) = \phi_x J_{1x} + \epsilon G(\phi_x, J_{1x}; s), \quad (11.38)$$

where  $J_{1x}$  is a new action variable, and  $G$  is a function to be determined.  $J_{1x}$  is the momentum conjugate to a new co-ordinate (angle variable)  $\phi_{1x}$ . The new variables  $(\phi_{1x}, J_{1x})$  obey equations of motion derived from a new Hamiltonian  $H_1$ . The old and new variables and Hamiltonians are related by:

$$J_x = \frac{\partial F_2}{\partial \phi_x} = J_{1x} + \epsilon \frac{\partial G}{\partial \phi_x}, \quad (11.39)$$

$$\phi_{1x} = \frac{\partial F_2}{\partial J_{1x}} = \phi_x + \epsilon \frac{\partial G}{\partial J_{1x}}, \quad (11.40)$$

$$H_1 = H + \frac{\partial F_2}{\partial s} = H + \epsilon \frac{\partial G}{\partial s}. \quad (11.41)$$

Note that in the limit  $\epsilon \rightarrow 0$ , the transformation from the old to the new variables approaches the identity.

In terms of the variables  $\phi_x$  and  $J_{1x}$ , the new Hamiltonian can be written:

$$H_1 = H_0 \left( J_{1x} + \epsilon \frac{\partial G}{\partial \phi_x}; s \right) + \epsilon V \left( \phi_x, J_{1x} + \epsilon \frac{\partial G}{\partial \phi_x}; s \right) + \epsilon \frac{\partial G}{\partial s}. \quad (11.42)$$

In this form, the Hamiltonian is expressed in terms of the old angle  $\phi_x$  and the new action  $J_{1x}$ . Eventually, to solve the equations of motion it will be necessary to substitute for  $\phi_x$  so that the Hamiltonian is expressed purely in terms of the new variables. However, it is convenient for the moment to leave the Hamiltonian in the form (11.42). Expanding the terms in the Hamiltonian  $H_1$  as a Taylor series gives:

$$H_1 = H_0(J_{1x}; s) + \epsilon H^{(1)}(\phi_x, J_{1x}; s) + \epsilon^2 H^{(2)}(\phi_x, J_{1x}; s) + O(\epsilon^3), \quad (11.43)$$

where:

$$H^{(1)}(\phi_x, J_{1x}; s) = \frac{\partial H_0(J_{1x}; s)}{\partial J_{1x}} \frac{\partial G}{\partial \phi_x} + \frac{\partial G}{\partial s} + V(\phi_x, J_{1x}; s), \quad (11.44)$$

and:

$$H^{(2)}(\phi_x, J_{1x}; s) = \frac{1}{2} \frac{\partial^2 H_0(J_{1x}; s)}{\partial J_{1x}^2} \left( \frac{\partial G}{\partial \phi_x} \right)^2 + \frac{\partial V(\phi_x, J_{1x}; s)}{\partial J_{1x}} \frac{\partial G}{\partial \phi_x}. \quad (11.45)$$

If the generating function is chosen such that:

$$H^{(1)}(\phi_x, J_{1x}; s) = \frac{\partial H_0(J_{1x}; s)}{\partial J_{1x}} \frac{\partial G}{\partial \phi_x} + \frac{\partial G}{\partial s} + V(\phi_x, J_{1x}; s) = 0, \quad (11.46)$$

then the terms in  $H_1$  that are first order in  $\epsilon$  vanish, and only terms of second order (and higher) in  $\epsilon$  remain:

$$H_1 = H_0(J_{1x}; s) + \epsilon^2 H^{(2)}(\phi_x, J_{1x}; s) + O(\epsilon^3). \quad (11.47)$$

Dropping the terms that are second order and higher in  $\epsilon$  gives the approximate Hamiltonian:

$$H_1(\phi_{1x}, J_{1x}; s) \approx H_0(J_{1x}; s). \quad (11.48)$$

Since  $H_0$  is integrable,  $H_1$  (to first order in  $\epsilon$ ) is also integrable. It is then possible to write down approximate solutions for the equations of motion for  $\phi_{1x}$  and  $J_{1x}$ ; and, assuming that (11.46) can be solved for the function  $G$ , approximate solutions for the original variables  $\phi_x$  and  $J_x$  can be found from (11.39) and (11.40). We have made an approximation to the Hamiltonian rather than to the solutions to the equations of motion: for this reason, and because the new variables are related to the original variables by a canonical transformation, the solutions obtained for the evolution of the variables  $\phi_x$  and  $J_x$  will be symplectic.

To eliminate (to some order) the perturbation  $V$  by transforming to new variables, it is necessary to solve (11.46) for the function  $G$ . The solution can be obtained as follows. In an accelerator beam line,  $V$  can be written

in terms of the (Cartesian) co-ordinate  $x$ , and, from (4.36),  $x$  is a periodic function of  $\phi_x$ :

$$x = \sqrt{2\beta_x(s)J_x} \cos(\phi_x). \quad (11.49)$$

It follows that  $V$  must also be a periodic function of  $\phi_x$ . Thus, it is possible to write  $V$  as a sum over modes:

$$V(\phi_x, J_{1x}; s) = \sum_{m=-\infty}^{\infty} \tilde{V}_m(J_{1x}; s) e^{im\phi_x}. \quad (11.50)$$

We assume that it is possible also to write  $G$  as a sum over modes:

$$G(\phi_x, J_{1x}; s) = \sum_{m=-\infty}^{\infty} \tilde{G}_m(J_{1x}; s) e^{im\phi_x}. \quad (11.51)$$

Then, (11.46) becomes:

$$\frac{\partial \tilde{G}_m}{\partial s} + im \frac{\partial H_0}{\partial J_{1x}} \tilde{G}_m = -\tilde{V}_m. \quad (11.52)$$

Note that  $\tilde{G}_m$ ,  $\tilde{V}_m$  and  $H_0$  are all functions of  $J_{1x}$  and  $s$ . Let us try a solution to (11.52) of the form:

$$\tilde{G}_m(J_{1x}; s) = A(J_{1x}) \int_s^{s+C_0} e^{im(\psi(J_{1x}; s') - \psi(J_{1x}; s))} \tilde{V}_m(J_{1x}; s') ds'. \quad (11.53)$$

$A$  is a function of  $J_{1x}$  (but is independent of  $s$ ) and  $\psi$  is a function of  $J_{1x}$  and  $s$ . Differentiating with respect to  $s$  gives:

$$\begin{aligned} \frac{\partial \tilde{G}_m(J_{1x}; s)}{\partial s} + im \frac{\partial \psi(J_{1x}; s)}{\partial s} \tilde{G}_m(J_{1x}; s) = \\ \left( e^{-2\pi im\nu_x(J_{1x})} - 1 \right) A(J_{1x}) \tilde{V}_m(J_{1x}; s), \end{aligned} \quad (11.54)$$

where:

$$2\pi\nu_x(J_{1x}) = \psi(J_{1x}; s + C_0) - \psi(J_{1x}; s), \quad (11.55)$$

and we have used the fact that  $\tilde{V}_m$  is periodic in  $s$ , so that  $\tilde{V}_m(J_{1x}, s + C_0) = \tilde{V}_m(J_{1x}, s)$ . Comparing (11.54) with (11.52), we see that (11.53) provides a solution to (11.52) if:

$$\frac{\partial \psi(J_{1x}; s)}{\partial s} = \frac{\partial H_0}{\partial J_{1x}}, \quad (11.56)$$

and:

$$A(J_{1x}) = \frac{ie^{-\pi im\nu_x}}{2\sin(\pi m\nu_x)}. \quad (11.57)$$

Hence, the solution to (11.52) is:

$$\tilde{G}_m(J_{1x}; s) = \frac{i}{2 \sin(\pi m \nu_x)} \int_s^{s+C_0} e^{im(\psi(J_{1x}; s') - \psi(J_{1x}; s) - \pi \nu_x)} \tilde{V}_m(J_{1x}; s') ds'. \quad (11.58)$$

Equation (11.58) is the main result from perturbation theory: it shows how to construct a generating function for a transformation such that a given perturbation is eliminated (to first order) from the new Hamiltonian.

An important feature of this equation is the dependence of the generating function on the quantity  $\nu_x$ . From (11.55), (11.56) and (11.28), it appears that we can identify  $\psi(J_{1x}; s') - \psi(J_{1x}; s)$  with the betatron phase advance from  $s$  to  $s'$ , and  $\nu_x(J_{1x})$  with the betatron tune. To consider this point in more explicit detail, observe that the equation of motion for the action variable  $\phi_{1x}$  is, to first order in the perturbation parameter  $\epsilon$ :

$$\frac{d\phi_{1x}}{ds} = \frac{\partial H_0}{\partial J_{1x}} = \frac{\partial \psi(J_{1x}; s)}{\partial s}. \quad (11.59)$$

Since  $J_{1x}$  is constant (to first order in the perturbation) it follows that:

$$\nu_x = \frac{1}{2\pi} (\psi(J_{1x}; s + C_0) - \psi(J_{1x}; s)) = \frac{1}{2\pi} (\phi_{1x}(s + C_0) - \phi_{1x}(s)). \quad (11.60)$$

Therefore  $\psi(J_{1x}; s') - \psi(J_{1x}; s)$  is the betatron phase advance between two points  $s$  and  $s'$  along the reference trajectory, and  $\nu_x$  is the number of betatron oscillations performed by a particle with a given betatron action over one period of the beam line. Strictly speaking, the betatron phase advance and betatron action are here defined in terms of the new canonical variables  $\phi_{1x}$  and  $J_{1x}$ . In general, the phase advance and betatron tune are functions of the betatron action.

We observe from (11.58) that, if  $m \nu_x$  is close to an integer for some particular value of  $J_{1x}$  and  $m$ , then unless the integral vanishes, the value of  $\tilde{G}_m(J_{1x}; s)$  will be large. This implies that a large transformation will be needed to eliminate the perturbation from the Hamiltonian to first order. In that case, the perturbation will be expected to have a significant impact on the dynamics. When  $m \nu_x$  is close to an integer, a particle is said to be close to an  *$m$ th-order resonance*.

The results obtained from perturbation theory in one degree of freedom can be generalised to more than one degree of freedom. We denote quantities with  $n$  components, with each component associated with a degree of freedom, using bold face symbols. Thus, in three degrees of freedom, the action vector  $\mathbf{J}$  is:

$$\mathbf{J} = (J_x, J_y, J_z). \quad (11.61)$$

The Hamiltonian is written:

$$H = H_0(\mathbf{J}; s) + \epsilon V(\phi, \mathbf{J}; s), \quad (11.62)$$

where  $H_0$  contains all the terms in  $H$  that are independent of the angle variables. The mixed-variable generating function (which removes the perturbation to first order in  $\epsilon$ ) is:

$$F_2(\phi, \mathbf{J}_1; s) = \phi \cdot \mathbf{J}_1 + \epsilon G(\phi, \mathbf{J}_1; s). \quad (11.63)$$

The perturbing potential  $V$  is expanded in a multivariate Fourier series:

$$V(\phi, \mathbf{J}_1; s) = \sum_{m_1=-\infty}^{\infty} \cdots \sum_{m_n=-\infty}^{\infty} \tilde{V}_{\mathbf{m}}(\mathbf{J}_1; s) e^{i\mathbf{m} \cdot \phi}, \quad (11.64)$$

where  $m_1, \dots, m_n$  are the components of  $\mathbf{m}$ . Similarly the function  $G$ , which appears in the generating function (11.38), is expanded as:

$$G(\phi, \mathbf{J}_1; s) = \sum_{m_1=-\infty}^{\infty} \cdots \sum_{m_n=-\infty}^{\infty} \tilde{G}_{\mathbf{m}}(\mathbf{J}_1; s) e^{i\mathbf{m} \cdot \phi}. \quad (11.65)$$

Following the same reasoning as in the case of one degree of freedom, it is found that:

$$\tilde{G}_{\mathbf{m}}(\mathbf{J}_1; s) = \frac{i}{2 \sin(\pi \mathbf{m} \cdot \boldsymbol{\nu})} \int_s^{s+C_0} e^{i\mathbf{m} \cdot (\psi(\mathbf{J}_1; s') - \psi(\mathbf{J}_1; s) - \pi \boldsymbol{\nu})} \tilde{V}_{\mathbf{m}}(\mathbf{J}_1; s') ds'. \quad (11.66)$$

The vector  $\psi(\mathbf{J}_1; s)$  (with components corresponding to the phase advances) is defined by:

$$\frac{\partial \psi(\mathbf{J}_1; s)}{\partial s} = \frac{\partial H_0}{\partial \mathbf{J}_1}. \quad (11.67)$$

The vector  $\boldsymbol{\nu}$  (with components corresponding to the tunes) is defined by:

$$\boldsymbol{\nu} = \int_0^{C_0} \frac{\partial \psi(\mathbf{J}_1; s)}{\partial s} ds = \psi(\mathbf{J}_1, C_0) - \psi(\mathbf{J}_1, 0). \quad (11.68)$$

The Hamiltonian with the perturbation removed to first order is:

$$H_1 = H_0(\mathbf{J}_1; s) + \epsilon^2 \left( \frac{1}{2} \left( \frac{\partial G}{\partial \phi} \right)^T \frac{\partial^2 H_0(\mathbf{J}_1; s)}{\partial \mathbf{J}_1 \partial \mathbf{J}_1} \frac{\partial G}{\partial \phi} + \frac{\partial V(\phi, \mathbf{J}_1; s)}{\partial \mathbf{J}_1} \frac{\partial G}{\partial \phi} \right). \quad (11.69)$$

Note that  $\partial G / \partial \phi$  is a vector, each component of which is the derivative of  $G$  with respect to the corresponding component of  $\phi$ ; and  $\partial^2 H_0 / \partial \mathbf{J}_1 \partial \mathbf{J}_1$  is a matrix, the  $(i, j)$  element of which is the derivative of  $H_0$  with respect to  $J_{1i}$  and  $J_{1j}$ .

To complete this section, we shall discuss some applications of perturbation theory to specific cases, starting with the effect of a dipole perturbation on the closed orbit in a storage ring.

### 11.2.1 Dipole perturbations: closed orbit distortion

Consider a storage ring with a set of horizontal dipole field perturbations. The Hamiltonian for vertical particle motion in the storage ring can be written:

$$H = \frac{J_y}{\beta_y(s)} - \frac{q}{P_0} \Delta B_x(s) y, \quad (11.70)$$

where  $q$  is the particle charge,  $P_0$  is the reference momentum, and  $\Delta B_x(s)$  the horizontal component of the dipole field as a function of position along the reference trajectory. The action  $J_y$  and the co-ordinate  $y$  are implicitly functions of  $s$ : the goal in solving the equations of motion is to find the explicit form for the dependence of these quantities on  $s$ . For simplicity, we consider the motion in only the vertical degree of freedom, and assume that this is uncoupled with either the horizontal or longitudinal degrees of freedom. Writing the co-ordinate  $y$  in terms of the action-angle variables  $J_y$  and  $\phi_y$ :

$$y = \sqrt{2\beta_y(s)J_y} \cos(\phi_y), \quad (11.71)$$

the Hamiltonian becomes:

$$H = \frac{J_y}{\beta_y(s)} - \sqrt{2\beta_y(s)J_y} \frac{q}{2P_0} \Delta B_x(s)(e^{-i\phi_y} + e^{i\phi_y}). \quad (11.72)$$

Hence, it can be seen immediately that the Fourier components of the perturbing potential are (11.50):

$$\tilde{V}_{\pm 1}(J_{1y}, s) = -\sqrt{2\beta_y(s)J_{1y}} \frac{q}{2P_0} \Delta B_x(s). \quad (11.73)$$

All other components of  $\tilde{V}_m$  (i.e. for  $m \neq \pm 1$ ) are zero. The Fourier components of the function  $G$  (11.38) are then:

$$\tilde{G}_{\pm 1} = \mp i \frac{\sqrt{J_{1y}}}{2 \sin(\pi\nu_y)} \int_s^{s+C_0} \sqrt{2\beta_y(s')} \frac{q}{2P_0} \Delta B_x(s') e^{\pm i\theta_y} ds', \quad (11.74)$$

where:

$$\theta_y = \psi_y(s') - \psi_y(s) - \pi\nu_y. \quad (11.75)$$

The function  $G$  itself is:

$$G = \frac{\sqrt{J_{1y}}}{2 \sin(\pi\nu_y)} \int_s^{s+C_0} \sqrt{2\beta_y(s')} \frac{q}{P_0} \Delta B_x(s') \sin(\theta_y + \phi_y) ds'. \quad (11.76)$$

From (11.39), the solution to the equation of motion for the action  $J_y$  can then be written (to first order in the perturbation):

$$J_y = J_{1y} + \frac{\sqrt{J_{1y}}}{2 \sin(\pi\nu_y)} \int_s^{s+C_0} \sqrt{2\beta_y(s')} \frac{q}{P_0} \Delta B_x(s') \cos(\theta_y + \phi_y) ds', \quad (11.77)$$

where  $J_{1y}$  is constant (but we again emphasise that the dynamical variables  $J_y$  and  $\phi_y$  are functions of  $s$ ). Using (11.71), the co-ordinate  $y$  evolves as:

$$\begin{aligned} y \approx & \sqrt{2\beta_y(s)J_{1y}} \cos(\phi_y) \\ & + \frac{1}{2\sin(\pi\nu_y)} \int_s^{s+C_0} \sqrt{\beta_y(s)\beta_y(s')} \frac{q}{P_0} \Delta B_x(s') \cos(\theta_y + \phi_y) \cos(\phi_y) ds'. \end{aligned} \quad (11.78)$$

If required, the canonical momentum  $p_y$  can be found from the usual relationship with the action-angle variables:

$$p_y = -\sqrt{\frac{2J_y}{\beta_y}} (\sin(\phi_y) + \alpha_y \cos(\phi_y)), \quad (11.79)$$

with  $J_y$  given by (11.77). Using an appropriate trigonometric identity to take the dependence on  $\phi_y$  as a factor outside the integral, the equation for the co-ordinate (11.78) becomes:

$$y \approx \sqrt{2\beta_y(s)J_{1y}} \cos(\phi_y) + \frac{a(s)}{2\sin(\pi\nu_y)} \cos(\phi_y + \phi_0(s)) \cos(\phi_y), \quad (11.80)$$

where:

$$a(s)e^{i\phi_0(s)} = \int_s^{s+C_0} \sqrt{\beta_y(s)\beta_y(s')} \frac{q}{P_0} \Delta B_x(s') e^{i\theta_y} ds'. \quad (11.81)$$

$a(s)$  is a real number characterising the amplitude of the perturbation, and  $\phi_0(s)$  is a fixed phase. Both  $a(s)$  and  $\phi_0(s)$  are functions of  $s$ , the position around the ring.

Since  $J_{1y}$  is constant, the first term on the right-hand side of (11.80) appears as a regular betatron oscillation, with invariant action  $J_{1y}$ . To understand the physical significance of the second term, consider how the co-ordinate  $y$  changes over successive turns at a given point  $s$  in the ring. In the absence of the perturbation, the co-ordinate over successive turns oscillates between a minimum value:

$$y(\phi_y = \pi; s) = -\sqrt{2\beta_y(s)J_{1y}}, \quad (11.82)$$

and a maximum value:

$$y(\phi_y = 0; s) = \sqrt{2\beta_y(s)J_{1y}}. \quad (11.83)$$

When the perturbation is included, these values (for the same values of the betatron phase) become:

$$y(\phi_y = \pi; s) = -\sqrt{2\beta_y(s)J_{1y}} + \frac{a(s) \cos(\phi_0(s))}{2\sin(\pi\nu_y)}, \quad (11.84)$$

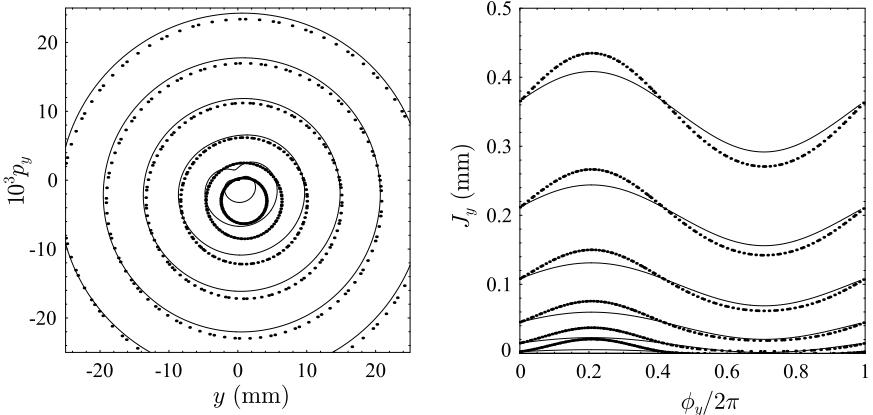


Fig. 11.1 Phase space portraits for a storage ring with small dipole field errors, in Cartesian variables (left) and action–angle variables (right). The points show the results of tracking 100 turns in a storage ring with tune  $\nu_y = 0.414$ , with observation point at a location immediately before a dipole kick  $\Delta p_y = 0.006$ . The Courant–Snyder parameters at the observation point are  $\beta_y = 1 \text{ m}$  and  $\alpha_y = 0$ . Each circular set of points (in Cartesian variables) corresponds to a different betatron action, from  $1.5 \mu\text{m}$  to  $350 \mu\text{m}$ . The lines show the phase space trajectories computed from perturbation theory, with  $y$  and  $p_y$  given by (11.71) and (11.79), respectively; the action  $J_y$  is given by (11.77) with invariant  $J_{1y}$ .

and:

$$y(\phi_y = 0; s) = \sqrt{2\beta_y(s)J_{1y}} + \frac{a(s) \cos(\phi_0(s))}{2 \sin(\pi\nu_y)}. \quad (11.85)$$

Although (11.84) and (11.85) are not exactly the minimum and maximum values of  $y$  in the presence of the perturbation, if the perturbation is small then these expressions will give the approximate limits of the oscillation of  $y$ : this is illustrated in the phase space portrait in Fig. 11.1. The oscillations appear to be shifted by an amount  $y_0$ :

$$\begin{aligned} y_0(s) &= \frac{a(s) \cos(\phi_0(s))}{2 \sin(\pi\nu_y)} \\ &= \frac{1}{2 \sin(\pi\nu_y)} \int_s^{s+C_0} \sqrt{\beta_y(s)\beta_y(s')} \frac{q}{P_0} \Delta B_x(s') \cos(\theta_y) ds'. \end{aligned} \quad (11.86)$$

We interpret the result (11.80) as representing a betatron oscillation with invariant amplitude  $J_{1y}$ , taking place around a fixed closed orbit given by (11.86). The expression (11.86) for the closed orbit is the same as the formula (6.21) that was derived, using rather more straightforward means, in Section 6.2.

### 11.2.2 Quadrupole perturbations: focusing errors

Focusing errors in a storage ring can easily arise from variations in the current flowing through the coils in the quadrupoles. In Section 6.4 it was shown that focusing errors lead to changes in the beta functions and the tunes. In this section, the effects of quadrupole errors will be derived using perturbation theory. For simplicity, we shall consider only the case of vertical motion, which we assume to be uncoupled from horizontal and longitudinal motion.

The Hamiltonian describing the vertical dynamics is:

$$H = \frac{J_y}{\beta_y(s)} - \Delta k_1(s) \frac{y^2}{2}, \quad (11.87)$$

where:

$$\Delta k_1(s) = \frac{q}{P_0} \Delta \left( \frac{\partial B_y(s)}{\partial x} \right) \quad (11.88)$$

is a small variation in the linear focusing at different locations around the ring. As usual,  $q$  is the particle charge,  $P_0$  is the reference momentum, and  $\partial B_y(s)/\partial x$  is the gradient of the magnetic field. Substituting for  $y$  from (11.71) into (11.87) gives:

$$H = \frac{J_y}{\beta_y(s)} - \frac{1}{2} \beta_y(s) J_y \Delta k_1(s) (1 + \cos(2\phi_y)). \quad (11.89)$$

The perturbation includes a term independent of the angle variable  $\phi_y$ . Therefore, the term  $H_0$  is defined:

$$H_0 = \frac{J_y}{\beta_y(s)} - \frac{1}{2} \beta_y(s) J_y \Delta k_1(s), \quad (11.90)$$

in terms of which the Hamiltonian is written:

$$H = H_0 - \frac{1}{2} \beta_y(s) J_y \Delta k_1(s) \cos(2\phi_y). \quad (11.91)$$

The tune of the lattice (i.e. the phase advance over one period of length  $C_0$ , in units of  $2\pi$ ) is given by:

$$\begin{aligned} \nu_y &= \frac{1}{2\pi} \int_s^{s+C_0} \frac{\partial H_0}{\partial J_y} ds \\ &= \frac{1}{2\pi} \int_s^{s+C_0} \frac{ds'}{\beta_y(s')} - \frac{1}{4\pi} \int_s^{s+C_0} \beta_y(s') \Delta k_1(s') ds'. \end{aligned} \quad (11.92)$$

There is a tune shift  $\Delta\nu_y$  arising from the focusing errors, given by:

$$\Delta\nu_y = -\frac{1}{4\pi} \int_s^{s+C_0} \beta_y(s') \Delta k_1(s') ds'. \quad (11.93)$$

This is in agreement with the expression (6.33) that was derived using transfer matrices.

To find an (approximate) solution to the equations of motion for the dynamical variables, perturbation theory can be applied in the same way as was done for dipole perturbations in Section 11.2.1. The perturbation is given by the potential:

$$V = -\frac{1}{2}\beta_y(s)J_y \Delta k_1(s) \cos(2\phi_y) = -\frac{1}{4}\beta_y(s)J_y \Delta k_1(s) (e^{-2i\phi} + e^{2i\phi}). \quad (11.94)$$

This leads to an expression for the function  $G$  (11.38) that appears in the generating function (transforming to new variables  $J_{1y}$  and  $\phi_{1y}$ , where  $J_{1y}$  is constant):

$$G = \frac{J_{1y}}{2\sin(2\pi\nu_y)} \int_s^{s+C_0} \frac{1}{2}\beta_y(s')\Delta k_1(s') \sin(2(\theta_y + \phi_y)) ds', \quad (11.95)$$

where:

$$\theta_y = \psi_y(s') - \psi_y(s) - \pi\nu_y, \quad (11.96)$$

and:

$$\psi_y(s) = \int_0^s \frac{1}{\beta_y(s')} - \frac{1}{2}\beta_y(s')\Delta k_1(s') ds'. \quad (11.97)$$

It is now possible, using (11.39), to write an expression for the co-ordinate  $y$  of a particle as a function of position  $s$ :

$$J_y = J_{1y} + \frac{J_{1y}}{2\sin(2\pi\nu_y)} \int_s^{s+C_0} \beta_y(s')\Delta k_1(s') \cos(2(\theta_y + \phi_y)) ds', \quad (11.98)$$

where  $J_{1y}$  is constant. Using (11.71), the co-ordinate  $y$  evolves as:

$$\begin{aligned} y &\approx \sqrt{2\beta_y(s)J_{1y}} \cos(\phi_y) \\ &\times \left( 1 + \frac{1}{4\sin(2\pi\nu_y)} \int_s^{s+C_0} \beta_y(s')\Delta k_1(s') \cos(2(\theta_y + \phi_y)) ds' \right). \end{aligned} \quad (11.99)$$

Using an appropriate trigonometric identity to take the dependence on  $\phi_y$  as a factor outside the integral, this becomes:

$$y \approx \sqrt{2\beta_y(s)J_{1y}} \left( 1 + \frac{b(s)}{4\sin(2\pi\nu_y)} \cos(2(\phi_y + \phi_0(s))) \right) \cos(\phi_y), \quad (11.100)$$

where:

$$b(s)e^{2i\phi_0(s)} = \int_s^{s+C_0} \beta_y(s')\Delta k_1(s')e^{2i\theta_y} ds'. \quad (11.101)$$

$b(s)$  is a real number characterising the amplitude of the perturbation, and  $\phi_0(s)$  is a phase constant. Both  $b(s)$  and  $\phi_0(s)$  are functions of  $s$ , the position around the ring.

In the case of a dipole perturbation, it was found that the perturbation introduced a shift (11.86) in the limits of the betatron oscillations (where the shift was a function of position around the ring). Applying similar arguments to the case of a quadrupole perturbation, we find from (11.100) that a quadrupole perturbation introduces a variation in the *amplitude* of the betatron oscillations, with the centre of the oscillations now remaining the same as without the perturbation. Since the action variable  $J_{1y}$  is constant around the ring, the variation in amplitude arising from the perturbation can be interpreted as a change  $\Delta\beta_y(s)$  in the beta function, given by:

$$\begin{aligned}\frac{\Delta\beta_y(s)}{\beta_y(s)} &\approx \frac{b(s) \cos(2\phi_0(s))}{2 \sin(2\pi\nu_y)} \\ &\approx \frac{1}{2 \sin(2\pi\nu_y)} \int_s^{s+C_0} \beta_y(s') \Delta k_1(s') \cos(2\theta_y) ds'.\end{aligned}\quad (11.102)$$

For small  $\Delta k_1 L$ , equation (11.102) is consistent with the formula that was derived using transfer matrices (6.34).

### 11.2.3 Skew quadrupole perturbations: coupling

Consider a planar storage ring in which the only coupling effects are from (small) skew quadrupole fields distributed around the ring. Using perturbation theory, it is possible to obtain approximate solutions to the equations of motion in the presence of the skew quadrupole fields, which reveal interesting aspects of the dynamics. The Hamiltonian describing particle motion in the ring is:

$$H = \frac{J_x}{\beta_x(s)} + \frac{J_y}{\beta_y(s)} + k_1^{(s)}(s)xy,\quad (11.103)$$

where  $k_1^{(s)}(s)$  is the skew quadrupole gradient multiplied by the particle charge and divided by the reference momentum:

$$k_1^{(s)}(s) = -\frac{q}{P_0} \frac{\partial B_x(s)}{\partial x}.\quad (11.104)$$

Using:

$$x = \sqrt{2\beta_x(s)J_x} \cos(\phi_x),\quad (11.105)$$

$$y = \sqrt{2\beta_y(s)J_y} \cos(\phi_y),\quad (11.106)$$

the Hamiltonian can be written:

$$H = H_0(J_x, J_y; s) + V(\phi_x, J_x, \phi_y, J_y; s), \quad (11.107)$$

where:

$$H_0(J_x, J_y; s) = \frac{J_x}{\beta_x(s)} + \frac{J_y}{\beta_y(s)}, \quad (11.108)$$

and:

$$V(\phi_x, J_x, \phi_y, J_y; s) = 2k_1^{(s)}(s)\sqrt{\beta_x(s)\beta_y(s)J_xJ_y}\cos(\phi_x)\cos(\phi_y). \quad (11.109)$$

The Fourier components of the perturbation potential are:

$$\tilde{V}_{\pm 1, \pm 1}(J_x, J_y; s) = \frac{1}{2}k_1^{(s)}(s)\sqrt{\beta_x(s)\beta_y(s)J_xJ_y}, \quad (11.110)$$

with all other  $\tilde{V}_{m_x, m_y}$  (i.e. for  $m_x \neq \pm 1$  and  $m_y \neq \pm 1$ ) equal to zero. Using (11.66), the generating function to remove the perturbation (the skew quadrupole term in the Hamiltonian) to first order is given by:

$$G = -\frac{\sqrt{J_{1x}J_{1y}}}{2\sin(\pi(\nu_x - \nu_y))}\Theta_- - \frac{\sqrt{J_{1x}J_{1y}}}{2\sin(\pi(\nu_x + \nu_y))}\Theta_+, \quad (11.111)$$

where:

$$\Theta_{\pm} = \int_s^{s+C_0} k_1^{(s)}(s')\sqrt{\beta_x(s')\beta_y(s')}\sin(\theta_x + \phi_x \pm \theta_y \pm \phi_y)ds', \quad (11.112)$$

and:

$$\theta_x = \psi_x(s') - \psi_x(s) - \pi\nu_x, \quad (11.113)$$

$$\theta_y = \psi_y(s') - \psi_y(s) - \pi\nu_y. \quad (11.114)$$

Also, using (11.67):

$$\psi_x(s) = \int_0^s \frac{ds'}{\beta_x(s')}, \quad (11.115)$$

$$\psi_y(s) = \int_0^s \frac{ds'}{\beta_y(s')}. \quad (11.116)$$

The betatron tunes are given by:

$$\nu_x = \frac{1}{2\pi}\psi_x(C_0), \quad (11.117)$$

$$\nu_y = \frac{1}{2\pi}\psi_y(C_0). \quad (11.118)$$

It appears that the skew quadrupole perturbation has no effect on the tunes, which seems inconsistent with the results found from the transfer matrix analysis in Section 6.6. However, (11.117) and (11.118) give the tunes to

first order in the perturbation; and from (6.68) and (6.69), the lowest order tune shifts are second order in the perturbation. We shall return to this point later.

Inspecting (11.111), we see that the generating function to remove the skew quadrupole perturbation to first order contains two terms. The first term becomes large when:

$$\nu_x - \nu_y = \text{integer}. \quad (11.119)$$

A storage ring with tunes approximately satisfying (11.119) is said to be operating close to a *difference resonance*. The second term in the generating function becomes large when:

$$\nu_x + \nu_y = \text{integer}. \quad (11.120)$$

The condition (11.120) defines the *sum resonance*. The behaviour of the betatron actions near a difference resonance is very different from that near a sum resonance. Consider first the difference resonance. Neglecting the term corresponding to the sum resonance, the betatron actions are given by:

$$\begin{aligned} J_x &= J_{1x} + \frac{\partial G}{\partial \phi_x} \\ &= J_{1x} - \frac{\sqrt{J_{1x} J_{1y}}}{2 \sin(\pi(\nu_x - \nu_y))} \\ &\times \int_s^{s+C_0} k_1^{(s)}(s') \sqrt{\beta_x(s') \beta_y(s')} \cos(\theta_x + \phi_x - \theta_y - \phi_y) ds', \end{aligned} \quad (11.121)$$

and:

$$\begin{aligned} J_y &= J_{1y} + \frac{\partial G}{\partial \phi_y} \\ &= J_{1y} + \frac{\sqrt{J_{1x} J_{1y}}}{2 \sin(\pi(\nu_x - \nu_y))} \\ &\times \int_s^{s+C_0} k_1^{(s)}(s') \sqrt{\beta_x(s') \beta_y(s')} \cos(\theta_x + \phi_x - \theta_y - \phi_y) ds'. \end{aligned} \quad (11.122)$$

Note that  $J_{1x}$  and  $J_{1y}$  are constant. Equations (11.121) and (11.122) can

be written in the form:

$$J_x = J_{1x} - \sqrt{J_{1x}J_{1y}} \frac{\kappa(s)}{2\sin(\pi(\nu_x - \nu_y))} \cos(\phi_x - \phi_y + \phi_0(s)), \quad (11.123)$$

$$J_y = J_{1y} + \sqrt{J_{1x}J_{1y}} \frac{\kappa(s)}{2\sin(\pi(\nu_x - \nu_y))} \cos(\phi_x - \phi_y + \phi_0(s)), \quad (11.124)$$

where:

$$\kappa(s)e^{i\phi_0(s)} = \int_s^{s+C_0} k_1^{(s)}(s') \sqrt{\beta_x(s')\beta_y(s')} e^{i(\theta_x - \theta_y)} ds'. \quad (11.125)$$

$\kappa(s)$  is a real number that characterises the strength of the coupling perturbation, and  $\phi_0(s)$  is a phase constant.

From (11.123) and (11.124), it is possible to see some significant features of the dynamics in a storage ring in the presence of a skew quadrupole perturbation, with the tunes near a difference resonance. First, it is clear that, since  $J_{1x}$  and  $J_{1y}$  are constant:

$$J_x + J_y = \text{constant}. \quad (11.126)$$

Second, we see that at a given point in the ring the betatron actions  $J_x$  and  $J_y$  oscillate around mean values given by  $J_{1x}$  and  $J_{1y}$  respectively. The oscillation corresponds to a transfer of betatron amplitude between the horizontal and vertical directions. The amplitude of the oscillations depends on the values of  $J_{1x}$ ,  $J_{1y}$ , on the strength of the skew quadrupole perturbation  $\kappa(s)$ , and on the proximity to the difference resonance. The frequency of the oscillations depends on the difference in tunes: after each turn, the phase of the oscillation advances by  $2\pi(\nu_x - \nu_y)$ . Close to the difference resonance, large amplitude, slow oscillations in the actions  $J_x$  and  $J_y$  are observed; as the distance from the resonance increases, the oscillations become more rapid and smaller in amplitude.

It is interesting to consider the expressions (11.123) and (11.124) for the betatron actions of a particle in a storage ring with coupling in the context of the equations of motion (7.159) derived in Section 7.4.1, where we discussed the impact of coupling on the equilibrium horizontal and vertical emittances in a storage ring. The equations of motion (7.159) were derived with a number of simplifications and approximations: in particular, we assumed that the coupling was driven by skew quadrupole fields that could be treated as a small perturbation to the optics, that the effect of the skew quadrupole fields could be represented by an average of the skew

quadrupole strength around the storage ring, and that the effects on the dynamics were dominated by a difference resonance, rather than a sum resonance. These assumptions are consistent with those we used in applying perturbation theory to derive the formulae (11.123) and (11.124). With some further approximations, we can see that these expressions for the betatron actions derived from perturbation theory are consistent with the equations of motion (7.159).

To show the consistency between the results from Section 7.4.1 and those from perturbation theory, first observe that from comparison of (7.156) and (11.125), we have:

$$K_\Delta = \frac{\kappa(0)}{C_0}, \quad (11.127)$$

$$\chi_\Delta = \phi_0(0) + \pi\nu_x - \pi\nu_y, \quad (11.128)$$

where we used  $K_\Delta$  and  $\chi_\Delta$  to characterise the strength and the phase of the coupling (averaged around the storage ring) in deriving the equations of motion (7.159), and we use  $\kappa(s)$  and  $\phi_0(s)$  to characterise the strength and the phase of the coupling (as functions of position around the ring) in perturbation theory. Based on these relationships, let us assume that in the expressions (11.123) and (11.124) we can replace  $\kappa(s)$  by  $C_0 K_\Delta$ , and (that since the phase  $\chi_\Delta$  does not appear in the final expressions (7.168) and (7.169) for the equilibrium emittances) that we can neglect  $\phi_0$ . Now we use (11.127) and take the derivatives of (11.123) and (11.124) to give:

$$\frac{dJ_x}{ds} = \sqrt{J_{1x} J_{1y}} \frac{C_0 K_\Delta}{2 \sin(\pi(\nu_x - \nu_y))} \sin(\phi_x - \phi_y) \frac{d\Delta\phi}{ds}, \quad (11.129)$$

$$\frac{dJ_y}{ds} = -\sqrt{J_{1x} J_{1y}} \frac{C_0 K_\Delta}{2 \sin(\pi(\nu_x - \nu_y))} \sin(\phi_x - \phi_y) \frac{d\Delta\phi}{ds}, \quad (11.130)$$

where  $\Delta\phi = \phi_x - \phi_y$ . We assume that the storage ring is tuned close to a difference resonance, so that:

$$2 \sin(\pi(\nu_x - \nu_y)) \approx 2\pi(\nu_x - \nu_y) \approx C_0 \frac{d\Delta\phi}{ds}, \quad (11.131)$$

where (for simplicity) we neglect any integer part of the difference in the betatron tunes. We then have:

$$\frac{dJ_x}{ds} = K_\Delta \sqrt{J_{1x} J_{1y}} \sin(\phi_x - \phi_y), \quad (11.132)$$

$$\frac{dJ_y}{ds} = -K_\Delta \sqrt{J_{1x} J_{1y}} \sin(\phi_x - \phi_y), \quad (11.133)$$

which closely resemble the equations of motion (7.159). Finally, we have already observed that in the perturbation theory, the skew quadrupole fields

have no impact on the betatron tunes, so that, averaging around the storage ring, we have:

$$\frac{d\phi_x}{ds} = \frac{2\pi\nu_x}{C_0} = \omega_x, \quad (11.134)$$

$$\frac{d\phi_y}{ds} = \frac{2\pi\nu_y}{C_0} = \omega_y, \quad (11.135)$$

where  $\omega_x$  and  $\omega_y$  are the average betatron frequencies (7.155).

The main differences between the equations of motion from perturbation theory (equations (11.132)–(11.135)) and the equations of motion derived in Section 7.4.1 (equations (7.159)) are first, that in the equations from perturbation theory, the constant values  $J_{1x}$  and  $J_{1y}$  appear instead of the action variables  $J_x$  and  $J_y$ , and second, that the betatron frequencies are unaffected by the skew quadrupole fields. Both of these differences are a result of the fact that we have carried the perturbation theory only to first order. Although this is sufficient to illustrate the consistency of the results with an analysis based on the full equations of motion, this example also serves to show that some caution is needed when applying perturbation theory, since significant higher-order effects may be overlooked.

In this context, it is worth looking a little more closely at the tune shifts that result from a skew quadrupole perturbation. As we have observed, to first order a skew quadrupole perturbation has no effect on the tunes: this is because the perturbation in this case has no component independent of the angle variables  $\phi_x$  and  $\phi_y$ . However, the canonical transformation from the old variables  $(\phi_x, J_x, \phi_y, J_y)$  to the new variables  $(\phi_{1x}, J_{1x}, \phi_{1y}, J_{1y})$  leads to a Hamiltonian in which the perturbation appears in second order. The second-order terms, which we previously neglected, *can* lead to a tune shift. We shall calculate this second-order tune shift for a skew quadrupole perturbation, and compare the result with the tune shift calculated using the matrix method in Section 6.6.

The general expression for the second-order terms in the transformed Hamiltonian is given in (11.69). For the particular case of skew quadrupole perturbations, from (11.108), (11.109) and (11.111), and neglecting the sum resonance term in  $G$ , it is found that the second-order term in the transformed Hamiltonian is:

$$H^{(2)} = \frac{\sqrt{\beta_x(s)\beta_y(s)}k_1^{(s)}(s)}{2\sin(\pi\Delta\nu)}(J_{1x} - J_{1y})\kappa(s)\cos(\phi_x)\cos(\phi_y)\cos(\phi_x - \phi_y - \phi_0(s)). \quad (11.136)$$

Before proceeding, we should substitute for  $\phi_x$  and  $\phi_y$  in terms of the new variables  $\phi_{1x}$  and  $\phi_{1y}$ , using (11.40). However, the difference between the old and new variables is first order in the perturbation, and the term  $H^{(2)}$  is second order in the perturbation; therefore, making the substitution will introduce new third-order terms, while in  $H^{(2)}$  we can simply replace  $\phi_x$  and  $\phi_y$  by  $\phi_{1x}$  and  $\phi_{1y}$ . Then, including the zeroth-order term  $H_0$ , it is found that:

$$\begin{aligned} \frac{d\phi_{1x}}{ds} &= \frac{\partial H_1}{\partial J_{1x}} = \\ \frac{1}{\beta_x(s)} + \frac{\sqrt{\beta_x(s)\beta_y(s)}k_1^{(s)}(s)}{2\sin(\pi\Delta\nu)}\kappa(s) \cos(\phi_{1x}) \cos(\phi_{1y}) \cos(\phi_{1x} - \phi_{1y} - \phi_0(s)). \end{aligned} \quad (11.137)$$

Similarly, for  $\phi_{1y}$ :

$$\begin{aligned} \frac{d\phi_{1y}}{ds} &= \frac{\partial H_1}{\partial J_{1y}} = \\ \frac{1}{\beta_y(s)} - \frac{\sqrt{\beta_x(s)\beta_y(s)}k_1^{(s)}(s)}{2\sin(\pi\Delta\nu)}\kappa(s) \cos(\phi_{1x}) \cos(\phi_{1y}) \cos(\phi_{1x} - \phi_{1y} - \phi_0(s)). \end{aligned} \quad (11.138)$$

The tune is the total phase advance over one complete turn around the ring, divided by  $2\pi$ . Thus, the tunes in the presence of the perturbation are (to second-order in the perturbation):

$$\nu'_x = \nu_x + \frac{\kappa}{4\pi\sin(\pi\Delta\nu)}I_k, \quad (11.139)$$

$$\nu'_y = \nu_y - \frac{\kappa}{4\pi\sin(\pi\Delta\nu)}I_k, \quad (11.140)$$

where we have replaced  $\kappa(s)$  by its mean value  $\kappa$ .  $\nu_x$  and  $\nu_y$  are the horizontal and vertical betatron tunes in the absence of the perturbation, and are given by:

$$\nu_x = \frac{1}{2\pi} \int_0^{C_0} \frac{ds}{\beta_x(s)}, \quad (11.141)$$

$$\nu_y = \frac{1}{2\pi} \int_0^{C_0} \frac{ds}{\beta_y(s)}. \quad (11.142)$$

The quantity  $I_k$  is defined by:

$$I_k = \int_0^{C_0} \sqrt{\beta_x(s)\beta_y(s)}k_1^{(s)}(s) \cos(\phi_{1x}) \cos(\phi_{1y}) \cos(\phi_{1x} - \phi_{1y} - \phi_0(s)) ds. \quad (11.143)$$

Averaging over  $\phi_{1x}$  and  $\phi_{1y}$  gives:

$$\langle I_k \rangle = \frac{1}{4} \int_0^{C_0} \sqrt{\beta_x(s)\beta_y(s)} k_1^{(s)}(s) \cos(\phi_0(s)) ds. \quad (11.144)$$

In the case that the perturbation arises from small skew quadrupole errors distributed randomly around the ring, then we expect that:

$$\int_0^{C_0} \sqrt{\beta_x(s)\beta_y(s)} k_1^{(s)}(s) \cos(\phi_0(s)) ds = \kappa, \quad (11.145)$$

where  $\kappa$  is the mean of the coupling strength  $\kappa$  defined by (11.125). This gives the final expressions for the tune shifts to second order in the perturbation:

$$\nu'_x = \nu_x + \frac{\kappa^2}{16\pi \sin(\pi \Delta\nu)}, \quad (11.146)$$

$$\nu'_y = \nu_y - \frac{\kappa^2}{16\pi \sin(\pi \Delta\nu)}. \quad (11.147)$$

If we associate  $\nu'_x$  and  $\nu'_y$  with the normal mode tunes  $\nu_I$  and  $\nu_{II}$ , and the coupling strength  $\kappa$  with the integrated strength  $k_1^{(s)} L$  of a single skew quadrupole weighted by the beta functions:

$$\kappa = \sqrt{\beta_x \beta_y} |k_1^{(s)} L|, \quad (11.148)$$

then (11.146) and (11.147) are consistent with the results (6.68) and (6.69) obtained using matrix methods (and with the tune shifts found in Section 7.4.1 by considering the equations of motion of a particle in a storage ring with a single skew quadrupole).

To complete our discussion of skew quadrupole perturbations, it is worth making a few brief remarks about the sum resonance. If the tunes in a storage ring are such that the sum resonance condition (11.120) is approximately satisfied but the difference resonance condition (11.119) is not, then the second term in the expression (11.111) for the generating function dominates over the first. Neglecting the first term, the equations for the actions  $J_x$  and  $J_y$  in terms of constants  $J_{1x}$  and  $J_{1y}$  are:

$$J_x = J_{1x} - \sqrt{J_{1x} J_{1y}} \frac{\bar{\kappa}(s)}{2 \sin(\pi(\nu_x + \nu_y))} \cos(\phi_x + \phi_y + \bar{\phi}_0(s)), \quad (11.149)$$

$$J_y = J_{1y} - \sqrt{J_{1x} J_{1y}} \frac{\bar{\kappa}(s)}{2 \sin(\pi(\nu_x + \nu_y))} \cos(\phi_x + \phi_y + \bar{\phi}_0(s)), \quad (11.150)$$

where:

$$\bar{\kappa}(s)e^{i\bar{\phi}_0(s)} = \int_s^{s+C_0} k_1^{(s)}(s') \sqrt{\beta_x(s')\beta_y(s')} e^{i(\theta_x+\theta_y)} ds'. \quad (11.151)$$

$\bar{\kappa}(s)$  is a real number that characterises the strength of the coupling perturbation, and  $\bar{\phi}_0(s)$  is a phase constant. Comparing (11.149) and (11.150) with (11.123) and (11.124), it is clear that the dynamics near a sum resonance are quite different from those near a difference resonance. Near a difference resonance, the sum of the actions  $J_x + J_y$  is constant: this means that the amplitude of the betatron oscillation in one plane can grow only if the oscillation amplitude in the other plane gets smaller. The actions themselves oscillate (slowly, with a period typically taking many turns of the ring) in antiphase; and the fact that the minimum possible value for the betatron action is zero means that there is a natural limit on the amount by which either betatron action can grow. However, near a sum resonance, the *difference* in the betatron actions  $J_x - J_y$  is constant. Inspecting (11.149) and (11.150), the actions oscillate in much the same way as they do near a difference resonance, except that the oscillations are now in phase. In the case of an electron storage ring, if the horizontal and vertical emittances are identified with mean values of the horizontal and vertical betatron action variables, this implies that there is effectively no limit on the emittances, which may become extremely large close to a sum resonance: skew quadrupole fields can have a strong adverse impact on beam quality in an electron storage ring operated near a sum resonance.

#### 11.2.4 Sextupole perturbations

The Hamiltonian describing particle motion in a storage ring with a sextupole perturbation is:

$$H = \frac{J_x}{\beta_x(s)} + \frac{J_y}{\beta_y(s)} + V(\phi_x, J_x, \phi_y, J_y; s), \quad (11.152)$$

where:

$$\begin{aligned} V(\phi_x, J_x, \phi_y, J_y; s) &= \frac{1}{6} k_2(s)(x^3 - 3xy^2) \\ &= \frac{\sqrt{2}}{3} k_2(s) \beta_x(s)^{3/2} J_x^{3/2} \cos^3(\phi_x) \\ &\quad - \sqrt{2} k_2(s) \sqrt{\beta_x(s) J_x} \beta_y(s) J_y \cos(\phi_x) \cos^2(\phi_y). \end{aligned} \quad (11.153)$$

$k_2(s)$  is the sextupole gradient multiplied by the particle charge and divided by the reference momentum:

$$k_2(s) = \frac{q}{P_0} \frac{\partial^2 B_y(s)}{\partial x^2}. \quad (11.154)$$

In general, the generating function that removes a perturbation to first order is given by (11.66). In the particular case of a sextupole perturbation, the function  $G$  is given by:

$$\begin{aligned} G(\phi_x, J_{1x}, \phi_y, J_{1y}; s) = & \\ & - \int_s^{s+C_0} \frac{(\beta_x(s') J_{1x})^{3/2}}{4\sqrt{2} \sin(\pi\nu_x)} k_2(s') \sin(\theta_x + \phi_x) ds' \\ & - \int_s^{s+C_0} \frac{(\beta_x(s') J_{1x})^{3/2}}{12\sqrt{2} \sin(3\pi\nu_x)} k_2(s') \sin(3(\theta_x + \phi_x)) ds' \\ & + \int_s^{s+C_0} \frac{\beta_y(s') J_{1y} \sqrt{\beta_x(s') J_{1x}}}{2\sqrt{2} \sin(\pi\nu_x)} k_2(s') \sin(\theta_x + \phi_x) ds' \\ & + \int_s^{s+C_0} \frac{\beta_y(s') J_{1y} \sqrt{\beta_x(s') J_{1x}}}{4\sqrt{2} \sin(\pi(\nu_x + 2\nu_y))} k_2(s') \sin(\theta_x + \phi_x + 2\theta_y + 2\phi_y) ds' \\ & + \int_s^{s+C_0} \frac{\beta_y(s') J_{1y} \sqrt{\beta_x(s') J_{1x}}}{4\sqrt{2} \sin(\pi(\nu_x - 2\nu_y))} k_2(s') \sin(\theta_x + \phi_x - 2\theta_y - 2\phi_y) ds', \end{aligned} \quad (11.155)$$

where:

$$\theta_x = \psi_x(s') - \psi_x(s) - \pi\nu_x, \quad (11.156)$$

$$\theta_y = \psi_y(s') - \psi_y(s) - \pi\nu_y, \quad (11.157)$$

and:

$$\psi_x(s) = \int_0^s \frac{ds'}{\beta_x(s')}, \quad (11.158)$$

$$\psi_y(s) = \int_0^s \frac{ds'}{\beta_y(s')}. \quad (11.159)$$

We see that the generating function contains driving terms for integer resonances, i.e. terms that become infinitely large when:

$$\nu_x = \text{integer}. \quad (11.160)$$

Also there are driving terms for third-integer resonances:

$$3\nu_x = \text{integer}, \quad (11.161)$$

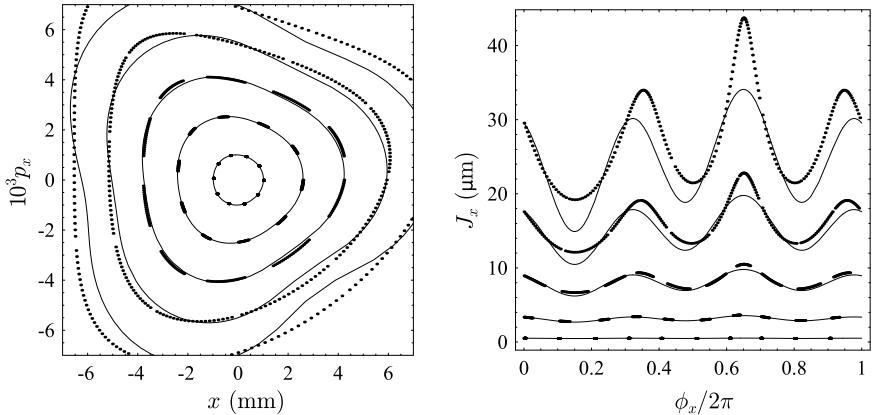


Fig. 11.2 Phase space portrait for a sextupole perturbation in a storage ring, in Cartesian variables (left) and action–angle variables (right). The points show the results of tracking 200 turns in a storage ring with tune  $\nu_x = 0.30$ , with observation point at a location immediately before a sextupole with integrated strength  $k_2 L = 100 \text{ m}^{-2}$ . The Courant–Snyder parameters at the observation point are  $\beta_x = 1 \text{ m}$  and  $\alpha_x = 0$ . Each closed loop of points (in Cartesian variables) corresponds to a different value of the invariant  $J_{1x}$ , from  $0.5 \mu\text{m}$  to  $30 \mu\text{m}$ . The lines show the phase space trajectories computed from perturbation theory, with horizontal action given by (11.164).

and for coupling resonances:

$$\nu_x + 2\nu_y = \text{integer}, \quad (11.162)$$

$$\nu_x - 2\nu_y = \text{integer}. \quad (11.163)$$

The variation of the betatron actions  $J_x$  and  $J_y$  are obtained (as usual) from:

$$J_x = J_{1x} + \frac{\partial G}{\partial \phi_x}, \quad (11.164)$$

$$J_y = J_{1y} + \frac{\partial G}{\partial \phi_y}, \quad (11.165)$$

where  $J_{1x}$  and  $J_{1y}$  are constant. For a given sextupole distribution it is possible to generate phase space portraits by plotting the Cartesian variables ( $(x, p_x)$  in the horizontal plane, and  $(y, p_y)$  in the vertical plane) for fixed  $J_{1x}$  and  $J_{1y}$ , as  $\phi_x$  and  $\phi_y$  are varied over the range from 0 to  $2\pi$ . The Cartesian variables are related to the action–angle variables by the usual relations of the form (4.36) and (4.37). Some examples for the horizontal plane (i.e. with  $J_{1y} = 0$ ) are shown in Fig. 11.2.

There are no tune shifts to first order in the perturbation. However, an approximation to the tune shifts (to second order in the perturbation)

can be obtained from the second-order term in (11.69), averaged over  $\phi_x$  and  $\phi_y$ . The calculation in two degrees of freedom gets very cumbersome; however, working in one degree of freedom by setting  $J_{1y} = 0$  makes things more manageable. In this case, since  $H_0$  only has a first-order dependence on  $J_{1x}$ , the horizontal tune is given by:

$$\begin{aligned}\nu'_x &= \frac{1}{2\pi} \int_0^{C_0} \frac{\partial H_1}{\partial J_{1x}} ds \\ &= \frac{1}{2\pi} \int_0^{C_0} \frac{1}{\beta_x(s)} + \frac{\partial}{\partial J_{1x}} \left\langle \frac{\partial G}{\partial \phi_x} \frac{\partial V}{\partial J_{1x}} \right\rangle ds,\end{aligned}\quad (11.166)$$

where the brackets  $\langle \cdot \rangle$  here indicate an average over  $\phi_x$  and  $\phi_y$ . Using (11.153) and (11.155), we find, with  $J_{1y} = 0$ :

$$\begin{aligned}\nu'_x &= \nu_x - \frac{J_{1x}}{64\pi} \int_0^{C_0} ds \int_s^{s+C_0} ds' \\ &\quad \times \beta_x(s)^{3/2} k_2(s) \beta_x(s')^{3/2} k_2(s') \left( \frac{3 \cos(\theta_x)}{\sin(\pi\nu_x)} + \frac{\cos(3\theta_x)}{\sin(3\pi\nu_x)} \right).\end{aligned}\quad (11.167)$$

$\nu_x$  is the horizontal tune in the absence of the sextupole perturbation. In the case of a single sextupole magnet in a storage ring, with integrated strength  $k_2 L$ , the horizontal tune (with zero vertical action) is given by:

$$\nu'_x = \nu_x - \frac{J_{1x}}{64\pi} \beta_x^3 (k_2 L)^2 (3 \cot(\pi\nu_x) + \cot(3\pi\nu_x)),\quad (11.168)$$

where  $\beta_x$  is the horizontal beta function at the location of the sextupole. As might be expected, if the tune is close to an integer or a third integer, then the tune shift with amplitude  $J_{1x}$  becomes extremely large.

### 11.3 Resonances and Dynamic Aperture

The results from perturbation theory developed in Section 11.2 reveal some significant features of single-particle dynamics in beam lines (and, in particular, storage rings) that have important consequences for accelerator design. It is worth discussing these features and consequences in more detail. First, to summarise briefly the key points of perturbation theory, recall that the basic idea is to construct a generating function for a transformation of variables that puts the Hamiltonian into a simpler form — a form where the equations of motion can be solved exactly if terms above some order in the angle-dependent part of the Hamiltonian are neglected. In two degrees of freedom, the Hamiltonian is written as:

$$H = H_0(J_x, J_y; s) + \epsilon V(\phi_x, J_x, \phi_y, J_y; s),\quad (11.169)$$

where  $J_x, \phi_x, J_y, \phi_y$  are the dynamical (action-angle) variables, and  $s$  is the independent variable.  $\epsilon$  is a parameter introduced to help to keep track of the order of the perturbation  $V$ . The generating function to transform to new variables  $J_{1x}, \phi_{1x}, J_{1y}, \phi_{1y}$  is:

$$F_2 = \phi_x J_{1x} + \phi_y J_{1y} + \epsilon G(\phi_x, J_{1x}, \phi_y, J_{1y}; s). \quad (11.170)$$

For  $G = 0$ , the transformation generated by  $F_2$  is the identity. In the new variables, the Hamiltonian is:

$$H_1 = H + \frac{\partial F_2}{\partial s} = H + \epsilon \frac{\partial G}{\partial s}. \quad (11.171)$$

First-order perturbation theory shows how to construct the function  $G$  from the perturbation potential  $V$ , so that:

$$H_1 = H_0(J_{1x}, J_{1y}; s) + O(\epsilon^2). \quad (11.172)$$

Neglecting terms of order  $\epsilon^2$ , the Hamiltonian  $H_1$  is integrable. The equations of motion in the new variables can be solved exactly (for given initial conditions); and the generating function can then be used to transform the solution into the original variables.

Applied to a periodic lattice, such as a storage ring with circumference  $C_0$ , terms appearing in the generating function are typically of the form:

$$\begin{aligned} G \sim & \int_s^{s+C_0} \frac{(\beta_x(s')J_{1x})^{|m_x|/2}(\beta_y(s')J_{1y})^{|m_y|/2}}{\sin(\pi(m_x\nu_x + m_y\nu_y))} \\ & \times k_n(s') \sin(m_x\theta_x + m_x\phi_x + m_y\theta_y + m_y\phi_y) ds', \end{aligned} \quad (11.173)$$

where  $k_n(s)$  is a multipole component of the magnet field as a function of position around the ring,  $m_x$  and  $m_y$  are integers, the dynamical variables  $J_{1x}$  and  $J_{1y}$  are constants of the motion,  $\beta_x$  and  $\beta_y$  are the beta functions, and  $\nu_x$  and  $\nu_y$  are the tunes (in the absence of the perturbation). The functions  $\theta_x$  and  $\theta_y$  are given by:

$$\theta_x = \psi_x(s') - \psi_x(s) - \pi\nu_x, \quad (11.174)$$

$$\theta_y = \psi_y(s') - \psi_y(s) - \pi\nu_y, \quad (11.175)$$

where:

$$\psi_x(s) = \int_0^s \frac{ds'}{\beta_x(s')}, \quad (11.176)$$

$$\psi_y(s) = \int_0^s \frac{ds'}{\beta_y(s')}, \quad (11.177)$$

so that  $\nu_x = \psi(C_0)/2\pi$ , and similarly for  $\nu_y$ . Synchrotron motion can be included by extending these formulae to three degrees of freedom.

The generating function typically involves many terms of the form given in (11.173). The generating function removes terms in the Hamiltonian to first order in the perturbation; but the new Hamiltonian includes higher-order terms. If  $G$  is small (so that the transformation to the new variables is close to the identity) these terms can be neglected, and the equations of motion derived from the new Hamiltonian can be solved exactly. But if  $G$  contains terms that are large, the higher-order terms cannot be neglected, and perturbation theory fails. By inspection of (11.173), terms in  $G$  can become large if the tunes are such that:

$$m_x \nu_x + m_y \nu_y = \text{integer}. \quad (11.178)$$

It is often sufficient to consider only betatron resonances; however, in some cases it is necessary to include the effects of coupling between the transverse and longitudinal motion, in which case (11.178) can be generalised to include synchro-betatron resonances:

$$m_x \nu_x + m_y \nu_y + m_z \nu_z = \text{integer}, \quad (11.179)$$

where  $\nu_z$  is the synchrotron tune, and  $m_z$  is an integer. For simplicity, in the following discussion we shall consider only betatron resonances. The condition (11.178) is said to correspond to a resonance of order  $|m_x| + |m_y|$ . Physically, when a storage ring is tuned onto a resonance, the kicks that particles receive from a perturbation at some given point add coherently from turn to turn, leading to a large change in the betatron action. Resonances are generally harmful for operation of a storage ring, but the impact of any given resonance depends on a number of factors. In particular, there are three important considerations for the choice of working point (i.e. the particular tune values) for a storage ring: first, the proximity of the working point to different resonances; second, the strength with which the various resonances are driven (broadly speaking, the driving strength is given by the numerator in the integral in (11.173)); and finally, the dynamical behaviour resulting from the nearby resonances.

Consider first the proximity of a given working point with tunes  $\nu_{x0}$  and  $\nu_{y0}$ , to a particular resonance:

$$m_x \nu_x + m_y \nu_y = \ell, \quad (11.180)$$

where  $\ell$  is an integer. It is assumed that the greatest common divisor of  $m_x$ ,  $m_y$  and  $\ell$  is 1. The condition (11.180) defines a line in a co-ordinate

space with axes  $\nu_x$  and  $\nu_y$ . Suppose that the working point  $(\nu_{x0}, \nu_{y0})$  is a distance  $d$  from the line defined by (11.180), so that:

$$\frac{|m_x\nu_{x0} + m_y\nu_{y0} - \ell|}{\sqrt{m_x^2 + m_y^2}} = d. \quad (11.181)$$

It follows that:

$$m_x\nu_{x0} + m_y\nu_{y0} = \ell \pm d\sqrt{m_x^2 + m_y^2}, \quad (11.182)$$

and therefore, for small  $d$ :

$$\sin(\pi(m_x\nu_{x0} + m_y\nu_{y0})) \approx \pm\pi d\sqrt{m_x^2 + m_y^2}. \quad (11.183)$$

The higher the order of the resonance (i.e. the larger  $|m_x| + |m_y|$ ), the closer the working point has to be to the resonance before the resonance starts to have a significant impact. This is fortunate, since any point in tune space will be close to a resonance of some order. The fact that the effects of low-order resonances extend further from the resonance than the effects of high-order resonances means that, when choosing the working point for a storage ring, more attention can be given to low-order resonances.

The second issue that needs to be considered when choosing the tunes of a storage ring is the strength of the driving terms of different resonances. Factoring out the dynamical variables and the dependence  $1/\sin(\pi(m_x\nu_x + m_y\nu_y))$  from (11.173), the strength with which a particular resonance is driven can be characterised by a value  $r$ , where:

$$re^{i\chi} = \int_s^{s+C_0} \beta_x(s')^{|m_x|/2} \beta_y(s')^{|m_y|/2} k_n(s') e^{i(m_x\theta_x + m_y\theta_y)} ds'. \quad (11.184)$$

The magnitude of  $r$  will depend on the details of the lattice design and the errors that are present in the lattice. This makes it difficult to draw any general conclusions. However, note that the terms in the generating function (11.173) include factors  $J_{1x}^{|m_x|/2}$  and  $J_{1y}^{|m_y|/2}$ . This implies that for sufficiently small betatron actions, the generating function will approach the identity; the higher the order of the resonance, the faster the corresponding term in the generating function will approach zero as the actions are reduced. This means that very high order resonances can effectively be ignored, as long as it is not intended for the storage ring to store particles that have betatron actions larger than some given limit.

Finally, the characteristics of the dynamical behaviour resulting from various resonances need to be considered. For example, we have seen in Section 11.2.3 that the effects of a skew quadrupole perturbation near a

difference resonance are very different from the effects of the same perturbation near a sum resonance. Near a difference resonance, the beam may become fully coupled, but, at least, the amplitudes of the betatron oscillations will be limited. On the other hand, near a sum resonance, the amplitudes of the betatron oscillations become larger, without limit, as the resonance is approached. Perturbation theory can be helpful in understanding the likely effects of a given resonance, but it should be remembered that the theory is only valid if the tunes are sufficiently far from a resonance.

A useful tool for choosing the working point of a storage ring is the *resonance diagram*, also known as the *tune grid*. This is simply a plot showing different resonance lines (11.180) in a co-ordinate space with axes  $\nu_x$  and  $\nu_y$ . Usually, all resonances up to some order are shown. Note that only the fractional parts of the tunes are significant, so the axes only need to range from 0 to 1. An example of a resonance diagram, displaying resonance lines up to fourth order, is Fig. 11.3. The working point of a given storage ring design can be shown as a point on the resonance diagram: the proximity of the working point to different resonances can then be seen immediately. Of course, the resonance diagram only shows the *presence* of a resonance: on its own, it gives no indication of how strongly a particular resonance might be driven in a given lattice design, or what the dynamical effects of the resonance might be.

An important issue to consider when choosing the working point of a storage ring is the periodicity of the lattice. Storage rings can be (and often are) constructed from repeated unit cells, which may be FODO cells, for example, or double-bend achromat cells. If a storage ring lattice has a high degree of periodicity, so that multipoles of a given order and strength occur at regular intervals around the ring (with the same phase advance between each adjacent pair), then there will be a tendency for the driving strengths of most resonances to be suppressed. This can be understood in terms of the integral (11.184). If a storage ring consists of  $N$  identical cells, then for a particular family of multipoles with strength  $k_n$ , it follows from (11.184) that:

$$re^{i\chi} = \beta_x^{|m_x|/2} \beta_y^{|m_y|/2} k_n \sum_{p=1}^N e^{\frac{p}{N} 2\pi i (m_x \nu_x + m_y \nu_y)}, \quad (11.185)$$

where we have used the fact that the horizontal and vertical phase advances over a single cell are  $\nu_x/N$  and  $\nu_y/N$  respectively. Hence:

$$r = \beta_x^{|m_x|/2} \beta_y^{|m_y|/2} k_n \frac{\sin(\pi(m_x \nu_x + m_y \nu_y))}{\sin(\pi(m_x \nu_x + m_y \nu_y)/N)}. \quad (11.186)$$

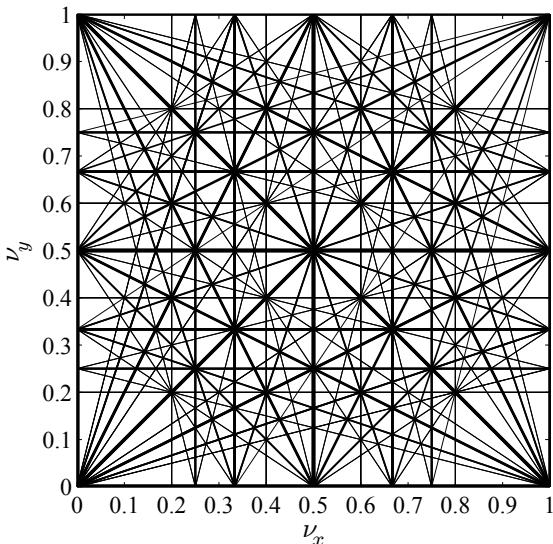


Fig. 11.3 Resonance diagram, showing resonances up to fifth order. Each line shows points satisfying (11.180), with  $|m_x| + |m_y| \leq 5$ . Lower-order resonances are shown with bolder lines.

The contribution to the generating function (11.173) becomes essentially the same as the corresponding term that we would write for a single cell in the lattice. This is as we should expect: the dynamics in a storage ring consisting of  $N$  identical cells is essentially the same as the dynamics in a single cell (except that the phase advance is  $N$  times as large over the ring as over a single cell). To put another way, in a periodic lattice, the effects of the sextupoles (for example) do not add up to give any stronger effects than those that occur in just a single cell of the lattice. If the periodicity of the lattice is broken, either accidentally (for example, by focusing errors in the quadrupoles leading to variations in the phase advance) or deliberately (for example, by inserting cells in the lattice that are not identical to the other cells), then the value of  $r$  will no longer be given by (11.186). In some cases, depending on how the periodicity is broken, the tunes and the resonance involved, the value of  $r$  may become smaller; but there will almost certainly be resonances for which the value of  $r$  becomes larger, and possibly much larger. Generally, to avoid increasing the strengths of any resonances over those found in a single cell, a storage ring should be constructed from an exact repetition of the unit cell.

Unfortunately, various design constraints mean that most storage rings cannot have a high degree of periodicity. For example, the periodicity in a storage ring designed for a collider will be broken by the interaction regions, and light sources often need some sections where the optics are adjusted to match the requirements of particular insertion devices. Also, there are inevitably regions for beam injection, and for systems such as the rf cavities, that will be different from other regions of the ring. However, there are ways to avoid the damaging effects associated with breaking the periodicity in a storage ring. One common technique is to ensure that sections inserted in an otherwise periodic lattice have phase advances that are exact multiples of  $2\pi$  in both planes. If these sections are also linear (i.e. contain no higher-order multipoles) then they are effectively ‘transparent’ in terms of any contribution to the resonance driving terms. Unfortunately, the effects of chromaticity mean that the ideal phase advance can only be achieved for particles with energy exactly equal to the reference energy: the result is that the periodicity of the lattice is still broken for particles with non-zero energy deviation.

The perturbation theory developed in Section 11.2 can describe the effects of a perturbation on particle motion, as long as the tunes are not too close to a resonance. Close to a resonance, the transformations required to eliminate the perturbation terms from the Hamiltonian become large (i.e. distant from the identity), and the description of the dynamics provided by perturbation theory is not accurate. However, since it is almost impossible to avoid resonances completely, it is of interest to have some understanding of the particle dynamics in the presence of a resonance. With some simplifying assumptions, it is possible to construct a phase space portrait for dynamics in the vicinity of a resonance that provides some insight; to achieve this, we proceed as follows.

In one degree of freedom, the Hamiltonian for particle motion in a storage ring can be written in the general form:

$$H = H_0(J_x; s) + \sum_{m,n=-\infty}^{\infty} \tilde{f}_{mn}(J_x) e^{im\phi_x} e^{-2\pi ins/C_0}, \quad (11.187)$$

where  $C_0$  is the circumference of the storage ring, and  $\tilde{f}_{mn}(J_x)$  are a set of Fourier coefficients that are functions of the action  $J_x$ . Note that we have made use of the periodicity of the Hamiltonian in  $\phi_x$  and  $s$ . In the vicinity of a strongly driven resonance, we can assume that there are a few terms in the Fourier series that dominate over the others. Thus, the Hamiltonian

can be approximated as:

$$H = H_0(J_x; s) + \tilde{f}(J_x) \cos(m\phi_x - 2\pi ns/C_0), \quad (11.188)$$

where  $\tilde{f}(J_x)$  is a real function of the action, related to the dominant coefficients from the set  $f_{mn}$ . To eliminate the explicit dependence on  $s$  in the second term of the Hamiltonian (11.188), we can make a canonical transformation to new variables  $\phi_{1x}$ ,  $J_{1x}$  using the generating function:

$$F_2 = \phi_x J_{1x} - 2\pi \frac{n}{m} \frac{s}{C_0} J_{1x}. \quad (11.189)$$

Then, the original and the new variables are related by:

$$J_x = \frac{\partial F_2}{\partial \phi_x} = J_{1x}, \quad (11.190)$$

$$\phi_{1x} = \frac{\partial F_2}{\partial J_{1x}} = \phi_x - 2\pi \frac{n}{m} \frac{s}{C_0}, \quad (11.191)$$

and the new Hamiltonian is given by:

$$\begin{aligned} H_1 &= H + \frac{\partial F_2}{\partial s} \\ &= H_0(J_{1x}, s) - 2\pi \frac{n}{m} \frac{1}{C_0} J_{1x} + \tilde{f}(J_{1x}) \cos(m\phi_{1x}). \end{aligned} \quad (11.192)$$

To remove completely the explicit dependence of  $H_1$  on  $s$ , we assume that we can replace  $H_0(J_{1x}; s)$  by its average over the circumference of the ring:

$$H_0(J_{1x}) = \frac{1}{C_0} \int_0^{C_0} H_0(J_{1x}; s) ds = \frac{2\pi\nu_x}{C_0} J_{1x} + \alpha(J_{1x}), \quad (11.193)$$

where  $\nu_x$  is the tune of the lattice in the limit  $J_{1x} \rightarrow 0$ , and  $\alpha(J_{1x})$  is a function containing powers of  $J_{1x}$  higher than 1 (not to be confused with the Courant–Snyder parameter  $\alpha_x$ ). Note that it is also assumed that  $\tilde{f}(J_{1x})$  contains powers of  $J_{1x}$  higher than 1. Then, the Hamiltonian becomes:

$$H_1 = \frac{2\pi}{C_0} \left( \nu_x - \frac{n}{m} \right) J_{1x} + \alpha(J_{1x}) + \tilde{f}(J_{1x}) \cos(m\phi_{1x}). \quad (11.194)$$

The equations of motion derived from the Hamiltonian (11.194) still do not admit of simple analytical solutions; however, since the explicit dependence on  $s$  has been removed from the Hamiltonian, the Hamiltonian must be a constant of the motion. This fact alone allows us to derive interesting properties of the dynamics; in particular, for any given values of the various parameters and functions appearing in the Hamiltonian (11.194), we can draw a phase space portrait simply by plotting contours with  $H_1 = \text{constant}$  in the phase space. It is convenient to do this using Cartesian variables:

$$x_1 = \sqrt{2J_{1x}} \cos(\phi_{1x}), \quad (11.195)$$

$$p_{x1} = -\sqrt{2J_{1x}} \sin(\phi_{1x}). \quad (11.196)$$

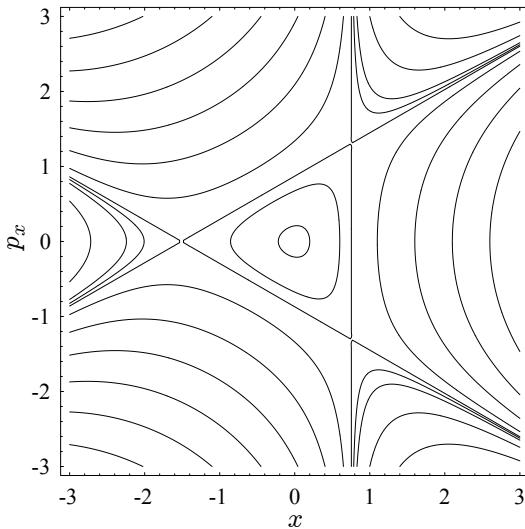


Fig. 11.4 Phase space portrait for the resonance Hamiltonian (11.200). The separatrices bound the region of stable, periodic motion, and intersect at unstable fixed points.

The standard Cartesian variables are related to the (original) action–angle variables  $J_x$  and  $\phi_x$  by:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (11.197)$$

$$p_x = -\sqrt{\frac{2J_x}{\beta_x}} (\sin(\phi_x) + \alpha_x \cos(\phi_x)). \quad (11.198)$$

There are two significant differences between the variables  $(x, p_x)$  and  $(x_1, p_{x1})$ . The first is that the variables  $(x_1, p_{x1})$  are normalised, so that for  $\dot{f}(J_{1x}) = 0$ , the particle motion maps out circles rather than ellipses in phase space. The second, more important difference, is that a phase advance  $\Delta\phi_x$  in the variables  $(x, p_x)$  corresponds to a phase advance:

$$\Delta\phi_{1x} = \Delta\phi_x - 2\pi \frac{n}{m} \frac{s}{C_0}, \quad (11.199)$$

in the variables  $(x_1, p_{x1})$ . We shall see the significance of this point shortly.

An example phase space portrait in the variables  $(x_1, p_{x1})$  is shown in Fig. 11.4. The Hamiltonian for this particular case is:

$$H_1 = 1.6J_{1x} + J_{1x}^{3/2} \cos(3\phi_{1x}). \quad (11.200)$$

For small amplitudes, the motion is dominated by the linear term (proportional to  $J_{1x}$ ) in the Hamiltonian, and particle trajectories follow circles

in phase space centred on the origin. However, for larger amplitudes, the motion becomes distorted and contour lines develop a triangular shape, characteristic of a third-order resonance. If the amplitude of the motion is increased further, eventually a boundary is crossed beyond which the motion no longer follows closed loops, but the contour lines (lines of constant value for the Hamiltonian) extend out to infinity. In the case of a storage ring, a particle on such a trajectory would be lost from the beam pipe.

Along the boundary of the stable motion, there are points where the contour lines appear to intersect. If the phase space is viewed as a surface where the height is given by the value of the Hamiltonian, then the contour lines intersect at saddle points in the phase space. At these points, the Hamiltonian satisfies:

$$\frac{\partial H_1}{\partial J_{1x}} = \frac{\partial H_1}{\partial \phi_{1x}} = 0, \quad (11.201)$$

and it then follows from Hamilton's equations that:

$$\frac{d\phi_{1x}}{ds} = \frac{dJ_{1x}}{ds} = 0. \quad (11.202)$$

In other words, at points where contour lines intersect, the dynamical variables are constants of the motion: such points are known as *fixed points*. In this case, the fixed points are unstable, since a small displacement of a particle from a fixed point will lead to it making a large excursion away from the fixed point. The intersecting contour lines are known as *separatrices*; they divide regions of phase space with qualitatively different behaviour.

A second example of a phase space portrait for a system close to a resonance is shown in Fig. 11.5. In this case, we have used the Hamiltonian:

$$H_1 = 1.6J_{1x} - 4J_{1x}^2 + J_{1x}^{3/2} \cos(3\phi_{1x}). \quad (11.203)$$

Note the introduction (compared to (11.200)) of the second-order term in  $J_{1x}$ : this term describes a variation in tune with betatron amplitude, but is not associated with any resonance (since it is independent of the angle variable  $\phi_{1x}$ ). Near  $J_{1x} = 0$ , the motion follows regular circles in phase space, as would be expected without any resonance terms. But as the amplitude increases, a chain of ‘islands’ appears, with each island centered on a local maximum in the Hamiltonian. The islands are bounded by separatrices. Since maxima (and, in general, minima) in the Hamiltonian satisfy the same conditions (11.201) as saddle points, it follows immediately that local maxima and minima also correspond to fixed points of the motion. However, since particles follow contours (lines of constant value of the Hamiltonian) in phase space, maxima and minima correspond to *stable*

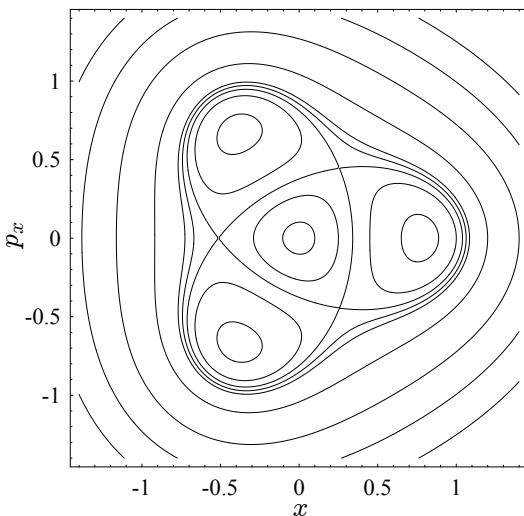


Fig. 11.5 Phase space portrait for the resonance Hamiltonian (11.203).

*fixed points*, where small displacements lead to oscillations around the fixed point. Saddle points correspond to *unstable fixed points*, where small displacements lead to particles moving into the vicinity of an adjacent fixed point, or out to infinity.

The introduction of the  $-4J_{1x}^2$  term in the Hamiltonian (11.203) leading to a tune shift with amplitude has improved the stability of the dynamics: there are no longer contour lines that extend out to infinity. Judicious use of a nonlinear term in this case has had a beneficial effect. Unfortunately, this result cannot be generalised. In (11.203), the tune shift with amplitude stabilises the dynamics because at large amplitudes the dominant term (which has a dependence on  $J_{1x}^2$ ) is independent of  $\phi_{1x}$ . Generally, in more realistic cases, there will always be higher-order resonance-driving terms, i.e. terms with a higher-order dependence on both the action and the angle variable. This means that, as well as carefully selecting the working point (the tunes at zero betatron action) when designing a storage ring lattice, attention also needs to be given to the tune shifts with amplitude that can lead to the tunes of particles approaching resonance lines in tune space even for quite modest betatron amplitudes.

At this stage, we can understand the significance of the main difference between the variables  $(x, p_x)$  and  $(x_1, p_{x1})$ . A particle at a fixed point of the Hamiltonian  $H_1$  has fixed values for the variables  $(x_1, p_{x1})$ , and hence a

fixed value for  $\phi_{1x}$ . Thus, from (11.199), the change in  $\phi_x$  over one complete turn is:

$$\Delta\phi_x = 2\pi \frac{n}{m}. \quad (11.204)$$

Since  $J_x = J_{1x}$ , if the particle is at a fixed point in  $(x_1, p_{x1})$ , the value of the action stays the same in both sets of variables,  $(x, p_x)$  and  $(x_1, p_{x1})$ . Therefore, in the variables  $(x, p_x)$ , after at most  $m$  complete turns of the storage ring, a particle at a fixed point in  $(x_1, p_{x1})$  will return to its original position in  $(x, p_x)$ . The net effect is that in  $(x, p_x)$ , the phase space portrait for a particular location in the ring may look the same as the phase space portrait in  $(x_1, p_{x1})$ ; but a particle close to a fixed point will jump from one fixed point to another on successive turns of the storage ring. If the fixed point is a stable fixed point, over many turns, as the particle moves from one fixed point to the next, it will map out loops around the fixed points. Each island in the phase space can be characterised by a width (the maximum distance between the separatrices enclosing the island) and an ‘island tune’ (the fraction of a complete loop around the island completed over one turn of the storage ring). The widths of the islands and the island tune can be determined from the Hamiltonian (11.203) [Ruth (1986a,b)].

The phase space portraits shown in Figs. 11.4 and 11.5 were derived from Hamiltonians approximating the motion in the presence of a single resonance: a key initial assumption was that only a single Fourier frequency in  $\phi_x$  dominated the terms in the Hamiltonian (11.187). If multiple resonances need to be considered, then the dynamics can become significantly more complicated. For example, it has been found that if resonance islands occur with widths larger than their separation, then chaotic motion occurs in the overlap region between them: this condition for the onset of chaos is known as the *Chirikov criterion* [Chirikov (1979)]. In practice, it is not unusual for more than one resonance to have a significant effect on the dynamics of particles in storage rings. This is most likely to occur near the intersection of resonance lines in the resonance diagram (Fig. 11.3) and working points close to such intersections are often best avoided.

Generally, beam properties in a storage ring are most easily controlled when the particle motion is not affected significantly by resonances. In that case, particle trajectories in phase space appear as smooth ellipses. If there is coupling, then the ellipses will only be evident if dynamical variables corresponding to the normal modes are used (see Section 5.2). In one degree of freedom, linear stable motion appears as a circle in normalised phase space (i.e. phase space with axes  $\sqrt{2J_I} \cos(\phi_I)$  and  $-\sqrt{2J_I} \sin(\phi_I)$ ).

In two degrees of freedom, linear stable motion is characterised by two amplitudes;  $J_I$  and  $J_{II}$ , and two corresponding phases (angle variables). Since the amplitudes are constants of the motion, a set of points with constant  $J_I$  and  $J_{II}$  is topologically equivalent to a torus. Geometrically, the shape of the torus remains the same as the system evolves; therefore, it is known as an *invariant torus*. Mathematically, the picture can be extended to any number of degrees of freedom, though it becomes difficult to visualise. Linear and nonlinear perturbations lead to distortions in the shapes of the tori, though, for small perturbations and sufficiently small amplitudes, invariant tori still exist. This is evident, for example, in Fig. 11.2 where we see that, although the value of the action variable changes as the system evolves, it can still be given as a function of the angle variable.

The statement that invariant tori persist for small perturbations is only true for systems that are sufficiently far from a resonance. The exact conditions under which invariant tori persist in the presence of nonlinear perturbations are given by the Kolmogorov–Arnold–Moser (KAM) theorem [Arnold (1989)]. Unfortunately, the conditions are not easily stated; nevertheless, the KAM theorem is an important result in nonlinear dynamics, simply because it shows the possibility that invariant tori can exist in the presence of perturbations. For storage rings, the significant result is that, as long as the tunes are not too close to a resonance and the nearby resonances are not driven too strongly, the particle motion, at least at small amplitude, can be approximated by a linear model of the lattice. Unfortunately, the maximum amplitude for a particle in a storage ring for which the motion is approximately linear may be very small: this can be important for injection into the storage ring (when initially, particles often follow trajectories with large amplitudes) and for operation of the storage ring (when scattering process can drive particles out to large amplitudes). Although nonlinear effects can be very complex, as we have seen, there is generally an amplitude limit beyond which the particle motion becomes unstable, and particles exceeding this limit are lost from the storage ring. This limit is known as the *dynamic aperture* [Irwin and Yan (1999); Wolski (2013)]. Nonlinearities in a storage ring can come from the sextupoles used to correct the chromaticity, from insertion devices used to generate intense beams of synchrotron radiation in light sources, and from systematic and random higher-order multipoles in the various magnets. The dynamic aperture depends on the strengths of the higher-order multipoles, the distribution of the multipoles around the ring, and the tunes (working point) of the lattice. For low emittance lattices (such as double-bend achromats) in particular,

which usually require very strong chromatic sextupoles, achieving a sufficient dynamic aperture to ensure efficient operation of a storage ring can be a considerable challenge.

Analytical formulae for dynamic aperture have been developed [Guignard and Hagel (1986); Gao (2000)]; however, dynamic aperture calculations are still performed using long-term particle tracking. In an electron storage ring it is usually sufficient to track particles for a few thousand turns, since radiation damping tends to reduce the amplitudes of particles with initially large amplitudes on that time scale; but in proton storage rings, it may be necessary to track particles for many tens or even hundreds of thousands of turns. It is important to use symplectic maps for tracking, since symplectic errors can lead to artificial changes in oscillation amplitudes that will give a misleading indication of the dynamic aperture [Kleiss *et al.* (1992)]. Tracking may be done with or without inclusion of synchrotron oscillations. Generally, some artificially large aperture (much larger than the real physical aperture of the beam pipe) is used to collimate particles whose oscillation amplitudes exceed some limit. The range of different conditions that can be applied in a dynamic aperture study is quite large, and it is important to specify all the conditions carefully since they can sometimes have a significant impact on the results.

The dynamic aperture characterises the amplitudes of particles that remain stable within a storage ring for some specified period of time, taking into account nonlinear effects. Creating a visualisation of the dynamic aperture can be challenging, since in general it will be a multi-dimensional region that usually does not have a smooth boundary (and may not even be a connected region). Often, an indication of the dynamic aperture is given by showing a set of points in Cartesian co-ordinate space representing the initial co-ordinates of particles at some point in the accelerator that have survived some number of turns. The dynamic aperture represented in this way will generally be a function of the energy deviation, or (if synchrotron motion is included in the tracking) of the synchrotron amplitude. The range of energy deviation over which the dynamic aperture remains above some specified size is often referred to as the *dynamic energy acceptance*. The maximum amplitude of stable synchrotron oscillations may be limited by effects (such as the finite rf voltage) other than nonlinearities in the magnetic lattice.

If the dynamic aperture and energy acceptance of a storage ring are too small, then it may be difficult to achieve injection, or the lifetime (limited by scattering effects, such as Touschek scattering) may be very short.

Optimising the dynamic aperture in a lattice design is often not an easy task. There are some guidelines that can be followed: for example, optimising the linear lattice to reduce the necessary strengths of the chromatic sextupoles is often an important first step. The working point should be chosen to avoid significant resonances. Arranging chromatic sextupoles in  $-I$  pairs can help suppress their nonlinear effects, as discussed in Section 11.1: for an example, see [Cai *et al.* (1995)]. Ensuring that the lattice has a high degree of periodicity can also help suppress resonance driving terms, as discussed earlier in this section. However, various design constraints can make it difficult to apply these (and similar) techniques. Despite the availability of some powerful mathematical tools for analysis of nonlinear dynamics, optimising a storage ring lattice to achieve a good dynamic aperture while satisfying a range of other design constraints remains something of an art form.

## 11.4 Normal Form Analysis

In Section 11.2 we discussed canonical perturbation theory, and showed how a generating function could be constructed for a canonical transformation to new dynamical variables in which the equations of motion have a simpler form. Specifically, if the original Hamiltonian contains a non-integrable term of order  $\epsilon$ , the Hamiltonian expressed in the new variables contains a non-integrable term of lowest order  $\epsilon^2$ . Generally, non-integrable terms are those that have some dependence on the angle variables: a Hamiltonian that depends only on the action variables is integrable, i.e. the equations of motion derived from such a Hamiltonian have exact analytical solutions. It is possible to extend the perturbation theory (showing how to construct an appropriate generating function) to higher order; however, the analysis becomes increasingly complicated.

An alternative approach to finding approximate solutions to non-integrable Hamiltonians is provided by *normal form analysis* [Dragt and Finn (1979); Neri and Dragt (1988)]. In this section, we shall discuss a version of normal form analysis in which Lie transformations are used to simplify a transfer map [Forest *et al.* (1989)]. We know from Section 9.1 that it is possible to construct a Lie transformation representing the transfer map for any given (time independent) Hamiltonian. Lie transformations representing the transfer maps for successive elements can be combined (by making use of the BCH formula) to give a Lie transformation representing the transfer map for some section of beam line. Unfortunately, application

of Lie transformations for particle tracking is not completely straightforward. Lie transformations can be expanded as power series to some order in the dynamical variables; however, except for special cases, the exact representation will involve a power series to infinite order. Truncating the power series at finite order results in the loss of desirable properties of the map (such as symplecticity).

Normal form analysis is a technique for constructing a similarity transformation that simplifies a given Lie transformation. The goal is to find a similarity transformation that results in a (new) Lie transformation that has an exact representation as a power series of finite order in the dynamical variables. In other words, the dynamical system represented by the new Lie transformation has an exact solution: applying the inverse similarity transformation to the solution then provides a solution to the original problem. Normal form analysis is analogous to the perturbation theory developed in Section 11.2. Presented with a problem for which we cannot write an exact solution, we construct a transformation that casts the problem in a form that, with some approximations, can be solved exactly. Making the inverse transformation then provides us with an approximate solution to the original problem. Normal form methods using Lie transformations are discussed, for example, in [Forest (1998); Yan (1999); Dragt (2020)].

Suppose that  $\mathcal{M}$  is a transfer map for a section of an accelerator beam line, expressed in the form of a Lie transformation:

$$\mathcal{M} = e^{\cdot m(\phi, \mathbf{J}) \cdot}, \quad (11.205)$$

where the generator  $m(\phi, \mathbf{J})$  is a function of the dynamical (action–angle) variables. The problem, in general, is to find an explicit expression for the transfer map that is obtained by applying  $\mathcal{M}$  to the dynamical variables:

$$\phi_1(\phi, \mathbf{J}) = \mathcal{M}\phi, \quad (11.206)$$

$$\mathbf{J}_1(\phi, \mathbf{J}) = \mathcal{M}\mathbf{J}. \quad (11.207)$$

Normal form analysis addresses the problem by showing how to construct a Lie transformation  $\mathcal{N}$  (*a normalising transformation*) such that:

$$\bar{\mathcal{M}} = \mathcal{N}\mathcal{M}\mathcal{N}^{-1} = e^{\cdot \bar{m}(\mathbf{J}) \cdot}, \quad (11.208)$$

where  $\bar{m}(\mathbf{J})$  is a function of the action variables only. We define the functions  $\bar{\phi}(\phi, \mathbf{J})$  and  $\bar{\mathbf{J}}(\phi, \mathbf{J})$  as the results of applying the normalising transformation  $\mathcal{N}$  to the dynamical variables  $\phi$  and  $\mathbf{J}$ :

$$\bar{\phi}(\phi, \mathbf{J}) = \mathcal{N}\phi, \quad (11.209)$$

$$\bar{\mathbf{J}}(\phi, \mathbf{J}) = \mathcal{N}\mathbf{J}. \quad (11.210)$$

Now, since  $\bar{m}$  is a function of the action variables only, it follows that:

$$\bar{\mathcal{M}}\bar{\phi}(\phi, \mathbf{J}) = \bar{\phi}(\bar{\mathcal{M}}\phi, \bar{\mathcal{M}}\mathbf{J}) = \bar{\phi}(\phi + \Delta\phi, \mathbf{J}), \quad (11.211)$$

$$\bar{\mathcal{M}}\bar{\mathbf{J}}(\phi, \mathbf{J}) = \bar{\mathbf{J}}(\bar{\mathcal{M}}\phi, \bar{\mathcal{M}}\mathbf{J}) = \bar{\mathbf{J}}(\phi + \Delta\phi, \mathbf{J}), \quad (11.212)$$

where:

$$\Delta\phi = -\frac{d\bar{m}}{d\mathbf{J}}. \quad (11.213)$$

Then, using:

$$\bar{\mathcal{M}}\bar{\phi}(\phi, \mathbf{J}) = \mathcal{N}\mathcal{M}\mathcal{N}^{-1}\mathcal{N}\phi = \mathcal{N}\mathcal{M}\phi, \quad (11.214)$$

$$\bar{\mathcal{M}}\bar{\mathbf{J}}(\phi, \mathbf{J}) = \mathcal{N}\mathcal{M}\mathcal{N}^{-1}\mathcal{N}\mathbf{J} = \mathcal{N}\mathcal{M}\mathbf{J}, \quad (11.215)$$

we can write:

$$\mathcal{M}\phi = \mathcal{N}^{-1}\bar{\mathcal{M}}\bar{\phi}(\phi, \mathbf{J}) = \bar{\phi}(\mathcal{N}^{-1}\phi|_{\phi+\Delta\phi}, \mathcal{N}^{-1}\mathbf{J}), \quad (11.216)$$

$$\mathcal{M}\mathbf{J} = \mathcal{N}^{-1}\bar{\mathcal{M}}\bar{\mathbf{J}}(\phi, \mathbf{J}) = \bar{\mathbf{J}}(\mathcal{N}^{-1}\phi|_{\phi+\Delta\phi}, \mathcal{N}^{-1}\mathbf{J}), \quad (11.217)$$

where  $\mathcal{N}^{-1}$  is the inverse transformation of  $\mathcal{N}$ . Note that the expression  $\mathcal{N}^{-1}\phi|_{\phi+\Delta\phi}$  represents the result of applying the Lie transformation  $\mathcal{N}^{-1}$  to the angle variables  $\phi$ , and then substituting  $\phi + \Delta\phi$  for  $\phi$ .

At this stage, it appears that all that has been achieved is to change the problem from applying the transformation  $\mathcal{M}$  to applying the Lie transformation  $\mathcal{N}$ . There seems to be little benefit in this. In general, the transformation  $\mathcal{N}$  that normalises  $\mathcal{M}$  will not be exactly solvable; that is, it will not be possible to express  $\mathcal{N}\phi$  and  $\mathcal{N}\mathbf{J}$  as analytical expressions in closed form. However, the benefit of the normal form analysis becomes apparent if we consider what happens if the transformation  $\mathcal{M}$  is applied multiple times, representing repeated passes through a periodic beam line (for example, multiple turns of a storage ring). The transfer map for  $n$  passes can be written:

$$\mathcal{M}^n\phi = \bar{\phi}(\mathcal{N}^{-1}\phi|_{\phi+n\Delta\phi}, \mathcal{N}^{-1}\mathbf{J}), \quad (11.218)$$

$$\mathcal{M}^n\mathbf{J} = \bar{\mathbf{J}}(\mathcal{N}^{-1}\phi|_{\phi+n\Delta\phi}, \mathcal{N}^{-1}\mathbf{J}). \quad (11.219)$$

The transformation  $\mathcal{N}$  and its inverse  $\mathcal{N}^{-1}$  still only need to be each applied once, while the  $n$  passes are effectively represented by an increase  $n\Delta\phi$  in the angle variable (where the increase in the angle variable from a single pass through the beam line is  $\Delta\phi$ ). This suggests that it should be possible to use non-symplectic approximations for  $\mathcal{N}$  and  $\mathcal{N}^{-1}$  when carrying out multi-pass tracking, without accumulating a symplectic error on every pass.

Normal form analysis provides a convenient way to construct a phase space portrait for a map  $\mathcal{M}$ . To see how this works, first observe that from (11.208):

$$\mathcal{M}\mathcal{N}^{-1}\mathbf{J} = \mathcal{N}^{-1}\bar{\mathcal{M}}\mathbf{J} = \mathcal{N}^{-1}\mathbf{J}, \quad (11.220)$$

where the final step follows from the fact that the generator of  $\bar{\mathcal{M}}$  is a function of  $\mathbf{J}$  only, i.e. is independent of the angle variables  $\phi$ . Hence, the quantities  $\mathcal{N}^{-1}\mathbf{J}$  are invariants of the map  $\mathcal{M}$ . Suppose that:

$$\mathcal{N}^{-1}\mathbf{J}|_{\mathbf{J}=\mathbf{J}_0, \phi=\phi_0} = \mathbf{I}_0, \quad (11.221)$$

for some values  $\mathbf{J}_0$  of the action variables and values  $\phi_0$  of the angle variables.  $\mathbf{I}_0$  is a constant vector. It then follows that the equation:

$$\mathbf{J} = \mathcal{N}\mathbf{J}|_{\mathbf{J}=\mathbf{I}_0, \phi} \quad (11.222)$$

gives the actions  $\mathbf{J}$  as functions of the angle variables  $\phi$ . Phase space portraits can be obtained by plotting  $\mathcal{N}\mathbf{J}$  as functions of the angle variables  $\phi$  for different values of  $\mathbf{I}_0$ .

Normal form analysis can provide information about the dynamics that may not be immediately obvious from the transfer map. To illustrate this, consider the case of a linear transfer map for a periodic beam line in one degree of freedom. It was shown in Section 4.2 that, working in Cartesian variables, the transfer map for one period may be written in matrix form as (4.20):

$$R = \begin{pmatrix} \cos(\mu_x) + \alpha_x \sin(\mu_x) & \beta_x \sin(\mu_x) \\ -\gamma_x \sin(\mu_x) & \cos(\mu_x) - \alpha_x \sin(\mu_x) \end{pmatrix}, \quad (11.223)$$

where  $\mu_x$  is the phase advance, and  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  are the Courant–Snyder parameters. If we define the matrix  $N_x$  by (4.62):

$$N_x = \begin{pmatrix} \frac{1}{\sqrt{\beta_x}} & 0 \\ \frac{\alpha_x}{\sqrt{\beta_x}} & \sqrt{\beta_x} \end{pmatrix}, \quad (11.224)$$

then:

$$N_x R N_x^{-1} = \begin{pmatrix} \cos(\mu_x) & \sin(\mu_x) \\ -\sin(\mu_x) & \cos(\mu_x) \end{pmatrix}. \quad (11.225)$$

Therefore, the matrix  $N_x$  represents a normalising transformation for the transfer matrix  $R$ : applied as a similarity transformation,  $N_x$  transforms  $R$  into a pure rotation. Applied to the dynamical variables:

$$N_x \begin{pmatrix} x \\ p_x \end{pmatrix} = \begin{pmatrix} x/\sqrt{\beta_x} \\ \alpha_x x/\sqrt{\beta_x} + \sqrt{\beta_x} p_x \end{pmatrix} = \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix}. \quad (11.226)$$

The effect on the normalised variables of transport through one periodic section is:

$$\begin{pmatrix} \cos(\mu_x) & \sin(\mu_x) \\ -\sin(\mu_x) & \cos(\mu_x) \end{pmatrix} \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix} = \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x + \mu_x) \\ -\sqrt{2J_x} \sin(\phi_x + \mu_x) \end{pmatrix}. \quad (11.227)$$

The effect on the normalised variables of transport through  $n$  periodic sections is:

$$\begin{pmatrix} \cos(\mu_x) & \sin(\mu_x) \\ -\sin(\mu_x) & \cos(\mu_x) \end{pmatrix}^n \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix} = \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x + n\mu_x) \\ -\sqrt{2J_x} \sin(\phi_x + n\mu_x) \end{pmatrix}. \quad (11.228)$$

The normalising transformation puts the transfer map into the form of a pure rotation; the rotation angle (which is found by inspection of the normal form of the transfer map) is the phase advance corresponding to a single application of the transfer map. The Courant–Snyder parameters, which describe the shape of the ellipse in  $(x, p_x)$  phase space mapped out by repeated applications of the transfer map, can be obtained from the normalising transformation.

Now let us consider how normal form analysis can be applied to a nonlinear map:

$$\mathcal{M} = e^{:m(\phi, \mathbf{J}):}. \quad (11.229)$$

The goal is to construct a transformation  $\mathcal{F}$  that removes the dependence of the generator  $m(\phi, \mathbf{J})$  on the angle variables in the generator of the nonlinear map, to some order. If the dependence on the angle variables is removed completely, then:

$$\bar{\mathcal{M}} = \mathcal{F} \mathcal{M} \mathcal{F}^{-1} = e^{:\bar{m}(\mathbf{J}):}. \quad (11.230)$$

The generator  $\bar{m}(\mathbf{J})$  will contain information on the phase advance (including its dependence on the action variables,  $\mathbf{J}$ ); the normalising transformation  $\mathcal{F}$  will contain the remaining information on the dynamics.

We assume that the map is represented in the form of a Dragt–Finn factorisation [Dragt and Finn (1976)]:

$$\mathcal{M} = \mathcal{R} e^{:m_3:} e^{:m_4:}. \quad (11.231)$$

The transformation  $\mathcal{R}$  contains the linear part of the map:

$$\mathcal{R} = e^{:m_2:}. \quad (11.232)$$

Usually, the map will be initially constructed using Cartesian variables, in which a phase space vector  $\vec{x}$  has components:

$$\vec{x} = \begin{pmatrix} x \\ p_x \\ y \\ p_y \\ z \\ \delta \end{pmatrix}. \quad (11.233)$$

In that case, the generator of the linear part of the map can be represented as:

$$m_2 = \vec{x}^T M_2 \vec{x}, \quad (11.234)$$

where  $M_2$  is a  $6 \times 6$  matrix (in three degrees of freedom). It is convenient (though not strictly necessary) to work in action-angle variables  $(\phi, \mathbf{J})$ . The relationship between the Cartesian variables and the action-angle variables is:

$$\vec{J} = N_x \vec{x}, \quad (11.235)$$

where, in three degrees of freedom:

$$\vec{J} = \begin{pmatrix} \sqrt{2J_1} \cos(\phi_1) \\ -\sqrt{2J_1} \sin(\phi_1) \\ \sqrt{2J_{II}} \cos(\phi_{II}) \\ -\sqrt{2J_{II}} \sin(\phi_{II}) \\ \sqrt{2J_{III}} \cos(\phi_{III}) \\ -\sqrt{2J_{III}} \sin(\phi_{III}) \end{pmatrix}, \quad (11.236)$$

and  $N$  is a matrix constructed from the eigenvectors of the transfer matrix as described in Section 5.2 (see in particular equation (5.89)).  $N$  provides a linear normalising transformation: in the variables defined by the components of  $\vec{J}$ , the motion of a particle on repeated application of the linear transfer map describes a circle in the phase space corresponding to each degree of freedom. In three degrees of freedom, the vectors  $\phi$  and  $\mathbf{J}$  have components  $(\phi_1, \phi_{II}, \phi_{III})$  and  $(J_1, J_{II}, J_{III})$  respectively. The linear part of the transfer map,  $\mathcal{R}$ , transforms  $\phi$  and  $\mathbf{J}$  as follows:

$$\mathcal{R}\mathbf{J} = \mathbf{J}, \quad (11.237)$$

$$\mathcal{R}\phi = \phi + \boldsymbol{\mu}, \quad (11.238)$$

where  $\boldsymbol{\mu} = (\mu_1, \mu_{II}, \mu_{III})$  is a vector whose components correspond to the phase advance in each degree of freedom. The generator  $m_2$  of the linear part of the transfer map can be written in action-angle variables as:

$$m_2 = - \sum_k \mu_k J_k, \quad (11.239)$$

where the index  $k$  ranges over all the degrees of freedom.

If the generators  $m_3$ ,  $m_4$ , etc. of the nonlinear part of the transfer map are known in Cartesian variables, then they can be expressed in action-angle variables using the linear normalising transformation (11.236). The generator  $m_3$  contains terms of order  $J_k^{3/2}$ , and  $m_4$  contains terms of order  $J_k^2$ . In general, there will be factors in the Dragt–Finn factorisation of  $\mathcal{M}$  that have generators containing terms of order  $J_k^{5/2}$ ,  $J_k^3$  and so on: we shall assume that these factors do not have a significant effect on the dynamics, and can be replaced by the identity transformation.

Now consider a transformation:

$$\mathcal{F} = e^{:f:}. \quad (11.240)$$

The transfer map  $\mathcal{M}$  is transformed into:

$$\bar{\mathcal{M}} = \mathcal{F}\mathcal{M}\mathcal{F}^{-1} = e^{:f:}\mathcal{R}e^{:m_3:}e^{:m_4:}e^{-:f:}. \quad (11.241)$$

Inserting identity transformations  $\mathcal{I} = \mathcal{R}\mathcal{R}^{-1}$  and  $\mathcal{I} = e^{-:f:}e^{:f:}$ , and applying the formula (9.19) for a similarity transformation, we obtain:

$$\bar{\mathcal{M}} = \mathcal{R}\mathcal{R}^{-1}e^{:f:}\mathcal{R}e^{:m_3:}e^{-:f:}e^{:f:}e^{:m_4:}e^{-:f:}, \quad (11.242)$$

$$= \mathcal{R}e^{:\mathcal{R}^{-1}f:}e^{:m_3:}e^{-:f:}e^{:f:}e^{:m_4:}. \quad (11.243)$$

Now suppose that  $f$  contains terms of order  $J_k^{3/2}$ . Then, applying the BCH and Zassenhaus formulae (10.34) and (10.38), we obtain:

$$\bar{\mathcal{M}} = \mathcal{R}e^{:\bar{m}_3:}e^{:m'_4:}, \quad (11.244)$$

where:

$$\bar{m}_3 = \mathcal{R}^{-1}f + m_3 - f, \quad (11.245)$$

and  $m'_4$  is a function containing terms of order  $J_k^2$ : we shall not be concerned with the exact form of  $m'_4$ .

Rearranging (11.245) gives:

$$f = (\mathcal{I} - \mathcal{R}^{-1})^{-1}(m_3 - \bar{m}_3). \quad (11.246)$$

Equation (11.246) provides a formula for the generator  $f$  of a Lie transformation that puts the generator  $m_3$  (i.e. the generator of order  $J_k^{3/2}$  in the Dragt–Finn factorisation of  $\mathcal{M}$ ) into some other form  $\bar{m}_3$ . There is some freedom in choosing the form of the generator  $\bar{m}_3$ , though there are also some constraints that we shall mention in a moment. Since  $m_3$  must be periodic in the angle variables  $\phi$ , we can write:

$$m_3 = \sum_{\mathbf{p}=-\infty}^{\infty} m_{3,\mathbf{p}}(\mathbf{J}) e^{i\mathbf{p}\cdot\phi}. \quad (11.247)$$

The summation in (11.247) is actually a triple summation over the components of the vector  $\mathbf{p}$ . Now, let us choose  $\bar{m}_3$  so that it is a function of the action variables only. In particular:

$$\bar{m}_3 = m_{3,0}(\mathbf{J}). \quad (11.248)$$

Then it follows from (11.246) that:

$$f = \sum_{\mathbf{p} \neq 0} \frac{m_{3,\mathbf{p}}(\mathbf{J}) e^{i\mathbf{p}\cdot\phi}}{1 - e^{-i\mathbf{p}\cdot\mu}}. \quad (11.249)$$

That (11.249) does indeed provide a solution for  $f$  can easily be checked by substituting for  $f$  in (11.246), and using the fact that:

$$\mathcal{R}^{-1}g(\phi, \mathbf{J}) = g(\phi - \mu, \mathbf{J}), \quad (11.250)$$

for any function  $g$ . The constraints on the choice of  $\bar{m}_3$  are apparent from (11.249). If  $m_3$  contains driving terms for a resonance close to the working point, so that:

$$\mathbf{p} \cdot \mu \approx 2\pi \times \text{integer}, \quad (11.251)$$

then applying (11.249) produces terms in  $f$  with large coefficients. In this case, it will be difficult to apply the map  $\mathcal{F}$  (with generator  $f$ ) with good accuracy. In general, resonance driving terms cannot be removed from the map  $\mathcal{M}$  if the working point (determined by the linear part of the map) is close to the corresponding resonance. Clearly, terms in  $m_3$  that depend only on the action variables also cannot be removed: hence our choice (11.248) for  $\bar{m}_3$ .

Using  $\mathcal{F} = e^{\cdot f \cdot}$  with  $f$  given by (11.249), the map  $\bar{\mathcal{M}}$  (11.241) is:

$$\bar{\mathcal{M}} = \mathcal{F} \mathcal{M} \mathcal{F}^{-1} = \mathcal{R} e^{\cdot \bar{m}_3(\mathbf{J}) \cdot} e^{\cdot m'_4 \cdot}. \quad (11.252)$$

To lowest order in the nonlinear terms, the map  $\bar{\mathcal{M}}$  contains only a tune shift with amplitude, represented by a Lie transformation with generator that is a function of the action variables  $\mathbf{J}$  (and has no dependence on the angle variables  $\phi$ ). As already mentioned, it is not possible to remove the terms independent of the angle variables. However, it is clear that  $\mathcal{F}$  provides a normalising transformation for the map  $\mathcal{M}$ , to order  $J_k^{3/2}$  in the generator of the map. Neglecting terms of higher order, the normalised map is:

$$\bar{\mathcal{M}} \approx \mathcal{R} e^{\cdot \bar{m}_3(\mathbf{J}) \cdot}, \quad (11.253)$$

and the transformation of the action–angle variables (under transport through  $n$  periodic sections of the beam line) can be written:

$$\mathcal{M}^n \mathbf{J} \approx \mathcal{F}^{-1} \mathcal{R}^n e^{:n\bar{m}_3(\mathbf{J}):} \mathcal{F} \mathbf{J}, \quad (11.254)$$

$$\mathcal{M}^n \phi \approx \mathcal{F}^{-1} \mathcal{R}^n e^{:n\bar{m}_3(\mathbf{J}):} \mathcal{F} \phi. \quad (11.255)$$

The benefit of writing the map in this way is that the (nonlinear) rotation is very easy to apply:

$$\mathcal{R}^n e^{:n\bar{m}_3(\mathbf{J}):} \mathbf{J} = \mathbf{J}, \quad (11.256)$$

$$\mathcal{R}^n e^{:n\bar{m}_3(\mathbf{J}):} \phi = \phi + n\mu + n \frac{\partial \bar{m}_3}{\partial \mathbf{J}}. \quad (11.257)$$

Tracking a particle for  $n$  periodic sections of a beam line then involves one application of the map  $\mathcal{F}^{-1}$ , a phase space rotation (11.257), and a final application of  $\mathcal{F}$ .

The above procedure normalises the map to lowest order in the nonlinearity: this corresponds to first-order perturbation theory. It is possible to carry the normalisation procedure to higher order [Forest (1998)].

To illustrate the technique of normal form analysis, we shall use a conventional example: a single sextupole in a storage ring that (apart from the sextupole) consists of linear components, and has no coupling. We shall construct the horizontal phase space portrait. In one degree of freedom, the single-turn transfer map at a location immediately following the sextupole can be represented by the Lie transformation:

$$\mathcal{M} = \mathcal{R} e^{:m_3:}, \quad (11.258)$$

where:

$$\mathcal{R} = e^{:-\mu_x J_x:}, \quad (11.259)$$

and:

$$m_3 = -\frac{1}{6} k_2 L x^3 = -\frac{1}{6} k_2 (2\beta_x J_x)^{3/2} \cos^3(\phi_x). \quad (11.260)$$

Here,  $\mu_x$  is the horizontal phase advance over one turn of the ring (neglecting the sextupole), and  $\beta_x$  is the horizontal beta function at the location of the sextupole. It is assumed that the sextupole can be treated in the limit that its length approaches zero, but with fixed integrated strength  $k_2 L$ . Since there is no linear coupling, the normal mode action–angle variables are  $J_x$  and  $\phi_x$ ; in the more general case when coupling is present, the substitution for  $x$  in (11.260) will be more complicated, and will involve elements of the linear normalising transformation matrix  $N$  in place of the Courant–Snyder parameter  $\beta_x$ .

Applying (11.249), the generator of the Lie transformation that removes the sextupole to first order in the sextupole strength is:

$$f = \frac{k_2 L}{16} \frac{(2\beta_x J_x)^{3/2}}{\sin(3\mu_x/2)} \left( \sin(\mu_x/2 - \phi_x) - \sin(\mu_x/2 + \phi_x) - \sin(3\mu_x/2 + \phi_x) - \frac{1}{3} \sin(3\mu_x/2 + 3\phi_x) \right). \quad (11.261)$$

From (11.247) and (11.248), it is found that:

$$\bar{m}_3 = 0, \quad (11.262)$$

and hence the normalised map, to first order in the sextupole strength, is:

$$\bar{\mathcal{M}} \approx e^{:f:} \mathcal{M} e^{-:f:} \approx \mathcal{R}. \quad (11.263)$$

Using (11.221), an (approximate) invariant of the map is:

$$e^{:f:} J_x = J_x + \frac{k_2 L}{8} (2\beta_x J_x)^{3/2} \frac{(\cos(3\mu_x/2 + 2\phi_x) + \cos(\mu_x/2))}{\sin(3\mu_x/2)} + O(J_x^2). \quad (11.264)$$

If  $I_0$  is a constant with value equal to the invariant, then using (11.222), trajectories in phase space (i.e. lines mapped out by particles on successive turns in the storage ring) are given by:

$$\begin{aligned} J_x &\approx e^{:f:} J_x \Big|_{J_x=I_0} \\ &\approx I_0 - \frac{k_2 L}{8} (2\beta_x I_0)^{3/2} \frac{(\cos(3\mu_x/2 + 2\phi_x) + \cos(\mu_x/2))}{\sin(3\mu_x/2)} + O(I_0^2). \end{aligned} \quad (11.265)$$

One of the main reasons for choosing this particular example of a (thin) sextupole, in an otherwise linear storage ring in one degree of freedom, is that it is straightforward to produce a phase space portrait simply by applying the map (11.258), allowing a direct comparison with the results of the normal form analysis. Figure 11.6 shows such a comparison. The horizontal phase space is computed for different values of the phase advance  $\mu_x$ ; the points show the result of numerical tracking by direct application of the map (11.258), while the lines show the trajectories obtained by plotting (11.265) for different values of the invariant  $I_0$ . For all cases, we have used  $\beta_x = 1$  m, and  $k_2 L = 1$  m<sup>-2</sup>. It can be seen that the normal form analysis agrees with the results of the numerical tracking reasonably well, even close to the third-integer resonance where there is (as expected) a strong nonlinear distortion of the phase space.

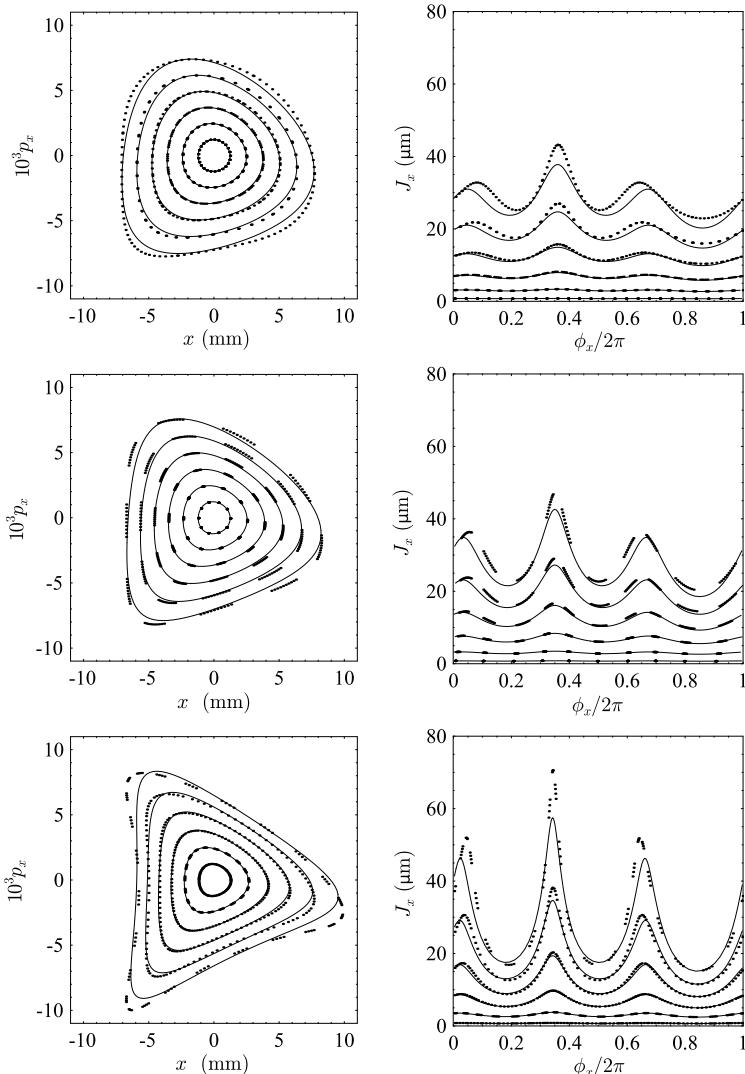


Fig. 11.6 Horizontal phase space portraits for a storage ring with a single sextupole, in Cartesian variables (left) and action-angle variables (right). The storage ring lattice is, apart from the sextupole, linear and without coupling. The fractional part of the horizontal tune  $\nu_x$  is 0.28 (top), 0.30 (middle) and 0.315 (bottom). The points show the result of numerical tracking using a direct application of the map (11.258), and the lines show the contours obtained by plotting (11.265) for different values of the invariant  $I_0$  from 0.75  $\mu\text{m}$  to 27  $\mu\text{m}$ . In all cases,  $\beta_x = 1 \text{ m}$ ,  $\alpha_x = 0$  and  $k_2 L = 100 \text{ m}^{-2}$ .

## 11.5 A Numerical Method: Frequency Map Analysis

In a storage ring, the betatron and synchrotron tunes are important characteristics of the dynamics. If the dynamics can be treated as being essentially linear (for example, in the limit of small betatron and synchrotron amplitudes) then the tunes are constant. However, in a lattice constructed from only dipoles, quadrupoles and rf cavities, the betatron tunes have a strong dependence on particle energy: this dependence is characterised by the chromaticity. Reducing the chromaticity, which is often necessary for operational reasons, requires the introduction of nonlinear elements such as sextupoles, and this leads to a dependence of the tunes on the betatron amplitudes. Furthermore, in the presence of nonlinearities, the amplitudes (actions) are not constant, so the tunes can vary over some number of turns. If the nonlinearities are not too strong, so that there remain invariant tori in the phase space (as allowed by the KAM theorem), then there may be only small variations in the values of the action variables, accompanied by small changes in the tunes. However, in general, there will be initial conditions for which the action variables are driven to very large values by the nonlinearities, resulting in significant changes in tune. This can occur if the initial values of the action variables are large, so that the starting point in phase space is beyond any of the invariant tori. However, for systems with more than one degree of freedom, the motion can be unstable (in the sense that the action variables reach arbitrarily large values) even if the initial values of the action variables are small compared to the values on nearby invariant tori: this phenomenon is known as *Arnold diffusion* [Arnold (1964)]. Since nonlinearities usually lead to a dependence of the tunes on the actions, it is possible to obtain some indication of the stability of a system over a range of initial conditions by studying the behaviour of the tunes associated with those initial conditions.

Techniques such as perturbation theory and normal form analysis can be useful in understanding properties of the dynamics in the case that nonlinearities in the lattice are not too strong. However, it is quite common for accelerator storage rings to operate in a parameter regime where the approximations required for these techniques start to break down for even modest betatron and synchrotron amplitudes. Aspects of the dynamical behaviour can still be understood, however, by observing the behaviour of the tunes either in simulation or experiment. Tracking results can be analysed numerically to determine the tunes as a function of betatron amplitude and energy deviation, and to estimate the rate of change of the tunes with

respect to time. In principle, all that is required is to apply a discrete Fourier transform to a set of tracking data. In practice, however, the number of turns of data available may be limited. For example, experimental data may be collected in a storage ring using a fast pulsed magnet or stripline kicker to excite betatron oscillations of a bunch; however, the oscillations can start to decay, sometimes quite rapidly, because of decoherence or synchrotron radiation effects, or particle losses. In tracking simulations, if the beam behaviour over a wide range of horizontal and vertical amplitudes needs to be investigated, then tracking over a large number of turns for each amplitude can take considerable computing power. Regardless of the reason, where the available number of turns of data is restricted, then a discrete Fourier transform has limited accuracy that limits the value of the results. Then, an alternative technique such as *numerical analysis of the fundamental frequencies* [Laskar (1990, 1993); Dumas and Laskar (1993)] may be of benefit.

In a conventional discrete Fourier transform (DFT), a set of  $N_t$  turns of tracking data is decomposed into the amplitudes of  $N_t$  Fourier modes. The mode with frequency closest to the actual tune in the storage ring will have the largest amplitude. Therefore, simply performing a DFT and identifying the mode with the largest amplitude will give the tune with precision  $1/N_t$ . Ideally, the data to which the DFT is applied should consist of the normalised phase space co-ordinates. For example, in one degree of freedom, the normalised phase space co-ordinates are given by:

$$\begin{pmatrix} x_{\text{norm}} \\ p_{x,\text{norm}} \end{pmatrix} = \begin{pmatrix} 1/\sqrt{\beta_x} & 0 \\ \alpha_x/\sqrt{\beta_x} & \sqrt{\beta_x} \end{pmatrix} \begin{pmatrix} x \\ p_x \end{pmatrix} = \begin{pmatrix} \sqrt{2J_x} \cos(\phi_x) \\ -\sqrt{2J_x} \sin(\phi_x) \end{pmatrix}. \quad (11.266)$$

If the values for  $x$  and  $p_x$  are obtained from a tracking simulation, then both the co-ordinate  $x$  and the momentum  $p_x$  will be available directly. However, experimentally, beam position monitors (BPMs) generally provide only the beam co-ordinates and not the momenta. It may be possible to estimate the value of  $p_x$  on each turn from readings on a pair of BPMs: for example, if two BPMs are separated by a drift space of length  $L$ , then:

$$p_x \approx \frac{dx}{ds} = \frac{x_B - x_A}{L}, \quad (11.267)$$

where  $x_A$  and  $x_B$  are the readings on the first and second BPM respectively. If appropriate pairs of BPMs are not available, then it is still possible to apply a DFT to the co-ordinate data alone. In that case, it will not be possible to determine from the results whether the fractional part of the tune lies above or below a half-integer; however, if the magnets in a storage

ring have strengths close to their design values, then the tunes may be estimated from a model of the lattice.

The normalised co-ordinate and momentum can be combined to form a complex quantity:

$$w = x_{\text{norm}} - ip_{x,\text{norm}} = \sqrt{2J_x} e^{i\phi_x}. \quad (11.268)$$

If the initial phase is  $\phi_{x0}$  and the horizontal tune is  $\nu_x$ , then on the  $n$ th turn (in the absence of nonlinearities):

$$w_n = \sqrt{2J_x} e^{i\phi_{x0}} e^{2\pi i n \nu_x}. \quad (11.269)$$

The discrete Fourier transform  $\bar{w}_m$  of  $w_n$  may be defined:

$$\bar{w}_m = \frac{1}{N_t} \sum_{n=0}^{N_t-1} e^{-2\pi i m n / N_t} w_n, \quad (11.270)$$

where  $m$  is an integer in the range 0 to  $N_t - 1$ . Using (11.269) for  $w_n$  gives:

$$\bar{w}_m = \sqrt{2J_x} e^{i\phi_{x0}} \frac{1 - e^{2\pi i \Delta}}{1 - e^{2\pi i \Delta / N_t}}, \quad (11.271)$$

where:

$$\Delta = N_t \text{frac}(\nu_x) - m. \quad (11.272)$$

$\text{frac}(\nu_x)$  is the fractional part of the horizontal tune. The absolute value of  $\bar{w}_m$  has a maximum for  $\Delta = 0$ :

$$|\bar{w}_m(\Delta = 0)| = \sqrt{2J_x}. \quad (11.273)$$

If  $m$  is restricted to be an integer, then unless  $N_t \text{frac}(\nu_x)$  happens to be an integer,  $\Delta$  given by (11.272) will not be exactly zero. However, if we allow  $m$  to take any real value (not just integer values) then we can search numerically for the value of  $m$  that maximises  $\bar{w}_m$  given by (11.271). If this value of  $m$  is denoted  $\hat{m}$ , then the fractional part of the tune can be estimated from:

$$\text{frac}(\nu_x) = \frac{\hat{m}}{N_t}. \quad (11.274)$$

Figure 11.7 shows a plot of  $|\bar{w}_m|/\sqrt{2J_x}$  for  $N_t = 100$  and  $\nu_x = 0.2134$ , for a range  $m/N_t = 0.1$  to  $m/N_t = 0.3$ . The continuous line shows the results for a continuous range of  $m$ , while the points show the results for integer values of  $m$ . The peak of the curve gives a much more accurate estimate for the tune than the position of the highest of the points corresponding to integer values of  $m$ . Essentially, finding the value of  $m$  (not restricted to an integer) that maximises  $|\bar{w}_m|$  involves searching for a frequency that

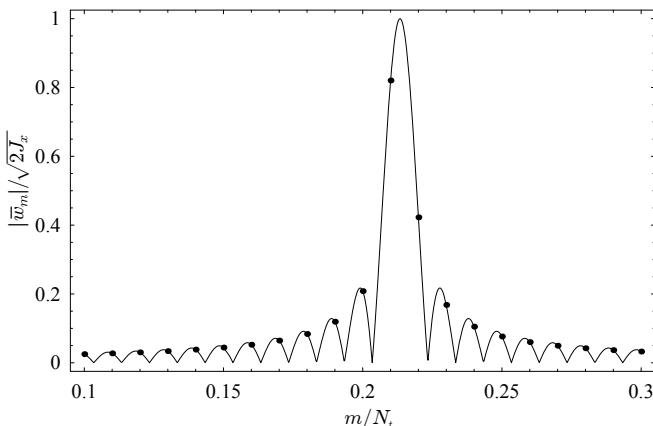


Fig. 11.7 A plot of  $|\bar{w}_m|/\sqrt{2J_x}$  given by equation (11.271), for  $N_t = 100$  and  $\nu_x = 0.2134$ . The line shows the results using a continuous range of  $m$ , while the points show the results for integer values of  $m$ .

gives the maximum overlap between the measured signal and a signal at the given frequency. This process is the basis of a technique known as the *numerical analysis of the fundamental frequencies* (NAFF). Briefly, the dominant frequency in a set of data  $w_n$  is identified by searching for the maximum of  $|\bar{w}_m|$ ; a signal of the appropriate frequency and amplitude is then subtracted from the original data. The process is repeated to find the next largest frequency component, which is again subtracted from the data. This can be repeated any number of times: the result is a set of relatively few frequencies and amplitudes (far fewer than the number of values in the data set  $w_n$ ) that can be used to reconstruct the original data with good accuracy.

In the presence of focusing errors, coupling effects and nonlinearities, the turn-by-turn data obtained by tracking particles in a storage ring will contain a number of frequencies in addition to the ‘principal’ frequencies corresponding to the tunes. For example, some spread in frequency will be introduced if the values of the Courant–Snyder parameters used to normalise the turn-by-turn data are not exactly correct. Betatron coupling will lead to the presence of a signal at the vertical tune appearing in the horizontal data, and vice versa. The presence of different frequency signals will affect the accuracy with which the frequency of the main component can be identified: to minimise the effects of other frequency components, a filter can be applied to the data before carrying out a numerical search for

the peak in  $|\bar{w}_m|$ . For example, a Hanning filter of the form:

$$\chi_p(n) = \frac{2^p(p!)^2}{(2p)!} \left(1 + \cos\left(\frac{2\pi n}{N_t}\right)\right)^p, \quad (11.275)$$

where  $N_t$  is the number of data points, can be used to calculate the quantity  $\bar{w}_m^{(p)}$ :

$$\bar{w}_m^{(p)} = \frac{1}{N_t} \sum_{n=0}^{N_t-1} e^{-2\pi i mn/N_t} w_n \chi_p(n). \quad (11.276)$$

Laskar has shown that, in this case, searching numerically for the value of  $m$  that maximises  $|\bar{w}_m^{(p)}|$  gives the dominant frequency in  $w_n$  with an accuracy that scales as  $1/N_t^{2p+2}$  [Laskar (1995)]. Simply looking for the peak in the discrete Fourier transform of  $w_n$  gives the dominant frequency with an accuracy that scales as  $1/N_t$ . Without any filter, i.e. with  $p = 0$  in (11.276), a numerical search (allowing  $m$  to vary continuously) for the maximum of  $|\bar{w}_m^{(1)}|$  gives the dominant frequency with an accuracy that scales as  $1/N_t^2$ . The scaling can be improved to  $1/N_t^4$  by using a Hanning filter with  $p = 1$ . Since it is often possible to track particles even through large rings for several thousand turns quite quickly, use of a Hanning filter can lead to a very significant improvement in the accuracy with which the dominant frequencies in a set of tracking data can be determined.

NAFF can be used in a technique known as *frequency map analysis* (FMA) to provide a powerful tool for the analysis of dynamics in nonlinear systems. Originally developed and applied in the context of astronomy [Laskar (1990); Laskar and Robutel (1993); Papaphilippou and Laskar (1996, 1998)], FMA is now widely used in accelerator physics to characterise the nonlinear dynamics in periodic beam lines, such as storage rings [Laskar (1994, 1995); Laskar and Robin (1996); Nadolski and Laskar (2003)]. In FMA, the tunes of particles with a range of horizontal and vertical betatron amplitudes are determined using NAFF. Most often, the turn-by-turn co-ordinates and momenta are produced in a tracking simulation, but the technique may also be applied to experimental data (see, for example, [Robin *et al.* (2000)]). The tunes obtained are plotted on a resonance diagram, such as that shown in Fig. 11.3. For small horizontal and vertical amplitudes, the tunes approach the nominal working point determined by the linear components in the beam line. As the amplitudes increase, points on the resonance diagram corresponding to different amplitudes spread out: the area of the resonance diagram covered by points over a given range of amplitudes is characteristic of the tune shift with amplitude. Neglecting

resonant terms, and considering only betatron motion, the transfer map may be represented by a Lie transformation:

$$\mathcal{M} = e^{im}, \quad (11.277)$$

where:

$$m = -\mu_x J_x - \mu_y J_y - \mu_{xx} J_x^2 - \mu_{xy} J_x J_y - \mu_{yy} J_y^2. \quad (11.278)$$

When this map is applied to the dynamical variables, the action variables are constant, while the angle variables change as follows:

$$\phi_x = \phi_x + \mu_x + 2\mu_{xx} J_x + \mu_{xy} J_y, \quad (11.279)$$

$$\phi_y = \phi_y + \mu_y + 2\mu_{yy} J_y + \mu_{xy} J_x. \quad (11.280)$$

The tune shifts with amplitude are characterised by the coefficients  $\mu_{xx}$ ,  $\mu_{xy}$  and  $\mu_{yy}$ . Note that the change in horizontal tune with vertical amplitude is equal to the change in vertical tune with horizontal amplitude: this is a consequence of the constraints imposed on the dynamics by the symplectic condition. If the tunes are far from any driven resonances, then the points for different horizontal and vertical amplitudes form a regular array on the resonance diagram.

The presence of a strongly driven resonance is indicated by the tune points for different amplitudes clustering around resonance lines on the resonance diagram. This behaviour is associated with the ‘resonance islands’ that can appear in phase space close to resonances: see, for example, Fig. 11.5. Since resonances appear in the motion in multiple degrees of freedom, it can be difficult to identify resonances by plotting phase space portraits. However, constructing a frequency map allows the effects of the resonances to be seen in the clustering of points around particular resonance lines. The strength of a resonance can be judged by the width over which the otherwise regular array of points in tune space is disrupted. An example of a frequency map analysis is shown in Fig. 11.8.

Resonances can lead to the variation of the tunes with time, either through distortion of the invariant tori in phase space, or (if the nonlinearities are strong enough) through the invariant tori breaking up altogether. The rate of change of the tunes over some specified number of turns can be used to characterise the regularity of the motion of a particle: if the tune remains constant or varies only slowly (or periodically) over time, then the motion is likely to be associated with an invariant torus in phase space, i.e. there are conserved quantities equal to or approximated by the action variables. However, if there is a large change in tune over a short time, then

the motion is likely to be irregular or chaotic. Irregular motion often occurs near the boundary of the dynamic aperture. Identifying and mitigating the resonances driving the irregular motion can sometimes help to improve the dynamic aperture and energy acceptance of a storage ring (see, for example, [Steier *et al.* (2002)]). The rate of change of tune for a particle with given

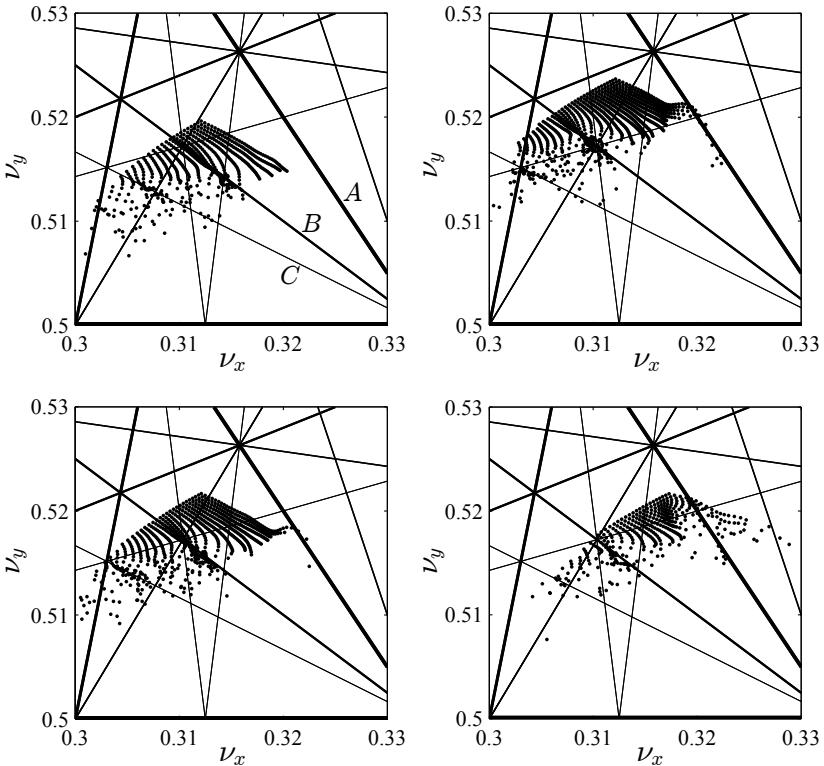


Fig. 11.8 Frequency map analysis. The points show the fractional parts of the betatron tunes for particles with different betatron amplitudes, determined by applying NAFF to tracking data for a single periodic cell of a synchrotron storage ring. (The lattice cell is a double-bend achromat, similar to that shown in Fig. 7.3.) The different plots show the results when the lattice is tuned to different points in tune space; that is, the quadrupole strengths are adjusted to give different tunes in the limit of zero amplitude. Linear changes in tune with betatron action would lead to a regular array of points in tune space. The clustering of points around particular resonance lines shows the presence of driving terms in the map for the corresponding resonances. There appear to be significant driving terms for the resonances corresponding to the lines labelled A ( $3\nu_x + 2\nu_y = 2$ ), B ( $3\nu_x + 4\nu_y = 3$ ) and C ( $3\nu_x + 6\nu_y = 4$ ). Other resonances are either too weakly driven or too far from the working point to have any clearly observable effect.

initial amplitude can be estimated by applying NAFF over different parts of a given set of tracking data. The diffusion rates obtained in this way (usually, the sum in quadrature of the rates of change of the horizontal and vertical tunes) can be used to determine the colours of points plotted in tune space: this provides an effective way to illustrate the effects of various resonances. Points corresponding to the starting conditions for the tracking data can be plotted in co-ordinate space, with the diffusion rates again indicated using different colours: this can indicate regions of the dynamic aperture where particle motion may be stable over a large number of turns, but where the motion is irregular. Such regions may become unstable in the presence of machine errors. Frequency map analysis can be used to provide a much deeper level of information on the particle dynamics than provided by a dynamic aperture plot that simply indicates regions of stable motion.

**This page intentionally left blank**

PART IV

## Collective Effects

**This page intentionally left blank**

## Chapter 12

# Space Charge

In previous chapters we discussed the dynamics of particles in accelerator beam lines considering only the forces on particles from beam line components such as magnets and rf cavities. We neglected the effects of forces on a particle from other particles within the beam line, and as a consequence, the various equations we derived to describe the dynamics were independent of the total number of particles within the beam line. However, since the particles in an accelerator carry electric charge, there will always be some interaction between the particles. In the limit of low current and high energy, the effects of the forces between particles will usually be negligible. However, accelerators are often required to work in regimes of high beam current and low or moderate particle energy: in such cases, the forces between particles can have very important effects that cannot be neglected in the design or operation of the accelerator.

The effects arising from interactions between particles within a beam are generally known as *collective effects*. The phenomena associated with collective effects are diverse, and their analysis is often quite complicated. We shall not attempt a completely rigorous or comprehensive study of these effects, but will rather focus on some of the key principles and concepts. In this chapter, we shall begin the discussion by considering space-charge effects, and in Chapter 13 we shall discuss scattering effects. Space-charge and scattering may be regarded as ‘incoherent’ effects in the sense that they result principally in changes to the overall size (or emittance) of the beam without generating or amplifying substructure within the beam. In Chapters 14 and 15 we shall consider the effects of wake fields, which can lead to the development of structure in the charge distribution, and are therefore sometimes known as ‘coherent’ effects. The defining feature of any collective effect, whether coherent or incoherent, is a dependence of its

behaviour on the overall charge (or charge density) within a beam or bunch of particles.

Consider a collection of charged particles moving along an accelerator beam line. In addition to any electric and magnetic fields generated by components along the beam line (such as magnets and rf cavities), there will be fields generated by the particles themselves. Assuming that the beam can be approximated by a continuum of charge, with charge density  $\rho$  and current density  $\mathbf{J}$ , the fields generated may be found by solving Maxwell's equations with  $\rho$  and  $\mathbf{J}$  as the source terms:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \quad (12.1)$$

$$\nabla \times \mathbf{B} - \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}, \quad (12.2)$$

where  $\mathbf{E}$  and  $\mathbf{B}$  are (respectively) the electric and magnetic fields, and  $\mu_0$  and  $\varepsilon_0$  are the permeability and permittivity of free space. The effects from the electric and magnetic fields on individual particles are described by the Lorentz force equation:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (12.3)$$

where  $\mathbf{F}$  is the force on a particle with charge  $q$  and velocity  $\mathbf{v}$ . Space-charge effects within a beam can be described by electric and magnetic fields calculated in the approximation that the beam consists of ‘smooth’ charge and current distributions. In reality, the charge distributions are not smooth, but consist of large numbers of individual particles. Scattering effects (arising from collisions between pairs of particles) and effects arising from small-scale variations in charge or current density are not accurately described by the fields arising from smooth distributions, and are treated separately.

One of the main effects of space-charge forces is a change in the distribution of charge within a beam. Since the forces depend on the distribution, the equations describing the evolution of the distribution should be solved self-consistently; that is, the changes in the fields resulting from a change in the distribution have to be taken into account when calculating how the distribution evolves. In most cases, this makes the problem very complicated, especially since the forces (in general) have a nonlinear dependence on position within the beam. Only a small number of special cases can be solved exactly. However, these cases are instructive, and it is possible to make a few generalisations that provide some insight into the behaviour of a beam in real situations.

Our analysis of space-charge effects in this chapter will proceed as follows. We shall first consider a continuous (unbunched) beam with a special distribution in phase space: the Kapchinsky–Vladimirsky, or KV distribution. We shall show that within a beam described by the KV distribution, the space-charge forces have a linear dependence on position within the beam. This allows us to derive a set of differential equations (the envelope equations) describing how the transverse beam sizes of a beam with a KV distribution evolve along a beam line. We shall then show that, with one significant change, the same equations describe the evolution of the transverse beam sizes of continuous beams with more general distributions. As an example, we shall consider the evolution of a continuous beam in a drift space, and derive the conditions for matching a beam to a solenoid channel so that the charge distribution remains invariant along the solenoid, even under the influence of space-charge forces. Finally, we shall discuss the behaviour of bunched beams, considering the impact of space-charge forces on both the longitudinal and transverse dynamics in storage rings.

## 12.1 The Kapchinsky–Vladimirsky Distribution

The Kapchinsky–Vladimirsky distribution [Kapchinsky and Vladimirsky (1959)] is defined by the particle density function:

$$\Psi = \frac{n_q}{4\pi^2 J_{x0} J_{y0}} \delta_D \left( 1 - \frac{J_x}{J_{x0}} - \frac{J_y}{J_{y0}} \right). \quad (12.4)$$

Here,  $\Psi = \Psi(\phi_x, \phi_y, J_x, J_y)$  is the density of particles in phase space (i.e. the number of particles per unit phase space volume),  $n_q$  is the number of particles per unit distance along the beam line,  $\delta_D(x)$  is the Dirac delta function, and  $J_{x0}$  and  $J_{y0}$  are positive constants.  $J_x$  is the horizontal action variable, defined by:

$$2J_x = \gamma_x x^2 + 2\alpha_x x p_x + \beta_x p_x^2, \quad (12.5)$$

where  $x$  and  $p_x$  are the horizontal co-ordinate and canonical momentum respectively, and  $\gamma_x$ ,  $\alpha_x$  and  $\beta_x$  are the horizontal Courant–Snyder parameters that define the exact shape of the particle distribution in phase space.  $J_y$  is the vertical action variable, defined in a similar way to  $J_x$ .

The number of particles per unit length is obtained by integrating the particle density function (12.4) over the horizontal and vertical action-angle variables. Using the properties of the Dirac delta function, we find:

$$\int_0^{2\pi} d\phi_x \int_0^{2\pi} d\phi_y \int_0^{J_{x0}} dJ_x \int_0^{J_{y0}} dJ_y \Psi = n_q. \quad (12.6)$$

The limits on the integrals over the action variables are determined by the fact that particles can only have values for the action variables from 0 to  $J_{x0}$  in the horizontal plane, and from 0 to  $J_{y0}$  in the vertical plane.

An important property of the KV distribution is that the particle density in the  $x$ - $y$  plane is *uniform* within an ellipse with half-axis  $a_x = \sqrt{2\beta_x J_{x0}}$  along the  $x$  axis, and  $a_y = \sqrt{2\beta_y J_{y0}}$  along the  $y$  axis. Outside this ellipse, the particle density is zero. To show that this is the case, consider the beam moments:

$$\langle x^{2m} \rangle = \frac{1}{n_q} \int_0^{2\pi} d\phi_x \int_0^{2\pi} d\phi_y \int_0^{J_{x0}} dJ_x \int_0^{J_{y0}} dJ_y x^{2m} \Psi, \quad (12.7)$$

where  $m$  is any positive integer. Using:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (12.8)$$

we find that:

$$\langle x^{2m} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos^{2m}(\phi_x) d\phi_x \frac{(2\beta_x J_{x0})^m}{m+1}. \quad (12.9)$$

The integral over  $\phi_x$  may be evaluated, but we do not need to do so for our present purposes. A similar expression is obtained for the moments  $\langle y^{2m} \rangle$ . Now consider a distribution of particles with uniform charge density  $\rho = qn_q/\pi a_x a_y$  within an ellipse defined by:

$$\frac{x^2}{a_x^2} + \frac{y^2}{a_y^2} = 1. \quad (12.10)$$

Outside the ellipse, the particle density is zero. The moments of this distribution are:

$$\langle x^{2m} \rangle = \frac{\iint \rho x^{2m} dx dy}{\iint \rho dx dy} = \frac{1}{\pi a_x a_y} \iint x^{2m} dx dy, \quad (12.11)$$

where the limits of integration are defined by the ellipse (12.10). We can perform the final integral by changing to polar co-ordinates  $(r, \theta)$ :

$$x = a_x r \cos(\theta), \quad (12.12)$$

$$y = a_y r \sin(\theta). \quad (12.13)$$

Within the ellipse (12.10),  $\theta$  ranges from 0 to  $2\pi$ , and  $r$  ranges from 0 to 1. The Jacobian of the transformation gives:

$$dx dy = a_x a_y r dr d\theta. \quad (12.14)$$

Hence we find that:

$$\langle x^{2m} \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos^{2m}(\theta) d\theta \frac{a_x^{2m}}{m+1}. \quad (12.15)$$

Comparing (12.15) with (12.9), we see immediately that the beam moments  $\langle x^{2m} \rangle$  for the KV distribution are the same as the moments for the uniform co-ordinate space distribution, if:

$$a_x = \sqrt{2\beta_x J_{x0}}. \quad (12.16)$$

A similar result holds for the vertical beam moments  $\langle y^{2m} \rangle$ . Also, for both distributions:

$$\langle x^{2m+1} \rangle = \langle y^{2m+1} \rangle = 0, \quad (12.17)$$

where  $m$  is zero or any positive integer. Finally, the cross-correlations  $\langle x^{m_x} y^{m_y} \rangle$  for integers  $m_x$  and  $m_y$  are also the same for both the KV and the uniform co-ordinate space distributions. Since all the moments of the two distributions are the same, we conclude that the projection of the KV distribution onto the  $x-y$  plane gives a uniform particle density with  $n_q/\pi a_x a_y$  particles per unit volume within the ellipse (12.10), and no particles outside this ellipse.

If we define the horizontal beam emittance  $\epsilon_x$  as the average of the betatron action over all particles in the beam, then we find for the KV distribution:

$$\epsilon_x = \langle J_x \rangle = \frac{J_{x0}}{2}. \quad (12.18)$$

Then, the second-order beam moment is related to the emittance by:

$$\sigma_x^2 = \langle x^2 \rangle = \frac{a_x^2}{4} = \frac{\beta_x J_{x0}}{2} = \beta_x \epsilon_x. \quad (12.19)$$

This is the relation we would normally expect between the rms beam size, the beta function and the emittance. Of course, a similar relation holds in the vertical plane.

The KV distribution is important in studies of space charge precisely because its projection onto the  $x-y$  plane gives a uniform particle density: in this case, the force seen by any particle in a continuous beam varies linearly with distance from the centre of the ellipse (12.10), and the particle dynamics can then be treated using the techniques of linear dynamics. To calculate the space-charge force on a particle in the beam, it is convenient first to work in the frame moving longitudinally with the particles. For simplicity, we assume that all particles have longitudinal momentum close to the reference momentum:

$$P_0 = \beta_0 \gamma_0 m c, \quad (12.20)$$

where  $m$  is the mass of each particle. If there are  $n_q$  particles per unit length in the laboratory frame, there will be  $n_q/\gamma_0$  particles per unit length

in the frame moving with speed  $\beta_0 c$  in the direction of the beam. We refer to this frame as the *beam frame*. Transversely, the distribution of the particles is unaffected by the Lorentz boost. We assume that we can neglect the magnetic forces that come from the (small) velocities of the particles in the beam frame. To find the electric field, we have to solve Maxwell's equation (12.1) for a two-dimensional static charge distribution with elliptical symmetry. For a charge distribution with no dependence on the longitudinal co-ordinate  $z$ , and with density dependence on the transverse variables  $x$  and  $y$  given by:

$$\rho = \rho(u), \quad (12.21)$$

where:

$$u = \frac{x^2}{a_x^2} + \frac{y^2}{a_y^2}, \quad (12.22)$$

the solution to Maxwell's equation (12.1) is [Kellogg (1953)]:

$$E_{x,\text{beam}}(x, y) = x \frac{a_x a_y}{2\varepsilon_0} \int_{\chi}^{\infty} \frac{\rho(\bar{u})}{(a_x^2 + \zeta)^{3/2}(a_y^2 + \zeta)^{1/2}} d\zeta, \quad (12.23)$$

$$E_{y,\text{beam}}(x, y) = y \frac{a_x a_y}{2\varepsilon_0} \int_{\chi}^{\infty} \frac{\rho(\bar{u})}{(a_x^2 + \zeta)^{1/2}(a_y^2 + \zeta)^{3/2}} d\zeta, \quad (12.24)$$

where:

$$\bar{u}(x, y; \zeta) = \frac{x^2}{a_x^2 + \zeta} + \frac{y^2}{a_y^2 + \zeta}. \quad (12.25)$$

The subscript 'beam' indicates that fields are given in the rest frame of the beam. The function  $\chi = \chi(x, y)$  is defined so that:

$$\chi = 0 \quad \text{if } u \leq 1, \quad (12.26)$$

and:

$$\bar{u}(x, y; \chi) = 1 \quad \text{if } u > 1. \quad (12.27)$$

For a continuous beam we expect the parameters  $a_x$  and  $a_y$  to vary with longitudinal position: strictly speaking, equations (12.23) and (12.24) are valid for constant  $a_x$  and  $a_y$ . However, we can assume that, in practice,  $a_x$  and  $a_y$  will vary slowly with position along the beam (which will certainly be the case if  $\gamma_0$  is reasonably large), and then equations (12.23) and (12.24) will provide a good approximation to the field.

For the KV distribution, we are interested in the case:

$$\rho(u) = \begin{cases} qn_q/\gamma_0 \pi a_x a_y & \text{if } u \leq 1, \\ 0 & \text{if } u > 1, \end{cases} \quad (12.28)$$

where  $q$  is the charge on a particle, and  $n_q$  (as before) is the number of particles per unit length along the beam line in the laboratory frame. Performing the integrals in equations (12.23) and (12.24) gives, for  $u \leq 1$ :

$$E_{x,\text{beam}}(x, y) = \frac{qn_q}{\gamma_0 \pi \varepsilon_0} \frac{x}{a_x(a_x + a_y)}, \quad (12.29)$$

$$E_{y,\text{beam}}(x, y) = \frac{qn_q}{\gamma_0 \pi \varepsilon_0} \frac{y}{a_y(a_x + a_y)}. \quad (12.30)$$

These are the fields in a frame in which the longitudinal velocity of the reference particle (moving with momentum  $P_0$  along the reference trajectory) is zero. In this frame, we assume that we can neglect any magnetic field arising from the (small) transverse velocities of the particles. To find the electric and magnetic fields in the laboratory, we perform a Lorentz boost to a frame moving with velocity  $-\beta_0 c$  along the  $z$  axis, relative to the beam. The Lorentz transformation with such a boost for a general electromagnetic field is:

$$E_x = \gamma_0 (E_{x,\text{beam}} + \beta_0 c B_{y,\text{beam}}), \quad (12.31)$$

$$E_y = \gamma_0 (E_{y,\text{beam}} - \beta_0 c B_{x,\text{beam}}), \quad (12.32)$$

$$E_z = E_{z,\text{beam}}, \quad (12.33)$$

$$B_x = \gamma_0 \left( B_{x,\text{beam}} - \frac{\beta_0}{c} E_{y,\text{beam}} \right), \quad (12.34)$$

$$B_y = \gamma_0 \left( B_{y,\text{beam}} + \frac{\beta_0}{c} E_{x,\text{beam}} \right), \quad (12.35)$$

$$B_z = B_{z,\text{beam}}. \quad (12.36)$$

Applying the Lorentz transformation to the electric field (12.29) and (12.30) gives electric field components in the laboratory frame:

$$E_x(x, y) = \frac{qn_q}{\pi \varepsilon_0} \frac{x}{a_x(a_x + a_y)}, \quad (12.37)$$

$$E_y(x, y) = \frac{qn_q}{\pi \varepsilon_0} \frac{y}{a_y(a_x + a_y)}, \quad (12.38)$$

and magnetic field components in the laboratory frame:

$$B_x(x, y) = -\frac{\beta_0}{c} E_y, \quad (12.39)$$

$$B_y(x, y) = \frac{\beta_0}{c} E_x. \quad (12.40)$$

## 12.2 The Envelope Equations for the KV Distribution

Equations (12.37) and (12.38) state that in the case of a continuous beam with a KV transverse distribution, the electric field in the beam varies linearly with distance from the axis of the beam. This means that the space-charge forces have a similar effect on the beam as the magnetic field in a quadrupole, and it is possible to describe the behaviour of the beam using the techniques of linear dynamics. The main difference between space-charge forces and the forces from a quadrupole field is that in the case of space-charge forces the size of the force depends on the size of the beam and the particle density within the beam (as well as on the position of a particle experiencing the space-charge force). However, in both cases the dynamics of particles in the beam can be derived from a Hamiltonian. The terms in the Hamiltonian describing the space-charge effects are dependent on the longitudinal position along the beam line (to take account of variations in the beam size) and have a similar form to the terms describing the effects of a quadrupole; although in the case of space-charge effects, the forces are *simultaneously defocusing* in the horizontal and vertical planes.

Since the particle dynamics are linear and described by a Hamiltonian, the transverse actions  $J_x$  and  $J_y$  are conserved quantities. Since the KV distribution is a function purely of  $J_x$  and  $J_y$ , this means that the KV distribution is invariant as the beam is transported along a linear beam line, even with space-charge forces taken into account. The beam size will of course vary along the beam line, but this can be described by the variation of the beta functions. In the case of the KV distribution, therefore, it is possible to find a self-consistent solution to the equations of motion for a beam in the presence of space-charge forces. The variation of the beta function will be affected by the space-charge forces, but the distribution will remain a KV distribution.

Another important feature of the KV distribution is that, since the particle actions  $J_x$  and  $J_y$  are conserved, the emittances  $\epsilon_x = \langle J_x \rangle$  and  $\epsilon_y = \langle J_y \rangle$  will also be conserved. A consequence of this, as we shall now show, is that it is possible to write down self-contained equations describing how the horizontal and vertical beam sizes vary along a beam line in the presence of external focusing forces and space-charge forces. By ‘self-contained’, we mean that the only quantities in the equations that are functions of position along the beam line are the horizontal and vertical beam sizes. For more general (i.e. non-KV) particle distributions, the emittances are not conserved quantities. We shall show later that for distributions with

elliptical symmetry (whether KV or non-KV distributions) the envelope equations describing the variation of the beam sizes take the same form. However, the emittance appears as a parameter in the equations; so for cases where the emittance is not conserved, the equations become significantly more difficult to solve.

In this section, we shall derive the envelope equations describing the evolution of the beam size for continuous beams with transverse KV distribution, in the presence of space-charge effects. To begin, let us consider the evolution of a general distribution in the presence only of external linear forces, i.e. forces that depend linearly on the position of a particle with respect to a given reference trajectory. In the small angle approximation, we can write:

$$x' \approx \frac{\gamma_0 m \dot{x}}{P_0} = p_x, \quad (12.41)$$

where the prime denotes the derivative with respect to distance along the reference trajectory,  $\gamma_0$  is the relativistic factor for the particle,  $m$  is the mass of the particle, the dot denotes the derivative with respect to time, and  $P_0$  is a given reference momentum. We assume that the  $x$  component of the vector potential is zero, and that the total momentum of the particle is equal to the reference momentum. In the presence of a focusing force  $k_1$  (for example, arising from a quadrupole field), the equation of motion for the particle in the  $x$  direction is:

$$x'' = -k_1 x. \quad (12.42)$$

In general,  $k_1 = k_1(s)$  is a function of position along the reference trajectory. It was shown in Section 4.5 that equation (12.42) has solution:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (12.43)$$

where  $J_x$  (the action variable) is constant, and  $\phi_x$  (the angle variable) is a function of  $s$ . The beta function  $\beta_x$  is a function of  $s$  that satisfies (4.82):

$$\beta_x'' - \frac{4 + \beta_x'^2}{2\beta_x} + 2k_1\beta_x = 0. \quad (12.44)$$

If the lattice is periodic with period  $C_0$ , that is:

$$k_1(s + C_0) = k_1(s), \quad (12.45)$$

then imposing the same periodicity condition on the beta function:

$$\beta_x(s + C_0) = \beta_x(s), \quad (12.46)$$

$$\beta_x'(s + C_0) = \beta_x'(s), \quad (12.47)$$

leads to a unique solution for  $\beta_x$ . Consider a beam consisting of a large number of particles with uniform distribution in  $\phi_x$ , and with a random distribution of  $J_x$  with some maximum value  $J_{x,\max}$ . At any point along the beam line, we can assume that there will be some particle with action close to  $J_{x,\max}$  and with a value for  $\phi_x$  close to 0 or  $\pi$ ; then the beam radius  $a_x$  at any point is given by:

$$a_x = \max(x) = \sqrt{2\beta_x J_{x,\max}}. \quad (12.48)$$

In general, the radius of a beam injected into a periodic beam line may not evolve with the periodicity of the beam line. The beam can still be described by a beta function satisfying (12.44); however, the beta function will not satisfy the periodicity conditions (12.46) and (12.47). When the radius of the beam (and therefore the beta function) *does* satisfy the periodicity conditions (12.46) and (12.47), the beam is said to be properly *matched* to the beam line. For our purposes, it does not actually matter whether the beam is correctly matched to the beam line or not: the important result is that the beta function always satisfies equation (12.44).

Since  $J_{x,\max}$  is a constant, the beam radius  $a_x$  is proportional to the square root of the beta function. Then, substituting for  $\beta_x$  from (12.48) into (12.44) gives the equation for the evolution of the beam radius along the beam line:

$$a''_x + k_1 a_x - \frac{4J_{x,\max}^2}{a_x^3} = 0. \quad (12.49)$$

Equation (12.49) is known as the *envelope equation*. To derive this equation, we have made no assumptions as to the particular distribution of the action variable for particles in the beam. For a distribution that has a hard ‘edge’ (for example, a distribution that is uniform in co-ordinate space within the ellipse defined by (12.10), and zero outside this ellipse), it can be appropriate to characterise the distribution using a maximum value of the action variable  $J_{x,\max}$ . However, for other distributions (for example, a Gaussian distribution), it may be more appropriate to characterise the distribution in terms of an average value  $\langle J_x \rangle$ . Assuming that the angle variables for the particles in a beam are uniformly distributed and uncorrelated with the action variables, the mean square value of the co-ordinate  $x$  is:

$$\sigma_x^2 = \langle x^2 \rangle = \beta_x \langle J_x \rangle = \beta_x \epsilon_x, \quad (12.50)$$

where the rms emittance  $\epsilon_x$  is defined by:

$$\epsilon_x = \langle J_x \rangle. \quad (12.51)$$

Since the angle variables for the particles in a beam are constant, the emittance is also constant. Then, substituting for  $\beta_x$  from (12.50) into (12.44) gives:

$$\sigma''_x + k_1 \sigma_x - \frac{\epsilon_x^2}{\sigma_x^3} = 0. \quad (12.52)$$

Equation (12.52) is an alternative form of the envelope equation, expressed in terms of the rms beam size  $\sigma_x$ , rather than by the ‘edge’ beam size  $a_x$ .

The envelope equation can readily be adapted to include the effects of space-charge forces in the case of the KV distribution. The KV distribution gives rise to an electric field that is proportional to the distance from the centre of the beam: the space-charge force on a particle in the beam is therefore linear in the distance of the particle from the centre of the beam, and can be taken into account simply by modifying the focusing force  $k_1$ . From the Lorentz force equation (12.3), and using the expressions (12.37)–(12.40) for the fields in a beam with a KV distribution, the components of the force on a particle in the beam are, in the laboratory frame:

$$F_x = \frac{qE_x}{\gamma_0^2} = \frac{q^2 n_q}{\pi \varepsilon_0} \frac{x}{\gamma_0^2 a_x (a_x + a_y)}, \quad (12.53)$$

$$F_y = \frac{qE_y}{\gamma_0^2} = \frac{q^2 n_q}{\pi \varepsilon_0} \frac{y}{\gamma_0^2 a_y (a_x + a_y)}. \quad (12.54)$$

At high beam energy (large  $\gamma_0$ ) the force from the electric field is nearly cancelled by the force from the magnetic field: the resultant force is equal to the electric force divided by  $\gamma_0^2$ .

Assuming that the trajectory of a particle makes only a small angle with the axis, we can write:

$$x'' \approx \frac{\ddot{x}}{\beta_0^2 c^2} = \frac{F_x}{\beta_0^2 \gamma_0 m c^2}. \quad (12.55)$$

Taking the  $x$  component of the space-charge force (12.53) we find:

$$x'' = \frac{2K}{a_x (a_x + a_y)} x, \quad (12.56)$$

where the *perveance*  $K$  (characterising the impact of spacecharge forces on the beam dynamics) is given by:

$$K = \frac{2I}{\beta_0^3 \gamma_0^3 I_c}. \quad (12.57)$$

$I$  is the total beam current:

$$I = q n_q \beta_0 c, \quad (12.58)$$

and  $I_c$  is the characteristic current:

$$I_c = \frac{4\pi\varepsilon_0 mc^3}{q}. \quad (12.59)$$

For electrons,  $I_c$  is the Alfvén current,  $I_A \approx 17.045$  kA. Including linear focusing in the beam line (for example, from quadrupoles) gives the equation of motion:

$$x'' = \left( \frac{2K}{a_x(a_x + a_y)} - k_1 \right) x. \quad (12.60)$$

The corresponding equation for the vertical motion is:

$$y'' = \left( \frac{2K}{a_y(a_x + a_y)} + k_1 \right) y. \quad (12.61)$$

The space-charge force is simultaneously defocusing in both the horizontal and the vertical directions, whereas the quadrupole focusing term when focusing in  $x$  ( $k_1 > 0$ ) is defocusing in  $y$ , and vice versa.

The permeance  $K$  is (for a beam consisting of particles of a given type) a function only of the beam energy and current, which can be assumed to be constant along the beam line. The beam radii  $a_x$  and  $a_y$  will, in general, vary along the beam line, as will the linear focusing  $k_1$ . Since the equation of motion is linear in  $x$ , it can be solved by a Floquet transformation, in the same way that we solved Hill's equation (in Chapter 4) when there was only the term in  $k_1$  on the right-hand side. In other words, the motion of an individual particle can be written, with or without space-charge, as:

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (12.62)$$

where  $J_x$  is the action, and  $\phi_x$  is the angle for the particle. Equation (12.44) for the horizontal beta function becomes:

$$\beta_x'' - \frac{4 + \beta_x'^2}{2\beta_x} + 2 \left( k_1 - \frac{2K}{a_x(a_x + a_y)} \right) \beta_x = 0. \quad (12.63)$$

For the vertical beta function we have:

$$\beta_y'' - \frac{4 + \beta_y'^2}{2\beta_y} - 2 \left( k_1 + \frac{2K}{a_y(a_x + a_y)} \right) \beta_y = 0. \quad (12.64)$$

Substituting as before for  $\beta_x$  from (12.48), we find the envelope equations for a KV beam with (in co-ordinate space) horizontal radius  $a_x$  and vertical radius  $a_y$  in the presence of space-charge forces:

$$a_x'' + k_1 a_x - \frac{4J_{x,\max}^2}{a_x^3} - \frac{2K}{a_x + a_y} = 0, \quad (12.65)$$

$$a_y'' - k_1 a_y - \frac{4J_{y,\max}^2}{a_y^3} - \frac{2K}{a_x + a_y} = 0. \quad (12.66)$$

Using (12.18) and (12.19), the envelope equations may be written in terms of the rms beam sizes  $\sigma_x = a_x/2$  and  $\sigma_y = a_y/2$ , and emittances  $\epsilon_x = \langle J_x \rangle = J_{x,\max}/2$  and  $\epsilon_y = \langle J_y \rangle = J_{y,\max}/2$ :

$$\sigma''_x + k_1 \sigma_x - \frac{\epsilon_x^2}{\sigma_x^3} - \frac{K}{2(\sigma_x + \sigma_y)} = 0, \quad (12.67)$$

$$\sigma''_y - k_1 \sigma_y - \frac{\epsilon_y^2}{\sigma_y^3} - \frac{K}{2(\sigma_x + \sigma_y)} = 0. \quad (12.68)$$

In the case of a continuous beam with a transverse KV distribution, the particle actions (and hence the beam emittances) are conserved under linear symplectic transport, even when space-charge forces are included. Then, the envelope equations (12.67) and (12.68) may be integrated to find the evolution of the rms beam sizes along a beam line. It is possible to use these equations to distinguish between two different regimes: if the action term dominates over the perveance term, the beam is said to be *emittance dominated*; if the perveance term dominates over the action term, the beam is said to be *space-charge dominated*.

### 12.3 Elliptically Symmetric Non-KV Distributions

We showed in Section 12.2 that for a beam with a KV distribution the emittance was conserved under transport along a beam line under the influence of linear focusing forces, including space-charge forces. This is a consequence of the fact that in co-ordinate space the particles are uniformly distributed inside an ellipse (the cross-section of the beam), which means that the horizontal and vertical components of the space-charge force vary linearly with distance from the centre of the ellipse. For other distributions, we expect a nonlinear variation of the space-charge force with position in the beam, leading to a variation of the emittance as the beam is transported along a beam line. Finding the change in the emittance in the general case is a complicated problem. Here, we shall consider a family of elliptically symmetric distributions of a particular form, for which the KV distribution and a Gaussian distribution constitute special cases. By considering the transfer of energy from electrical potential energy to the kinetic energy of particles in the beam, we shall show that the emittance growth for distributions within this family is related to the perveance of the beam. We shall also derive the interesting result that, despite the variation in the emittance, beams with elliptically symmetric non-KV distributions obey envelope equations of the same form as those for beams with a KV distribution.

Consider a distribution:

$$\Psi_m = \frac{n_q m(m+1)}{4\pi^2 J_{x0} J_{y0}} \left(1 - \frac{J_x}{J_{x0}} - \frac{J_y}{J_{y0}}\right)^{m-1}, \quad (12.69)$$

where  $n_q$  is the number of particles per unit distance along the reference trajectory, and  $m$  is an integer, greater than or equal to zero. The case  $m = 0$  corresponds to the KV distribution, and as  $m \rightarrow \infty$  the form of the distribution approaches a Gaussian. By integrating over the distribution (in the same way as we did for the KV distribution), it is possible to show that the moments of the distribution (12.69) are:

$$\langle x^{\ell_x} y^{\ell_y} \rangle = (m+1) \frac{a_x^{\ell_x} a_y^{\ell_y}}{\pi} \int_0^{2\pi} \cos^{\ell_x}(\theta) \sin^{\ell_y}(\theta) d\theta \int_0^1 r^{\ell_x + \ell_y + 1} (1 - r^2)^m dr. \quad (12.70)$$

These are the same as the moments of the distribution:

$$\frac{\rho_m}{q} = \frac{n_q(m+1)}{\pi a_x a_y} \left(1 - \frac{x^2}{a_x^2} - \frac{y^2}{a_y^2}\right)^m, \quad (12.71)$$

where:

$$a_x = \sqrt{2\beta_x J_{x0}}, \quad (12.72)$$

$$a_y = \sqrt{2\beta_y J_{y0}}. \quad (12.73)$$

In particular:

$$\langle x^2 \rangle = \frac{a_x x^2}{2(m+2)} = \frac{\beta_x J_{x0}}{m+2}. \quad (12.74)$$

We also find from (12.69) that the mean value of the horizontal action is given by:

$$\langle J_x \rangle = \frac{J_{x0}}{m+2}. \quad (12.75)$$

Therefore, if we define the horizontal emittance in the usual way:

$$\epsilon_x = \langle J_x \rangle, \quad (12.76)$$

then we have the familiar relationship between the rms beam size and the emittance:

$$\langle x^2 \rangle = \beta_x \epsilon_x. \quad (12.77)$$

Similar relationships hold for the vertical variables. Two distributions (with different  $m$ ) are said to be *equivalent* if they have the same Courant–Snyder parameters and emittances.

Since the space-charge forces in a beam with a distribution of the form (12.69) are nonlinear (for  $m > 0$ ), the distribution will not be invariant as the beam moves along the beam line. In general, we can expect the space-charge forces to act in such a way as to minimise the electric potential energy of the particles in the beam. As we shall see below, a distribution with a given  $m$  has a higher potential energy than an equivalent distribution (i.e. a distribution with the same Courant–Snyder parameters and emittances) with a lower  $m$ . A distribution with  $m = 0$  corresponds to a KV distribution, which is invariant under space-charge forces. This suggests that a beam in a distribution with a given value of  $m$  will evolve towards a KV distribution. The loss in electric potential energy will appear as an increase in the kinetic energy of the beam: this can be related to an increase in the emittance of the beam. Therefore, we can estimate the emittance growth from space-charge forces in a beam with a distribution of the form (12.69) by calculating the change in potential energy if the charge is redistributed into a KV distribution.

This is a simplified model of the emittance evolution; nevertheless, we expect that it will provide some reasonable estimate of the emittance growth in a given case. Let us proceed with the analysis, as follows. First, we can write down the rate of change of the horizontal mechanical momentum  $P_x = \gamma_0 m \dot{x}$  of a particle in terms of the electric and magnetic forces acting on the particle:

$$\frac{dP_x}{dt} = q(E_x - \beta_0 c B_y), \quad (12.78)$$

where we assume that the longitudinal velocity is approximately equal to the total velocity of the particle, i.e.  $\beta_s \approx \beta_0$ , and we assume that the beam is moving in a drift space (i.e. there are no external fields). If we assume that the magnetic forces in the rest frame of the beam are negligible, then we can relate the magnetic field  $B_y$  to the electric field  $E_x$ . The fields in the laboratory frame are related to the fields in the beam frame by a Lorentz boost (12.31)–(12.36). Assuming that the magnetic field in the rest frame of the beam is zero, i.e.  $\mathbf{B}_{\text{beam}} = 0$ , then:

$$E_x - \beta_0 c B_y = \gamma_0 (1 - \beta_0^2) E_{x,\text{beam}} = \frac{E_x}{\gamma_0^2}. \quad (12.79)$$

In terms of the canonical momentum  $p_x = P_x/P_0$ , equation (12.78) for the rate of change of the momentum becomes:

$$\frac{dp_x}{dt} = \frac{qE_x}{\gamma_0^2 P_0} = \frac{qE_x}{\beta_0 \gamma_0^3 mc}. \quad (12.80)$$

Since  $ds = \beta_0 c dt$ , the rate of change of momentum with respect to distance along the beam line becomes:

$$\frac{dp_x}{ds} = \frac{qE_x}{\beta_0^2 \gamma_0^3 mc^2}, \quad (12.81)$$

and hence:

$$\frac{dp_x^2}{ds} = \frac{2qp_x E_x}{\beta_0^2 \gamma_0^3 mc^2}. \quad (12.82)$$

Adding this to the equivalent expression for the vertical motion gives:

$$\frac{d}{ds}(p_x^2 + p_y^2) = \frac{2q}{\beta_0^2 \gamma_0^3 mc^2}(p_x E_x + p_y E_y). \quad (12.83)$$

Now, for small values of the transverse momenta,  $p_x \approx dx/ds$  and  $p_y \approx dy/ds$ . Hence, the quantity  $q(p_x E_x + p_y E_y)$  represents the rate at which the electric field does work on the particle with respect to distance along the reference trajectory, that is:

$$q(p_x E_x + p_y E_y) = -\frac{du}{ds}, \quad (12.84)$$

where  $u$  is the electric potential energy of a particle. Hence:

$$\frac{d}{ds}(p_x^2 + p_y^2) = -\frac{2}{\beta_0^2 \gamma_0^3 mc^2} \frac{du}{ds}. \quad (12.85)$$

Averaging over the particles in the beam at some given location along the reference trajectory, and using:

$$\langle p_x^2 \rangle = \gamma_x \epsilon_x, \quad (12.86)$$

where  $\gamma_x$  is the Courant–Snyder gamma function and  $\epsilon_x$  is the emittance (and similarly for the vertical motion):

$$\Delta(\gamma_x \epsilon_x + \gamma_y \epsilon_y) = -\frac{2}{\beta_0^2 \gamma_0^3 mc^2} \Delta \langle u \rangle, \quad (12.87)$$

where  $\Delta$  indicates a change in the given quantity between two points in the beam line.

For a small change  $\Delta\rho$  in the charge density (using SI units,  $\Delta\rho$  is in  $\text{C m}^{-3}$ ), the change in the average electric potential energy per particle is:

$$\Delta \langle u \rangle = \frac{1}{n_q} \iint dx dy \Delta\rho(x, y) \Phi(x, y), \quad (12.88)$$

where we assume that the electric potential  $\Phi(x, y)$  remains approximately constant. Let us suppose that we redistribute the charge from an initial distribution  $\rho_m$ , given by (12.71), to a final distribution  $\rho_0$ . As the charge

is redistributed, the potential will change from  $\Phi_m$  to  $\Phi_0$ . The total change in the average potential energy per particle during the charge redistribution is:

$$\Delta\langle u \rangle = \frac{1}{n_q} \iint dx dy \int_0^1 d\zeta (\rho_0 - \rho_m) ((1 - \zeta)\Phi_m + \zeta\Phi_0). \quad (12.89)$$

Performing the integrals in (12.89) is non-trivial, but can be done for a given value of  $m$ , using the expressions (12.23) and (12.24) for the electric fields in a beam with elliptical symmetry: the potential (with  $\Phi = 0$  at the centre of the beam,  $x = y = 0$ ) can be obtained by integrating the fields. Assuming equivalent distributions (i.e. distributions with the same emittance), the result of the integration is independent of the emittance, and can be expressed as:

$$\Delta\langle u \rangle = -\frac{q^2 n_q}{4\pi\epsilon_0} \xi_m, \quad (12.90)$$

where  $\xi_m$  is a (numerical) constant, depending on the value of  $m$ . The values of  $\xi_m$  for  $1 \leq m \leq 6$  are shown in Fig. 12.1. The coefficients may be approximated by:

$$\xi_m \approx 0.024(1 - e^{-0.246m}). \quad (12.91)$$

Note that the change in energy is negative (i.e. the potential energy is reduced in a redistribution from  $m > 0$  to  $m = 0$ ), and the magnitude becomes larger for increasing  $m$ . This suggests that the KV distribution is the most ‘stable’ of the family (12.69) of distributions (assuming that the distributions are equivalent), and justifies our assumption that the charge in a non-KV distribution will be redistributed towards a KV distribution. The largest change in potential energy is achieved for  $m \rightarrow \infty$ , i.e. for a Gaussian distribution.

Using (12.87) and (12.90) we can write:

$$\Delta(\gamma_x \epsilon_x + \gamma_y \epsilon_y) = \frac{2}{\beta_0^2 \gamma_0^3 m c^2} \frac{q^2 n_q}{4\pi\epsilon_0} \xi_m = \xi_m K, \quad (12.92)$$

where  $K$  is the perveance, given by (12.57). Equation (12.92) indicates the emittance growth that can be expected in a beam with a given initial distribution of the form (12.69), as a result of space-charge forces. However, (12.92) does not tell us how the emittance actually evolves (i.e. the rate of emittance growth along the beam line); nor does it tell us how the emittance growth will be partitioned between the horizontal and the vertical. For a more accurate and detailed analysis of beam dynamics in the presence of space-charge effects, it is necessary to perform numerical simulations.

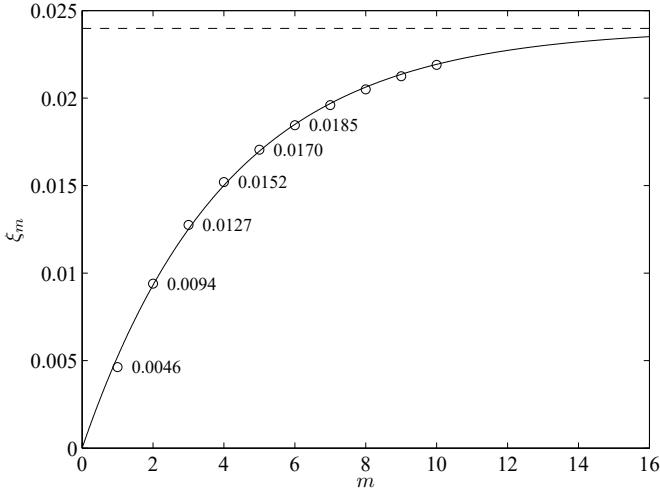


Fig. 12.1 Coefficients describing the change in energy associated with redistribution of charge in an elliptically symmetric charge distribution of the form (12.71). The circles show the results of a numerical integration of (12.89), with  $\xi_m$  defined by (12.90); the values of  $\xi_m$  for  $1 \leq m \leq 6$  are shown on the plot. The solid line shows a fit to the points, of the form (12.91). The dashed line shows the asymptotic value of  $\xi_m$  in the limit  $m \rightarrow \infty$ .

Despite the variation in the emittance, the horizontal and vertical rms beam sizes for a non-KV distribution in an uncoupled beam line obey envelope equations of the same form as for a KV distribution, (12.67) and (12.68). This can be shown as follows [Sacherer (1970, 1971)]. In the absence of coupling, the horizontal emittance can be written:

$$\epsilon_x^2 = \langle x^2 \rangle \langle p_x^2 \rangle - \langle xp_x \rangle^2. \quad (12.93)$$

With the approximation  $p_x \approx x' = dx/ds$ , we can write:

$$\frac{d}{ds} \langle x^2 \rangle = 2 \langle xp_x \rangle, \quad (12.94)$$

and:

$$\frac{1}{2} \frac{d^2}{ds^2} \langle x^2 \rangle = \frac{d}{ds} \langle xp_x \rangle = \langle p_x^2 \rangle + \left\langle x \frac{dp_x}{ds} \right\rangle. \quad (12.95)$$

Then, substituting for  $\langle p_x^2 \rangle$  and  $\langle xp_x \rangle$  in (12.93) gives:

$$\frac{\epsilon_x^2}{\sigma_x^2} = \frac{1}{2} \frac{d^2 \sigma_x^2}{ds^2} - \left\langle x \frac{dp_x}{ds} \right\rangle - \frac{1}{4\sigma_x^2} \left( \frac{d\sigma_x^2}{ds} \right)^2, \quad (12.96)$$

where  $\sigma_x^2 = \langle x^2 \rangle$ . Then, using:

$$\frac{1}{2} \frac{d^2 \sigma_x^2}{ds^2} = \left( \frac{d\sigma_x}{ds} \right)^2 + \sigma_x \frac{d^2 \sigma_x}{ds^2}, \quad (12.97)$$

and:

$$\frac{1}{4\sigma_x^2} \left( \frac{d\sigma_x^2}{ds} \right)^2 = \left( \frac{d\sigma_x}{ds} \right)^2, \quad (12.98)$$

we find:

$$\frac{\epsilon_x^2}{\sigma_x^2} = \sigma_x \frac{d^2 \sigma_x}{ds^2} - \left\langle x \frac{dp_x}{ds} \right\rangle. \quad (12.99)$$

Now, the rate of change of the value of the momentum  $p_x$  of a particle is related to the force acting on the particle:

$$\frac{dp_x}{ds} = -k_1 x + \frac{qE_x}{\beta_0 \gamma_0^2 P_0 c}, \quad (12.100)$$

where  $E_x$  is the horizontal component of the electric field. Therefore:

$$\left\langle x \frac{dp_x}{ds} \right\rangle = -k_1 \sigma_x^2 + \frac{q}{\beta_0 \gamma_0^2 P_0 c} \langle x E_x \rangle, \quad (12.101)$$

and (12.99) becomes:

$$\frac{d^2 \sigma_x}{ds^2} - \frac{\epsilon_x^2}{\sigma_x^2} + k_1 \sigma_x - \frac{q}{\beta_0^2 \gamma_0^3 mc} \frac{\langle x E_x \rangle}{\sigma_x} = 0, \quad (12.102)$$

where, in the final step, we have used  $P_0 = \beta_0 \gamma_0 mc$ .

To make further progress, we need to make some assumption about the distribution, which will allow us to evaluate the quantity  $\langle x E_x \rangle$ . Sacherer [Sacherer (1970, 1971)] has shown that for elliptically symmetric charge distributions in continuous beams,  $\langle x E_x \rangle$  is independent of the exact form of the distribution. Following Sacherer's approach, we first write the horizontal component of the electric field in the form (12.23):

$$E_x = x \frac{a_x a_y}{2\varepsilon_0} \int_0^\infty \frac{\rho(\bar{u})}{(a_x^2 + \zeta)^{3/2} (a_y^2 + \zeta)^{1/2}} d\zeta, \quad (12.103)$$

where:

$$\bar{u}(x, y; \zeta) = \frac{x^2}{a_x^2 + \zeta} + \frac{y^2}{a_y^2 + \zeta}. \quad (12.104)$$

Using the fact that:

$$\int_{-\infty}^\infty dx \int_{-\infty}^\infty dy \rho(u) = q n_q, \quad (12.105)$$

where:

$$u(x, y) = \frac{x^2}{a_x^2} + \frac{y^2}{a_y^2}, \quad (12.106)$$

it follows that:

$$\langle xE_x \rangle = \frac{a_x a_y}{2\varepsilon_0 q n_q} \int_0^\infty d\zeta \int_{-\infty}^\infty dx \int_{-\infty}^\infty dy \frac{x^2 \rho(\bar{u}) \rho(u)}{(a_x^2 + \zeta)^{3/2} (a_y^2 + \zeta)^{1/2}}. \quad (12.107)$$

We now change the variables of integration from  $x$  and  $y$  to  $r$  and  $\theta$ , where:

$$x = \sqrt{a_x^2 + \zeta} r \cos(\theta), \quad (12.108)$$

$$y = \sqrt{a_y^2 + \zeta} r \sin(\theta). \quad (12.109)$$

Note that:

$$r^2 = \bar{u}. \quad (12.110)$$

We define:

$$\bar{r}^2 = u = \frac{x^2}{a_x^2} + \frac{y^2}{a_y^2}. \quad (12.111)$$

Since:

$$\zeta = \left( \frac{\bar{r}^2}{r^2} - 1 \right) \frac{a_x^2 a_y^2}{a_x^2 \sin^2(\theta) + a_y^2 \cos^2(\theta)}, \quad (12.112)$$

equation (12.107) becomes, in the new variables:

$$\langle xE_x \rangle = \frac{a_x a_y}{2\varepsilon_0 q n_q} \int_0^{2\pi} \frac{2a_x^2}{1 + (a_x^2/a_y^2) \tan^2(\theta)} d\theta \int_0^\infty \rho(r^2) r dr \int_r^\infty \rho(\bar{r}^2) \bar{r} d\bar{r}. \quad (12.113)$$

Performing the integral over  $\theta$  gives:

$$\langle xE_x \rangle = \frac{a_x^3 a_y^2}{2\pi \varepsilon_0 q n_q (a_x + a_y)} \int_0^\infty \rho(r^2) 2\pi r dr \int_r^\infty \rho(\bar{r}^2) 2\pi \bar{r} d\bar{r}. \quad (12.114)$$

A given value of  $r$  (or  $\bar{r}$ ) defines an ellipse in the  $x$ - $y$  plane on which the charge density  $\rho$  is constant. Integrating the charge density over  $r$  from zero to infinity gives the charge per unit length of the beam line:

$$qn_q = a_x a_y \int_0^\infty \rho(r^2) 2\pi r dr. \quad (12.115)$$

We define the function  $\bar{n}_q(r)$  such that  $qn_q(r)$  is the charge per unit length within an ellipse corresponding to the given value of  $r$ :

$$qn_q(r) = a_x a_y \int_0^r \rho(\bar{r}^2) 2\pi \bar{r} d\bar{r}. \quad (12.116)$$

It follows that:

$$a_x a_y \int_r^\infty \rho(\bar{r}^2) 2\pi \bar{r} d\bar{r} = q(n_q - \bar{n}_q(r)), \quad (12.117)$$

and:

$$a_x a_y \rho(r^2) 2\pi r = \frac{d}{dr} q \bar{n}_q(r). \quad (12.118)$$

Equation (12.114) can now be written:

$$\langle xE_x \rangle = \frac{qa_x}{2\pi\varepsilon_0 n_q(a_x + a_y)} \int_0^\infty (n_q - \bar{n}_q(r)) \frac{d}{dr} \bar{n}_q(r) dr. \quad (12.119)$$

The final integral is readily performed, to give the result:

$$\langle xE_x \rangle = \frac{qn_q a_x}{4\pi\varepsilon_0 (a_x + a_y)}. \quad (12.120)$$

Note that this is independent of the form of the charge distribution  $\rho(u)$ . Since the parameters  $a_x$  and  $a_y$  define the aspect ratio of the beam cross-section, we can write:

$$\frac{a_x}{a_x + a_y} = \frac{\sigma_x}{\sigma_x + \sigma_y}, \quad (12.121)$$

where  $\sigma_x$  and  $\sigma_y$  are the rms transverse beam sizes. Finally, substituting for  $\langle xE_x \rangle$  in (12.102) gives the envelope equation for a continuous beam with elliptically symmetric charge distribution:

$$\frac{d^2\sigma_x}{ds^2} + k_1 \sigma_x - \frac{\epsilon_x^2}{\sigma_x^3} - \frac{K}{2(\sigma_x + \sigma_y)} = 0, \quad (12.122)$$

where  $K$  is the perveance (12.57). There is of course a similar equation describing the variation of the vertical beam size.

With an elliptically symmetric charge distribution, the envelope equations take the same form as we found for the KV distribution, (12.67) and (12.68). The KV distribution is of course a special case of a distribution with elliptical symmetry: what distinguishes the KV distribution is that the emittance of the beam is constant as the beam is transported along the beam line (assuming that the external fields do not introduce any non-linear effects). In other cases, there will be some emittance growth (or at least some variation in the emittance with distance); but since the KV distribution represents a stationary distribution with the minimum electrical potential energy for a given emittance, we expect that the distribution will ultimately tend towards a KV distribution.

## 12.4 Space-Charge Tune Shifts

Since space-charge forces have defocusing effects on beams in accelerators, we expect optical quantities such as the beta functions and phase advances to depend on the beam current. The effects of space charge on the beta functions are implicit in the envelope equation (12.122). Since the forces are (in general) nonlinear and depend on the beam size, it is not possible to write simple, exact expressions for the changes in the beta functions and phase advances resulting from space-charge effects, even in the case of simple periodic beam lines such as FODO lattices. However, if the space-charge forces are not too strong, so that the changes in beta functions and phase advances are small, then some indication of the size of space-charge effects may be obtained by making some approximations.

An estimate for the tune shifts from space-charge forces in a storage ring can be made by approximating the effects of external fields (i.e. from quadrupole and other magnets in the lattice) as a continuous focusing force. In the case of a continuous beam with a KV distribution, the equation of motion of a particle can be written as:

$$\frac{d^2x}{ds^2} = -k_1 x + \frac{2Kx}{a_x(a_x + a_y)}, \quad (12.123)$$

where  $k_1$  represents the focusing from external fields,  $K$  is the perveance (characterising the strength of the space-charge forces), and  $a_x$  and  $a_y$  are the horizontal and vertical half-axes respectively of the elliptical distribution in the  $x$ - $y$  plane. The co-ordinate  $x$  of a particle can be written in terms of action-angle variables  $J_x$  and  $\phi_x$ :

$$x = \sqrt{2\beta_x J_x} \cos(\phi_x), \quad (12.124)$$

where  $\beta_x$  is the horizontal beta function. In the smooth-focusing approximation, the beta function is treated as constant; then, combining (12.123) with (12.124) gives:

$$\frac{d\phi_x}{ds} \approx \sqrt{k_1 - \frac{2K}{a_x(a_x + a_y)}}. \quad (12.125)$$

Let  $\beta_{x0}$  represent the beta function in the absence of space-charge forces, i.e. in the limit of low beam current. From (4.50), it follows that (again using the smooth-focusing approximation):

$$k_1 \approx \frac{1}{\beta_{x0}^2}. \quad (12.126)$$

The horizontal tune  $\nu_x$  is the change in the horizontal angle variable over one turn of the storage ring, divided by  $2\pi$ :

$$\nu_x = \frac{1}{2\pi} \int_0^{C_0} \frac{d\phi_x}{ds} ds \approx \frac{1}{2\pi} \int_0^{C_0} \sqrt{\frac{1}{\beta_{x0}^2} - \frac{2K}{a_x(a_x + a_y)}} ds. \quad (12.127)$$

If the space-charge forces are small, we can expand the square root to first order in  $K$ , giving:

$$\Delta\nu_x \approx -\frac{K}{2\pi} \int_0^{C_0} \frac{\beta_{x0}}{a_x(a_x + a_y)} ds, \quad (12.128)$$

where:

$$\nu_x = \nu_{x0} + \Delta\nu_x. \quad (12.129)$$

The horizontal tune in the absence of space-charge forces (i.e. in the limit of low beam current) is:

$$\nu_{x0} = \frac{1}{2\pi} \int_0^{C_0} \frac{ds}{\beta_{x0}}. \quad (12.130)$$

Equation (12.128) for the horizontal space-charge tune shift can also be obtained by treating the space-charge force as a focusing error. Using (6.33) the change in horizontal tune arising from a focusing error in a storage ring can be written as:

$$\Delta\nu_x \approx \frac{1}{4\pi} \Delta K \beta_{x0}, \quad (12.131)$$

where  $\Delta K$  is the integrated quadrupole strength of the focusing error (not to be confused with the permeance  $K$ ), and  $\beta_{x0}$  is the horizontal beta function at the location of the error. In the case of a continuous beam with a KV distribution, the space-charge forces may be represented by a horizontal focusing strength:

$$k_{1x,sc} = -\frac{2K}{a_x(a_x + a_y)}, \quad (12.132)$$

where  $K$  is the permeance, and  $a_x$  and  $a_y$  are (respectively) the horizontal and vertical radii of the beam in co-ordinate space. If the space-charge forces are not too strong, then their effect on the tune can be obtained from (12.131), where the focusing error  $\Delta K$  arises from the space-charge focusing at every point around the ring. This gives the result:

$$\Delta\nu_x \approx -\frac{K}{2\pi} \int_0^{C_0} \frac{\beta_{x0}}{a_x(a_x + a_y)} ds, \quad (12.133)$$

which is the same as (12.128).

It is sometimes convenient to express the charge distribution in terms of the rms beam sizes rather than the half-axes. Using  $a_x = 2\sigma_x = 2\sqrt{\beta_x \epsilon_x}$ , and similarly for  $a_y$ , the horizontal tune shift can be written:

$$\Delta\nu_x \approx -\frac{K}{8\pi} \int_0^{C_0} \frac{\beta_{x0}}{\sigma_x(\sigma_x + \sigma_y)} ds = -\frac{K}{8\pi\epsilon_x} \int_0^{C_0} \frac{ds}{1 + \sqrt{\beta_{y0}\epsilon_y/\beta_{x0}\epsilon_x}}. \quad (12.134)$$

The vertical tune shift is:

$$\Delta\nu_y \approx -\frac{K}{8\pi} \int_0^{C_0} \frac{\beta_{y0}}{\sigma_y(\sigma_x + \sigma_y)} ds = -\frac{K}{8\pi\epsilon_y} \int_0^{C_0} \frac{ds}{1 + \sqrt{\beta_{x0}\epsilon_x/\beta_{y0}\epsilon_y}}. \quad (12.135)$$

Treating the space-charge force as a focusing error allows us to write an expression for the change in the beta functions arising from space-charge forces. We again assume a continuous beam with a KV distribution. From (6.34), the change in the horizontal beta function resulting from a (small) focusing error  $\Delta K$  can be written:

$$\frac{\Delta\beta_x}{\beta_{x0}} \approx -2\pi\Delta\nu_{x0} \cot(2\pi\nu_{x0}). \quad (12.136)$$

The change in the vertical beta function is given by a similar expression.

Strictly speaking, equations (12.128) and (12.136) for the tune shift and change in beta function resulting from space-charge forces are valid only where the space-charge forces are linear (i.e. for a KV distribution) and small enough that the effects may be regarded as a perturbation. In more general cases, the lattice parameters (including betatron tunes and beta functions) are best obtained by numerical solutions of the equations of motion. However, in the case of a beam with a Gaussian transverse profile, some indication of the space-charge tune shift can be made by comparing the charge density at the centre of a KV beam with the charge density at the centre of a Gaussian beam. In the case of a KV beam, the charge per unit volume within the beam envelope is:

$$\rho_{KV} = \frac{qn_q}{\pi a_x a_y} = \frac{qn_q}{4\pi\sigma_x\sigma_y}, \quad (12.137)$$

where we have used  $a_x = 2\sigma_x$  and  $a_y = 2\sigma_y$ . For a Gaussian beam, the charge per unit volume is:

$$\rho_G = \frac{qn_q}{2\pi\sigma_x\sigma_y} e^{-x^2/2\sigma_x^2} e^{-y^2/2\sigma_y^2}. \quad (12.138)$$

At the centre of the beam ( $x = y = 0$ ), the charge density in a Gaussian beam is twice the charge density in a KV beam with the same rms beam size

and the same charge per unit length. The tune shifts in a Gaussian beam are therefore expected to be twice as large as the tune shifts in a KV beam. It is also possible to estimate the space-charge tune shifts in a Gaussian beam by making a linear approximation to the fields within a Gaussian charge distribution. For ultra-relativistic beams, we need consider only a two-dimensional Gaussian distribution. An expression in closed form for the electric field in such a case has been derived by Bassetti and Erskine [Bassetti and Erskine (1980)].

At this point, we should make a clear distinction between three separate but related quantities: the *incoherent tune shift*, the *coherent tune shift* and the *tune spread*. The expression in (12.128) is said to give the incoherent tune shift from space-charge forces, since the quantity  $\Delta\nu_x$  in this case is the change in betatron tune for each individual particle in the beam. The betatron tunes of individual particles can be different from each other, and different also from the tune corresponding to the betatron oscillations performed by an entire bunch. Consider, for example, a particle at the centroid of a bunch. The motion of this particle will not be affected by space-charge forces, so its motion will be determined by the external fields from beam line components such as magnets and rf cavities. If the entire bunch is deflected from the closed orbit in a storage ring, then as long as a particle at the centroid of the bunch remains at the centroid, the tune of the bunch oscillations will be the same as the tune in the absence of space-charge forces. A change in the tune of the oscillations of an entire bunch is known as a *coherent tune shift*. If we neglect image charges induced in the vacuum chamber, space-charge forces lead only to incoherent tune shifts, and not to coherent tune shifts. We shall discuss coherent tune shifts from image charges shortly.

Measurement of the tunes in a storage ring is usually achieved by resonant excitation of *coherent* motion of particles within a bunch. The common procedure is to use an oscillating field at one location in the storage ring to excite betatron or synchrotron oscillations in the beam. When the frequency of the field matches the fractional part of the betatron (or synchrotron) frequency, the motion of the bunch resonates with the field and the beam oscillations are driven to large amplitude. The motion of the bunch can be detected using electrodes within the beam pipe that are sensitive to the position of the centroid of the bunch. This technique can be used to investigate the overall behaviour of the bunch, rather than the behaviour of individual particles; the result is a measurement of the *coherent tune*.

For a continuous beam with a KV distribution, the space-charge forces vary linearly with distance from the centre of the beam: the incoherent tune shift is therefore the same for all particles. More generally, for bunched beams or beams with non-KV transverse distributions, the tune shift of any given particle will depend on the amplitude of the betatron oscillations performed by that particle, and on the local charge density at the position of the particle with the bunch. Each particle within the bunch will then have a different tune: the range of tunes is described by the *tune spread*. Consider, for example, a bunch with a Gaussian transverse and longitudinal profile. The maximum horizontal and vertical tune shifts for any particle will be given by twice the values of (12.134) and (12.135), with the perveance corresponding to the peak current in the bunch: this is because the charge density at the centre of a Gaussian beam is twice the charge density at the centre of a KV beam with the same rms beam sizes. The minimum tune shifts for any particle will be close to zero, since particles at the extreme ends of the bunch see very low longitudinal charge density. Therefore, the horizontal and vertical tune spreads in a Gaussian bunched beam will be:

$$\Delta\nu_{x,\text{spread}} \approx -\frac{K}{4\pi} \int_0^{C_0} \frac{\beta_{x0}}{\sigma_x(\sigma_x + \sigma_y)} ds, \quad (12.139)$$

$$\Delta\nu_{y,\text{spread}} \approx -\frac{K}{4\pi} \int_0^{C_0} \frac{\beta_{y0}}{\sigma_y(\sigma_x + \sigma_y)} ds. \quad (12.140)$$

As we have mentioned, if image charges in the walls of the vacuum chamber can be neglected (as we have assumed so far), there will be no coherent tune shifts from space-charge forces. However, image charges lead to additional ‘external’ forces affecting the motion of the bunch as a whole, and can therefore cause coherent tune shifts. Image charges (and currents) can also contribute to the incoherent space-charge tune shifts. In general, the tune shifts resulting from image charges are known as *Laslett tune shifts* [Laslett (1963)], and depend on the geometry of the vacuum chamber and the material from which the chamber is made. The effects of image charges can be understood by considering some simple examples: in the remainder of this section we shall discuss the case of a continuous beam in a vacuum chamber consisting of two infinitely wide, perfectly conducting plates. By considering the fields resulting from the image charges, we shall derive expressions for the incoherent and the coherent space-charge tune shifts.

Let us first consider the incoherent tune shifts. For this case, we assume that the beam is centred on the line  $x = y = 0$ , and that the walls of the vacuum chamber are defined by the planes  $y = \pm h$ . We assume that the

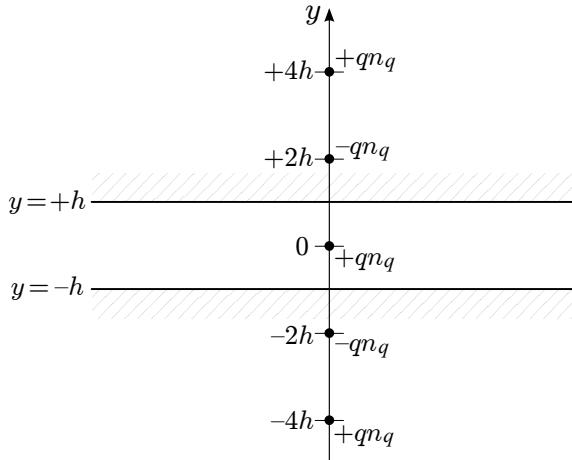


Fig. 12.2 Image charges created by a line charge between two parallel, perfectly conducting plates. The beam is mid-way between the plates, which are located at  $y = \pm h$ .

radius of the beam is small compared to the height of the vacuum chamber, so that it can be considered, from the point of view of the image charges, as a line charge of uniform density. From the discussion in Section 1.1.2, we know that if the walls of the vacuum chamber have perfect conductivity, the electric field lines at the surface of the chamber wall must be perpendicular to the wall. For a line of charge at a distance  $h$  from the conductor, the boundary conditions are satisfied if the electric field takes the form of that generated by two parallel lines of charge on opposite sides of the surface of the conductor, and with charge density equal in magnitude but opposite in sign. Where there are two plates present, then the field is the same as the field that would be generated by an infinite series of image charges, as indicated in Fig. 12.2.

Let us now calculate the contribution to the electric field from the image charges, at a point  $x = 0, y = y_0$ . The contribution to the field from the line of charge itself is the ‘direct’ space-charge field: we have already discussed the effects of this contribution to the field. At a distance  $r$  from an infinite line of charge with charge density  $qn_q$ , the electric field strength is:

$$E = \frac{qn_q}{2\pi\epsilon_0 r}. \quad (12.141)$$

Using this expression for the field generated by the image charges, we find

that the electric field at  $x = 0, y = y_0$  is given by:

$$E_{y,\text{image}} = \sum_{n=1}^{\infty} (-1)^n \frac{qn_q}{2\pi\varepsilon_0(2nh + y_0)} - \sum_{n=1}^{\infty} (-1)^n \frac{qn_q}{2\pi\varepsilon_0(2nh - y_0)}. \quad (12.142)$$

The first summation in (12.142) comes from the image charges in the plate at  $y = -h$ , and the second summation comes from the image charges in the plate at  $y = +h$ . In the case that  $y_0 \ll h$ , (12.142) can be written:

$$E_{y,\text{image}} \approx -\frac{qn_q y_0}{4\pi\varepsilon_0 h^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = \frac{qn_q y_0}{4\pi\varepsilon_0 h^2} \frac{\pi^2}{12}. \quad (12.143)$$

By symmetry, the vertical component of the electric field from the image charges must be (approximately) independent of the horizontal position. Therefore, in the region of the beam, the vertical component of the electric field can be written as a function of position:

$$E_{y,\text{image}} \approx \frac{qn_q}{4\pi\varepsilon_0 h^2} \frac{\pi^2}{12} y. \quad (12.144)$$

Since the electric field from the image charges must satisfy (in the region between the conducting plates):

$$\nabla \cdot \mathbf{E}_{\text{image}} = 0, \quad (12.145)$$

we can immediately write down an approximate expression for the horizontal component of the electric field from the image charges, in the region of the beam:

$$E_{x,\text{image}} \approx -\frac{qn_q}{4\pi\varepsilon_0 h^2} \frac{\pi^2}{12} x. \quad (12.146)$$

The vertical field from the image charges leads to a defocusing force, while the horizontal field leads to a focusing force. The forces result in incoherent tune shifts, since the forces apply to each particle individually. By analogy with the focusing strength in a magnetic quadrupole (1.45), we define the focusing strength of the image charges:

$$k_1 = -\frac{q}{\beta_0 c P_0} \frac{\partial E_x}{\partial x}. \quad (12.147)$$

We include a factor  $1/\beta_0 c$  since the force on a particle in a magnetic field depends on the velocity of the particle, as well as on the charge of the particle and the field; in an electric field, the force depends only on the charge and the field. Using (12.146) in (12.147) gives for the incoherent focusing:

$$k_{1,\text{image}} = \frac{\pi^2}{24} \frac{\gamma_0^2 K}{h^2}, \quad (12.148)$$

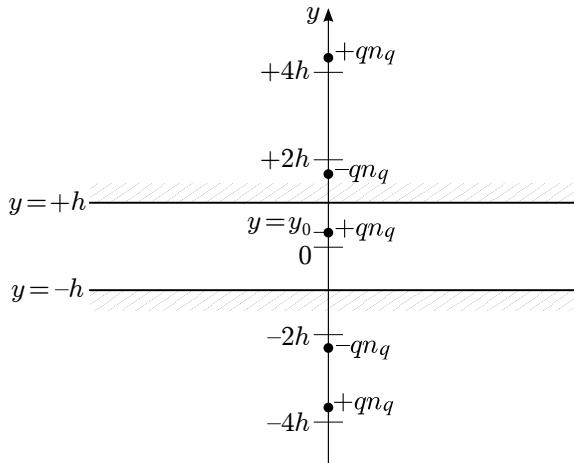


Fig. 12.3 Image charges created by a line charge between two parallel, perfectly conducting plates. The beam is displaced vertical by  $y_0$  from the mid-plane between the plates.

where  $K$  is the permeance (12.57). Compared with the direct space-charge focusing (12.132), we see that because of the dependence on the relativistic factor  $\gamma_0$ , the image charges are potentially more significant than the direct space-charge forces at high energy. The relationship between the two contributions also depends, however, on the beam size and aperture of the vacuum chamber.

Now let us consider the coherent tune shift from image charges. In this case, we consider a displacement of the entire beam, so that the centre of the beam is at  $x = 0, y = y_0$ . Again, we calculate the electric field at  $x = 0, y = y_0$ : in contrast to the previous case (of incoherent tune shifts), every particle in the beam sees the same force from the image charges. Therefore, all the particles in the beam will move together (coherently) in response to the force from the image charges. The vertical field is different from the incoherent case, because the image charges move in response to movements of the beam; see Fig. 12.3. We find that, when the beam is at any location  $(x, y)$  between the plates, the vertical field from the image charges at the position of the beam is:

$$E_{y,\text{image}} \approx \frac{qn_q}{4\pi\epsilon_0 h^2} \frac{\pi^2}{16} y. \quad (12.149)$$

By symmetry, the horizontal field from the image charges at the position

of the beam is zero:

$$E_{x,\text{image}} = 0. \quad (12.150)$$

Defining the vertical focusing strength in this case:

$$k_{1y} = -\frac{q}{\beta_0 c P_0} \frac{\partial E_y}{\partial y}, \quad (12.151)$$

we find from (12.149) that the coherent focusing is described by separate horizontal and vertical focusing strengths:

$$k_{1x,\text{image}} = 0, \quad (12.152)$$

$$k_{1y,\text{image}} \approx -\frac{\pi^2}{32} \frac{\gamma_0^2 K}{h^2}. \quad (12.153)$$

The fact that all particles in the beam respond together to the image forces in this case means that the resulting vertical tune shift will be coherent: if the beam is ‘kicked’ at one location, the measured betatron oscillations will have a frequency determined by the focusing in the ring, including the focusing (12.153) from the image charges.

Using the results for the focusing from the direct space-charge forces and the forces from the image charges, we can summarise the incoherent and coherent tune shifts as follows. The results apply to the case of a KV beam in a storage ring, with a vacuum chamber consisting of two horizontal, perfectly conducting parallel plates. The incoherent tune shifts are given by direct and image charge contributions:

$$\Delta\nu_{x,\text{incoherent}} = -\frac{1}{4\pi} \int_0^{C_0} \frac{2\beta_{x0}K}{a_x(a_x + a_y)} - \frac{\pi^2}{24} \frac{\gamma_0^2}{h^2} \beta_{x0}K ds, \quad (12.154)$$

$$\Delta\nu_{y,\text{incoherent}} = -\frac{1}{4\pi} \int_0^{C_0} \frac{2\beta_{y0}K}{a_y(a_x + a_y)} + \frac{\pi^2}{24} \frac{\gamma_0^2}{h^2} \beta_{y0}K ds. \quad (12.155)$$

The coherent tune shift has only an image charge contribution:

$$\Delta\nu_{y,\text{coherent}} = -\frac{1}{4\pi} \int_0^{C_0} \frac{\pi^2}{32} \frac{\gamma_0^2}{h^2} \beta_{y0}K ds. \quad (12.156)$$

In our analysis of image charges, we have taken into account only the effects of electric fields, not the effects of magnetic fields. If the vacuum chamber is constructed from a non-magnetic material, then the chamber will have negligible impact on the magnetic fields around the beam. However, when the beam passes through magnetic components in the beam line, then the magnetic materials used in the construction of these components can have a significant influence on the magnetic field generated by

the beam. The result is the presence of additional terms in the expressions for the tune shifts. The expressions given above will also be modified for different chamber geometries: in particular, for chambers with finite horizontal width, there will be some coherent horizontal tune shift as well as coherent vertical tune shift. The tune shifts occurring in certain standard geometries (for example, chambers with circular or elliptical cross-section, and including the presence of dipole or quadrupole magnets) are described by various *Laslett coefficients* [Laslett (1963)].

Space-charge forces lead (in general) to coherent and incoherent tune shifts. Coherent tune shifts can be compensated in an accelerator by adjusting quadrupole strengths: incoherent tune shifts cannot. In extreme cases, the tune spread resulting from space-charge forces can limit the current of individual bunches within a storage ring. For example, if the charge within a bunch is large enough that the tunes of some particles are pushed onto integer or half-integer resonances, then these particles will quickly be lost from the storage ring.

## 12.5 Matching a Continuous Beam to a Solenoid Field

A case of particular interest is that of beam transport in a solenoid. Solenoids provide constant transverse focusing: it was shown in Chapter 5 that it is possible to construct a distribution for which the transverse part remains invariant as the particles are transported along a given solenoid. We now show how the matched distribution is modified if space-charge forces are taken into account. We shall consider only the case of a continuous beam with a KV distribution, so that the space-charge forces are linear.

It is convenient to work with the canonical momenta  $p_x$  and  $p_y$ , defined by:

$$p_x = \frac{\gamma_0 m \dot{x}}{P_0} + \frac{q A_x}{P_0}, \quad (12.157)$$

$$p_y = \frac{\gamma_0 m \dot{y}}{P_0} + \frac{q A_y}{P_0}, \quad (12.158)$$

where, in a solenoid field of strength  $B_0$ , the transverse components of the vector potential are:

$$A_x = -\frac{B_0}{2} y, \quad (12.159)$$

$$A_y = \frac{B_0}{2} x. \quad (12.160)$$

In the absence of space-charge forces, the equations of motion in a solenoid are:

$$\frac{dx}{ds} = p_x + \omega_0 y, \quad (12.161)$$

$$\frac{dp_x}{ds} = -\omega_0^2 x + \frac{1}{2} \omega_0 p_y, \quad (12.162)$$

$$\frac{dy}{ds} = p_y - \omega_0 x, \quad (12.163)$$

$$\frac{dp_y}{ds} = -\omega_0^2 y - \frac{1}{2} \omega_0 p_x, \quad (12.164)$$

where:

$$\omega_0 = \frac{q}{P_0} \frac{B_0}{2}. \quad (12.165)$$

$q$  is the particle charge,  $P_0$  the reference momentum (equal to the particle momentum in this case, since we are neglecting the longitudinal dynamics), and  $B_0$  is the solenoid field. Note that  $x' = dx/ds$  and  $y' = dy/ds$  are equal to the respective components of the *mechanical* momentum, scaled by the reference momentum. The presence of space-charge forces introduces additional terms to the equations of motion. For a beam with a KV distribution and horizontal and vertical radii  $a_x$  and  $a_y$  respectively, (12.162) and (12.164) become:

$$\frac{dp_x}{ds} = -\omega_0^2 x + \frac{2K}{(a_x + a_y)} \frac{x}{a_x} + \frac{1}{2} \omega_0 p_y, \quad (12.166)$$

$$\frac{dp_y}{ds} = -\omega_0^2 y + \frac{2K}{(a_x + a_y)} \frac{y}{a_y} - \frac{1}{2} \omega_0 p_x. \quad (12.167)$$

From (5.161), we see that in the absence of space-charge forces, a matched beam in a solenoid always has circular cross-section. Therefore, we consider only the case with circular cross-section, and set  $a_x = a_y = a$ . If  $a$  is constant, the solution to the equations of motion can be expressed as a transfer matrix:

$$R_{\text{sol}} = \begin{pmatrix} c_0 c_1 & \frac{1}{\omega_1} c_0 s_1 & s_0 c_1 & \frac{1}{\omega_1} s_0 s_1 \\ -\omega_1 c_0 s_1 & c_0 c_1 & -\omega_1 s_0 s_1 & s_0 c_1 \\ -s_0 c_1 & -\frac{1}{\omega_1} s_0 s_1 & c_0 c_1 & \frac{1}{\omega_1} c_0 s_1 \\ \omega_1 s_0 s_1 & -s_0 c_1 & -\omega_1 c_0 s_1 & c_0 c_1 \end{pmatrix}, \quad (12.168)$$

where:

$$c_0 = \cos(\omega_0 L), \quad (12.169)$$

$$s_0 = \sin(\omega_0 L), \quad (12.170)$$

$$c_1 = \cos(\omega_1 L), \quad (12.171)$$

$$s_1 = \sin(\omega_1 L). \quad (12.172)$$

$L$  is the length of the solenoid, and:

$$\omega_1 = \sqrt{\omega_0^2 - \frac{K}{a^2}}. \quad (12.173)$$

Taking the transfer matrix (12.168) as the transfer matrix for one periodic section of a beam line, we can look for a matched distribution using the techniques of Section 5.2. The matched distribution will be expressed in terms of a  $4 \times 4$  matrix  $\Sigma$  with elements  $\langle x_i x_j \rangle$ .

We have assumed in deriving the transfer matrix (12.168) that there exists a KV distribution that is invariant under transport along the solenoid. Strictly speaking, the KV distribution (12.4) is defined in the absence of coupling, which excludes the case of a beam in a solenoid. However, we can define a KV distribution at the entrance of a solenoid, and look for the conditions for the distribution to remain invariant as it travels along the solenoid. (Recall that the transfer map for the fringe field of a solenoid, in the approximation that the fringe field has zero length, is the identity.) Since the initial distribution must be uncoupled, the elements of the sigma matrix outside the  $2 \times 2$  block diagonals ( $\langle xp_x \rangle$ ,  $\langle xy \rangle$  etc.) must all be equal to zero. From (5.161), this means that we must have:

$$\epsilon_I = \epsilon_{II} = \epsilon, \quad (12.174)$$

i.e. the normal mode emittances of the beam must be equal. For a KV distribution, it is usual to specify the beam size in terms of the beam radii  $a_x$  and  $a_y$ . Using (12.15) with  $m = 1$ , we have for a KV distribution with circular cross-section of radius  $a$ :

$$\langle x^2 \rangle = \langle y^2 \rangle = \frac{a^2}{4} = \beta\epsilon, \quad (12.175)$$

where  $\beta$  is the beta function (which must be the same in the horizontal direction as in the vertical direction). Following similar reasoning to that leading to (12.15), we find that for a KV distribution with circular cross-section, the mean square transverse momenta are:

$$\langle p_x^2 \rangle = \langle p_y^2 \rangle = \frac{a^2}{4\beta^2} = \frac{b^2}{4}. \quad (12.176)$$

From (5.161), an invariant distribution in a solenoid has  $\langle xp_x \rangle = \langle yp_y \rangle = 0$ ; we therefore assume this condition for the KV distribution at the entrance to the solenoid. Then, the sigma matrix describing the KV distribution is:

$$\Sigma = \begin{pmatrix} \frac{1}{4}a^2 & 0 & 0 & 0 \\ 0 & \frac{1}{4}b^2 & 0 & 0 \\ 0 & 0 & \frac{1}{4}a^2 & 0 \\ 0 & 0 & 0 & \frac{1}{4}b^2 \end{pmatrix}. \quad (12.177)$$

Applying the analysis of Section 5.4.1 to the transfer matrix (12.168) shows that a matched distribution for a solenoid including space-charge forces from a uniform (co-ordinate space) charge distribution has the general form:

$$\Sigma = \begin{pmatrix} \frac{1}{2\omega_1}(\epsilon_I + \epsilon_{II}) & 0 & 0 & \frac{1}{2}(\epsilon_I - \epsilon_{II}) \\ 0 & \frac{\omega_1}{2}(\epsilon_I + \epsilon_{II}) & -\frac{1}{2}(\epsilon_I - \epsilon_{II}) & 0 \\ 0 & -\frac{1}{2}(\epsilon_I - \epsilon_{II}) & \frac{1}{2\omega_1}(\epsilon_I + \epsilon_{II}) & 0 \\ \frac{1}{2}(\epsilon_I - \epsilon_{II}) & 0 & 0 & \frac{\omega_1}{2}(\epsilon_I + \epsilon_{II}) \end{pmatrix}. \quad (12.178)$$

Comparing (12.177) and (12.178) we see that the KV distribution will be invariant under transport along the solenoid if  $\epsilon_I = \epsilon_{II} = \epsilon$ , and:

$$\omega_1 = \frac{4\epsilon}{a^2} = \frac{b^2}{4\epsilon}. \quad (12.179)$$

Using (12.173) and (12.165), a KV distribution with circular cross-section will be invariant under transport along a solenoid with field strength  $B_0$  if:

$$\omega_0 = \frac{q}{P_0} \frac{B_0}{2} = \sqrt{\frac{16\epsilon^2}{a^4} + \frac{K}{a^2}}, \quad (12.180)$$

where  $K$  is the perveance,  $a$  the beam radius,  $\epsilon$  the emittance and  $\beta$  the beta function. Using (12.175), the matching condition can be expressed in terms of the beta function:

$$\frac{q}{P_0} \frac{B_0}{2} = \sqrt{\frac{1}{\beta^2} + \frac{K}{a^2}}. \quad (12.181)$$

It should be emphasised that there are many types of distribution (uniform, Gaussian, parabolic, etc.) that can be matched to the transfer matrix (12.168); however, this transfer matrix is valid only for particle distributions that have a uniform charge density in co-ordinate space. For consistency, therefore, a distribution can only be said to be properly matched to the solenoid if it is invariant under transformations given by the transfer matrix (12.168), and if it also has uniform charge density in co-ordinate space. A KV distribution with beta function, beam radius and perveance satisfying (12.181) provides a solution satisfying both the condition for invariance under the transfer matrix (12.168) and the condition on uniform charge density. A real beam is unlikely to be described exactly by a KV distribution; in general, the space-charge forces will be nonlinear, and it will not be possible to find a solenoid strength such that the distribution is invariant under transport along the solenoid. Furthermore, we have neglected effects associated with the energy spread of the beam. However, if the energy spread is small, and if space-charge effects are not too strong, (12.181) may be used to give an estimate of the solenoid field needed for (approximate) invariant transport of a beam in the presence of space-charge forces.

## 12.6 Longitudinal Dynamics with Space Charge

So far, we have considered space-charge effects for the case of a continuous beam, where the beam current is independent of longitudinal position along the beam line. However, particles in accelerators are often collected into bunches, with gaps of zero (or almost zero) current between bunches. In a continuous beam, we could neglect longitudinal forces: if the transverse distribution is independent of longitudinal position, then the space-charge force has no longitudinal component (the contributions to this component from all particles in the beam cancel exactly). Even where there is some longitudinal variation in transverse beam size, the longitudinal component of the space-charge force will usually be small. In particular, the relativistic ‘flattening’ of the fields around a particle towards a plane perpendicular to the particle trajectory means that, for high energy beams, particles will only experience forces from each other when their longitudinal separation is small. Variations in transverse beam size therefore have no significant longitudinal effect unless the variations occur over short distances.

In bunched beams, however, the longitudinal symmetry is broken: the system is not invariant under longitudinal translations. Space-charge forces can then have significant longitudinal effects. In this section, we shall consider the case of a bunched beam in a synchrotron storage ring, where the longitudinal focusing from rf cavities naturally leads to bunching of the beam. We shall show that longitudinal space-charge forces lead to a modification of the focusing potential, which in turn leads to a change in the longitudinal profile of an equilibrium (stationary) particle distribution within a bunch, depending on the number of particles within the bunch. If the bunch contains a sufficient number of particles, then the motion of particles within the storage ring can become unstable. Additionally, we expect that there will be a change in the transverse effects of space charge. However, because of the relativistic flattening of the fields around particles in a high energy beam, and because synchrotron motion in a storage ring is usually slow compared to the betatron motion, the transverse effects of space charge in bunched beams can be approximated by treating the beam as a continuous beam, but with a current that depends on the position within the bunch. This model will be discussed further towards the end of this section.

The equations of motion for longitudinal dynamics in a synchrotron storage ring were discussed in Chapter 5. Consider a single particle within a bunch in a synchrotron storage ring. The rate of change of the longitudinal

co-ordinate  $z$  of the particle is related to the energy deviation  $\delta$  by the phase slip factor  $\eta_p$ . The rate of change of the energy deviation depends on the rf voltage (and the phase at which the particle arrives at the rf cavities), and on the rate of energy loss through synchrotron radiation. Taking space-charge forces into account, there is an additional term in the equation for the rate of change of energy loss, representing the longitudinal component of the electric field within the bunch. The equations of motion for the longitudinal dynamics (5.41) and (5.43) become:

$$\frac{dz}{ds} = -\eta_p \delta, \quad (12.182)$$

$$\frac{d\delta}{ds} = \frac{qV_{\text{rf}}}{cP_0C_0} \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}} z}{c}\right) - \frac{U_0}{cP_0C_0} + \frac{qV(z)}{cP_0C_0}, \quad (12.183)$$

where  $V_{\text{rf}}$  is the amplitude of the rf voltage,  $\phi_{\text{rf}}$  is a constant phase,  $P_0$  is the reference momentum,  $C_0$  is the circumference of the storage ring, and  $U_0$  is the energy loss per turn from synchrotron radiation. The potential difference  $V(z)$  represents the integral (over one turn of the storage ring) of the longitudinal component of the space-charge force:

$$V(z) = \int_{s_0}^{s_0+C_0} E_s(z) ds. \quad (12.184)$$

We assume that the space-charge force is a function of longitudinal position within the bunch, and that the variation in longitudinal co-ordinate of particles within the bunch is small over one period of the lattice. To simplify the analysis, we neglect the dependence of the energy loss per turn  $U_0$  on the energy of the particle: this dependence leads to synchrotron radiation damping, as discussed in Chapter 7. Radiation damping is usually a slow effect on the time scale of the synchrotron oscillations.

The analysis of the longitudinal dynamics now proceeds along similar lines to that in Section 5.1.3. Taking the derivative of (12.182) with respect to  $s$  and substituting for  $d\delta/ds$  from (12.183) gives:

$$\frac{d^2z}{ds^2} = -\frac{qV_{\text{rf}}}{cP_0C_0} \eta_p \sin\left(\phi_{\text{rf}} - \frac{\omega_{\text{rf}} z}{c}\right) + \eta_p \frac{U_0}{cP_0C_0} - \eta_p \frac{qV(z)}{cP_0C_0}. \quad (12.185)$$

If the rf phase is set so that  $\phi_{\text{rf}} = \phi_s$ , where:

$$\sin(\phi_s) = \frac{U_0 + qV(0)}{qV_{\text{rf}}}, \quad (12.186)$$

then expanding (12.185) to first order in the dynamical variable  $z$  gives:

$$\frac{d^2z}{ds^2} = -k_z^2 z, \quad (12.187)$$

where:

$$k_z^2 = -\frac{qV_{\text{rf}} \cos(\phi_s)}{cP_0} \frac{\omega_{\text{rf}}}{cC_0} \eta_p + \eta_p \frac{q}{cP_0 C_0} \left. \frac{\partial V(z)}{\partial z} \right|_{z=0}. \quad (12.188)$$

The longitudinal motion of a particle is (approximately) that of a harmonic oscillator. The synchrotron tune is the number of synchrotron oscillations completed per revolution:

$$\nu_z = \frac{k_z C_0}{2\pi}. \quad (12.189)$$

One effect of space-charge forces will be to shift the synchrotron tune from the value in the limit of low bunch charge, given by (5.49). To calculate the change in tune, we need to know the voltage  $V(z)$ : the voltage depends on the distribution of charge within the bunch.

In an electron storage ring, the equilibrium distribution of particles within a bunch will be determined by various lattice parameters (including the rf voltage and the phase slip factor) and synchrotron radiation effects. Space-charge forces will lead to some change in the equilibrium distribution. Without including radiation effects in our analysis, we cannot completely determine the equilibrium distribution; however, we can calculate the shape of an invariant distribution, that is, a distribution of particles in phase space that remains unchanged as the particles move around the storage ring. The equilibrium distribution will be an invariant distribution covering a certain area of phase space. We can estimate the area from considering synchrotron radiation effects separately. In other words, we determine the shape of the equilibrium distribution from space-charge effects, and the size of the distribution from synchrotron radiation effects. Taking this approach avoids the need to solve equations that include simultaneously space-charge and synchrotron radiation effects: the justification for having some confidence in the validity of the results will be made as we develop the arguments in detail.

The first step is to find an invariant distribution in longitudinal phase space; that is, a distribution of particles that remains unchanged as a bunch moves around the storage ring, and individual particles within the bunch perform synchrotron oscillations. The invariant distribution can be found as follows. First, note that the equations of motion (12.182) and (12.183) may be obtained from the Hamiltonian:

$$H = -\frac{qV_{\text{rf}}}{\omega_{\text{rf}} P_0 C_0} \cos\left(\phi_s - \frac{\omega_{\text{rf}} z}{c}\right) + \frac{U_0 z}{cP_0 C_0} - \frac{q}{cP_0 C_0} \int_{-\infty}^z V(z') dz' - \frac{\eta_p \delta^2}{2}. \quad (12.190)$$

Space-charge effects are included through the term involving an integral over  $V(z)$ . Since the Hamiltonian has no explicit dependence on the independent variable  $s$ ,  $H$  is a constant of the motion. Therefore, any function of  $H$  only is also a constant of the motion. As a result, any distribution that can be written purely as a function of  $H$  will be an invariant distribution. For example, the distribution:

$$\Psi(z, \delta) = \hat{\Psi} \exp\left(\frac{H(z, \delta) - H_0}{\langle H \rangle}\right) \quad (12.191)$$

(for constants  $\hat{\Psi}$ ,  $H_0$  and  $\langle H \rangle$ ) is an invariant distribution. If we set  $\langle H \rangle = \eta_p \sigma_\delta^2$  and  $H_0 = H(z = 0, \delta = 0)$ , then the invariant distribution is:

$$\Psi(z, \delta) = \hat{\Psi} \exp(-f(z)) \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right), \quad (12.192)$$

where:

$$\begin{aligned} f(z) = & \frac{qV_{\text{rf}}}{\eta_p \sigma_\delta^2 \omega_{\text{rf}} P_0 C_0} \left( \cos\left(\phi_s - \frac{\omega_{\text{rf}} z}{c}\right) - \cos(\phi_s) \right) - \frac{U_0 z}{\eta_p \sigma_\delta^2 c P_0 C_0} \\ & + \frac{q}{\eta_p \sigma_\delta^2 c P_0 C_0} \int_0^z V(z') dz'. \end{aligned} \quad (12.193)$$

Note that  $V(z)$  will affect the longitudinal distribution (the dependence of  $\Psi(z, \delta)$  on  $z$ ), but not the momentum distribution (the dependence of  $\Psi(z, \delta)$  on  $\delta$ ).  $V(z)$  in turn depends on the longitudinal distribution. It is convenient to express this dependence as a dependence on the local beam current  $I(z)$ , given by:

$$I(z) = \beta_0 c q n_q = \beta_0 c q \int_{-\infty}^{\infty} \Psi(z, \delta) d\delta. \quad (12.194)$$

This can be written:

$$I(z) = I_0 e^{-f(z)}, \quad (12.195)$$

where  $I_0$  is the current at  $z = 0$ , and  $f(z)$  is given by (12.193). Equation (12.195) is known as the *Haissinski equation* [Haissinski (1973)]. Since  $f(z)$  depends on the current  $I(z)$  (through the voltage  $V(z)$ ), equation (12.195) must usually be solved numerically to find the current  $I(z)$ . To find a solution, the dependence of the voltage  $V(z)$  on the current  $I(z)$  must be known explicitly: we shall shortly find this dependence for the case of space-charge forces. Before proceeding to this step, however, let us consider briefly the case of an invariant longitudinal phase space distribution in the absence of space-charge forces. If we assume that the number of particles per unit phase space area is significant only for  $\omega_{\text{rf}} z / c \ll 1$ ,

then we can write the distribution in an approximate form by expanding the Hamiltonian (12.190) to second order in  $z$ :

$$H(z, \delta) \approx -\frac{qV_{\text{rf}} \cos(\phi_s)}{\omega_{\text{rf}} P_0 C_0} \left( 1 - \frac{\omega_{\text{rf}}^2 z^2}{2c^2} \right) - \frac{\eta_p \delta^2}{2}. \quad (12.196)$$

If we set:

$$H_0 = -\frac{qV_{\text{rf}} \cos(\phi_s)}{\omega_{\text{rf}} P_0 C_0}, \quad (12.197)$$

then the distribution (12.191) can be written:

$$\Psi(z, \delta) = \hat{\Psi} \exp \left( -\frac{z^2}{2\sigma_z^2} - \frac{\delta^2}{2\sigma_\delta^2} \right), \quad (12.198)$$

where:

$$\sigma_z^2 = \frac{c^2 P_0 C_0}{qV_{\text{rf}} \omega_{\text{rf}} \cos(\phi_s)} \langle H \rangle, \quad (12.199)$$

and:

$$\sigma_\delta^2 = -\frac{\langle H \rangle}{\eta_p}. \quad (12.200)$$

We see that the bi-Gaussian distribution (12.198) is an invariant distribution if the rms bunch length and energy spread are in the ratio:

$$\frac{\sigma_z}{\sigma_\delta} = \sqrt{-\frac{c^2 P_0 C_0 \eta_p}{qV_{\text{rf}} \omega_{\text{rf}} \cos(\phi_s)}} = \frac{C_0 |\eta_p|}{2\pi \nu_z}. \quad (12.201)$$

Still neglecting space-charge effects, it can be seen that the invariant distribution is generally (i.e. including all higher-order terms in the Hamiltonian) described by contours of constant value of the Hamiltonian in longitudinal phase space. A longitudinal phase space portrait is shown in Fig. 5.2.

Now let us return to the discussion of space-charge effects. We have two goals: first, to determine an expression for the space-charge potential  $V(z)$  in terms of the beam (and beam line) parameters; and second, to use the potential  $V(z)$  to determine the effect of space-charge forces on the longitudinal dynamics. We shall see, in particular, how the synchrotron tune and invariant phase space distribution are affected by space charge.

We start by finding an expression for the space-charge potential  $V(z)$ . For simplicity, we shall consider the case where the transverse distribution is a KV distribution with equal horizontal and vertical beam sizes: that is, in any plane perpendicular to the reference trajectory there is a uniform charge density up to some given radius  $a$  (beyond which the charge density

is zero). We shall assume that the beam has high energy, and that longitudinal variations in charge density and beam size are slow. Then, the radial component of the electric field at any longitudinal position can be approximated by the field around a continuous beam with charge density and beam size equal to the local density and beam size of the bunched beam. The radial component of the electric field can then be written:

$$E_r = \begin{cases} \frac{qn_q r}{2\pi\epsilon_0 a^2} & \text{if } r \leq a, \\ \frac{qn_q}{2\pi\epsilon_0 r} & \text{if } r \geq a, \end{cases} \quad (12.202)$$

where  $n_q = n_q(s, t)$  is the number of particles per unit length at some location  $s$  along the reference trajectory at time  $t$ . The magnetic field has an azimuthal component:

$$B_\theta = \frac{\beta_0}{c} E_r. \quad (12.203)$$

In addition to the radial component of the electric field, there will be a longitudinal component arising from longitudinal variations of the charge density in the beam. We want to calculate the longitudinal component of the electric field within the beam itself. To do this, we first assume that the beam is within a beam pipe of circular cross-section of radius  $b$ . We take a surface  $\mathcal{S}$  bounded by a loop  $\mathcal{C}$ , as shown in Fig. 12.4. Two sides of the loop  $\mathcal{C}$  are parallel to the axis of the beam pipe and have length  $\Delta s$ ; the other two sides are formed by radial lines from the axis of the beam pipe to the wall. Integrating Maxwell's equation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (12.204)$$

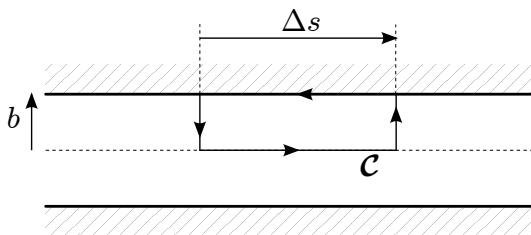


Fig. 12.4 Geometry for calculating the longitudinal electric field within a beam in an accelerator beam pipe. The beam travels along the axis of a beam pipe with circular cross-section of radius  $b$ . The integral of the electric field around the curve  $\mathcal{C}$  is equal to the rate of change of the magnetic flux through the area bounded by  $\mathcal{C}$ . For a perfectly conducting beam pipe, the longitudinal component of the electric field vanishes at the wall of the beam pipe.

over the area  $\mathcal{A}$  and applying Stokes' theorem gives:

$$\int_C \mathbf{E} \cdot d\mathbf{l} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{A}. \quad (12.205)$$

This can be written as:

$$\Delta s E_s|_{r=0} + \Delta s \frac{\partial}{\partial s} \int_0^b E_r dr - \Delta s E_s|_{r=b} = -\Delta s \frac{\partial}{\partial t} \int_0^b B_\theta dr. \quad (12.206)$$

On the left-hand side, the first and third terms represent the integral of the longitudinal component of the electric field along the axis of the beam pipe and along the wall of the beam pipe respectively; and the second term represents the sum of the integrals of the electric field along the radial lines from the axis of the beam pipe to the wall of the beam pipe.

The analysis is simplified if we consider the case of a beam pipe with perfectly conducting walls: then, the longitudinal component of the electric field vanishes at the walls of the beam pipe, i.e.  $E_s|_{r=b} = 0$ . We shall consider the case where the walls have a finite conductivity in Chapter 14; but for now (12.206) becomes:

$$E_s|_{r=0} = - \int_0^b \left( \frac{\partial B_\theta}{\partial t} + \frac{\partial E_r}{\partial s} \right) dr. \quad (12.207)$$

To proceed, it is convenient to work in frequency space, so that the number of particles per unit length of the beam line is written as:

$$n_q(s, t) = \int_{-\infty}^{\infty} \tilde{n}_q(\omega) e^{i(ks - \omega t)} d\omega. \quad (12.208)$$

If the synchrotron tune is small (i.e.  $\nu_z \ll 1$ ), then over a single turn of the ring we can assume that all particles move with the same velocity  $\beta_0 c$ . Then, the wave number  $k$  and frequency  $\omega$  of the particle density satisfy the dispersion relation:

$$\frac{\omega}{k} = \beta_0 c. \quad (12.209)$$

Using (12.208), and equations (12.202) and (12.203) for the electric and magnetic fields, we can take the derivatives and perform the integrals in (12.207) to find:

$$E_s|_{r=0} = -i \frac{qg}{4\pi\epsilon_0\beta_0\gamma_0^2 c} \int_{-\infty}^{\infty} \omega \tilde{n}_q(\omega) e^{i(ks - \omega t)} d\omega, \quad (12.210)$$

where the geometry factor  $g$  is:

$$g = 1 + 2 \ln \left( \frac{b}{a} \right). \quad (12.211)$$

As a particle moves along the beam line, the force from the electric field  $E_s$  acting on the particle will lead to a change in energy of the particle. The energy change will depend on the longitudinal co-ordinate of the particle  $z$ , which is given by (2.54):

$$z = \frac{s}{\beta_0} - ct. \quad (12.212)$$

From the dispersion relation (12.209), it then follows that:

$$ks - \omega t = \frac{\omega z}{c}. \quad (12.213)$$

Over one revolution (from  $s = s_0$  to  $s = s_0 + C_0$ ), the change in the energy deviation of a particle with longitudinal co-ordinate  $z$  will be:

$$\Delta\delta(z) = \frac{q}{cP_0} \int_{s_0}^{s_0+C_0} E_s|_{ct=s/\beta_0-z} ds. \quad (12.214)$$

If the synchrotron tune is small so that  $\nu_z \ll 1$ , then the co-ordinate  $z$  can be treated as approximately constant, and we can write:

$$\Delta\delta(z) = -i \frac{q^2 g C_0}{4\pi\varepsilon_0\beta_0\gamma_0^2 c^2 P_0} \int_{-\infty}^{\infty} \omega \tilde{n}_q(\omega) e^{i\omega z/c} d\omega. \quad (12.215)$$

It is convenient to express the change in the energy deviation in terms of the beam current  $I(s, t)$ , instead of the number of particles per unit length  $n_q(s, t)$ . These two quantities are related by:

$$I(s, t) = q n_q(s, t) \beta_0 c. \quad (12.216)$$

If we write the current in terms of a sum over Fourier modes:

$$I(s, t) = \int_{-\infty}^{\infty} \tilde{I}(\omega) e^{i(k s - \omega t)} d\omega, \quad (12.217)$$

then the mode amplitudes are given by:

$$\tilde{I}(\omega) = q \tilde{n}_q \beta_0 c. \quad (12.218)$$

The change in the energy deviation of a particle over one revolution becomes:

$$\Delta\delta(z) = -i \frac{qg}{2\varepsilon_0\beta_0\gamma_0^2 c P_0} \int_{-\infty}^{\infty} \frac{\omega}{\omega_0} \tilde{I}(\omega) e^{i\omega z/c} d\omega, \quad (12.219)$$

where we have introduced the angular revolution frequency  $\omega_0$ :

$$\omega_0 = 2\pi \frac{\beta_0 c}{C_0}. \quad (12.220)$$

We define the *longitudinal space-charge impedance*  $Z_{\parallel sc}(\omega)$ , as:

$$Z_{\parallel sc}(\omega) = i \frac{gZ_0}{2\beta_0\gamma_0^2} \frac{\omega}{\omega_0}, \quad (12.221)$$

where  $Z_0$  is the impedance of free space:

$$Z_0 = \frac{1}{\varepsilon_0 c} = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 376.73 \Omega. \quad (12.222)$$

The subscript  $\parallel$  indicates that this is a *longitudinal* impedance, i.e. a quantity characterising the effect of a longitudinal electric field. The effects of transverse fields generated by bunches in an accelerator can be characterised using transverse impedances: this will be discussed further in Chapter 14. In terms of the longitudinal space-charge impedance, the change in the energy deviation per turn of a particle in the beam with longitudinal co-ordinate  $z$  can be written:

$$\Delta\delta(z) = -\frac{q}{cP_0} \int_{-\infty}^{\infty} \tilde{I}(\omega) Z_{\parallel sc}(\omega) e^{i\omega z/c} d\omega. \quad (12.223)$$

The space-charge impedance plays the role expected of an impedance in an electric circuit, relating currents and voltages oscillating at particular frequencies. To see that this is the case, recall that the voltage  $V(z)$  is defined as the integral (over one revolution of the storage ring) of the longitudinal electric field seen by a particle in the beam with longitudinal co-ordinate  $z$ :

$$V(z) = \int_{s_0}^{s_0+C_0} E_s|_{ct=s/\beta_0-z} ds, \quad (12.224)$$

so that:

$$\Delta\delta(z) = \frac{qV(z)}{cP_0}. \quad (12.225)$$

$V(z)$  is positive if the particle gains energy from the space-charge forces. Using (12.210), (12.213), (12.218) and (12.221), the voltage  $V(z)$  can be written as:

$$V(z) = \int_{-\infty}^{\infty} \tilde{V}(\omega) e^{i\omega z/c} d\omega, \quad (12.226)$$

where:

$$\tilde{V}(\omega) = -\tilde{I}(\omega) Z_{\parallel sc}(\omega). \quad (12.227)$$

The Fourier transform of the current  $\tilde{I}(\omega)$  can be found by inverting (12.217), and making use of (12.213):

$$\tilde{I}(\omega) = \frac{1}{2\pi c} \int_{-\infty}^{\infty} I(z) e^{-i\omega z/c} dz. \quad (12.228)$$

The space-charge impedance (12.221) can be written in the form [Hofmann (1977)]:

$$\frac{Z_{\parallel sc}}{n} = i Z_0 \frac{g}{2\beta_0 \gamma_0^2}, \quad (12.229)$$

where  $n = \omega/\omega_0$ . For a beam with a current modulation of frequency  $\omega$  (when observed at a particular location in the ring),  $n$  is the number of cycles of the modulation in one revolution period. ( $n$  should not be confused with  $n_q$ , the number of particles in the beam per unit length of the beam line.) We have calculated the space-charge impedance assuming that both the beam and the beam pipe have circular cross-section, with radii  $a$  and  $b$  respectively: in this case, the geometry factor  $g$  is given by (12.211). For different geometries of the beam and the beam pipe, the form of the impedance (12.229) can still be used, but with a modified geometry factor (see, for example, [Koscieleniak (2001)]).

Equations (12.226), (12.227), (12.228) and (12.229) give the dependence of the potential  $V(z)$  on the beam current  $I(z)$ . The Haissinski equation (12.195) can now be solved (numerically) to find the effect of space-charge forces on the invariant charge distribution; with a given invariant charge distribution, it is possible to find the effect of space-charge forces on the synchrotron tune. Some assumptions are needed for the energy distribution and for the various lattice parameters. An example for the case of a Gaussian energy distribution is shown in Fig. 12.5.

In the case that the space-charge forces are not too strong (i.e. the change in the tune is small compared to the tune), it is possible to derive an approximate expression for the change in the synchrotron tune arising from space-charge forces. If the energy spread is much smaller than rf energy acceptance, then an invariant distribution can be approximated by the bi-Gaussian distribution:

$$\Psi(z, \delta) = \hat{\Psi} \exp\left(-\frac{z^2}{2\sigma_z^2}\right) \exp\left(-\frac{\delta^2}{2\sigma_\delta^2}\right), \quad (12.230)$$

where the ratio of the rms bunch length  $\sigma_z$  to the rms energy spread  $\sigma_\delta$  is given by (12.201). The beam current is:

$$I(z) = I_0 \exp\left(-\frac{z^2}{2\sigma_z^2}\right), \quad (12.231)$$

and the Fourier transform of the current is:

$$\tilde{I}(\omega) = \frac{I_0}{2\pi c} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2\sigma_z^2}\right) e^{-i\omega z/c} dz = \frac{\sigma_z I_0}{\sqrt{2\pi c}} \exp\left(-\frac{\sigma_z^2 \omega^2}{2c^2}\right). \quad (12.232)$$

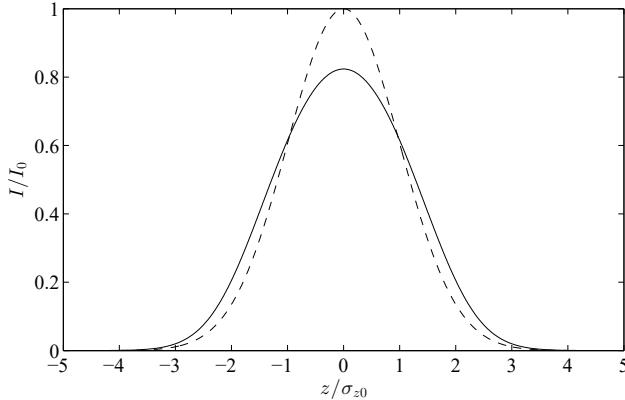


Fig. 12.5 Effect of space-charge forces on the longitudinal distribution of particles in a bunch in a synchrotron storage ring. The dashed line shows the beam current as a function of longitudinal position within a bunch in the absence of space-charge forces: the peak current and rms bunch length are  $I_0$  and  $\sigma_{z0}$  respectively. The solid line shows the beam current in the presence of space-charge forces. The storage ring is operating below transition, and has parameters such that  $qI_0gZ_0/2\eta_p\sigma_\delta^2E_0\beta_0^2\gamma_0^2 = -1$ .

Using the space-charge impedance (12.221), we find:

$$\begin{aligned} V(z) &= - \int_{-\infty}^{\infty} \tilde{I}(\omega) Z_{||\text{sc}}(\omega) e^{i\omega z/c} d\omega \\ &= - \int_{-\infty}^{\infty} \frac{\sigma_z I_0}{\sqrt{2\pi c}} \exp\left(-\frac{\sigma_z^2 \omega^2}{2c^2}\right) i \frac{gZ_0}{2\beta_0\gamma_0^2} \frac{\omega}{\omega_0} e^{i\omega z/c} d\omega \\ &= \frac{I_0 Z_0 g c}{2\beta_0\gamma_0^2\sigma_z^2\omega_0} z \exp\left(-\frac{z^2}{2\sigma_z^2}\right). \end{aligned} \quad (12.233)$$

Then, using (12.188):

$$k_z^2 = \left(\frac{2\pi\nu_z}{C_0}\right)^2 = -\frac{qV_{\text{rf}} \cos(\phi_s)}{cP_0} \frac{\omega_{\text{rf}}}{cC_0} \eta_p + \frac{\eta_p q}{cP_0 C_0} \frac{I_0 Z_0 g c}{2\beta_0\gamma_0^2\sigma_z^2\omega_0}. \quad (12.234)$$

After some algebraic manipulation, we find that the synchrotron tune is given by:

$$\nu_z^2 = \nu_{s0}^2 + \frac{\eta_p r_0 g N_b C_0^2}{(2\pi)^{\frac{5}{2}} \beta_0^2 \gamma_0^3 \sigma_z^3}, \quad (12.235)$$

where  $\nu_{s0}$  is the synchrotron tune in the absence of space-charge effects (5.49),  $N_b$  is the total number of particles in the bunch, and  $r_0$  is the classical radius of the particles in the bunch, given by:

$$r_0 = \frac{q^2}{4\pi\varepsilon_0 mc^2}, \quad (12.236)$$

where  $m$  is the mass of a single particle. It should be remembered that the expression in (12.235) for the synchrotron tune is valid only for bi-Gaussian distributions in longitudinal phase space, and where space-charge effects are sufficiently weak that there is no significant distortion of the longitudinal bunch profile. However, in cases where these conditions are satisfied, we see that below transition ( $\eta_p < 0$ ) the synchrotron tune is reduced; and above transition ( $\eta_p > 0$ ) the synchrotron tune is increased. This is easily understood using a simple physical model. A particle near the head of the bunch will see a longitudinal space-charge force acting in the direction of motion, i.e. to increase the energy of the particle. Below transition, an increase in energy results in an increase in revolution frequency; in the absence of other forces (in particular, in the absence of rf cavities) the bunch would tend to spread out longitudinally as a result of space-charge forces. This behaviour is associated with a defocusing force, which will reduce the oscillation frequency. Above transition, an increase in particle energy results in a decrease in the revolution frequency: in the absence of other forces, space-charge will tend to reduce the length of the bunch, and therefore acts like a focusing force. An increase in the focusing force will increase the oscillation frequency. If space-charge forces are strong enough, they can lead to the development of a longitudinal instability [Nielsen and Sessler (1959a)] known as the *negative-mass instability*: this is discussed in Chapter 15.

## 12.7 Beam-Beam Effects

For a beam at high energy, space-charge effects are strongly suppressed because of the cancellation that occurs between the forces from the electric and magnetic fields generated by the beam. The situation is very different, however, at the interaction point of a collider, where particles in one beam are travelling in the opposite direction to the particles from the other beam. Consider, for example, the case of an electron-positron collider. For simplicity, we shall consider only the case where both beams have the same energy, although asymmetric colliders have been built and operated. To begin with, we shall assume that the beams have a continuous longitudinal distribution, and a KV transverse distribution. In the laboratory (centre-of-mass) frame, the fields around the positron beam (moving with speed  $\beta_0 c$  in the  $+s$  direction) are given by (12.37)–(12.40), with the particle charge  $q = +e$ . Using the Lorentz force equation (12.3), the force on an electron

(charge  $q = -e$ ) moving with speed  $\beta_0 c$  in the  $-s$  direction is:

$$F_x \approx -\frac{2e^2 n_q}{\pi \varepsilon_0 a_x (a_x + a_y)} x, \quad (12.237)$$

$$F_y \approx -\frac{2e^2 n_q}{\pi \varepsilon_0 a_y (a_x + a_y)} y, \quad (12.238)$$

where  $n_q$  is the number of particles per unit length in the positron beam, and the cross-section of the positron beam has half-axes  $a_x$  and  $a_y$ . The approximations are valid for  $\beta_0 \approx 1$ , and it is assumed that the electron is within the envelope of the positron beam during the collision. Because the beams are moving in opposite directions, the forces from the electric and magnetic fields add up rather than cancel out.

For beams of particles with opposite sign of charge, the beam-beam force acts as a focusing force in both the  $x$  and the  $y$  direction: if the colliding beams consist of particles with the same sign of charge, the beam-beam force is defocusing in both the  $x$  and the  $y$  direction. The focusing (or defocusing) is strongest in the direction in which the dimension of the opposing beam is smallest. The focusing effect of the beam-beam force will lead to tune shifts and changes in the beta functions, which may be estimated (if the forces are not too strong, and if the unperturbed tunes are not too close to an integer or half-integer resonance) using the usual formulae for quadrupole perturbations, as follows. Beams in colliders are usually bunched. Suppose the bunches have a Gaussian longitudinal profile, with rms bunch length  $\sigma_z$ . Since the fields around an ultra-relativistic particle are flattened into a plane perpendicular to the direction of motion of the particle, we can estimate the force seen by an electron (for example) from the local longitudinal charge density in the opposing positron bunch. If the electron is a longitudinal distance  $z$  from the centre of the positron bunch, the local longitudinal particle density in the positron bunch will be:

$$n_q(z) = \frac{N_b}{\sqrt{2\pi}\sigma_z} e^{-z^2/2\sigma_z^2}, \quad (12.239)$$

where  $N_b$  is the number of positrons in a bunch. The transverse force on the electron can be found by substituting this expression for the local longitudinal charge density into (12.237) and (12.238). To calculate the deflection of the electron, we need to integrate the force over the length of the positron bunch. In doing this, we have to take account of the fact that the electron and the positron bunch are moving in opposite directions. Let us calculate the vertical deflection. We start with:

$$\Delta p_y = \frac{1}{P_0} \int_{-\infty}^{\infty} F_y dt \approx \frac{1}{cP_0} \int_{-\infty}^{\infty} F_y ds, \quad (12.240)$$

where  $P_0$  is the reference momentum (of both the electron and positron beams). Suppose that the interaction point is at  $s = 0$ , so that the electron and the centre of the positron bunch are at  $s = 0$  simultaneously. Then, taking account of the motion of the positron bunch, the local longitudinal positron density seen by the electron can be expressed as a function of  $s$ :

$$n_q(s) = \frac{N_b}{\sqrt{2\pi}\sigma_z} e^{-2s^2/\sigma_z^2}. \quad (12.241)$$

Substituting this expression for  $n_q$  into (12.238), and integrating over  $s$  gives:

$$\Delta p_y \approx -\frac{N_b}{cP_0} \frac{e^2}{4\pi\varepsilon_0\sigma_y(\sigma_x + \sigma_y)} y. \quad (12.242)$$

Using  $cP_0 \approx E_0 = \gamma_0 mc^2$ , where  $E_0$  is the reference energy, the deflection can be expressed as:

$$\Delta p_y \approx -\frac{N_b r_e}{\gamma_0 \sigma_y (\sigma_x + \sigma_y)} y, \quad (12.243)$$

where:

$$r_e = \frac{e^2}{4\pi\varepsilon_0 mc^2} \quad (12.244)$$

is the classical radius of the electron.

The beam-beam kick (12.243) assumes that the positron beam has a KV transverse distribution. In a more realistic case, the transverse distribution might be more closely approximated by a Gaussian. From (12.137) and (12.138), the charge density at the centre of a Gaussian beam is twice the charge density at the centre of a KV beam with the same longitudinal density and rms transverse beam size. We therefore write the effective integrated vertical focusing strength from the positron beam as:

$$\Delta K_y \approx \frac{2N_b r_e}{\gamma_0 \sigma_y (\sigma_x + \sigma_y)}. \quad (12.245)$$

This will be valid for electrons with small transverse offsets  $y \ll \sigma_y$ : for larger offsets, the nonlinearities in the force will play a significant role. Nevertheless, we can estimate from (12.245) and (6.33) the beam-beam tune shift for particles in the electron beam:

$$\Delta\nu_y \approx \frac{N_b r_e \beta_y^*}{2\pi\gamma_0 \sigma_y (\sigma_x + \sigma_y)}, \quad (12.246)$$

where  $\beta_y^*$  is the beta function for the electron beam at the interaction point (in the absence of beam-beam forces). Similarly, there is a horizontal beam-beam tune shift:

$$\Delta\nu_x \approx \frac{N_b r_e \beta_x^*}{2\pi\gamma_0 \sigma_x (\sigma_x + \sigma_y)}. \quad (12.247)$$

In (12.246) and (12.247), the beta functions and the relativistic factor  $\gamma_0$  refer to the electron beam, while the other parameters (bunch population  $N_b$ , and beam sizes  $\sigma_x$  and  $\sigma_y$ ) refer to the positron beam. For these approximations for the tune shifts to be valid, the tunes in the absence of beam-beam forces should not be close to either an integer or a half-integer resonance; however, it is common to operate colliders slightly above a half-integer resonance (for collisions of beams with opposite signs of charge) in order to enhance the luminosity by means of the ‘dynamic squeeze’ of the beta functions that results from the focusing effects of the beam-beam force.

For bunches colliding with nominally zero transverse offset, the beam-beam forces lead to a coherent tune shift: if a bunch in one of the colliding beams performs coherent betatron oscillations of some small amplitude as it moves around the ring, then the frequency of these oscillations will be affected by the focusing force from the opposing beam in the same way as the betatron frequencies of individual particles within the bunch. In principle, it should be possible to measure the beam-beam tune shift by resonant excitation of coherent betatron motion. In practice, except in special (contrived) situations, the nonlinear nature of the beam-beam force leads to a tune spread within the beams, with the betatron frequencies of different particles being shifted by different amounts according to their betatron amplitudes.

The strength of the linear focusing effects of the beam-beam forces are sometimes expressed in terms of the *beam-beam parameter*, which is defined so that, for the vertical motion:

$$\xi_y = -\frac{\beta_y^*}{4\pi} \left. \frac{d\Delta p_y}{dy} \right|_{y=0} = \frac{\beta_y^*}{4\pi} \Delta K_y. \quad (12.248)$$

The horizontal beam-beam parameter  $\xi_x$  is defined in a similar way. The vertical beam-beam tune shift is given (for tunes far from an integer or half-integer resonance) simply by:

$$\Delta\nu_y \approx \xi_y, \quad (12.249)$$

and similarly for the horizontal.

The beam-beam parameter is related to one of the principal figures of merit for a collider: the luminosity. The luminosity is a measure of the rate of collision between particles in opposing beams. For beams colliding head-on, the luminosity  $\mathcal{L}$  is given by:

$$\mathcal{L} = \frac{N_b^+ N_b^- f_0}{2\pi\Sigma_x\Sigma_y}, \quad (12.250)$$

where  $N_b^+$  and  $N_b^-$  are the bunch populations in the opposing beams,  $f_0$  is the collision frequency, and  $\Sigma_x$  and  $\Sigma_y$  are the sums in quadrature of the transverse sizes of the opposing beams:

$$\Sigma_x^2 = (\sigma_x^+)^2 + (\sigma_x^-)^2, \quad (12.251)$$

and similarly for  $\Sigma_y$ . To achieve a high luminosity, it is helpful to operate with small beam sizes, high bunch populations and high collision rates. However, small beam sizes and high bunch populations also lead to high beam-beam parameters, with the result that there can be a large tune spread resulting from the collisions. The situation is made worse by the fact that nonlinearities in the beam-beam force drive resonances in tune space: this can result in strong coupling, or reduced dynamic aperture and short beam lifetime. Generally, it is difficult to operate with beam-beam tune shifts much larger than roughly 0.01 in hadron colliders [Ohmi (2012)], or much larger than roughly 0.05 in lepton colliders (where synchrotron radiation damping helps to suppress some of the adverse effects from the beam-beam interaction) [Ohmi *et al.* (2004)]. The desire to minimise beam-beam tune shifts conflicts with the requirement to maximise the luminosity: one solution is to operate with flat beams, so that the beam size in one dimension (usually the horizontal) is kept relatively large, while the beam size in the other dimension is made as small as possible. In lepton rings, flat beams occur naturally because of synchrotron radiation effects.

The strong nonlinear forces occurring in beam-beam interactions make it very difficult to predict with any accuracy what the effects will be on the beams in a collider under given conditions. The situation is made more complex if there is a crossing-angle at the interaction point, if crab cavities are used, and if there are ‘parasitic’ interactions between bunches approaching or leaving the interaction point. If the beta function at the interaction point is of the same order as the bunch length, then there will be a significant change in beam size over the length of the bunch at the interaction point: this *hourglass effect* further complicates studies of beam-beam interactions. It is possible to develop analytical treatments for a range of conditions and including many different effects (see, for example, [Chao (2022d)]), but ultimately, numerical simulations, often requiring significant computational resources, are needed to model the processes involved in beam-beam interactions. Understanding — and controlling — beam-beam effects presents considerable challenges, and is an active area of research (see, for example, [Herr (2001); Alexahin (2005); Valishev (2008)]).

## Chapter 13

# Scattering Effects

In the discussion of space-charge effects in Chapter 12, we considered the motion of individual particles in the electromagnetic field generated by all other particles in the beam. When the density of particles is large, the field can be regarded as smooth; in particular, it is possible to neglect local variations in field strength arising from the point-like nature of particles in the beam. However, if two particles approach to within a very short distance of each other, the interaction between them can be much larger than that described by the approximation of a continuous charge distribution. Such interactions are usually referred to as ‘collisions’. Although they are ignored in the analysis of space-charge, collisions between pairs of particles do occur within a beam, and lead to observable effects. A collision between two particles can be treated as a scattering process leading to a change in the momentum of each particle. The changes in momentum resulting from collisions can be much larger than the changes that occur from the ‘background’ field generated collectively by all particles in the beam.

Collisions between particles within a beam happen randomly as a result of the betatron and synchrotron motion of the particles. The probability of a collision depends on the density of particles; the momentum change resulting from a collision depends on the positions and momenta of the interacting particles. Generally, collisions leading to large changes in momentum are less likely than collisions leading to small changes in momentum. Collisions leading to small changes in momentum happen relatively frequently: the result of such collisions is to exchange momentum between the different degrees of freedom, without loss of particles from the beam. This process is known as *intrabeam scattering*, and is discussed in Section 13.2. A collision resulting in a large momentum change can lead to particles being lost from the beam altogether: in storage rings, this occurs when the energy deviation

of a particle following the collision is outside the energy acceptance of the ring. The loss of particles from the beam as a result of collisions between particles within the beam is known as the *Touschek effect* [Bernardini *et al.* (1946)]. In Section 13.1, we describe a simple model of scattering between particles within the beam, and derive a formula for the rate of particle loss from the Touschek effect.

### 13.1 Touschek Effect

Our goal in this section is to derive a formula for the rate of loss of particles from a beam in a storage ring, where the losses result from scattering between pairs of particles within the beam. To begin, consider the particles in the rest frame of a bunch. To simplify the model, we shall ignore the effects of dispersion, and assume that, in the rest frame of a bunch, particles within the bunch are moving non-relativistically (this assumption may not be satisfied exactly in all cases of interest). We shall further suppose that the vertical and longitudinal velocities of the particles are negligible, so that we need consider only the horizontal motion. This is often a reasonable approximation for ‘flat’ beams, such as those in electron storage rings with little betatron coupling. More general formulae, including the effects of dispersion and coupling, have been derived: see, for example, [Piwinski (1998)] and references therein.

Consider a sample of  $n_1$  point-like particles incident on a ‘target’ of similar particles, where the density of the target is  $\rho_2$  (particles per unit volume). Let us work, for the moment, in the rest frame of the centre of mass of the colliding particles. The number of particles from the sample  $n_1$  scattered into a solid angle  $\Omega$  per unit distance of travel through the target will be:

$$\frac{dn_1}{dx} = \int_{\Omega} n_1 \rho_2 d\sigma, \quad (13.1)$$

where the differential cross-section for scattering between two point-like particles (electrons or positrons) is given by the Møller scattering formula [Roqué (1991)]:

$$d\sigma = \frac{4r_0^2}{(v/c)^4} \left( \frac{4}{\sin^4(\theta)} - \frac{3}{\sin^2(\theta)} \right) d\Omega. \quad (13.2)$$

Here,  $d\Omega = \sin(\theta) d\theta d\phi$  is an element of solid angle,  $r_0$  is the classical radius of the particles, and  $v$  is the velocity of the particles in the reference frame in which their centre of mass is at rest.  $\theta$  is the angle through

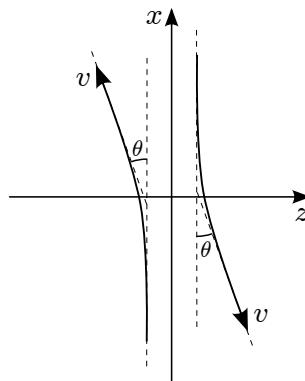


Fig. 13.1 Scattering of two particles within a bunch. In the frame in which the centre of mass of the two particles is at rest, the particles have velocity  $v$  parallel to the  $x$  axis. After scattering, the trajectory of each particle makes an angle  $\theta$  to the  $x$  axis. The differential cross-section for the scattering event is given by the Møller scattering formula (13.2).

which the trajectories of the particles are deflected: if the particles are initially travelling parallel to the  $x$  axis, then  $\theta$  is the polar angle of the final trajectory with respect to the  $x$  axis (see Fig. 13.1).

Consider a particle that (in the rest frame of the bunch) has initial momentum  $P_x$  parallel to the  $x$  axis, and is scattered through an angle  $\theta$ . After scattering, the longitudinal component of the momentum will be  $P_z = P_x \cos(\psi)$ , where  $\psi$  is the angle of the trajectory with respect to the reference trajectory. If, in the rest frame of the bunch (i.e. of the reference particle) the scattered particle is moving non-relativistically, then its energy will be close to  $mc^2$ . Applying a Lorentz boost (with relativistic factor  $\gamma_0$ ) into the laboratory frame, the energy of the scattered particle will be (in the laboratory frame):

$$E_0 = \gamma_0 mc^2 \quad (13.3)$$

before scattering, and

$$E = \gamma_0 mc^2 + \beta_0 \gamma_0 P_z c \quad (13.4)$$

after scattering. Hence, the scattering process leads to the particle acquiring an energy deviation:

$$\delta = \frac{E}{P_0 c} - \frac{1}{\beta_0} = \frac{P_z}{mc}, \quad (13.5)$$

where  $P_0$  is the reference momentum. Transverse momenta are unchanged by Lorentz boosts in the longitudinal direction. Hence, we can write:

$$P_z = P_x \cos(\psi) = P_0 p_x \cos(\psi) = \beta_0 \gamma_0 m c p_x \cos(\psi), \quad (13.6)$$

where  $p_x$  is the horizontal component of the canonical momentum (before scattering) in the laboratory frame. Therefore, the energy deviation resulting from the scattering process is:

$$\delta = \beta_0 \gamma_0 p_x \cos(\psi). \quad (13.7)$$

Now suppose that the storage ring has energy acceptance  $\delta_{\max}$ , so that any particle having energy deviation with magnitude greater than  $\delta_{\max}$  will be lost from the beam. Any scattering event for which:

$$|p_x \cos(\psi)| > \frac{\delta_{\max}}{\beta_0 \gamma_0} \quad (13.8)$$

will result in the loss of the two particles involved in the scattering. If we can calculate the rate at which such scattering events occur, then we can calculate the rate at which particles will be lost from the beam by Touschek scattering. The first step is to find the total cross-section  $\sigma_T$  for Touschek scattering. This is given by:

$$\sigma_T = \int_0^{\psi_{\max}} \sin(\psi) d\psi \int_0^{2\pi} d\chi \frac{d\sigma}{d\Omega}, \quad (13.9)$$

where  $\psi$  and  $\chi$  are polar and azimuthal angles with respect to the longitudinal axis, and  $\psi_{\max}$  is defined by:

$$\cos(\psi_{\max}) = \frac{\delta_{\max}}{\beta_0 \gamma_0 p_x}. \quad (13.10)$$

By considering the  $x$  component of any vector, we can write the relationship between angles in the different polar co-ordinate systems:

$$\cos(\theta) = \sin(\psi) \cos(\chi). \quad (13.11)$$

Then, using (13.2), the total cross-section becomes:

$$\begin{aligned} \sigma_T = & \int_0^{\psi_{\max}} \sin(\psi) d\psi \int_0^{2\pi} d\chi \\ & \frac{4r_0^2}{(v/c)^4} \left( \frac{4}{(1 - \sin^2(\psi) \cos^2(\chi))^2} - \frac{3}{1 - \sin^2(\psi) \cos^2(\chi)} \right). \end{aligned} \quad (13.12)$$

Note that we integrate over  $\psi$  only from 0 to  $\psi_{\max}$ , and not also over  $\pi - \psi_{\max}$  to  $\pi$ : although the latter range would also lead to the loss of the particles taking part in the scattering event, it is already accounted for by the scattering of the second particle (that is indistinguishable from the first).

The integral in (13.12) can be performed. The result is:

$$\sigma_T(v) = \frac{4r_0^2}{(v/c)^4} 2\pi \left( \frac{1}{\mu^2} - 1 + \ln \mu \right), \quad (13.13)$$

where  $\mu$  is defined as:

$$\mu = \cos(\psi_{\max}) = \frac{\delta_{\max}}{\beta_0 \gamma_0 p_x}. \quad (13.14)$$

Recall that the velocity  $v$  in (13.13) is the relative velocity of the colliding particles in the reference frame in which their centre of mass is at rest. We wish to calculate the scattering rate in the laboratory frame, i.e. the frame in which the centre of mass of two colliding particles is travelling at speed  $\beta_0 c$  perpendicular to their relative velocity. To do this, we need to apply a Lorentz boost to the scattering cross-section: this requires some care. Consider a sample of  $n_1$  particles incident with relative velocity  $v$  on a target with density  $\rho_2$  particles per unit volume. From (13.1), the number of particles from the sample that are scattered during a time interval  $dt$  (measured in the centre of mass of the collision) is:

$$dn_1 = v \sigma_T(v) n_1 \rho_2 dt. \quad (13.15)$$

Since the number of particles in any given sample is the same in any frame, we can say that in the laboratory frame the number of scattered particles is:

$$dn_1 = v' \sigma'_T(v') n_1 \rho'_2 dt', \quad (13.16)$$

where  $v'$  is the relative velocity of the colliding particles in the laboratory frame,  $\rho'_2$  is the density of particles in the target in the laboratory frame,  $dt'$  the time interval in the laboratory frame corresponding to the interval  $dt$  in the centre of mass frame and  $\sigma'_T(v')$  is the collision cross-section in the laboratory frame. From the relationships:

$$v' = \frac{v}{\gamma_0}, \quad (13.17)$$

$$\rho' = \gamma_0 \rho, \quad (13.18)$$

$$dt' = \gamma_0 dt, \quad (13.19)$$

and comparing (13.15) and (13.16), we see that:

$$\sigma'_T(v') = \frac{\sigma_T(v)}{\gamma_0}. \quad (13.20)$$

Finally, from the dependence of the cross-section on the relative velocity expressed in (13.13), we have:

$$\sigma'_T(v') = \frac{\sigma_T(\gamma_0 v')}{\gamma_0} = \frac{\sigma_T(v')}{\gamma_0^5}. \quad (13.21)$$

We can now write an expression for the total rate of change of the number of particles  $N_b$  in the beam in the laboratory frame as follows:

$$\begin{aligned} \frac{dN_b}{dt} = & \\ -\frac{2}{\gamma_0^5} \int (v_{x2} - v_{x1}) \sigma_T \Psi(\mathbf{x}_1, \mathbf{p}_1) \Psi(\mathbf{x}_2, \mathbf{p}_2) \delta_D(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{p}_1 d\mathbf{p}_2. \end{aligned} \quad (13.22)$$

Here, all quantities are measured in the laboratory frame (we drop the prime in order to simplify the notation):  $v_{x2} - v_{x1} = v$  is the relative velocity of two particles involved in a scattering event,  $\Psi(\mathbf{x}, \mathbf{p})$  is the phase space density of particles, and the cross-section  $\sigma_T = \sigma_T(v_{x2} - v_{x1})$  is given by (13.13). The Dirac delta function  $\delta_D(\mathbf{x}_1 - \mathbf{x}_2)$  appears because only particles at the same location in space will scatter off each other. The factor  $1/\gamma_0^5$  takes account of the Lorentz transformation of the scattering rate from the rest frame of the bunch to the laboratory frame, and follows from (13.21). Finally, the factor of 2 in front of the integral in (13.22) takes account of the fact that each scattering event leads to the loss of two particles from the beam.

Let us assume a Gaussian distribution of charge, for which the phase space density can be written:

$$\Psi(\mathbf{J}, \phi) = \frac{N_b}{(2\pi)^3 \epsilon_x \epsilon_y \epsilon_z} e^{-\frac{J_x}{\epsilon_x}} e^{-\frac{J_y}{\epsilon_y}} e^{-\frac{J_z}{\epsilon_z}}, \quad (13.23)$$

where  $\mathbf{J} = (J_x, J_y, J_z)$  and  $\phi = (\phi_x, \phi_y, \phi_z)$  are (respectively) the action and the angle variables, and  $\epsilon_x$ ,  $\epsilon_y$  and  $\epsilon_z$  are the emittances. The action variables can be expressed in terms of the Cartesian variables using the usual relationship (4.34). Then, the integrals over the co-ordinates  $\mathbf{x}_1$  and  $\mathbf{x}_2$  and over the vertical and longitudinal conjugate momenta can be performed, leaving only an integral over  $p_{x1}$  and  $p_{x2}$ :

$$\begin{aligned} \frac{dN_b}{dt} = & -\frac{2}{\gamma_0^5} \frac{N_b^2}{16\pi^3} \frac{1}{\epsilon_x^2 \sigma_y \sigma_z} \sqrt{\frac{\pi \epsilon_x}{\gamma_x}} \times \\ & \int (v_{x2} - v_{x1}) \sigma_T \exp\left(\frac{\alpha_x^2(p_{x1} + p_{x2})^2}{4\gamma_x \epsilon_x}\right) \exp\left(-\frac{\beta_x(p_{x1}^2 + p_{x2}^2)}{2\epsilon_x}\right) dp_{x1} dp_{x2}. \end{aligned} \quad (13.24)$$

Here,  $\alpha_x$ ,  $\beta_x$  and  $\gamma_x$  are the horizontal Courant–Snyder parameters (and we use the relationship  $\beta_x \gamma_x - \alpha_x^2 = 1$ ). The quantities  $\sigma_y = \sqrt{\beta_y \epsilon_y}$  and  $\sigma_z = \sqrt{\beta_z \epsilon_z}$  are the vertical beam size and the bunch length, respectively. We neglect any effects from betatron coupling and dispersion.

The integrals over  $p_{x1}$  and  $p_{x2}$  remain because the velocities  $v_{x1}$  and  $v_{x2}$  are related to these variables by:

$$v_{x1} \approx cp_{x1}, \quad (13.25)$$

$$v_{x2} \approx cp_{x2}. \quad (13.26)$$

Note also that  $v = v_{x2} - v_{x1}$ , where  $v$  is the velocity that appears in the expression for the cross-section  $\sigma_T$ . To proceed, define a new variable:

$$\zeta = p_{x2} - p_{x1}, \quad (13.27)$$

in terms of which the loss rate can be written:

$$\begin{aligned} \frac{dN_b}{dt} = & -\frac{2}{\gamma_0^5} \frac{N_b^2 c}{16\pi^3} \frac{1}{\epsilon_x^2 \sigma_y \sigma_z} \sqrt{\frac{\pi \epsilon_x}{\gamma_x}} \times \\ & \int_{-\infty}^{\infty} dp_{x1} \int_{\frac{2\delta_{\max}}{\beta_0 \gamma_0}}^{\infty} d\zeta \sigma_T(\zeta) \zeta \exp\left(-\frac{p_{x1}^2 + p_{x1}\zeta}{\gamma_x \epsilon_x}\right) \exp\left(-\frac{(1 + \beta_x \gamma_x)\zeta^2}{4\gamma_x \epsilon_x}\right). \end{aligned} \quad (13.28)$$

The cross-section is a function of  $\zeta$ , and not  $p_{x1}$ , since we can set  $v = \zeta c$  in (13.13), and  $p_x = \zeta/2$  in (13.14). Explicitly:

$$\sigma_T(\zeta) = \frac{4r_0^2}{\zeta^4} 2\pi \left( \frac{\beta_0^2 \gamma_0^2 \zeta^2}{4\delta_{\max}^2} - 1 + \ln\left(\frac{2\delta_{\max}}{\beta_0 \gamma_0 \zeta}\right) \right). \quad (13.29)$$

The lower limit on the integral over  $\zeta$  in (13.28) comes from the fact that  $\cos(\psi_{\max}) < 1$ , and hence (from (13.14)):

$$\zeta > \frac{2\delta_{\max}}{\beta_0 \gamma_0}. \quad (13.30)$$

Physically, the lower limit on  $\zeta$  means that particles need a minimum amount of transverse momentum if, after scattering, they are to have an energy deviation larger than the acceptance of the storage ring.

Since  $\sigma_T$  does not depend on  $p_{x1}$ , the integral over  $p_{x1}$  in (13.28) can be performed:

$$\frac{dN_b}{dt} = -\frac{2}{\gamma_0^5} \frac{N_b^2 c}{16\pi^2} \frac{1}{\epsilon_x \sigma_y \sigma_z} \int_{\frac{2\delta_{\max}}{\beta_0 \gamma_0}}^{\infty} \sigma_T(\zeta) \zeta \exp\left(-\frac{\beta_x \zeta^2}{4\epsilon_x}\right) d\zeta. \quad (13.31)$$

The final step is to change the variable of integration from  $\zeta$  to  $u$ , given by:

$$u = \frac{\beta_x \zeta^2}{4\epsilon_x}. \quad (13.32)$$

For convenience, we also define the parameter:

$$\xi = \frac{\delta_{\max}^2 \beta_x}{\beta_0^2 \gamma_0^2 \epsilon_x}. \quad (13.33)$$

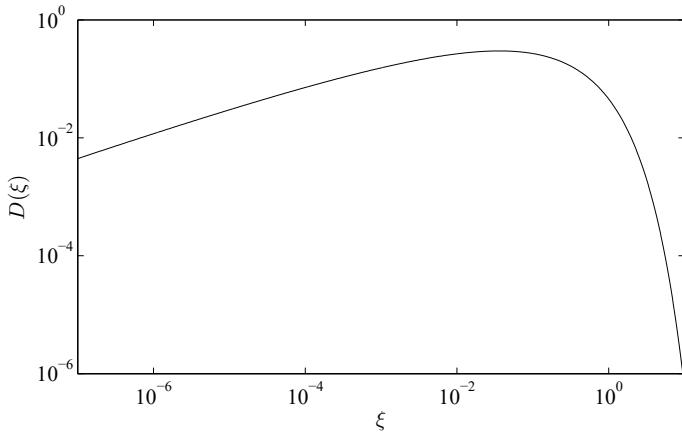


Fig. 13.2 The Touschek function  $D(\xi)$ , given by (13.35).

Then, the rate of loss of particles from the beam can be written:

$$\frac{dN_b}{dt} = -\frac{N_b^2 \beta_0^3 c r_0^2}{8\pi\gamma_0^2 \sigma_x \sigma_y \sigma_z \delta_{\max}^3} D(\xi), \quad (13.34)$$

where the function  $D(\xi)$  is given by:

$$D(\xi) = \xi^{3/2} \int_{\xi}^{\infty} \frac{e^{-u}}{u^2} \left( \frac{u}{\xi} - 1 - \frac{1}{2} \ln \left( \frac{u}{\xi} \right) \right) du. \quad (13.35)$$

The function  $D(\xi)$  is plotted in Fig. 13.2.

Equations (13.33), (13.34) and (13.35) are standard expressions for the rate of particle loss from the Touschek effect. Strictly speaking, they apply in the case of flat Gaussian beams without coupling or dispersion, when the particle motion in the rest frame of the reference particle is non-relativistic. Formulae have been derived for the loss rate in more general cases. For example, Piwinski gives expressions taking into account coupling and dispersion [Piwinski (1998)].

The Touschek lifetime  $\tau_T$  is usually defined as:

$$\frac{1}{\tau_T} = -\frac{1}{N_b} \frac{dN_b}{dt} = \frac{N_b \beta_0^3 c r_0^2}{8\pi\gamma_0^2 \sigma_x \sigma_y \sigma_z \delta_{\max}^3} D(\xi). \quad (13.36)$$

Since the loss rate  $dN_b/dt$  depends on the *square* of the number of particles, the decay of the number of particles is not exponential, although over a short time (for which the change in  $N_b$  is small)  $\tau_T$  gives the decay time of

an exponential curve fitting the decay in the bunch population. The actual decay of bunch population from the Touschek effect is given by:

$$N_b(t) = \frac{N_b(0)}{1 + N_b(0)t/\bar{\tau}_T}, \quad (13.37)$$

where:

$$\frac{1}{\bar{\tau}_T} = -\frac{1}{N_b^2} \frac{dN_b}{dt} = \frac{\beta_0^3 c r_0^2}{8\pi\gamma_0^2 \sigma_x \sigma_y \sigma_z \delta_{\max}^3} D(\xi). \quad (13.38)$$

In practice, it is more common to state the Touschek lifetime  $\tau_T$  at a particular bunch population (or beam current) than to give a value for  $\bar{\tau}_T$ .

In low-emittance storage rings, such as electron rings in third-generation synchrotron light sources, the Touschek effect is usually the dominant limitation on the beam lifetime. At higher emittances (or lower beam currents) other effects, such as scattering from residual gas molecules in the vacuum chamber, can make a significant contribution to the beam lifetime [Bocchetta (1998)]. However, at the vacuum levels typically achieved in modern accelerators, the gas scattering lifetime may be some tens of hours, while the Touschek lifetime may be only a few hours. The Touschek effect is therefore of considerable importance for operating a low-emittance storage ring, and needs to be given attention at the design stage. Users of synchrotron light often need highly stable, high brightness beams of radiation for long periods, and a Touschek lifetime that is too short will have a detrimental impact on the experiments. User requirements will generally determine the beam energy, beam current (i.e. bunch population), and horizontal and vertical beam sizes with which a machine must operate. Then, the only parameters available to improve the Touschek lifetime are the bunch length and the energy acceptance. The bunch length is determined by the momentum compaction factor of the lattice and the rf parameters (frequency and voltage). The momentum compaction factor is largely determined by the type of lattice and the size of the storage ring, but there may be some flexibility to increase the momentum compaction factor by adjusting the length and field strength of the dipoles. Reducing the rf frequency and rf voltage will increase the bunch length; however, reducing the rf voltage will also reduce the rf acceptance (which may be a limitation on  $\delta_{\max}$ ), so reducing the voltage too far will reduce the Touschek lifetime.

The most effective way to improve the Touschek lifetime in a storage ring is to increase the energy acceptance  $\delta_{\max}$ : since the function  $D(\xi)$  has a rather weak dependence on  $\xi$  in the typical operating regime of a low-emittance electron (or positron) storage ring, we see from (13.34) that the

lifetime has a very strong dependence on  $\delta_{\max}$ . There are two contributions to  $\delta_{\max}$ : the rf acceptance  $\delta_{\text{rf}}$ , and the dynamic energy acceptance  $\delta_{\text{dyn}}$ . The energy acceptance  $\delta_{\max}$  is the smaller of the rf acceptance, given by (5.59), and the dynamic energy acceptance. Typically, in an electron storage ring of the type used in third-generation synchrotron light sources, the rf acceptance is a few per cent. Usually, this is larger than the dynamic energy acceptance  $\delta_{\text{dyn}}$ , which is therefore the limiting factor for the Touschek lifetime. The dynamic energy acceptance is determined by nonlinear effects in the lattice, usually dominated by the sextupole magnets but with higher-order multipoles in the dipole and quadrupole magnets also making some contribution. Although the sextupole magnets are usually adjusted so that the first-order chromaticity is close to zero, the higher-order chromaticity can be significant. Then, the betatron tunes of particles with energy deviation of only a few per cent can reach resonances, leading to the betatron motion becoming unstable.

In an operating storage ring, a measurement of the dynamic energy acceptance can be made by recording the Touschek lifetime as a function of rf voltage (see, for example, [Steier *et al.* (2002)]). While the rf voltage is large enough that the energy acceptance  $\delta_{\max}$  is limited by the dynamic energy acceptance, the Touschek lifetime (at a given beam current) will tend to *increase* as the rf voltage is reduced: this is because reducing the rf voltage increases the bunch length, reducing the density of particles within the bunch. However, once the rf voltage becomes low enough that the rf acceptance is the limiting factor on the energy acceptance, then the Touschek lifetime will fall as the voltage is reduced. If the various parameters in the formula for  $\delta_{\text{rf}}$  (5.59) are known (as is often the case), then it is possible to calculate the rf acceptance as a function of rf voltage.

In reality, the dynamic energy acceptance is a function of position around the ring. This is because the effect of a scattering event that changes the energy deviation of a particle depends on the lattice functions (especially the dispersion) at the point that the scattering takes place. Also, because of nonlinear effects, the energy acceptance may be asymmetric; that is, whether the subsequent motion is stable depends on whether the change in energy deviation is positive or negative, as well as on the size of the change. To make an accurate analysis of the Touschek effect using a model of a machine, it is necessary to construct a detailed model of the energy acceptance of the lattice. This involves tracking particles for some (large) number of turns, starting with different energy deviations, from as many different points around the ring as possible. Ideally, an accurate

description of the physical apertures within the ring will also be included in the model. Particles that survive the tracking for more than a couple of synchrotron radiation damping times are assumed to be within the energy acceptance at the point from where the tracking started. Constructing a model of the energy acceptance in this way is computationally expensive, but often necessary if more than a very rough estimate of the Touschek lifetime is needed.

Optimisation of the energy acceptance of a lattice is a rather complex task, both at the design stage, and in operation of a storage ring. Usually, random errors in the construction of a storage ring mean that the energy acceptance achieved in practice is smaller (sometimes much smaller) than that expected from the model. If, after optimisation of the energy acceptance, the Touschek lifetime is still limiting performance, there are some measures that can be taken. For example, it is possible to install a higher-harmonic rf system that ‘flattens’ the focusing potential seen by the beam [Gram and Morton (1967); Hofmann and Myers (1980); Kohaupt (1983); Byrd *et al.* (2000)]. The goal is to increase the bunch length (or, more generally, re-shape the longitudinal charge distribution to reduce the average density) without reducing the rf acceptance. Another possibility is to use a top-up injection system, whereby small amounts of charge are injected into the ring at regular intervals, rather than waiting for the beam current to fall by a significant amount before injecting [Emery and Borland (1999)]. Installation and operation of either a higher-harmonic rf system or a top-up injection system is a non-trivial task, which may be expensive in terms of both equipment costs and staff effort. However, a number of facilities have successful experience with systems of this kind.

Finally, it is worth mentioning that the Touschek lifetime has a dependence on the beam polarisation. Although the spin dependence is not included in the formula given in (13.2), the cross-section for Møller scattering is slightly lower for particles with parallel spins than for particles with anti-parallel spins. This means that the Touschek lifetime will be longer for a polarised beam than for an unpolarised beam. Although the effect is not large (the Touschek lifetime for a fully polarised beam would be about 10% longer than for a completely depolarised beam), it is observable in electron storage rings, and has some useful application in making precise measurements of the beam energy. As discussed in Chapter 7, when a relativistic electron emits a photon in a magnetic field, there is a (small) probability that the spin of the electron will reverse direction. The probability of a spin flip has some dependence on the initial orientation of the

spin with respect to the magnetic field: over time, an electron beam in a storage ring will become polarised antiparallel to the magnetic field, while a positron beam will become polarised parallel to the magnetic field. The polarisation of a beam in this way is known as the *Sokolov–Ternov effect*. Polarisation times are typically of order some tens of minutes. In a uniform magnetic field, the polarisation of a beam of electrons will reach an equilibrium of about 92%, independent of the beam energy. As an electron moves through a magnetic field, the spin precesses about the direction of the field, with a rate given by the Thomas–BMT equation (3.169). In an electron storage ring, this leads to a spin tune, which is the total precession angle made by the spin vector in one turn of the lattice. The spin tune for an ultra-relativistic electron (i.e. with relativistic factor  $\gamma \gg 1$ ) that sees only the vertical magnetic field from the main dipole magnets in a storage ring is  $\gamma G$  (3.180), where  $G \approx 0.00116$  is the anomalous magnetic moment of the electron. If an oscillating horizontal or longitudinal magnetic field is introduced in a storage ring, then if the frequency of oscillation matches the spin tune, the result is a resonant depolarisation of the beam. If the beam initially had a high level of spin polarisation, then the depolarisation will be accompanied by a drop in the Touschek lifetime. By measuring the oscillation frequency of the field required to induce a drop in the lifetime, it is possible to determine the spin tune, and hence the beam energy, with high precision [Zholentz *et al.* (1980); Melissinos (1995); Klein *et al.* (1997); Minty and Zimmermann (2003a)].

## 13.2 Intrabeam Scattering

In Section 13.1 we discussed the Touschek effect, the loss of particles from a beam in an accelerator as a result of scattering between particles. The rate of loss of particles depends on the probability of scattering events that lead to particles having energy deviation larger than the energy acceptance of the ring. Scattering events resulting in particle losses are relatively uncommon: the Touschek lifetime in the kind of electron storage ring used in third-generation synchrotron light sources is typically several hours. However, it can be seen from the Møller scattering formula (13.2) that the cross-section increases rapidly as the deflection angle decreases. In other words, small-angle scattering is much more likely than large-angle scattering. Although small-angle scattering does not lead to particle loss, it can have an effect on the beam by transferring particle momentum between the different degrees of freedom. In practice, this leads to an increase in

the energy spread of the beam. If the scattering takes place at a location where there is horizontal (or vertical) dispersion, then there will also be an excitation of the horizontal (or vertical) emittance, in much the same way as photon emission at a location where there is dispersion leads to quantum excitation of the transverse emittance. The excitation of the longitudinal and transverse emittances by small-angle scattering is sometimes called the *multiple Touschek effect*, but is also known as *intrabeam scattering*.

The first detailed analysis of intrabeam scattering (IBS) was performed by Piwinski [Piwinski (1974)], who showed that very different behaviour resulted from scattering in a storage ring, depending on whether the ring was operated below or above transition. Below transition, small-angle scattering between particles in a beam is analogous to collisions between particles in a gas. In the case of a gas, collisions between particles lead to the energy being shared equally between the different degrees of freedom. If the energy distribution associated with a particular degree of freedom (which may be, for example, horizontal, vertical or longitudinal motion) is characterised by a temperature, then a gas that has initially a higher temperature in one degree of freedom will, through collisions between the particles, reach an equilibrium where all degrees of freedom have the same temperature. Similarly, in a storage ring below transition, IBS will lead to an equilibrium in which the distribution of particles within a bunch can be characterised by an equal ‘temperature’ in each degree of freedom. This is most evident in proton storage rings, where synchrotron radiation effects are negligible. In electron storage rings (which, however, usually operate above transition), synchrotron radiation effects dominate over IBS, which only becomes significant when the particle density within a bunch becomes very large.

In storage rings above transition, the longitudinal dynamics become very different from the dynamics below transition. In particular, an increase in energy means an effective reduction in longitudinal velocity (the revolution frequency decreases). The analogy with particles in a gas is no longer valid. It is found that, above transition, IBS does not lead to an equilibrium charge distribution; instead, the horizontal, vertical and longitudinal emittances can all grow indefinitely, with the energy for the increasing oscillation amplitudes being provided by the rf system.

The principal effect of IBS is to cause some change in the beam emittances; the main results of the theory are formulae giving the growth rates in the different degrees of freedom. The IBS growth rates are defined so

that:

$$\frac{1}{T_i} = \frac{1}{\epsilon_i^{1/2}} \frac{d\epsilon_i^{1/2}}{dt} = \frac{1}{2\epsilon_i} \frac{d\epsilon_i}{dt}, \quad (13.39)$$

where  $i = x, y$  or  $z$ ;  $T_x$ ,  $T_y$  and  $T_z$  are respectively the horizontal, vertical and longitudinal IBS growth times;  $\epsilon_x$ ,  $\epsilon_y$  and  $\epsilon_z$  are respectively the horizontal, vertical and longitudinal emittances. Neglecting coupling between the longitudinal and transverse planes, the relative energy spread  $\sigma_\delta$  is related to the longitudinal emittance  $\epsilon_z$  by:

$$\sigma_\delta^2 = \gamma_z \epsilon_z, \quad (13.40)$$

where  $\gamma_z$  is the longitudinal Courant–Snyder gamma function. Assuming that the Courant–Snyder parameters are constant in time, the longitudinal growth rate can be written (in the more conventional form):

$$\frac{1}{T_z} = \frac{1}{\sigma_\delta} \frac{d\sigma_\delta}{dt}. \quad (13.41)$$

In practice, emittance growth rates from intrabeam scattering in electron storage rings tend to be slow compared to effects from synchrotron radiation (damping and quantum excitation). This means that IBS is difficult to observe in electron storage rings; but in hadron machines, where radiation effects are generally negligible, intrabeam scattering can be a significant issue. In a hadron collider, for example, emittance growth from IBS can contribute to limitations on the luminosity. Early observations and measurements of intrabeam scattering effects were made in proton storage rings [Conte and Martini (1985); Evans and Gareyte (1985)], and experiments aimed at validating and extending the understanding and theories of IBS continued, both in proton beams [Bhat and Marriner (1991); Bhat *et al.* (1999); Rao and Hermansson (2000); Hu and Nagaitsev (2005)] and in ion beams [Fischer *et al.* (2001); Burkhardt *et al.* (2004); Artikova *et al.* (2011)]. In the years following the initial observations in hadron machines, electron storage rings reached regimes of high bunch charge and low transverse and longitudinal emittance that made observations of intrabeam scattering possible, even in the presence of synchrotron radiation effects. Measurements have been made at a number of electron storage rings [Steier *et al.* (2001); Kubo *et al.* (2002); Bane *et al.* (2002); Antoniou *et al.* (2012); Ehrlichman *et al.* (2013)], and it is possible that as electron machines continue to aim at ever more ambitious parameters, intrabeam scattering could, in some cases, be a limiting effect on performance. More recently, evidence has been found for effects of intrabeam scattering in a linac used for driving a free-electron laser [Di Mitri *et al.* (2020, 2022)].

The mathematical analysis of IBS is sufficiently complicated that it is necessary to make some approximations to obtain results that are of practical use. For the full derivations, the reader is referred to the literature; in the following, we simply present the main results, and discuss the relationships between the formulae derived by different authors. Many authors have contributed to the theory over the years, but the most notable work has been done by Piwinski [Piwinski (1974)], and by Bjorken and Mtingwa [Bjorken and Mtingwa (1983)]. Kubo and Oide [Kubo and Oide (2001)] have developed formulae based on those of Bjorken and Mtingwa, but expressed in form making a close analogy with synchrotron radiation effects as discussed in Section 7.4. Compared to synchrotron radiation, an additional complication arises in the case of IBS from the fact that the scattering rates depend on the bunch dimensions, whereas radiation effects do not. This means that some iteration is needed in electron machines to determine the equilibrium emittances in the presence of IBS (in proton machines, of course, synchrotron radiation effects can usually be neglected, so the goal is generally just to calculate the IBS growth rates for some given emittances). Nevertheless, the work of Kubo and Oide provides a convenient approach for calculating the impact of IBS in the presence of effects such as betatron coupling and vertical dispersion.

### 13.2.1 *Piwinski formulae*

Assuming a Gaussian charge distribution, the Piwinski formulae for the IBS growth rates in a synchrotron storage ring are as follows [Piwinski (1974)]:

$$\frac{1}{T_x} = A \left\langle f\left(\frac{1}{\tilde{a}}, \frac{\tilde{b}}{\tilde{a}}, \frac{\tilde{q}}{\tilde{a}}\right) + \frac{\eta_x^2 \sigma_h^2}{\beta_x \epsilon_x} f(\tilde{a}, \tilde{b}, \tilde{q}) \right\rangle, \quad (13.42)$$

$$\frac{1}{T_y} = A \left\langle f\left(\frac{1}{\tilde{b}}, \frac{\tilde{a}}{\tilde{b}}, \frac{\tilde{q}}{\tilde{b}}\right) + \frac{\eta_y^2 \sigma_h^2}{\beta_y \epsilon_y} f(\tilde{a}, \tilde{b}, \tilde{q}) \right\rangle, \quad (13.43)$$

$$\frac{1}{T_z} = A \left\langle \frac{\sigma_h^2}{\sigma_\delta^2} f(\tilde{a}, \tilde{b}, \tilde{q}) \right\rangle, \quad (13.44)$$

where:

$$A = \frac{\pi r_0^2 c N_b}{8\gamma_0 \Gamma}, \quad (13.45)$$

$$\frac{1}{\sigma_h^2} = \frac{1}{\sigma_\delta^2} + \frac{\eta_x^2}{\beta_x \epsilon_x} + \frac{\eta_y^2}{\beta_y \epsilon_y}, \quad (13.46)$$

$$\tilde{a} = \frac{\sigma_h}{\gamma_0} \sqrt{\frac{\beta_x}{\epsilon_x}}, \quad (13.47)$$

$$\tilde{b} = \frac{\sigma_h}{\gamma_0} \sqrt{\frac{\beta_y}{\epsilon_y}}, \quad (13.48)$$

$$\tilde{q} = \beta_0 \sigma_h \sqrt{\frac{2d}{r_0}}. \quad (13.49)$$

The function  $f$  is given by:

$$f(\tilde{a}, \tilde{b}, \tilde{q}) = 8\pi \int_0^1 (\log)_P \frac{1 - 3u^2}{\tilde{P}\tilde{Q}} du, \quad (13.50)$$

where:

$$(\log)_P = 2 \ln \left( \frac{\tilde{q}}{2} \left( \frac{1}{\tilde{P}} + \frac{1}{\tilde{Q}} \right) \right) - 0.577 \dots, \quad (13.51)$$

$$\tilde{P}^2 = \tilde{a}^2 + (1 - \tilde{a}^2)u^2, \quad (13.52)$$

$$\tilde{Q}^2 = \tilde{b}^2 + (1 - \tilde{b}^2)u^2. \quad (13.53)$$

The various quantities appearing in the above formulae are as follows:  $\beta_x$  ( $\beta_y$ ) is the horizontal (vertical) beta function;  $\eta_x$  ( $\eta_y$ ) is the horizontal (vertical) dispersion;  $\sigma_\delta$  is the rms energy deviation, and  $\sigma_z$  is the rms bunch length;  $N_b$  is the number of particles in the bunch;  $\gamma_0$  is the relativistic factor;  $r_0$  is the classical radius for a particle of charge  $q$  and mass  $m$ :

$$r_0 = \frac{q^2}{4\pi\varepsilon_0 mc^2}. \quad (13.54)$$

$\Gamma$  is the normalised phase space volume of the bunch, given (for a Gaussian beam) by:

$$\Gamma = \prod_{i=x,y,z} \int_0^{2\pi} d\phi_i \int_0^\infty dJ_i \beta_0 \gamma_0 e^{-J_i/\epsilon_i} = 8\pi^3 \beta_0^3 \gamma_0^3 \epsilon_x \epsilon_y \epsilon_z. \quad (13.55)$$

Note that  $\beta_0 = v/c$  (also appearing in (13.49)) where  $v$  is the velocity of the reference particle, so  $\beta_0 \gamma_0 \epsilon_i$  (for  $i = x, y$  or  $z$ ) is the normalised (horizontal, vertical or longitudinal) emittance. The parameter  $d$  is the maximum impact parameter in the collision between any two particles, and

is usually taken as the smaller of the horizontal and vertical beam sizes. The brackets  $\langle \rangle$  denote an average around the circumference of the storage ring.

Note that the longitudinal growth rate (13.44) contains a single term, whereas the horizontal and vertical growth rates (13.42) and (13.43) contain two terms, with the second term in each case proportional to the longitudinal growth rate and to the dispersion (squared). The first term in each of the expressions (13.42) and (13.43) can be interpreted as a direct excitation of (respectively) the horizontal or vertical emittance. The second term can be interpreted as a coupling of the longitudinal growth into the horizontal or vertical plane. The physics is similar to the quantum excitation of horizontal or vertical betatron motion by synchrotron radiation. In the approximation that the radiation is emitted directly along the instantaneous direction of motion of a particle, then in the absence of dispersion there is no excitation of betatron motion by the radiation. However, a particle losing energy by radiation at a location where the dispersion is non-zero experiences a change of its transverse position (and momentum) with respect to the new off-energy closed orbit, which can be described as a change in betatron amplitude. Similarly, a change in the energy of a particle by scattering off another particle at a location where there is non-zero dispersion leads to an excitation of betatron motion. There is of course also a change in betatron amplitude associated with the transfer of momentum into a different degree of freedom: this effect is independent of the dispersion, and is described by the first terms in the expressions on the right-hand sides of (13.42) and (13.43).

The similarity between quantum excitation from synchrotron radiation and emittance growth from intrabeam scattering has led Bane [Bane (2002)] to propose a modification to the Piwinski formulae, in which  $\eta_x^2/\beta_x$  is replaced by the dispersion invariant  $\mathcal{H}_x$  given by (7.74):

$$\mathcal{H}_x = \gamma_x \eta_x^2 + 2\alpha_x \eta_x \eta_{px} + \beta_x \eta_{px}^2, \quad (13.56)$$

and the quantity  $\eta_y^2/\beta_y$  is replaced by  $\mathcal{H}_y$  (defined in a similar way to  $\mathcal{H}_x$ ). The replacements are made wherever the dispersion appears, i.e. in equations (13.42), (13.43) and (13.46). The justification for this modification of the Piwinski formulae is that the change in horizontal (vertical) betatron action resulting from a change in particle energy is determined by  $\mathcal{H}_x$  ( $\mathcal{H}_y$ ). The mechanism leading to an increase in betatron action as a result of a change in energy is the same in the case of intrabeam scattering as in quantum excitation from synchrotron radiation; so the change in betatron

action should be described by the same quantity in both cases. Note that in the modified Piwinski formulae, there is a dependence of the growth rates on the gradient of the dispersion (i.e. on  $\eta_{px}$  and  $\eta_{py}$ ), which is otherwise absent from the original Piwinski formulae.

The function  $f$  satisfies:

$$f(\tilde{a}, \tilde{b}, \tilde{q}) + \frac{1}{\tilde{a}^2} f\left(\frac{1}{\tilde{a}}, \frac{\tilde{b}}{\tilde{a}}, \frac{\tilde{q}}{\tilde{a}}\right) + \frac{1}{\tilde{b}^2} f\left(\frac{1}{\tilde{b}}, \frac{\tilde{a}}{\tilde{b}}, \frac{\tilde{q}}{\tilde{b}}\right) = 0. \quad (13.57)$$

Using this property, it is possible to show from (13.42), (13.43) and (13.44) that in the presence of intrabeam scattering (and neglecting synchrotron radiation and other effects that change the emittances) the beam emittances satisfy:

$$\left( \frac{1}{\gamma_0^2} - \left\langle \frac{\eta_x^2}{\beta_x^2} \right\rangle - \left\langle \frac{\eta_y^2}{\beta_y^2} \right\rangle \right) \epsilon_z + \left\langle \frac{1}{\beta_x} \right\rangle \epsilon_x + \left\langle \frac{1}{\beta_y} \right\rangle \epsilon_y = \text{constant}. \quad (13.58)$$

If the coefficient of  $\epsilon_z$  in (13.58) is positive, then each term on the left-hand side of (13.58) is positive; in that case, there is an upper limit on the value that any one of the emittances can reach as a result of intrabeam scattering. This implies that the beam can reach an equilibrium. However, if the coefficient of  $\epsilon_z$  in (13.58) is negative, then the longitudinal emittance and one or both of the transverse emittances can grow indefinitely.

Whether the IBS emittance growth has an upper limit or not depends on whether the storage ring is operated below or above transition. We can show that this is so in the case of a FODO storage ring, as follows. First, using (7.108) and (7.109), we find that the peak beta function and dispersion in a FODO storage ring satisfy:

$$\frac{\beta_x^2}{\eta_x} = \frac{2f\rho(2f + \rho\tan(\theta))}{(2f + \rho\tan(\theta/2))(2f - \rho\tan(\theta))}, \quad (13.59)$$

where  $f$  is the focal length of the quadrupoles,  $\rho$  is the bending radius of the dipoles, and  $\theta$  is the bending angle of a single dipole. In a ring consisting of a large number of FODO cells, we can expand the right-hand side of (13.59) as a series in  $\theta$ :

$$\frac{\beta_x^2}{\rho\eta_x} = 1 + \frac{3\rho}{4f}\theta + O(\theta^2). \quad (13.60)$$

Neglecting terms of first order (and higher) in  $\theta$ , this gives:

$$\frac{\eta_x^2}{\beta_x^2} \approx \frac{\eta_x}{\rho}. \quad (13.61)$$

Hence, the momentum compaction factor (5.20) can be written:

$$\alpha_p = \left\langle \frac{\eta_x}{\rho} \right\rangle \approx \left\langle \frac{\eta_x^2}{\beta_x^2} \right\rangle. \quad (13.62)$$

Finally, neglecting vertical dispersion (which in a planar storage ring has the design value zero), the coefficient of  $\epsilon_z$  in (13.58) can be written:

$$\frac{1}{\gamma_0^2} - \left\langle \frac{\eta_x^2}{\beta_x^2} \right\rangle \approx \frac{1}{\gamma_0^2} - \alpha_p = -\eta_p, \quad (13.63)$$

where  $\eta_p$  is the phase slip factor, and we have used (5.30). Below transition, the phase slip is negative (by definition); hence, the coefficient of  $\epsilon_z$  in (13.58) will be positive, and there will be an upper limit on the IBS emittance growth. Above transition (i.e. for sufficiently high beam energy), the phase slip will be positive: the coefficient of  $\epsilon_z$  in (13.58) will be negative, and the emittances can grow (in principle) indefinitely as a result of intrabeam scattering.

High energy electron storage rings usually operate above transition, since the relativistic factor  $\gamma_0$  is large. Intrabeam scattering will cause some emittance growth, but the effects are mitigated by synchrotron radiation damping. Taking into account quantum excitation, radiation damping and IBS, the emittance in an electron storage ring will evolve as:

$$\frac{d\epsilon_i}{dt} = \frac{2}{\tau_i}(\epsilon_{i0} - \epsilon_i) + \frac{2}{T_i}\epsilon_i, \quad (13.64)$$

where  $\tau_i$  is the radiation damping time,  $T_i$  is the IBS growth time, and  $\epsilon_{i0}$  is the equilibrium emittance in the absence of IBS (i.e. the equilibrium emittance determined just by quantum excitation and radiation damping). If  $\epsilon_{i1}$  is the equilibrium emittance when IBS is taken into account, then:

$$\frac{2}{\tau_i}(\epsilon_{i0} - \epsilon_{i1}) + \frac{2}{T_i}\epsilon_{i1} = 0, \quad (13.65)$$

and hence:

$$\epsilon_{i1} = \frac{\epsilon_{i0}}{1 - \tau_i/T_i}. \quad (13.66)$$

Since the IBS growth time  $T_i$  is a function of the beam emittances, to find the equilibrium emittance in the presence of IBS it is necessary to use an iterative procedure to solve (13.66). Intrabeam scattering is normally not significant in electron storage rings: at the emittances and bunch charges that are usually achieved, the IBS growth times are long compared to the radiation damping times. In rings operating with ultra-low vertical emittance (perhaps of a few picometres), the IBS growth times can be short enough to

have an observable effect on the equilibrium emittance. In extreme cases, it is theoretically possible for the IBS growth time to be shorter than the radiation damping time: equation (13.66) then suggests a *negative* value for the equilibrium emittance. However, since the IBS growth time depends on the emittance, the emittance will simply increase until the IBS growth time is less than the radiation damping time. Equation (13.66) applies to the equilibrium situation, and tells us that no equilibrium is possible where the IBS growth time is shorter than the radiation damping time.

### 13.2.2 Bjorken–Mtingwa formulae

Bjorken and Mtingwa used a different approach from Piwinski to obtain the following formula for the IBS growth rates [Bjorken and Mtingwa (1983)]:

$$\frac{1}{T_i} = 4\pi A(\log)_{\text{BM}} \times \left\langle \int_0^\infty d\lambda \sqrt{\frac{\lambda}{\det(L + \lambda I)}} \left[ \text{Tr}(L_i) \text{Tr}\left(\frac{1}{L + \lambda I}\right) - 3 \text{Tr}\left(\frac{L_i}{L + \lambda I}\right) \right] \right\rangle, \quad (13.67)$$

where  $A$  is given by (13.45),  $I$  is the identity matrix, the matrices  $L_i$  are given by:

$$L_x = \frac{\gamma_0 \sqrt{\beta_x \mathcal{H}_x}}{\epsilon_x} \begin{pmatrix} \frac{1}{\gamma_0} \sqrt{\frac{\beta_x}{\mathcal{H}_x}} \sin(\varphi_x) & 0 \\ \sin(\varphi_x) & \gamma_0 \sqrt{\frac{\mathcal{H}_x}{\beta_x}} 0 \\ 0 & 0 \end{pmatrix}, \quad (13.68)$$

$$L_y = \frac{\gamma_0 \sqrt{\beta_y \mathcal{H}_y}}{\epsilon_y} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \gamma_0 \sqrt{\frac{\mathcal{H}_y}{\beta_y}} \sin(\varphi_y) & \\ 0 & \sin(\varphi_y) & \frac{1}{\gamma_0} \sqrt{\frac{\beta_y}{\mathcal{H}_y}} \end{pmatrix}, \quad (13.69)$$

$$L_z = \frac{\gamma_0^2}{\sigma_\delta^2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (13.70)$$

and  $L$  is defined as:

$$L = L_x + L_y + L_z. \quad (13.71)$$

$\mathcal{H}_x$  is the dispersion invariant (7.74):

$$\mathcal{H}_x = \gamma_x \eta_x^2 + 2\alpha_x \eta_x \eta_{px} + \beta_x \eta_{px}^2, \quad (13.72)$$

with a similar expression for  $\mathcal{H}_y$ . The function  $\varphi_x$  is defined by:

$$\tan(\varphi_x) = -\beta_x \frac{\eta_{px}}{\eta_x} - \alpha_x, \quad (13.73)$$

with a similar definition for  $\varphi_y$ .

The meaning of the quantities  $\mathcal{H}_x$  and  $\varphi_x$  can be understood as follows. If we compare (13.72) and (13.73) with the action  $J_x$  (4.34) and angle  $\phi_x$  (4.35), and note that the dispersion  $\eta_x \delta$  gives the trajectory of a particle with energy deviation  $\delta$ , then it follows from (4.36) and (4.37) that:

$$\eta_x = \sqrt{\beta_x \mathcal{H}_x} \cos(\varphi_x), \quad (13.74)$$

$$\eta_{px} = -\sqrt{\frac{\mathcal{H}_x}{\beta_x}} (\sin(\varphi_x) + \alpha_x \cos(\varphi_x)). \quad (13.75)$$

That is,  $\mathcal{H}_x$  describes the ‘amplitude’ of the horizontal dispersion, and  $\varphi_x$  the ‘phase’ of the horizontal dispersion.

The quantity  $(\log)_{\text{BM}}$ , sometimes known as the *Coulomb log* is defined as:

$$(\log)_{\text{BM}} = \ln\left(\frac{b_{\max}}{b_{\min}}\right), \quad (13.76)$$

where  $b_{\max}$  is the maximum impact parameter that can occur during the collision of two particles in the beam, and  $b_{\min}$  is the minimum impact parameter that can occur during collision. The values for these parameters are sometimes chosen as follows:

$$b_{\max} = d, \quad (13.77)$$

$$b_{\min} = r_0, \quad (13.78)$$

where  $d$  is the smaller of the horizontal and vertical beam size, and  $r_0$  is the classical radius of the particles.

The expressions for the IBS growth rates in the Bjorken–Mttingwa formula (13.67) take a form that looks very different from the expressions for the growth rates in the Piwinski formulae (13.42), (13.43) and (13.44). However, Bane [Bane (2002)] has shown that, with certain assumptions, the Piwinski formulae and the Bjorken–Mttingwa formulae are in good agreement with each other. This is discussed further in the following section, where we consider approximations to the formulae for the IBS growth rates that simplify the computations of the growth rates for high energy beams.

### 13.2.3 High energy approximation

The formulae (13.42)–(13.44) and (13.67) give expressions for the growth rates from intrabeam scattering that can be evaluated if the various lattice and beam parameters are known. However, evaluation involves performing the integrals appearing in these formulae, and averaging the results around the entire storage ring. This can be computationally expensive, and it is often useful to have simplified, though approximate, formulae that can be more easily and quickly applied. Various approximations have been presented for the high energy case, notably by Bane [Bane (2002)], and by Kubo *et al.* [Kubo *et al.* (2005)]. One possible approach is as follows.

Let us start with the longitudinal growth rate given by the Bjorken–Mttingwa formula (13.67) with the index  $i = z$ . Note that the off-diagonal elements of  $L$  are in the range  $-1$  to  $1$ ; the elements on the diagonal, however, can become much larger than  $1$ . This can be the case, in particular, at high energy. We shall assume that the off-diagonal elements of  $L$  are negligible compared to the diagonal elements; then,  $L$  becomes:

$$L = \frac{\gamma_0^2}{\sigma_\delta^2} \begin{pmatrix} a^2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & b^2 \end{pmatrix}, \quad (13.79)$$

where:

$$a = \frac{\sigma_H}{\gamma_0} \sqrt{\frac{\beta_x}{\epsilon_x}}, \quad (13.80)$$

$$b = \frac{\sigma_H}{\gamma_0} \sqrt{\frac{\beta_y}{\epsilon_y}}, \quad (13.81)$$

and:

$$\frac{1}{\sigma_H^2} = \frac{1}{\sigma_\delta^2} + \frac{\mathcal{H}_x}{\epsilon_x} + \frac{\mathcal{H}_y}{\epsilon_y}. \quad (13.82)$$

Note that  $\sigma_H$ ,  $a$  and  $b$  correspond to the quantities  $\sigma_h$ ,  $\tilde{a}$  and  $\tilde{b}$  defined in the Piwinski formulation (13.46), (13.47) and (13.48), but with Bane's modification in which  $\eta_x^2/\beta_x$  is replaced by  $\mathcal{H}_x$ , and similarly for  $y$ . Then, after some algebraic manipulation, we find that the longitudinal growth rate given by (13.67) can be written:

$$\frac{1}{T_z} = 4\pi A (\log)_{\text{BM}} \left\langle \frac{\sigma_H^2}{\sigma_\delta^2} I_{\text{BM}}(a, b) \right\rangle, \quad (13.83)$$

where the function  $I_{\text{BM}}(a, b)$  is the integral:

$$I_{\text{BM}}(a, b) = \int_0^\infty d\lambda \sqrt{\frac{\lambda}{(a^2 + \lambda)(1 + \lambda)(b^2 + \lambda)}} \left( \frac{1}{a^2 + \lambda} - \frac{2}{1 + \lambda} + \frac{1}{b^2 + \lambda} \right). \quad (13.84)$$

Before proceeding to write an explicit expression for the integral (in terms of special functions), consider the longitudinal IBS growth rate given by the modified Piwinski formula:

$$\frac{1}{T_z} = A \left\langle \frac{\sigma_H^2}{\sigma_\delta^2} f(a, b, q) \right\rangle, \quad (13.85)$$

where the function  $f$  is given by (13.50). Let us assume that we can take the factor  $(\log)_P$  outside the integral in (13.50), by re-defining the Coulomb log as:

$$(\log)_P = 2 \ln \left( \frac{q}{2} \left( \frac{1}{a} + \frac{1}{b} \right) \right) - 0.577. \quad (13.86)$$

Then the function  $f$  becomes:

$$f(a, b, q) = 4\pi (\log)_P I_P(a, b), \quad (13.87)$$

and the longitudinal IBS growth rate is:

$$\frac{1}{T_z} = 4\pi A (\log)_P \left\langle \frac{\sigma_H^2}{\sigma_\delta^2} I_P(a, b) \right\rangle, \quad (13.88)$$

where the function  $I_P(a, b)$  is the integral:

$$I_P(a, b) = 2 \int_0^1 du \frac{1 - 3u^2}{\sqrt{(a^2 + (1 - a^2)u^2)(b^2 + (1 - b^2)u^2)}}. \quad (13.89)$$

If we equate the Coulomb log in the Piwinski formulae  $(\log)_P$  with the Coulomb log in the Bjorken–Mttingwa formula  $(\log)_{\text{BM}}$ , we see that the longitudinal growth rates (13.83) and (13.89) will agree if  $I_{\text{BM}}(a, b) = I_P(a, b)$ . This seems to be the case, as can be verified by numerical integration for different values of  $a$  and  $b$ . In the form (13.89), the integral may be written in terms of elliptic integrals:

$$I_P(a, b) = \frac{2}{ab \sin(\psi)} \left( F(\psi|m) - \frac{3}{\sin^2(\bar{\psi})} (F(\psi|m) - E(\psi|m)) \right), \quad (13.90)$$

where  $E(\psi|m)$  is an elliptic integral of the first kind,  $F(\psi|m)$  is an elliptic integral of the second kind, and:

$$\sin(\psi) = \frac{\sqrt{a^2 - 1}}{a}, \quad (13.91)$$

$$\sin(\bar{\psi}) = \frac{\sqrt{b^2 - 1}}{b}, \quad (13.92)$$

$$m = \frac{\sin^2(\bar{\psi})}{\sin^2(\psi)}. \quad (13.93)$$

The elliptic integrals are defined by [Callaghan and Maslen (1960); Abramowitz and Stegun (1964); Gradshteyn and Ryzhik (2014a)]:

$$E(\psi|m) = \int_0^\psi \sqrt{1 - m \sin^2(\theta)} d\theta, \quad (13.94)$$

$$F(\psi|m) = \int_0^\psi \frac{d\theta}{\sqrt{1 - m \sin^2(\theta)}}. \quad (13.95)$$

There are function libraries available in different computer codes and programming languages for efficient computation of the elliptic integrals, i.e. without the need for performing numerical integration. Then, equation (13.88), with  $(\log)_P$  given by (13.86) and  $I_P(a, b)$  given by (13.90), provides an expression for the longitudinal IBS growth rate where the only integration required is in averaging over the circumference of the storage ring.

Using Piwinski's formulae, the horizontal and vertical IBS growth rates are again expressed in terms of the function  $f$ , given by (13.87), (13.86) and (13.90). With Bane's modification, the expressions are, from (13.42) and (13.43):

$$\frac{1}{T_x} = A \left\langle f\left(\frac{1}{a}, \frac{b}{a}, \frac{q}{a}\right) + \frac{\eta_x^2 \sigma_H^2}{\beta_x \epsilon_x} f(a, b, q) \right\rangle, \quad (13.96)$$

$$\frac{1}{T_y} = A \left\langle f\left(\frac{1}{b}, \frac{a}{b}, \frac{q}{b}\right) + \frac{\eta_y^2 \sigma_H^2}{\beta_y \epsilon_y} f(a, b, q) \right\rangle. \quad (13.97)$$

## Chapter 14

# Wake Fields, Wake Functions and Impedance

As a bunch of particles moves along an accelerator beam line, each particle within the bunch will act as a source of electric and magnetic fields that will affect the motion of other particles. In Chapter 12 we discussed space-charge effects, treating the bunch (as a source of electromagnetic fields) as a continuous charge distribution. This approximation neglects the possibility of collisions between pairs of particles within the bunch; such collisions do have observable effects, including the loss of particles from the beam (the Touschek effect) and the increase in beam emittances by transfer of momentum between different degrees of freedom (intrabeam scattering). Scattering effects were discussed in Chapter 13.

Space-charge and scattering effects occur even in the absence of a beam pipe, i.e. assuming that the bunch is travelling in an unbounded region of space. When a beam pipe is present, there are image charges induced in the walls of the pipe that can affect the dynamics of the beam: some of these effects are described by the Laslett tune shifts. However, there are also more complicated effects resulting from the presence of a vacuum chamber. If the particles in an accelerator are ultra-relativistic, then fields interacting with the vacuum chamber can only affect particles behind those generating the fields. These fields are therefore known as *wake fields*. At high energy, space-charge and scattering effects are often relatively weak compared to the effects of wake fields.

Wake fields tend to depend on position and time in complicated ways. To understand the impact of wake fields on the beam dynamics in an accelerator, it is necessary to work with simplified descriptions of the fields. An appropriate simplification is provided by *wake functions*, which can be used to describe the force on a particle following a localised charge distribution (such as, for example, a single point charge) through a given section of the

accelerator. In general, a wake function expresses the wake fields as a function of the distance between a localised source and the charge experiencing the wake field, with the assumption that all particles can be treated in the ultra-relativistic limit.

The effects of wake fields are quite diverse, and include longitudinal and transverse emittance growth, distortion of the charge distribution, and beam instability. The goal of an analysis of wake fields in an accelerator is generally to characterise the various effects under different operational conditions. Beam instabilities in particular can impose strong limits on parameters such as the beam current. In some situations, above a certain threshold value of the beam current, there no longer exists a steady equilibrium distribution of charge within the beam: individual particles (or entire bunches) start to perform oscillations of increasing amplitude, leading eventually to loss of some fraction of the beam current. Understanding the limits on beam stability, and designing the accelerator so that the specified performance can be achieved, is an important aspect of machine design, commissioning and operation. Even if the beam does not become unstable, wake field effects can make it difficult to achieve the desired beam quality (in terms, for example, of the emittance).

For analysis of the beam dynamics in the presence of wake fields, it is often convenient to work in the frequency domain, where one considers the behaviour of a beam with a density perturbation that oscillates at a particular frequency. A beam instability is indicated by the existence of perturbations with amplitudes that grow over time. Wake functions provide a description of the wake fields in the time domain; the Fourier transform of a wake function is an *impedance*, which describes the wake fields in the frequency domain. By writing the equation of motion for the charge distribution in terms of the impedance, it is possible to derive an eigenfunction equation for density perturbations oscillating at well-defined frequencies. The imaginary part of the oscillation frequency gives the growth rate of the amplitude of the corresponding perturbation. In this chapter, we shall discuss wake fields, wake functions and impedances; the effects of wake fields on the beam dynamics will be considered in Chapter 15.

Computation of wake fields, wake functions and impedances for a given accelerator component is usually a complex task requiring the numerical solution of Maxwell's equations for a given source (charge distribution). However, there are simple cases where the calculation of the wake function can be performed analytically. A full discussion of the various situations and techniques that can be used is beyond the scope of this book, but we

shall give examples of the calculation of wake fields (and the corresponding wake functions and impedances) in two cases: first, a resonant cavity, and second, a long, straight vacuum chamber with circular cross-section and walls with specified conductivity. For further information on the calculation and properties of wake fields, the reader is referred to more specialised texts and articles, for example [Chao (1993a); Zotter and Kheifets (1998); Gluckstern and Kurennoy (1999); Weiland (1999); Bane *et al.* (2004)].

### 14.1 Wake Fields in a Resonant Cavity

As an illustration of the generation of wake fields in an accelerator, we shall discuss a simplified model for a point charge moving through a resonant cavity. Even if the cavity has a simple geometry, such as a cylinder with axis lying along the reference trajectory, a careful treatment of the full physical system becomes fairly complicated. We will not attempt a rigorous analysis here: however, it will be convenient later on to use the wake field for a resonant cavity as a standard example when discussing wake functions, impedance and instabilities. We therefore present a simplified model by way of justification for the formulae used in later discussion; for a more rigorous treatment of the wake field in a resonant cavity, the reader is referred to the original papers by Lawson [Lawson (1968, 1990)], or specialised texts, for example [Zotter and Kheifets (1998)].

A resonant cavity provides an example of a *geometric wake field*, i.e. a wake field arising from variations in the size or shape of the cross-section of the chamber through which a beam is moving. Wake fields also occur in long, straight beam pipes with uniform cross-section, as a result of the finite conductivity of the material from which the beam pipe is made. Resistive-wall wake fields are discussed in Section 14.2.

When a particle moving along an accelerator beam pipe enters a cavity (which could be an accelerating cavity, or simply a section of the beam pipe with larger radius than the sections on either side), the fields induced in the cavity can be calculated by considering the diffraction of the fields around the particle at the aperture through which the particle enters the cavity. As a result of the diffraction, the fields spread out to fill the volume of the cavity, and form the wake fields observed by particles entering the cavity some time after the initial particle has passed through. The evolution of the fields is determined by Maxwell's equations, with initial conditions given by the fields around the charged particle as it passes through the cavity, and subject to the boundary conditions imposed by the walls of the cavity. The

fields in a cavity may generally be expressed as resonant modes with particular frequencies; for simple geometries, there are analytical expressions for the resonant modes (examples for rectangular and cylindrical cavities were discussed in Chapter 1). By writing the initial fields as a sum over the resonant modes with appropriate amplitudes and phases, explicit expressions are obtained for the evolution of the fields within the cavity after the particle has passed through. The practical value of the results found by this method is limited by the rather simple model we use for the initial conditions; nevertheless, it provides some basis for the formulae used later for the wake fields in a resonant cavity. More rigorous treatments (which can be carried out for simple geometries) lead to formulae that differ in some respects, in particular for the amplitudes of the induced wake fields, though the essential features are the same. For almost any real case, the wake fields must be found from computer modelling of the fields around a charge distribution moving through a system with a given geometry.

Consider a particle with charge  $q$  moving at constant velocity  $v$  in a straight line in free space. In the limit  $v \rightarrow c$ , relativistic effects flatten the fields around the particle so that they lie entirely within a plane perpendicular to the trajectory of the particle. Applying Gauss' theorem:

$$\int_{\mathcal{V}} \nabla \cdot \mathbf{E} dV = \int_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{A}, \quad (14.1)$$

to Maxwell's equation:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, \quad (14.2)$$

gives:

$$2\pi r \int_{-\infty}^{\infty} E_r ds = \frac{q}{\epsilon_0}. \quad (14.3)$$

Here, the integral on the left-hand side of (14.1) is over the volume  $\mathcal{V}$  of a cylinder of radius  $r$  and infinite length, coaxial with the trajectory of the particle; the integral on the right-hand side of (14.1) is over the surface  $\mathcal{S}$  of this cylinder;  $E_r$  is the radial component of the electric field on the surface of the cylinder; and we have used the rotational symmetry of the system about the axis of the cylinder to perform the integral over the azimuthal angle  $\theta$  (where  $dA = r d\theta ds$ ).  $s$  is the distance along the trajectory of the particle. Since the fields lie in a plane perpendicular to the trajectory of the particle, we can write, in terms of the Dirac delta function  $\delta_D(z)$ :

$$E_r = \frac{q}{2\pi\epsilon_0 r} \delta_D(s - ct). \quad (14.4)$$

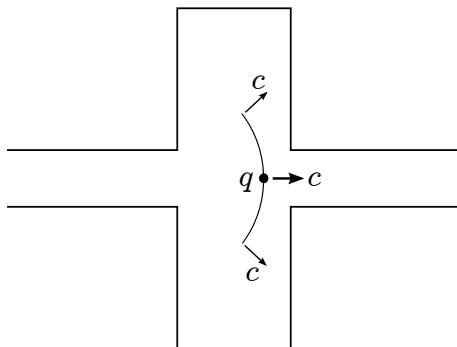


Fig. 14.1 Generation of wake fields by a particle moving at (approximately) the speed of light through a cylindrical cavity. The field is represented by a curved wave front that extends an increasing distance from the axis of the cavity as the particle moves from the entrance to the exit of the cavity. The field is reflected from the end wall of the cavity and forms a wake field seen by a later particle.

Now suppose that the particle enters a cavity with conducting walls. The shape of the cavity is a cylinder of length  $L$  and radius  $R$ , and the cavity is oriented so that its axis lies along the (straight line) trajectory of the charged particle. The cavity enters one end of the cavity through a small aperture. At the aperture, the field around the particle will be diffracted, so that the field within the cavity will no longer be described by (14.4). If the size of the aperture is small, then the field may be represented as a curved wavefront extending over some small area of the surface of a sphere centred on the entrance aperture. At the far end of the cavity, the particle exits through another small aperture. However, the field around the particle at that point extends beyond the exit aperture. The field will then be reflected from the end wall of the cavity, and forms the wake fields observed by later particles: see Fig. 14.1.

For a particle travelling close to the speed of light, at the exit of a cavity of length  $L$  the field around the particle will extend roughly a distance  $L$  from the axis. The longitudinal separation between the ends of the wavefront and the centre is therefore:

$$\Delta s = L(1 - \cos(\psi)), \quad (14.5)$$

with  $\psi \approx 1$ ; hence,  $\Delta s \approx L/2$ . Using a simplified model, we represent the curved wavefront generated by the particle as a plane wavefront travelling a distance  $L/2$  behind the particle, but with amplitude equal to the field around the particle in free space. The initial condition on the field in the

cavity is then:

$$E_r = \frac{q}{2\pi\epsilon_0 r} \delta_D(s - ct + L/2). \quad (14.6)$$

To determine how this field evolves, we assume that it can be expressed as a sum over TM modes in the cavity. The electric field in a TM mode in a cylindrical cavity can be written (1.319)–(1.321):

$$E_r = E_{\ell mn} \frac{k_\ell}{k_{mn}} J'_n(k_{mn}r) \cos(n\theta) \cos(k_\ell s) \cos(\omega_{\ell mn} t - \phi_{\ell mn}), \quad (14.7)$$

$$E_\theta = -E_{\ell mn} \frac{n k_\ell}{k_{mn}^2 r} J_n(k_{mn}r) \sin(n\theta) \cos(k_\ell s) \cos(\omega_{\ell mn} t - \phi_{\ell mn}), \quad (14.8)$$

$$E_s = E_{\ell mn} J_n(k_{mn}r) \cos(n\theta) \sin(k_\ell s) \cos(\omega_{\ell mn} t - \phi_{\ell mn}). \quad (14.9)$$

The corresponding magnetic field can be written (1.322)–(1.324):

$$B_r = E_{\ell mn} \frac{n \omega_{\ell mn}}{c^2 k_{mn}^2 r} J_n(k_{mn}r) \sin(n\theta) \sin(k_\ell s) \sin(\omega_{\ell mn} t - \phi_{\ell mn}), \quad (14.10)$$

$$B_\theta = E_{\ell mn} \frac{\omega_{\ell mn}}{c^2 k_{mn}^2} J'_n(k_{mn}r) \cos(n\theta) \sin(k_\ell s) \sin(\omega_{\ell mn} t - \phi_{\ell mn}), \quad (14.11)$$

$$B_s = 0. \quad (14.12)$$

The quantities  $\phi_{\ell mn}$  are fixed phases. Maxwell's equations are satisfied if:

$$\omega_{\ell mn}^2 = c^2 (k_\ell^2 + k_{mn}^2). \quad (14.13)$$

Assuming perfectly conducting walls, the tangential component of the electric field and the normal component of the magnetic field must vanish at the walls of the cavity. The boundary conditions on the curved wall are satisfied if:

$$k_{mn} = \frac{p_{mn}}{R}, \quad (14.14)$$

where  $m$  and  $n$  are positive integers (with  $m \geq 1$  and  $n \geq 0$ ), and  $p_{mn}$  is the  $m$ th zero of the Bessel function  $J_n(x)$ . The boundary conditions on the ends of the cavity (at  $s = \pm L/2$ ) are satisfied if:

$$k_\ell = (2\ell - 1) \frac{\pi}{L}, \quad (14.15)$$

for  $\ell$  any positive integer ( $\ell \geq 1$ ).

For fields excited by a charge moving along the axis of the cavity, the symmetry of the system determines that the fields have no dependence on the azimuthal angle  $\theta$ . Therefore, we must have:

$$E_{\ell mn} = 0 \quad \text{if} \quad n \neq 0. \quad (14.16)$$

The radial field in the cavity can then be expressed as a series:

$$E_r = \sum_{\ell,m=1}^{\infty} E_{\ell m} \frac{k_l}{k_m} J'_0(k_m r) \cos(k_\ell s) \cos(\omega_{\ell m} t - \phi_{\ell m}), \quad (14.17)$$

where we drop the subscript  $n$  on the various quantities.

We now look for field amplitudes  $E_{\ell m}$  that make the radial field component in the TM mode expansion (14.17) match the initial condition (14.6). The Dirac delta function in (14.6) can be expressed as a series:

$$\delta_D(s - ct + L/2) = \frac{4}{L} \sum_{\ell=1}^{\infty} \cos(k_\ell s) \cos(k_\ell(ct - L/2)). \quad (14.18)$$

A comparison of (14.18) with (14.17) suggests that the phases  $\phi_{\ell m}$  are given by:

$$\phi_{\ell m} = (2\ell - 1) \frac{\pi}{2}. \quad (14.19)$$

We shall see shortly that, based on physical considerations, (14.19) is an appropriate expression for the phase values, and this provides some justification for our model (14.6) for the initial condition for the field in the cavity. To continue with our search for the field amplitudes  $E_{\ell m}$ , we write the dependence of the radial field on the co-ordinate  $r$  in (14.6) as a summation over Bessel functions. This can be done using the properties of Fourier–Bessel series. If a function  $f(\bar{r})$  is expressed as a series:

$$f(\bar{r}) = \sum_{m=1}^{\infty} C_{mn} J_n(p_{mn} \bar{r}), \quad (14.20)$$

for  $0 < \bar{r} < 1$ , then it follows from the orthogonality of Bessel functions that the coefficients  $C_{mn}$  are given by:

$$C_{mn} = \frac{2}{J_{n\pm 1}(p_{mn})^2} \int_0^1 \bar{r} f(\bar{r}) J_n(p_{mn} \bar{r}) d\bar{r}, \quad (14.21)$$

where either sign (in  $J_{n\pm 1}(p_{mn})$ ) is equally valid. If we choose:

$$f(\bar{r}) = \ln(\bar{r}), \quad (14.22)$$

where  $\bar{r} = r/R$ , then:

$$\ln(\bar{r}) = \sum_{m=1}^{\infty} C_{mn} J_n(p_{mn} \bar{r}), \quad (14.23)$$

and:

$$\frac{R}{r} = \sum_{m=1}^{\infty} p_{mn} C_{mn} J'_n\left(p_{mn} \frac{r}{R}\right), \quad (14.24)$$

where:

$$C_{mn} = \frac{2}{J_{n+1}(p_{mn})^2} \int_0^1 x \ln(x) J_n(p_{mn}x) dx. \quad (14.25)$$

The series in (14.24) is rather slow to converge. Nevertheless, using (14.18), (14.24) and (14.25) (with  $n = 0$ ), we find that (14.6) can be written:

$$E_r = \frac{q}{2\pi\varepsilon_0} \frac{4}{LR} \sum_{\ell,m=1}^{\infty} p_m C_m J'_0\left(p_m \frac{r}{R}\right) \cos(k_\ell s) \cos(k_\ell(ct - L/2)). \quad (14.26)$$

As usual, we drop the index  $n$  where  $n = 0$  is implied, i.e.  $C_m = C_{m0}$ . If  $L \ll R$  (i.e. for a short cavity), then for small values of  $\ell$  and  $m$ , we can write:

$$\omega_{\ell m} \approx k_\ell c. \quad (14.27)$$

With the approximation  $p_m C_m \approx -\pi$ , comparing (14.26) and (14.17) leads to:

$$E_{\ell m} = -\frac{q}{4\pi\varepsilon_0 R^2} \frac{8p_m}{(2\ell - 1)} = -\frac{Z_0 c}{4\pi} \frac{q}{R^2} \frac{8p_m}{(2\ell - 1)}. \quad (14.28)$$

Note that we write the quantity  $Z_0 c / 4\pi$  (where  $Z_0 = \sqrt{\mu_0/\varepsilon_0}$  is the impedance of free space) as a separate factor in the expression for  $E_{\ell m}$ . Analysis of wake fields (and wake functions and impedances) is often carried out in cgs units, while here we work in SI units. The conversion from SI to cgs units is readily performed using:

$$\frac{Z_0 c}{4\pi} = 1 \quad (\text{cgs units}). \quad (14.29)$$

Using (14.28), the electric and magnetic fields in a cavity excited by a point charge travelling along the axis can be calculated as functions of time and position within the cavity. For example, from (14.9), the longitudinal field is:

$$E_s = -\frac{Z_0 c}{4\pi} \frac{q}{R^2} \sum_{\ell,m=1}^{\infty} \frac{8p_m}{(2\ell - 1)} J_0(k_m r) \sin(k_\ell s) \cos(\omega_{\ell m}(t - L/2c)), \quad (14.30)$$

where  $\omega_{\ell m}$ ,  $k_m$  and  $k_\ell$  are given by (14.13), (14.14) and (14.15) respectively (with  $n = 0$ ). Given the approximations we have made, the expression for the field in (14.30) is valid for  $L \ll R$ , and for small  $\ell$  and  $m$ . Also, to satisfy causality, we restrict the value of  $t$  such that  $t > L/2c$ , i.e. (14.30) gives the field after the particle has passed through the cavity. The fields are zero ahead of the particle.

It appears from (14.30) that the fields in the cavity will persist indefinitely after a charged particle has passed through. However, there are of course damping mechanisms that will lead to the decay of the field: in particular, some of the energy in the field (especially that associated with components oscillating at higher frequencies) may ‘leak out’ of the cavity along the beam pipe, and some of the energy will be dissipated by currents induced in the walls of the cavity. Processes leading to the decay of the field are not included in our model, but will always be present in practice. The rate of decay of the field in a particular mode (oscillating at a single frequency) is described by the quality factor  $Q$ :

$$Q = 2\pi \frac{\mathcal{E}}{U}, \quad (14.31)$$

where  $\mathcal{E}$  is the energy stored in the cavity in the given mode, and  $U$  is the energy lost per oscillation. For  $Q \gg 1$  (i.e.  $U \ll \mathcal{E}$ ), the energy will decay exponentially:

$$\mathcal{E}(t) = \mathcal{E}(0) \exp\left(-\frac{\omega t}{Q}\right), \quad (14.32)$$

where  $\omega$  is the oscillation frequency. The quality factor of a particular mode in a given cavity depends on the field pattern in the mode, the geometry of the cavity and the properties of the material making up the walls of the cavity. In accelerator beam pipes, cavities that occur ‘incidentally’ (i.e. as a consequence of transitions between different sections, or from the inclusion of vacuum pumps, instrumentation, etc.) tend to have rather low quality factors, so that the induced fields decay quickly, even falling significantly in the time between two successive bunches in the beam. Nevertheless, where cavities do occur in the beam pipe (for example, in the bellows linking two sections of vacuum chamber that need to be allowed some relative movement), it is usually necessary to ‘shield’ the cavities using strips of metal arranged to make the beam pipe look as smooth and continuous as possible. Cavities designed as accelerating structures of course cannot be shielded, and can have very large quality factors; for superconducting cavities,  $Q$  for the accelerating mode may be of order  $10^9$ . The fields induced in these cavities, whether in the fundamental mode or in higher-order modes (HOMs),

must be taken into account when considering the beam dynamics in the accelerator. Cavities are routinely designed with features to damp (i.e. reduce the quality factor of) HOMs, to avoid beam instabilities. HOM damping may be achieved by structures on the cavity acting as short waveguides, which allow the energy in the HOMs to propagate out of the cavity, to be safely dissipated in ferrite loads.

Returning to equation (14.30) for the longitudinal electric field excited in a cavity by a charged particle, the effect of damping can be included by introducing an exponential decay factor. Note that damping also shifts the resonant frequency of an oscillator, since the solution to the equation of motion:

$$\frac{d^2x}{dt^2} + 2\alpha \frac{dx}{dt} + \omega_r^2 x = 0, \quad (14.33)$$

is:

$$x(t) = Ae^{-\alpha t} \cos(\bar{\omega}_r t - \phi), \quad (14.34)$$

where:

$$\bar{\omega}_r = \sqrt{|\omega_r^2 - \alpha^2|}. \quad (14.35)$$

In resonant cavities, higher frequency modes tend to be damped more quickly than lower frequency modes. We therefore assume that the dominant effects of the wake fields can be described just by taking the lowest mode  $\ell = m = 1$  in (14.30). Then,  $\omega_{\ell m} L / 2c \approx \pi/2$ . If we further restrict ourselves to the case  $r = 0$ , the longitudinal electric field will be given by:

$$E_s = -\frac{Z_0 c}{4\pi} \frac{8p_1 q}{R^2} e^{-\alpha t} \sin(k_0 s) \sin(\bar{\omega}_r t), \quad (14.36)$$

where:

$$k_0 = \frac{\pi}{L}, \quad (14.37)$$

$$\omega_r = k_0 c, \quad (14.38)$$

$$\alpha = \frac{\omega_r}{2Q}, \quad (14.39)$$

and  $\bar{\omega}_r$  is given by (14.35). Note that the exponential decay rate of the field amplitude is half the exponential decay rate of the energy in the field.

The longitudinal component of the electric field will lead to longitudinal forces on charged particles following the particle that initially excites fields in the cavity. The longitudinal force depends on the integral of the longitudinal field over the length of the cavity, subject to the constraint:

$$z = s - ct = \text{constant}. \quad (14.40)$$

Performing the integral of the longitudinal field (14.36) over the length of the cavity gives (with various approximations), for fixed  $z < 0$ :

$$\int_{-L/2}^{L/2} E_s(s - ct = z) ds \approx -\frac{Z_0 c}{4\pi} 3q \frac{L^2}{R^2} \frac{\omega_r}{c} e^{\alpha z/c} \cos\left(\frac{\bar{\omega}_r z}{c}\right). \quad (14.41)$$

For  $z > 0$  (i.e. for points ahead of the charge exciting the field in the cavity) the fields are zero. Strictly speaking, the result (14.41) is valid for  $z < -L/2$ ; nevertheless, taking the limit  $z \rightarrow 0^-$ , we see that the integral of the field over the length of the cavity leads to a net force in the  $-s$  direction on a particle just behind the particle generating the wake field, and with the same sign of charge. This is a general property that we expect of wake fields, and in this case is ensured by the ansatz (14.6) for the initial condition for the field in the cavity.

It should be remembered that the expression on the right-hand side of (14.41) is based on a highly simplified model of the fields induced in a cavity by a point charge. It is possible to derive more reliable expressions using more sophisticated models and methods (see, for example, [Zotter and Kheifets (1998)]); however, to obtain proper estimates of the wake fields in particular cases, it is generally necessary to solve Maxwell's equations numerically using a computer code.

In general, there will also be transverse wake forces arising from the fields induced in the cavity. The transverse force on a particle experiencing the wake fields in a cavity will depend on the transverse position of the particle, as well as its distance behind the particle generating the wake fields. The transverse forces and the longitudinal forces are related by Maxwell's equations: this will be discussed further in Sections 14.3 and 14.4. The consequences of the wake forces for the beam dynamics in an accelerator will be discussed in Chapter 15.

## 14.2 Resistive-Wall Wake Fields

Our goal in this section is to calculate the electromagnetic fields around a specified charge distribution in an accelerator beam pipe. The particular case we shall consider is that of a long, straight vacuum chamber with circular cross-section and infinitely thick walls of finite conductivity. The wake fields in this case depend directly on the conductivity of the vacuum chamber, and are known as resistive-wall wake fields. By choosing certain specific charge distributions as sources of the wake field, we are able to construct simplified descriptions of the wake fields in terms of wake functions.

Wake functions provide a convenient representation of the wake fields for studies of beam dynamics, and will be discussed further in Section 14.3.

Geometric wake fields arise from variations in the size or shape of the cross-section of the vacuum chamber through which a beam of particles is moving: an example is the wake field in a cavity discussed in Section 14.1. However, wake fields can also be generated even in a perfectly smooth and uniform beam pipe, simply by virtue of the finite conductivity of the material from which the beam pipe is made. Resistive-wall wake fields are often of significant importance in accelerators; in this section, we derive formulae describing these wake fields (following the approach of Chao [Chao (1993a)]). The fields generated depend not just on the properties (size, shape and electrical conductivity) of the vacuum chamber, but also on the distribution of the charge generating the wake field. We will not consider the most general case; however, to go beyond the simplest situation of a point charge, we shall consider a charge distribution of the form:

$$\rho = \sum_{m=0}^{\infty} \frac{I_m}{\pi(1 + \delta_{m0})a^{m+1}} \cos(m\theta) \delta_D(r - a) \delta_D(s - ct), \quad (14.42)$$

where  $\rho$  is the charge density,  $r$  and  $\theta$  are polar co-ordinates around the reference trajectory  $s$  (with  $\theta$  the angle with respect to the  $x$  axis),  $t$  is the time,  $c$  is the speed of light, and  $a$  is a constant.  $\delta_D(x)$  is a Dirac delta function. In this distribution, all the charge is located within a ring of radius  $a$  centred on the reference trajectory at  $s = ct$  (i.e. the charge is moving along the reference trajectory at the speed of light). The coefficients  $I_m$  represent the (normal) multipole moments of the charge distribution, defined by:

$$I_m = \text{Re} \int (x + iy)^m \rho dV, \quad (14.43)$$

where the integral is taken over all space.

For a uniform ring of total charge  $q$ :

$$\rho = \frac{q}{2\pi a} \delta_D(r - a) \delta_D(s - ct), \quad (14.44)$$

we find:

$$I_m = \begin{cases} q & \text{for } m = 0, \\ 0 & \text{for } m \neq 0. \end{cases} \quad (14.45)$$

For a point-like particle with charge  $q$  moving along the line  $r = a$ ,  $\theta = 0$  (i.e. along the line  $x = a$ ,  $y = 0$ ):

$$\rho = q \frac{\delta_D(\theta)}{a} \delta_D(r - a) \delta_D(s - ct), \quad (14.46)$$

the moments are given by:

$$I_m = qa^m. \quad (14.47)$$

Let us now calculate the electric and magnetic fields around a charge distribution of the form (14.42), moving along the axis of a long, straight vacuum chamber of circular cross-section, radius  $b$ . The electric and magnetic fields must satisfy Maxwell's equations:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}, \quad (14.48)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (14.49)$$

$$\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{j}, \quad (14.50)$$

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad (14.51)$$

where:

$$\mathbf{j} = c\rho\hat{s} \quad (14.52)$$

is the current density. ( $\hat{s}$  is a unit vector in the direction of the reference trajectory.) To find a solution with the specified source, we first expand the fields in Fourier modes in  $\theta$  and  $s$ :

$$\mathbf{E} = \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} (\cos(m\theta)\tilde{E}_r, \sin(m\theta)\tilde{E}_{\theta}, \cos(m\theta)\tilde{E}_s), \quad (14.53)$$

$$\mathbf{B} = \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} (\sin(m\theta)\tilde{B}_r, \cos(m\theta)\tilde{B}_{\theta}, \sin(m\theta)\tilde{B}_s), \quad (14.54)$$

where we have introduced the variable:

$$z = s - ct. \quad (14.55)$$

The dependence on  $\theta$  of the various components is determined by Maxwell's equations. Substituting (14.53) and (14.54) into Maxwell's equations, we find a set of equations for the Fourier coefficients of the fields. From (14.48) and (14.49) we have, respectively:

$$\frac{1}{r} \frac{\partial(r\tilde{E}_r)}{\partial r} + ik\tilde{E}_s + \frac{m}{r}\tilde{E}_{\theta} = \frac{\tilde{\rho}_m(r, k)}{\varepsilon_0}, \quad (14.56)$$

$$\frac{1}{r} \frac{\partial(r\tilde{B}_r)}{\partial r} + ik\tilde{B}_s - \frac{m}{r}\tilde{B}_{\theta} = 0. \quad (14.57)$$

Equation (14.50) leads to:

$$-\frac{m}{r}\tilde{E}_s - ik\tilde{E}_\theta - ikc\tilde{B}_r = 0, \quad (14.58)$$

$$ik\tilde{E}_r - \frac{\partial\tilde{E}_s}{\partial r} - ikc\tilde{B}_\theta = 0, \quad (14.59)$$

$$\frac{m}{r}\tilde{E}_r + \frac{1}{r}\frac{\partial(r\tilde{E}_\theta)}{\partial r} - ikc\tilde{B}_s = \mu_0 c \tilde{\rho}_m(r, k). \quad (14.60)$$

Finally, (14.51) gives:

$$\frac{m}{r}\tilde{B}_s - ik\tilde{B}_\theta + \frac{ik}{c}\tilde{E}_r = 0, \quad (14.61)$$

$$ik\tilde{B}_r - \frac{\partial\tilde{B}_s}{\partial r} + \frac{ik}{c}\tilde{E}_\theta = 0, \quad (14.62)$$

$$-\frac{m}{r}\tilde{B}_r + \frac{1}{r}\frac{\partial(r\tilde{B}_\theta)}{\partial r} + \frac{ik}{c}\tilde{E}_s = 0. \quad (14.63)$$

The quantity  $\tilde{\rho}_m(r, k)$  is defined as the Fourier transform of the charge density, so that:

$$\rho = \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} \cos(m\theta) \tilde{\rho}_m(r, k). \quad (14.64)$$

From (14.42), it follows that:

$$\tilde{\rho}_m(r, k) = \frac{I_m}{\pi(1 + \delta_{m0})a^{m+1}} \delta_D(r - a). \quad (14.65)$$

Since  $\tilde{\rho}_m(r, k) = 0$  except when  $r = a$ , we can integrate the equations (14.56)–(14.63) to give the solution for  $r < a$ :

$$\tilde{E}_r = C_m^{(0)}\bar{r}^{-m-1} + C_m^{(1)}\bar{r}^{m-1} - \frac{ikbC_m^{(2)}}{2(m+1)}\bar{r}^{m+1}, \quad (14.66)$$

$$\tilde{E}_\theta = C_m^{(0)}\bar{r}^{-m-1} - C_m^{(1)}\bar{r}^{m-1} - \frac{ikbC_m^{(2)}}{2(m+1)}\bar{r}^{m+1}, \quad (14.67)$$

$$\tilde{E}_s = C_m^{(2)}\bar{r}^m, \quad (14.68)$$

$$c\tilde{B}_r = -C_m^{(0)}\bar{r}^{-m-1} + \left( C_m^{(1)} + \frac{imC_m^{(2)}}{kb} \right) \bar{r}^{m-1} + \frac{ikbC_m^{(2)}}{2(m+1)}\bar{r}^{m+1}, \quad (14.69)$$

$$c\tilde{B}_\theta = C_m^{(0)}\bar{r}^{-m-1} + \left( C_m^{(1)} + \frac{imC_m^{(2)}}{kb} \right) \bar{r}^{m-1} - \frac{ikbC_m^{(2)}}{2(m+1)}\bar{r}^{m+1}, \quad (14.70)$$

$$c\tilde{B}_s = -C_m^{(2)}\bar{r}^m, \quad (14.71)$$

where:

$$\bar{r} = \frac{r}{b}, \quad (14.72)$$

$C_m^{(0)}$ ,  $C_m^{(1)}$  and  $C_m^{(2)}$  are constants of integration (with units of potential difference) to be determined from the boundary conditions. For  $r > a$  there is an exactly similar set of equations, but with constants of integration  $\bar{C}_m^{(0)}$ ,  $\bar{C}_m^{(1)}$  and  $\bar{C}_m^{(2)}$ . In the particular case  $m = 0$  we need only a reduced set of constants of integration; this case may be included in the expressions (14.66)–(14.71) simply by setting:

$$C_0^{(1)} = \bar{C}_0^{(1)} = 0. \quad (14.73)$$

In the limit  $b \rightarrow \infty$  (i.e. the charge distribution is travelling in free space, without any vacuum chamber) the only non-zero constants of integration are  $C_m^{(0)}$  and  $\bar{C}_m^{(0)}$ ; their values may be obtained as follows. Consider a point-like particle with electric charge  $q$  moving along the line  $r = a$  and  $\theta = 0$  in free space (i.e. in the absence of a vacuum chamber). Integrating Maxwell's equation (14.48) over an infinitely long cylinder of radius  $r_0$  with axis along the path of the charge gives:

$$\int \nabla \cdot \mathbf{E} dV = 2\pi r_0 \int_{-\infty}^{\infty} E_r dz = \frac{q}{\varepsilon_0}, \quad (14.74)$$

where we have used Gauss' theorem to change the integral over the volume of the cylinder to an integral over the surface of the cylinder, and we have used the fact that the electric field has rotational symmetry about the axis of the cylinder. Using (14.53), we can write:

$$\int_{-\infty}^{\infty} E_r dz = \sum_{m=0}^{\infty} \int_{-\infty}^{\infty} dz \int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} \cos(m\theta) \tilde{E}_r = \frac{q}{2\pi\varepsilon_0 r_0}. \quad (14.75)$$

The integral over  $k$  is a Dirac delta function; so the double integral (over  $k$  then over  $z$ ) is simply equal to unity. Using the fact that  $\theta = 0$  and  $r_0 = r - a$ , (14.75) gives:

$$\sum_{m=0}^{\infty} \tilde{E}_r = \frac{q}{2\pi\varepsilon_0 r} \sum_{m=0}^{\infty} \left(\frac{a}{r}\right)^m = \frac{q}{2\pi\varepsilon_0 a} \sum_{m=0}^{\infty} \left(\frac{a}{r}\right)^{m+1}. \quad (14.76)$$

Comparing with (14.66), we deduce that:

$$\bar{C}_m^{(0)} = \frac{q}{2\pi\varepsilon_0 a} \left(\frac{a}{b}\right)^{m+1}. \quad (14.77)$$

Finally, using the fact that the  $m$ th multipole moment of the charge distribution in this case is  $I_m = qa^m$ , this becomes:

$$\bar{C}_m^{(0)} = \frac{I_m}{2\pi\varepsilon_0 b^{m+1}}. \quad (14.78)$$

Although we have shown this result only for the special case of a point charge, it generalises to an arbitrary distribution of charge around a ring  $r = a$  moving at speed  $c$ . Using similar arguments, it can be shown that:

$$C_m^{(0)} = 0 \quad (14.79)$$

for all  $m$ .

Let us assume that the presence of the vacuum chamber does not affect  $C_m^{(0)}$  and  $\bar{C}_m^{(0)}$ . When the chamber is present, we can make the following observations regarding the other constants of integration. First, since the field components  $E_s$  and  $B_s$  must be continuous across  $r = a$ , we can say at once that:

$$C_m^{(2)} = \bar{C}_m^{(2)}. \quad (14.80)$$

Second, since the field components  $\tilde{E}_\theta$  and  $\tilde{B}_r$  must be continuous across  $r = a$ , we find:

$$C_m^{(1)} = \bar{C}_m^{(1)} - \bar{C}_m^{(0)} \left(\frac{b}{a}\right)^{2m}. \quad (14.81)$$

Hence, there are only two unknown constants of integration (for each value of  $m$ ):  $\bar{C}_m^{(1)}$  and  $\bar{C}_m^{(2)}$ . These must be determined from the boundary conditions on the inner surface of the vacuum chamber, i.e. at  $r = b$ .

Let us assume that the material from which the vacuum chamber is made is an ohmic conductor with conductivity  $\sigma$ . Then, the current density in the chamber will be:

$$\mathbf{j} = \sigma \mathbf{E}, \quad (14.82)$$

where  $\mathbf{E}$  is the electric field. To solve Maxwell's equations within the walls of the vacuum chamber, we can again decompose the fields into Fourier components, just as we did in (14.53) and (14.54). We substitute into Maxwell's equations, with zero charge density, and current density given by (14.82). Unfortunately, it is not possible to write an exact solution for the fields in this case. An *approximate* solution is given for  $r > b$  by:

$$\tilde{E}_r = -C_m^{(3)} \frac{k}{\lambda} e^{i\lambda(r-b)}, \quad (14.83)$$

$$\tilde{E}_\theta = C_m^{(4)} \frac{k}{\lambda} e^{i\lambda(r-b)}, \quad (14.84)$$

$$\tilde{E}_s = C_m^{(3)} e^{i\lambda(r-b)}, \quad (14.85)$$

$$c\tilde{B}_r = -C_m^{(4)} \frac{k}{\lambda} e^{i\lambda(r-b)}, \quad (14.86)$$

$$c\tilde{B}_\theta = -C_m^{(3)} \left(\frac{k}{\lambda} + \frac{\lambda}{k}\right) e^{i\lambda(r-b)}, \quad (14.87)$$

$$c\tilde{B}_s = C_m^{(4)} e^{i\lambda(r-b)}, \quad (14.88)$$

where:

$$\lambda^2 = ic\mu_0\sigma k, \quad (14.89)$$

and  $C_m^{(3)}$  and  $C_m^{(4)}$  are constants of integration (with units of potential difference). In order for the fields to decay to zero in the limit  $r \rightarrow \infty$ ,  $\lambda$  (given by the square root of the expression on the right-hand side of (14.89)) must be chosen such that the imaginary part is positive, i.e.

$$\lambda = (i + \text{sgn}(k)) \sqrt{\frac{c\mu_0\sigma|k|}{2}} = \frac{i + \text{sgn}(k)}{\delta_{\text{skin}}}, \quad (14.90)$$

where  $\text{sgn}(k)$  is the sign of  $k$ , i.e. +1 for positive  $k$ , and -1 for negative  $k$ .  $\delta_{\text{skin}}$  is the skin depth of the vacuum chamber wall, i.e. the distance over which the amplitude of an electromagnetic wave propagating through the chamber wall will fall by a factor  $1/e$ . The fields (14.83)–(14.88) satisfy Maxwell's equations in the limit  $r \rightarrow \infty$ .

The remaining constants of integration can now be determined by matching the fields at  $r = b$ . Since there can be surface charges induced by the fields on the surface of the vacuum chamber, we do not expect  $E_r$  to be continuous across  $r = b$ .  $E_s$  and  $B_s$  should be continuous across the boundary at  $r = b$ ; hence:

$$C_m^{(3)} = \bar{C}_m^{(2)}, \quad (14.91)$$

$$C_m^{(4)} = -\bar{C}_m^{(2)}. \quad (14.92)$$

Then, matching  $E_\theta$  and  $B_\theta$  across the boundary at  $r = b$  gives, for  $m > 0$ :

$$\bar{C}_m^{(2)} = \frac{2\bar{C}_m^{(0)}}{\frac{ikb}{m+1} - \frac{\lambda}{k} - \frac{im}{kb} - 2\frac{k}{\lambda}}, \quad (14.93)$$

$$\bar{C}_m^{(1)} = -\frac{im + b\lambda}{2kb}\bar{C}_m^{(2)}. \quad (14.94)$$

For  $m = 0$ , there is only one constant of integration to be found ( $\bar{C}_0^{(2)}$ ), since we have already set  $\bar{C}_0^{(1)} = 0$ . Matching the single field component  $B_\theta$  across the boundary at  $r = b$  gives:

$$\bar{C}_0^{(2)} = \frac{2\bar{C}_0^{(0)}}{ikb - 2\frac{\lambda}{k} - 2\frac{k}{\lambda}}. \quad (14.95)$$

It is convenient to write approximate expressions for  $\bar{C}_m^{(2)}$  and  $\bar{C}_m^{(1)}$  by making a series expansion in terms of a dimensionless parameter  $\xi$  defined by:

$$\xi = \sqrt{\frac{2}{c\mu_0\sigma b}}. \quad (14.96)$$

Substituting for  $\lambda$  from (14.96) in (14.93) and (14.94), and expanding to first order in  $\xi$  gives for  $m > 0$ :

$$\frac{\bar{C}_m^{(2)}}{\bar{C}_m^{(0)}} = (i \operatorname{sgn}(k) - 1)(|k|b)^{1/2}\xi + O((|k|b)^2\xi^2), \quad (14.97)$$

$$\frac{\bar{C}_m^{(1)}}{\bar{C}_m^{(0)}} = 1 + (i \operatorname{sgn}(k) + 1)\frac{(|k|b)^{3/2}}{2(m+1)}\xi + O((|k|b)^3\xi^2). \quad (14.98)$$

For  $m = 0$ , we find:

$$\frac{\bar{C}_0^{(2)}}{\bar{C}_0^{(0)}} = (i \operatorname{sgn}(k) - 1)\frac{(|k|b)^{1/2}}{2}\xi + O((|k|b)^2\xi^2), \quad (14.99)$$

$$\bar{C}_0^{(1)} = 0. \quad (14.100)$$

Note that  $\xi$  is independent of  $k$ : the dependence of  $\bar{C}_m^{(2)}$  and  $\bar{C}_m^{(1)}$  on  $k$  appears explicitly in (14.97) and (14.98). This will be convenient when we come to do the Fourier transforms to produce expressions for the fields. Note also that inspecting higher-order terms in the series expansions for  $\bar{C}_m^{(2)}$  and  $\bar{C}_m^{(1)}$ , there is a convergence condition:

$$|k|b \ll \xi^{-2/3}. \quad (14.101)$$

When we transform from frequency ( $k$ ) space back to co-ordinate ( $z$ ) space, this is equivalent to the condition:

$$|z| \gg b\xi^{2/3} = \sqrt[3]{\frac{2b^2}{c\mu_0\sigma}}. \quad (14.102)$$

That is, there is a minimum value of  $|z|$ , determined by the conductivity and radius of the beam pipe, for which our results will be valid. We shall discuss the short-range wake further below; but for now, we observe that for an aluminium vacuum chamber (conductivity  $\sigma = 3.55 \times 10^7 \Omega^{-1}\text{m}^{-1}$ ) with 40 mm diameter, the value of  $\xi$  is approximately  $10^{-4}$ . The wake field expressions we derive using the first terms in the series expansions (14.99) and (14.100) will be valid for  $|z| \gg 40 \mu\text{m}$ . It may not be possible to use them to understand the effects of the wake fields on the dynamics of particles within individual bunches, but they will be valid for describing the coupling between different bunches in a storage ring arising from resistive-wall wake fields.

We now have expressions for the fields inside the beam pipe (and approximate expressions for the field in the wall of the vacuum chamber). A summary of the coefficients appearing in the expressions (14.66)–(14.71) is given in Table 14.1.

Table 14.1 Coefficients in the expressions (14.66)–(14.71) for resistive-wall wake fields following a charge distribution with multipole moment  $I_m$ . The expressions are valid for  $kb \ll \xi^{-2/3}$ , where  $\xi$  is given by (14.96). Coefficients  $C_m^{(0)}$ ,  $C_m^{(1)}$  and  $C_m^{(2)}$  are for the range  $0 < r \leq a$ ; and coefficients  $\bar{C}_m^{(0)}$ ,  $\bar{C}_m^{(1)}$  and  $\bar{C}_m^{(2)}$  are for the range  $a \leq r < b$ .

field coefficients	$m = 0$	$m > 0$
$C_m^{(0)}$	0	0
$C_m^{(1)}/\bar{C}_m^{(0)}$	0	$1 + (i \operatorname{sgn}(k) + 1) \frac{( k b)^{3/2}}{2(m+1)} \xi - \left(\frac{b}{a}\right)^{2m}$
$C_m^{(2)}/\bar{C}_m^{(0)}$	$(i \operatorname{sgn}(k) - 1) \frac{( k b)^{1/2}}{2} \xi$	$(i \operatorname{sgn}(k) - 1) ( k b)^{1/2} \xi$
$\bar{C}_m^{(0)}$	$q/2\pi\varepsilon_0 b$	$I_m/2\pi\varepsilon_0 b^{m+1}$
$\bar{C}_m^{(1)}/\bar{C}_m^{(0)}$	0	$1 + (i \operatorname{sgn}(k) + 1) \frac{( k b)^{3/2}}{2(m+1)} \xi$
$\bar{C}_m^{(2)}/\bar{C}_m^{(0)}$	$(i \operatorname{sgn}(k) - 1) \frac{( k b)^{1/2}}{2} \xi$	$(i \operatorname{sgn}(k) - 1) ( k b)^{1/2} \xi$

The electric and magnetic fields are given in terms of their Fourier coefficients by (14.53) and (14.54). For the longitudinal components, we can make use of the results:

$$\int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} (i \operatorname{sgn}(k) - 1) |k|^{1/2} = \frac{1 - \operatorname{sgn}(z)}{2\sqrt{2\pi}|z|^{3/2}}, \quad (14.103)$$

$$\int_{-\infty}^{\infty} e^{ikz} \frac{dk}{2\pi} (i \operatorname{sgn}(k) + 1) |k|^{3/2} = -\frac{3(1 - \operatorname{sgn}(z))}{4\sqrt{2\pi}|z|^{5/2}}. \quad (14.104)$$

Note that the form of the expressions on the right-hand side of (14.103) and (14.104) ensures that the fields around the charge vanish for  $z > 0$ : that is, the fields satisfy causality. We find that the longitudinal components of the field can be written (for  $0 < r < b$ ,  $z < 0$  and  $|z| \gg b\xi^{2/3}$ ):

$$E_s = \sqrt{\left(\frac{Z_0 c}{4\pi}\right) \frac{c}{\sigma}} \sum_{m=0}^{\infty} \frac{I_m r^m \cos(m\theta)}{(1 + \delta_{m0}) \pi b^{2m+1}} \frac{1}{|z|^{3/2}}, \quad (14.105)$$

$$cB_s = -\sqrt{\left(\frac{Z_0 c}{4\pi}\right) \frac{c}{\sigma}} \sum_{m=0}^{\infty} \frac{I_m r^m \sin(m\theta)}{\pi b^{2m+1}} \frac{1}{|z|^{3/2}}, \quad (14.106)$$

where  $Z_0$  is the impedance of free space:

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}. \quad (14.107)$$

Equation (14.105) has been written so that the factor:

$$\frac{Z_0 c}{4\pi} = \frac{1}{4\pi\varepsilon_0} \quad (\text{SI units}) \quad (14.108)$$

appears explicitly. As previously mentioned, analysis of wake fields and impedances is often carried out using cgs units. Here, we always use SI units; however, the conversion of formulae to cgs units can easily be accomplished using:

$$\frac{Z_0c}{4\pi} = 1 \quad (\text{cgs units}). \quad (14.109)$$

We can perform the Fourier transforms required to find the remaining field components. In the plane  $z = 0$ , delta functions appear corresponding to the ‘pancake’ field around the charge distribution; however, we drop these terms because they are excluded by the condition  $|z| \gg b\xi^{2/3}$ .

In the region  $0 < r < b$ ,  $z < 0$  and  $|z| \gg b\xi^{2/3}$ , the azimuthal component of the electric field is:

$$E_\theta = -\sqrt{\left(\frac{Z_0c}{4\pi}\right)\frac{c}{\sigma}} \sum_{m=0}^{\infty} \frac{3I_m}{4\pi b^{2m+1}} \frac{r^{m-1}(r^2 - b^2)}{m+1} \sin(m\theta) \frac{1}{|z|^{5/2}}, \quad (14.110)$$

and the radial component of the electric field is:

$$E_r = -\sqrt{\left(\frac{Z_0c}{4\pi}\right)\frac{c}{\sigma}} \times \left( \frac{3q}{8\pi b} \frac{r}{|z|^{5/2}} + \sum_{m=1}^{\infty} \frac{3I_m}{4\pi b^{2m+1}} \frac{r^{m-1}(r^2 + b^2)}{m+1} \cos(m\theta) \frac{1}{|z|^{5/2}} \right). \quad (14.111)$$

Finally, the transverse components of the magnetic field are, for  $0 < r < b$ ,  $z < 0$  and  $|z| \gg b\xi^{2/3}$ :

$$cB_\theta = E_r - \sqrt{\left(\frac{Z_0c}{4\pi}\right)\frac{c}{\sigma}} \sum_{m=1}^{\infty} \frac{2I_m m r^{m-1} \cos(m\theta)}{\pi b^{2m+1}} \frac{1}{|z|^{1/2}}, \quad (14.112)$$

and:

$$cB_r = -E_\theta - \sqrt{\left(\frac{Z_0c}{4\pi}\right)\frac{c}{\sigma}} \sum_{m=0}^{\infty} \frac{2I_m m r^{m-1} \sin(m\theta)}{\pi b^{2m+1}} \frac{1}{|z|^{1/2}}. \quad (14.113)$$

It is worth making a few remarks about these results for the resistive-wall wake fields. First, recall that for the fields in the chamber wall to vanish in the limit  $r \rightarrow \infty$ , we had to make a choice for  $\lambda$  such that the imaginary part was positive. This leads to expressions for the electric and magnetic fields such that the fields vanish for  $z > 0$ , i.e. in the region ahead of the charge generating the fields. In other words the choice for the complex phase of  $\lambda$  is determined not just by the requirement that the

fields decay with radial distance in the walls of the vacuum chamber, but also by the need to satisfy causality.

Second, notice that although we were able to write down only *approximate* expressions for the fields in the walls of the vacuum chamber, these expressions were used only to determine (by matching the fields at the boundary  $r = b$ ) the coefficients for the expressions (14.66)–(14.71) for the fields within the vacuum chamber. The expressions (14.66)–(14.71) provide exact solutions to Maxwell's equations for any given values of the coefficients. We made further approximations to the coefficients when performing the Fourier transform from frequency ( $k$ ) space to co-ordinate ( $z$ ) space. However, the approximation is again in the coefficients of the various terms, rather than in the form of the dependence on the co-ordinates. The consequence is that although the fields (14.105)–(14.113) satisfy Maxwell's equations exactly, for some charge distribution and conducting boundary, they do not necessarily provide correct solutions for the particular case we are considering, i.e. a thin ring of charge with multipole moments  $I_m$ , travelling at speed  $c$  along a cylindrical vacuum chamber.

An explicit condition on the validity of the expressions for the fields (14.66)–(14.71) is that  $z < 0$ , and  $|z| \gg b\xi^{2/3}$ . Note that taking the limit  $z \rightarrow 0^-$  in (14.105) leads to an expression for an electric field producing a force in the same direction as the charge is moving. If this were the case, the charge would be accelerated by its own wake field, which would violate conservation of energy. To show that this does not happen, we need to consider more exact expressions for the fields, valid for  $|z| \ll b\xi^{2/3}$ . Appropriate expressions can be obtained from the Fourier transform of (14.93) and (14.95). The results for the fields, obtained by Bane [Bane (1991); Bane and Sands (1995)], are, for  $m = 0$  and  $z < 0$ :

$$E_s = -\frac{16q}{4\pi\varepsilon_0 b^2} \left( \frac{e^u}{3} \cos(\sqrt{3}u) - \frac{\sqrt{2}}{\pi} \int_0^\infty \frac{x^2 e^{ux^2}}{x^6 + 8} dx \right), \quad (14.114)$$

$$\begin{aligned} E_r = cB_\theta &= \frac{8qr}{4\pi\varepsilon_0 b^3 \xi^{2/3}} \\ &\times \left( \frac{e^u}{3} \cos(\sqrt{3}u) - \frac{e^u}{\sqrt{3}} \sin(\sqrt{3}u) - \frac{\sqrt{2}}{\pi} \int_0^\infty \frac{x^4 e^{ux^2}}{x^6 + 8} dx \right), \end{aligned} \quad (14.115)$$

where:

$$u = \frac{z}{b\xi^{2/3}}. \quad (14.116)$$

For the longitudinal field, we find:

$$\lim_{z \rightarrow 0^-} E_s = -\frac{q}{\pi \varepsilon_0 b^2}. \quad (14.117)$$

As we shall see later, an ultra-relativistic particle in an accelerator beam pipe experiences a retarding force corresponding to half the wake field in the limit  $z \rightarrow 0^-$ . This result is known as the *fundamental theorem of beam loading* [Wilson (1978); Wilson and Griffin (1981)], and applies generally, not just to the case of resistive-wall wake fields. A consequence of the fundamental theorem of beam loading is that a particle generating wake fields will lose energy to those wake fields. In the case of resistive-wall wake fields, the energy is dissipated by the currents driven by the electric field in the walls of the vacuum chamber. We see from (14.117) that in the limit  $z \rightarrow 0^-$ , the longitudinal field  $E_s$  is independent of the conductivity of the vacuum chamber. This seems curious, since we would expect that in the limit of perfect conductivity there would be no energy dissipated in the chamber walls. However, we are in fact prevented from taking the limit  $\sigma \rightarrow \infty$ , since in that case  $\xi \rightarrow 0$ , and the variable  $u$  in (14.114) is not well defined. In other words, the expressions for the wake fields given by (14.114) and (14.115) are valid only for finite conductivity.

Finally, it is worth noting that although we derived the expressions for the wake fields with the assumption that the walls of the vacuum chamber are infinitely thick, it is possible to apply these expressions even if the walls have finite thickness, as long as an upper limit is placed on the range of  $|z|$ . Consider a vacuum chamber with walls of thickness  $w$ . If:

$$w \gg \delta_{\text{skin}} = \sqrt{\frac{b}{|k|}} \xi, \quad (14.118)$$

where  $\delta_{\text{skin}}$  is the skin depth for an electromagnetic wave (of frequency  $|k|c$  in free space), then the fields will be very similar to those in the case  $w \rightarrow \infty$ . Using  $z = 2\pi/k$ , (14.118) leads to an upper limit on  $|z|$ , given by:

$$|z| \ll \frac{2\pi w^2}{b\xi^2}. \quad (14.119)$$

For an aluminium vacuum chamber with 40 mm diameter ( $\xi \approx 10^{-4}$ ), and 2 mm wall thickness, this gives  $|z| \ll 167$  km. This is a considerable distance; however, in storage rings it is possible that resistive-wall wake fields can have significant effects even after many revolution periods. Therefore, although it is possible to use the expressions (14.66)–(14.71) in many practical cases, it is worth bearing in mind both the lower and the upper limits

of the range of validity on  $|z|$ . The expressions for the wake field from a ‘thick-walled’ chamber can even be applied when a thin dielectric or semiconductor coating is applied on the inner surface of the vacuum chamber, as is sometimes done to improve the vacuum properties, or suppress electron cloud effects (where electrons emitted from the chamber surface can collect in the potential of a proton or positron beam). The skin depth in a semiconductor coating is typically very large except at very high frequencies; thus, although a semiconductor coating may affect the wake fields at high frequencies (i.e. short  $|z|$ ) [Hock and Wolski (2011)], the impact of such a coating on beam instabilities driven by resistive-wall wake fields will often be limited.

### 14.3 Wake Functions

The analysis carried out in Sections 14.1 and 14.2 shows that, even in relatively simple cases, calculating the wake fields in an accelerator is no easy matter, and the expressions for the fields can be rather complicated. In certain cases, there are standard expressions that can be used, but mostly it is necessary to compute the electric and magnetic fields numerically, using a computer code to solve Maxwell’s equations with specified boundary conditions.

In Chapter 15, we shall discuss the impact of wake fields on the beam dynamics of particles in accelerators. Having obtained the wake fields (by whatever method) for the various components in the accelerator, the next step is to determine the forces on particles within the beam. In the case of ultra-relativistic beams, we can make use of some simplifying approximations that lead to useful properties of the forces on particles arising from wake fields. These properties, which ultimately derive from Maxwell’s equations, are used to define *wake functions* that provide a convenient description (in the time domain) of the forces on particles resulting from wake fields. As we shall see, it is often useful when investigating the beam dynamics to have a description of the wake forces in the frequency domain: such a description is obtained by a Fourier transform of a wake function, which leads to the impedance. Our goal in the remainder of this chapter is to define wake fields and impedances, and to understand some of their main properties. We discuss wake functions in the present section; impedances are discussed in Section 14.4.

Consider a charge distribution  $\rho$  moving along an accelerator beam line, with  $\rho$  given by (14.42):

$$\rho = \frac{I_m}{\pi(1 + \delta_{m0})a^{m+1}} \cos(m\theta) \delta_D(r - a) \delta_D(s - ct). \quad (14.120)$$

The charge forms a ring of radius  $a$ , with multipole moment  $I_m$  (14.43). At time  $t$ , the entire charge is in the plane  $s = ct$ , where  $s$  is the distance along the reference trajectory. As the charge distribution passes through a section of the vacuum chamber, it generates wake fields such that a point-like particle of charge  $q$ , and travelling some distance behind the leading charge, experiences a force  $\mathbf{F}$ . If the trailing particle is moving with velocity  $\mathbf{v}$ , the force is given by the Lorentz force equation:

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (14.121)$$

where  $\mathbf{E}$  is the electric field and  $\mathbf{B}$  is the magnetic field. With some simplifying assumptions, it is possible to find a set of equations for  $\mathbf{F}$  that allow the force on the trailing particle to be expressed in terms of a single function of the distance between the particle and the charge distribution: this function (the *wake function*) is related to the wake fields, examples of which were discussed in Sections 14.1 and 14.2. The simplifying assumptions that we need are that the (leading) charge distribution and the (trailing) point-like particle are both moving at the speed of light; and that the (trailing) particle is moving directly along the reference trajectory.

In this section, we shall first derive, starting from Maxwell's equations, a set of equations for the force on a particle resulting from the wake fields generated by a charge distribution travelling ahead of it. This will lead to a definition of the wake function as a solution to the equations for the force. In Section 14.4 we shall define the impedance as the Fourier transform of the wake field, i.e. as a description of the wake fields in frequency space. Then, in Chapter 15 we shall use wake functions and impedances to discuss the impact of wake fields on the beam dynamics.

Consider a particular section of beam pipe of length  $L$  and with the centre on the reference trajectory at  $s = 0$ . A charge distribution generating a wake field in the beam pipe is at  $s = 0$  at time  $t = 0$ , and is travelling at speed close to  $c$ , parallel to the reference trajectory. We define the variable  $z$ :

$$z = s - ct, \quad (14.122)$$

so that the trailing particle has  $z < 0$ ; if both the particle and the charge distribution are moving along the reference trajectory at the speed of light,

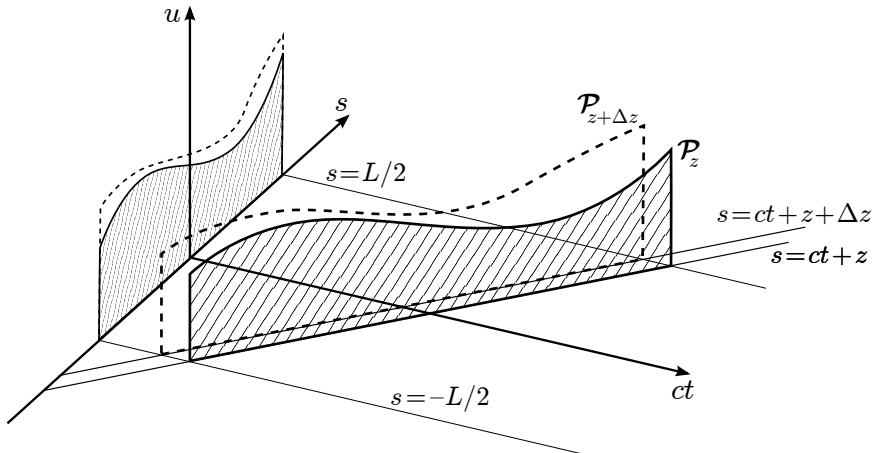


Fig. 14.2 Integration of the force on a charged particle from a wake field. The wake field exists in a beam line component extending from  $-L/2$  to  $L/2$  along the  $s$  axis. The force  $f$  on a particle from the wake field defines a surface  $u = f(s, ct)$  in the space shown. The particle experiencing the force is at a fixed value of  $z = s - ct$ , so the integral of the force from the wake field is represented by the area of the projection onto the plane  $ct = 0$  of the shaded area  $\mathcal{P}_z$ .

then  $z$  is constant, and  $|z|$  is the distance between the charge distribution and the trailing particle.

The electric and magnetic fields in the beam pipe are, in general, functions of the co-ordinates  $r$ ,  $\theta$  and  $s$ , and of the time  $t$ . Using the fact that the particle and the charge distribution are travelling at the speed of light, the equations describing the fields seen by the trailing particle can be expressed in terms of the single variable  $z$ , rather than the two variables  $s$  and  $t$ . This is done as follows. Let  $[f]_z$  represent the integral of a quantity  $f$  (which may be a component of the electromagnetic field, or a derivative of a field component) over a path along the reference trajectory from  $s = -L/2$  to  $s = +L/2$ . At each step along the path,  $f$  is evaluated at position  $s$  and time  $ct = s - z$ , for constant  $z$ :

$$[f]_z = \int_{-L/2}^{L/2} f(s, ct = s - z) ds. \quad (14.123)$$

Neglecting (for the moment) the dependence of  $f$  on  $r$  and  $\theta$ , we can represent  $f$  as a surface in a three-dimensional space with axes  $ct$  and  $s$  in the horizontal plane, and a vertical axis that we call  $u$  (so that the surface representing  $f$  is given by  $u = f(s, ct)$ ). The integral in (14.123) is taken along a path that is a diagonal line in a (horizontal) plane with axes  $ct$

and  $s$ : the integral gives the area, projected onto the plane  $ct = 0$ , of a (vertical) plane  $\mathcal{P}_z$  on which  $s = z + ct$ , with upper and lower boundaries given by  $u = f$  and  $u = 0$ , and bounded at the sides by  $s = -L/2$  and  $s = +L/2$  (see Fig. 14.2). The quantity  $[\partial f / \partial ct]_z \Delta z$  can be represented by the difference in the projected areas of two planes  $\mathcal{P}_z$  and  $\mathcal{P}_{z+\Delta z}$ . Then, it can be seen that:

$$\frac{1}{c} \left[ \frac{\partial f}{\partial t} \right]_z \Delta z = \int_{-L/2}^{L/2} f(s, ct = s - (z + \Delta z)) - f(s, ct = s - z) ds, \quad (14.124)$$

which leads to:

$$\frac{1}{c} \left[ \frac{\partial f}{\partial t} \right]_z = -\frac{\partial}{\partial z} [f]_z. \quad (14.125)$$

Similarly, we find that:

$$\left[ \frac{\partial f}{\partial s} \right]_z \approx \frac{\partial}{\partial z} [f]_z. \quad (14.126)$$

Note that in equation (14.126) there is an approximation, while equation (14.125) is exact. The approximation arises because  $[\partial f / \partial s]_z \Delta z$  represents the difference in projected area of two planes displaced along the  $s$  axis: both planes cannot have the same boundaries  $s = -L/2$  and  $s = +L/2$  that define the limits of the integration. In general, this will lead to some error, but we can assume that the size of the error will be reduced if we increase the range of the integration (i.e. integrate over a larger length  $L$ ). We observe, incidentally, that if  $f$  is such that  $f(s, ct) = f(z)$  (i.e.  $f$  can be expressed purely as a function of  $z = s - ct$ ), then (14.126) is an exact relationship. This will be the case if the system is invariant under translations along  $s$ : an example is the resistive-wall wake field generated by a particle moving along a long, straight beam pipe with uniform cross-section.

We now take Maxwell's equations, and integrate along the reference trajectory from  $s = -L/2$  to  $s = +L/2$ , subject to the constraint  $s - ct = z = \text{constant}$ . Using (14.125) and (14.126) we can replace derivatives (inside the integrals) with respect to  $s$  and  $t$  with derivatives (outside the integrals) with respect to  $z$ . Derivatives with respect to other co-ordinates can be taken outside the integrals, and left unchanged. It is straightforward to write down the resulting equations. We do not give them all; only, as an example, Maxwell's equation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad (14.127)$$

after applying the integration and substituting for the derivatives with respect to  $s$  and  $t$ , yields:

$$\frac{1}{r} \frac{\partial [E_s]_z}{\partial \theta} + \frac{\partial [E_\theta]_z}{\partial z} = c \frac{\partial [B_r]_z}{\partial z}, \quad (14.128)$$

$$-\frac{\partial [E_r]_z}{\partial z} - \frac{\partial [E_s]_z}{\partial r} = c \frac{\partial [B_\theta]_z}{\partial z}, \quad (14.129)$$

$$\frac{1}{r} \frac{\partial (r[E_\theta]_z)}{\partial r} - \frac{1}{r} \frac{\partial [E_r]_z}{\partial \theta} = c \frac{\partial [B_s]_r}{\partial z}. \quad (14.130)$$

For the inhomogeneous equations, since all particles are moving parallel to the  $s$  axis at the speed of light, the current density  $\mathbf{j}$  is related to the charge density  $\rho$  by:

$$\mathbf{j} = (0, 0, c\rho). \quad (14.131)$$

If the trailing particle is moving at the speed of light along the  $s$  axis, so that  $\mathbf{v} = (0, 0, c)$ , integrating the components of the Lorentz force gives:

$$[F_r]_z = q([E_r]_z - [B_\theta]_z), \quad (14.132)$$

$$[F_\theta]_z = q([E_\theta]_z + [B_r]_z), \quad (14.133)$$

$$[F_s]_z = q [E_s]_z. \quad (14.134)$$

The next step is to substitute the force for the electric field in Maxwell's equations (integrated along  $s$ ), and eliminate  $[B_r]_z$  and  $[B_\theta]_z$ . The resulting equations are:

$$\frac{\partial [F_r]_z}{\partial z} = \frac{\partial [F_s]_z}{\partial r} = -\frac{qc}{r} \frac{\partial [B_s]_z}{\partial \theta}, \quad (14.135)$$

$$\frac{\partial [F_\theta]_z}{\partial z} = \frac{1}{r} \frac{\partial [F_s]_z}{\partial \theta} = qc \frac{\partial [B_s]_z}{\partial r}, \quad (14.136)$$

$$\frac{\partial (r[F_r]_z)}{\partial r} = -\frac{\partial [F_\theta]_z}{\partial \theta}, \quad (14.137)$$

$$\frac{\partial (r[F_\theta]_z)}{\partial r} = \frac{\partial [F_r]_z}{\partial \theta}. \quad (14.138)$$

By substitution, it is straightforward to show that these equations have a solution that can be expressed as:

$$[F_r]_z = -qI_m W_m(z) mr^{m-1} \cos(m\theta), \quad (14.139)$$

$$[F_\theta]_z = qI_m W_m(z) mr^{m-1} \sin(m\theta), \quad (14.140)$$

$$[F_s]_z = -qI_m W'_m(z) r^m \cos(m\theta), \quad (14.141)$$

$$[B_s]_z = \frac{I_m}{c} W'_m(z) r^m \sin(m\theta). \quad (14.142)$$

Note that the prime indicates the derivative of  $W_m(z)$  with respect to  $z$ . The functions  $W_m(z)$  are the *wake functions*: by solving Maxwell's equations with appropriate boundary conditions for the fields generated by the charge distribution with moment  $I_m$ , as we did for the resistive-wall wake fields in Section 14.2, we can find an explicit expression for the wake functions. For example, in the case of the resistive-wall wake field, with longitudinal electric field component given by (14.105), the wake function over a length  $L$  of the beam pipe is (for  $z < 0$  and  $|z| \gg b\xi^{2/3}$ ):

$$W_m(z) = -\sqrt{\left(\frac{Z_0c}{4\pi}\right)\frac{c}{\sigma}} \frac{2L}{(1 + \delta_{m0})\pi b^{2m+1}} \frac{1}{|z|^{1/2}}. \quad (14.143)$$

Using (14.139) and (14.140), we then obtain transverse components of the integrated force that are consistent with the fields given by (14.112) and (14.113). Also, using (14.142), we obtain an expression for the integrated longitudinal magnetic field component that is consistent with (14.106).

In the case of a resonant cavity, the  $m = 0$  wake function can be found by integrating equation (14.41) with respect to  $z$ . The result is, for  $z < 0$ :

$$W_0(z) = \frac{Z_0c}{4\pi} \frac{3L^2}{R^2} \frac{\omega_r}{\bar{\omega}_r} e^{\alpha z/c} \sin\left(\frac{\bar{\omega}_r z}{c}\right), \quad (14.144)$$

where  $R$  is the radius of the cavity,  $L \ll R$  is the length of the cavity,  $\omega_r$  is the frequency of the fundamental resonant mode,  $\alpha = \omega_r/2Q$  characterises the damping of the fundamental mode (with quality factor  $Q$ ), and:

$$\bar{\omega}_r = \sqrt{|\omega_r^2 - \alpha^2|}. \quad (14.145)$$

As discussed in Section 14.1, this result is based on a simplified model of the fields induced in a resonant cavity by a point charge. However, the dependence on  $z$  (effectively, a damped oscillation with parameters related to the resonant frequency and quality factor of the cavity) is a convenient and commonly used model for resonator wake fields [Ng (1999)].

The solutions (14.139)–(14.142) are appropriate for axially symmetric boundary conditions within the accelerator beam pipe. More complicated expressions are needed to express the integrated fields in the case of structures that are not axially symmetric. It should also be remembered that the equations leading to the solutions (14.139)–(14.142) were derived with the assumption that the charge density generating the wake fields and the charge ‘observing’ the wake fields travel at the speed of light along the vacuum chamber. We therefore expect (14.139)–(14.142) to be valid only for high energy beams, with relativistic factor  $\gamma_0 \gg 1$ .

The quantity  $[\mathbf{F}]_z$  has units of energy: it is generally referred to as the *wake potential*. In SI units, the wake function  $W_m(z)$  has units of  $\text{VC}^{-1}\text{m}^{1-2m}$  (or, equivalently,  $\Omega\text{s}^{-1}\text{m}^{1-2m}$ ). In particular,  $W_1(z)$  (which often dominates the transverse effects) has units of  $\text{VC}^{-1}\text{m}^{-1}$  (or  $\Omega\text{s}^{-1}\text{m}^{-1}$ ), and  $W'_1(z)$  (which often dominates the longitudinal effects) has units of  $\text{VC}^{-1}$  (or  $\Omega\text{s}^{-1}$ ).

There are some interesting (and useful) properties of wake functions that follow from (14.139)–(14.141). First, we observe that the transverse and longitudinal wake potentials are directly related. If we write:

$$[\mathbf{F}_\perp]_z = [F_r]_z \hat{r} + [F_\theta]_z \hat{\theta}, \quad (14.146)$$

where  $\hat{r}$  and  $\hat{\theta}$  are unit vectors in the radial and azimuthal directions respectively, then it follows from (14.139), (14.140) and (14.141) that the relationship between the transverse and longitudinal parts of the wake potentials can be expressed as:

$$\nabla_\perp [F_s]_z = \frac{\partial}{\partial z} [\mathbf{F}_\perp]_z, \quad (14.147)$$

where  $\nabla_\perp$  is the transverse part of the grad operator. Equation (14.147) is known as the *Panofsky–Wenzel theorem* [Panofsky and Wenzel (1956)].

A number of other interesting general properties of wake functions follow from various physical constraints. First, by causality:

$$W_m(z) = 0 \quad \text{if} \quad z > 0. \quad (14.148)$$

Second, if two particles have a very short separation  $|z|$ , the second particle should be *decelerated* by the wake fields from the first particle. This means that:

$$W'_m(z) > 0 \quad \text{as} \quad z \rightarrow 0^-. \quad (14.149)$$

Although the longitudinal wake potential can be non-zero in the limit  $z \rightarrow 0^-$ , the transverse wake potential should approach zero for small  $z$ : the longitudinal wake potential in the limit  $z \rightarrow 0^-$  is related to the energy lost by a particle to the wake fields; but we do not expect (in an axially symmetric beam pipe) to see a particle experience a transverse deflection from its own wake field. This suggests that:

$$W_m(z) \rightarrow 0 \quad \text{as} \quad z \rightarrow 0^-. \quad (14.150)$$

Since  $W'_m(z)$  is positive for small (negative)  $z$ ,  $W_m(z)$  must approach zero at  $z = 0$  with positive gradient in the region  $z < 0$ . Hence, for  $z < 0$ ,  $W_m(z)$  should be a sine-like function of  $z$ , and  $W'_m(z)$  should be a cosine-like function of  $z$ .

Now consider the case of a resonant cavity in which the wake potential is described by:

$$[F_s]_z = F_0 \cos(kz), \quad (14.151)$$

for some constants  $F_0$  and  $k$  (and  $z < 0$ ). This is based on the simple model discussed in Section 14.1, and neglects energy dissipation in the cavity. Suppose that a particle sees a fraction  $\eta$  of the wake field that it generates in the limit  $z \rightarrow 0^-$ . The energy change of the particle as it passes through the cavity is:

$$\Delta\mathcal{E}_1 = \eta[F_s]_{z \rightarrow 0^-} = \eta F_0. \quad (14.152)$$

Now suppose a second particle crosses the cavity at a distance  $z = \pi/k$  behind the first. This particle sees the same fraction of its own wake field as the first particle; but it also sees the wake field from the first particle. Therefore, the total change in energy of the second particle as it crosses the cavity is:

$$\Delta\mathcal{E}_2 = \eta[F_s]_{z \rightarrow 0^-} - F_0 = (\eta - 1)F_0. \quad (14.153)$$

Since the field in the cavity changes phase by  $\pi$  in the time between the two particles, the fields induced by the second particle exactly cancel the fields induced by the first particle. Then, after the second particle leaves the cavity, there is no electromagnetic energy remaining in the cavity. By conservation of energy, the total energy change of the two particles must also be zero; hence:

$$\Delta\mathcal{E}_1 + \Delta\mathcal{E}_2 = (2\eta - 1)F_0 = 0. \quad (14.154)$$

Therefore:

$$\eta = \frac{1}{2}. \quad (14.155)$$

In other words, a particle sees half of its own wake field: this is the *fundamental theorem of beam loading* [Wilson (1978); Wilson and Griffin (1981)].

Finally, consider the resistive-wall wake fields generated by a point charge moving at a radial distance  $a$  from the reference trajectory (the axis of the beam pipe). We see from (14.139)–(14.141), (14.143) and (14.47) that in this case the longitudinal wake potential scales with  $m$  as  $a^m r^m / b^{2m+1}$ , and the transverse wake potentials scale as  $ma^m r^{m-1} / b^{2m+1}$ , where  $b$  is the beam pipe radius, and  $r$  is the radial distance from the beam pipe axis at which the wake fields are observed. This implies that if  $a \ll b$  and  $r \ll b$ , the dominant contribution to the longitudinal wake field comes from the mode  $m = 0$ , and the dominant contribution to the transverse wake field

comes from the mode  $m = 1$ . It turns out that this is a common (though by no means universal) situation, including in particular the case of geometric wake fields generated by structures of a similar size to the beam pipe radius. The analysis of the beam dynamics can be greatly simplified by including only the dominant wake functions. This leads to a definition of the *longitudinal wake function*  $W_{\parallel}(z)$  (with SI units  $\Omega \text{ s}^{-1}$ , or  $\text{V C}^{-1}$ ) as:

$$W_{\parallel}(z) = W'_0(z), \quad (14.156)$$

and the *transverse wake function*  $W_{\perp}(z)$  (with SI units  $\Omega \text{ s}^{-1} \text{ m}^{-1}$ , or  $\text{V C}^{-1} \text{ m}^{-1}$ ) as:

$$W_{\perp}(z) = W_1(z). \quad (14.157)$$

## 14.4 Impedance

We have seen in Section 14.3 that the wake fields in a section of accelerator beam line can be represented in terms of wake functions. A wake function gives the integrated force seen by a particle as it passes through the specified section, at a specified distance behind a delta-function beam. The wake function is expressed in terms of the variable  $z = s - ct$  (where the delta-function beam is at position  $s$  along the reference trajectory at time  $t$ ); this means that the wake function provides a time-domain description of the wake fields. As we shall see in Chapter 15, for understanding the effects of wake fields on the stability of a beam in an accelerator, it is often useful to have a frequency-domain description of the wake fields. Transforming the wake function to the frequency domain produces a function (the impedance) that characterises the strength of the wake fields generated by a beam with a current modulation of a particular frequency. The wake fields will affect the evolution of the current modulation; by solving an equation of motion for the current modulation, it is possible to determine whether the modulation damps or grows with time. A modulation that grows with time is likely to lead to a beam instability.

In this final part of the current chapter, we shall first define the impedance of a section of an accelerator beam pipe in terms of the wake function, and then give some examples of impedances. To begin, consider a current  $I = I(t)$  at the entrance of a section of an accelerator with longitudinal wake function  $W_{\parallel}(z)$ . The current  $I$  (in amperes) should not be confused with the multipole moment  $I_m$  of a charge distribution. The  $m = 0$  moment  $I_0$  is the total charge in a given distribution, in coulombs; the total amount of charge passing the entrance of the specified section of

the accelerator between time  $t$  and time  $t + dt$  is  $I_0 = I(t) dt$ . Then, from (14.141), the potential across the specified section at time  $t'$  is:

$$V(t') = \frac{[F_s]_z(t')}{q} = - \int_0^\infty I(t' - t) W_{\parallel}(-ct) dt. \quad (14.158)$$

The potential  $V(t')$  is defined as the energy gain per unit charge for a particle as it passes through the specified section of the accelerator, if it enters the section at time  $t'$ . Defining a variable  $z = -ct$ , the potential can be written:

$$V(t') = - \int_{-\infty}^0 I(t' + z/c) W_{\parallel}(z) \frac{dz}{c}. \quad (14.159)$$

The wake function acts as a Green's function for calculating the wake potential in a given section of an accelerator beam pipe as a function of time. A positive voltage  $V$  means that a particle with charge  $q$  gains energy  $qV$  as it passes through the section of beam pipe. Let us write the current as an integral over a frequency spectrum:

$$I(t) = \int_{-\infty}^\infty e^{-i\omega t} \tilde{I}(\omega) \frac{d\omega}{2\pi}. \quad (14.160)$$

Then, the voltage becomes:

$$\begin{aligned} V(t) &= - \int_{-\infty}^0 \frac{dz}{c} \int_{-\infty}^\infty \frac{d\omega}{2\pi} e^{-i\omega t} \tilde{I}(\omega) e^{-i\omega z/c} W_{\parallel}(z) \\ &= - \int_{-\infty}^\infty e^{-i\omega t} \tilde{I}(\omega) Z_{\parallel}(\omega) \frac{d\omega}{2\pi}, \end{aligned} \quad (14.161)$$

where we define the *longitudinal impedance*  $Z_{\parallel}(\omega)$ :

$$Z_{\parallel}(\omega) = \int_{-\infty}^\infty e^{-i\omega z/c} W_{\parallel}(z) \frac{dz}{c}. \quad (14.162)$$

Note that we have made use of the fact that  $W_{\parallel}(z) = 0$  for  $z > 0$  to change the upper limit on the integral over  $z$  from 0 to infinity. Now we multiply both sides of (14.161) by  $e^{i\omega't}$ , and integrate over  $t$ :

$$\begin{aligned} \tilde{V}(\omega') &= \int_{-\infty}^\infty e^{i\omega't} V(t) dt \\ &= - \int_{-\infty}^\infty dt \int_{-\infty}^\infty \frac{d\omega}{2\pi} e^{i(\omega' - \omega)t} \tilde{I}(\omega) Z_{\parallel}(\omega) \\ &= - \int_{-\infty}^\infty \delta_D(\omega - \omega') \tilde{I}(\omega) Z_{\parallel}(\omega) d\omega. \end{aligned} \quad (14.163)$$

Thus, we have:

$$\tilde{V}(\omega) = -\tilde{I}(\omega)Z_{\parallel}(\omega). \quad (14.164)$$

In other words, multiplying the current spectrum by the impedance (defined as the Fourier transform of the wake function) gives the voltage spectrum. This is exactly as we would expect from AC circuit theory, except for the minus sign (which comes from the fact we have defined the voltage as the energy *gain* per unit charge in the given section of beam pipe).

Equation (14.164) makes clear the significance of the impedance. For the wake fields from a particular part of the beam line to have any real impact on the performance of an accelerator, there should be some overlap (in the frequency domain) between the impedance and the beam current spectrum. There may be a large impedance for some range of frequencies, but unless the current spectrum has a significant component in the same frequency range, the wake fields associated with the impedance will not have a significant effect on the beam.

In (14.162), we defined the longitudinal impedance as the Fourier transform of the longitudinal wake function  $W_{\parallel}(z) \equiv W'_0(z)$ . The definition may be generalised for longitudinal wake functions  $W'_m$  related to higher-order beam moments  $I_m$ :

$$Z_m^{\parallel}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega z/c} W'_m(z) \frac{dz}{c}. \quad (14.165)$$

The longitudinal effects of wake fields are often dominated by the monopole ( $m = 0$ ) moment of the beam; so for convenience, we define:

$$Z_{\parallel}(\omega) = Z_0^{\parallel}(\omega). \quad (14.166)$$

We can similarly define the transverse impedance as the Fourier transform of the wake function  $W_m(z)$ : the transverse impedance describes the transverse force on particles in a beam generated by a beam with a given multipole moment and modulated at a given frequency. Following convention, we include a factor of  $i \equiv \sqrt{-1}$  in the definition:

$$Z_m^{\perp}(\omega) = i \int_{-\infty}^{\infty} e^{-i\omega z/c} W_m(z) \frac{dz}{c}. \quad (14.167)$$

The transverse effects are often dominated by the dipole ( $m = 1$ ) mode; it is convenient to define:

$$Z_{\perp}(\omega) = Z_1^{\perp}(\omega). \quad (14.168)$$

The wake functions  $W'_m(z)$  and  $W_m(z)$  are of course given by the inverse Fourier transforms of the impedances:

$$W'_m(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega z/c} Z_m^{\parallel}(\omega) d\omega, \quad (14.169)$$

$$W_m(z) = -\frac{i}{2\pi} \int_{-\infty}^{\infty} e^{i\omega z/c} Z_m^{\perp}(\omega) d\omega. \quad (14.170)$$

The transverse and longitudinal wake functions are related by the Panofsky–Wenzel theorem (14.147). The corresponding relationship between the transverse and longitudinal impedances can be obtained by integrating (14.165) by parts, which gives:

$$Z_m^{\parallel}(\omega) = \frac{\omega}{c} Z_m^{\perp}(\omega). \quad (14.171)$$

In Section 14.3 we found the wake functions for resistive-wall wake fields and for the wake fields in a resonant cavity. The corresponding impedances can be found by performing an inverse Fourier transform. However, for the resistive-wall case, the impedance can be obtained more or less directly from the expressions for the wake fields that were discussed in Section 14.2. This is because we found the wake fields by working initially in the frequency domain, and then performing a Fourier transform to produce the wake fields in the time domain. Consider a particle of charge  $q$  travelling with speed  $c$  parallel to the axis of an accelerator beam pipe, following a charge distribution also travelling at speed  $c$ , with multipole moment  $I_m$ , and delta-function distribution in the longitudinal direction. If  $E_s$  is the longitudinal electric field seen by the charge  $q$ , and  $W_m(z)$  the corresponding wake function over a distance  $L$  of the beam pipe, then from (14.53), (14.68), (14.141) and (14.165), we can write:

$$\begin{aligned} E_s &= \int_{-\infty}^{\infty} e^{ikz} \cos(m\theta) \bar{C}_m^{(2)} \left(\frac{r}{b}\right)^m \frac{dk}{2\pi} \\ &= -\frac{I_m}{L} W'_m(z) r^m \cos(m\theta) \\ &= -\frac{I_m}{L} \int_{-\infty}^{\infty} e^{i\omega z/c} Z_m^{\parallel}(\omega) r^m \cos(m\theta) \frac{d\omega}{2\pi}, \end{aligned} \quad (14.172)$$

where  $\bar{C}_m^{(2)}$  is given in (14.93) and (14.95). Hence:

$$\int_{-\infty}^{\infty} e^{ikz} \frac{\bar{C}_m^{(2)}}{cb^m} dk = -\frac{I_m}{L} \int_{-\infty}^{\infty} e^{i\omega z/c} Z_m^{\parallel}(\omega) \frac{d\omega}{c}. \quad (14.173)$$

Therefore, if we put  $k = \omega/c$ , it follows that:

$$Z_m^{\parallel}(\omega) = -\frac{L \bar{C}_m^{(2)}}{c I_m b^m}. \quad (14.174)$$

Using the expressions for  $\bar{C}_m^{(2)}$  in (14.93) and (14.95) (with  $\lambda$  given by (14.90)), the impedance can be written:

$$Z_m^{\parallel}(\omega) = \left( \frac{Z_0 c}{4\pi} \right) \frac{4L}{b^{2m}} \times \\ \left( (1 + \delta_{m0})(1 + i \operatorname{sgn}(\omega))bc \sqrt{\left( \frac{Z_0 c}{4\pi} \right) \frac{2\pi\sigma}{|\omega|}} - \frac{ib^2\omega}{m+1} + \frac{ic^2 m}{\omega} \right)^{-1}. \quad (14.175)$$

Note that we drop a term  $2k/\lambda$  that appears in the denominator in the expressions for  $\bar{C}_m^{(2)}$  (14.93) and (14.95) (and makes negligible contribution to the impedance).

The expression for the longitudinal impedance in (14.175) is valid even at high frequency, corresponding to short distances  $z$ . When working in the time domain, we applied the condition  $|z| \gg b\xi^{2/3}$  (with  $\xi$  given by (14.96)) to obtain an approximate expression (14.143) for the resistive-wall wake function. Making a low frequency (large  $|z|$ ) approximation to the impedance provides some simpler expressions for the impedance. For frequencies such that  $|\omega| \ll \xi^{-2/3}c/b$ , (14.175) gives:

$$Z_m^{\parallel}(\omega) \approx \frac{(1 - i \operatorname{sgn}(\omega))}{(1 + \delta_{m0})} \frac{L}{b^{2m+1}c} \sqrt{\left( \frac{Z_0 c}{4\pi} \right) \frac{2|\omega|}{\pi\sigma}}. \quad (14.176)$$

The corresponding transverse impedance is readily obtained by applying (14.171). Assuming that the modes  $m = 0$  and  $m = 1$  dominate in the longitudinal and transverse directions respectively, the resistive-wall impedance in a beam pipe of length  $L$ , radius  $b$  and conductivity  $\sigma$  can be written as:

$$Z_{\parallel}(\omega) \equiv Z_0^{\parallel}(\omega) \approx (1 - i \operatorname{sgn}(\omega)) \frac{L}{2bc} \sqrt{\left( \frac{Z_0 c}{4\pi} \right) \frac{2|\omega|}{\pi\sigma}}, \quad (14.177)$$

$$Z_{\perp}(\omega) \equiv Z_1^{\perp}(\omega) \approx (1 - i \operatorname{sgn}(\omega)) \frac{c}{\omega b^3 c} \sqrt{\left( \frac{Z_0 c}{4\pi} \right) \frac{2|\omega|}{\pi\sigma}}. \quad (14.178)$$

These expressions apply for  $|\omega| \ll \xi^{-2/3}c/b$ , and if the skin depth is much smaller than the thickness of the wall of the beam pipe.

We observe that for resistive-wall wake fields, the transverse and longitudinal impedances are related by:

$$Z_{\perp} = \frac{2c}{\omega b^2} Z_{\parallel} = \frac{C_0}{\pi b^2} \frac{Z_{\parallel}}{n}, \quad (14.179)$$

where  $n = \omega C_0 / 2\pi c$  is the machine circumference divided by the wavelength of an electromagnetic wave with frequency  $\omega$ . We shall see in Chapter 15

that the quantity  $Z_{\parallel}/n$  provides a convenient (though rather crude) way to characterise the impedance in a storage ring for the purposes of estimating instability thresholds. The transverse impedance  $Z_{\perp}$  and longitudinal impedance  $Z_{\parallel}$  characterise the wake field response to different moments (i.e. different values of  $m$ ) of the charge distribution; there is no reason, in principle, to expect any particular relationship between the transverse impedance  $Z_{\perp}$  and longitudinal impedance  $Z_{\parallel}$ . However, the relationship (14.179) is sometimes generalised to impedances arising from sources other than resistive-wall.

Let us now consider the impedance of a resonant cavity. The  $m = 0$  wake function, based on a simple model of a cylindrical cavity, is given by (14.144), and can be written in the form (for  $z < 0$ ):

$$W_0(z) = \frac{cR_s}{Q} e^{\alpha z/c} \sin\left(\frac{\bar{\omega}_r z}{c}\right), \quad (14.180)$$

where  $\omega_r$  is the resonant frequency,  $Q$  is the quality factor,  $\alpha = \omega_r/2Q$ , and  $\bar{\omega}_r = \sqrt{|\omega_r^2 - \alpha^2|}$ .  $R_s$  is a constant with (SI) units of ohms, characterising the amplitude of the wake function. For  $z > 0$ , the wake function is zero. Taking the Fourier transform of  $W'_0(z)$  gives the longitudinal impedance:

$$Z_0^{\parallel}(\omega) \approx \frac{R_s}{1 + iQ\left(\frac{\omega_r}{\omega} - \frac{\omega}{\omega_r}\right)}. \quad (14.181)$$

The approximation is valid for  $2Q \gg 1$ . Although we have discussed only the  $m = 0$  mode for a resonant cavity, (14.181) is often generalised to other modes:

$$Z_m^{\parallel}(\omega) \approx \frac{R_{s,m}^{\parallel}}{1 + iQ\left(\frac{\omega_r}{\omega} - \frac{\omega}{\omega_r}\right)}. \quad (14.182)$$

The corresponding transverse impedance is:

$$Z_m^{\perp}(\omega) \approx \frac{c}{\omega} \frac{R_{s,m}^{\perp}}{1 + iQ\left(\frac{\omega_r}{\omega} - \frac{\omega}{\omega_r}\right)}. \quad (14.183)$$

The resonator impedances (14.181) and (14.183) are plotted as a function of frequency in Fig. 14.3.

The impedance (14.181) is the same as would be found in an electrical circuit, consisting of a resistance  $R_s$ , inductance  $L$  and capacitance  $C$  in parallel (an LCR circuit):

$$\frac{1}{Z} = \frac{1}{R_s} + \frac{i}{\omega L} - i\omega C = \frac{1 + iQ\left(\frac{\omega_r}{\omega} - \frac{\omega}{\omega_r}\right)}{R_s}. \quad (14.184)$$

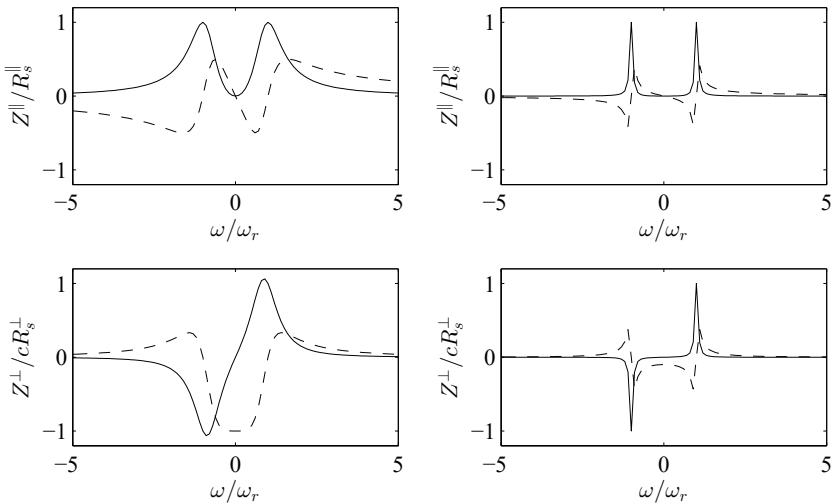


Fig. 14.3 Longitudinal (top) and transverse (bottom) impedances for a resonant cavity. The left-hand plots show the case  $Q = 1$ , and the right-hand plots show the case  $Q = 10$ . The solid lines show the real part of the impedance, and the dashed lines show the imaginary part.

The resonant frequency in an LCR circuit is:

$$\omega_r = \frac{1}{\sqrt{LC}}, \quad (14.185)$$

and the quality factor is:

$$Q = R_s \sqrt{\frac{C}{L}}. \quad (14.186)$$

In an LCR circuit, the impedance is dominated by the inductance for  $|\omega| \ll \omega_r$ , and by the capacitance for  $|\omega| \gg \omega_r$ . For  $\omega \approx \omega_r$ , the impedance is dominated by the resistance,  $Z \approx R_s$ . Impedances in accelerators are referred to as ‘inductive’ or ‘capacitive’ according to whether the sign of the imaginary part of the impedance (when  $\omega$  is positive) is negative or positive, respectively. For a resonant cavity, the impedance is inductive for  $\omega \ll \omega_r$ , and capacitive for  $\omega \gg \omega_r$ . The resistive-wall impedance (longitudinal (14.178) and transverse (14.178)) is partly resistive and partly inductive.

In Section 12.6 we defined the longitudinal space-charge impedance (12.221) for a storage ring:

$$Z_{\parallel sc}(\omega) = i \frac{gZ_0}{2\beta\gamma^2} \frac{\omega}{\omega_0}, \quad (14.187)$$

where  $g = 1 + 2 \ln(b/a)$ ,  $b$  is the beam pipe radius,  $a$  is the beam radius (assumed to have a KV distribution) and  $\omega_0$  is the revolution frequency. Strictly speaking, space-charge forces cannot be treated in exactly the same way as wake forces arising from fields generated by interaction between the beam and the vacuum chamber. However, the space-charge impedance (12.221) has the same physical significance as the impedance derived from a wake function: combined with the frequency spectrum of the beam current, it gives the frequency spectrum of the energy change of particles in the beam. This is convenient when discussing beam instabilities in terms of impedances. Note that the space-charge impedance is a purely (positive) imaginary quantity, so is referred to as a capacitive impedance, despite the fact that the magnitude of the impedance increases with frequency (which is a characteristic of an inductance in an electrical circuit).

In Section 14.3, we saw that some general properties of wake functions could be derived from physical constraints. In the same way, it is possible to apply physical (or mathematical) considerations to derive some general properties of impedances. Here, we give two examples. First, since the wake functions are real, it follows from (14.169) and (14.170) that:

$$Z_m^{\parallel}(\omega)^* = Z_m^{\parallel}(-\omega), \quad (14.188)$$

$$Z_m^{\perp}(\omega)^* = -Z_m^{\perp}(-\omega), \quad (14.189)$$

where  $Z_m^{\parallel}(\omega)^*$  is the complex conjugate of  $Z_m^{\parallel}(\omega)$ .

The second general property of impedances follows from the conservation of energy. Consider a bunch with longitudinal charge distribution  $\rho(z)$ , where  $z = s - ct$ . The change in energy of the bunch as it passes through a section of the beam pipe with longitudinal wake function  $W_{\parallel}(z) = W'_0(z)$  is:

$$\Delta\mathcal{E} = - \int_{-\infty}^{\infty} dz' \int_{z'}^{\infty} dz \rho(z') \rho(z) W_{\parallel}(z' - z). \quad (14.190)$$

The charge distribution in frequency space is:

$$\tilde{\rho}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega z/c} \rho(z) dz. \quad (14.191)$$

Since  $\rho(z)$  is real:

$$\tilde{\rho}(-\omega) = \tilde{\rho}(\omega)^*. \quad (14.192)$$

The charge distribution can be expressed as an integral over Fourier modes:

$$\rho(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega z/c} \tilde{\rho}(\omega) \frac{d\omega}{c}. \quad (14.193)$$

Substituting for  $\rho(z)$  from (14.193) and for  $W_{\parallel}(z)$  from (14.169), and using (14.192) and (14.188), the energy change (14.190) becomes:

$$\Delta\mathcal{E} = -\frac{1}{2\pi} \int_{-\infty}^{\infty} |\tilde{\rho}(\omega)|^2 \operatorname{Re} Z_{\parallel}(\omega) d\omega, \quad (14.194)$$

where  $Z_{\parallel}(\omega) = Z_0^{\parallel}(\omega)$ . The energy change depends on only the real part of the impedance. Furthermore, the energy change must be negative (i.e. the beam loses energy) for *any* charge distribution  $\rho(z)$ . It follows that:

$$\operatorname{Re} Z_{\parallel}(\omega) \geq 0 \quad \text{for all } \omega. \quad (14.195)$$

Wake fields in accelerators have a number of adverse effects. As we shall see in Chapter 15, wake fields can have significant impact on the beam dynamics; but the energy transferred via wake fields from the beam to components in the accelerator can also be important, especially in machines operating with high beam currents. In certain circumstances, the power deposition associated with wake fields can damage components in the vacuum chamber, such as the electrodes for beam position monitors. The transfer of energy from the beam to components in the vacuum chamber through wake fields is known as *parasitic loss*, or *higher-order mode (HOM) heating*. We finish the present chapter by defining the *loss factor*, which is often used to characterise the power load associated with a beam passing through a particular section of an accelerator.

Consider a bunch of particles passing through a given section of beam pipe. In terms of the potential  $V(t')$  (14.158) across this section of the beam pipe, the total energy gained by the particles in the bunch can be expressed as:

$$\Delta\mathcal{E} = \int_{-\infty}^{\infty} I(t')V(t') dt', \quad (14.196)$$

where  $I(t')$  is the beam current at the entrance of the section at time  $t'$ . Since the beam loses energy to wake fields in the accelerator, we expect  $\Delta\mathcal{E}$  to be negative. Substituting for  $V(t')$  from (14.158) the energy gain can be written in terms of the wake field:

$$\Delta\mathcal{E} = - \int_{-\infty}^{\infty} dt' \int_0^{\infty} dt I(t')I(t'-t)W_{\parallel}(-ct). \quad (14.197)$$

If the current in the bunch (as a function of time) is known, and if the wake function is also known, then the energy transferred from the bunch to the beam pipe can be calculated from (14.197). It is sometimes more convenient to express the energy transfer in terms of the impedance rather

than the wake function. This can be done as follows. Writing (14.197) in terms of the current spectrum by substituting for the current from (14.160) gives:

$$\begin{aligned} \Delta\mathcal{E} = & - \int_{-\infty}^{\infty} dt' \int_0^{\infty} dt \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} e^{-i\omega't'} \tilde{I}(\omega') \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega(t'-t)} \tilde{I}(\omega) W_{\parallel}(-ct). \end{aligned} \quad (14.198)$$

Performing the integral over  $t'$  gives:

$$\Delta\mathcal{E} = -2\pi \delta_D(\omega + \omega') \int_0^{\infty} dt \int_{-\infty}^{\infty} \frac{d\omega'}{2\pi} \tilde{I}(\omega') \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega t} \tilde{I}(\omega) W_{\parallel}(-ct). \quad (14.199)$$

Since the current as a function of time is a purely real quantity, the current spectrum must have the property:

$$\tilde{I}(-\omega) = \tilde{I}(\omega)^*. \quad (14.200)$$

Then, (14.199) becomes:

$$\Delta\mathcal{E} = - \int_{-\infty}^{\infty} \frac{dz}{c} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} |\tilde{I}(\omega)|^2 e^{-i\omega z/c} W_{\parallel}(z), \quad (14.201)$$

where we have also changed a variable of integration from  $t$  to  $z = -ct$ , and used the fact that  $W_{\parallel}(z) = 0$  for  $z > 0$  to change the upper limit of the integral. Finally, using the definition of the impedance (14.162), the energy gained by the bunch can be written:

$$\Delta\mathcal{E} = - \int_{-\infty}^{\infty} |\tilde{I}(\omega)|^2 Z_{\parallel}(\omega) \frac{d\omega}{2\pi}. \quad (14.202)$$

Again, we see the significance of the impedance: the change in the energy of the bunch depends on the size of the impedance in the frequency range of the current spectrum. Parasitic losses can be large if the bunch length is comparable to the size of cavities in the vacuum chamber.

The *loss factor*  $\kappa_{\parallel}$  (with SI units  $\text{V C}^{-1}$ ) is defined so that for a bunch with total charge  $q_{\text{bunch}}$ :

$$\Delta\mathcal{E} = -\kappa_{\parallel} q_{\text{bunch}}^2. \quad (14.203)$$

In terms of the impedance, the loss factor can be written:

$$\kappa_{\parallel} = \frac{1}{q_{\text{bunch}}^2} \int_{-\infty}^{\infty} |\tilde{I}(\omega)|^2 Z_{\parallel}(\omega) \frac{d\omega}{2\pi}. \quad (14.204)$$

The loss factor depends on the longitudinal distribution of the charge within a bunch. In a storage ring with total beam current  $I_{\text{beam}}$  (consisting of some

number of identical bunches), the power load  $P$  on a component with loss factor  $\kappa_{\parallel}$  will be:

$$P = \kappa_{\parallel} q_{\text{bunch}} I_{\text{beam}}. \quad (14.205)$$

In the case of short bunches in a resonant cavity with high quality factor, we can derive a simple expression for the loss factor in terms of the cavity parameters, as follows. Consider a Gaussian bunch, with current as a function of time given by:

$$I(t) = q_{\text{bunch}} c \frac{e^{-c^2 t^2 / 2\sigma_z^2}}{\sqrt{2\pi}\sigma_z}, \quad (14.206)$$

where  $\sigma_z$  is the rms bunch length. From (14.160), the current spectrum is:

$$\tilde{I}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t} I(t) dt = q_{\text{bunch}} e^{-\omega^2 \sigma_z^2 / 2c^2}. \quad (14.207)$$

Suppose that the bunch passes through a resonant cavity with longitudinal impedance (14.181):

$$Z_{\parallel}(\omega) = \frac{R_s}{1 + iQ \left( \frac{\omega_r}{\omega} - \frac{\omega}{\omega_r} \right)}. \quad (14.208)$$

Then, from (14.202) the energy gained by the bunch is:

$$\Delta\mathcal{E} = - \int_{-\infty}^{\infty} q_{\text{bunch}}^2 \frac{R_s e^{-\omega^2 \sigma_z^2 / c^2}}{1 + iQ \left( \frac{\omega_r}{\omega} - \frac{\omega}{\omega_r} \right)} \frac{d\omega}{2\pi}. \quad (14.209)$$

For  $Q \gg 1$  and  $\omega_r \sigma_z / c \ll 1$ , the integral can be approximated to give:

$$\Delta\mathcal{E} \approx -q_{\text{bunch}}^2 \frac{\omega_r}{2} \frac{R_s}{Q} e^{-\omega_r^2 \sigma_z^2 / c^2}. \quad (14.210)$$

Hence, from (14.203), the loss factor for short bunches in a high- $Q$  resonant cavity is:

$$\kappa_{\parallel} \approx \frac{\omega_r}{2} \frac{R_s}{Q} e^{-\omega_r^2 \sigma_z^2 / c^2}. \quad (14.211)$$

The *transverse loss factor*  $\kappa_{\perp}$  is defined in a similar way to the loss factor  $\kappa_{\parallel}$ :

$$\kappa_{\perp} = \frac{1}{q_{\text{bunch}}^2} \int_{-\infty}^{\infty} |\tilde{I}(\omega)|^2 Z_{\perp}(\omega) \frac{d\omega}{2\pi i}. \quad (14.212)$$

Using (14.189), this can be written:

$$\kappa_{\perp} = \frac{1}{q_{\text{bunch}}^2} \int_0^{\infty} |\tilde{I}(\omega)|^2 \text{Im}(Z_{\perp}(\omega)) \frac{d\omega}{\pi}. \quad (14.213)$$

The transverse loss factor is also called the *kick factor*, since it can be used to characterise the transverse kick experienced by a bunch from the wake fields in a given section of beam pipe. Assuming that all the particles in a bunch have a horizontal displacement  $x$  from the reference trajectory, the average change in the horizontal momentum (scaled by the reference momentum  $P_0$ ) can be written:

$$\langle \Delta p_x \rangle = -\frac{1}{cP_0} \int_{-\infty}^{\infty} dt' \int_0^{\infty} dt x I(t') I(t' - t) W_{\perp}(-ct). \quad (14.214)$$

Following the same procedure as for the loss factor  $\kappa_{\parallel}$ , this becomes, in terms of the transverse impedance:

$$\langle \Delta p_x \rangle = -\frac{x}{cP_0} \int_{-\infty}^{\infty} \left| \tilde{I}(\omega) \right|^2 Z_{\perp}(\omega) \frac{d\omega}{2\pi i}. \quad (14.215)$$

In terms of the kick factor:

$$\langle \Delta p_x \rangle = -\frac{x}{cP_0} q_{\text{bunch}}^2 \kappa_{\perp}. \quad (14.216)$$

It should be remembered, however, that not all particles will receive the same deflection: particles at the tail of the bunch will experience much stronger wake fields than particles at the head of the bunch, and will therefore experience a much larger deflection.

## Chapter 15

# Coherent Instabilities

After having discussed wake fields in Chapter 14, and having seen how the resulting forces on particles in an accelerator can be described in terms of wake functions and impedances, we now consider the impact of wake fields on the beam dynamics in an accelerator. The effects of wake fields depend on the operating regime and the characteristics of the wake fields that are present. The range of phenomena associated with wake fields is quite large, and even classifying the various effects is not straightforward; but it is convenient, at least for an initial analysis, to identify two regimes: coupled-bunch instabilities and single-bunch instabilities. Although the distinction between these regimes is not always completely clear, in the case of coupled-bunch instabilities each bunch may be considered to perform coherent betatron and (in a storage ring) synchrotron motion. In other words, each bunch can be treated as a single ‘macroparticle’ with charge and mass equal to the total charge and mass of all the particles within the bunch. In general, macroparticles can have attributes other than mass and charge, such as transverse or longitudinal dimensions. However, if the spacing between bunches is large compared with the size of a bunch, then it may be possible to model each bunch (or macroparticle) as a point-like particle. For coupled-bunch instabilities, the dynamics in the presence of wake fields depends on the long-range behaviour of the wake functions (i.e. the low frequency region of the impedance).

In the case of single-bunch instabilities, on the other hand, we are concerned with the short-range wake fields (high frequency impedance), and how these affect the distribution of charge within a bunch. Usually, in this regime, each bunch within the accelerator can be considered in isolation, neglecting the effect of one bunch on another. We shall discuss single-bunch instabilities starting in Section 15.2; but we begin in Section 15.1 by

discussing coupled-bunch instabilities. In particular, we shall develop a model for the motion of bunches in a storage ring in the presence of wake fields. We shall find that, under rather general conditions, the wake fields can lead to an increase in the amplitudes of betatron and synchrotron oscillations of bunches within the accelerator. Without some appropriate mitigation, the oscillation amplitudes can grow indefinitely, leading to a loss of beam quality and, ultimately, a loss of particles from the beam. Our goal is to derive an expression for the growth rates of the oscillations driven by the wake fields, in terms of the machine parameters.

## 15.1 Coupled-Bunch Instabilities

### 15.1.1 *Transverse modes*

Consider a storage ring containing  $n_b$  equally spaced bunches. We assume that we can treat each bunch as a single point-like particle (a macroparticle) with total charge  $qN_b$ , where  $N_b$  is the number of (real) particles in each bunch. In this section, we consider only the betatron motion of the bunches; longitudinal motion will be discussed in Section 15.1.2. In the absence of wake fields, each bunch can perform coherent betatron motion (i.e. the centroid of each bunch can oscillate) with frequency  $\omega_\beta$ . The betatron frequency is determined by the arrangement of quadrupoles in the lattice, and depends on the position of the bunch around the ring. If we assume that we can use a ‘smooth-focusing’ approximation, so that the betatron frequency is constant, then in the presence of wake fields the equation of motion for the  $n$ th bunch can be written:

$$\frac{d^2x_n}{dt^2} + \omega_\beta^2 x_n = \frac{F_x}{\gamma_0 m N_b} = \frac{F_x c^2}{E_0 N_b}, \quad (15.1)$$

where  $x$  is the horizontal co-ordinate of the bunch centroid,  $F_x$  is the force from the wake field,  $\gamma_0$  is the relativistic factor for particles in the storage ring, and  $m$  is the mass of each particle.  $E_0 = \gamma_0 mc^2$  is the reference energy. For simplicity, we neglect any coupling between the horizontal and the vertical or longitudinal motion.

In a further approximation, we assume that the force from the wake field can be found by dividing the wake potential for the entire storage ring by the ring circumference. This effectively averages the force over the entire circumference. The wake field is the sum of the fields generated by all bunches in the beam, over all previous turns of the storage ring. Therefore,

we write:

$$F_x = -\frac{(qN_b)^2}{C_0} \sum_{k=0}^{\infty} \sum_{n'=0}^{n_b-1} W_1(z) x_{n'} \left( t + \frac{z}{c} \right), \quad (15.2)$$

where  $C_0$  is the circumference,  $n'$  is an index over all bunches in the storage ring, and  $k$  is an index over all the (previous) turns. The magnitude of  $z$  is the distance between bunches  $n$  and  $n'$ , plus  $k$  times the circumference:

$$z = -\frac{(n' - n)}{n_b} C_0 - kC_0. \quad (15.3)$$

Note that the wake field from a given bunch depends on the co-ordinate of the bunch at the time that it generated the wake field, and also that each bunch sees the wake fields that it generated itself on previous turns.

Substituting for  $F_x$  from (15.2) into (15.1) gives the equation of motion:

$$\frac{d^2 x_n}{dt^2} + \omega_\beta^2 x_n = -\frac{q^2 N_b c^2}{E_0 C_0} \sum_{k=0}^{\infty} \sum_{n'=0}^{n_b-1} W_1(z) x_{n'} \left( t + \frac{z}{c} \right). \quad (15.4)$$

We try a solution to (15.4) of the form:

$$x_n^\mu(t) = A e^{i(2\pi\mu n/n_b - \Omega_\mu t)}, \quad (15.5)$$

where  $A$  and  $\mu$  are constants, and  $\Omega_\mu$  is a frequency to be determined. The real part of  $\Omega_\mu$  will give the betatron frequency of each bunch in the presence of wake fields: in general, the wake fields will lead to some shift in the frequency so that it is different from  $\omega_\beta$ . The imaginary part will determine whether the amplitude of the mode grows or damps, depending on whether it is positive or negative, respectively. The magnitude of the imaginary part of  $\Omega_\mu$  gives the exponential growth or damping rate.

The constant  $\mu$  describes the horizontal positions of the bunches in the accelerator at a particular time  $t$ . If  $\mu = 0$ , then at any given time all bunches have the same  $x$  co-ordinate. If  $\mu = 1$ , then at a given time the bunches are arranged to form a cosine wave, with period equal to the ring circumference. With  $\mu = 2$ , the period is half the ring circumference, and so on. The different patterns are illustrated in Fig. 15.1. In general, any arrangement of bunches (i.e. any set of  $x$  co-ordinates) can be constructed by a summation over solutions of the form (15.5), with different values of  $\mu$  and  $A$ . We refer to  $\mu$  as a ‘mode index’ for the bunch motion.

Substituting the trial solution (15.5) into the equation of motion (15.4) gives:

$$\omega_\beta^2 - \Omega_\mu^2 = -\frac{q^2 N_b c^2}{E_0 C_0} \sum_{k=-\infty}^{\infty} \sum_{n'=0}^{n_b-1} W_1(z) e^{2\pi i \mu(n'-n)/n_b} e^{-i\Omega_\mu z/c}. \quad (15.6)$$

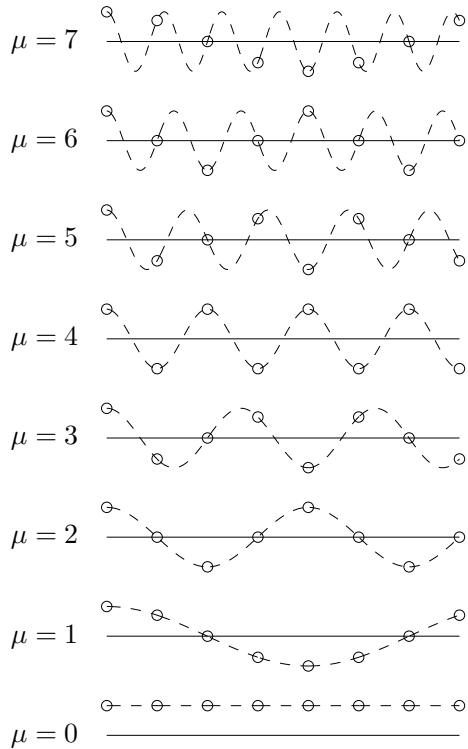


Fig. 15.1 Positions of bunches in a storage ring with different mode numbers for coupled-bunch oscillations. In this case, there are  $n_b = 8$  bunches (represented by circles in the figure) and hence the mode number  $\mu$  is in the range  $0 \leq \mu < 8$ . The transverse position of the  $n$ th bunch is given by  $x_n = A \cos(2\pi\mu n/n_b + \varphi)$  where  $A$  is a constant amplitude,  $\mu$  is the mode number, and  $\varphi$  is a phase (which depends on the time). The solid horizontal line in each case indicates  $x_n = 0$ ; the dashed lines show the values that  $x_n$  would take for non-integer values of  $n$ .

Note that we have used the property of the wake function that  $W_1(z) = 0$  for  $z > 0$  to extend the range on the sum over  $k$ . Now we substitute the impedance  $Z_1^\perp(\omega)$  for the wake function  $W_1(z)$  using (14.170), to give:

$$\omega_\beta^2 - \Omega_\mu^2 = i \frac{q^2 N_b c^2}{2\pi E_0 C_0} \sum_{k=-\infty}^{\infty} \sum_{n'=0}^{n_b-1} \int_{-\infty}^{\infty} Z_1^\perp(\omega) e^{i\omega z/c} e^{2\pi i \mu(n'-n)/n_b} e^{-i\Omega_\mu z/c} d\omega. \quad (15.7)$$

Substituting for  $z$  from (15.3) and collecting various exponential factors

together gives:

$$\omega_\beta^2 - \Omega_\mu^2 = i \frac{q^2 N_b c^2}{2\pi E_0 C_0} \sum_{k=-\infty}^{\infty} \sum_{n'=0}^{n_b-1} \int_{-\infty}^{\infty} Z_1^\perp(\omega) e^{2\pi i k (\Omega_\mu - \omega)/\omega_0} e^{2\pi i (n' - n)p} d\omega, \quad (15.8)$$

where:

$$\omega_0 = 2\pi \frac{c}{C_0} \quad (15.9)$$

is the revolution (angular) frequency, and  $p$  is defined so that:

$$n_b p = \frac{\Omega_\mu - \omega}{\omega_0} + \mu. \quad (15.10)$$

Using the fact that:

$$\sum_{k=-\infty}^{\infty} e^{2\pi i k u / u_0} = u_0 \sum_{p'=-\infty}^{\infty} \delta_D(u - p' u_0), \quad (15.11)$$

equation (15.8) becomes:

$$\omega_\beta^2 - \Omega_\mu^2 = i \frac{q^2 N_b \omega_0^2 c}{4\pi^2 E_0} \sum_{p'=-\infty}^{\infty} \sum_{n'=0}^{n_b-1} Z_1^\perp(\omega_p) e^{2\pi i (n' - n)p}, \quad (15.12)$$

where  $\omega_p$  is defined so that:

$$\frac{\Omega_\mu - \omega_p}{\omega_0} = p', \quad (15.13)$$

and now:

$$n_b p = \frac{\Omega_\mu - \omega_p}{\omega_0} + \mu. \quad (15.14)$$

For large  $n_b$ , the summation over  $n'$  vanishes unless  $p$  is an integer, in which case it simply gives a factor of  $n_b$ . This means we can replace the sum over  $p'$  with a sum over  $p$  (with  $\omega_p$  given by (15.14)), and replace the sum over  $n'$  by a factor  $n_b$ :

$$\omega_\beta^2 - \Omega_\mu^2 = i \frac{q^2 n_b N_b \omega_0^2 c}{4\pi^2 E_0} \sum_{p=-\infty}^{\infty} Z_1^\perp(\omega_p). \quad (15.15)$$

Finally, we assume that the betatron frequency shift is small. This enables us to make two simplifications. First, we can write:

$$\omega_\beta^2 - \Omega_\mu^2 = (\omega_\beta - \Omega_\mu)(\omega_\beta + \Omega_\mu) \approx 2\omega_\beta(\omega_\beta - \Omega_\mu). \quad (15.16)$$

Second, we can replace  $\Omega_\mu$  where it appears in the argument of the impedance by  $\omega_\beta$ . Then, (15.15) gives:

$$\Omega_\mu - \omega_\beta = -i \frac{q^2 n_b N_b \omega_0 c}{8\pi^2 E_0 \nu_x} \sum_{p=-\infty}^{\infty} Z_1^\perp(\omega_\beta + (\mu - n_b p) \omega_0), \quad (15.17)$$

where  $\nu_x = \omega_\beta / \omega_0$  is the betatron tune.

Equation (15.17) is the result we were seeking: it gives the betatron frequency  $\Omega_\mu$  of a given mode in terms of the beam and storage ring lattice parameters, and the total impedance. The real part of  $\Omega_\mu$  gives the betatron oscillation frequency in the presence of wake fields, and the imaginary part of  $\Omega_\mu$  gives the growth or damping rate of the mode amplitude.

It is sometimes convenient to write (15.17) in terms of the classical radius of the particles in the beam  $r_0$ , given by:

$$r_0 = \frac{Z_0 c}{4\pi} \frac{q^2}{mc^2}. \quad (15.18)$$

In terms of  $r_0$ , (15.17) is written:

$$\Omega_\mu - \omega_\beta = -i \frac{4\pi}{Z_0 c} \frac{n_b N_b r_0 \omega_0 c}{8\pi^2 \gamma_0 \nu_x} \sum_{p=-\infty}^{\infty} Z_1^\perp(\omega_\beta + (\mu - n_b p) \omega_0). \quad (15.19)$$

Note that the impedance is evaluated at frequencies  $\omega_p$ :

$$\omega_p = \omega_\beta + (\mu - n_b p) \omega_0. \quad (15.20)$$

To understand the significance of these frequencies, consider the first-order moment of the beam current observed (as a function of time) at a particular location in the storage ring:

$$xI(t) = (xI)_0 \sum_{n=-\infty}^{\infty} e^{i(\omega_\beta + \mu \omega_0)t} \delta_D(t - nT_0/n_b), \quad (15.21)$$

where  $(xI)_0$  is the amplitude of the signal, and  $T_0$  is the revolution period. Each bunch is performing betatron oscillations of frequency  $\omega_\beta$  as it moves around the ring: we neglect the frequency shift from wake field forces. If the bunches in the beam are arranged in a mode  $\mu$ , the frequency observed at a fixed point in the ring is shifted from the betatron frequency by  $\mu \omega_0$ . With  $n_b$  equally spaced bunches in the ring, a current ‘spike’ is observed at regular intervals of  $T_0/n_b$ , as each bunch passes the observation point. Using (15.11), and the fact that  $\omega_0 = 2\pi/T_0$ , the beam signal can be written:

$$xI(t) = (xI)_0 \frac{n_b}{T_0} \sum_{p=-\infty}^{\infty} e^{i\omega_p t}, \quad (15.22)$$

where  $\omega_p$  is given by (15.20). In other words, the frequencies  $\omega_p$  are the frequencies of the components present in the spectrum of the first-order moment of the beam current, observed at a fixed point in the storage ring. These are exactly the frequencies at which a wake field will be excited, and hence the frequencies at which the impedance should be evaluated when calculating the effects of the wake field on the beam.

Let us consider the particular case of a (transverse) coupled-bunch instability driven by the resistive-wall wake fields. The impedance is given by (14.178), with  $L$  equal to the total circumference of the storage ring, i.e.  $L = C_0$ . The growth or damping rates of the different modes are plotted in Fig. 15.2. Since the impedance is largest at low frequencies, the beam mode with the highest growth rate will be the mode with the most *negative* value of  $\omega_p$ : a negative value of  $\omega_p$  will mean that the real part of  $Z_1^\perp(\omega_p)$  will be negative, which will mean that the imaginary part of  $\Omega_\mu$  will be positive. From (15.20), the most negative value of  $\omega_p$  occurs for values of  $\mu$  and  $p$  such that:

$$\mu - n_b p = -\text{int}(\nu_x) - 1, \quad (15.23)$$

where  $\text{int}(\nu_x)$  is the integer part of the betatron (in this case, horizontal) tune. Since  $0 \leq \mu < n_b$ , assuming that  $\nu_x < n_b$  we must have  $p = 1$ , so that:

$$\mu = n_b - \text{int}(\nu_x) - 1, \quad (15.24)$$

and:

$$\frac{\omega_{p,\min}}{\omega_0} = \text{frac}(\nu_x) - 1, \quad (15.25)$$

where  $\text{frac}(\nu_x)$  is the fractional part of the betatron tune. Using  $\omega_{p,\min}$  for  $\omega$  in the impedance (14.178), we find from (15.19) that the growth rate for the *fastest-growing* mode is:

$$\frac{1}{\tau} = \text{Im}(\Omega_\mu) = \frac{n_b N_b r_0 c}{4\pi^2 \gamma_0 \nu_x b^3} \sqrt{\left(\frac{4\pi}{Z_0 c}\right)} \frac{c C_0}{\sigma} \frac{1}{\sqrt{1 - \text{frac}(\nu_x)}}. \quad (15.26)$$

We notice that the growth rate for the transverse resistive-wall coupled-bunch instability has a strong dependence on the beam pipe radius, but only a weak dependence on the conductivity of the material from which the beam pipe is made. We also notice that because of the dependence on the tune, it is beneficial to make the fractional part of the tune as small as possible, while making the integer part of the tune as large as possible.

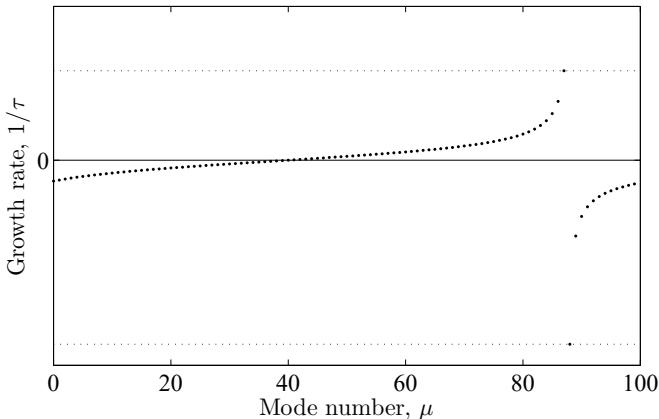


Fig. 15.2 Growth and damping rates for transverse coupled-bunch modes driven by resistive-wall wake fields, calculated from (15.19) with the impedance (14.178). The storage ring contains 100 equally spaced bunches, so the mode numbers are in the range  $0 \leq \mu < 100$ . The betatron tune is 12.2. The integer part of the tune determines which mode has the highest growth rate. The maximum growth and damping rates are indicated by the dotted lines. Since the fractional part of the tune is less than 0.5, the maximum growth rate, given by (15.26), is less than the maximum damping rate.

It is sometimes convenient to express the growth rate (15.26) in terms of the average current in the storage ring:

$$\langle I \rangle = \frac{n_b N_b q c}{C_0}. \quad (15.27)$$

Using the expression for the characteristic current  $I_c$ :

$$I_c = \frac{qc}{r_0}, \quad (15.28)$$

the growth rate (15.26) can be written:

$$\frac{1}{\tau} = \frac{\langle I \rangle}{I_c} \frac{(cC_0)^{3/2}}{4\pi^2 \gamma_0 \nu_x b^3} \sqrt{\left(\frac{4\pi}{Z_0 c}\right)} \frac{1}{\sigma} \frac{1}{\sqrt{1 - \text{frac}(\nu_x)}}. \quad (15.29)$$

For electrons,  $I_c$  is the Alfvén current  $I_A \approx 17.045$  kA. By writing the growth rate in the form (15.29), we see that the growth rate is proportional to the average current in the ring. For a given type of lattice, the tune is roughly proportional to the circumference; therefore, we expect the growth rate to scale as the square root of the circumference of the storage ring. Finally, we note that the growth rate is inversely proportional to the beam energy.

The evolution of the amplitude of a coupled-bunch mode in a storage ring depends on the mode, the wake fields in the ring, and the machine and beam parameters. In the case of the resistive-wall wake field, the coupled-bunch modes have a characteristic behaviour where mode numbers close to  $n_b - \nu_x$  (where  $n_b$  is the number of bunches and  $\nu_x$  is the betatron tune) are strongly damped or antidamped, depending on whether the mode number is above or below  $n_b - \nu_x$ : see Fig. 15.2. In general, in a storage ring with a large number of bunches, there will always be some modes that are antidamped. There is no threshold for the instability: the growth rate increases with increasing beam current, but there will always be a non-zero growth rate, even at very low current. In the storage rings typical of third-generation synchrotron light sources, the growth times from resistive-wall wake fields may be of the order of some hundreds of turns of the ring, and could be much shorter in some cases. This may at first seem surprising, since storage rings clearly do operate without the stored bunches performing large coherent betatron oscillations, which would make the beam unusable.

Fortunately, there are mechanisms that naturally damp the oscillations. In electron storage rings in particular, synchrotron radiation effects will naturally tend to damp the motion of any particles starting to make large betatron (or synchrotron) oscillations, and will therefore also damp the coherent motion of bunches of particles. Also, because there are variations in oscillation frequency of individual particles depending on the oscillation amplitude and energy deviation, coherent oscillations of a bunch will tend to decohere, resulting in an increase in the size of the bunch, but with the bunch centroid following the closed orbit. Synchrotron radiation and decoherence may act quickly enough to suppress coupled-bunch instabilities at low beam current, but the growth rates from wake field effects will increase with beam current, while the damping rates from synchrotron radiation and decoherence will remain fixed. As a result, coupled-bunch instabilities can impose a limit on the beam current that can be injected into a storage ring. In an electron storage ring, a simple estimate of the current limit might be made by calculating the current at which the growth rate of the fastest-growing mode just exceeds the damping rate from synchrotron radiation.

If a beam current is required that is higher than the limit from coupled-bunch instabilities, then it will be necessary to provide some additional damping of the coherent oscillations of individual bunches within the beam. This can be done using a *fast feedback system*, consisting of a pick-up that detects the position of a bunch at a particular location in the storage ring,

an amplifier, and a set of electrodes that provides a corrective ‘kick’ based on the amplified signal from the pick-up. Because the signal in the feedback system cannot be transmitted faster than the speed of light, the kick is applied one or more turns after the signal from the pick-up is detected. Three important parameters of a fast feedback system are the bandwidth, the gain and the power. The required bandwidth is related to the spacing between bunches in the storage ring: the shorter the spacing between bunches, the higher the frequencies of the coupled-bunched modes that can occur in the beam, and the higher the frequency at which the feedback system has to operate. Modern feedback systems can achieve bandwidths sufficient to suppress single-bunch instabilities occurring in proton machines with bunch lengths of order 1 ns [Fox *et al.* (2013)]. The gain determines the maximum damping rate that can be achieved using the feedback system. A system with very high gain can be sensitive to noise from the pick-up, and can excite coherent motion in the beam, rather than damping it. With modern systems, very rapid damping rates can be achieved: for example, the longitudinal feedback system for PEP-II operated with damping rates of more than  $3\text{ ms}^{-1}$  (corresponding to damping times of order 45 turns) [Fox *et al.* (2008)]. The maximum power of the feedback system determines the largest amplitude of beam motion for which the maximum damping rate can be achieved. If the beam is properly damped, so that the amplitude of coherent motion is small, then the feedback system will require very little power. However, if some motion is excited (for example, during the injection of particles into the storage ring), then the power needed by the feedback system will increase. Fast feedback systems, particularly those based on digital technology, can provide useful information on the beam behaviour as well as providing damping for coupled-bunch instabilities [Fox *et al.* (1999)]. For further information on fast feedback systems, see for example [Lonza (2009)].

### **15.1.2 *Longitudinal modes***

Let us now consider the synchrotron motion of bunches in a storage ring in the presence of wake fields. We shall again assume that the ring is filled with  $n_b$  equally spaced bunches, and that each bunch can be treated as a point-like particle with charge  $qN_b$ . If the wake function for the entire storage ring is  $W_0(z)$ , then from (14.141) the change in the energy deviation

of a single particle in bunch  $n$  over one turn is given by:

$$\Delta\delta_n = -\frac{q}{E_0} \sum_{k=0}^{\infty} \sum_{n'=0}^{n_b-1} qN_b W'_0(Z), \quad (15.30)$$

where:

$$Z = -\frac{(n' - n)}{n_b} C_0 - kC_0 - z_n(t') + z_n(t), \quad (15.31)$$

and:

$$t' = t - \frac{(n' - n)}{n_b} C_0 - kC_0. \quad (15.32)$$

The sum over  $n'$  includes the contribution to the wake field from all bunches in the ring; and the sum over  $k$  includes the contribution from each bunch over all turns of the ring. The last two terms in (15.31) account for the synchrotron motion of the bunches:  $z_n(t)$  is the longitudinal position of bunch  $n$  at time  $t$ , relative to a synchronous (reference) particle within the bunch. A positive value for  $z_n(t)$  means that at time  $t$ , bunch  $n$  is ahead of the reference particle. The rate of change of the longitudinal position of a bunch is related to the energy deviation by the phase slip factor  $\eta_p$ :

$$\frac{dz_n}{dt} = -\eta_p c \delta_n. \quad (15.33)$$

In addition to energy changes from the wake fields, there will also be energy changes resulting from the rf system, and (in electron machines) from synchrotron radiation. Suppose that, in the absence of wake fields, bunches perform synchrotron oscillations (in the focusing potential provided by the rf system) with frequency  $\omega_z$ . Then, including wake fields (but neglecting synchrotron radiation effects, which are assumed to be relatively slow), the equation of motion for  $z_n$  can be written:

$$\frac{d^2 z_n}{dt^2} + \omega_z^2 z_n = \frac{q^2 N_b \eta_p c^2}{E_0 C_0} \sum_{k=0}^{\infty} \sum_{n'=0}^{n_b} W'_0(Z). \quad (15.34)$$

To solve the equation of motion (15.34), we follow the same procedure as we did in Section 15.1.1. We try a solution of the form:

$$z_n^\mu(t) = A e^{i(2\pi\mu n/n_b - \Omega_\mu t)}. \quad (15.35)$$

The goal is to find an expression for the frequency  $\Omega_\mu$ : the real part of  $\Omega_\mu$  will give the (synchrotron) oscillation frequency in the presence of wake fields for a given mode  $\mu$ , and the imaginary part of  $\Omega_\mu$  will give the exponential growth or damping rate for the given mode. The next step is to

substitute for  $z_n$  in (15.34) and substitute also for  $W'_0(Z)$  from (14.169). For small  $z_n$ , we can use the approximation:

$$e^{i\omega Z} \approx -i\frac{\omega}{c} (z_{n'}(t') - z_n(t)) e^{-i\omega((n'-n)/n_b+k)C_0/c}. \quad (15.36)$$

After substituting for  $z_n$  and  $W'_0(Z)$ , and using (15.36), the equation of motion becomes:

$$\Omega_\mu^2 - \omega_z^2 = i \frac{q^2 n_b N_b \omega_0^2 \eta_p}{4\pi^2 E_0} \left( \sum_{p=-\infty}^{\infty} \omega_p Z_0^{\parallel}(\omega_p) - \sum_{p_0=-\infty}^{\infty} \omega_{p0} Z_0^{\parallel}(\omega_{p0}) \right), \quad (15.37)$$

where:

$$\omega_p = \Omega_\mu + (\mu + n_b p) \omega_0, \quad (15.38)$$

$$\omega_{p0} = n_b p_0 \omega_0, \quad (15.39)$$

and  $\omega_0 = 2\pi c/C_0$  is the revolution frequency.

The sum over  $p_0$  in (15.37) is independent of the mode number  $\mu$  and the synchrotron frequency. Also, from the property (14.188) of the impedance, the sum over  $p_0$  gives a value that is purely imaginary; hence, we can write:

$$\Omega_\mu^2 - \omega_z^2 = W_s^2 + i \frac{q^2 n_b N_b \omega_0^2 \eta_p}{4\pi^2 E_0} \sum_{p=-\infty}^{\infty} \omega_p Z_0^{\parallel}(\omega_p), \quad (15.40)$$

where:

$$W_s^2 = \frac{q^2 n_b N_b \omega_0^2 \eta_p}{4\pi^2 E_0} \operatorname{Im} \sum_{p_0=-\infty}^{\infty} \omega_{p0} Z_0^{\parallel}(\omega_{p0}). \quad (15.41)$$

Note that  $W_s^2$  is a real number: it contributes to a change in the synchrotron frequency, but does not lead to any growth or damping of the synchrotron oscillations. Since  $W_s^2$  is also independent of the mode of the beam motion (i.e. it is independent of  $\mu$ ), it can be interpreted physically as a modification of the focusing potential seen by the beam.

To proceed, we need an explicit expression for the impedance  $Z_0^{\parallel}(\omega)$ . There will be some contribution to the longitudinal impedance from the resistive-wall wake fields; however, in storage rings a major contribution to the longitudinal impedance usually comes from the rf cavities. The impedance can often be assumed to be that of a resonator (14.181) with some shunt impedance  $R_s$  and quality factor  $Q \gg 1$ :

$$Z_0^{\parallel}(\omega) = \frac{R_s}{1 + iQ \left( \frac{\omega_r}{\omega} - \frac{\omega}{\omega_r} \right)}. \quad (15.42)$$

The rf cavities in a storage ring are generally operated at a frequency  $h\omega_0$  close to, but not exactly equal to, their resonant frequency  $\omega_r$ . Hence, we write:

$$\omega_r \approx h\omega_0 + \Delta\omega, \quad (15.43)$$

where  $h$  is the harmonic number (the rf frequency divided by the beam revolution frequency). Since the impedance in a resonator peaks at frequencies close to  $\pm\omega_r$ , the effects of the cavity wake fields are largest for mode  $\mu = 0$  and  $n_b p = h$ , in which case we need only include terms in the sums over  $p$  and  $p_0$  in (15.37), for which:

$$\omega_p = \omega_z \pm h\omega_0, \quad (15.44)$$

$$\omega_{p0} = \pm h\omega_0. \quad (15.45)$$

Then, substituting the resonator impedance (14.181) into (15.37), keeping only the dominant terms in the sums over  $p$  and  $p_0$ , and expanding to first order in  $\omega_z$  and  $\Delta\omega$  gives:

$$\Omega_\mu^2 - \omega_z^2 = i \frac{q^2 n_b N_b \omega_0^2 \eta_p R_s}{4\pi^2 E_0} \left( 2\omega_z + \frac{16Q^2 \Delta\omega \omega_z}{h\omega_0} + \frac{2iQ\omega_z^2}{h\omega_0} \right). \quad (15.46)$$

Assuming that the synchrotron tune shift is small, we can make the approximation:

$$\Omega_\mu^2 - \omega_z^2 = (\Omega_\mu + \omega_z)(\Omega_\mu - \omega_z) \approx 2\omega_z(\Omega_\mu - \omega_z). \quad (15.47)$$

With this approximation, we find from (15.46):

$$\Omega_\mu \approx \omega_z + \Delta\omega_z + \frac{i}{\tau}, \quad (15.48)$$

where the synchrotron tune shift is:

$$\Delta\omega_z = - \frac{q^2 n_b N_b \omega_0^2 \eta_p Q R_s}{4\pi^2 E_0 h}, \quad (15.49)$$

and the synchrotron oscillations have an exponential growth rate:

$$\frac{1}{\tau} = \frac{q^2 n_b N_b \omega_0^2 \eta_p R_s}{4\pi^2 E_0} \left( 1 + \frac{8Q^2 \Delta\omega}{h\omega_0} \right). \quad (15.50)$$

It is convenient to express the synchrotron tune shift and the growth rate in terms of the classical radius  $r_0$  of the particles in the accelerator (15.18). The tune shift becomes:

$$\Delta\omega_z = - \frac{4\pi}{Z_0 c} \frac{r_0 n_b N_b \eta_p Q R_s}{\gamma_0 T_0^2 h}, \quad (15.51)$$

where  $\gamma_0$  is the relativistic factor for a particle with the reference energy, and  $T_0$  is the revolution period. For large  $Q$ , the second term in the brackets in (15.50) dominates over the first term, and the exponential growth rate becomes:

$$\frac{1}{\tau} = \frac{4\pi}{Z_0 c} \frac{4r_0 n_b N_b \eta_p Q^2 R_s \Delta\omega}{\pi \gamma_0 T_0 h}. \quad (15.52)$$

From (15.52) we see that the coherent synchrotron oscillations are damped if  $\eta_p \Delta\omega < 0$  and are antidamped if  $\eta_p \Delta\omega > 0$ . Antidamping corresponds to a beam instability: this particular type of instability, a growth in the amplitude of synchrotron oscillations driven by the longitudinal impedance of the rf cavities, is known as the *Robinson instability* [Robinson (1964)]. To avoid the Robinson instability, a ring operating below transition ( $\eta_p < 0$ ) must have the rf cavities tuned so that the resonant frequency is slightly above the frequency at which the cavities are operated ( $\Delta\omega > 0$ ); above transition, the resonant frequency of the cavities should be slightly below the operating frequency. To understand the physical mechanism of the Robinson instability, recall that the longitudinal impedance of a structure indicates the energy *lost* by the beam across the structure, when the beam current contains a component at a given frequency. Consider a ring operating below transition, so that the beam revolution frequency increases with increasing energy. For simplicity, suppose that the ring contains a single bunch, and that the harmonic number (the ratio of the rf frequency to the ring revolution frequency) is  $h = 1$ . As a bunch performs (coherent) synchrotron oscillations, the rf cavities see a higher frequency beam signal for positive energy deviations ( $\delta > 0$ ), and a lower frequency beam signal for negative energy deviations ( $\delta < 0$ ). If the cavity impedance also increases with frequency, then at higher energy the beam will lose more energy to the cavities through the impedance than it does at lower energy: see Fig. 15.3. The net effect will be to damp the synchrotron oscillations. Given the nature of the impedance of a resonant structure, the impedance increases with frequency  $\omega$  if  $\omega < \omega_r$  (i.e. the frequency is below the resonant frequency), and decreases with frequency if  $\omega > \omega_r$ . Hence, below transition, to avoid the Robinson instability the rf cavities should operate at a frequency below their resonant frequency. Above transition, the frequency of the beam signal decreases with increasing beam energy, so to avoid instability the rf cavities should operate at a frequency above their resonant frequency.

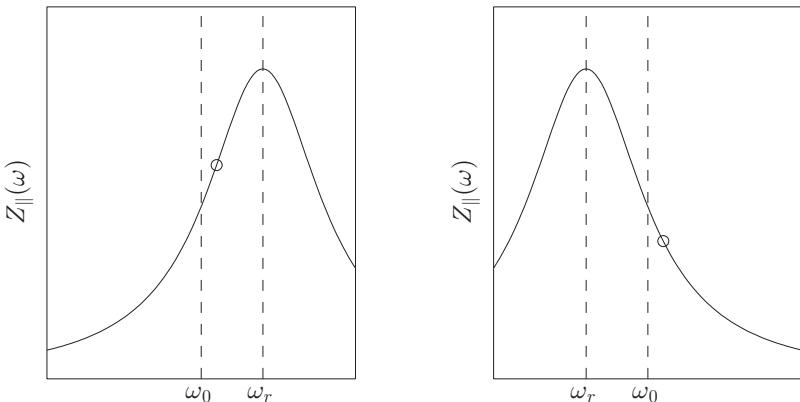


Fig. 15.3 Mechanism of the Robinson instability.  $\omega_0$  is both the oscillation frequency of the field in the cavity and the revolution frequency of a bunch in the storage ring with mean energy deviation  $\langle \delta \rangle = 0$ . In the plot on the left, the cavity is tuned so that its resonant frequency  $\omega_r$  is slightly above  $\omega_0$ . If the ring is below transition, then a bunch with positive mean energy deviation (represented by the circle) has a revolution frequency above  $\omega_0$ : particles in the bunch see a higher impedance in the cavity, and so lose more energy to the cavity than particles in a bunch with zero mean energy deviation. Coherent synchrotron oscillations of the bunch will be damped. In the plot on the right, the cavity is tuned so that its resonant frequency is slightly below  $\omega_0$ . In this case, particles in a bunch with positive mean energy deviation will see a lower impedance, and lose less energy to the cavity than particles in a bunch with zero energy deviation. Coherent synchrotron oscillations of the bunch will be antidamped. In a ring above transition, the situation will be reversed: coherent synchrotron oscillations will be antidamped if  $\omega_r > \omega_0$  and damped if  $\omega_r < \omega_0$ .

## 15.2 Potential-Well Distortion

In Section 15.1.2, we found that the longitudinal impedance from resonant cavities in a storage ring can lead to a shift in the frequency of coherent synchrotron oscillations. The effect could be interpreted as a variation in the focusing potential from the rf cavities. For the coherent motion of bunches of particles, we were concerned with the long-range wake fields, corresponding to the low frequency range of the impedance. When studying the dynamics of particles within a bunch, we need to consider the short-range wake fields, corresponding to the impedance at high frequencies. In this section, we shall see that the short-range wake fields can have effects on the longitudinal focusing from the rf cavities that are similar in some respects to the effects from long-range wake fields. Generally, the variation in longitudinal focusing resulting from wake fields is known as *potential-well distortion*.

*distortion.* Our goal here is to understand the effect that potential-well distortion has on the charge distribution within a single bunch.

We shall consider only the longitudinal motion of particles within a bunch in a synchrotron storage ring. Suppose that  $W_{\parallel}(z) \equiv W_0(z)$  is the longitudinal wake function for the entire ring. Including the phase slip, rf focusing and wake fields, the longitudinal equations of motion for a particle within a bunch are:

$$\frac{dz}{ds} = -\eta_p \delta, \quad (15.53)$$

$$\frac{d\delta}{ds} = \frac{\omega_z^2}{\eta_p c^2} z + \frac{qV(z)}{E_0 C_0}, \quad (15.54)$$

where  $z$  is the longitudinal co-ordinate relative to a reference particle with energy  $E_0$ ,  $\delta$  is the energy deviation,  $\eta_p$  is the phase slip factor,  $C_0$  is the circumference of the storage ring,  $\omega_z$  is the synchrotron frequency in the absence of wake fields and  $q$  is the charge of a particle. The potential  $V(z)$  characterises the effects of the wake fields within the storage ring, and is defined by:

$$V(z) = - \int_z^\infty \lambda(z') W_{\parallel}(z - z') dz', \quad (15.55)$$

where  $\lambda$  is the longitudinal distribution (in  $\text{C m}^{-1}$ ) of charge within the bunch. The second term on the right-hand side of (15.54) follows from the definition of the wake potential in terms of the wake function (14.141). The integral in (15.55) extends over all charge ahead of a particle with longitudinal co-ordinate  $z$ .

The equations of motion (15.53) and (15.54) can be derived from a Hamiltonian:

$$H = -\frac{\eta_p}{2} \delta^2 - \frac{\omega_z^2}{2\eta_p c^2} z^2 - \frac{q}{E_0 C_0} \int_0^z V(z') dz'. \quad (15.56)$$

We have averaged the phase slip, the longitudinal focusing and the wake function around the entire circumference of the storage ring: this provides a good approximation to the equations of motion if the synchrotron motion and the wake field effects are slow compared to the revolution frequency. In practice, this will often be the case. The Hamiltonian (15.56) has no explicit dependence upon  $s$ , and is therefore a constant of the motion:

$$\frac{dH}{ds} = 0. \quad (15.57)$$

Since the Hamiltonian is a constant of motion, any function of the dynamical variables  $z$  and  $\delta$  that can be expressed as a function of the Hamiltonian

must also be a constant of the motion. In particular, let  $\Psi(z, \delta)$  represent the particle distribution in longitudinal phase space. If:

$$\Psi(z, \delta) = \Psi(H), \quad (15.58)$$

then the particle distribution will be constant in time.

In the absence of wake fields, a bunch of particles in an electron (or positron) storage ring will be damped to a Gaussian distribution:

$$\begin{aligned} \Psi(z, \delta) &= \frac{N_b}{2\pi\sigma_z\sigma_\delta} e^{-\delta^2/2\sigma_\delta^2} e^{-z^2/2\sigma_z^2} \\ &= \frac{N_b}{2\pi\sigma_z\sigma_\delta} e^{H/\eta_p\sigma_\delta^2}, \end{aligned} \quad (15.59)$$

where  $N_b$  is the total number of particles in the bunch. The equilibrium energy spread  $\sigma_\delta$  is determined by synchrotron radiation effects (7.94), and the equilibrium bunch length  $\sigma_z$  is given by (7.96):

$$\sigma_z = -c \frac{\eta_p}{\omega_z} \sigma_\delta, \quad (15.60)$$

where we have taken the ultra-relativistic limit  $\beta_0 \rightarrow 1$ . In an electron storage ring, changing the rf voltage changes the equilibrium bunch length, but has no impact on the equilibrium energy spread. The rf system and the wake fields are similar in that they both provide a longitudinal force on particles in a bunch, depending on the longitudinal co-ordinate  $z$ . We can therefore assume that the wake fields do not change the equilibrium energy spread, but do have an effect on the longitudinal charge distribution. In particular, we write the distribution in longitudinal phase space in exactly the form (15.59), but with the Hamiltonian given by (15.56). Since the longitudinal charge distribution  $\lambda(z)$  is related to the phase space distribution by:

$$\lambda(z) = q \int_{-\infty}^{\infty} \Psi(z, \delta) d\delta, \quad (15.61)$$

it follows that, in the presence of longitudinal wake fields, the longitudinal charge distribution will satisfy the equation:

$$\lambda(z) = A \exp\left(-\frac{z^2}{2\sigma_z^2} - \frac{q}{\eta_p\sigma_\delta^2 E_0 C_0} \int_0^z V(z') dz'\right), \quad (15.62)$$

where  $\sigma_z$  is given by (15.60). The constant  $A$  is determined by the normalisation condition:

$$\int_{-\infty}^{\infty} \lambda(z) dz = qN_b, \quad (15.63)$$

where  $qN_b$  is the total charge within the bunch.

Equation (15.62) is the Haissinski equation [Haissinski (1973)]: since the potential  $V(z)$  depends on the charge distribution  $\lambda(z)$ , the Haissinski equation is an integral equation for the equilibrium longitudinal charge distribution in a bunch, in a storage ring with given longitudinal wake function. There are certain special cases for which the Haissinski equation can be solved exactly; however, in general, it is necessary to solve the Haissinski equation numerically (for example, using an iterative technique). An example of the distortion of the longitudinal charge distribution resulting from longitudinal wake fields is shown in Fig. 15.4.

Note that there is no guarantee that, for given machine parameters and wake function, a solution to the Haissinski equation exists. If there is no solution to the Haissinski equation, then there will not be an equilibrium longitudinal charge distribution. This means that in a storage ring the charge distribution will change continuously: in other words, the beam will be unstable. Inspecting equation (15.62), we might suppose that instability might occur when the wake function (at short range) is large, when there

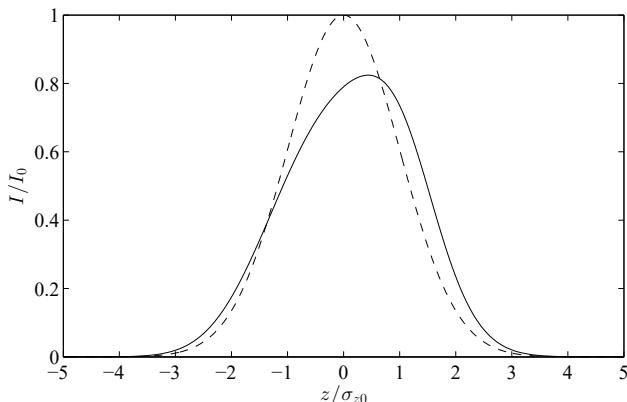


Fig. 15.4 Potential-well distortion in a synchrotron storage ring. The dashed line shows the beam current as a function of longitudinal position within a bunch in the absence of wake fields: the peak current and rms bunch length are  $I_0$  and  $\sigma_{z0}$  respectively. The solid line shows the beam current in the presence of short-range wake fields, calculated by solving the Haissinski equation (15.62) numerically for known wake fields. In the case shown, the wake fields are described in the frequency domain by a resonator impedance of the form (14.181), with  $Q \approx 1$  and  $\omega_r \approx 2c/\sigma_{z0}$ . We observe that the charge distribution becomes asymmetric, and has a tendency to tilt forward, towards the head of the bunch. The storage ring is above transition: particles at the head of the bunch arrive at the rf cavities earlier than the reference particle, and gain more energy from the cavities. This compensates the energy lost to the wake fields.

is a high bunch charge, or when the phase slip or the energy spread are small. Under any of these conditions, the term involving the wake function in (15.62) will become large, and have a significant impact on the beam behaviour. In fact, all these conditions are associated with the onset of beam instability, though it is not completely clear from the Haissinski equation how this happens. Mathematically, it is difficult to prove (for a given set of parameters, and a given wake function) whether a solution to the Haissinski equation exists or not. An alternative approach to determining the stability of the longitudinal charge distribution in a bunch can be developed starting with the Vlasov equation, and is discussed in Section 15.3.

### 15.3 Coasting Beams: Microwave Instability

We saw in Section 15.2 that longitudinal wake fields can lead to a change in the longitudinal focusing potential seen by particles in a storage ring. The change in the focusing potential depends on the charge distribution within a bunch, which itself depends on the focusing potential. The equilibrium distribution (if it exists) satisfies the Haissinski equation (15.62). In practice, it is found that when the charge in a bunch exceeds a certain threshold, the bunch becomes unstable, and no equilibrium distribution exists. It is important to know, when designing or operating a storage ring, the instability threshold: however, this can be difficult to determine from the Haissinski equation.

An alternative approach to finding the instability threshold is to use the Vlasov equation, which is a partial differential equation for the evolution of the charge distribution. If it is assumed that the charge within a storage ring consists of a uniform distribution (i.e. an unbunched beam) with a small density modulation of a given wavelength, then the Vlasov equation leads to a dispersion relation giving the frequency (i.e. the time dependence) of the density modulation. If the frequency has an imaginary part, then the amplitude of the modulation is either damped or antidamped. Thus, solving the dispersion relation for the frequency of the density modulation provides information on the stability of the beam. With some assumptions, it is possible to apply this model to bunched beams in a synchrotron storage ring. In this section, we shall develop the analysis based on the Vlasov equation and the dispersion relation, with the goal of determining the current at which the beam becomes unstable, for given machine parameters and wake function.

It is convenient to work with time as the independent variable, and with

dynamical variables  $\theta$  and  $\delta$ .  $\delta$  is the energy deviation (2.55):

$$\delta = \frac{E}{cP_0} - \frac{1}{\beta_0} \approx \frac{E - E_0}{E_0}, \quad (15.64)$$

where  $E$  is the energy of a particle, and  $E_0$  is the reference energy (corresponding to reference momentum  $P_0$  and reference velocity  $\beta_0 c$ ).  $\theta$  is the particle co-ordinate, giving the longitudinal position around the ring:

$$\theta = 2\pi \frac{s}{C_0}, \quad (15.65)$$

where  $s$  is the distance along the reference trajectory, and  $C_0$  is the circumference of the storage ring. We assume the ultra-relativistic limit,  $\beta_0 \rightarrow 1$ .

In the absence of any rf system, synchrotron radiation or wake fields, the equations of motion for a particle in the storage ring are:

$$\frac{d\theta}{dt} = \omega_0(1 - \eta_p \delta), \quad (15.66)$$

$$\frac{d\delta}{dt} = 0. \quad (15.67)$$

Here,  $\eta_p$  is the phase slip factor, and  $\omega_0$  is the revolution frequency for a particle with  $\delta = 0$ . We shall continue, for the moment, to assume that there is no rf system or synchrotron radiation; then, the particles in the storage ring form a *coasting beam*, and do not perform synchrotron oscillations. Under such conditions, the beam can be expected to fill the storage ring, so that the current at any point is (approximately) constant in time, with some small random fluctuations.

In the presence of wake fields, particles will experience some change in energy as a result of longitudinal forces from the wake fields, and the equation of motion (15.67) will be modified. Let us assume that the longitudinal wake fields can be described by an impedance  $Z_{||}(\omega) \equiv Z_0^{||}(\omega)$ , such that  $Z_{||}(0) = 0$ . That is, the dc component of the beam current will not excite any wake fields (this excludes resistive-wall wake fields, for example). Let us further assume that the beam current at a given location in the ring at a given time is:

$$I(\theta, t) = I_0 + \Delta I e^{i(\theta n - \omega_n t)}. \quad (15.68)$$

The current has a modulation of amplitude  $\Delta I$ , which takes the form of a wave with wavelength  $C_0/n$  (where  $C_0$  is the machine circumference), oscillating (at a given location) with frequency  $\omega_n$ . If  $\omega_n$  has an imaginary part, then the amplitude of the modulation will be either damped or anti-damped. To determine the stability of the beam, we therefore need to find an expression for the frequency of oscillation  $\omega_n$  of the current modulation.

To proceed, we first need to write down the correct equation of motion for  $\delta$ , taking into account the effects of the wake fields. Using (14.161), the variation in the energy deviation of a given particle is:

$$\frac{d\delta}{dt} = -\frac{q}{E_0 T_0} \int_{-\infty}^{\infty} e^{-i\omega t} \tilde{I}(\omega) Z_{||}(\omega) \frac{d\omega}{2\pi}, \quad (15.69)$$

where  $T_0$  is the revolution period. From (15.68), the current spectrum contains a single component at frequency  $\omega_n$ :

$$\tilde{I}(\omega) = 2\pi \Delta I e^{in\theta} \delta_D(\omega - \omega_n), \quad (15.70)$$

where  $\delta_D(\omega - \omega_n)$  is a Dirac delta function. Substituting for  $\tilde{I}(\omega)$  in (15.69) gives:

$$\frac{d\delta}{dt} = -\frac{q}{E_0 T_0} \Delta I Z_{||}(\omega_n) e^{i(n\theta - \omega_n t)}. \quad (15.71)$$

Equations (15.66) and (15.71) provide the equations of motion for a particle in the storage ring in the presence of wake fields. Having obtained the equations of motion for a single particle, the next step is to write down an equation describing the evolution of the charge distribution. When the dynamics of individual particles obey Hamilton's equations (as in this case), the evolution of the particle distribution in phase space is given by the Vlasov equation. In the present case (in one degree of freedom, with dynamical variables  $\theta$  and  $\delta$ ), the Vlasov equation can be written:

$$\frac{\partial \Psi}{\partial t} + \frac{d\theta}{dt} \frac{\partial \Psi}{\partial \theta} + \frac{d\delta}{dt} \frac{\partial \Psi}{\partial \delta} = 0, \quad (15.72)$$

where  $\Psi(\theta, \delta; t)$  is the density of particles in phase space. The Vlasov equation (15.72) is essentially an expression of Liouville's theorem (2.93). Using explicit expressions for  $d\theta/dt$  and  $d\delta/dt$  from (15.66) and (15.71), we obtain a partial differential equation for  $\Psi(\theta, \delta; t)$ . We shall see that this partial differential equation can be manipulated to give an integral equation relating the frequency  $\omega_n$  of the current modulation to the lattice parameters, the impedance and the energy distribution of the beam. This allows us, for a given lattice and impedance, to determine the stability of a beam with a given energy distribution.

To work with the Vlasov equation, it is convenient to replace the amplitude of the current modulation  $\Delta I$  by a quantity describing a modulation of the phase space density. This can be done as follows. The beam current is related to the phase space density by:

$$I(\theta, t) = q\omega_0 \int_{-\infty}^{\infty} \Psi(\theta, \delta; t) d\delta. \quad (15.73)$$

Writing:

$$\Psi(\theta, \delta; t) = \frac{\Psi_0(\delta)}{2\pi} + \frac{\Delta\Psi(\delta)}{2\pi} e^{i(n\theta - \omega_n t)}, \quad (15.74)$$

we see that the current modulation can be directly related to a phase space density modulation:

$$\Delta I = \frac{q\omega_0}{2\pi} \int_{-\infty}^{\infty} \Delta\Psi(\delta') d\delta'. \quad (15.75)$$

Note that the amplitude of the modulation of particle density in co-ordinate space is (in general) a function of the energy deviation. Substituting for  $\Delta I$  from (15.75) in (15.71) gives the equation of motion for  $\delta$  in the form:

$$\frac{d\delta}{dt} = -\frac{q^2}{T_0^2 E_0} \left( \int_{-\infty}^{\infty} \Delta\Psi(\delta') d\delta' \right) Z_{\parallel}(\omega_n) e^{i(n\theta - \omega_n t)}. \quad (15.76)$$

Substituting for  $d\theta/dt$  and  $d\delta/dt$  from (15.66) and (15.76) in the Vlasov equation (15.72), and keeping terms up to first order in the (small) density modulation  $\Delta\Psi(\delta)$  gives:

$$\Delta\Psi(\delta) = -i \frac{q^2}{T_0^2 E_0} \left( \int_{-\infty}^{\infty} \Delta\Psi(\delta') d\delta' \right) Z_{\parallel}(\omega_n) \frac{\partial\Psi_0/\partial\delta}{n\omega_0(1 - \eta_p\delta) - \omega_n}. \quad (15.77)$$

The phase space density modulation as a function of energy deviation  $\Delta\Psi(\delta)$  is unknown; however, it can be eliminated from (15.77) by integrating both sides of the equation over  $\delta$ ; the integral of  $\Delta\Psi(\delta)$  over all  $\delta$  can then be cancelled, leaving:

$$1 = -i \frac{q^2}{2\pi T_0 E_0} \frac{Z_{\parallel}(\omega_n)}{n} \int_{-\infty}^{\infty} \frac{\partial\Psi_0/\partial\delta}{1 - \omega_n/n\omega_0 - \eta_p\delta} d\delta. \quad (15.78)$$

Equation (15.78) is a dispersion relation: in principle, this equation can be solved to find the frequency  $\omega_n$  of a modulation in beam current that has wavelength  $C_0/n$ . The imaginary part of  $\omega_n$  determines whether the modulation is damped (negative imaginary part) or antidamped (positive imaginary part).

It should be remembered that the dispersion relation (15.78) applies to coasting beams. We shall discuss shortly how to apply the results to bunched beams; but first, let us consider two example cases for the energy distribution  $\Psi_0$ . These examples will emphasise the importance of the energy spread for determining the stability of the beam.

As a first example, let us take the case that the beam has zero energy spread; that is, the distribution is a Dirac delta function:

$$\Psi_0(\delta) = N_b \delta_D(\delta), \quad (15.79)$$

where  $N_b$  is the total number of particles in the beam. Integrating by parts, we find:

$$\int_{-\infty}^{\infty} \frac{\partial \Psi_0 / \partial \delta}{1 - \omega_n / n\omega_0 - \eta_p \delta} d\delta = -\frac{N_b \eta_p}{(1 - \omega_n / n\omega_0)^2}. \quad (15.80)$$

Then, the dispersion relation (15.78) gives:

$$\frac{\omega_n}{n\omega_0} = 1 \pm \sqrt{i \frac{qI_0 \eta_p}{2\pi E_0} \frac{Z_{||}(n\omega_0)}{n}}, \quad (15.81)$$

where the average beam current is:

$$I_0 = \frac{qN_b}{T_0}, \quad (15.82)$$

and we assume that we can write  $Z_{||}(\omega_n) \approx Z_{||}(n\omega_0)$ . For a purely capacitive impedance (i.e.  $\text{Im}(Z_{||}(\omega)) > 0$ ), the beam is stable below transition ( $\eta_p < 0$ ), but unstable above transition ( $\eta_p > 0$ ). The space-charge impedance (12.229) is an example of a purely capacitive impedance. The instability of the beam that occurs in this case is known as the *negative-mass instability* [Nielsen and Sessler (1959a)], and can be understood from some simple physical arguments. Particles just ahead of a ‘peak’ in the particle density gain energy from the space-charge forces, while particles just behind the peak lose energy. Below transition, particles with higher energy have a higher revolution frequency, and will therefore tend to move forward of the density peak. Particles with lower energy will slip back relative to the peak. The net effect is to ‘smear out’ the peak in the particle density. Above transition, however, particles ahead of the peak that gain energy from the space-charge forces will slip back towards the peak, while those behind the peak will move forwards. As a result, any peak in the particle density will tend to increase rather than decrease.

In general, the longitudinal impedance in an accelerator will have a real part as well as an imaginary part. From (15.81), we see that with a purely inductive impedance, or an impedance that has a real (resistive) part, there will always be a solution for  $\omega_n$  with a positive imaginary part. Therefore, a beam with zero energy spread will always be unstable.

As a second example, consider a coasting beam with a Gaussian energy spread:

$$\Psi_0(\delta) = \frac{N_b}{\sqrt{2\pi}\sigma_\delta} e^{-\delta^2/2\sigma_\delta^2}. \quad (15.83)$$

Substituting into the dispersion relation (15.78), and changing the variable of integration from  $\delta$  to  $\zeta = \delta/\sigma_\delta$  gives:

$$1 = -i \frac{qI_0}{(2\pi)^{3/2} E_0 \eta_p \sigma_\delta^2} \frac{Z_{||}(\omega_n)}{n} \int_{-\infty}^{\infty} \frac{\zeta e^{-\zeta^2/2}}{\Delta_n + \zeta} d\zeta, \quad (15.84)$$

where  $I_0$  is the average beam current (15.82), and we define the quantity  $\Delta_n$ :

$$\Delta_n = \frac{\omega_n/n\omega_0 - 1}{\eta_p\sigma_\delta}. \quad (15.85)$$

At this stage, we should solve (15.84) for  $\Delta_n$  (for a given impedance), and thus obtain the imaginary part of  $\omega_n$ . However, this is difficult to do, so we take a different approach. If we define a function  $F(\Delta_n)$ :

$$\frac{1}{F(\Delta_n)} = -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\zeta e^{-\zeta^2/2}}{\Delta_n + \zeta} d\zeta, \quad (15.86)$$

then (15.84) can be written in the form:

$$\frac{qI_0}{2\pi E_0 \eta_p \sigma_\delta^2} \frac{Z_{||}(\omega_n)}{n} = F(\Delta_n). \quad (15.87)$$

Now, we can plot the real and imaginary parts of  $F(\Delta_n)$  for a range of values of  $\Delta_n$  with fixed imaginary part. Since the imaginary part of  $\Delta_n$  can only come from the imaginary part of  $\omega_n$ , lines in the complex plane constructed in this way must correspond to fixed growth or damping rates of the beam current modulation.

The solid curve in Fig. 15.5 shows:

$$u + iv = F(\Delta_n), \quad (15.88)$$

for a range of real values of  $\Delta_n$ . Values of  $\omega_n$  with a negative imaginary part (corresponding to density modulations that are damped) lead to points within the solid curve; values of  $\omega_n$  with a positive imaginary part (corresponding to undamped density modulations) lead to points outside the solid curve. To determine whether a beam will be stable with given parameters and with a given impedance, we can plot a curve defined by:

$$u + iv = \frac{qI_0}{2\pi E_0 \eta_p \sigma_\delta^2} \frac{Z_{||}(\omega)}{\omega/\omega_0}, \quad (15.89)$$

for a range of values of  $\omega$ . If  $\omega_n \approx n\omega_0$ , then the curve (15.89) will be close to the curve defined by:

$$u + iv = \frac{qI_0}{2\pi E_0 \eta_p \sigma_\delta^2} \frac{Z_{||}(\omega_n)}{n}, \quad (15.90)$$

for a range of values of  $n$ . If the curve defined by (15.89) for a range of  $\omega$  crosses the boundary defined by (15.88) for real  $\Delta_n$ , then the dispersion relation will have solutions for  $\omega_n$  with negative imaginary part. In that case, the beam will be unstable.

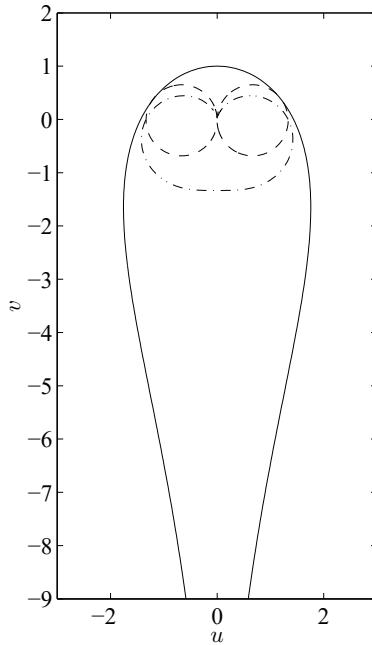


Fig. 15.5 Derivation of the Keil–Schnell criterion (15.95), providing an estimate of the microwave instability threshold. The solid curve shows  $u + iv = F(\Delta_n)$ , defined by (15.86), for real values of  $\omega_n$ . Inside the curve, the real part of  $\omega_n$  is negative (i.e. a beam density modulation with frequency  $\omega_n$  is damped); outside the curve, the real part of  $\omega_n$  is positive (a beam density modulation with frequency  $\omega_n$  is antidamped). The dash-dotted and dashed lines show the curves defined by (15.89) with the resonator impedance (15.91) and, respectively,  $Q = 1$  and  $Q = 20$ . The value of the shunt resistance  $R_s$  is chosen to satisfy (15.92).

For example, consider a resonator impedance (14.181):

$$Z_{\parallel}(\omega) = \frac{R_s}{1 + iQ \left( \frac{\omega_r}{\omega} - \frac{\omega}{\omega_r} \right)}. \quad (15.91)$$

The dashed line in Fig. 15.5 shows the curve defined by (15.89) with the impedance (15.91) calculated over a range of values of  $\omega$ , with:

$$\frac{qI_0}{2\pi E_0 \eta_p \sigma_{\delta}^2} \frac{\omega_0}{\omega_r} R_s = \frac{4}{3}. \quad (15.92)$$

With the condition (15.92), the curve defined by (15.89) comes close to (for small  $Q \approx 1$ ), or crosses over (for larger  $Q \gg 1$ ), the curve defined by (15.88). Thus, a condition for the beam to be stable can be written:

$$I_0 < \frac{8\pi E_0 |\eta_p| \sigma_{\delta}^2}{3q\omega_0 R_s / \omega_r}. \quad (15.93)$$

The dashed curve in Fig. 15.5 crosses over the solid curve for frequencies  $n\omega_0 \approx \omega_r$ , where  $\omega_r$  is the resonant frequency of the impedance source. The unstable modes in the beam density will be those modes with frequency close to the resonant frequency of the impedance driving the instability. In a storage ring, contributions to the impedance come from many different components; resonances with wavelengths of order of a millimetre are quite common. A longitudinal instability leading to a density modulation on the beam of this length scale can lead to the generation of significant amounts of microwave radiation, hence the term *microwave instability*.

The resonator impedance (15.91) is a special case. More generally, the instability threshold is sometimes expressed in terms of a limit on the magnitude of the impedance, without reference to a particular form of impedance. Using the fact that the imaginary part of  $\Delta_n$  is always negative if:

$$|F(\Delta_n)| < 1, \quad (15.94)$$

we can say that the beam should be stable if:

$$\left| \frac{Z_{||}(n\omega_0)}{n} \right| < \frac{2\pi E_0 |\eta_p| \sigma_\delta^2}{qI_0}, \quad (15.95)$$

for all  $n$ . The condition (15.95) is known as the *Keil–Schnell criterion* [Keil and Schnell (1969)].

A fundamental assumption that we made in deriving the Keil–Schnell criterion was that the beam was a coasting (unbunched) beam, with no longitudinal focusing, i.e. with no synchrotron oscillations. It is possible to relax this condition, to find a (modified) Keil–Schnell criterion that may be applied to bunched beams, as follows. Consider a bunched beam, with a modulation in the particle density along the length of each bunch. For this model to be meaningful, the wavelength of the modulation should be small compared to the bunch length. For example, if the bunches have a Gaussian longitudinal profile with standard deviation  $\sigma_z$ , then we consider only modulations in the density with wavelength  $C_0/n$ , where  $C_0$  is the ring circumference, and:

$$n \gg \frac{C_0}{\sigma_z}. \quad (15.96)$$

Also, the time scale of the development of any instability should be short compared with the synchrotron period. For density modulations satisfying this condition, we assume that each bunch can be regarded as a continuous

beam, with current equal to the peak current in the bunch. Then, the Keil–Schnell criterion is modified by setting:

$$I_0 = \frac{qN_b c}{\sqrt{2\pi}\sigma_z}, \quad (15.97)$$

where  $N_b$  is the bunch population. The resulting stability condition is known as the *Keil–Schnell–Boussard criterion* [Boussard (1975)]:

$$\left| \frac{Z_{||}(n\omega_0)}{n} \right| < \frac{(2\pi)^{3/2} E_0 |\eta_p| \sigma_\delta^2 \sigma_z}{q^2 N_b c}. \quad (15.98)$$

While the Keil–Schnell criterion (or Keil–Schnell–Boussard criterion) can sometimes be useful in indicating the likelihood of an instability occurring, the results should not, in general, be taken too literally. In particular, note that the Keil–Schnell criterion is stated as a condition that should be satisfied for the beam to be stable: it does not imply that if the condition is not satisfied, then the beam will be unstable. The reason for this is simply that the boundary of the region with  $\text{Im}(\omega_n) < 0$  is not a simple circle in the complex plane. In particular, as we see in Fig. 15.5, all points along the negative imaginary axis lie within the region  $\text{Im}(\omega_n) < 0$ . This means that a beam with a Gaussian energy distribution will always be stable no matter how large the magnitude of the impedance, as long as the impedance is purely capacitive below transition, or purely inductive above transition. This is the situation that we found for the space-charge impedance, in the case of a beam with zero energy spread. However, if the absolute value of the impedance is large, even a small real part of the impedance can lead to a beam instability.

A proper assessment of beam stability requires a detailed knowledge of the machine impedance. Unfortunately, calculation of the impedance in a storage ring is not an easy task, and (especially for large rings) can require a considerable amount of effort. The high frequency regime of the impedance is often significant; but an accurate calculation of the high frequency impedance of a section of beam line depends upon solution of Maxwell's equations over large spatial volumes with high resolution. Measurement of the machine impedance in an operational storage ring also presents considerable challenges. Usually, it is found that the observed behaviour of the beam in a storage ring is consistent with a higher impedance than would be expected from calculations based on a model of the vacuum chamber. The impedance of a vacuum chamber is sometimes characterised as a single value of  $Z_{||}/n$ . A broad-band resonator (i.e. a resonator with

quality factor  $Q \approx 1$ ) behaves for frequencies  $\omega < \omega_r$  as:

$$\frac{|Z_{\parallel}(\omega)|}{\omega} \approx \frac{R_s}{\omega_r}, \quad (15.99)$$

and so:

$$\frac{|Z_{\parallel}(n\omega_0)|}{n} \approx \frac{\omega_0 R_s}{\omega_r} = \frac{2\pi c R_s}{\omega_r C_0}, \quad (15.100)$$

where  $R_s$  is the shunt impedance. Hence,  $Z_{\parallel}/n$  is essentially the average value of the impedance per unit length of the storage ring. Modern storage rings designed for high beam currents typically aim for values  $Z_{\parallel}/n$  of the order of 100 mΩ or less.

Although the Keil–Schnell criterion (15.95) does not provide a rigorous condition for beam stability, it is still interesting to discuss the dependence of the instability threshold indicated by (15.95) on the various machine parameters. First of all, note that there is a linear dependence on the beam energy: a higher energy beam has a higher beam rigidity, and will be less sensitive to external fields affecting the motion of particles in the beam. Second, we see that the limit on the impedance is inversely proportional to the particle charge and the beam current. The strength of the wake fields generated in the beam pipe depends on the beam current; the force on a particle from the wake fields depends on the particle charge. Finally, notice that the limit on the impedance depends on the product of the phase slip factor and the square of the energy spread: the dependence of the beam stability on these quantities is important for machine design and operation, and can be understood from some simple physical arguments. Consider a beam that starts to develop a density modulation close to the resonant frequency of some impedance source in the ring. If the impedance has the right characteristics, the wake field generated by the density modulation acts back on the beam to increase the amplitude of the modulation: the result will be an exponential growth in the amplitude of the density modulation. However, particles in the beam in a storage ring travel at different rates around the ring, depending on the particle energy and the phase slip factor. If the ring has a large phase slip factor and the beam has a large energy spread, then the range in particle revolution frequencies prevents the development of any density modulation. A density modulation of significant amplitude will only appear if the impedance effects are large enough to drive a growth in the amplitude of the modulation that is fast compared to the rate at which the modulation is smoothed out by the range in revolution frequencies of the particles. This results in a threshold effect, where the

impedance (or the beam current) has to exceed a certain value before an instability (characterised by a density modulation of increasing amplitude) occurs.

The suppression of an instability by the spread in revolution (or oscillation) frequencies of particles in a beam in an accelerator is known as *Landau damping*. Landau damping was first described in the context of plasmas [Landau (1946)], where decoherence from a spread in particle oscillation frequencies leads to a decay in the amplitudes of plasma waves. Strictly speaking, Landau damping in an accelerator does not describe the decoherence of the motion of particles within a bunch, but rather the suppression of the development of any coherent motion in the first place. The distinction is significant because decoherence generally leads to a growth in the emittance of a beam in an accelerator, whereas Landau damping does not.

Despite the differences between Landau damping in a plasma (describing the decay in amplitude of plasma waves) and Landau damping in an accelerator (describing the suppression of an instability), there are fundamental similarities in the mechanisms involved, which in both cases have their origin in the spread of particle oscillation frequencies [Nielsen and Sessler (1959b)]. If a beam in a storage ring has zero energy spread or if the phase slip factor is zero, then all particles move around the ring at the same rate and any density modulations that develop can remain coherent, i.e. there is no Landau damping. The result is that the instability threshold (in terms of either the impedance or the beam current) is effectively zero.

It is important to note that even when an instability does occur, it may not significantly affect the performance of a storage ring. The results we have derived, in particular the Keil–Schnell criterion (15.95), give some indication of the conditions under which an instability is likely to occur, but do not provide any real description of how the beam behaves above the instability threshold. One reason for this is that we made a linear approximation (in terms of the amplitude of the density modulation) to the Vlasov equation. Once the density modulation becomes large enough to be observable, it can no longer be regarded as a small perturbation, and the linear approximation to the Vlasov equation is no longer valid.

In practice, it is not uncommon for synchrotron storage rings to operate with bunch currents that are above the threshold for microwave instability. In an electron storage ring operating below the instability threshold, the charge distribution within a bunch achieves an equilibrium determined by the effects of synchrotron radiation. In the longitudinal phase space, the

distribution will be a smooth two-dimensional Gaussian distribution. If more particles are injected into a bunch, so that the current increases to above the instability threshold, density modulations will start to appear. Because of the synchrotron oscillations, density variations that are initially functions only of the longitudinal co-ordinate  $z$  are ‘rotated’ in the longitudinal phase space, so that there are also density modulations in the energy spread. The overall dynamics can become quite complicated, but the observable result may be little more than some increase in the bunch length and energy spread, an effect sometimes known as *turbulent bunch lengthening*. In other words, there is an overall increase in the longitudinal emittance but the charge distribution still achieves an equilibrium. If the beam current is further increased, then more complicated effects can start to appear; for example, the distribution of particles within a bunch may vary periodically (or quasi-periodically) and fail to reach any equilibrium. For coupled-bunch instabilities, it is possible to use fast feedback systems to suppress the development of coherent bunch oscillations. However, for single-bunch instabilities, the time scales for the development of the instability are often too short for feedback systems to provide much means of control. The main mitigation is in the design of the storage ring for appropriate values of phase slip factor and energy spread, and in careful design and manufacture of the beam pipe to keep the impedance as low as possible.

#### 15.4 Single-Bunch Instabilities

In Section 15.1 we discussed the effects of wake fields on the beam using a model in which each bunch was treated as a single macroparticle. In other words, it was assumed that all particles within a bunch moved in the same way in response to the forces from the wake fields. This model is able to describe some significant effects; however, as we have seen in Sections 15.2 and 15.3, there are also important effects from wake fields that impact the distribution of charge within a bunch. Since the number of particles in a bunch can be very large (typically, of order  $10^9$  or  $10^{10}$ ), it can be difficult to develop an accurate description of the collective dynamics of particles within a bunch. Two approaches are possible. One is to represent the bunch as a continuum of charge, and perform an analysis using techniques from fluid dynamics. This is essentially the approach we took to our analysis of potential well distortion in Section 15.2, and to the microwave instability in Section 15.3: in both cases, although we began with the equations of

motion for individual particles, ultimately we developed formulae describing the equilibrium or evolution of the charge density, treating the bunch (or the beam) as a continuum of charge. A second approach, which can be more appropriate in some cases, is to represent the bunch as a collection of a much smaller number of particles — perhaps just two! Neither the continuum model nor the ‘few-particle’ model is strictly accurate, but each approach can, when applied in an appropriate context, provide some insights into the beam behaviour, and lead to some useful results. In this section, we shall describe a variety of different phenomena: some are more easily understood using a continuum model, while others are better treated using a macroparticle model. We shall use the most appropriate model in each case. Our goal is generally to understand the impact that the wake fields have on the charge distribution within a single bunch and, in particular, to determine the conditions under which a bunch may fail to achieve a steady, equilibrium distribution; i.e. we shall try to determine the conditions under which an instability develops.

#### 15.4.1 *Head-tail instability*

We shall begin our discussion of single-bunch instabilities by considering the betatron motion of particles in a bunch moving around a synchrotron storage ring in the presence of transverse wake fields. The wake fields lead to a coupling between the betatron oscillations of different particles within the bunch: because the force on a trailing particle from the wake field of a leading particle depends on the distance between the particles, to understand the dynamics in this case, the synchrotron motion has to be taken into account. We shall analyse the dynamics by representing the bunch as two macroparticles, each with half the total charge of the bunch. Although this is a very simple model, it nevertheless provides some useful results: in particular, we find that above a certain threshold bunch charge, the motion of the macroparticles becomes unstable, with the betatron oscillation amplitudes growing exponentially. Even below the instability threshold, the wake fields modify the dynamics by causing a change in the betatron tune: our simplified model provides the frequencies of the normal modes of oscillation, and shows how they vary with the bunch charge.

To help develop an understanding of the dynamics, we shall initially discuss the motion of the macroparticles in the absence of chromaticity. This simplifies the analysis by allowing us to neglect any change of betatron oscillation frequency resulting from the energy deviation of the particles.

However, chromaticity does play an important role in the dynamics of the head-tail instability, since the energy deviation of a particle will change over the course of a synchrotron period as a natural consequence of the synchrotron motion, and we shall therefore extend the analysis by considering the case where the chromaticity is non-zero. We find that the presence of chromaticity modifies the nature of the instability, which no longer exhibits a distinct threshold, but occurs even in the limit of low bunch charge. Although it now appears that if the chromaticity is non-zero the betatron motion in a storage ring will *always* be unstable, the growth rates at low charge are generally slow enough that the instability can be suppressed by fairly weak, naturally occurring effects (such as decoherence), as long as the chromaticity is not too large.

**Zero chromaticity: fast head-tail instability** We initially consider the case in which the storage ring has zero chromaticity: this allows us to neglect any effects from the energy deviation of the particles in the ring, so we can use a simplified form for the equations of motion, involving only the transverse co-ordinate  $x$  and longitudinal co-ordinate  $z$ . We further simplify the system by representing the bunch by two macroparticles, as shown in Fig. 15.6, with each macroparticle having half the total charge of the bunch. Each macroparticle performs (coherent) betatron and synchrotron oscillations as it moves around the storage ring.

Assume first that macroparticle 1 is ahead of macroparticle 2. By causality, macroparticle 1 is not affected by short-range wake fields from macroparticle 2, so the equation of motion of macroparticle 1 can be written:

$$\frac{d^2x_1}{ds^2} + k_\beta^2 x_1 = 0, \quad (15.101)$$

where  $x_1$  is the transverse co-ordinate (which we assume to be horizontal, but may equally be vertical),  $s$  is the distance around the storage ring, and  $2\pi/k_\beta$  is the betatron wavelength. The equation of motion (15.101) simply describes betatron oscillations with wavelength  $2\pi/k_\beta = 2\pi\beta_x$ , where  $\beta_x$  is the beta function: note that we use the ‘smooth focusing’ approximation, in which we treat the beta function as constant around the ring.

Macroparticle 2 will experience wake fields from macroparticle 1, which means that the equation of motion for macroparticle 2 will have a driving term. Assuming that the particles are ultra-relativistic, so that they are

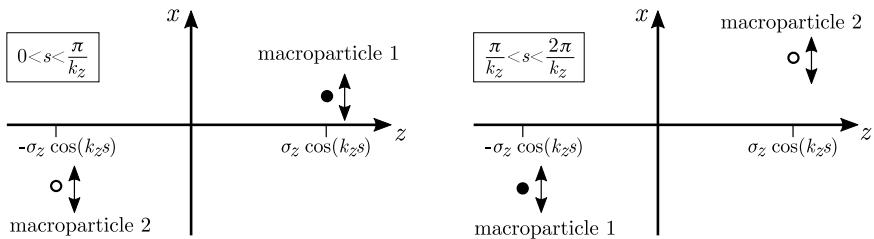


Fig. 15.6 Macroparticle model of the dynamics in the head-tail instability. The bunch is represented as two macroparticles, each with half the total charge of the bunch. The macroparticles perform betatron and synchrotron oscillations as they move around the ring. In the first half synchrotron period (left), macroparticle 1 is ahead of macroparticle 2, so that the motion of macroparticle 2 is affected by the wake fields from macroparticle 1. In the second half synchrotron period (right) the situation is reversed.

travelling close to the speed of light, we can write the equation of motion:

$$\frac{d^2x_2}{ds^2} + k_\beta^2 x_2 = \frac{F_\perp}{\gamma m_2 c^2}, \quad (15.102)$$

where  $F_\perp$  is the force on macroparticle 2 from the wake field generated by macroparticle 1, macroparticle 2 has mass  $m_2 = N_b m/2$ , where there are a total  $N_b$  real particles of mass  $m$  in the full bunch, and  $\gamma$  is the relativistic factor.

To express the driving term in terms of the wake function for the storage ring, we make two further assumptions. First, we assume that the wake field can be averaged over the circumference of the ring: this can be a reasonable assumption in many cases, but in some situations the wake fields of particular components may dominate (so that particles receive a ‘kick’ at discrete points in the ring), in which case a different approach to the analysis might be more appropriate. Second, we assume that the synchrotron motion is slow, so that over a single turn of the storage ring, there is only a small change in the distance between the macroparticles. With these assumptions, we can write the driving force (using Eq. (14.139), with  $m = 1$  and  $I_m = q_1 x_1$ ) as:

$$F_\perp = -q_1 q_2 \frac{W_\perp(-\Delta z)}{C_0} x_1 = -\frac{N_b^2 q^2}{4} \frac{W_\perp(-\Delta z)}{C_0} x_1, \quad (15.103)$$

where  $q_1 = q_2 = N_b q/2$  are the charges of macroparticles 1 and 2 (with  $q$  the charge of each of the real particles in the bunch), and  $W_\perp(-\Delta z)$  is the wake function for the ring, which has circumference  $C_0$ . Macroparticle 2 is a distance  $\Delta z$  behind macroparticle 1. Because of the synchrotron motion of the macroparticles,  $\Delta z$  will vary as a function of the distance  $s$  travelled

by the bunch; but we assumed (as already mentioned) that the change in  $\Delta z$  over one turn of the ring is small.

The next step is to solve the equations of motion. In principle, since the separation of the macroparticles varies as they continue to travel around the ring, we need to know the form of the wake function  $W_{\perp}(-\Delta z)$ . For simplicity, we shall assume that we can make a linear approximation, so that:

$$W_{\perp}(-\Delta z) = \frac{\Delta z}{\sigma_z} W_{\perp}(-\sigma_z), \quad (15.104)$$

where  $\sigma_z$  is the rms bunch length. Using (15.104), we can find a solution to the equations of motion. It is convenient to use complex notation for the macroparticle co-ordinates, so that, taking the synchrotron motion into account, we write:

$$\Delta z = 2i\sigma_z e^{-ik_z s}, \quad (15.105)$$

where the synchrotron frequency is  $\omega_s = ck_z$ , and we assume that each macroparticle has synchrotron amplitude equal to the rms bunch length (so that the maximum separation between the macroparticles is twice the bunch length). The physical separation is the real part of the complex expression in (15.105): the complex phase (factor of  $i$ ) is chosen, so that at  $s = 0$  the macroparticles have zero longitudinal separation, with macroparticle 1 moving towards the head of the bunch, and macroparticle 2 moving towards the rear of the bunch. The equation of motion for macroparticle 2 is then:

$$\frac{d^2 x_2}{ds^2} + k_{\beta}^2 x_2 = -iA k_z k_{\beta} x_1 e^{-ik_z s}, \quad (15.106)$$

where the dimensionless constant  $A$  is defined:

$$A = \frac{N_b q^2 W_{\perp}(-\sigma_z)}{k_z k_{\beta} E_0 C_0}, \quad (15.107)$$

with  $E_0 = \gamma mc^2$  the particle energy.

The equations of motion (15.101) and (15.106) have solution:

$$x_1(s) = x_{10} e^{-ik_{\beta}s}, \quad (15.108)$$

$$x_2(s) = B_1 e^{-ik_{\beta}s} + B_2 e^{-i(k_{\beta}+k_z)s}, \quad (15.109)$$

where  $x_{10}$  is the value of  $x_1(s)$  at  $s = 0$ . The constants  $B_1$  and  $B_2$  are determined from the initial conditions for macroparticle 2, which can be written:

$$x_2(0) = x_{20}, \quad x'_2(0) = -ik_{\beta} x_{20}. \quad (15.110)$$

Here,  $x_2(0)$  is the value of  $x_2(s)$  at  $s = 0$ , and  $x'_2(0)$  is the rate of change of  $x_2$  with  $s$  at  $s = 0$ , in the approximation that the macroparticle simply performs betatron oscillations with wavelength  $2\pi/k_\beta$ , i.e. in the limit  $k_z/k_\beta \rightarrow 0$ . The full expressions for  $B_1$  and  $B_2$  are quite complicated, but can be simplified by making some approximations. In particular, our assumption that the synchrotron motion is slow compared to the betatron motion means that  $k_z \ll k_\beta$ , in which case we find:

$$B_1 \approx x_{20} - \frac{iA}{2}x_{10}, \quad \text{and} \quad B_2 \approx \frac{iA}{2}x_{10}. \quad (15.111)$$

Equation (15.109) expresses the fact that in the presence of the wake field from macroparticle 1, the transverse motion of macroparticle 2 is not simply sinusoidal at the betatron frequency, but includes a component at a synchrotron sideband of the betatron frequency, i.e. at a frequency  $\omega_\beta + \omega_s$ . The amplitude of this component (given by  $B_1$ ) depends on the strength of the wake field, which in turn depends on the wake function in the storage ring and the betatron amplitude of macroparticle 1.

Each half synchrotron period, the macroparticles swap longitudinal positions: if macroparticle 2 is behind macroparticle 1 in the first half synchrotron period, and experiences the wake field of macroparticle 1 over this time, then in the second half synchrotron period, macroparticle 1 will be behind macroparticle 2 and will experience the wake field of macroparticle 2. We therefore need to consider the motion in each half synchrotron period separately. In the first half synchrotron period, the solution to the equations of motion can be written:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{s=\pi/|k_z|} \approx e^{-\pi ik_\beta/|k_z|} \begin{pmatrix} 1 & 0 \\ -iA & 1 \end{pmatrix} \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix}. \quad (15.112)$$

Note that in writing half the synchrotron wavelength as  $\pi/|k_z|$ , we take the modulus of  $k_z$  to allow for the possibility that  $k_z$  may be negative, depending on the operating mode of the storage ring. By symmetry, we can write the solution to the equations of motion in the second half synchrotron period:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{s=2\pi/|k_z|} \approx e^{-\pi ik_\beta/|k_z|} \begin{pmatrix} 1 & -iA \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{s=\pi/|k_z|}. \quad (15.113)$$

Combining (15.112) and (15.113), we find that, for a full synchrotron period:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{s=2\pi/|k_z|} \approx R \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix}, \quad (15.114)$$

where the matrix  $R$  is given by:

$$R = e^{-2\pi i k_\beta / |k_z|} \begin{pmatrix} 1 - A^2 & -iA \\ -iA & 1 \end{pmatrix}. \quad (15.115)$$

$R$  is a transfer matrix for the co-ordinates of the macroparticles over one synchrotron period. For the motion to be stable, the trace of  $R$  must have magnitude less than 2 (see the discussion in Section 4.4):

$$|\text{Tr}(R)| = |2 - A^2| \leq 2. \quad (15.116)$$

Hence, we obtain the stability condition:

$$|A| = \frac{N_b q^2 |W_\perp(-\sigma_z)|}{|k_z| k_\beta E_0 C_0} \leq 2. \quad (15.117)$$

This gives a limit on the total bunch charge  $N_b q$  for the bunch to be stable, in terms of the strength of the wake field, the betatron and synchrotron wavelengths, the beam energy and the circumference of the ring.

Further analysis of the matrix  $R$  can provide additional insight into the dynamics in the presence of wake fields. In particular, using the eigenvalues of  $R$  we can find the frequencies of the normal modes of oscillation of the macroparticles, i.e. the frequencies of the modes in which the macroparticles perform transverse oscillations described by single, well-defined frequencies. To obtain the normal mode frequencies, we construct quantities  $x_+$  and  $x_-$  as linear combinations of  $x_1$  and  $x_2$  that satisfy the equation of motion:

$$\frac{d^2}{ds^2} \begin{pmatrix} x_+ \\ x_- \end{pmatrix} = -\frac{1}{c^2} \begin{pmatrix} \Omega_+^2 & 0 \\ 0 & \Omega_-^2 \end{pmatrix} \begin{pmatrix} x_+ \\ x_- \end{pmatrix}. \quad (15.118)$$

The motion described by  $x_+$  (with  $x_- = 0$ ), and by  $x_-$  (with  $x_+ = 0$ ) are the normal modes, and the oscillation frequencies are (respectively)  $\Omega_+$  and  $\Omega_-$ . The solution over one synchrotron period for the normal mode equation of motion (15.118) is:

$$\begin{pmatrix} x_+ \\ x_- \end{pmatrix}_{s=2\pi/|k_z|} = R_{\text{NM}} \begin{pmatrix} x_+ \\ x_- \end{pmatrix}_{s=0}, \quad (15.119)$$

where  $R_{\text{NM}}$  is the ‘transfer matrix’ for the normal modes:

$$R_{\text{NM}} = \begin{pmatrix} e^{-2\pi i \Omega_+ / |\omega_s|} & 0 \\ 0 & e^{-2\pi i \Omega_- / |\omega_s|} \end{pmatrix}. \quad (15.120)$$

$\omega_s = ck_z$  is the synchrotron frequency, so that  $2\pi\Omega_\pm / |\omega_s|$  are the phase advances of the normal modes over one synchrotron period. Since  $x_+$  and  $x_-$  are linear combinations of  $x_1$  and  $x_2$ , equation (15.119) must be obtained by a diagonalization of (15.114). This means that the eigenvalues of  $R$  are

the same as the eigenvalues of  $R_{\text{NM}}$ , and hence are related to the phase advances of the normal modes of oscillation over one synchrotron period:

$$\text{eigenvalues}(R) = \text{eigenvalues}(R_{\text{NM}}) = e^{-2\pi i \Omega_{\pm}/|\omega_s|}. \quad (15.121)$$

Since there are two macroparticles in our model, there are two normal modes, and therefore two eigenvalues, which from (15.120) are found to be:

$$e^{-2\pi i \Omega_{\pm}/|\omega_s|} = \left(1 - \frac{A^2}{2} \mp i \frac{A}{2} \sqrt{4 - A^2}\right) e^{-2\pi i k_{\beta}/|k_z|}. \quad (15.122)$$

It then follows that:

$$\frac{\Omega_{\pm} - \omega_{\beta}}{|\omega_s|} = \frac{i}{2\pi} \ln \left(1 - \frac{A^2}{2} \mp i \frac{A}{2} \sqrt{4 - A^2}\right), \quad (15.123)$$

where  $\omega_{\beta} = ck_{\beta}$  is the betatron frequency in the limit of low bunch charge (i.e. in the absence of wake fields). Note that for  $|A| < 2$ , we have:

$$\left|1 - \frac{A^2}{2} \mp i \frac{A}{2} \sqrt{4 - A^2}\right| = 1. \quad (15.124)$$

In this case, we can write:

$$\frac{\Omega_{\pm} - \omega_{\beta}}{|\omega_s|} = \pm \frac{1}{2\pi} \tan^{-1} \left( \frac{A\sqrt{4 - A^2}}{2 - A^2} \right), \quad (15.125)$$

and it follows that the normal mode frequencies  $\Omega_{\pm}$  are purely real (have zero imaginary part). This means that as they move around the storage ring, the macroparticles perform betatron oscillations described by some combination of the normal modes, with constant amplitudes: for  $|A| < 2$ , analysis of the normal modes leads to the conclusion that the motion of the particles is stable, which is consistent with the stability condition (15.117) that we derived above. The fact that the normal mode frequencies  $\Omega_{\pm}$  are different from  $\omega_{\beta}$  (the betatron frequency in the limit of low bunch charge) means that there is a tune shift as a result of the wake field, with the size of the tune shift dependent on the value of  $A$ .

In the case that  $|A| > 2$ , we find:

$$\frac{\Omega_{\pm} - \omega_{\beta}}{|\omega_s|} = \pm \frac{1}{2} + \frac{i}{2\pi} \ln \left( \left|1 - \frac{A^2}{2} \pm \frac{A}{2} \sqrt{A^2 - 4}\right| \right). \quad (15.126)$$

Since:

$$\left(1 - \frac{A^2}{2} + \frac{A}{2} \sqrt{A^2 - 4}\right)^{-1} = 1 - \frac{A^2}{2} - \frac{A}{2} \sqrt{A^2 - 4}, \quad (15.127)$$

it follows that:

$$\frac{\Omega_{\pm} - \omega_{\beta}}{|\omega_s|} = \pm \frac{1}{2} \pm \frac{i}{2\pi} \ln \left( \left|1 - \frac{A^2}{2} + \frac{A}{2} \sqrt{A^2 - 4}\right| \right). \quad (15.128)$$

The normal mode frequencies are again ‘shifted’ from the low-charge betatron frequencies: the tune shift is given by the real part of  $\Omega_{\pm}$ , which for  $|A| > 2$  is always equal to half the synchrotron frequency, and independent of the exact value of  $A$ . We also see, however, that in the case  $|A| > 2$  the normal mode frequencies have a non-zero imaginary part. Since the matrix  $R$  is symplectic, the eigenvalues will always occur in reciprocal pairs, and where the eigenvalues are complex quantities this means that there will always be one mode with positive imaginary part, and one mode with negative imaginary part. Depending on the sign of the imaginary part of the oscillation frequency, the amplitude of the motion will grow (positive imaginary part) or decay (negative imaginary part) exponentially. Since the motion will be, in general, described by some superposition of the normal modes, for  $|A| > 2$  we can expect always to observe an exponential increase in oscillation amplitude: in other words, there will be a beam instability, with threshold given by  $|A| = 2$ . The conclusion is again consistent with the stability condition (15.117). However, from analysis of the normal modes, we also find the growth rate of the instability: this is given by the imaginary part of the normal mode frequency, with larger values of  $|A|$  (for example, from larger bunch charge, stronger wake fields or lower beam energy) corresponding to faster exponential growth.

Figure 15.7 shows the variation of the normal mode frequencies as functions of the parameter  $A$ . Features of the plot can be understood from equations (15.125) and (15.128). In the absence of wake fields ( $W_{\perp}(z) = 0$ , or in the limit of low bunch charge), the frequencies of both normal modes are equal to the nominal betatron frequency:

$$\Omega_+ = \Omega_- = \omega_{\beta}. \quad (15.129)$$

In this case, the modes correspond to both macroparticles oscillating (independently) at the betatron frequency. If the parameter  $A$  is increased (for example, by increasing the bunch charge), the normal mode frequencies shift from the betatron frequency. One mode increases in frequency, while the other mode decreases in frequency. From (15.125):

$$\Omega_{\pm} = \omega_{\beta} \pm \frac{|\omega_s|}{2\pi} \tan^{-1} \left( \frac{A\sqrt{4 - A^2}}{2 - A^2} \right). \quad (15.130)$$

As the value of  $|A|$  approaches 2, the mode frequencies approach values:

$$\Omega_{\pm} = \omega_{\beta} \pm \frac{1}{2} |\omega_s|. \quad (15.131)$$

If  $|A|$  is increased further, the real parts of the mode frequencies  $\Omega_{\pm}$  remain fixed at these values, but the frequency of each mode acquires an imaginary

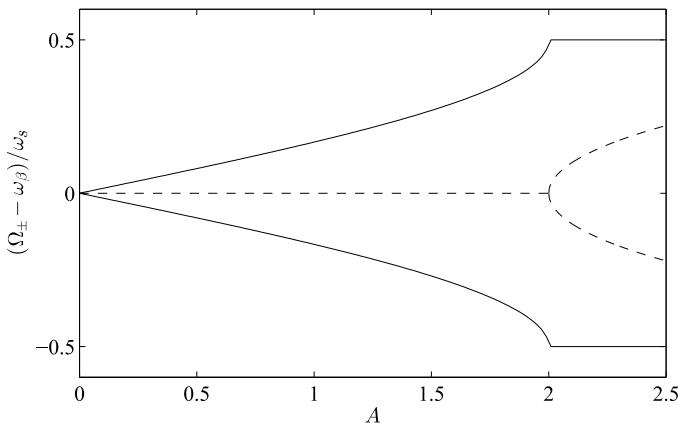


Fig. 15.7 Frequencies of the head-tail modes in a two-macroparticle model, as a function of the parameter  $A$  (15.107) that characterises the transverse wake fields and storage ring parameters. The solid lines show the real parts of the mode frequencies  $\Omega_{\pm}$ , and the dashed lines show the imaginary parts (corresponding to the growth or damping rates of the head-tail oscillation amplitudes). In the absence of wake fields, the frequencies of the head-tail modes are equal to the betatron frequency  $\omega_{\beta}$ . For  $|A| \geq 2$ , the frequencies of the head-tail modes differ from the betatron frequency by half the synchrotron frequency.

part that describes an exponential growth (i.e. instability) or damping of the amplitude of the corresponding mode of oscillation. The instability appears above a particular threshold bunch population, which can be found from the condition  $|A| = 2$ : from (15.117), the threshold bunch population is:

$$N_b = \frac{2|k_z|k_{\beta}E_0C_0}{q^2|W_{\perp}(-\sigma_z)|}. \quad (15.132)$$

Above this threshold, the amplitude of the betatron oscillations can grow very rapidly. For this reason, the instability is known as the *fast head-tail instability* [Kohaupt (1972)].

**Effect of non-zero chromaticity** So far, we have neglected the fact that as a particle in a storage ring performs synchrotron oscillations the energy of the particle changes as well as its longitudinal position within a bunch. In general, the betatron frequency of a particle, and hence the value of  $k_{\beta}$ , depends on its energy. Using (8.35), we can write:

$$k_{\beta} = k_{\beta 0} + \frac{2\pi}{C_0}\xi\delta + O(\delta^2), \quad (15.133)$$

where  $\delta$  is the energy deviation,  $k_{\beta 0}$  is the betatron wave number ( $= \omega_{\beta}/c$ ) for a particle with zero energy deviation,  $C_0$  is the circumference of the

storage ring, and  $\xi$  is the chromaticity. In a synchrotron storage ring, the natural chromaticity (i.e. the chromaticity in a lattice considering only dipole and quadrupole magnets) is always negative, and can be large enough to give a significant spread in the betatron frequencies for particles in the ring. However, the chromaticity is usually corrected using sextupole magnets so that it is zero, or close to zero. There are two reasons for correcting the chromaticity in a storage ring. The first is to reduce the spread of betatron frequencies, so that resonances in tune space can be avoided. The second is that a non-zero chromaticity is associated with an instability related to (but different from) the fast head-tail instability. We shall discuss the instability associated with the chromaticity (known simply as the *head-tail instability* [Pellegrini (1969); Sands (1969)]) in the remainder of this section.

For our analysis of the head-tail instability, it is convenient to represent the transverse co-ordinate and momentum of a particle in normalised phase space using a complex variable:

$$\bar{x} = \frac{x}{\sqrt{\beta_x}} + i\sqrt{\beta_x}p_x. \quad (15.134)$$

In the smooth-focusing (constant beta function) approximation,  $\bar{x}$  is related to the action-angle variables  $J_x$  and  $\phi_x$  by:

$$\bar{x} = \sqrt{2J_x}e^{-i\phi_x}. \quad (15.135)$$

We shall again (as for the fast head-tail instability) use a simple model in which a bunch consists of two macroparticles, each with half the total charge of the bunch. If macroparticle 1 is ahead of macroparticle 2, so that only macroparticle 2 feels the effect of wake fields, then in terms of the complex variables  $\bar{x}_1$  and  $\bar{x}_2$  for macroparticles 1 and 2, the equations of motion (corresponding to (15.101) and (15.102) in our previous analysis) can be approximated by:

$$\frac{d\bar{x}_1}{ds} = -ik_\beta\bar{x}_1, \quad (15.136)$$

$$\frac{d\bar{x}_2}{ds} = -ik_\beta\bar{x}_2 - \frac{1}{2}iAk_z \sin(k_z s)\bar{x}_1, \quad (15.137)$$

where we have used:

$$\frac{d\phi_x}{ds} = k_\beta = \frac{1}{\beta_x}, \quad (15.138)$$

and the parameter  $A$  (characterising the strength of the wake field) is given by (15.107). Equations (15.136) and (15.137) provide approximate (rather than exact) descriptions of the motion since in these equations, the effects

of the forces on the macroparticles are ‘shared’ between the rate of change of  $x$  and the rate of change of  $p_x$ . For the case of zero chromaticity in the preceding discussion, we used a different (but again, approximate) model for the equations of motion (15.101) and (15.102): in those equations, the forces appear in the second-order rate of change of  $x$  (effectively, in the rate of change of  $p_x$ ). But despite the fact that the mathematical description is somewhat different, equations (15.136) and (15.137) lead, in the special case of zero chromaticity, to the same solution as equations (15.101) and (15.102), and the same physical conclusions, as we now show.

For zero chromaticity,  $k_\beta$  is constant, and the solution to (15.136) and (15.137) can be expressed in terms of a transfer matrix:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s+\Delta s} = e^{-ik_\beta \Delta s} \begin{pmatrix} 1 & 0 \\ -\frac{1}{2}iA(1-\cos(k_z \Delta s)) & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_s. \quad (15.139)$$

For one half of a synchrotron period (during which macroparticle 1 is ahead of macroparticle 2), the transfer matrix can be written:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=\pi/|k_z|} = e^{-\pi ik_\beta / |k_z|} \begin{pmatrix} 1 & 0 \\ -iA & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=0}. \quad (15.140)$$

For the second half synchrotron period (during which macroparticle 2 is ahead of macroparticle 1):

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=2\pi/|k_z|} = e^{-\pi ik_\beta / |k_z|} \begin{pmatrix} 1-iA & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=\pi/|k_z|}. \quad (15.141)$$

Multiplying the transfer matrices for the first half and second half of one synchrotron period, we find the transfer matrix for one complete synchrotron period:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=2\pi/|k_z|} = R \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=0}, \quad (15.142)$$

where the matrix  $R$  is given by (15.115). Hence, in the special case of zero chromaticity, we reach the same conclusions regarding the modes and the stability of the system, whether we use the real variables  $(x_1, p_{x1})$  and  $(x_2, p_{x2})$ , or the complex variables  $\bar{x}_1$  and  $\bar{x}_2$ .

Using the complex variables, it is relatively straightforward to include the effects of chromaticity. Let us first consider the equation of motion for macroparticle 1 during the first half synchrotron period, i.e. while macroparticle 1 is ahead of macroparticle 2. The rate of change of the betatron phase is:

$$\frac{d\phi_{x1}}{ds} \approx (1 + \bar{\xi}\delta_1(s))k_{\beta 0}, \quad (15.143)$$

where, for convenience, we define the ‘normalised’ chromaticity:

$$\bar{\xi} = \frac{2\pi}{C_0} \frac{\xi}{k_{\beta 0}} = \frac{\xi}{\nu_0}. \quad (15.144)$$

Here,  $\nu_0$  is the betatron tune (in the absence of wake fields) for a particle with zero energy deviation. Integrating over  $s$ , we find that the phase advance from  $s$  to  $s + \Delta s$  is:

$$\Delta\phi_{x1} = k_{\beta 0}\Delta s + \bar{\xi}k_{\beta 0} \int_s^{s+\Delta s} \delta_1(s) ds. \quad (15.145)$$

The equation of motion for the longitudinal co-ordinate  $z_1(s)$  for macroparticle 1 (as it performs synchrotron oscillations) is:

$$\frac{dz_1}{ds} = -\eta_p \delta_1(s), \quad (15.146)$$

where  $\eta_p$  is the phase slip factor. Using (15.146) in (15.145), we can perform the integral to give:

$$\Delta\phi_{x1} = k_{\beta 0}\Delta s - \frac{\bar{\xi}}{\eta_p} k_{\beta 0} \Delta z_1, \quad (15.147)$$

where:

$$\Delta z_1 = z_1(s + \Delta s) - z_1(s). \quad (15.148)$$

Thus, in a storage ring with chromaticity, the betatron phase advance over any section of the ring can be expressed in terms of the change in the longitudinal co-ordinate, without an explicit dependence on the energy deviation. After one half a synchrotron period starting with  $z_1 = 0$ , the final value of the longitudinal co-ordinate is also  $z_1 = 0$ ; thus, over half a synchrotron period, the betatron phase advance is independent of the chromaticity.

We can therefore include the effects of chromaticity in our two-macroparticle model by writing the equations of motion as follows. During the first half synchrotron period, macroparticle 1 experiences no wake forces. Hence, using (15.135), and (15.147) for the phase advance, we can write over any distance  $\Delta s$ :

$$\bar{x}_1(s + \Delta s) = \exp\left(-ik_{\beta 0}\Delta s + i\frac{\bar{\xi}}{\eta_p} k_{\beta 0} \Delta z_1\right) \bar{x}_1(s). \quad (15.149)$$

Assuming that:

$$\left| \frac{\bar{\xi}}{\eta_p} k_{\beta 0} \Delta z_1 \right| \ll 1, \quad (15.150)$$

we can make the approximation:

$$\bar{x}_1(s + \Delta s) \approx e^{-ik_{\beta 0}\Delta s} \left( 1 + i \frac{\bar{\xi}}{\eta_p} k_{\beta 0} \Delta z_1 \right) \bar{x}_1(s). \quad (15.151)$$

Using this solution to the equation of motion for macroparticle 1, the equation of motion (15.137) for macroparticle 2 during the first half synchrotron period (over which time it experiences wake forces from macroparticle 1) can be written:

$$\frac{d\bar{x}_2}{ds} \approx -ik_{\beta 0}\bar{x}_2 - \frac{1}{2}iAk_z \sin(k_z s) e^{-ik_{\beta 0}s} \left( 1 + i \frac{\bar{\xi}}{\eta_p} k_{\beta 0} z_1(s) \right) \bar{x}_1(0). \quad (15.152)$$

Note that in the first term on the right-hand side of (15.152), we neglect any effect of the chromaticity. This is justified by the fact that this term gives the usual betatron oscillations, and we know from (15.147) that after half a synchrotron period, starting with  $z_2 = 0$ , the betatron phase advance is independent of the chromaticity. In other words, the effect of the chromaticity on the usual betatron oscillations averages to zero. The second term on the right-hand side of (15.152), however, includes a factor  $\sin(k_z s)$ , which arises from the variation in the longitudinal distance between the two macroparticles. The presence of this factor means that in this term, the effects of the chromaticity will not average to zero over one half synchrotron period.

Using:

$$z_1(s) = \sigma_z \sin(|k_z|s), \quad (15.153)$$

we find that (15.152) has solution, for  $0 < s < \pi/|k_z|$ :

$$x_2(s) = e^{-ik_{\beta 0}s} x_2(0) - \frac{1}{2}iA \left( 1 - \cos(k_z s) + i \frac{\bar{\xi}}{2\eta_p} k_{\beta 0} \sigma_z \left( |k_z|s - \frac{1}{2} \sin(2|k_z|s) \right) \right) e^{-ik_{\beta 0}s} x_1(0). \quad (15.154)$$

Putting  $s = \pi/|k_z|$ , the phase space co-ordinates for macroparticles 1 and 2 after half a synchrotron period can be written in terms of a transfer matrix:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=\pi/|k_z|} = e^{-\pi ik_{\beta 0}/|k_z|} \begin{pmatrix} 1 & 0 \\ -iA\xi & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=0}, \quad (15.155)$$

where:

$$A_\xi = \left( 1 + i \frac{\pi}{4} \frac{\bar{\xi}}{\eta_p} k_{\beta 0} \sigma_z \right) A. \quad (15.156)$$

For the second half synchrotron period (during which macroparticle 2 is ahead of macroparticle 1), the particle co-ordinates transform as:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=2\pi/|k_z|} = e^{-\pi i k_{\beta 0}/|k_z|} \begin{pmatrix} 1 & -iA_\xi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=\pi/|k_z|}. \quad (15.157)$$

The transformation for a full synchrotron period is then:

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=2\pi/|k_z|} = R_\xi \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix}_{s=0}, \quad (15.158)$$

where the transfer matrix  $R_\xi$  is given by:

$$R_\xi = e^{-2\pi i k_{\beta 0}/|k_z|} \begin{pmatrix} 1 - A_\xi^2 & -iA_\xi \\ -iA_\xi & 1 \end{pmatrix}. \quad (15.159)$$

Equation (15.158) is the same as (15.142), except that  $A$  is replaced by  $A_\xi$ . This has implications for the behaviour of the normal modes of oscillation of the macroparticles, compared to the case of zero chromaticity. We can (as before) obtain the frequencies of the normal modes, and the growth or damping rates of the oscillation amplitudes, from the eigenvalues of the transfer matrix. In the present case, the imaginary part of  $A_\xi$  (when the chromaticity  $\xi$  is non-zero) means that the eigenvalues of the matrix  $R_\xi$  do not have modulus 1. For small  $A_\xi$ :

$$\text{eigenvalues}(R_\xi) \approx e^{-2\pi i \frac{k_{\beta 0}}{|k_z|}} e^{\pm i A_\xi} = e^{-2\pi i \frac{k_{\beta 0}}{|k_z|} \pm i \text{Re}(A_\xi)} e^{\mp \text{Im}(A_\xi)}. \quad (15.160)$$

The mode corresponding to the upper sign in (15.160) will be damped if the imaginary part of  $A_\xi$  is positive, i.e. if  $\xi/\eta_p > 0$ , and antidamped if  $\xi/\eta_p < 0$ . Conversely, the mode corresponding to the lower sign in (15.160) is damped if  $\xi/\eta_p < 0$ , and antidamped if  $\xi/\eta_p > 0$ . If the chromaticity is not exactly zero, there is always one damped mode and one antidamped mode: which mode is damped depends on the sign of the chromaticity, and whether the storage ring is above or below transition. For example, above transition ( $\eta_p > 0$ ), the mode corresponding to the upper sign in (15.160) is damped if the chromaticity is positive, and antidamped if the chromaticity is negative. The growth or damping rate is given by the imaginary part of  $A_\xi$  multiplied by the synchrotron frequency.

It appears from this analysis that any non-zero value of the chromaticity will lead to an instability. The instability in this case is known as the *head-tail instability*; in contrast to the *fast head-tail instability* discussed above, the head-tail instability does not have a threshold, but occurs for any current if the chromaticity (and transverse wake fields) are non-zero.

Fortunately, in practice the chromaticity in a storage ring does not have to be exactly zero in order to mitigate the effects of the head-tail instability, and it is usually possible, by use of sextupoles in the lattice, to reduce the chromaticity sufficiently for effects such as synchrotron radiation damping and decoherence to prevent head-tail oscillations reaching large amplitudes. Furthermore, the simple two-particle model we have developed in this section tends to over-estimate the instability growth rates. A more sophisticated model based on the Vlasov equation (treating the charge within a bunch as a continuous distribution) gives a more accurate estimate of the head-tail growth rates. It is also found that there is some asymmetry in the growth rates for positive and negative chromaticity: for storage rings operating below transition, it can be beneficial to operate with slightly negative chromaticity; above transition, it can be better to operate with slightly positive chromaticity. We discuss the Vlasov equation, and its application for analysis of longitudinal instabilities, in the following sections. A treatment of transverse instabilities using the Vlasov equation can be found in [Chao (1993b)].

#### **15.4.2 Sacherer's integral equation**

In Section 15.3 we discussed the microwave instability, in which a small density modulation in a beam in a storage ring could grow exponentially, driven by the impedance in the ring. Our model was based on an unbunched beam, in which there were no synchrotron oscillations, and the only change in the energy of a particle came from the wake fields induced in the vacuum chamber by the beam. Using the Vlasov equation to describe the evolution of the particle distribution in phase space, we derived the conditions under which the amplitude of a small density modulation could grow exponentially. Strictly speaking, the criteria for beam stability following from this model should only be applied in the case of a coasting beam. Nevertheless, if the wavelength of a density modulation is small compared with the length of a bunch, and if the growth rate of the instability is large compared with the synchrotron frequency, then it is possible to generalise the stability condition to cases where the beam is bunched.

However, beam instabilities in bunched beams do occur where the beam cannot be treated as a coasting beam. In particular, density perturbations on the scale of the bunch length can be driven to large amplitudes by the machine impedance. To describe such instabilities, and in particular to understand the conditions under which they occur, we need to revisit our

analysis, making somewhat different initial assumptions for the basic beam behaviour. In this section, and in Sections 15.4.3 and 15.4.4, our goal will be to perform an analysis similar to that carried out in Section 15.3, but using a model in which the beam is bunched (and particles perform synchrotron oscillations) rather than a model in which the unperturbed distribution is uniform around the entire ring. The evolution of the charge distribution is still described by the Vlasov equation, so the underlying physics is fundamentally the same. We again describe the charge distribution in terms of a prescribed density distribution (corresponding to the equilibrium distribution in the absence of the impedance) on which we superpose a small modulation. If the modulation has an appropriate shape, then it oscillates at a well-defined frequency. The first step in the analysis is to identify the modulations that have well-defined frequencies: these modulations we refer to as *modes*. The next step is to identify the frequency of each mode. As usual, the imaginary part of the frequency indicates whether a mode is damped or antidamped. If any modes are antidamped, then the beam can become unstable. We shall find that the mode frequencies depend on the beam current; as the current is increased, the real parts of the frequencies of two different modes can approach the same value. An instability is associated with the real parts of the frequencies of two different modes becoming equal; this type of instability is therefore known as a *mode-coupling instability*.

In this section, our goal is to derive Sacherer's integral equation [Sacherer (1972)], starting from the Vlasov equation. Sacherer's integral equation plays a similar role to the dispersion relation (15.78), in relating the frequency of a perturbation to the wavelength of the density modulation that constitutes the perturbation. Here, however, we take account of the fact that the beam is bunched (with particles performing synchrotron oscillations) from the start, rather than attempting to generalise results obtained for a coasting beam.

We shall use canonical longitudinal phase space variables  $z$  and  $\delta$ , where  $z$  is the co-ordinate representing the longitudinal position of a particle relative to a reference particle in a storage ring, and  $\delta$  is the conjugate momentum, representing the energy deviation. In these variables, the longitudinal phase space density  $\Psi(z, \delta; s)$  evolves according the Vlasov equation:

$$\frac{\partial \Psi}{\partial s} + \frac{dz}{ds} \frac{\partial \Psi}{\partial z} + \frac{d\delta}{ds} \frac{\partial \Psi}{\partial \delta} = 0, \quad (15.161)$$

where  $s$  is the independent variable, corresponding to the distance along the reference trajectory in the storage ring. Including the phase slip, rf focusing

and wake fields, the dynamical variables  $z$  and  $\delta$  of a single particle evolve according to the equations of motion:

$$\frac{dz}{ds} = -\eta_p \delta, \quad (15.162)$$

$$\frac{d\delta}{ds} = \frac{\omega_z^2}{\eta_p c^2} z + \frac{qV(z; s)}{E_0 C_0}, \quad (15.163)$$

where  $\eta_p$  is the phase slip factor,  $\omega_z$  is the synchrotron frequency,  $E_0$  is the reference energy, and  $C_0$  is the circumference of the ring (i.e. the total length of the reference trajectory). The synchrotron frequency is related to the synchrotron phase  $\phi_z$  by:

$$\omega_z = \beta_0 c \frac{d\phi_z}{ds}, \quad (15.164)$$

where  $\beta_0 c$  is the speed of the reference particle. We shall take the ultra-relativistic limit  $\beta_0 \rightarrow 1$ . With our usual definition for the synchrotron phase  $\phi_z$  (as given in (5.64) and (5.65)),  $\omega_z < 0$  if  $\eta_p > 0$ , i.e. the synchrotron frequency is negative in a storage ring operating above transition.

The potential  $V(z; s)$  is defined by:

$$V(z; s) = - \int_z^\infty \lambda(z'; s) W_{\parallel}(z - z') dz', \quad (15.165)$$

where  $W_{\parallel}(z)$  is the longitudinal wake function, and  $\lambda(z)$  is the longitudinal charge distribution, given by:

$$\lambda(z; s) = \int_{-\infty}^\infty \Psi(z, \delta; s) d\delta. \quad (15.166)$$

Let us assume that there is a ‘zero-frequency’ mode, i.e. a solution to the Vlasov equation such that the distribution function  $\Psi(z, \delta; s)$  has no explicit dependence on the independent variable  $s$ :

$$\Psi(z, \delta; s) = \Psi_0(z, \delta). \quad (15.167)$$

Using  $\partial\Psi_0/\partial s = 0$ , and (15.162) and (15.163), the Vlasov equation (15.161) leads to:

$$-\eta_p \delta \frac{\partial \Psi_0}{\partial z} + \frac{\omega_z^2}{\eta_p c^2} z \frac{\partial \Psi_0}{\partial \delta} + \frac{qV(z)}{E_0 C_0} \frac{\partial \Psi_0}{\partial \delta} = 0. \quad (15.168)$$

Equation (15.168) is a more general form of the Haissinski equation (15.62). Let us assume for the moment that the beam has a Gaussian energy spread, and that there is no correlation between the longitudinal co-ordinate and

the energy deviation. Then, the distribution function  $\Psi_0$  can be written in the form:

$$\Psi_0(z, \delta) = \frac{e^{-\delta^2/2\sigma_\delta^2}}{\sqrt{2\pi}\sigma_\delta} \lambda_0(z). \quad (15.169)$$

Substituting for  $\Psi_0$  from (15.169) into (15.168) gives:

$$\frac{d\lambda_0(z)}{dz} = -\frac{\omega_z^2}{\eta_p^2 c^2 \sigma_\delta^2} z \lambda_0(z) - \frac{qV(z)}{\eta_p \sigma_\delta^2 E_0 C_0} \lambda_0(z). \quad (15.170)$$

Equation (15.62), which describes the effect of potential-well distortion, can be seen to provide a solution to (15.170).

Now let us return to the more general case (i.e. where we do not assume a particular form for the energy spread). Consider a perturbation to the stationary solution, so that the phase space density  $\Psi(z, \delta; s)$  can be written:

$$\Psi(z, \delta; s) = \Psi_0(z, \delta) + \Delta\Psi(z, \delta) e^{-i\Omega s/c}, \quad (15.171)$$

for small  $\Delta\Psi$ . If  $\Psi_0(z, \delta)$  satisfies the Vlasov equation in the limit  $\Delta\Psi(z, \delta) \rightarrow 0$  (i.e. we take into account the effects of potential-well distortion), then substitution of (15.171) into the Vlasov equation (15.161), and keeping terms up to first order in the perturbation leads to:

$$-i\frac{\Omega}{c} \Delta\Psi - \eta_p \delta \frac{\partial \Delta\Psi}{\partial z} + \frac{\omega_z^2}{\eta_p c^2} z \frac{\partial \Delta\Psi}{\partial \delta} + \frac{q\Delta V(z)}{E_0 C_0} \frac{\partial \Psi_0}{\partial \delta} + \frac{qV_0(z)}{E_0 C_0} \frac{\partial \Delta\Psi}{\partial \delta} = 0, \quad (15.172)$$

where:

$$V_0(z) = - \int_z^\infty \lambda_0(z') W_{||}(z - z') dz', \quad (15.173)$$

$$\Delta V(z) = - \int_z^\infty \Delta\lambda(z') W_{||}(z - z') dz', \quad (15.174)$$

and:

$$\Delta\lambda(z') = \int_{-\infty}^\infty \Delta\Psi(z, \delta) d\delta. \quad (15.175)$$

The frequency  $\Omega$  is the oscillation frequency of a perturbation of the phase-space charge density, with a given dependence on  $z$  and  $\delta$ . For a given distribution  $\Psi_0$  there are only particular perturbations  $\Delta\Psi(z, \delta)$  that solve the differential equation (15.172): our goal is to find these perturbations, and the corresponding frequencies  $\Omega$ . If there are frequencies for which the imaginary part is positive, then the corresponding perturbations will be unstable (i.e. will grow exponentially in time). In order to find the solutions to (15.172) we first manipulate it into a more manageable form: in

the remainder of this section, we show how it leads to an integral equation, known as Sacherer's integral equation. Then, we discuss solutions to Sacherer's integral equation in Sections 15.4.3 and 15.4.4.

In equation (15.172), there are two terms in  $\partial\Delta\Psi/\partial\delta$ . Depending on the wake function and the perturbation, the potential  $V(z)$  can be approximated as a linear function of  $z$ ; then the two terms in  $\partial\Delta\Psi/\partial\delta$  can be combined by defining  $\bar{\omega}_z$  such that:

$$\frac{\bar{\omega}_z^2}{\eta_p c^2} = \frac{\omega_z^2}{\eta_p c^2} + \frac{q}{E_0 C_0} \left. \frac{dV_0(z)}{dz} \right|_{z=0}. \quad (15.176)$$

It is also convenient to introduce the longitudinal action-angle variables  $J_z$  and  $\phi_z$  defined, as in (5.64) and (5.65), so that:

$$z = \sqrt{2\beta_z J_z} \cos(\phi_z), \quad (15.177)$$

$$\delta = -\sqrt{\frac{2J_z}{\beta_z}} \sin(\phi_z). \quad (15.178)$$

$\beta_z$  is the longitudinal Courant-Snyder beta function, and we assume that the longitudinal Courant-Snyder alpha function  $\alpha_z = 0$ , i.e.  $\beta_z$  is approximately constant: this will be a valid approximation if the longitudinal focusing is not too strong. The value of  $\beta_z$  is determined by the shape of the ellipse traced by a single particle in longitudinal phase space. In the absence of wake fields, the longitudinal equations of motion (15.162) and (15.163) can be derived from the Hamiltonian:

$$H = -\frac{\eta_p}{2} \delta^2 - \frac{\omega_z^2}{2\eta_p c^2} z^2. \quad (15.179)$$

Since the Hamiltonian is a constant of the motion, it follows that:

$$\beta_z = \frac{z_{\max}}{\delta_{\max}} = -\frac{\eta_p c}{\omega_z}. \quad (15.180)$$

The minus sign comes from the fact that  $\eta_p/\omega_z < 0$ . In the presence of wake fields, we replace  $\omega_z$  in (15.180) by  $\bar{\omega}_z$ , given by (15.176).

Using (15.176), (15.177), (15.178) and (15.180), equation (15.172) becomes:

$$-i\frac{\Omega}{c}\Delta\Psi + \frac{\bar{\omega}_z}{c} \frac{\partial\Delta\Psi}{\partial\phi_z} + \frac{q\Delta V(z)}{E_0 C_0} \frac{\partial\Psi_0}{\partial\delta} = 0. \quad (15.181)$$

Since  $\Psi_0$  is an invariant distribution, it is a function of  $J_z$  only, i.e. it has no explicit dependence on  $\phi_z$ . Hence, we can write for the derivative of  $\Psi_0$  with respect to  $\delta$ :

$$\frac{\partial\Psi_0}{\partial\delta} = \frac{\partial J_z}{\partial\delta} \frac{\partial\Psi_0}{\partial J_z} = -\sqrt{2\beta_z J_z} \sin(\phi_z) \frac{\partial\Psi_0}{\partial J_z}. \quad (15.182)$$

Equations (15.173) and (15.174) are valid in a storage ring only if the wake fields decay completely within one revolution period. It is possible in some cases for the wake fields to persist for longer than a revolution period, in which case we need to take account of the wake fields generated by a bunch on previous turns of the storage ring. This can be done by including a summation over an index  $k$ , representing the turn number:

$$\Delta V(z) = - \sum_{k=-\infty}^{\infty} \int_z^{z+C_0} \Delta\lambda(z') e^{i\Omega k C_0/c} W_{\parallel}(z - z' - kC_0) dz'. \quad (15.183)$$

Note that the factor  $e^{i\Omega k C_0/c}$  accounts for the phase of the perturbation on the phase space density on different turns. It is convenient to express  $\Delta V(z)$  in terms of the impedance rather than the wake function. This can be done using:

$$\Delta\lambda(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Delta\tilde{\lambda}(\omega) e^{i\omega z/c} d\omega, \quad (15.184)$$

$$W_{\parallel}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z_{\parallel}(\omega) e^{i\omega z/c} d\omega. \quad (15.185)$$

Substituting into (15.183) gives:

$$\Delta V(z) = - \frac{c}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} \Delta\tilde{\lambda}(\omega) Z_{\parallel}(\omega) e^{i\omega z/c} e^{2\pi i k (\Omega - \omega)/\omega_0} d\omega, \quad (15.186)$$

where  $\omega_0 = 2\pi c/C_0$ . Then, using (15.11) gives:

$$\Delta V(z) = - \frac{c^2}{C_0} \sum_{p=-\infty}^{\infty} \Delta\tilde{\lambda}(\Omega_p) Z_{\parallel}(\Omega_p) e^{i\Omega_p z/c}, \quad (15.187)$$

where:

$$\Omega_p = \Omega + p\omega_0. \quad (15.188)$$

In terms of action-angle variables:

$$\Delta V = \Delta V(\phi_z, J_z) = - \frac{c^2}{C_0} \sum_{p=-\infty}^{\infty} \Delta\tilde{\lambda}(\Omega_p) Z_{\parallel}(\Omega_p) e^{i\Omega_p \sqrt{2\beta_z J_z} \cos(\phi_z)/c}. \quad (15.189)$$

Equation (15.181) describing the evolution of the perturbation  $\Delta\Psi$  can now be written:

$$-i\Omega\Delta\Psi + \bar{\omega}_z \frac{\partial\Delta\Psi}{\partial\phi_z} - \frac{qc}{E_0 C_0} \sqrt{2\beta_z J_z} \sin(\phi_z) \Delta V \frac{\partial\Psi_0}{\partial J_z} = 0. \quad (15.190)$$

This equation is not easy to solve in general. To proceed, we expand the perturbation in a set of Fourier modes in the angle variable  $\phi_z$  (which is possible since  $\Delta\Psi$  must be periodic in  $\phi_z$ , with period  $2\pi$ ):

$$\Delta\Psi(\phi_z, J_z) = \sum_{\ell'=-\infty}^{\infty} R_{\ell'}(J_z) e^{i\ell'\phi_z}. \quad (15.191)$$

Substituting into (15.181) gives:

$$i \sum_{\ell'=-\infty}^{\infty} R_{\ell'}(J_z) e^{i\ell'\phi_z} (\Omega - \ell' \bar{\omega}_z) + \frac{qc}{E_0 C_0} \sqrt{2\beta_z J_z} \sin(\phi_z) \Delta V \frac{\partial \Psi_0}{\partial J_z} = 0. \quad (15.192)$$

To eliminate the summation over  $\ell'$ , we multiply by  $e^{-i\ell\phi_z}/2\pi$ , and integrate over  $\phi_z$ . Taking into account the dependence of  $\Delta V$  on  $\phi_z$ , we need the result:

$$\frac{1}{2\pi} \int_0^{2\pi} e^{-i\ell\phi_z} \sin(\phi_z) e^{iu \cos(\phi_z)} d\phi_z = -i\ell \frac{\ell}{u} J_\ell(u), \quad (15.193)$$

where  $J_\ell(u)$  is a Bessel function. After performing the integral over  $\phi_z$ , (15.192) becomes:

$$iR_\ell(J_z) (\Omega - \ell \bar{\omega}_z) + i\ell \frac{q\ell c^4}{E_0 C_0^2} \frac{\partial \Psi_0}{\partial J_z} \sum_{p=-\infty}^{\infty} \Delta \tilde{\lambda}(\Omega_p) \frac{Z_{||}(\Omega_p)}{\Omega_p} J_\ell\left(\frac{\Omega_p \hat{z}}{c}\right) = 0, \quad (15.194)$$

where, for convenience, we define the amplitude  $\hat{z}$  of synchrotron oscillations of a particle with synchrotron action  $J_z$ :

$$\hat{z} = \sqrt{2\beta_z J_z}. \quad (15.195)$$

The next step is to express the beam current spectrum  $\Delta \tilde{\lambda}(\omega)$  in terms of the phase space density perturbation  $\Delta\Psi$  (15.191). Using (15.175) and (15.184), we can write:

$$\Delta \tilde{\lambda}(\omega) = \int_{-\infty}^{\infty} \frac{dz}{c} \int_{-\infty}^{\infty} d\delta e^{-i\omega z/c} \Delta\Psi. \quad (15.196)$$

Since  $(z, \delta)$  and  $(\phi_z, J_z)$  both form canonical pairs of phase space variables, the Jacobian of the transformation from one pair of variables to another is equal to unity. Hence, the integral is easily converted to an integral over  $\phi_z$  and  $J_z$ . Then, substituting for  $\Delta\Psi$  from (15.191) gives:

$$\Delta \tilde{\lambda}(\omega) = \frac{1}{c} \int_0^{2\pi} d\phi_z \int_0^{\infty} dJ_z e^{-i\omega z/c} \sum_{\ell'=-\infty}^{\infty} R_{\ell'}(J_z) e^{i\ell'\phi_z}. \quad (15.197)$$

The integral over  $\phi_z$  can be performed with the help of the formula:

$$\frac{1}{2\pi} \int_0^{2\pi} e^{i\ell'\phi_z} e^{-iu \cos(\phi_z)} d\phi_z = i^{-\ell'} J_{\ell'}(u), \quad (15.198)$$

giving the result:

$$\Delta\tilde{\lambda}(\omega) = \frac{2\pi i^{-\ell'}}{c} \int_0^\infty dJ_z \sum_{\ell'=-\infty}^{\infty} R_{\ell'}(J_z) J_{\ell'}\left(\frac{\omega\hat{z}}{c}\right). \quad (15.199)$$

Then, substituting for  $\Delta\tilde{\lambda}(\Omega_p)$  in (15.194) gives:

$$\begin{aligned} R_\ell(J_z) (\Omega - \ell\bar{\omega}_z) &= i^{1+\ell-\ell'} \frac{qc\omega_0^2\ell}{2\pi E_0} \frac{\partial\Psi_0}{\partial J_z} \\ &\times \sum_{\ell',p=-\infty}^{\infty} \int_0^\infty R_{\ell'}(J'_z) \frac{Z_{||}(\Omega_p)}{\Omega_p} J_\ell\left(\frac{\Omega_p\hat{z}}{c}\right) J_{\ell'}\left(\frac{\Omega_p\hat{z}'}{c}\right) dJ'_z. \end{aligned} \quad (15.200)$$

Equation (15.200) may appear rather intimidating; however, the advantage of writing the equation for the mode frequencies  $\Omega$  in this way is that the problem becomes one of finding an eigensystem. This becomes clear after some further manipulation. First, we write (15.200) in the form:

$$\left(\frac{\Omega}{\bar{\omega}_z} - \ell\right) R_\ell(J_z) = \sum_{\ell'=-\infty}^{\infty} W(J_z) \int_0^\infty G_{\ell\ell'}(J_z, J'_z) R_{\ell'}(J'_z) dJ'_z, \quad (15.201)$$

where we define the weight function (with dimensions of  $1/\text{length}^3$ ):

$$W(J_z) = -\frac{1}{qN_b\beta_z} \frac{\partial\Psi_0}{\partial J_z}, \quad (15.202)$$

and the kernel (with dimensions of  $\text{length}^2$ ):

$$G_{\ell\ell'}(J_z, J'_z) = i^{1+\ell-\ell'} \frac{q^2 N_b c^2 \omega_0^2 \eta_p \ell}{2\pi E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \frac{Z_{||}(\Omega_p)}{\Omega_p} J_\ell\left(\frac{\Omega_p\hat{z}}{c}\right) J_{\ell'}\left(\frac{\Omega_p\hat{z}'}{c}\right). \quad (15.203)$$

Here,  $\beta_z$  is the longitudinal Courant–Snyder beta function (15.180), and  $\eta_p$  is the phase slip factor. The minus sign is introduced in the definition of the weight function (15.202) so that in the common situation, where the particle density in phase space decreases with increasing longitudinal action, the value of the weight function is positive.

If the machine parameters, impedance and charge distribution in longitudinal phase space are known, then the weight function and kernel can both be found (and expressed as functions of  $J_z$ ). Then, (15.201) provides

an integral equation for the functions  $R_\ell(J_z)$  and the mode frequencies  $\Omega$ . Equation (15.201) is known as *Sacherer's integral equation* [Sacherer (1972)]. The right-hand side of (15.201) is both an integral operator (acting on the functions  $R_\ell(J_z)$ ) and a matrix multiplication (acting on a vector with components given by the functions  $R_\ell(J_z)$ ).

To complete the reduction of (15.201) to an eigenvalue problem, let us assume that the functions  $R_\ell(J_z)$  may be expanded in terms of a set of functions  $f_k(J_z)$ :

$$R_\ell(J_z) = W(J_z) \sum_{k=0}^{\infty} a_{\ell k} f_k(J_z), \quad (15.204)$$

where the functions  $f_k(J_z)$  are defined by the orthogonality condition:

$$\int_0^\infty W(J_z) f_k(J_z) f_{k'}(J_z) dJ_z = \delta_{kk'}. \quad (15.205)$$

We substitute for  $R_\ell(J_z)$  (and  $R_{\ell'}(J'_z)$ ) from (15.204) in (15.201); multiply both sides by  $f_{k'}(J_z)$ , and integrate over  $J_z$ . After applying the orthogonality condition (15.205) (and relabelling  $k$  and  $k'$  to  $k'$  and  $k$ , respectively), we obtain:

$$\frac{\Omega}{\bar{\omega}_z} a_{\ell k} = \sum_{\ell'=-\infty}^{\infty} \sum_{k'=0}^{\infty} M_{\ell k \ell' k'} a_{\ell' k'}, \quad (15.206)$$

where  $M_{\ell k \ell' k'}$  is the *interaction matrix*, given by:

$$\begin{aligned} M_{\ell k \ell' k'} &= \ell \delta_{\ell \ell'} \delta_{k k'} \\ &+ \int_0^\infty dJ_z \int_0^\infty dJ'_z G_{\ell \ell'}(J_z, J'_z) W(J_z) f_k(J_z) W(J'_z) f_{k'}(J'_z). \end{aligned} \quad (15.207)$$

Equation (15.206) states that the perturbation frequencies  $\Omega$  may be obtained from the eigenvalues of a four-dimensional matrix; the eigenvector corresponding to a particular eigenvalue (frequency) gives the form of the perturbation that oscillates at that frequency. We cannot directly solve the eigensystem equation, because (15.206) is written in terms of a four-dimensional matrix. However, we shall show in Section 15.4.4 that, simply by relabelling the indices, the equation can be reduced to an eigensystem equation for a two-dimensional matrix. First, though, we note that the system is relatively easy to solve in the case that  $M_{\ell k \ell' k'}$  is zero (or close to zero) unless  $\ell = \ell'$ : this corresponds to the situation that the azimuthal modes (indexed by  $\ell$ ) are independent of each other. We shall discuss this case in Section 15.4.3, and return to the more general case, where there is some coupling between modes, in Section 15.4.4.

### 15.4.3 Discrete modes: Robinson instability

Sacherer's integral equation provides an equation for the frequencies and phase space structure of perturbations to the longitudinal phase space distribution of particles within a bunch. The imaginary part of the frequency for a particular perturbation will tell us whether that perturbation will be damped or antidamped. While it is difficult to find a general solution to Sacherer's integral equation, we can consider particular cases that are of practical interest in accelerators.

Let us first consider the case that the only non-zero components in the interaction matrix  $M_{\ell k \ell' k'}$  (15.207) are those with  $\ell' = \ell$ . In that case, (15.206) becomes:

$$\frac{\Omega}{\bar{\omega}_z} a_{\ell k} = \sum_{k'=0}^{\infty} M_{\ell k \ell k'} a_{\ell k'}, \quad (15.208)$$

which we can write in matrix form:

$$\frac{\Omega}{\bar{\omega}_z} \mathbf{a}^{(\ell)} = M^{(\ell)} \mathbf{a}^{(\ell)}. \quad (15.209)$$

$M^{(\ell)}$  is a matrix with elements  $M_{\ell k \ell k'}$ , and  $\mathbf{a}^{(\ell)}$  is a vector with elements  $a_{\ell k'}$ . Modes with different values of  $\ell$  can be considered independently: this greatly simplifies the problem, but will not cover all cases of interest. Nevertheless, let us proceed to investigate the mode frequencies (and structure) in this case.

First, we note from (15.207) and (15.203) that the matrix  $M^{(\ell)}$  has elements:

$$M_{kk'}^{(\ell)} = \ell \delta_{kk'} + i \frac{q^2 N_b c^2 \omega_0^2 \eta_p \ell}{2\pi E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \frac{Z_{\parallel}(\Omega_p)}{\Omega_p} g_{\ell k}(\Omega_p) g_{\ell k'}(\Omega_p), \quad (15.210)$$

where:

$$g_{\ell k}(\Omega_p) = \int_0^{\infty} W(J_z) J_{\ell} \left( \frac{\Omega_p \hat{z}}{c} \right) f_k(J_z) dJ_z. \quad (15.211)$$

Calculation of the mode frequencies  $\Omega$  proceeds as follows. For a given unperturbed distribution, the weight function  $W(J_z)$  is calculated from (15.202). The functions  $f_k(J_z)$  are then determined by (15.205), and the functions  $g_{\ell k}(\Omega_p)$  are found from (15.211). Then, for a given impedance and mode number  $\ell$ , the matrix  $M^{(\ell)}$  is computed from (15.210). The eigenvalues of  $M^{(\ell)}$  are equal to  $\Omega/\bar{\omega}_z$ . For a given mode number  $\ell$ , there are a number of different frequencies  $\Omega$  corresponding to different eigenvalues:

each frequency is associated with a different dependence of the perturbation on  $J_z$ , i.e. with a different function  $R_\ell(J_z)$ . From (15.204), we see that the different  $R_\ell(J_z)$  (which can be considered as different radial modes in longitudinal phase space) are determined by the eigenvectors of  $M^{(\ell)}$ . For any given  $\ell$ , the matrix  $M^{(\ell)}$  usually has an infinite number of elements: to calculate the frequency of a particular mode, it is necessary to truncate the matrix. The eigenvalues of the truncated matrix will not be exactly the same as the eigenvalues of the full matrix  $M^{(\ell)}$ ; however, if there are a sufficient number of elements retained in the truncated matrix, the eigenvalues of the truncated matrix should give a reasonable approximation for the frequencies of at least some of the modes.

Some statements about beam stability can be made for certain cases simply from the form of (15.210), without solving for the mode frequencies explicitly. For example, if the impedance is purely imaginary, then  $M^{(\ell)}$  is a real symmetric matrix: the eigenvalues are then purely real, and beam instability cannot occur. This seems similar to the result we found in Section 15.3, where we concluded (following an analysis of a coasting beam, based on the Vlasov equation) that with a purely imaginary impedance the beam was always stable, depending on the sign of the impedance. In the present case, however, with a purely imaginary impedance, the beam is always stable irrespective of the sign of the impedance. The inconsistency comes from the fact that in the present analysis we are neglecting any possibility of coupling between perturbations with different mode numbers  $\ell$ . This will be rectified in Section 15.4.4.

Another observation from (15.210) is that for short bunches a beam instability is most likely to occur if the impedance has a sharp peak at some particular frequency; for example, in the case of a resonator with high quality factor. This can be shown as follows. The short bunch condition can be written explicitly as:

$$\frac{\omega_{\max}\sigma_z}{c} \ll 1, \quad (15.212)$$

where the impedance extends up to a frequency  $\omega_{\max}$ , and  $\sigma_z$  is the bunch length. When this condition is satisfied, we can expand the Bessel function in (15.211) to lowest order in its argument:

$$J_\ell\left(\frac{\Omega_p \hat{z}}{c}\right) \approx \frac{1}{2^\ell \ell!} \left(\frac{\Omega_p \hat{z}}{c}\right)^\ell. \quad (15.213)$$

Then,  $g_{\ell k}(\Omega_p)g_{\ell k'}(\Omega_p)$  is an even function of  $\Omega_p$ . For a broad-band impedance, the summation over  $p$  in (15.210) can be replaced by an integral; then, it follows from (14.188) that  $M^{(\ell)}$  is a real symmetric matrix.

Therefore, the eigenvalues of  $M^{(\ell)}$  must be real: the mode frequencies  $\Omega$  have zero imaginary part, and the beam must be stable. Beam instability can only occur for a sharply peaked impedance (as we shall discuss in the remainder of this section), or for long bunches, or when coupling occurs between different mode numbers  $\ell$  (as we shall discuss in Section 15.4.4).

As an illustration of the procedure we have described for finding the frequencies of the different modes, let us consider the particular case of a beam with a ‘water-bag’ distribution in longitudinal phase space, in a storage ring where the impedance can be modelled as a sharply peaked (high quality factor) resonator. A water-bag distribution is uniform up to some value  $J_{z0}$  of the action, and then is zero if the action is larger than  $J_{z0}$ :

$$\Psi_0(J_z) = \begin{cases} qN_b/2\pi J_{z0} & \text{if } J_z \leq J_{z0}, \\ 0 & \text{if } J_z > J_{z0}. \end{cases} \quad (15.214)$$

Note that the distribution is normalised so that:

$$\int_0^{2\pi} d\phi_z \int_0^\infty dJ_z \Psi_0(J_z) = qN_b. \quad (15.215)$$

The distribution will depend on the bunch charge because of the impedance in the storage ring; we assume that the distribution (15.214) includes the effects of potential-well distortion. The derivative of the distribution with respect to  $J_z$  is a delta function:

$$\frac{\partial \Psi_0}{\partial J_z} = -\frac{qN_b}{2\pi J_{z0}} \delta_D(J_z - J_{z0}). \quad (15.216)$$

Hence, using (15.202), the weight function in this case is:

$$W(J_z) = \frac{\delta_D(J_z - J_{z0})}{2\pi\beta_z J_{z0}}. \quad (15.217)$$

Since the weight function is a delta function, the only perturbation we can express in the form (15.204) is a perturbation around the edge of the distribution in longitudinal phase space. Thus, although we are allowed any number of azimuthal modes, we only have one radial mode. As a consequence, the set of orthonormal functions  $f_k(J_z)$  has only a single member:

$$f_0(J_z) = \sqrt{2\pi\beta_z J_{z0}}. \quad (15.218)$$

Then, from (15.211):

$$g_{\ell 0}(\Omega_p) = \int_0^\infty \frac{\delta_D(J_z - J_{z0})}{\sqrt{2\pi\beta_z J_{z0}}} J_\ell\left(\frac{\Omega_p \hat{z}}{c}\right) dJ_z = \frac{1}{\sqrt{2\pi\beta_z J_{z0}}} J_\ell\left(\frac{\Omega_p \hat{z}_0}{c}\right), \quad (15.219)$$

where:

$$\hat{z}_0 = \sqrt{2\beta_z J_{z0}}. \quad (15.220)$$

If the dominant contribution to the impedance comes from the rf cavities in the ring, then the main contribution to the summation in (15.210) will come from frequencies  $\Omega_p$  such that:

$$\Omega_p \approx \pm h\omega_0, \quad (15.221)$$

where  $h$  is the harmonic number of the storage ring. The rf wavelength must be large in comparison with the bunch length, so:

$$\frac{\Omega_p \hat{z}_0}{c} \approx \frac{h\omega_0 \hat{z}_0}{c} \ll 1. \quad (15.222)$$

Thus, we can expand the Bessel function in (15.219) to first order in  $\hat{z}_0$ , to give:

$$g_{\ell 0}(\Omega_p) \approx \frac{1}{\sqrt{2\pi\beta_z J_{z0}}} \frac{1}{2^\ell \ell!} \left( \frac{\Omega_p \hat{z}_0}{c} \right)^\ell. \quad (15.223)$$

Then, from (15.210):

$$M_{00}^{(\ell)} = 1 + i \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 \ell! (\ell-1)! E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p) \left( \frac{\Omega_p^2 \beta_z J_{z0}}{2c^2} \right)^{\ell-1}. \quad (15.224)$$

The particular case  $\ell = 1$  corresponds to a dipole mode of oscillation: if the charge distribution in a bunch has a dipole mode perturbation, then the distribution in longitudinal phase space retains its elliptical shape, but it will not be centred on the origin. In co-ordinate space, it will appear that the entire bunch oscillates backwards and forwards relative to a reference particle, as the bunch moves round the storage ring. There are corresponding energy oscillations,  $90^\circ$  out of phase with the oscillations along the  $z$  axis. Putting  $\ell = 1$  in (15.224) gives:

$$M_{00}^{(1)} = 1 + i \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p), \quad (15.225)$$

where:

$$\Omega_p = \Omega + p\omega_0 \approx p\omega_0. \quad (15.226)$$

The approximation for  $\Omega_p$  is valid if the resonant frequency of the impedance is large compared to the oscillation frequency of the dipole perturbation. We expect the oscillation frequency of the dipole perturbation to be close to the synchrotron frequency, which in practice will indeed be

small compared to the resonant frequency of the rf cavities. Finally, using (15.209) we obtain the frequency of the dipole mode perturbation:

$$\Omega \approx \bar{\omega}_z + i \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 E_0 \bar{\omega}_z} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p). \quad (15.227)$$

The exponential growth rate of the amplitude of the mode is given by the imaginary part, hence:

$$\frac{1}{\tau} \approx \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 E_0 \bar{\omega}_z} \operatorname{Re} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p). \quad (15.228)$$

From (15.171), if the imaginary part of the frequency of the perturbation is positive, then the mode is antidamped; if the imaginary part is negative, then the mode is damped.

The expression for the dipole mode frequency (15.227) should be compared with the expression for the frequency of longitudinal coupled-bunch modes (15.40). In the latter case, if we consider just a single bunch in the storage ring, then  $n_b = 1$ , and the only possible mode of oscillation is a dipole mode: since we modelled the bunch as a single macroparticle, the charge must perform rigid oscillations in longitudinal phase space. In other words, the motion is the same as in the water-bag model. If we assume that the term  $W_s^2$  in (15.40) shifts the synchrotron frequency from  $\omega_z$  to  $\bar{\omega}_z$  (we argued at the time that this term represented a change in the longitudinal focusing potential, i.e. potential-well distortion), then equations (15.40) and (15.227) give the same result for the frequency of the dipole mode. This is what we should expect: although the approach used for the analysis was very different in the two cases, the underlying physics is the same.

Equations (15.40) and (15.227) both describe the Robinson instability, which was discussed in Section 15.1.2. However, in Section 15.1.2 we only considered a dipole instability, where each bunch in the storage ring behaved as a rigid object: our analysis was based on a model where each bunch was represented as a single macroparticle. From (15.225), we see that there are in fact many longitudinal modes (for individual bunches) that can be unstable, depending on the characteristics of the impedance. However, the higher modes are expected to have lower growth rates; and fortunately, if the impedance is dominated by the rf cavities, all the modes can be suppressed in the same way by detuning the cavities slightly so that  $\omega_r = h\omega_0 + \Delta\omega$  (with  $\Delta\omega$  positive below transition, and negative above transition), where the actual rf frequency is  $h\omega_0$ .

It is interesting to consider whether the form of the charge distribution in longitudinal phase space has any effect on the Robinson instability. So far, we have only considered a water-bag distribution, but we can apply the same technique (in principle) to any charge distribution. A more realistic distribution is a Gaussian, which in terms of the longitudinal action can be written:

$$\Psi_0(J_z) = \frac{qN_b}{2\pi\epsilon_z} e^{-J_z/\epsilon_z}, \quad (15.229)$$

where  $\epsilon_z = \langle J_z \rangle$  is the longitudinal emittance. Again, the distribution is normalised so that (15.215) holds. The distribution will depend on the bunch charge, because of impedance effects; we assume that the distribution (15.229) includes potential-well distortion. The weight function in this case is:

$$W(J_z) = -\frac{1}{qN_b\beta_z} \frac{\partial\Psi_0}{\partial J_z} = \frac{e^{-J_z/\epsilon_z}}{2\pi\beta_z\epsilon_z^2}. \quad (15.230)$$

The orthonormal functions  $f_k(J_z)$  defined by (15.205) are expressed in terms of Laguerre polynomials:

$$f_k(J_z) = \sqrt{2\pi\beta_z\epsilon_z} L_k(J_z/\epsilon_z). \quad (15.231)$$

Then, from (15.211), the functions  $g_{\ell k}(\Omega_p)$  are given by:

$$g_{\ell k}(\Omega_p) = \int_0^\infty \frac{e^{-J_z/\epsilon_z}}{\sqrt{2\pi\beta_z\epsilon_z}} J_\ell\left(\zeta\sqrt{\frac{J_z}{\epsilon_z}}\right) L_k\left(\frac{J_z}{\epsilon_z}\right) \frac{dJ_z}{\epsilon_z}, \quad (15.232)$$

where:

$$\zeta = \frac{\Omega_p}{c} \sqrt{2\beta_z\epsilon_z}. \quad (15.233)$$

If the impedance is dominated by the rf cavities, and is sharply peaked at the resonant frequency  $\omega_r \approx \hbar\omega_0$  of the cavities, then the only significant contribution to  $M_{kk'}^{(\ell)}$  (15.210) comes from frequencies  $\Omega_p \approx \omega_r$ . For short bunches,  $\zeta \ll 1$ , we can approximate the Bessel function in (15.232):

$$J_\ell\left(\zeta\sqrt{\frac{J_z}{\epsilon_z}}\right) \approx \frac{1}{2^\ell\ell!} \left(\zeta\sqrt{\frac{J_z}{\epsilon_z}}\right)^\ell. \quad (15.234)$$

We can now make use of the result:

$$\int_0^\infty e^{-x} x^\nu L_k(x) dx = \frac{\Gamma(1+\nu)}{\Gamma(-\nu)} \frac{\Gamma(k-\nu)}{k!}, \quad (15.235)$$

so that (15.232) becomes:

$$g_{\ell k}(\Omega_p) = \frac{1}{2^\ell\ell!} \left(\frac{\Omega_p}{c\sqrt{\pi}}\right)^\ell (2\pi\beta_z\epsilon_z)^{(\ell-1)/2} \frac{\Gamma(1+\frac{\ell}{2})}{\Gamma(-\frac{\ell}{2})} \frac{\Gamma(k-\frac{\ell}{2})}{k!}. \quad (15.236)$$

For the dipole mode  $\ell = 1$ , substituting from (15.236) into (15.210) gives:

$$M_{kk'}^{(1)} = \delta_{kk'} + i \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p) \frac{\Gamma(k - \frac{1}{2})}{4k!} \frac{\Gamma(k' - \frac{1}{2})}{4k'!}, \quad (15.237)$$

where  $\Omega_p$  is given by (15.226). The matrix  $M^{(1)}$ , truncated at  $k_{\max}$ , has a single eigenvalue different from 1, which in the limit of large  $k_{\max}$  gives:

$$\frac{\Omega}{\bar{\omega}_z} = 1 + i \frac{q^2 N_b \omega_0^2 \eta_p}{8\pi^2 E_0 \bar{\omega}_z^2} \sum_{p=-\infty}^{\infty} \Omega_p Z_{\parallel}(\Omega_p). \quad (15.238)$$

Comparing with (15.225), we see that the dipole mode frequency for a Gaussian distribution is the same as the dipole mode frequency for a water-bag distribution. Hence, the growth rate of the dominant (dipole) mode for the Robinson instability is the same for a point-like charge, a water-bag distribution and a Gaussian distribution.

#### 15.4.4 Mode coupling

In Section 15.4.3 we found that, if the longitudinal modes with frequencies  $\Omega$  could be treated entirely separately, then, for short bunches, beam instability could occur only for a sharply peaked impedance. In that case, mode amplitudes are able to grow as a result of interaction with a wake field that persists for many turns. The resulting instability is known as the Robinson instability. For a broad-band impedance, the mode frequencies are purely real. The situation changes, however, if we allow the possibility of coupling between different modes, i.e. if we include all the terms in the summation over  $\ell'$  in (15.206), not just the term  $\ell' = \ell$ . The instability that occurs in this case, which we shall discuss in the present section, is known as a *mode-coupling instability*.

To solve the eigenvalue problem, we need to express (15.206) as a standard eigenvector equation, in terms of a two-dimensional matrix, rather than a four-dimensional matrix. The first step is to truncate the summations over  $\ell'$  and  $k'$  at values  $\pm \bar{N}$  and  $2\bar{N} + 1$  respectively, so that (15.206) becomes:

$$\frac{\Omega}{\bar{\omega}_z} a_{\ell k} \approx \sum_{\ell'=-\bar{N}}^{\bar{N}} \sum_{k'=0}^{2\bar{N}} M_{\ell k \ell' k'} a_{\ell' k'}. \quad (15.239)$$

Now we define variables  $j$  and  $j'$ :

$$j = (2\bar{N} + 1)(\ell + \bar{N}) + k, \quad (15.240)$$

$$j' = (2\bar{N} + 1)(\ell' + \bar{N}) + k'. \quad (15.241)$$

In terms of  $j$  and  $j'$ , (15.239) can be written as:

$$\frac{\Omega}{\bar{\omega}_z} a_j \approx \sum_{j'=0}^{4\bar{N}(\bar{N}+1)} M_{jj'} a_{j'}, \quad (15.242)$$

where:

$$M_{jj'} = M_{\ell k \ell' k'}, \quad (15.243)$$

and:

$$a_j = a_{\ell k}, \quad a_{j'} = a_{\ell' k'}. \quad (15.244)$$

Note that for any given mode index  $\ell$ , there are  $2\bar{N} + 1$  mode frequencies  $\Omega$ , related to the eigenvalues of  $M_{jj'}$ . Each frequency corresponds to a particular perturbation of the charge density in longitudinal phase space.

To illustrate the behaviour of the modes and the appearance of an instability, let us consider the special case of a water-bag distribution in phase space (15.214):

$$\Psi_0(J_z) = \begin{cases} qN_b/2\pi J_{z0} & \text{if } J_z \leq J_{z0}, \\ 0 & \text{if } J_z > J_{z0}. \end{cases} \quad (15.245)$$

This charge distribution was already discussed in Section 15.4.3, where we found that the weight function was given by a Dirac delta function (15.217):

$$W(J_z) = \frac{\delta_D(J_z - J_{z0})}{2\pi\beta_z J_{z0}}, \quad (15.246)$$

and the set of functions  $f_k(J_z)$  had a single member:

$$f_0(J_z) = \sqrt{2\pi\beta_z J_{z0}}. \quad (15.247)$$

In the present case, where we wish to include the effect of coupling between different modes, the fact that the set of functions  $f_k(J_z)$  has a single member leads to a significant simplification, since equation (15.239) can then be written:

$$\frac{\Omega}{\bar{\omega}_z} a_\ell \approx \sum_{\ell'=0}^{2\bar{N}+1} M_{\ell\ell'} a_{\ell'}, \quad (15.248)$$

without the need to ‘flatten’ the four-dimensional matrix.

Using (15.207), (15.246) and (15.247), the interaction matrix is:

$$M_{\ell\ell'} = \ell\delta_{\ell\ell'} + i^{1+\ell-\ell'} \frac{q^2 N_b c^2 \omega_0^2 \eta_p \ell}{4\pi^2 E_0 \bar{\omega}_z^2 \beta_z J_{z0}} \sum_{p=-\infty}^{\infty} \frac{Z_{\parallel}(\Omega_p)}{\Omega_p} J_\ell\left(\frac{\Omega_p \hat{z}_0}{c}\right) J_{\ell'}\left(\frac{\Omega_p \hat{z}_0}{c}\right), \quad (15.249)$$

where  $\hat{z}_0 = \sqrt{2\beta_z J_{z0}}$ . Let us assume that the impedance covers a wide frequency range, rather than being sharply peaked. Then, we can replace the summation in (15.249) by an integral:

$$M_{\ell\ell'} = \ell\delta_{\ell\ell'} + i^{1+\ell-\ell'} \frac{q^2 N_b c^2 \omega_0^2 \eta_p \ell}{4\pi^2 E_0 \bar{\omega}_z^2 \beta_z J_{z0}} \int_{-\infty}^{\infty} \frac{Z_{\parallel}(\omega)}{\omega} J_{\ell}\left(\frac{\omega \hat{z}_0}{c}\right) J_{\ell'}\left(\frac{\omega \hat{z}_0}{c}\right) \frac{d\omega}{\omega_0}. \quad (15.250)$$

To proceed further, we need an explicit form for the impedance. Let us consider an impedance of the form:

$$Z_{\parallel}(\omega) = R_s (1 + i \operatorname{sgn}(\omega)) \sqrt{\frac{\omega_0}{|\omega|}}, \quad (15.251)$$

with constant  $R_s$ . The impedance (15.251) is known as the *diffraction model*, and in the regime  $\omega > \omega_r$  behaves roughly as a broad-band resonator with quality factor  $Q \approx 1$  and resonant frequency  $\omega_r$ . Changing the variable of integration in (15.250) from  $\omega$  to  $u = \omega \hat{z}_0 / c$ , the interaction matrix becomes:

$$M_{\ell\ell'} = \ell\delta_{\ell\ell'} + i^{1+\ell-\ell'} \ell \Upsilon \int_{-\infty}^{\infty} (1 + i \operatorname{sgn}(u)) \frac{J_{\ell}(u) J_{\ell'}(u)}{\sqrt{|u|}} \frac{du}{u}, \quad (15.252)$$

where the dimensionless parameter  $\Upsilon$  is determined by the machine and beam parameters:

$$\Upsilon = \frac{R_s q^2 N_b \omega_0 \eta_p}{2\pi^2 E_0 \bar{\nu}_z^2} \left( \frac{c}{\omega_0 \hat{z}_0} \right)^{\frac{3}{2}}. \quad (15.253)$$

We have used (15.220) to write  $\beta_z J_{z0} = \hat{z}_0^2 / 2$ , and the synchrotron tune  $\bar{\nu}_z$  is defined as the synchrotron frequency divided by the revolution frequency:

$$\bar{\nu}_z = \frac{\bar{\omega}_z}{\omega_0}. \quad (15.254)$$

The integral in (15.252) can be evaluated numerically, or analytically using [Gradshteyn and Ryzhik (2014c)]:

$$\int_0^{\infty} J_{\ell}(u) J_{\ell'}(u) u^{-\lambda} du = \frac{\Gamma(\lambda) \Gamma(\frac{\ell+\ell'-\lambda+1}{2})}{2^{\lambda} \Gamma(\frac{-\ell+\ell'+\lambda+1}{2}) \Gamma(\frac{\ell+\ell'+\lambda+1}{2}) \Gamma(\frac{\ell-\ell'+\lambda+1}{2})}, \quad (15.255)$$

for  $\operatorname{Re}(\ell + \ell' + 1) > \operatorname{Re}(\lambda) > 0$ . The eigenvalues of the interaction matrix  $M_{\ell\ell'}$  can then be computed, for any order of truncation of the matrix. It should be noted that this analysis neglects the effects of potential-well distortion, which will change the distribution in the longitudinal phase space (and the synchrotron tune) as a function of the bunch charge. Also, the truncation procedure is not guaranteed to converge; that is, significant

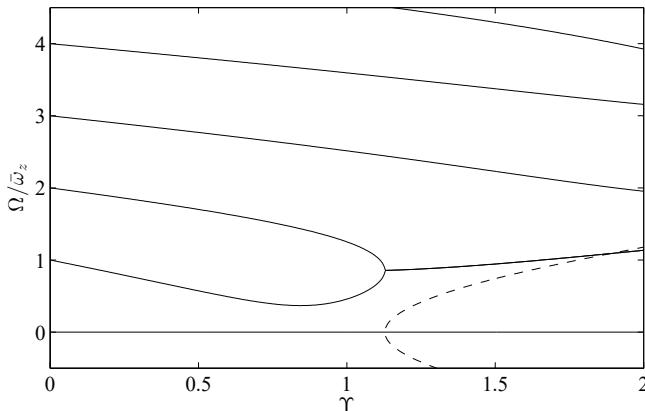


Fig. 15.8 Mode frequencies for perturbations in the longitudinal phase space for a water-bag beam. The frequencies are determined from the eigenvalues of the interaction matrix, with a diffraction model impedance (15.251). The parameter  $\Upsilon$  is defined by (15.253). The solid lines show the real parts of the eigenvalues, the dashed line shows the positive imaginary part that appears as a result of the coupling between modes  $\ell = 1$  and  $\ell = 2$ . A positive imaginary part of a mode frequency describes the growth rate of the amplitude of the perturbation. Only modes  $0 \leq \ell \leq 5$  are shown; modes with  $\ell < 0$  show the same structure.

changes in the behaviour of the low-order modes may be observed as a result of including higher-order modes in the interaction matrix. In particular, good convergence requires rapid reduction in the impedance at high frequencies: so, for example, instabilities driven by space-charge or resistive-wall impedance are not easily treated by this procedure.

Figure 15.8 shows the eigenvalues as a function of  $\Upsilon$  for the interaction matrix truncated so that  $-5 \leq \ell \leq 5$  (and similarly for  $\ell'$ ). Although the results are for the special case of a water-bag charge distribution in a storage ring with a low quality factor resonator impedance, the behaviour of the eigenvalues is characteristic for systems of this type. We see that for low values of  $\Upsilon$  (for example, small bunch population  $N_b$  or low shunt resistance  $R_s$ ), the eigenvalues are purely real, and take approximately integer values. In this regime, the frequencies of the ‘normal mode’ perturbations are approximately  $\ell\bar{\omega}_z$ , where  $\ell$  is an (integer) mode number. Inspecting the corresponding eigenvalues, and using (15.191), we find that the perturbation with mode number  $\ell$  has a dependence  $\sim e^{i\ell\phi_z}$  on the angle variable. Hence, from (15.171), the charge distribution in phase space is:

$$\Psi \approx \Psi_0 + \Psi_1 \delta_D(J_z - J_{z0}) e^{i\ell(\phi_z - \bar{\omega}_z t)}, \quad (15.256)$$

where  $\Psi_0$  is given by (15.245), and  $\Psi_1$  is a constant amplitude. Each mode takes the form of a ‘wave’ around the edge of the charge distribution in longitudinal phase space, with the wave having phase velocity equal to the synchrotron frequency.

As the bunch charge (or the impedance) is increased, the frequencies of the perturbations start to shift. At a value  $\Upsilon = \Upsilon_{\text{threshold}} \approx 1.13$ , the frequencies of the modes  $\ell = 1$  and  $\ell = 2$  become equal. If  $\Upsilon$  is increased further, the real parts of the frequencies of these modes remain equal, but the frequencies acquire imaginary parts: this means that the amplitudes of the perturbations corresponding to these modes will grow (or damp) exponentially. In general, one mode will have a frequency with a positive imaginary part, in which case the charge distribution will be unstable. Since our analysis is based on perturbation theory (that is, we assume infinitesimal perturbations) we cannot say exactly how the amplitude of the density perturbation develops. It may be, for example, that the beam quickly reaches a new equilibrium, with a different distribution. However, the imaginary part of the frequency, as a function of  $\Upsilon$ , quickly reaches a value comparable to the synchrotron frequency, and the exponential growth rate is therefore significant. In an electron storage ring, for example, the growth rate could easily become large compared with the synchrotron radiation damping rate. This type of instability, because it is characterised by the merging of two modes of density perturbation in phase space, is known as a *mode-coupling instability*.

Given the potentially rapid growth rates, it is usually desirable to operate a storage ring with parameters below the threshold for the instability. If there is a need for operation with high bunch charge, careful attention will likely need to be paid to the design and construction of the vacuum chamber, to keep the impedance as low as possible. This is not to say, however, that operation above an instability threshold is never possible. In fact, even if the growth rates of an instability are very large, the results in practice might be relatively innocuous. This is because the primary effect of a longitudinal mode-coupling instability is often to increase the longitudinal emittance, and hence to increase the bunch length. Since, as we see from (15.253), the threshold (in terms of the bunch charge) depends on the bunch length, if the bunch charge is increased above an instability threshold for a particular bunch length, it is possible for the bunch to reach a new equilibrium with the bunch length determined by the value of  $\Upsilon_{\text{threshold}}$ . From (15.253), we expect for a diffraction model impedance that the bunch

length above threshold will vary as:

$$\hat{z}_0 = \frac{c}{\omega_0} \left( \frac{R_s q^2 N_b \omega_0 \eta_p}{2\pi^2 E_0 \bar{\nu}_z^2 \Upsilon_{\text{threshold}}} \right)^{\frac{2}{3}}, \quad (15.257)$$

where  $\Upsilon_{\text{threshold}} \approx 1.13$ . Equation (15.257) provides a means of testing the theory since, in a given storage ring, it is possible to measure the bunch length as a function of bunch charge, phase slip factor, beam energy and synchrotron tune. Although the exact form of the dependence depends on the characteristics of the impedance, it turns out that under fairly general conditions it is expected that the bunch length above the instability threshold varies as:

$$\sigma_z \propto \xi^a, \quad (15.258)$$

where  $a$  is a constant parameter, and the quantity  $\xi$  is defined:

$$\xi = \frac{I_{\text{avg}} \eta_p}{E_0 \bar{\nu}_z^2}. \quad (15.259)$$

$I_{\text{avg}}$  is the average current in the ring. Equation (15.258) is known as the *scaling law* [Month and Messerschmid (1977); Chao (1993b)]. For the simplified model of instability developed in this section, and with an impedance given by the diffraction model, we expect from (15.257) that  $a = 2/3$ . Measurements at SPEAR found that the scaling law gave a good description of beam behaviour, with  $a \approx 0.76$ : this case is known as *SPEAR scaling* [Wilson *et al.* (1977)].

It should be mentioned that although we have considered here only the dynamics in the longitudinal phase space, mode coupling can occur also in the transverse phase space. The transverse mode-coupling instability is related to the head-tail instability discussed in Section 15.4.1, but an analysis based on the Vlasov equation for the evolution of a continuous charge distribution provides for a more realistic treatment than that based on a simple two-(macro)particle model. For further discussion, the reader is referred to Chao [Chao (1993b)].

**This page intentionally left blank**

# Bibliography

- Abell, D. T. (1995). *Analytic properties and Cremona approximation of transfer maps for Hamiltonian systems*, Ph.D. thesis, University of Maryland, College Park, MD, USA.
- Abi, B., Albahri, T., Al-Kilani, S., Allspach, D., Alonzi, L. P., Anastasi, A., Anisenkov, A., Azfar, F., Badgley, K., Baefbler, S., Bailey, I., Baranov, V. A., Barlas-Yucel, E., Barrett, T., Barzi, E., Basti, A., Bedeschi, F., Behnke, A., Berz, M., Bhattacharya, M., Binney, H. P., Bjorkquist, R., Bloom, P., Bono, J., Bottalico, E., Bowcock, T., Boyden, D., Cantatore, G., Carey, R. M., Carroll, J., Casey, B. C. K., Cauz, D., Ceravolo, S., Chakraborty, R., Chang, S. P., Chapelain, A., Chappa, S., Charity, S., Chislett, R., Choi, J., Chu, Z., Chupp, T. E., Convery, M. E., Conway, A., Corradi, G., Corrodi, S., Cotrozzi, L., Crnkovic, J. D., Dabagov, S., De Lurgio, P. M., Debevec, P. T., Di Falco, S., Di Meo, P., Di Sciascio, G., Di Stefano, R., Drendel, B., Driutti, A., Duginov, V. N., Eads, M., Eggert, N., Epps, A., Esquivel, J., Farooq, M., Fatemi, R., Ferrari, C., Fertl, M., Fiedler, A., Fienberg, A. T., Fioretti, A., Flay, D., Foster, S. B., Friedsam, H., Frlež, E., Froemming, N. S., Fry, J., Fu, C., Gabbanini, C., Galati, M. D., Ganguly, S., Garcia, A., Gastler, D. E., George, J., Gibbons, L. K., Gioiosa, A., Giovanetti, K. L., Girotti, P., Gohn, W., Gorringe, T., Grange, J., Grant, S., Gray, F., Haciomeroglu, S., Hahn, D., Halewood-Leagas, T., Hampai, D., Han, F., Hazen, E., Hempstead, J., Henry, S., Herrod, A. T., Hertzog, D. W., Hesketh, G., Hibbert, A., Hodge, Z., Holzbauer, J. L., Hong, K. W., Hong, R., Iacovacci, M., Incagli, M., Johnstone, C., Johnstone, J. A., Kammel, P., Kargiantoulakis, M., Karuza, M., Kaspar, J., Kawall, D., Kelton, L., Keshavarzi, A., Kessler, D., Khaw, K. S., Khechadoorian, Z., Khomutov, N. V., Kiburg, B., Kiburg, M., Kim, O., Kim, S. C., Kim, Y. I., King, B., Kinnaird, N., Korostelev, M., Kourbanis, I., Kraegeloh, E., Krylov, V. A., Kuchibhotla, A., Kuchinskiy, N. A., Labe, K. R., LaBounty, J., Lancaster, M., Lee, M. J., Lee, S., Leo, S., Li, B., Li, D., Li, L., Logashenko, I., Lorente Campos, A., Lucà, A., Lukicov, G., Luo, G., Lusiani, A., Lyon, A. L., MacCoy, B., Madrak, R., Makino, K., Marignetti, F., Mastroianni, S., Maxfield, S., McEvoy, M., Merritt,

W., Mikhailichenko, A. A., Miller, J. P., Miozzi, S., Morgan, J. P., Morse, W. M., Mott, J., Motuk, E., Nath, A., Newton, D., Nguyen, H., Oberling, M., Osofsky, R., Ostiguy, J.-F., Park, S., Paulette, G., Piacentino, G. M., Pilato, R. N., Pitts, K. T., Plaster, B., Počanić, D., Pohlman, N., Polly, C. C., Popovic, M., Price, J., Quinn, B., Raha, N., Ramachandran, S., Ramberg, E., Rider, N. T., Ritchie, J. L., Roberts, B. L., Rubin, D. L., Santi, L., Sathyam, D., Schellman, H., Schlesier, C., Schreckenberger, A., Semertzidis, Y. K., Shatunov, Y. M., Shemyakin, D., Shenk, M., Sim, D., Smith, M. W., Smith, A., Soha, A. K., Sorbara, M., Stöckinger, D., Stapleton, J., Still, D., Stoughton, C., Stratakis, D., Strohman, C., Stuttard, T., Swanson, H. E., Sweetmore, G., Sweigart, D. A., Syphers, M. J., Tarazona, D. A., Teubner, T., Tewsley-Booth, A. E., Thomson, K., Tishchenko, V., Tran, N. H., Turner, W., Valetov, E., Vasilkova, D., Venanzoni, G., Volnykh, V. P., Walton, T., Warren, M., Weisskopf, A., Welty-Rieger, L., Whitley, M., Winter, P., Wolski, A., Wormald, M., Wu, W., and Yoshikawa, C. (2021). Measurement of the positive muon anomalous magnetic moment to 0.46 ppm, *Physical Review Letters* **126**, 141801.

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs and mathematical tables*, chap. 17 Elliptic integrals (National Bureau of Standards, Washington, DC, USA).

Albahri, T., Anastasi, A., Anisenkov, A., Badgley, K., Baefbler, S., Bailey, I., Baranov, V. A., Barlas-Yucel, E., Barrett, T., Basti, A., Bedeschi, F., Berz, M., Bhattacharya, M., Binney, H. P., Bloom, P., Bono, J., Bottalico, E., Bowcock, T., Cantatore, G., Carey, R. M., Casey, B. C. K., Cauz, D., Chakraborty, R., Chang, S. P., Chapelain, A., Charity, S., Chislett, R., Choi, J., Chu, Z., Chupp, T. E., Corrodi, S., Cotrozzi, L., Crnkovic, J. D., Dabagov, S., Debevec, P. T., Di Falco, S., Di Meo, P., Di Sciascio, G., Di Stefano, R., Driutti, A., Duginov, V. N., Eads, M., Esquivel, J., Farooq, M., Fatemi, R., Ferrari, C., Fertl, M., Fienberg, A. T., Fioretti, A., Flay, D., Frlež, E., Froemming, N. S., Fry, J., Gabbanini, C., Galati, M. D., Ganguly, S., Garcia, A., George, J., Gibbons, L. K., Gioiosa, A., Giovanetti, K. L., Girotti, P., Gohn, W., Gorringe, T., Grange, J., Grant, S., Gray, F., Haciomeroglu, S., Halewood-Leagas, T., Hampai, D., Han, F., Hempstead, J., Herrod, A. T., Hertzog, D. W., Hesketh, G., Hibbert, A., Hodge, Z., Holzbauer, J. L., Hong, K. W., Hong, R., Iacovacci, M., Incagli, M., Kammel, P., Kargiantoulakis, M., Karuza, M., Kaspar, J., Kawall, D., Kelton, L., Keshavarzi, A., Kessler, D., Khaw, K. S., Khechadoorian, Z., Khomutov, N. V., Kiburg, B., Kiburg, M., Kim, O., Kim, Y. I., King, B., Kinnaird, N., Kraegeloh, E., Kuchibhotla, A., Kuchinskiy, N. A., Labe, K. R., LaBounty, J., Lancaster, M., Lee, M. J., Lee, S., Leo, S., Li, B., Li, D., Li, L., Logashenko, I., Lorente Campos, A., Lucà, A., Lukicov, G., Lussiani, A., Lyon, A. L., MacCoy, B., Madrak, R., Makino, K., Marignetti, F., Mastroianni, S., Miller, J. P., Miozzi, S., Morse, W. M., Mott, J., Nath, A., Nguyen, H., Osofsky, R., Park, S., Paulette, G., Piacentino, G. M., Pilato, R. N., Pitts, K. T., Plaster, B., Počanić, D., Pohlman, N., Polly, C. C., Price, J., Quinn, B., Raha, N., Ramachandran, S., Ramberg, E., Ritchie,

- J. L., Roberts, B. L., Rubin, D. L., Santi, L., Schlesier, C., Schreckenberger, A., Semertzidis, Y. K., Shemyakin, D., Smith, M. W., Sorbara, M., Stöckinger, D., Stapleton, J., Stoughton, C., Stratakis, D., Stuttard, T., Swanson, H. E., Sweetmore, G., Sweigart, D. A., Syphers, M. J., Tarazona, D. A., Teubner, T., Tewsley-Booth, A. E., Thomson, K., Tishchenko, V., Tran, N. H., Turner, W., Valetov, E., Vasilkova, D., Venanzoni, G., Walton, T., Weisskopf, A., Welty-Rieger, L., Winter, P., Wolski, A., and Wu, W. (2021). Measurement of the anomalous precession frequency of the muon in the Fermilab Muon  $g - 2$  Experiment, *Physical Review D* **103**, 072002.
- Alexahin, Y. (2005). Theory and reality of beam-beam effects at hadron colliders, in *Proceedings of the 2005 Particle Accelerator Conference* (Knoxville, TN, USA), pp. 544–548.
- Antoniou, F., Papaphilippou, Y., Aiba, M., Boege, M., Milas, N., Streun, A., and Demma, T. (2012). Intrabeam scattering studies at the swiss Light Source, in *Proceedings of the Third International Particle Accelerator Conference* (New Orleans, LA, USA), pp. 1951–1953.
- Arnold, V. I. (1964). Instability of dynamical systems with several degrees of freedom, *Soviet Mathematics* **5**, pp. 581–588.
- Arnold, V. I. (1989). *Mathematical methods of classical mechanics*, chap. Appendix 8 Theory of perturbations of conditionally periodic motion, and Kolmogorov's theorem, 2nd edn. (Springer, New York, NY, USA).
- Artikova, S., Grieser, M., and Ullrich, J. (2011). Investigation of intrabeam scattering in the heavy ion storage ring TSR, in *Proceedings of the Second International Particle Accelerator Conference* (San Sebastian, Spain), pp. 2490–2492.
- Bane, K. L. F. (1991). The short range resistive wall wakefields, Tech. Rep. SLAC/AP-87, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Bane, K. L. F. (2002). An accurate, simplified model of intrabeam scattering, Tech. Rep. SLAC-AP-141, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Bane, K. L. F., Hayano, H., Kubo, K., Naito, T., Okugi, T., and Urakawa, J. (2002). Intrabeam scattering analysis of measurements at KEK's Accelerator Test Facility damping ring, *Physical Review Special Topics – Accelerators and Beams* **5**, 084403.
- Bane, K. L. F., Oide, K., and Zobov, M. (2004). Impedance calculation and verification in storage rings, in *Proceedings of the First CARE-HHH-APD Workshop on Beam Dynamics in Future Hadron Colliders and Rapidly Cycling High-Intensity Synchrotrons*, CERN-2005-006; CARE-Conf-05-002-HHH (Geneva, Switzerland), pp. 143–158.
- Bane, K. L. F. and Sands, M. (1995). The short-range resistive wall wakefields, Tech. Rep. SLAC-PUB-95-7074, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Barber, D. P. (1999). Electron and positron spin polarisation in storage rings – an introduction, in P. Chen (ed.), *Advanced ICFA beam dynamics workshop on quantum aspects of beam physics* (World Scientific, Singapore), pp. 67–90.
- Barber, D. P., Heinemann, K., and Ripken, G. (1999). Notes on spin dynamics

- in storage rings (second revision), Tech. Rep. DESY M 92–04, DESY, Hamburg, Germany.
- Barber, D. P. and Ripken, G. (1999). Radiative polarization in electron storage rings, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 153–156.
- Bargmann, V., Michel, L., and Telegdi, V. L. (1959). Precession of the polarization of particles moving in a homogeneous electromagnetic field, *Physical Review Letters* **2**, pp. 435–436.
- Bassetti, M. and Erskine, G. A. (1980). Closed expression for the electrical field of a two-dimensional gaussian charge, Tech. Rep. CERN-ISR-TH/80-06, European Organization for Nuclear Research, Geneva, Switzerland.
- Berg, J. S., Warnock, R. L., Ruth, R. D., and Forest, E. (1994). Construction of symplectic maps for nonlinear motion of particles in accelerators, *Physical Review E* **49**, 1, pp. 722–739.
- Beringer, J. et al. (Particle Data Group) (2012). Review of particle physics, *Physical Review D* **86**, 010001.
- Bernardini, C., Corazza, G. F., Giugno, G. D., Ghigo, G., Haissinski, J., Marin, P., Querzoli, R., and Touschek, B. (1946). Lifetime and beam size in a storage ring, *Physical Review Letters* **10**, 9, pp. 407–409.
- Berz, M. (1987). The method of power series tracking for the mathematical description of beam dynamics, *Nuclear Instruments and Methods in Physics Research Section A* **258**, pp. 431–436.
- Berz, M. (1989). Differential algebraic description of beam dynamics to very high orders, *Particle Accelerators* **24**, pp. 109–124.
- Berz, M. (1991). Symplectic tracking through circular accelerators with high order maps, in W. Scandale and G. Turchetti (eds.), *Nonlinear problems in future particle accelerators* (World Scientific, Singapore), pp. 288–296.
- Berz, M. (1999). *Modern map methods in particle beam physics*, chap. 2 Differential algebraic techniques (Academic Press, London, UK).
- Berz, M., Makino, K., and Wan, W. (2015). *An introduction to beam physics*, chap. 5 Computation and properties of maps (CRC Press, Boca Raton, FL, USA).
- Bhat, C., Spentzouris, K., and Colestock, P. L. (1999). Measurements of intrabeam scattering rates below transition in the Fermilab Antiproton Accumulator, in *Proceedings of the 1999 Particle Accelerator Conference* (New York, NY, USA), pp. 3155–3157.
- Bhat, C. M. and Marriner, J. (1991). Intrabeam scattering in the Fermilab Antiproton Accumulator, in *Proceedings of the 1991 Particle Accelerator Conference* (San Francisco, CA, USA), pp. 1771–1775.
- Bjorken, J. D. and Mttingwa, S. K. (1983). Intrabeam scattering, *Particle Accelerators* **13**, pp. 115–143.
- Bocchetta, C. (1998). Lifetime and beam quality, in *Proceedings of the CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers*, CERN 98-04 (Grenoble, France (22–27 April 1996)), pp. 221–285.
- Boussard, D. (1975). Observation of microwave longitudinal instabilities in the CPS, Tech. Rep. LABII/RF/Int./75-2, CERN, Geneva, Switzerland.

- Bravin, E. (2009). Transverse beam profiles, in *Proceedings of the CERN Accelerator School, Beam Diagnostics*, CERN-2009-005 (Dourdan, France (28 May–6 June 2008)), pp. 377–406.
- Bravin, E. (2020). Transverse beam profiles, in *Proceedings of the 2018 CERN Accelerator School course on Beam Instrumentation*, arXiv:2005.07400 [physics.acc-ph] (Tuusula, Finland (2–15 June 2018)), pp. 318–354.
- Brinkmann, R., Derbenev, Y., and Flöttmann, K. (1999). A flat beam electron source for linear colliders, Tech. Rep. TESLA Note 99-09, DESY, Hamburg, Germany.
- Brown, K. L. (1982). A first- and second-order matrix theory for the design of beam transport systems and charged particle spectrometers, Tech. Rep. SLAC Report-75, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Brown, K. L., Rothacker, F., Carey, D. C., and Iselin, C. (1977). TRANSPORT: a computer program for designing charged particle beam transport systems, Tech. Rep. SLAC-91, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Burkhardt, H., Manglunki, D., Martini, M., Roncarolo, F., and Rumolo, G. (2004). Investigation of space charge effects and intrabeam scattering for lead ions in the SPS, in *Proceedings of the Ninth European Particle Accelerator Conference* (Lucerne, Switzerland), pp. 1846–1848.
- Butcher, J. C. (2008). *Numerical methods for ordinary differential equations*, chap. 3 Runge–Kutta methods, 2nd edn. (John Wiley and Sons, Chichester, UK).
- Byrd, J. M., Leemans, W. P., Loftsdottir, A., Marcelis, B., Martin, M. C., McKinney, W. R., Sannibale, F., Scarvie, T., and Steier, C. (2002). Observation of broadband self-amplified spontaneous coherent terahertz synchrotron radiation in a storage ring, *Physical Review Letters* **89**, 224801.
- Byrd, J. M., Santis, S. D., Georgsson, M., Stover, G., Fox, J. D., and Teytelman, D. (2000). Commissioning of a higher harmonic rf system for the Advanced Light Source, *Nuclear Instruments and Methods in Physics Research Section A* **455**, 2, pp. 271–282.
- Cai, Y., Bane, K., Hettel, R., Nosochkov, Y., Wang, M.-H., and Borland, M. (2012). Ultimate storage ring based on fourth-order geometric achromats, *Physical Review Special Topics – Accelerators and Beams* **15**, 054002.
- Cai, Y., Donald, M., Helm, R., Irwin, J., Nosochkov, Y., Ritson, D. M., Yan, Y. T., Forest, E., and Zholents, A. (1995). Low energy ring lattice of the PEP-II asymmetric B-Factory, in *Proceedings of the 1995 Particle Accelerator Conference* (Dallas, TX, USA), pp. 576–578.
- Callaghan, E. E. and Maslen, S. H. (1960). The magnetic field of a finite solenoid, Tech. Rep. NASA TN D-465, National Aeronautics and Space Administration, Lewis Research Center, Cleveland, OH, USA.
- Carey, D. C. (1992). *The optics of charged particle beams*, chap. 7 Fringe fields (Harwood Academic Publishers, New York, NY, USA).
- Chambers, E. E. (1968). Particle motion in a standing wave linear accelerator, in *Proceedings of the 1968 Summer Study on Superconducting Devices and Accelerators* (Brookhaven, NY, USA), pp. 79–100.

- Chambliss, A. and Franklin, J. (2020). A magnetic velocity Verlet method, *American Journal of Physics* **88**, pp. 1075–1082, see also arXiv:2008.11810 [physics.class-ph].
- Chao, A. W. (1979). Evaluation of beam distribution parameters in an electron storage ring, *Journal of Applied Physics* **50**, pp. 595–598.
- Chao, A. W. (1993a). *Physics of collective beam instabilities in high energy accelerators*, chap. 2 Wake fields and impedances (John Wiley and Sons, Inc., New York, NY, USA).
- Chao, A. W. (1993b). *Physics of collective beam instabilities in high energy accelerators*, chap. 6 Perturbation formalism (John Wiley and Sons, Inc., New York, NY, USA).
- Chao, A. W. (2022a). *Special topics in accelerator physics*, chap. 4 Lie algebra (World Scientific Publishing Company, New Jersey, USA).
- Chao, A. W. (2022b). *Special topics in accelerator physics*, chap. 3 Truncated power series algebra (World Scientific Publishing Company, New Jersey, USA).
- Chao, A. W. (2022c). *Special topics in accelerator physics*, chap. 2 Symplectification of maps (World Scientific Publishing Company, New Jersey, USA).
- Chao, A. W. (2022d). *Special topics in accelerator physics*, chap. 8 Beam-beam interaction (World Scientific Publishing Company, New Jersey, USA).
- Chasman, R., Green, G. K., and Rowe, E. M. (1975). Preliminary design of a dedicated synchrotron radiation facility, *IEEE Transactions on Nuclear Science* **NS-22**, 3, pp. 1765–1767.
- Chirikov, B. V. (1979). A universal instability of many-dimensional oscillator systems, *Physics Reports* **52**, 5, pp. 263–379.
- Clarke, J. A. (2004). *The science and technology of undulators and wigglers* (Oxford University Press, Oxford, UK).
- Conte, M. and Martini, M. (1985). Intrabeam scattering in the CERN Antiproton Accumulator, *Particle Accelerators* **17**, pp. 1–10.
- Conway, J. T. (2001). Exact solutions for the magnetic fields of axisymmetric solenoids and current distributions, *IEEE Transactions on Magnetics* **37**, 4, pp. 2977–2988.
- Courant, E. D. and Snyder, H. S. (1958). Theory of the alternating-gradient synchrotron, *Annals of Physics* **3**, pp. 1–48.
- Deacon, D. A. G., Elias, L. R., Madey, J. M. J., Ramian, G. J., Schwettman, H. A., and Smith, T. I. (1977). First operation of a free-electron laser, *Physical Review Letters* **38**, pp. 892–894.
- Derbenev, Y. (1998). Adapting optics for high energy electron cooling, Tech. Rep. UM HE 98–04, Randall Laboratory of Physics, University of Michigan, Ann Arbor, MI, USA.
- Derbenev, Y. S. and Kondratenko, A. M. (1973). *Soviet Physics: Journal of Experimental and Theoretical Physics* **37**, 6, p. 968.
- Derby, N. and Olbert, S. (2010). Cylindrical magnets and ideal solenoids, *American Journal of Physics* **78**, pp. 229–235.
- Di Mitri, S., Perosa, G., Brynes, A., Setija, I., Spampinati, S., Williams, P. H., Wolski, A., Allaria, E., Brussard, S., Giannessi, L., Penco, G., Rebernik,

- P. R., and Trovò, M. (2020). Experimental evidence of intrabeam scattering in a free-electron laser driver, *New Journal of Physics* **22**, 083053.
- Di Mitri, S., Perosa, G., Brynes, A., Setija, I., Spampinati, S., Williams, P. H., Wolski, A., Allaria, E., Brussard, S., Giannessi, L., Penco, G., Rebernik, P. R., and Trovò, M. (2022). Addendum: Experimental evidence of intra-beam scattering in a free-electron laser driver, *New Journal of Physics* **24**, 039401.
- Dijkstal, P., Arnold, M., Burandt, C., Hug, F., and Pietralla, N. (2015). Automated transverse beam emittance measurement using a slow wire scanner at the S-DALINAC, in *Proceedings of the Sixth International Particle Accelerator Conference* (Richmond, VA, USA), pp. 817–819.
- Dragt, A. J. (1990). Numerical third-order transfer map for solenoid, *Nuclear Instruments and Methods in Physics Research Section A* **298**, pp. 441–459.
- Dragt, A. J. (2020). Lie methods for nonlinear dynamics with applications to accelerator physics, <http://www.physics.umd.edu/dsat/dsatliemethods.html>.
- Dragt, A. J. and Finn, J. M. (1976). Lie series and invariant functions for analytic symplectic maps, *Journal of Mathematical Physics* **17**, pp. 2215–2227.
- Dragt, A. J. and Finn, J. M. (1979). Normal form for mirror machine Hamiltonians, *Journal of Mathematical Physics* **20**, pp. 2649–2660.
- Dragt, A. J. and Forest, E. (1983). Computation of nonlinear behaviour of Hamiltonian systems using Lie algebraic methods, *Journal of Mathematical Physics* **24**, pp. 2734–2744.
- Dumas, H. S. and Laskar, J. (1993). Global dynamics and long-time stability in Hamiltonian systems via numerical frequency analysis, *Physical Review Letters* **70**, 20, pp. 2975–2979.
- Edwards, D., Edwards, H., Holtkamp, N., Nagaitsev, S., Santucci, J., Brinkmann, R., Desler, K., Flöttmann, K., Bohnet, I., and Ferrario, M. (2000). The flat beam experiment at the FNAL photoinjector, in *Proceedings of the XX International Linac Conference* (Monterey, CA, USA), pp. 122–124.
- Edwards, D. A. and Syphers, M. (1999). Linear betatron motion, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 49–50.
- Edwards, D. A. and Teng, L. C. (1973). Parameterization of linear coupled motion in periodic systems, *IEEE Transactions on Nuclear Science* **20**, 3, pp. 885–888.
- Ehrlichman, M. P., Hartung, W., Heltsley, B., Peterson, D. P., Rider, N., Rubin, D., Sagan, D., Shanks, J., Wang, S. T., Campbell, R., and Holtzapple, R. (2013). Intrabeam scattering studies at the cornell electron storage ring test accelerator, *Physical Review Special Topics – Accelerators and Beams* **16**, 104401.
- Emery, L. and Borland, M. (1999). Top-up operation experience at the Advanced Photon Source, in *Proceedings of the 1999 Particle Accelerator Conference* (New York, NY, USA), pp. 200–202.
- Enge, H. A. (1967). Deflecting magnets, in A. Septier (ed.), *Focusing of charged particles*, Vol. 2 (Academic Press, Orlando, FL, USA), pp. 203–264.
- Erdélyi, B. (2006). Global theory of extended generating functions, *International Journal of Pure and Applied Mathematics* **33**, 4, pp. 553–578.

- Erdélyi, B. and Berz, M. (2001). Optimal symplectic approximation of Hamiltonian flows, *Physical Review Letters* **87**, 114302.
- Erdélyi, B. and Berz, M. (2004). Local theory and applications of extended generating functions, *International Journal of Pure and Applied Mathematics* **11**, 3, pp. 241–280.
- Eriksson, M., Lindgren, L., Sjöström, M., Wallén, E., Rivkin, L., and Streun, A. (2008). Some small-emittance light-source lattices with multi-bend achromats, *Nuclear Instruments and Methods in Physics Research Section A* **587**, pp. 221–226.
- Evans, L. R. and Gareyte, J. (1985). Performance limitations of the CERN SPS collider, *IEEE Transactions on Nuclear Science* **NS-32**, 5, pp. 2234–2236.
- Farvacque, L., Carmignani, N., Chavanne, J., Franchi, A., Le Bec, G., Liuzzo, S., Nash, B., Perron, T., and Raimondi, P. (2013). A low-emittance lattice for the ESRF, in *Proceedings of the Fourth International Particle Accelerator Conference* (Shanghai, China), pp. 79–81.
- Fischer, W., Bai, M., Blaskiewicz, M., Brennan, J. M., Cameron, P., Connolly, R., Lehrach, A., Parzen, G., Tepikian, S., Zeno, K., and van Zeijts, J. (2001). Measurements of intra-beam scattering growth times with gold beam below transition in RHIC, in *Proceedings of the 2001 Particle Accelerator Conference* (Chicago, IL, USA), pp. 2857–2859.
- Forest, E. (1998). *Beam dynamics: a new attitude and framework* (Harwood Academic Publishers, Amsterdam, Netherlands).
- Forest, E. (2006). Geometric integration for particle accelerators, *Journal of Physics A: Mathematical and Theoretical* **39**, pp. 5321–5377.
- Forest, E., Berz, M., and Irwin, J. (1989). Normal form methods for complicated periodic systems: a complete solution using differential algebra and Lie operators, *Particle Accelerators* **24**, pp. 91–107.
- Forest, E. and Ruth, R. D. (1990). Fourth-order symplectic integration, *Physica D* **43**, pp. 105–117.
- Fox, J. D., Cesaratto, J. M., Dusatko, J., Olsen, J., Pollock, K., Rivetta, C., Turgut, O., Bartosik, H., Hofle, W., Kotzian, G., and Wehrle, U. (2013). First results and analysis of the performance of a 4 GS/s intra-bunch vertical feedback system at the SPS, in *Proceedings of the Fourth International Particle Accelerator Conference* (Shanghai, China), pp. 3070–3072.
- Fox, J. D., Larsen, R., Prabhakar, S., Teytelman, D., Young, A., Drago, A., Serio, M., Barry, W., and Stover, G. (1999). Multi-bunch instability diagnostics via digital feedback systems at PEP-II, DAΦNE, ALS and SPEAR, in *Proceedings of the 1999 Particle Accelerator Conference* (New York, NY, USA), pp. 636–640.
- Fox, J. D., Mastorides, T., Rivetta, C., Winkle, D. V., and Teytelman, D. (2008). Lessons learned from PEP-II LLRF and longitudinal feedback, in *Proceedings of the Eleventh European Particle Accelerator Conference* (Genoa, Italy), pp. 1953–1955.
- Freund, H. P. and Antonson, Jr, T. M. (1996). *Principles of free-electron lasers*, 2nd edn. (Chapman and Hall, London, UK).
- Gao, J. (2000). Analytical estimation of the dynamic apertures of circular

- accelerators, *Nuclear Instruments and Methods in Physics Research Section A* **451**, 3, pp. 545–557.
- Garg, A. (2012). *Classical electromagnetism in a nutshell* (Princeton University Press, Princeton, NJ, USA).
- Gluckstern, R. L. and Kurennoy, S. S. (1999). Impedance calculation, frequency domain, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 195–199.
- Goldstein, H., Poole Jr., C. P., and Safko, J. L. (2001a). *Classical mechanics*, 3rd edn. (Addison–Wesley, Boston, MA, USA).
- Goldstein, H., Poole Jr., C. P., and Safko, J. L. (2001b). *Classical mechanics*, chap. 10 Hamilton–Jacobi theory and action-angle variables, 3rd edn. (Addison–Wesley, Boston, MA, USA).
- Goldstein, H., Poole Jr., C. P., and Safko, J. L. (2001c). *Classical mechanics*, chap. 9 Canonical transformations, 3rd edn. (Addison–Wesley, Boston, MA, USA).
- Goldstein, H., Poole Jr., C. P., and Safko, J. L. (2001d). *Classical mechanics*, chap. 12 Canonical perturbation theory, 3rd edn. (Addison–Wesley, Boston, MA, USA).
- Gradshteyn, I. S. and Ryzhik, I. M. (2014a). *Table of integrals, series and products*, chap. 8.1 Elliptic integrals and functions, 8th edn. (Academic Press, Burlington, MA, USA).
- Gradshteyn, I. S. and Ryzhik, I. M. (2014b). *Table of integrals, series and products*, chap. 10.611 Vector field theory: orthogonal curvilinear coordinates, 8th edn. (Academic Press, Burlington, MA, USA).
- Gradshteyn, I. S. and Ryzhik, I. M. (2014c). *Table of integrals, series and products*, chap. 6.574 Definite integrals of special functions: combinations of Bessel functions and algebraic functions, 8th edn. (Academic Press, Burlington, MA, USA).
- Gram, R. and Morton, P. (1967). Advantages of a set of second harmonic rf cavities, Tech. Rep. TN-67-30, Stanford Linear Accelerator Centre, Stanford, CA, USA.
- Granum, P., Madsen, M. L., McKenna, J. T. K., Hodgkinson, D. L., and Fajans, J. (2022). Efficient calculations of magnetic fields of solenoids for simulations, *Nuclear Instruments and Methods in Physics Research Section A* **1034**, 166706.
- Green, A. T. and Shin, Y. (2015). Implementation of quadrupole-scan emittance measurement at Fermilab's Advanced Superconducting Test Accelerator (ASTA), in *Proceedings of the Sixth International Particle Accelerator Conference* (Richmond, VA, USA), pp. 669–671.
- Guignard, G. and Hagel, J. (1986). Dynamic aperture: analytic derivation and sextupole algorithms, in *Nonlinear dynamics aspects of particle accelerators, Lecture Notes in Physics*, Vol. 247 (Springer, Berlin, Germany), pp. 367–389.
- Hairer, E., Lubich, C., and Wanner, G. (2006a). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, 2nd edn. (Springer, Berlin, Germany).

- Hairer, E., Lubich, C., and Wanner, G. (2006b). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, chap. I.1.4 The Störmer–Verlet scheme, 2nd edn. (Springer, Berlin, Germany).
- Hairer, E., Lubich, C., and Wanner, G. (2006c). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, chap. VI.4 Symplectic Runge–Kutta methods, 2nd edn. (Springer, Berlin, Germany).
- Hairer, E., Lubich, C., and Wanner, G. (2006d). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, chap. II.1.3 Gauss and Lobatto collocation, 2nd edn. (Springer, Berlin, Germany).
- Hairer, E., Wanner, G., and Nørsett, S. P. (1993). *Solving ordinary differential equations I: nonstiff problems*, chap. 2 Runge–Kutta and extrapolation methods (Springer, Berlin, Germany).
- Haissinski, J. (1973). Exact longitudinal equilibrium distribution of stored electrons in the presence of self-fields, *Il Nuovo Cimento B* **18**, pp. 72–82.
- Hellert, T., Steier, C., and Venturini, M. (2022). Lattice correction and commissioning simulation of the Advanced Light Source upgrade storage ring, *Physical Review Accelerators and Beams* **25**, 110701.
- Herr, W. (2001). Beam-beam issues in the LHC and relevant experience from the SPS proton antiproton collider and LEP, in T. Sen and M. Xiao (eds.), *Proceedings of a Workshop on Beam-Beam Effects in Circular Colliders*, FERMILAB-Conf-01/390-T (Batavia, IL, USA), pp. 6–16.
- Hock, K. M. and Wolski, A. (2011). The effect of NEG coating on resistive wall wake field, *Nuclear Instruments and Methods in Physics Research Section A* **657**, 1, pp. 94–98.
- Hofmann, A. (1977). Single-beam collective phenomena: longitudinal, in M. H. Blewett (ed.), *Theoretical aspects of the behaviour of beams in accelerators and storage rings*, CERN 77–13 (Geneva, Switzerland), pp. 139–174.
- Hofmann, A. (2004). *The physics of synchrotron radiation* (Cambridge University Press, Cambridge, UK).
- Hofmann, A. and Myers, S. (1980). Beam dynamics in a double rf system, in *Proceedings of the Eleventh International Conference on High Energy Accelerators* (Geneva, Switzerland), pp. 610–614.
- Hu, M. and Nagaitsev, S. (2005). Observation of longitudinal diffusion and cooling due to intrabeam scattering at the Fermilab Recycler Ring, in *Proceedings of the 2005 Particle Accelerator Conference* (Knoxville, TN, USA), pp. 1560–1562.
- Irwin, J. and Yan, Y. T. (1999). Dynamic aperture, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 104–107.
- Jackson, J. D. (1998). *Classical electrodynamics*, 3rd edn. (John Wiley and Sons, New York, NY, USA).
- Kapchinsky, I. M. and Vladimirsksy, V. V. (1959). Limitations of proton beam current in a strong focusing linear accelerator associated with the beam

- space charge, in *Proceedings of the International Conference on High Energy Accelerators* (Geneva, Switzerland), pp. 274–288.
- Keil, E. and Schnell, W. (1969). Concerning longitudinal instability in the ISR, Tech. Rep. CERN-ISR-TH-RF/69-48, CERN, Geneva, Switzerland.
- Keil, J., Ehrlichmann, H., Sahoo, G., and Wanzenberg, R. (2017). Quadrupole scan measurements in the beam transport line between DESY II and PETRA III, in *Proceedings of the Eighth International Particle Accelerator Conference* (Copenhagen, Denmark), pp. 174–176.
- Kellogg, O. D. (1953). *Foundations of potential theory* (Dover, New York, NY, USA).
- Kim, K.-J., Huang, Z., and Lindberg, R. (2017). *Synchrotron radiation and free-electron lasers* (Cambridge University Press, Cambridge, UK).
- Klein, R., Mayer, T., Kuske, P., Thornagel, R., and Ulm, G. (1997). Beam diagnostics at the BESSY I electron storage ring with Compton backscattered laser photons: measurement of the electron energy and related quantities, *Nuclear Instruments and Methods in Physics Research Section A* **384**, 2, pp. 293–298.
- Kleiss, R., Schmidt, F., Yan, Y. T., and Zimmermann, F. (1992). On the feasibility of tracking with differential-algebra maps in long-term stability studies for large hadron colliders, Tech. Rep. CERN SL/92-02 (AP), DESY HERA 92 01, SSCL Report SSCL 564, CERN, Geneva, Switzerland.
- Kohaupt, R. D. (1972). Excitation of a transverse instability by parasitic cavity modes, Tech. Rep. DESY H1-74/2, DESY, Hamburg, Germany.
- Kohaupt, R. D. (1983). The experimental study of a higher harmonic rf system in PETRA, *IEEE Transactions on Nuclear Science* **NS-30**, 4, pp. 2525–2527.
- Koscieliak, S. (2001). Longitudinal space-charge geometric factor for an elliptical beam, in *Proceedings of the 2001 Particle Accelerator Conference* (Chicago, IL, USA), pp. 2970–2972.
- Koziel, H. (2005). Beam diagnostics for accelerators, in *Proceedings of the CERN Accelerator School, Basic Course on General Accelerator Physics*, CERN-2005-004 (Loutraki, Greece (2–13 October 2000)), pp. 154–197.
- Kubo, K., Akemoto, M., Anderson, S., Aoki, T., Araki, S., Bane, K. L. F., Blum, P., Corlett, J., Dobashi, K., Emma, P., Frisch, J., Fukuda, M., Guo, Z., Hasegawa, K., Hayano, H., Higo, T., Higurashi, A., Honda, Y., Iimura, T., Imai, T., Jobe, K., Kamada, S., Karataev, P., Kashiwagi, S., Kim, E., Kobuki, T., Kotseroglou, T., Kurihara, Y., Kuriki, M., Kuroda, R., Kuroda, S., Lee, T., Luo, X., McCormick, D. J., McKee, B., Mimashi, T., Minty, M., Muto, T., Naito, T., Naumenko, G., Nelson, J., Nguyen, M. N., Oide, K., Okugi, T., Omori, T., Oshima, T., Pei, G., Potylitsyn, A., Qin, Q., Raubenheimer, T., Ross, M., Sakai, H., Sakai, I., Schmidt, F., Slaton, T., Smith, H., Smith, S., Smith, T., Suzuki, T., Takano, M., Takeda, S., Terunuma, N., Toge, N., Turner, J., Urakawa, J., Vogel, V., Woodley, M., Yocky, J., Young, A., and Zimmermann, F. (2002). Extremely low vertical-emittance beam in the Accelerator Test Facility at KEK, *Physical Review Letters* **88**, 194801.

- Kubo, K., Mttingwa, S. K., and Wolski, A. (2005). Intrabeam scattering formulas for high energy beams, *Physical Review Special Topics – Accelerators and Beams* **8**, 081001.
- Kubo, K. and Oide, K. (2001). Intrabeam scattering in electron storage rings, *Physical Review Special Topics – Accelerators and Beams* **4**, 124401.
- Landau, L. (1946). On the vibration of the electronic plasma, *Journal of Physics (USSR)* **10**, pp. 25–34.
- Laskar, J. (1990). The chaotic motion of the solar system: a numerical estimate of the size of the chaotic zones, *Icarus* **88**, pp. 266–291.
- Laskar, J. (1993). Frequency analysis for multi-dimensional systems: global dynamics and diffusion, *Physica D* **67**, pp. 257–281.
- Laskar, J. (1994). Frequency map analysis of a Hamiltonian system, in *Nonlinear dynamics in particle accelerators: theory and experiments* (Arcidosso, Italy), pp. 160–169.
- Laskar, J. (1995). Introduction to frequency map analysis, in *Proceedings of the NATO Advanced Studies Institute on Hamiltonian Systems with Three or More Degrees of Freedom* (S'Agaró, Spain), pp. 134–150.
- Laskar, J. and Robin, D. (1996). Application of frequency map analysis to the ALS, *Particle Accelerators* **54**, pp. 183–192.
- Laskar, J. and Robutel, P. (1993). The chaotic obliquity of the planets, *Nature* **361**, pp. 608–612.
- Laskar, J. and Robutel, P. (2001). High order symplectic integrators for perturbed Hamiltonian systems, *Celestial Mechanics and Dynamical Astronomy* **80**, pp. 39–62.
- Laslett, L. J. (1963). On intensity limitations imposed by transverse space-charge effects in circular particle accelerators, in *Summer Study on Storage Rings* (BNL Report 7534), pp. 325–367.
- Lawson, J. D. (1968). Radiation from a ring charge passing through a resonator, Tech. Rep. RHEL/M 144, Rutherford Appleton Laboratory, Chilton, Oxon, UK.
- Lawson, J. D. (1990). Radiation from a ring charge passing through a resonator, *Particle Accelerators* **25**, pp. 107–112.
- Lee, S. Y. (1997). *Spin dynamics and snakes in storage rings* (World Scientific, Singapore).
- Leemann, S., Andersson, Å., Eriksson, M., Lindgren, L.-J., Wallén, E., Bengtsson, J., and Streun, A. (2009). Beam dynamics and expected performance of Sweden's new storage-ring light source: MAX IV, *Physical Review Special Topics – Accelerators and Beams* **12**, 120701.
- Lonza, M. (2009). Multi-bunch feedback systems, in *Proceedings of the CERN Accelerator School, Beam Diagnostics*, CERN–2009–005 (Dourdan, France (28 May–6 June 2008)), pp. 467–511.
- Madey, J. (1971). Stimulated emission of bremsstrahlung in a periodic magnetic field, *Journal of Applied Physics* **42**, pp. 1906–1913.
- Maroju, P. K., Grazioli, C., Fraia, M. D., Moioli, M., Ertel, D., Ahmadi, H., Plekan, O., Finetti, P., Allaria, E., Giannessi, L., Ninno, G. D., Spezzani, C., Penco, G., Spampinati, S., Demidovich, A., Danailov, M. B., Borghes,

- R., Kourousias, G., Reis, C. E. S. D., Billé, F., Lutman, A. A., Squibb, R. J., Feifel, R., Carpeggiani, P., Reduzzi, M., Mazza, T., Meyer, M., Bengtsson, S., Ibrakovic, N., Simpson, E. R., Mauritsson, J., Csizmadia, T., Dumergue, M., Kühn, S., Gopalakrishna, H. N., You, D., Ueda, K., Labeye, M., Bækhøj, J. E., Schafer, K. J., Gryzlova, E. V., Grum-Grzhimailo, A. N., Prince, K. C., Callegari, C., and Sansone, G. (2020). Attosecond pulse shaping using a seeded free-electron laser, *Nature* **578**, pp. 386–391.
- McKee, C. B., O'Shea, P. G., and Madey, J. M. J. (1995). Phase space tomography of relativistic electron beams, *Nuclear Instruments and Methods in Physics Research Section A* **358**, pp. 264–267.
- McLachlan, R. I. (1995). Composition methods in the presence of small parameters, *BIT* **35**, pp. 258–268.
- Melissinos, A. C. (1995). Energy measurement by resonant depolarization, in *Proceedings of the CERN Accelerator School, Fifth Advanced Accelerator Physics Course*, CERN 95–06 (Rhodes, Greece (20 September–1 October 1993)), pp. 1051–1065.
- Mess, K.-H., Schmüser, P., and Wolff, S. (1996). *Superconducting accelerator magnets* (World Scientific, Singapore).
- Minty, M. G. and Zimmermann, F. (2003a). *Measurement and control of charged particle beams*, chap. 7 Longitudinal optics measurement and correction (Springer, Berlin, Germany).
- Minty, M. G. and Zimmermann, F. (2003b). *Measurement and control of charged particle beams*, chap. 4 Transverse beam emittance measurement and control (Springer, Berlin, Germany).
- Minty, M. G. and Zimmermann, F. (2003c). *Measurement and control of charged particle beams* (Springer, Berlin, Germany).
- Minty, M. G. and Zimmermann, F. (2003d). *Measurement and control of charged particle beams*, chap. 2 Transverse optics measurement and correction (Springer, Berlin, Germany).
- Minty, M. G. and Zimmermann, F. (2003e). *Measurement and control of charged particle beams*, chap. 3 Orbit measurement and correction (Springer, Berlin, Germany).
- Mitchell, C. E. and Dragt, A. J. (2010). Accurate transfer maps for realistic beam-line elements: straight elements, *Physical Review Special Topics – Accelerators and Beams* **13**, 064001.
- Month, M. and Messerschmid, E. (1977). A resistive theory of bunch lengthening, *IEEE Transactions on Nuclear Science* **NS-24**, 3, pp. 1208–1210.
- Muratori, B., Jones, J. K., and Wolski, A. (2015). Analytical expressions for fringe fields in multipole magnets, *Physical Review Special Topics – Accelerators and Beams* **18**, 064001.
- Nadolski, L. and Laskar, J. (2002). Application of a new class of symplectic integrators to accelerator tracking, in *Proceedings of the Eighth European Particle Accelerator Conference* (Paris, France), pp. 1276–1278.
- Nadolski, L. and Laskar, J. (2003). Review of single particle dynamics for third generation light sources through frequency map analysis, *Physical Review Special Topics – Accelerators and Beams* **6**, 114801.

- Nagaoka, R. and Wrulich, A. (2007). Emittance minimisation with longitudinal dipole field variation, *Nuclear Instruments and Methods in Physics Research Section A* **575**, 3, pp. 292–304.
- Neri, F. and Dragt, A. (1988). Analysis of tracking data using normal forms, in *Proceedings of the First European Particle Accelerator Conference* (Rome, Italy), pp. 675–677.
- Ng, K. Y. (1999). Explicit expressions of impedances and wake functions, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 203–208.
- Nielsen, C. E. and Sessler, A. M. (1959a). Longitudinal space charge effects in particle accelerators, *Review of Scientific Instruments* **30**, 2, pp. 80–89.
- Nielsen, C. E. and Sessler, A. M. (1959b). On the instability of a coasting beam, Tech. Rep. MURA-441, Midwestern Universities Research Association, Madison, Wisconsin.
- Ohmi, K. (2012). Beam-beam limit in a hadron collider, in *Proceedings of the Third International Particle Accelerator Conference* (New Orleans, LA, USA), pp. 3208–3212.
- Ohmi, K., Hirata, K., and Oide, K. (1994). From the beam-envelope matrix to synchrotron-radiation integrals, *Physical Review E* **49**, 1, pp. 751–765.
- Ohmi, K., Tawada, M., Cai, Y., Kamada, S., Oide, K., and Quiang, J. (2004). Beam-beam limit in  $e^+e^-$  circular colliders, *Physical Review Letters* **92**, 214801.
- Panofsky, W. K. H. and Wenzel, W. A. (1956). Some considerations concerning the transverse deflection of charged particles in radiofrequency fields, *Review of Scientific Instruments* **27**, 11, p. 967.
- Papaphilippou, Y. and Laskar, J. (1996). Frequency map analysis and global dynamics in a galactic potential with two degrees of freedom, *Astronomy and Astrophysics* **307**, pp. 427–449.
- Papaphilippou, Y. and Laskar, J. (1998). Global dynamics of triaxial galactic models through frequency map analysis, *Astronomy and Astrophysics* **329**, pp. 451–481.
- Pellegrini, C. (1969). On a new instability in electron–positron storage rings (the head-tail effect), *Il Nuovo Cimento A* **64**, 2, pp. 447–473.
- Pellegrini, C., Marinelli, A., and Reiche, S. (2016). The physics of x-ray free-electron lasers, *Reviews of Modern Physics* **88**, 015006.
- Piwinski, A. (1974). Intra-beam scattering, in *Proceedings of the 9th International Conference on High Energy Accelerators* (Stanford, CA, USA), pp. 405–408.
- Piwinski, A. (1998). The Touschek effect in strong focusing storage rings, Tech. Rep. DESY 98–179, DESY, Hamburg, Germany.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007a). *Numerical recipes: the art of scientific computing*, chap. 2.6 Singular value decomposition, 3rd edn. (Cambridge University Press, Cambridge, UK).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007b). *Numerical recipes: the art of scientific computing*, chap. 17.4 Second-order conservative equations, 3rd edn. (Cambridge University Press, Cambridge, UK).

- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007c). *Numerical recipes: the art of scientific computing*, chap. 17.1 Runge-Kutta method, 3rd edn. (Cambridge University Press, Cambridge, UK).
- Rao, Y. N. and Hermansson, L. (2000). Intra-beam scattering at CELSIUS, in *Proceedings of the Seventh European Particle Accelerator Conference* (Vienna, Austria), pp. 1549–1551.
- Raubenheimer, T. O. (1991). The generation and acceleration of low emittance flat beams for future linear colliders, Tech. Rep. SLAC-R-387, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Riemann, B. and Streun, A. (2019). Low emittance lattice design from first principles: reverse bending and longitudinal gradient bends, *Physical Review Accelerators and Beams* **22**, 021601.
- Robin, D., Steier, C., Laskar, J., and Nadolski, L. (2000). Global dynamics of the Advanced Light Source revealed through experimental frequency map analysis, *Physical Review Letters* **85**, 3, pp. 558–561.
- Robinson, K. W. (1958). Radiation effects in circular electron accelerators, *Physical Review* **111**, 2, pp. 373–380.
- Robinson, K. W. (1964). Stability of beam in radiofrequency system, Tech. Rep. CEAL-1010, CEA Saclay, Paris, France.
- Roqué, X. (1991). Møller scattering: a neglected application of early quantum electrodynamics, *Archive for History of Exact Sciences* **44**, 3, pp. 197–264.
- Rosenzweig, J. and Serafini, L. (1994). Transverse particle motion in radio-frequency linear accelerators, *Physical Review E* **49**, 2, pp. 1599–1602.
- Roser, T., Krisch, A. D., Barber, D. P., Ripken, G., and Yokoya, K. (1999). Polarization, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 148–165.
- Ruth, R. D. (1983). A canonical integration technique, *IEEE Transactions on Nuclear Science* **NS-30**, 4, pp. 2669–2671.
- Ruth, R. D. (1986a). Single particle dynamics and nonlinear resonances in circular accelerators, in *Nonlinear dynamics aspects of particle accelerators, Lecture Notes in Physics*, Vol. 247 (Springer, Berlin, Germany), pp. 37–63.
- Ruth, R. D. (1986b). Single particle dynamics in circular accelerators, Tech. Rep. SLAC-PUB-4103, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Sacherer, F. J. (1970). RMS envelope equations with space charge, Tech. Rep. CERN/SA/Int. DL/70-12, European Organization for Nuclear Research, Geneva, Switzerland.
- Sacherer, F. J. (1971). RMS envelope equations with space charge, in *Proceedings of the 1971 Particle Accelerator Conference* (Chicago, IL, USA), pp. 1105–1107.
- Sacherer, F. J. (1972). Methods for computing bunched-beam instabilities, Tech. Rep. CERN/SI-BR/72-5, European Organization for Nuclear Research, Geneva, Switzerland.
- Safranek, J. (1997). Experimental determination of storage ring optics using orbit response measurements, *Nuclear Instruments and Methods in Physics Research Section A* **388**, 1, pp. 27–36.
- Sagalovsky, L. (1989). Dipole fringe field optics in TRANSPORT, in *Proceedings of*

- the 1989 Particle Accelerator Conference (Chicago, IL, USA), pp. 1385–1387.
- Sagan, D. (2022). The Bmad reference manual (rev. November 6, 2022), <https://www.classe.cornell.edu/bmad/>.
- Sagan, D., Meller, R., Littauer, R., and Rubin, D. (2000). Betatron phase and coupling measurements at the Cornell Electron/Positron Storage Ring, *Physical Review Special Topics – Accelerators and Beams* **3**, 092801.
- Sagan, D. and Rubin, D. (1999). Linear analysis of coupled lattices, *Physical Review Special Topics – Accelerators and Beams* **2**, 074001.
- Saldin, E. L., Schneidmiller, E. A., and Yurkov, M. V. (2008). *The physics of free electron lasers* (Springer, Berlin, Germany).
- Sands, M. (1969). The head-tail effect: an instability mechanism in storage rings, Tech. Rep. SLAC-TN-69-8, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Sands, M. (1970). The physics of electron storage rings—an introduction, Tech. Rep. SLAC-121, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Schiff, L. I. (1946). Production of particle energies beyond 200 MeV, *Review of Scientific Instruments* **17**, p. 6.
- Schmüser, P., Dohlus, M., and Rossbach, J. (2008). *Ultraviolet and soft x-ray free-electron lasers* (Springer, Berlin, Germany).
- Schmüser, P., Dohlus, M., Rossbach, J., and Behrens, C. (2014). *Free-Electron Lasers in the Ultraviolet and X-Ray Regime* (Springer International Publishing, Cham, Switzerland).
- Schroer, C. G., Agapov, I., Brefeld, W., Brinkmann, R., Chul Chae, Y., Chun Chao, H., Eriksson, M., Keil, J., Gavaldà, X. N., Röhlsberger, R., Seeck, O. H., Sprung, M., Tischer, M., Wanzenberg, R., and Weckert, E. (2018). PETRA-IV: the ultralow-emittance source project at DESY, *Journal of Synchrotron Radiation* **25**.
- Schwinger, J. (1949). On the classical radiation of accelerated electrons, *Physical Review* **75**, 12, pp. 1912–1925.
- Seddon, E. A., Clarke, J. A., Dunning, D. J., Masciovecchio, C., Milne, C. J., Parmigiani, F., Rugg, D., Spence, J. C. H., Thompson, N. R., Ueda, K., Vinko, S. M., Wark, J. S., and Wurth, W. (2017). Short-wavelength free-electron laser sources and science: a review, *Reports on Progress in Physics* **80**, 115901.
- Shimosaki, Y., Soutome, K., Schimizu, J., Kaneki, K., Takao, M., Nakamura, T., and Ohkuma, H. (2011). Lattice design of a very low-emittance storage ring for SPring-8-II, in *Proceedings of the Second International Particle Accelerator Conference* (San Sebastian, Spain), pp. 942–944.
- Singh, B., Ghasem, H., Bengtsson, J., Streun, A., and Bartolini, R. (2018). Lattice options for DIAMOND-II, in *Proceedings of the Ninth International Particle Accelerator Conference* (Vancouver, BC, Canada), pp. 4050–4053.
- Skokos, C. and Papaphilippou, Y. (2008). Nonlinear dynamics study of the CLIC damping rings using symplectic integrators, in *Proceedings of the Eleventh European Particle Accelerator Conference* (Genoa, Italy), pp. 682–684.
- Slater, J. C. (1948). The design of linear accelerators, *Reviews of Modern Physics* **20**, 3, pp. 474–518.

- Sokolov, A. A. and Ternov, I. M. (1964). On polarization and spin effects in the theory of synchrotron radiation, *Soviet Physics: Doklady* **8**, 12, pp. 1203–1205.
- Steier, C., Atkinson, D., Byrd, J., Corlett, J., Nishimura, H., Robin, D., San-nibale, F., Wolski, A., Wu, Y., Bane, K., Raubenheimer, T., Ross, M., Sheppard, J., and Smith, T. (2001). Intra-beam scattering and minimum achievable emittance in the Advanced Light Source, in *Proceedings of the 2001 Particle Accelerator Conference* (Chicago, IL, USA), pp. 2938–2940.
- Steier, C., Robin, D., Nadolski, L., Decking, W., Wu, Y., and Laskar, J. (2002). Measuring and optimizing the momentum aperture in a particle accelerator, *Physical Review E* **65**, 056506.
- Stupakov, G. and Heifets, S. (2002). Beam instability and microbunching due to coherent synchrotron radiation, *Physical Review Special Topics – Accelerators and Beams* **5**, 054402.
- Takao, M. (2013). Beam dynamics and collective effects in “ultimate” storage rings, in *Proceedings of the Fourth International Particle Accelerator Conference* (Shanghai, China), pp. 3981–3985.
- Thomas, L. H. (1927). The kinematics of an electron with an axis, *Philosophical Magazine* **3**, pp. 1–22.
- Twiss, R. Q. and Frank, N. H. (1949). Orbital stability in a proton synchrotron, *Review of Scientific Instruments* **20**, p. 1.
- Valishev, A. (2008). Simulation of beam-beam effects and Tevatron experience, in *Proceedings of the Eleventh European Particle Accelerator Conference* (Genoa, Italy), pp. 2937–2941.
- Venturini, M. and Dragt, A. J. (1999). Accurate computation of transfer maps from magnetic field data, *Nuclear Instruments and Methods in Physics Research Section A* **427**, pp. 387–392.
- Wang, C. (2009). Minimum emittance in storage rings with uniform or nonuniform dipoles, *Physical Review Special Topics – Accelerators and Beams* **12**, 061001.
- Wang, G., Wei, Y., and Qiao, S. (2018). *Generalized inverses: theory and computations*, 2nd edn. (Springer, Singapore).
- Weiland, T. (1999). Impedance calculation, time domain, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 199–202.
- Wiedemann, H. (1999). Coherent radiation, in A. W. Chao and M. Tigner (eds.), *Handbook of accelerator physics and engineering* (World Scientific, Singapore), pp. 182–183.
- Wilson, P. B. (1978). Transient beam loading in electron–positron storage rings, Tech. Rep. SLAC–PEP–Note–276, Stanford Linear Accelerator Center, Stanford, CA, USA.
- Wilson, P. B. and Griffin, J. E. (1981). High energy electron linacs; application to storage ring rf systems and linear colliders, in R. A. Carrigan, M. Month, and F. R. Huson (eds.), *Proceedings of the 1981 Summer School on High Energy Particle Accelerators*, AIP Conference Proceedings 87 (Batavia, IL, USA), pp. 450–555.

- Wilson, P. B., Servranckx, R., Sabersky, A. P., Gareyte, J., Fischer, G. E., and Chao, A. W. (1977). Bunch lengthening and related effects in SPEAR-II, *IEEE Transactions on Nuclear Science* **NS-24**, 3, pp. 1211–1214.
- Winick, H. (1994). *Synchrotron radiation sources—a primer*, chap. 1 Introduction and overview (World Scientific, Singapore).
- Wolski, A. (2006). Alternative approach to general coupled linear optics, *Physical Review Special Topics – Accelerators and Beams* **9**, 024001.
- Wolski, A. (2013). Dynamic aperture, in A. W. Chao, K. H. Mess, M. Tigner, and F. Zimmermann (eds.), *Handbook of accelerator physics and engineering*, 2nd edn. (World Scientific, Singapore), pp. 111–115.
- Wolski, A., Christie, D. C., Militsyn, B. L., Scott, D. J., and Kockelbergh, H. (2020). Transverse phase space characterization in an accelerator test facility, *Physical Review Accelerators and Beams* **23**, 032804.
- Wolski, A. and Zimmermann, F. (2004). Closed orbit response to quadrupole strength variation, Tech. Rep. ATF-03-08, KEK High Energy Accelerator Research Organization, Tsukuba, Japan.
- Wu, Y. K., Forest, E., and Robin, D. S. (2003). Explicit symplectic integrator for  $s$ -dependent static magnetic field, *Physical Review E* **68**, 046502.
- Yan, Y. T. (1999). The use of Lie algebra methods to analyze and design accelerator beamlines, Lectures at the US Particle Accelerator School, Argonne, IL, USA.
- Yan, Y. T., Channell, P., and Syphers, M. (1992). An algorithm for symplectic implicit Taylor-map tracking, Tech. Rep. SSCL-Preprint-157, Superconducting Super Collider Laboratory, Dallas, TX, USA.
- Yang, P., Li, W., Ren, Z., Bai, Z., and Wang, L. (2021). Design of a diffraction-limited storage ring lattice using longitudinal gradient bends and reverse bends, *Nuclear Instruments and Methods in Physics Research Section A* **990**, 164968.
- Yokoya, K. (1986). The action-angle variables of classical spin motion in circular accelerators, Tech. Rep. DESY 86-057, DESY, Hamburg, Germany.
- Yoshida, H. (1990). Construction of higher order symplectic integrators, *Physics Letters A* **150**, pp. 262–268.
- Zangwill, A. (2013). *Modern electrodynamics* (Cambridge University Press, Cambridge, UK).
- Zholentz, A. A., Kurdadze, L. M., Lelchuk, M., Mishnev, S. I., Nikitin, S. A., Pakhtusova, E. V., Petrov, V. V., Protopopov, I., Saldin, E. L., Sidorov, V. A., Skrinsky, A. N., Tumaikin, G. M., Chilingarov, A. G., Shatunov, Y. M., Shwartz, B. A., Eidelman, S. I., and Eidelman, Y. I. (1980). High precision measurement of the  $\psi$ - and  $\psi'$ -meson masses, *Physics Letters B* **96**, 1, pp. 214–216.
- Zotter, B. W. and Kheifets, S. A. (1998). *Impedances and wakes in high-energy particle accelerators* (World Scientific, Singapore).

# Index

- achromat  
  double-bend (DBA), 277  
  multibend (MBA), 282
- action, 88
- action variable, 160, 205, 211
- action-angle variables, 162
- adiabatic damping, 140, 178
- Alfvén current, 464, 576
- angle variable, 161, 205
- anomalous magnetic moment, 142
- Arnold diffusion, 442
- Baker–Campbell–Hausdorff (BCH) formula, 370
- Bassetti–Erskine formula, 477
- beam loading, 135  
  fundamental theorem of, 548, 556
- beam rigidity, 115
- beam size (rms), 212
- beam-based alignment, 239
- beam-beam parameter, 501
- beta function, 158
- beta-beating, 176
- betatron function, 158
- betatron oscillation, 116, 161
- betatron phase advance, 167
- betatron resonance, 238, 399, 419  
  difference, 408  
  sum, 408
- betatron tune  
  measurement of, 477
- Bjorken–Mttingwa formulae, 522
- boundary condition  
  surface of ideal conductor, 11  
  surface of infinite permeability material, 8
- Boussard criterion, 595
- brightness (of a light source), 269
- broad-band resonator impedance, 595
- bunch compressor, 313
- bunch length, natural, 269
- Butcher tableau, 383
- canonical momentum, 87
- canonical perturbation theory, 394
- canonical transformation, 92
- canonical variables, 92
- capacitive impedance, 563
- characteristic current, 464, 576
- charge density, 5
- Chasman–Green cell, 277
- chicane, 315
- Chirikov criterion, 428
- chromaticity, 128, 323  
  FODO cell, 325  
  head-tail instability, 608  
  quadrupole magnet, 328  
  sextupole magnet, 332  
  storage ring, 332
- closed orbit, 184, 228, 403
- coasting beam, 588
- coherent instabilities, 453

- coherent synchrotron radiation, 248, 285
- coherent tune shift, 477
- combined function dipole, 25, 117
- commutator (of Lie operators), 369
- cooling, 104
- Coulomb log, 523, 525
- coupled-bunch feedback system, 577
- Courant–Snyder parameters, 157, 201
- Cremona map, 339
- critical frequency, 264
- current density, 5
- damping
  - adiabatic, 140, 178
  - Landau, 597
  - partition number, 257
  - synchrotron radiation, 252
- DBA cell, *see* double-bend achromat cell
- decoherence, 577
- difference resonance, 408
- diffraction limit, 270, 283
- diffraction model impedance, 630
- dipole magnet
  - combined function, 117
  - field, 14
  - fringe field, 52, 121
    - transfer map, 124
  - pole face rotation, 121
  - transfer map, 119
- dispersion, 115, 184
  - and rms beam size, 213
- dispersion  $\mathcal{H}$ -function, 264, 519
- dispersion relation, 590
- double-bend achromat (DBA) cell, 277
  - lattice functions, 277
  - natural emittance, 279
- Dragt–Finn factorisation, 435
- drift space
  - Hamiltonian, 109
  - transfer map, 110
  - transfer matrix, 112
- drift-kick-drift approximation, 372
- dynamic aperture, 429
- edge focusing, 125
- Edwards–Teng coupling formalism, 201
- electric displacement, 4
- electric field, 4
- elliptic integral, 51, 525
- emittance, 104, 175, 207
  - and rms beam size, 212
  - geometric, 178
  - longitudinal, 261
  - measurement, 219
  - natural, 265
  - normalised, 178
  - radiation beam, 284
- emittance dominated beam, 465
- energy acceptance, 195, 430, 506
  - dynamic, 512
  - measurement, 512
  - rf, 197
- energy deviation, 95
- energy loss per turn, 252
- energy spread
  - and rms beam size, 213
  - natural, 268
- Enge function
  - dipole fringe field, 53
  - multipole fringe field, 40, 64
  - solenoid fringe field, 41
- envelope equation, 462, 463
  - with space charge, 464, 473
- explicit midpoint method, 364
- fast feedback system, 577
- fast head-tail instability, 607
- FEL, 248, 284
- first-generation light source, 283
- flat-beam electron source, 215
- Floquet transformation, 172, 464
- FMA, *see* frequency map analysis
- FODO cell, 152
  - chromaticity, 325
  - lattice functions, 272
  - natural emittance, 274
  - transfer matrix, 154
- free-electron laser, 248, 284
- frequency map analysis (FMA), 446

- fringe field, 33, 39  
 dipole, 52, 123  
 multipole magnet, 39, 48  
 solenoid, 40, 45, 131  
 fundamental theorem of beam loading, 548, 556
- gauge transformation, 16  
 Gauss methods, 386  
 Gauss' theorem, 10  
 generalised gradient, 43, 44  
 geometric wake field, 529  
 gyromagnetic ratio, 142
- $\mathcal{H}$ -function, 264, 519  
 Haissinski equation, 490, 496, 586  
 Hamilton's equations, 85  
 Hamiltonian, 85  
   combined function dipole, 118  
   curved co-ordinate system, 99  
   dipole magnet, 114  
   drift space, 109  
   quadrupole magnet, 126  
   sextupole magnet, 333, 347, 367  
   skew quadrupole magnet, 129  
   synchrotron motion, 196, 584, 617  
 harmonic number, 198, 581, 625  
 head-tail instability, 608  
 higher-order mode heating, 565  
 Hill's equation, 171  
 hourglass effect, 502
- impedance, 528  
   broad-band resonator, 595  
   capacitive, 563  
   diffraction model, 630  
   inductive, 563  
   LCR circuit model, 562  
   longitudinal, 558  
   properties, 564  
   resistive-wall, 561  
   resonant cavity, 562  
   space-charge, 495, 563  
   transverse, 559  
 $Z_{\parallel}/n$ , 561, 596  
 incoherent collective effects, 453
- incoherent synchrotron radiation, 248  
 incoherent tune shift, 477  
 inductive impedance, 563  
 insertion device, 7, 33, 283  
 instability  
   coupled-bunch, 569  
   fast head-tail, 607  
   head-tail, 608  
   microwave, 594  
   mode-coupling, 614, 632  
   negative-mass, 498, 591  
   resistive-wall, 575  
   Robinson, 582, 626  
   single-bunch, 569  
 interaction matrix, 621, 629  
 intrabeam scattering (IBS), 503  
   growth rates, 516  
     Bjorken–Mtingwa formulae, 522  
     high energy approximations, 524  
     Piwinski formulae, 517  
 invariant spin field, 150  
 invariant tori, 429  
 isochronous beam line, 191
- Jacobi identity, 369  
 Jonqui re function, 63
- KAM theorem, *see*  
   Kolmogorov–Arnold–Moser theorem  
 Kapchinsky–Vladimirsky distribution, 455  
   emittance, 457  
 Keil–Schnell criterion, 594  
 Keil–Schnell–Boussard criterion, 595  
 kick factor, 568  
 kick-drift-kick approximation, 372  
 Kolmogorov–Arnold–Moser (KAM) theorem, 429  
 KV distribution, *see*  
   Kapchinsky–Vladimirsky distribution
- Landau damping, 597

- Larmor precession, 143, 305
- Laslett coefficients, 483
- Laslett tune shift, 478
- leap-frog method, 360
- Lie algebra, 370
- Lie operator, 340
- Lie transformation, 341
  - Baker–Campbell–Hausdorff (BCH) formula, 370
  - Zassenhaus formula, 370
- lifetime, 511
- Liouville's theorem, 103, 176
- longitudinal impedance, 558
- longitudinal wake function, 557
- Lorentz force, 4, 454
- Lorentz transformation of electromagnetic field, 459
- loss factor, 566
- luminosity, 270, 501
- macroparticle, 569, 598
- magnetic axis, 126
- magnetic field, 4
- magnetic intensity, 4
- magnetic moment, 140, 141
- magnetic monopole field components, 45
- magnetic multipole
  - see multipole, 14
- magnetic scalar potential, 43
- matched distribution, 206, 462
- Maxwell's equations, 4
- MBA cell, *see* multibend achromat cell
- microwave instability, 594
  - dispersion relation, 590
- midpoint method, 364
- mixed-variable generating function, 93, 98, 339, 351, 396
  - for action–angle and Cartesian variables, 166
- mode coupling, 614, 632
- Møller scattering, 504
- momentum
  - canonical, 87
  - reference, 92
- momentum compaction factor, 187, 190, 195, 511, 521
  - higher-order, 188
- momentum deviation, 184
- multibend achromat (MBA) cell, 279, 282
- multipole
  - allowed, 28
  - coefficients  $a_n$  and  $b_n$ , 16
  - decomposition, 27
  - errors, 27
  - field, 14
  - forbidden, 28
  - fringe field, 39
  - moment, 538
  - normal, 15
  - potential, 16
  - skew, 15
- muon  $g - 2$  experiment, 140
- NAFF, *see* numerical analysis of the fundamental frequencies
- natural bunch length, 269
- natural emittance, 265
- natural energy spread, 268
- negative-mass instability, 498, 591
- normal form, 431
- normalising matrix, 164, 167, 203
- numerical analysis of the fundamental frequencies (NAFF), 443
- orbit response matrix (ORM)
  - analysis, 234
- Panofsky–Wenzel theorem, 555, 560
- parasitic loss, 565
- paraxial approximation, 108, 378
- permeability, 4
- permittivity, 4
- perturbation theory, 394
- perveance, 463, 469, 473, 475, 481, 486
- phase advance, 167, 206
- phase slip, 116, 200
- phase slip factor, 189, 193, 195, 199, 259, 269, 488, 521, 584, 588, 596,

- 610, 614, 620, 633
- and transition, 190
- higher-order, 189
- phase space, 88
  - portrait, 334
  - tomography, 220
- phase stability, 197
- Piwinski formulae, 517
- Poincaré section, 334
- Poisson bracket, 344, 370
  - canonical invariance of, 345
- polylogarithm, 63
- potential-well distortion, 584
- Poynting vector, 73
- quadrupole magnet
  - alignment error, 235
  - field, 14, 126
  - gradient error, 236
  - Hamiltonian, 126
  - transfer map, 127, 353
- quality factor, 75, 535
- quantum radiation constant  $C_q$ , 265
- reference
  - momentum, 92
  - particle, 94
  - trajectory, 96
- resistive-wall
  - impedance, 561
  - instability, transverse growth rate, 575
  - wake fields, 546
  - wake function, 554
- resonance, 399, 419
  - difference, 408
  - sum, 408
  - synchro-betatron, 419
- resonance diagram, 421
- resonant depolarisation, 140, 309, 514
- rf (radiofrequency) acceptance, 197
- rf bucket, 197
- rf cavity, 67
  - field modes
    - cylindrical cavity, 77, 134
    - rectangular cavity, 71
- Hamiltonian, 137
- quality factor ( $Q$ ), 75
- shunt impedance, 76
- stored energy, 75
- transfer map, 137
- transit-time factor, 136
- Robinson damping theorem, 262
- Robinson instability, 582, 626
- Runge–Kutta integration, 382
  - explicit midpoint method, 364
  - explicit schemes, 383
  - fourth-order symplectic, 387
  - Gauss methods, 386
  - implicit midpoint rule, 385
  - implicit schemes, 384
  - midpoint method, 384
  - RK4, 384
  - sixth-order symplectic, 387
  - symplectic, 385
- SABA<sub>2</sub>, SABA<sub>2</sub>C integrators, 375
- Sacherer’s integral equation, 621
- scalar potential, magnetic, 43
- scaling law (for bunch length), 633
- second-generation light source, 283
- separatrix, 197, 426
- sextupole magnet
  - field, 14
  - Hamiltonian, 347
  - transfer map
    - drift-kick-drift approximation, 373
    - Lie transformation, 347
    - mixed-variable generating function, 357
    - power series, 348
- shunt impedance, 76
- Siberian snake, 148
- Sigma matrix ( $\Sigma$ ), 201, 206, 298
- similarity transformation, 343
- skin depth, 543, 548
- Sokolov–Ternov effect, 306, 514
- solenoid
  - field, 14, 40, 46, 49, 130
  - flat-beam electron source, 215
  - fringe field, 40, 47, 49, 131

- Hamiltonian, 132
- matched lattice functions, 214
- transfer map, 133, 214, 484
- space-charge
  - impedance, 495, 563
  - geometry factor, 493, 496
  - Laslett coefficients, 483
  - longitudinal tune shift, 497
  - perveance, 463
- space-charge dominated beam, 465
- SPEAR scaling, 633
- spin, 140
  - precession, 140
  - resonant depolarisation, 140, 309, 514
- spin polarisation, 146
- spin tune, 145, 514
- splitting method, 373
- Stokes' theorem, 9
- Störmer–Verlet method, 360
- sum resonance, 408
- symmetric factorisation, 373
- symplectic integrator, 342, 360
  - fourth-order, 373
  - Runge–Kutta, 385
  - SABA<sub>2</sub>, SABA<sub>2</sub>C, 375
  - second-order, 372
  - splitting method, 373
  - Störmer–Verlet method, 359
  - symmetric factorisation, 373
  - Wu–Forest–Robin, 377
  - Yoshida factorisation, 373
- symplectic matrix, 102
  - eigenvalues, 202
  - eigenvectors, 202
- symplectic transformation, 101
- synchronous phase, 194, 259
- synchrotron frequency, 199, 260
- synchrotron light source, 269, 283
  - brightness, 269
  - first, second and third generation, 283
- synchrotron motion Hamiltonian, 196
- synchrotron oscillations, 116, 192
- synchrotron phase advance, 167, 199
- synchrotron radiation
  - brightness, 269
  - coherent, 248
  - constant  $C_\gamma$ , 253
  - critical frequency, 264
  - energy loss per turn, 252, 253
  - horizontal damping partition number, 257
  - horizontal damping time, 257
  - incoherent, 248
  - intensity spectrum, 264
  - longitudinal damping partition number, 261
  - longitudinal damping time, 261
  - quantum constant  $C_q$ , 265
  - vertical damping time, 252
- synchrotron radiation integrals
  - $I_1$ , 188
  - $I_2$ , 253
  - $I_3$ , 268
  - $I_4$ , 257
  - $I_5$ , 265
- synchrotron tune, 194, 260
- TE and TM modes, 79
- theoretical minimum emittance (TME) cell, 279
  - lattice functions, 280
  - natural emittance, 280
- thin lens approximation, 153
- third-generation light source, 283
- Thomas–BMT equation, 141, 305, 514
- TME cell, *see* theoretical minimum emittance cell
- tomography
  - phase space, 220
- Touschek effect, 504
  - multiple, 515
- Touschek lifetime, 510
- TPSA, 339
- transfer map, 100
  - change in reference momentum, 139
  - dipole, 119
  - dipole fringe field, 124
  - drift space, 110

- quadrupole magnet, 129, 353  
rf cavity, 137  
sextupole magnet, 348, 349  
sextupole, thin lens, 333  
solenoid fringe field, 132  
synchrotron radiation, 302  
transfer matrix  
  Courant–Snyder parameters, 158,  
    166, 211  
  dipole, 119  
  dipole fringe field, 124  
  drift space, 112  
  FODO cell, 154  
  quadrupole magnet, 127  
  rf cavity, 137  
  solenoid, 133, 484  
transit-time factor, 136  
transition, 187  
  Robinson instability, 582  
transition energy, 187  
TRANSPORT notation, 319  
transverse impedance, 559  
transverse loss factor, 567  
transverse wake function, 557  
truncated power series, 338  
  TPSA, 339  
tune grid, 421  
tune shift  
  beam-beam, 500  
  coherent, 477  
  focusing error, 238, 404  
  incoherent, 477  
  Laslett, 478  
longitudinal space-charge, 497  
sextupole, 417  
skew quadrupole, 242, 411  
space-charge, 474  
  with amplitude, 396, 438, 446  
tune spread, 396, 477  
turbulent bunch lengthening, 598  
Twiss parameters, 158  
  
undulator magnet, 7, 33  
  field, 35, 376  
  
Verlet integration, 360  
Vlasov equation, 589, 614  
  
wake field, 527  
  geometric, 529  
  resistive-wall, 537  
  resonant cavity, 536  
wake function, 527, 554  
  longitudinal, 557  
  properties, 555  
  resistive-wall, 554  
  resonant cavity, 554  
  transverse, 557  
  units, 555  
wake potential, 555  
water-bag distribution, 624, 629  
weak focusing, 116  
wiggler magnet, 7, 33  
  field, 35, 376  
working point, 419  
Wu–Forest–Robin integrator, 377  
  
Yoshida factorisation, 373  
  
 $Z_{\parallel}/n$ , 561, 596  
  broad-band resonator, 596  
  Keil–Schnell criterion, 594  
Zassenhaus formula, 370