

Molecular Modelling for Beginners

Second Edition

ALAN HINCHLIFFE

The University of Manchester

 **WILEY**

A John Wiley and Sons, Ltd, Publication

This edition first published 2008
© 2008 John Wiley & Sons Ltd

Registered office

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex,
PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

The publisher and the author make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of fitness for a particular purpose. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for every situation. In view of ongoing research, equipment modifications, changes in governmental regulations, and the constant flow of information relating to the use of experimental reagents, equipment, and devices, the reader is urged to review and evaluate the information provided in the package insert or instructions for each chemical, piece of equipment, reagent, or device for, among other things, any changes in the instructions or indication of usage and for added warnings and precautions. The fact that an organization or Website is referred to in this work as a citation and/or a potential source of further information does not mean that the author or the publisher endorses the information the organization or Website may provide or recommendations it may make. Further, readers should be aware that Internet Websites listed in this work may have changed or disappeared between when this work was written and when it is read. No warranty may be created or extended by any promotional statements for this work. Neither the publisher nor the author shall be liable for any damages arising herefrom.

Library of Congress Cataloging in Publication Data

Hinchliffe, Alan.

Molecular modelling for beginners / Alan Hinchliffe. — 2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-51313-2 (cloth) — ISBN 978-0-470-51314-9 (pbk. : alk. paper)

1. Molecules—Mathematical models. 2. Molecules—Computer simulation.

I. Title.

QD480.H58 2008

541'.22015118—dc22

2008024099

British Library Cataloguing in Publication Data

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

CLOTH: 978-0-470-51313-2

PAPER: 978-0-470-51314-9

Set in 10/12pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India

Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire

Contents

Preface to the Second Edition	xv
Preface to the First Edition	xvii
1 Electric Charges and their Properties	1
1.1 Point Charges	1
1.2 Coulomb's Law	3
1.3 Pair Wise Additivity	4
1.4 Electric Field	5
1.5 Work	6
1.6 Charge Distributions	8
1.7 Mutual Potential Energy, U	8
1.8 Relationship between Force and Mutual Potential Energy	9
1.9 Electric Multipoles	10
1.9.1 <i>Continuous Charge Distributions</i>	12
1.9.2 <i>Electric Second Moment</i>	13
1.9.3 <i>Higher Electric Moments</i>	15
1.10 Electrostatic Potential	15
1.11 Polarization and Polarizability	16
1.12 Dipole Polarizability	17
1.12.1 <i>Properties of Polarizabilities</i>	18
1.13 Many-Body Forces	19
1.14 Problem Set	20
2 The Forces between Molecules	21
2.1 Pair Potential	21
2.2 Multipole Expansion	23
2.3 Charge–Dipole Interaction	23
2.4 Dipole–Dipole Interaction	25
2.5 Taking Account of the Temperature	26
2.6 Induction Energy	27
2.7 Dispersion Energy	28
2.8 Repulsive Contributions	29
2.9 Combination Rules	31
2.10 Comparison with Experiment	32
2.10.1 <i>Gas Imperfections</i>	32
2.10.2 <i>Molecular Beams</i>	32

2.11	Improved Pair Potentials	33
2.12	A Numerical Potential	33
2.13	Site–Site Potentials	34
2.14	Problem Set	35
	References	35
3	Balls on Springs	37
3.1	Vibrational Motion	38
3.2	The Force Law	41
3.3	A Simple Diatomic	41
3.4	Three Problems	43
3.5	The Morse Potential	45
3.6	More Advanced Potentials	46
	References	47
4	Molecular Mechanics	49
4.1	More About Balls on Springs	49
4.2	Larger Systems of Balls on Springs	51
4.3	Force Fields	53
4.4	Molecular Mechanics (MM)	53
4.4.1	<i>Bond Stretching</i>	54
4.4.2	<i>Bond Bending</i>	54
4.4.3	<i>Dihedral Motions</i>	55
4.4.4	<i>Out-of-Plane Angle Potential (Inversion)</i>	56
4.4.5	<i>Nonbonded Interactions</i>	56
4.4.6	<i>Coulomb Interactions</i>	57
4.5	Modelling the Solvent	57
4.6	Time- and Money-Saving Tricks	57
4.6.1	<i>United Atoms</i>	57
4.6.2	<i>Cut-Offs</i>	58
4.7	Modern Force Fields	58
4.7.1	<i>Variations on a Theme</i>	59
4.8	Some Commercial Force Fields	60
4.8.1	<i>Dreiding</i>	60
4.8.2	<i>MM1</i>	60
4.8.3	<i>MM2 (Improved Hydrocarbon Force Field)</i>	61
4.8.4	<i>AMBER</i>	61
4.8.5	<i>OPLS (Optimized Potentials for Liquid Simulations)</i>	62
4.8.6	<i>Johnson</i>	62
4.8.7	<i>Unified Force Field (UFF)</i>	63
	References	63
5	The Molecular Potential Energy Surface	65
5.1	Multiple Minima	65
5.2	Saddle Points	66
5.3	Characterization	68

5.4	Finding Minima	68
5.5	Multivariate Grid Search	69
	5.5.1 <i>Univariate Search</i>	69
5.6	Derivative Methods	70
5.7	First-Order Methods	70
	5.7.1 <i>Steepest Descent</i>	70
	5.7.2 <i>Conjugate Gradients</i>	71
5.8	Second-Order Methods	72
	5.8.1 <i>Newton–Raphson</i>	72
	5.8.2 <i>Block Diagonal Newton–Raphson</i>	75
	5.8.3 <i>Quasi-Newton–Raphson</i>	75
	5.8.4 <i>Fletcher–Powell Algorithm</i>	75
5.9	Choice of Method	76
5.10	The Z-Matrix	77
5.11	The End of the Z-Matrix	78
5.12	Redundant Internal Coordinates	80
	References	80
6	Molecular Mechanics Examples	81
6.1	Geometry Optimization	81
6.2	Conformation Searches	82
6.3	Amino Acids	84
	6.3.1 <i>Molecular Structure Databases</i>	85
	6.3.2 <i>The .pdb File Format</i>	85
6.4	QSAR	86
	6.4.1 <i>Atomic Partial Charges</i>	86
	6.4.2 <i>Polarizabilities</i>	87
	6.4.3 <i>Molecular Volume and Surface Area</i>	90
	6.4.4 <i>Log(P)</i>	92
6.5	Problem Set	93
	References	93
7	Sharing Out the Energy	95
7.1	Games of Chance	97
	7.1.1 <i>Two Dice</i>	99
7.2	Enumeration	99
7.3	Boltzmann Probability	100
7.4	Safety in Numbers	104
7.5	Partition Function	106
7.6	Two-Level Quantum System	107
7.7	Lindemann’s Theory of Melting	109
7.8	Problem Set	111
8	Introduction to Statistical Thermodynamics	113
8.1	The Ensemble	114
8.2	Internal Energy, U_{th}	116

8.3	Helmholtz Energy, A	116
8.4	Entropy S	117
8.5	Equation of State and Pressure	117
8.6	Phase Space	117
8.7	Configurational Integral	119
8.8	Virial of Clausius	120
9	Monte Carlo Simulations	123
9.1	An Early Paper	125
9.2	The First ‘Chemical’ Monte Carlo Simulation	127
9.3	Importance Sampling	128
9.4	Periodic Box	129
9.5	Cut-Offs	131
9.6	MC Simulation of Rigid Molecules	132
9.7	Flexible Molecules	133
	References	133
10	Molecular Dynamics	135
10.1	Radial Distribution Function	136
10.2	Pair Correlation Functions	139
10.3	Molecular Dynamics Methodology	140
	10.3.1 <i>Hard Sphere Potential</i>	140
	10.3.2 <i>Finite Square Well</i>	140
	10.3.3 <i>Lennard-Jonesium</i>	142
10.4	Algorithms for Time Dependence	142
	10.4.1 <i>Leapfrog Algorithm</i>	142
	10.4.2 <i>Verlet Algorithm</i>	143
10.5	Molten Salts	144
10.6	Liquid Water	144
	10.6.1 <i>Other Water Potentials</i>	147
10.7	Different Types of Molecular Dynamics	147
10.8	Uses in Conformational Studies	148
	References	150
11	Introduction to Quantum Modelling	151
11.1	The Schrödinger Equation	151
11.2	The Time-Independent Schrödinger Equation	153
11.3	Particles in Potential Wells	154
	11.3.1 <i>One-Dimensional Infinite Well</i>	154
11.4	Correspondence Principle	156
11.5	Two-Dimensional Infinite Well	157
11.6	Three-Dimensional Infinite Well	159
11.7	Two Noninteracting Particles	161
11.8	Finite Well	162
11.9	Unbound States	164

11.10	Free Particles	164
11.11	Vibrational Motion	165
12	Quantum Gases	169
12.1	Sharing Out the Energy	170
12.2	Rayleigh Counting	172
12.3	Maxwell–Boltzmann Distribution of Atomic Kinetic Energies	173
12.4	Black Body Radiation	175
12.5	Modelling Metals	177
	12.5.1 <i>The Drude Model</i>	178
	12.5.2 <i>The Pauli Treatment</i>	180
12.6	Indistinguishability	181
12.7	Spin	185
12.8	Fermions and Bosons	187
12.9	Pauli Exclusion Principle	187
12.10	Boltzmann’s Counting Rule	188
	References	189
13	One-Electron Atoms	191
13.1	Atomic Spectra	192
	13.1.1 <i>Bohr’s Theory</i>	192
13.2	Correspondence Principle	194
13.3	Infinite Nucleus Approximation	194
13.4	Hartree’s Atomic Units	195
13.5	Schrödinger Treatment of the Hydrogen Atom	195
13.6	Radial Solutions	197
13.7	Atomic Orbitals	199
	13.7.1 <i>$l = 0$ (<i>s</i>-orbitals)</i>	200
	13.7.2 <i>The <i>p</i>-orbitals</i>	202
	13.7.3 <i>The <i>d</i>-orbitals</i>	205
13.8	The Stern–Gerlach Experiment	205
13.9	Electron Spin	208
13.10	Total Angular Momentum	209
13.11	Dirac Theory of the Electron	210
13.12	Measurement in the Quantum World	212
	References	213
14	The Orbital Model	215
14.1	One- and Two-Electron Operators	216
14.2	Many-Body Problem	216
14.3	Orbital Model	217
14.4	Perturbation Theory	219
14.5	Variation Method	221
14.6	Linear Variation Method	223
14.7	Slater Determinants	226
14.8	Slater–Condon–Shortley Rules	228

14.9	Hartree Model	229
14.10	Hartree–Fock Model	230
14.11	Atomic Shielding Constants	231
	14.11.1 Zener’s Wavefunctions	232
	14.11.2 Slater’s Rules	233
14.12	Koopmans’ Theorem	234
	References	235
15	Simple Molecules	237
15.1	Hydrogen Molecule Ion, H_2^+	238
15.2	LCAO Model	240
15.3	Elliptic Orbitals	242
15.4	Heitler–London Treatment of Dihydrogen	244
15.5	Dihydrogen MO Treatment	246
15.6	James and Coolidge Treatment	247
15.7	Population Analysis	247
	15.7.1 Extension to Many-Electron Systems	249
	References	250
16	The HF-LCAO Model	253
16.1	Roothaan’s 1951 Landmark Paper	254
16.2	The \hat{J} and \hat{K} Operators	256
16.3	HF-LCAO Equations	256
	16.3.1 The HF-LCAO Equations	259
16.4	Electronic Energy	259
16.5	Koopmans’ Theorem	260
16.6	Open Shell Systems	260
16.7	Unrestricted Hartree–Fock (UHF) Model	262
	16.7.1 Three Technical Points	263
16.8	Basis Sets	264
	16.8.1 Clementi and Raimondi	264
	16.8.2 Extension to Second-Row Atoms	265
	16.8.3 Polarization Functions	266
16.9	Gaussian Orbitals	267
	16.9.1 STO/nG	270
	16.9.2 STO/4-31G	271
	16.9.3 Gaussian Polarization and Diffuse Functions	272
	16.9.4 Extended Basis Sets	273
	References	274
17	HF-LCAO Examples	275
17.1	Output	277
	17.1.1 GaussView, the Graphical User Interface for Gaussian 03	280
17.2	Visualization	281
17.3	Properties	282
	17.3.1 Electrostatic Potential	283

17.4	Geometry Optimization	285
17.5	Vibrational Analysis	287
	17.5.1 <i>Hellmann Feynman Theorem</i>	290
17.6	Thermodynamic Properties	291
	17.6.1 <i>Ideal Monatomic Gas</i>	292
	17.6.2 <i>Ideal Diatomic Gas</i>	293
	17.6.3 q_{rot}	293
	17.6.4 q_{vib}	294
17.7	Back to L-Phenylalanine	295
17.8	Excited States	296
17.9	Consequences of the Brillouin Theorem	300
17.10	Electric Field Gradients	301
17.11	Hyperfine Interactions	303
17.12	Problem Set	305
	References	306
18	Semiempirical Models	307
18.1	Hückel π -Electron Theory	308
	18.1.1 <i>Treatment of Heteroatoms</i>	310
18.2	Extended Hückel Theory	311
	18.2.1 <i>Wolfsberg and Helmholtz</i>	312
	18.2.2 <i>Roald Hoffmann</i>	314
18.3	Pariser, Parr and Pople	314
18.4	Zero Differential Overlap	315
18.5	Which Basis Functions Are They?	317
18.6	All Valence Electron ZDO Models	318
18.7	CNDO	318
18.8	CNDO/2	319
18.9	CNDO/S	320
18.10	INDO	320
18.11	NDDO (Neglect of Diatomic Differential Overlap)	321
18.12	The MINDO Family	321
	18.12.1 <i>MINDO/3</i>	321
18.13	MNDO	322
18.14	Austin Model 1 (AM1)	323
18.15	PM3	323
18.16	SAM1	323
18.17	ZINDO/1 and ZINDO/S	323
18.18	Effective Core Potentials	324
18.19	Problem Set	324
	References	324
19	Electron Correlation	327
19.1	Electron Density Functions	327
	19.1.1 <i>Fermi Correlation</i>	329

19.2	Configuration Interaction	329
19.3	Coupled Cluster Method	330
19.4	Møller–Plesset Perturbation Theory	331
19.5	Multiconfiguration SCF	335
	References	335
20	Density Functional Theory and the Kohn–Sham LCAO Equations	337
20.1	Pauli and Thomas–Fermi Models	338
20.2	Hohenberg–Kohn Theorems	340
20.3	Kohn–Sham (KS-LCAO) Equations	342
20.4	Numerical Integration (Quadrature)	343
20.5	Practical Details	343
20.6	Custom and Hybrid Functionals	345
20.7	An Example	345
	References	347
21	Accurate Thermodynamic Properties; the G_n Models	349
21.1	G_1 Theory	349
21.2	G_2 Theory	351
	21.2.1 $G_2(MP2)$	352
21.3	G_3 Theory	352
	21.3.1 G_3 Variations	353
	References	353
22	Transition States	355
22.1	An Example	357
22.2	The Reaction Path	359
	References	361
23	Dealing with the Solvent	363
23.1	Solvent Models	363
23.2	Langevin Dynamics	364
23.3	Continuum Solvation Models	366
	23.3.1 <i>The Onsager Model</i>	368
	23.3.2 <i>Polarizable Continuum Models</i>	370
23.4	Periodic Solvent Box	370
	References	371
24	Hybrid Models; the QM/MM Approach	373
24.1	Link Atoms	374
24.2	IMOMM	375
24.3	IMOMO	376
24.4	ONIOM (Our Own N -layered Integrated Molecular Orbital and Molecular Mechanics)	377
	References	377

Appendix A	A Mathematical <i>Aide-Mémoire</i>	379
Appendix B	Glossary	399
Appendix C	List of Symbols	401
Index		405

Preface to the Second Edition

It is five years since the first edition was published, and many things have moved on sufficiently to justify this second edition.

Some things never change; I have left the elementary chapters alone and I still believe that Appendix A on relevant mathematical methods is the correct place for you to *start* your studies.

Some topics have matured in the last five years. Density functional theory (and especially the B3LYP choice of functionals) has become the workhorse of modern computational chemistry. I have reworked all the problems and expanded the text as appropriate.

I have also said 'goodbye' to a few of the older topics. For example, everyone can now do chemical drawing, so I do not need to teach it. Thankfully that bane of our lives the Z-matrix has all but disappeared; I still have fond memories of struggling to get cyclic structures symmetrical and so it still gets a page of discussion.

I have completely rewritten the chapters dealing with Monte Carlo and molecular dynamics, the G_n models, transition states and solvent models. I have also added a completely new chapter called 'Sharing Out the Energy', and I hope you will enjoy reading it.

It is fashionable to have an associated website with any new teaching text, and I have therefore added a website at

<http://www.wileyurope.com/college/hinchliffe>

where you will find a number of problem sets and their solutions. Feel free to use them any way you like. I used them in my own teaching. Perhaps you have a corresponding set that you would like to share with the rest of us? Let me know.

I did all the illustrative calculations using either Gaussian 03 or HyperChem; these were done either on a beautiful Sony Vaio laptop or on the University of Manchester's High Performance Computing parallel computer, a Bull Itanium2 system.

As always, I welcome comments and can be reached at: Alan.Hinchliffe@manchester.ac.uk.

Alan Hinchliffe
Manchester, UK

Preface to the First Edition

There is nothing radically new about the techniques we use in modern molecular modelling. Classical mechanics hasn't changed since the time of Newton, Hamilton and Lagrange, the great ideas of statistical mechanics and thermodynamics were discovered by Ludwig Boltzmann and J. Willard Gibbs amongst others and the basic concepts of quantum mechanics appeared in the 1920s, by which time J.C. Maxwell's famous electromagnetic equations had long since been published.

The chemically inspired idea that molecules can profitably be treated as a collection of balls joined together with springs can be traced back to the work of D.H. Andrews in 1930. The first serious molecular Monte Carlo simulation appeared in 1953, closely followed by B.J. Alder and T.E. Wainwright's classic molecular dynamics study of hard discs in 1957.

The Hartrees' 1927 work on atomic structure is the concrete foundation of our everyday concept of atomic orbitals, whilst C.C.J. Roothaan's 1951 formulation of the HF-LCAO model arguably gave us the basis for much of modern molecular quantum theory.

If we move on a little, most of my colleagues would agree that the two recent major advances in molecular quantum theory have been density functional theory, and the elegant treatment of solvents using ONIOM. Ancient civilizations believed in the cyclic nature of time and they might have had a point for, as usual, nothing is new. Workers in solid-state physics and biology actually proposed these models many years ago. It took the chemists a while to catch up.

Scientists and engineers first got their hands on computers in the late 1960s. We have passed the point on the computer history curve where every ten years gave us an order of magnitude increase in computer power, but it is no coincidence that that growth in our understanding and application of molecular modelling has run in parallel with growth in computer power. Perhaps the two greatest driving forces in recent years have been the PC and the graphical user interface. I am humbled by the fact that my lowly 1.2 GHz AMD Athlon office PC is far more powerful than the world-beating mainframes that I used as a graduate student all those years ago, and that I can build a molecule on screen and run a B3LYP/6-311++G(3d,2p) calculation before my eyes (of which more in Chapter 20).

We have also reached a stage where tremendously powerful molecular modelling computer packages are commercially available, and the subject is routinely taught as part of undergraduate science degrees. I have made use of several such packages to produce the screenshots; obviously they look better in colour than the greyscale of this text.

There are a number of classic (and hard) texts in the field; if I'm stuck with a basic molecular quantum mechanics problem, I usually reach for Eyring, Walter and Kimball's *Quantum Chemistry* but the going is rarely easy.

Equally there are a number of beautifully produced elementary texts and software reference manuals that can apparently transform you into an expert overnight. It's a two-edged sword, and we are victims of our own success. One often meets self-appointed experts in the field who have picked up much of the jargon with little of the deep understanding. It's no use (in my humble opinion) trying to hold a conversation about gradients, Hessians and density functional theory with a colleague who has just run a molecule through one package or another but hasn't the slightest clue what the phrases or the output mean.

It therefore seemed to me (and to the reviewers who read my new book proposal) that the time was right for a middle course. I assume that you are a 'Beginner' in the sense of Chambers dictionary, *someone who begins; a person who is in the early stages of learning or doing anything*. . . , and I want to tell you how we go about modern molecular modelling, why we do it, and most important of all, explain much of the basic theory behind the mouse clicks. This involves mathematics and physics, and the book neither pulls punches nor aims at instant enlightenment. Many of the concepts and ideas are difficult ones, and you will have to think long and hard about them; if it's any consolation, so did the pioneers in our subject. I have given many of the derivations in full, and tried to avoid the dreaded phrase 'it can be shown that'.

There are various strands to our studies, all of which eventually intertwine. We start off with molecular mechanics, a classical treatment widely used to predict molecular geometries. In Chapter 8, I give a quick guide to statistical thermodynamics (if such a thing is possible), because we need to make use of the concepts when trying to model arrays of particles at nonzero temperatures. Armed with this knowledge, we are ready for an assault on Monte Carlo and molecular dynamics.

Just as we have to bite the bullet of statistical mechanics, so we have to bite the equally difficult one of quantum mechanics, which occupies Chapters 11 and 12. We then turn to the quantum treatment of atoms, where many of the sums can be done on a postcard if armed with knowledge of angular momentum.

The Hartree–Fock and HF-LCAO models dominate much of the next few chapters, as they should. The Hartree–Fock model is great for predicting many molecular properties, but it can't usually cope with bond breaking and bond making. Chapter 19 treats electron correlation and Chapter 20 deals with the very topical density functional theory (DFT). You won't be taken seriously if you have not done a DFT calculation on your molecule. Quantum mechanics, statistical mechanics and electromagnetism all have a certain well-deserved reputation amongst science students; they are hard subjects. Unfortunately all three all feature in this new text. In electromagnetism it is mostly a matter of getting to grips with the mathematical notation (although I have spared you Maxwell's beautiful equations), whilst in the other two subjects it is more a question of mastering hard concepts. In the case of quantum mechanics, the concepts are often in direct contradiction to everyday experience and common sense. I expect from you a certain level of mathematical competence; I have made extensive use of vectors and matrices not because I am perverse, but because such mathematical notation brings out the inherent simplicity and beauty of many of the equations. I have tried to help by giving a mathematical appendix, which should also make the text self-contained.

I have tried to put the text into historical perspective, and in particular I have quoted directly from a number of what I call *keynote papers*. It is interesting to read at first hand how the pioneers put their ideas across, and in any case they do it far better than me. For example, I am not the only author to quote Paul Dirac's famous statement

The underlying Physical Laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that exact application of these laws leads to equations much too complicated to be soluble.

I hope you have a profitable time in your studies, and at the very least begin to appreciate what all those options mean next time you run a modelling package!

Alan Hinchliffe
Manchester, UK

1

Electric Charges and their Properties

As far as we can tell, there are four fundamental types of interactions between physical objects. There is the *weak nuclear interaction* that governs the decay of beta particles, and the *strong nuclear interaction* that is responsible for binding together the particles in a nucleus. The familiar *gravitational* interaction holds the Earth very firmly in its orbit round the Sun, and finally we know that there is an *electromagnetic* interaction that is responsible for binding atomic electrons to nuclei and for holding atoms together when they combine to form molecules.

Of the four, the gravitational interaction is the only one we would normally come across in our everyday world. This is because gravitational interactions between bodies always add. The gravitational interaction between two atoms is negligible but when large numbers of fundamental particles such as atoms are aggregated together, the gravitational interaction becomes significant.

You may think it bizarre that there are four types of interaction, yet, conversely, you might wonder why there should be just four. Why not one, three or five? Should there not be a unifying theory to explain why there are four, and whether they are related? As I write, there is no such unifying theory despite tremendous research activity.

1.1 Point Charges

In this chapter I am going to concentrate on electric charges and their properties, since electrons and protons are fundamental building blocks for atoms and molecules.

It turns out that there are two types of electric charge in nature, which we might choose to call type X and type Y (or Red and Blue for that matter, but X and Y will do for

now). Experimental evidence shows the existence of an electrostatic force between electric charges; the force between two X-type charges is always repulsive, as is the force between two Y-type charges. The force between an X-type and a Y-type is always attractive. For this reason, the early experimenters decided to classify charges as positive or negative, because a positive quantity times a positive quantity gives a positive quantity, a negative quantity times a negative quantity gives a positive quantity whilst a negative quantity times a positive quantity gives a negative quantity. I am sure you know that the best known fundamental particles responsible for these charges are electrons and protons, and you are probably expecting me to tell you that the electrons are the negatively charged particles whilst protons are positively charged. It is actually just a convention that we take: we could just as well have called electrons positive.

Whilst on the subject, it is fascinating to note that the charge on the electron is exactly equal and opposite of that on a proton. Atoms and molecules generally contain exactly the same number of electrons and protons, and so the net charge on a molecule is almost always zero. Ions certainly exist in solutions of electrolytes, but the number of Na^+ ions in a solution of sodium chloride is exactly equal to the number of Cl^- ions and once again we are rarely aware of any imbalance of charge.

A thunderstorm results when nature separates out positive and negative charges on a macroscopic scale. It is thought that friction between moving masses of air and water vapour detaches electrons from some molecules and attaches them to others. This results in parts of clouds being left with an excess of charge, often with spectacular results. It was investigations into such atmospheric phenomena that gave the first clues about the nature of the electrostatic force.

We normally start any study of charges at rest (*electrostatics*) by considering the force between two point charges, as shown in Figure 1.1. The term ‘point charge’ is a mathematical abstraction; obviously electrons and protons have a finite size. Just bear with me for a few pages, and accept that a point charge is one whose dimensions are small compared to the distance between them. An electron is large if you happen to be a nearby electron, but can normally be treated as a point charge if you happen to be a human being a metre away.

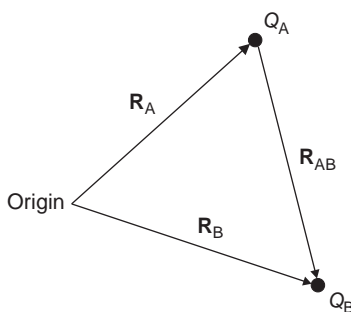


Figure 1.1 Point charges

The concept of a point charge may strike you as an odd one, but once we have established the magnitude of the force between two such charges, we can deduce the force between

any arbitrary charge distributions on the grounds that they are composed of a large number of point charges.

In Figure 1.1 we have point charge Q_A at position vector \mathbf{R}_A and Q_B at \mathbf{R}_B . From the laws of vector analysis, the vector $\mathbf{R}_{AB} = \mathbf{R}_B - \mathbf{R}_A$ joins Q_A to Q_B , and points from Q_A to Q_B as shown. I have indicated the direction of the vectors with arrows.

1.2 Coulomb's Law

In 1785, Charles Augustin de Coulomb became the first person to give a mathematical form to the force between point charges. He measured the force directly between two very small charged bodies, and was able to show that the force exerted by Q_A on Q_B was

- proportional to the inverse square of the distance between Q_A and Q_B when both charges were fixed;
- proportional to Q_A when Q_B and \mathbf{R}_{AB} were fixed;
- proportional to Q_B when Q_A and \mathbf{R}_{AB} were fixed.

He also noticed that the force acted along the line joining the centres of the two charges, and that the force was either attractive or repulsive depending on whether the charges were different or of the same type. The sign of the product of the charges therefore determines the direction of the force.

A mathematical result of these observations can be written in scalar form as

$$F_{A \text{ on } B} \propto \frac{Q_A Q_B}{R_{AB}^2} \quad (1.1)$$

Forces are vector quantities, and Equation (1.1) is better written in vector form as

$$\mathbf{F}_{A \text{ on } B} \propto \frac{Q_A Q_B}{R_{AB}^3} \mathbf{R}_{AB}$$

When Coulomb first established his law, he had no means of quantifying charge and so could not identify the proportionality constant. He took it to be unity, and thereby defined charge in terms of the force between charges. Modern practice is to regard charge and force as independent quantities, and because of this a dimensioned proportionality constant is necessary. For a reason that need not concern us, this is taken as $1/4\pi\epsilon_0$, where the permittivity of free space ϵ_0 is an experimentally determined quantity with the approximate value $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$. Coulomb's law is therefore

$$\mathbf{F}_{A \text{ on } B} = \frac{1}{4\pi\epsilon_0} \frac{Q_A Q_B}{R_{AB}^3} \mathbf{R}_{AB} \quad (1.2)$$

and it applies to measurements done in free space. If we repeat Coulomb's experiments with the charges immersed in different media, then we find that the law still holds but with a different proportionality constant. We modify the proportionality constant using a quantity ϵ_r called the *relative permittivity*. In older texts, ϵ_r is called the *dielectric constant*. Our final statement of Coulomb's law is therefore

$$\mathbf{F}_{A \text{ on } B} = \frac{1}{4\pi\epsilon_r\epsilon_0} \frac{Q_A Q_B}{R_{AB}^3} \mathbf{R}_{AB} \quad (1.3)$$

According to Newton's third law, we know that if Q_A exerts a force $\mathbf{F}_{A \text{ on } B}$ on Q_B , then Q_B should exert an equal and opposite force on Q_A . Coulomb's law satisfies this requirement, since

$$\mathbf{F}_{B \text{ on } A} = \frac{1}{4\pi\epsilon_r\epsilon_0} \frac{Q_A Q_B}{R_{BA}^3} \mathbf{R}_{BA}$$

(the vector \mathbf{R}_{BA} points in the opposite direction to \mathbf{R}_{AB} and so one force is exactly the negative of the other, as it should be).

1.3 Pair Wise Additivity

Suppose we now add a third point charge Q_C with position vector \mathbf{R}_C as shown in Figure 1.2. Since Q_A and Q_B are point charges, the addition of Q_C cannot alter the force between Q_A and Q_B .

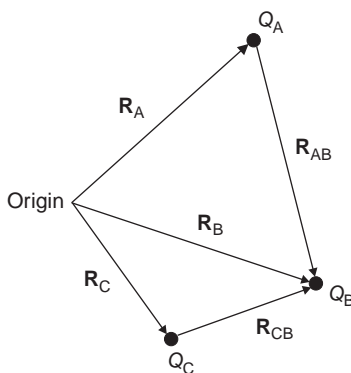


Figure 1.2 Third charge added

The total force on Q_B now comprises two terms, the force due to point charge Q_A and the force due to point charge Q_C . This total force is given by

$$\mathbf{F}_B = \frac{Q_B}{4\pi\epsilon_0} \left(Q_A \frac{\mathbf{R}_{AB}}{R_{AB}^3} + Q_C \frac{\mathbf{R}_{CB}}{R_{CB}^3} \right) \quad (1.4)$$

This may seem at first sight to be a trivial statement: surely all forces act this way. Not necessarily, for I have assumed that the addition of Q_C did not have any effect on Q_A and Q_B (and so did not influence the force between them).

The generic term *pair wise additive* describes things like forces that add as above. Forces between point electric charges are certainly pair wise additive, and so you might imagine that forces between atoms and molecules must therefore be pair wise additive, because atoms and molecules consist of (essentially) point charges. I am afraid that nature is not so kind, and we will shortly meet situations where forces between the composites of electrons and protons that go to make up atoms and molecules are far from being pair wise additive.

1.4 Electric Field

Suppose now we have a point charge Q at the coordinate origin, and we place another point charge q at point P that has position vector \mathbf{r} (Figure 1.3).

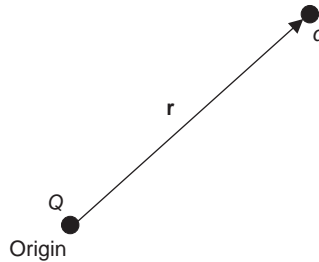


Figure 1.3 Field concept

The force exerted by Q on q is

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^3} \mathbf{r}$$

which I can rewrite trivially as

$$\mathbf{F} = \left(\frac{1}{4\pi\epsilon_0} \frac{Q}{r^3} \mathbf{r} \right) q$$

The point is that the term in brackets is to do with Q and the vector \mathbf{r} , and contains no mention of q . If we want to find the force on any arbitrary q at \mathbf{r} , we calculate the quantity in brackets once and then multiply by q . One way of thinking about this is to imagine that the charge Q creates a certain field at point \mathbf{r} , which determines the force on any other q when placed at position \mathbf{r} .

This property is called the *electric field* \mathbf{E} at that point. It is a vector quantity, like force, and the relationship is

$$\mathbf{F} \text{ (on } q \text{ at } \mathbf{r}) = q\mathbf{E} \text{ (at } \mathbf{r})$$

Comparison with Coulomb's law, Equation (1.3), shows that the electric field at point \mathbf{r} due to a point charge Q at the coordinate origin is

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q\mathbf{r}}{r^3} \quad (1.5)$$

\mathbf{E} is sometimes written $\mathbf{E}(\mathbf{r})$ to emphasize that the electric field depends on the position vector \mathbf{r} .

Electric fields are vector fields and they are often visualized as *field lines*. These are drawn such that their spacing is inversely proportional to the strength of the field, and their tangent is in the direction of the field. They start at positive charges and end at negative charges, and two simple examples are shown in Figure 1.4. Here the choice of eight lines is quite arbitrary.

Electric fields that do not vary with time are called *electrostatic* fields.

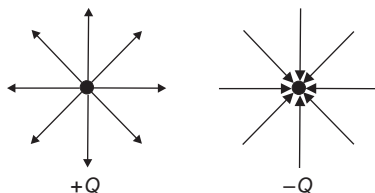


Figure 1.4 Field lines for point charges

1.5 Work

Look again at Figure 1.3, and suppose we move point charge q whilst keeping Q fixed in position. When a force acts to make something move, energy is transferred. There is a useful phrase in physical science that is to do with the energy transferred, and it is *work*. Work measures the energy transferred in any change, and can be calculated from the change in energy of a body when it moves through a distance under the influence of a force.

We have to be careful to take account of the energy balance. If a body gains energy, this energy has to come from somewhere, and that somewhere must lose energy. What we do is to divide the universe into two parts: the bits we are interested in called the *system* and the rest of the universe that we call the *surroundings*.

Some texts focus on the work done *by* the system, some concern themselves with the work done *on* the system. According to the law of conservation of energy, one is exactly the equal and opposite of the other, but we have to be clear which is being discussed. I am going to write w_{on} for the work done on our system.

If the system gains energy, then w_{on} will be positive. If the system loses energy then w_{on} will be negative.

We also have to be careful about the phrase ‘through a distance’. The phrase means ‘through a distance that is the projection of the force vector on the displacement vector’, and you should instantly recognize a vector scalar product (see Appendix A).

A useful formula that relates to the energy gained by a system (i.e. w_{on}) when a constant force \mathbf{F} moves its point of application through \mathbf{l} is

$$w_{\text{on}} = -\mathbf{F} \cdot \mathbf{l} \quad (1.6)$$

In the case where the force is not constant, we have to divide up the motion into differential elements $d\mathbf{l}$. The energy transferred is then given by the sum of all the corresponding differential elements dw_{on} . The corresponding formulae are

$$\begin{aligned} dw_{\text{on}} &= -\mathbf{F} \cdot d\mathbf{l} \\ w_{\text{on}} &= -\int \mathbf{F} \cdot d\mathbf{l} \end{aligned} \quad (1.7)$$

We now move q by an infinitesimal vector displacement $d\mathbf{l}$, as shown in Figure 1.5, so that it ends up at point $\mathbf{r} + d\mathbf{l}$. The work done on the system in that differential change is

$$dw_{\text{on}} = -\mathbf{F} \cdot d\mathbf{l}$$

If the angle between the vectors \mathbf{r}_1 and $d\mathbf{l}$ is θ , then we have

$$dw_{\text{on}} = -F dl \cos \theta$$

and examination of Figure 1.6 shows that $dl \cos \theta$ is the radial distance moved by charge q , which we will write dr .

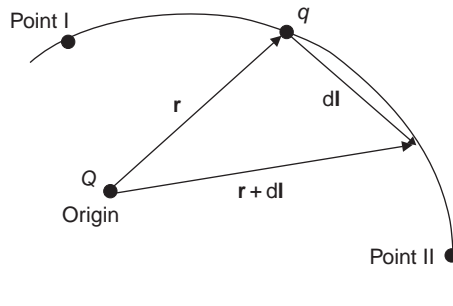


Figure 1.5 Electrostatic work

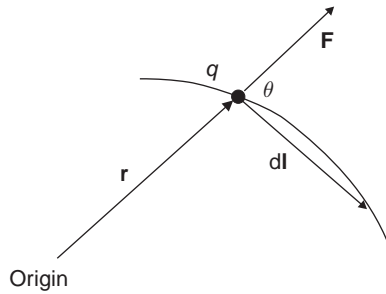


Figure 1.6 Relationship between vectors

Hence

$$dw_{\text{on}} = -\frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2} dr$$

The total work done moving from position I to position II is therefore found by integrating

$$\begin{aligned} w_{\text{on}} &= -\frac{1}{4\pi\epsilon_0} \int_I^{II} \frac{Qq}{r^2} dr \\ &= \frac{1}{4\pi\epsilon_0} Qq \left(\frac{1}{r_{II}} - \frac{1}{r_I} \right) \end{aligned} \tag{1.8}$$

The work done depends only on the initial and final positions of charge q ; it is independent of the way we make the change.

Another way to think about the problem is as follows. The force is radial, and we can divide the movement from position I to position II into infinitesimal steps, some of which are parallel to \mathbf{F} and some of which are perpendicular to \mathbf{F} . The perpendicular steps count 0 towards w_{on} , the parallel steps only depend on the change in the (scalar) radial distance.

1.6 Charge Distributions

So far I have concentrated on point charges, and carefully skirted round the question as to how we deal with continuous distributions of charge. Figure 1.7 shows a charge distribution Q_A . The density of charge need not be constant through space, and we normally write $\rho(\mathbf{r})$ for the density at the point whose position vector is \mathbf{r} . The charge contained within the volume element $d\tau$ at \mathbf{r} is therefore $\rho(\mathbf{r})d\tau$ and the relationship between $\rho(\mathbf{r})$ and Q_A is discussed in Appendix A. It is

$$Q_A = \int \rho(\mathbf{r}) d\tau \quad (1.9)$$

In order to find the force between the charge distribution and the point charge Q_B we simply extend our ideas about the force between two point charges; one of the point charges being $\rho(\mathbf{r})d\tau$ and the other Q_B .

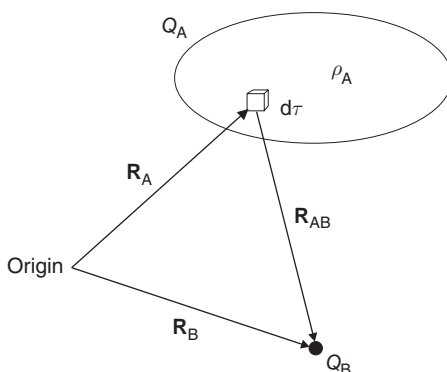


Figure 1.7 Charge distribution

The total force is given by the sum of all possible contributions from the elements of the continuous charge distribution Q_A with point charge Q_B . The practical calculation of such a force can be a nightmare, even for simple charge distributions. One of the reasons for the nightmare is that forces are vector quantities; we need to know about both their magnitude and their direction.

In the next section, I am going to tell you about a very useful scalar field called the mutual potential energy U . This field has the great advantage that it is a scalar field, and so we do not need to worry about direction in our calculations.

1.7 Mutual Potential Energy, U

Suppose now we start with charge q at infinity, and move it up to a point with vector position \mathbf{r} , as shown in Figure 1.3. The work done is

$$w_{\text{on}} = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r} \quad (1.10)$$

and this represents the energy change on building up the charge distribution, with the charges initially at infinite separation. It turns out that this energy change is an important property, and we give it a special name (the *mutual potential energy*) and a special symbol U (occasionally Φ).

Comparison of the equations for force, work and mutual potential energy given above suggests that there might be a link between the force and the mutual potential energy; at first sight, one expression looks like the derivative of the other.

I am going to derive a relationship between force and mutual potential energy. The relationship is perfectly general; it applies to all forces provided that they are constant in time.

1.8 Relationship between Force and Mutual Potential Energy

Consider a body of mass m that moves in (say) the x -direction under the influence of a constant force. Suppose that at some instant its speed is v . The kinetic energy is $\frac{1}{2}mv^2$. You are probably aware of the law of conservation of energy, and know that when I add the potential energy U to the kinetic energy, I will get a constant energy that I will denote ε (sorry about the clash of symbols with electric permittivity, but that is life):

$$\varepsilon = \frac{1}{2}mv^2 + U \quad (1.11)$$

I want to show you how to relate U to the force F . If the energy ε is constant in time, then $d\varepsilon/dt = 0$. Differentiation of Equation (1.11) with respect to time gives

$$\frac{d\varepsilon}{dt} = mv \frac{dv}{dt} + \frac{dU}{dt}$$

and so, by the chain rule

$$\frac{d\varepsilon}{dt} = mv \frac{dv}{dt} + \frac{dU}{dx} \frac{dx}{dt}$$

If the energy ε is constant then its first differential with respect to time is zero, and v is just dx/dt . Likewise dv/dt is the acceleration and so

$$0 = \left(m \frac{d^2x}{dt^2} + \frac{dU}{dx} \right) \frac{dx}{dt} \quad (1.12)$$

Equation (1.12) is true if the speed is zero, or if the term in brackets is zero. According to Newton's second law of motion, mass times acceleration is force, and so

$$F = - \frac{dU}{dx} \quad (1.13)$$

which gives us the link between force and mutual potential energy.

When working in three dimensions, we have to be careful to distinguish between vectors and scalars. We treat a body of mass m whose position vector is \mathbf{r} . The velocity is $\mathbf{v} = d\mathbf{r}/dt$ and the kinetic energy is $\frac{1}{2}m(d\mathbf{r}/dt)(d\mathbf{r}/dt)$.

Analysis along the lines given above shows that the force \mathbf{F} and U are related by

$$\mathbf{F} = -\text{grad } U \quad (1.14)$$

where the gradient of U is discussed in Appendix A and is given in Cartesian coordinates by

$$\text{grad } U = \frac{\partial U}{\partial x} \mathbf{e}_x + \frac{\partial U}{\partial y} \mathbf{e}_y + \frac{\partial U}{\partial z} \mathbf{e}_z \quad (1.15)$$

Here \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z are unit vectors pointing along the Cartesian axes.

1.9 Electric Multipoles

We can define exactly an array of point charges by listing the magnitudes of the charges, together with their position vectors. If we then wish to calculate (say) the force between one array of charges and another, we simply apply Coulomb's law (Equation (1.3)) repeatedly to each pair of charges. Equation (1.3) is exact, and can be easily extended to cover the case of continuous charge distributions.

For many purposes, it proves more profitable to describe a charge distribution in terms of certain quantities called the *electric moments*. We can then discuss the interaction of one charge distribution with another in terms of the interactions between the electric moments.

Consider first a pair of equal and opposite point charges, $+Q$ and $-Q$ separated by distance R (Figure 1.8). This pair of charges is usually said to form an *electric dipole* of magnitude QR . In fact, electric dipoles are vector quantities and a more rigorous definition is

$$\mathbf{p}_e = QR \quad (1.16)$$

where the vector \mathbf{R} points from the negative charge to the positive charge.

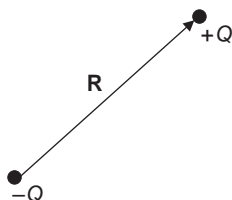


Figure 1.8 Simple electric dipole

We sometimes have to concern ourselves with a more general definition, one relating to an arbitrary array of charges such as that shown in Figure 1.9. We have four point charges: Q_1 whose position vector is \mathbf{R}_1 , Q_2 whose position vector is \mathbf{R}_2 , Q_3 whose position vector is \mathbf{R}_3 and Q_4 whose position vector is \mathbf{R}_4 . We define the *electric dipole moment* \mathbf{p}_e of these four charges as

$$\mathbf{p}_e = \sum_{i=1}^4 Q_i \mathbf{R}_i$$

It is a vector quantity with x , y and z Cartesian components

$$\sum_{i=1}^4 Q_i X_i, \quad \sum_{i=1}^4 Q_i Y_i \quad \text{and} \quad \sum_{i=1}^4 Q_i Z_i$$

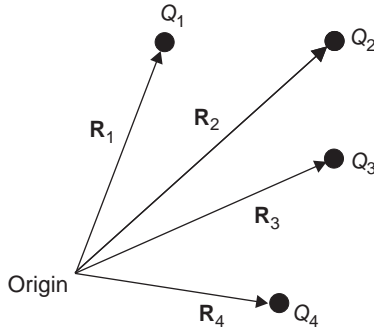


Figure 1.9 Generalized electric dipole

This is consistent with the elementary definition given above; in the case of two equal and opposite charges $+Q$ and $-Q$ a distance d apart, the electric dipole moment has magnitude Qd and points from the negative charge to the positive. The generalization to n charges is obvious: we substitute n for 4 in the above definition.

There are several general rules about electric moments of charge distributions, and we can learn a couple of the ones that apply to dipole moments by considering the simple arrays shown in Figure 1.10 and keeping the definitions in mind.

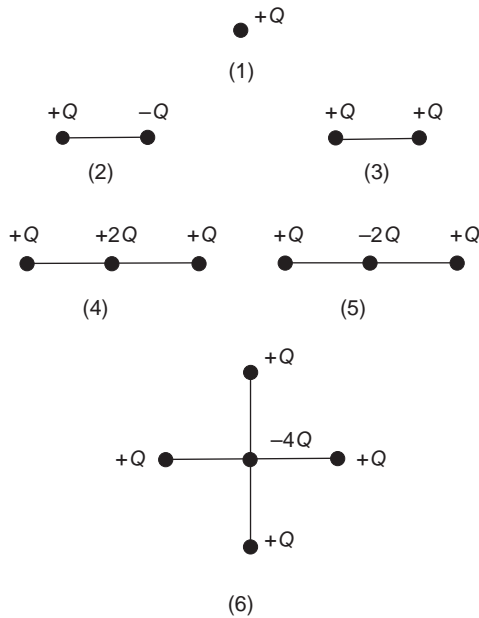


Figure 1.10 Simple arrays of point charges

I have joined up the charges with lines in order to focus attention on the charge systems involved; there is no implication of a ‘bond’. We do not normally discuss the electric dipole due to a point charge (1). Examination of the charge distributions (2) through (6) and calculation of their electric dipole moment for different coordinate origins suggest the general result; neutral arrays of point charges have a unique electric dipole moment that does not depend on where we take the coordinate origin. Otherwise, we have to state the coordinate origin when we discuss the electric dipole moment.

I can prove this from Equation (1.16), generalized to n point charges:

$$\mathbf{p}_e = \sum_{i=1}^n Q_i \mathbf{R}_i \quad (1.17)$$

Suppose that we move the coordinate origin so that each point charge Q_i has a position vector \mathbf{R}'_i where

$$\mathbf{R}_i = \mathbf{R}'_i + \mathbf{\Delta}$$

with $\mathbf{\Delta}$ a constant vector. From the definition of electric dipole moment we have

$$\mathbf{p}_e = \sum_{i=1}^n Q_i \mathbf{R}_i$$

and so, with respect to the new coordinate origin

$$\begin{aligned} \mathbf{p}'_e &= \sum_{i=1}^n Q_i \mathbf{R}'_i \\ &= \sum_{i=1}^n Q_i (\mathbf{R}_i - \mathbf{\Delta}) \\ &= \mathbf{p}_e - \mathbf{\Delta} \sum_{i=1}^n Q_i \end{aligned}$$

The two definitions only give the same vector if the sum of charges is zero. We often use the phrase *gauge invariant* to describe quantities that do not depend on the choice of coordinate origin.

Arrays (5) and (6) each have a centre of symmetry. There is a general result that any charge distribution having no overall charge but a centre of symmetry must have a zero dipole moment, and similar results follow for other highly symmetrical arrays of charges.

1.9.1 Continuous Charge Distributions

In order to extend the definition of an electric dipole to a continuous charge distribution such as that shown in Figure 1.7, we first of all divide the region of space into differential elements $d\tau$. If $\rho(\mathbf{r})$ is the charge density then the change in volume element $d\tau$ is $\rho(\mathbf{r})d\tau$. We then treat each of these volume elements as point charges and add (i.e. integrate). The electric dipole moment becomes

$$\mathbf{p}_e = \int \mathbf{r} \rho(\mathbf{r}) d\tau \quad (1.18)$$

1.9.2 Electric Second Moment

The electric dipole moment of an array of point charges is defined by the three sums

$$\sum_{i=1}^n Q_i X_i, \quad \sum_{i=1}^n Q_i Y_i \quad \text{and} \quad \sum_{i=1}^n Q_i Z_i$$

and we can collect them into a column vector in an obvious way as

$$\mathbf{p}_e = \begin{pmatrix} \sum_{i=1}^n Q_i X_i \\ \sum_{i=1}^n Q_i Y_i \\ \sum_{i=1}^n Q_i Z_i \end{pmatrix} \quad (1.19)$$

The six independent quantities

$$\sum_{i=1}^n Q_i X_i^2, \quad \sum_{i=1}^n Q_i X_i Y_i, \quad \sum_{i=1}^n Q_i X_i Z_i, \dots, \quad \sum_{i=1}^n Q_i Z_i^2$$

are said to define the *electric second moment* of the charge distribution. We usually collect them into a real symmetric 3×3 matrix \mathbf{q}_e :

$$\mathbf{q}_e = \begin{pmatrix} \sum_{i=1}^n Q_i X_i^2 & \sum_{i=1}^n Q_i X_i Y_i & \sum_{i=1}^n Q_i X_i Z_i \\ \sum_{i=1}^n Q_i Y_i X_i & \sum_{i=1}^n Q_i Y_i^2 & \sum_{i=1}^n Q_i Y_i Z_i \\ \sum_{i=1}^n Q_i Z_i X_i & \sum_{i=1}^n Q_i Z_i Y_i & \sum_{i=1}^n Q_i Z_i^2 \end{pmatrix} \quad (1.20)$$

The matrix is symmetric because of the obvious equalities of the off-diagonal sums such as

$$\sum_{i=1}^n Q_i X_i Y_i \quad \text{and} \quad \sum_{i=1}^n Q_i Y_i X_i$$

There are unfortunately many different definitions related to the second (and higher) moments in the literature. There is little uniformity of usage, and it is necessary to be crystal clear about the definition and choice of origin when dealing with these quantities.

Most authors prefer to work with a quantity called the *electric quadrupole moment* rather than the second moment, but even then there are several different conventions. A common choice is to use the symbol Θ_e and the definition

$$\Theta_e = \frac{1}{2} \begin{pmatrix} \sum_{i=1}^n Q_i (3X_i^2 - R_i^2) & 3 \sum_{i=1}^n Q_i X_i Y_i & 3 \sum_{i=1}^n Q_i X_i Z_i \\ 3 \sum_{i=1}^n Q_i Y_i X_i & \sum_{i=1}^n Q_i (3Y_i^2 - R_i^2) & 3 \sum_{i=1}^n Q_i Y_i Z_i \\ 3 \sum_{i=1}^n Q_i Z_i X_i & 3 \sum_{i=1}^n Q_i Z_i Y_i & \sum_{i=1}^n Q_i (3Z_i^2 - R_i^2) \end{pmatrix} \quad (1.21)$$

Notice that the diagonal elements of this matrix sum to zero and so the matrix has zero trace (the *trace* being the sum of the diagonal elements; see Appendix A). Some authors do not

use the factor of $1/2$ in their definition. Quadrupole moments are gauge invariant provided the electric dipole moment and the charge are both zero.

Figure 1.11 shows an octahedrally symmetrical array of point charges. Each point charge has magnitude Q , apart from the central charge that has magnitude $-6Q$ in order to make the system neutral. The distance between each axial point charge and the central one is a .

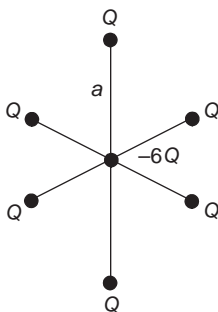


Figure 1.11 *Octahedral charge distribution*

If I choose to direct the Cartesian axes along the symmetry axes, then the second moment matrix is

$$\mathbf{q}_e = Qa^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

whilst the quadrupole moment matrix is zero.

If I now reduce the symmetry of the charge distribution by placing charges $2Q$ along the vertical axis (taken for the sake of argument as the x -axis) and $-8Q$ at the centre (to keep the electrical balance), the second moment matrix becomes

$$\mathbf{q}_e = Qa^2 \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

whilst the quadrupole moment matrix is now

$$\mathbf{\Theta}_e = Qa^2 \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

The electric quadrupole moment measures deviations from spherical symmetry. It is zero when the charge distribution has spherical symmetry.

It always has zero trace (because of the definition), but it is not always diagonal. Nevertheless, it can always be made diagonal by a rotation of the coordinate axes.

Finally, consider a linear array formed by the top ($+Q$), central ($+2Q$) and lower charges ($-3Q$). We find

$$\mathbf{q}_e = Qa^2 \begin{pmatrix} -2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \Theta_e = Qa^2 \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In cases where the symmetry of the problem determines that the second moment tensor only has one nonzero component, we speak colloquially of *the* second moment (which in this case is $-2Qa^2$).

1.9.3 Higher Electric Moments

The set of ten independent quantities

$$\sum_{i=1}^n Q_i X_i^3, \quad \sum_{i=1}^n Q_i X_i^2 Y_i, \dots, \quad \sum_{i=1}^n Q_i Z_i^3$$

defines the electric third moment of the charge distribution, and so on. We rarely encounter such higher moments of electric charge in chemistry.

1.10 Electrostatic Potential

Electrostatic forces are vector quantities, and we have to worry about their magnitude and direction. I explained earlier that it is more usual to work with the mutual potential energy U rather than the force \mathbf{F} , if only because U is a scalar quantity. In any case we can recover one from the other by the formula

$$\mathbf{F} = -\text{grad } U$$

Similar considerations apply when dealing with electrostatic fields. They are vector fields with all the inherent problems of having to deal with both a magnitude and a direction. It is usual to work with a scalar field called the *electrostatic potential* ϕ . This is related to the electrostatic field \mathbf{E} in the same way that U is related to \mathbf{F} :

$$\mathbf{E} = -\text{grad } \phi$$

We will hear more about the electrostatic potential in later sections. In the meantime, I will tell you that the electrostatic potential at field point \mathbf{r} due to a point charge Q at the coordinate origin is

$$\phi(\mathbf{r}) = \frac{Q}{4\pi\epsilon_0} \frac{1}{r} \tag{1.22}$$

The electric field and the electrostatic potential due to an electric dipole, quadrupole and higher electric moments are discussed in all the elementary electromagnetism texts. The expressions can be written exactly in terms of the various distances and charges involved. For many applications, including our own, it is worthwhile examining the mathematical form of these fields for points in space that are far away from the charge distribution. We then refer to (for example) the ‘small’ electric dipole and so on.

The electrostatic potential at field point \mathbf{r} due to a small electric dipole \mathbf{p}_e at the coordinate origin turns out to be

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{p}_e \cdot \mathbf{r}}{r^3} \quad (1.23)$$

which falls off as $1/r^2$. It falls off faster with r than the potential due to a point charge because of the cancellation due to plus and minus charges. This is in fact a general rule: the electrostatic potential for a small electric multipole of order l falls off as $r^{-(l+1)}$ so dipole moment potentials fall off faster than those due to point charges, and so on.

1.11 Polarization and Polarizability

In electrical circuits, charges are stored in capacitors, which at their simplest consist of a pair of conductors carrying equal and opposite charges. Michael Faraday (1837) made a great discovery when he observed that filling the space between the plates of a parallel plate capacitor with substances such as mica increased their ability to store charge. The multiplicative factor is called the relative permittivity and is given a symbol ϵ_r , as discussed above. I also told you that the older name is the dielectric constant.

Materials such as glass and mica differ from substances such as copper wire, in that they have few conduction electrons and so make poor conductors of electric current. We call materials such as glass and mica *dielectrics*, to distinguish them from metallic conductors.

Figure 1.12 shows a two-dimensional picture of a dielectric material, illustrated as positively charged nuclei each surrounded by a localized electron cloud.

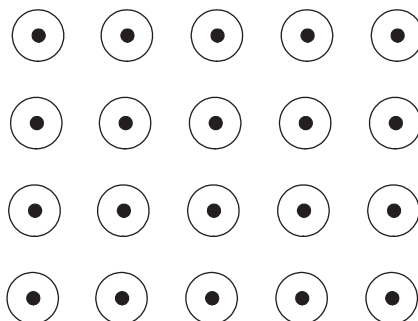


Figure 1.12 Dielectric slab

We now apply an electrostatic field, directed from the left to the right. There is a force on each charge, and the positive charges are displaced to the right whilst the negative charges move a corresponding distance to the left, as shown in Figure 1.13.

The macroscopic theory of this phenomenon is referred to as *dielectric polarization*, and we focus on the induced dipole moment $d\mathbf{p}_e$ per differential volume $d\tau$. Because it is a macroscopic theory, no attention is paid to atomic details; we assume that there are a large number of atoms or molecules within the volume element $d\tau$ (or that the effects caused by the discrete particles have somehow been averaged out).

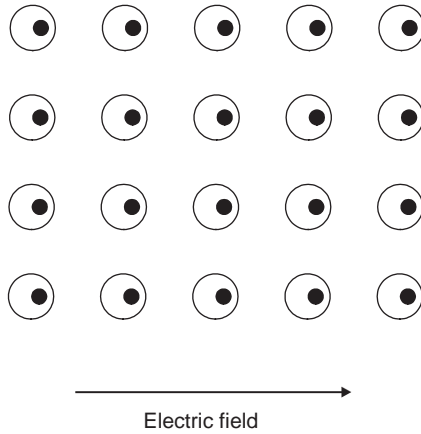


Figure 1.13 Dielectric with applied field

We relate the induced electric dipole to the volume of a differential element by

$$d\mathbf{p}_e = \mathbf{P}d\tau \quad (1.24)$$

where the dielectric polarization \mathbf{P} is an experimentally determined quantity. \mathbf{P} can depend on the applied field in all manner of complicated ways, but for very simple media and for low field strengths, it turns out that \mathbf{P} is directly proportional to \mathbf{E} . We write

$$\mathbf{P} = (\epsilon_r - 1) \epsilon_0 \mathbf{E} \quad (1.25)$$

where ϵ_r is the relative permittivity of the dielectric. The polarization acts so as to reduce the field inside a dielectric.

1.12 Dipole Polarizability

At the microscopic level, we concern ourselves with the various electric moments that are induced in each atom or molecule.

Consider the simple case shown in Figure 1.14, where we apply a weak electric field \mathbf{E} in a direction parallel to a molecule's electric dipole moment. This causes charge redistribution and we can write

$$p_e (\text{induced}) = p_e (\text{permanent}) + \alpha E \quad (1.26)$$

I have distinguished between the permanent electric dipole, the one a molecule has in free space with no fields present, from the induced dipole. I have also used the symbol α for the *dipole polarizability*.

In the general case, the field need not be weak and the induced dipole need not be in the same direction as either the applied field or the permanent dipole moment. This is shown in Figure 1.15. Parameter α cannot be a scalar, since the directions of the applied field and the induced dipole need not be the same.

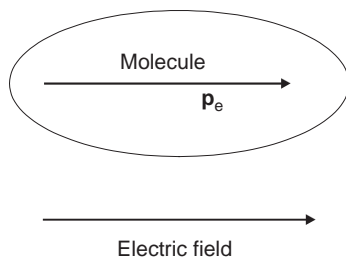


Figure 1.14 Induced molecular dipole

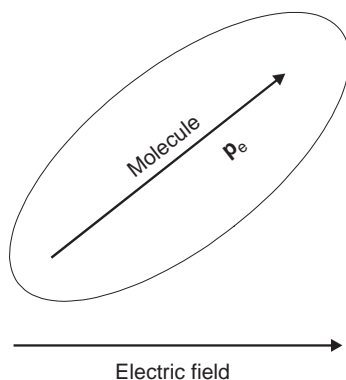


Figure 1.15 More general case

The dipole polarizability α is a special type of physical property called a tensor, just like the electric second moment. We can represent α as a 3×3 real symmetric matrix

$$\alpha = \begin{pmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{pmatrix} \quad (1.27)$$

and I will write the more general expression as

$$\mathbf{p}_e \text{ (induced)} = \mathbf{p}_e \text{ (permanent)} + \alpha \cdot \mathbf{E} + \text{higher order terms} \quad (1.28)$$

The higher order terms are the *hyperpolarizabilities*; they feature in advanced texts of this kind. We are not going to meet them again. We interpret Equation (1.28) as a matrix equation; the \mathbf{p} 's are column vectors.

1.12.1 Properties of Polarizabilities

The matrix α can always be written in diagonal form by a suitable rotation of the Cartesian axes (see Appendix A) to give

$$\alpha = \begin{pmatrix} \alpha_{aa} & 0 & 0 \\ 0 & \alpha_{bb} & 0 \\ 0 & 0 & \alpha_{cc} \end{pmatrix}$$

The quantities α_{aa} , α_{bb} and α_{cc} are called the *principal values* of the polarizability tensor. For molecules with symmetry, the principal axes of polarizability correspond to the molecular symmetry axes. For a linear molecule the components that refer to perpendicular axes are equal and usually different from the parallel component and the matrix is usually written

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{aa} & 0 & 0 \\ 0 & \alpha_{bb} & 0 \\ 0 & 0 & \alpha_{bb} \end{pmatrix}$$

1.13 Many-Body Forces

It is instructive to calculate the work done in building up an array of charges such as that shown in Figure 1.9. We will assume that all the charges are point charges and so cannot be polarized. We start with all the charges (n in total) at infinity, and move them in turn from infinity to their position as shown.

Moving Q_1 from infinity to position vector \mathbf{R}_1 takes no energy, because no other charges are present.

Moving Q_2 from infinity to position vector \mathbf{R}_2 involves an energy change:

$$U_{12} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_2}{R_{12}}$$

where R_{12} is the length of the vector joining the charges. We now move Q_3 from infinity to position vector \mathbf{R}_3 . This charge moves in the field due to Q_1 and Q_2 , involving an energy cost:

$$U_{13} + U_{23} = \frac{1}{4\pi\epsilon_0} \frac{Q_1 Q_3}{R_{13}} + \frac{1}{4\pi\epsilon_0} \frac{Q_2 Q_3}{R_{23}}$$

The total work done U_{tot} is seen to be

$$U_{\text{tot}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n U_{ij} \quad (1.29)$$

This expression holds because of the pair wise additivity of the forces between point charges. The expression would not necessarily hold if the charges were themselves charge distributions because the addition of further charges could polarize the existing ones and so alter the forces already calculated.

Whatever the case, U_{tot} will depend on the coordinates of all the charges present:

$$U_{\text{tot}} = U_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n) \quad (1.30)$$

and it is always possible to write a formal expansion

$$\begin{aligned} U_{\text{tot}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n) &= \sum_{\text{pairs}} U^{(2)}(\mathbf{R}_i, \mathbf{R}_j) + \sum_{\text{triples}} U^{(3)}(\mathbf{R}_i, \mathbf{R}_j, \mathbf{R}_k) + \dots \\ &+ U^{(n)}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n) \end{aligned} \quad (1.31)$$

involving distinct pairs, triples, etc. of particles. The $U^{(2)}$ terms are referred to as the pair contributions, the $U^{(3)}$ terms are the three-body terms, and so on. Terms higher than the second are identically zero for the interaction between point charges.

1.14 Problem Set

A set of problems, together with solutions, is available on the associated website.

2

The Forces between Molecules

When molecules are near enough to influence one another, we need to concern ourselves with the balance between the forces of attraction and repulsion. We know that such attractive forces exist, for otherwise there would be nothing to bring molecules together into the solid and liquid states, and all matter would be gaseous. A study of the forces between atomic or molecular species constitutes the subject of *intermolecular forces*.

People have speculated about the nature of intermolecular forces ever since the ideas of atoms and molecules first existed. Our present ideas, that molecules attract at long range but repel strongly at short range, began to emerge in the nineteenth century due to the experimental work of Rudolf J.E. Clausius and Thomas Andrews. It was not until the early twentieth century, when the principles of quantum mechanics became established, that we could truly say that we understood the detailed mechanism of intermolecular forces.

Although we talk about intermolecular *forces*, it is more usual and convenient to focus on the *mutual potential energy*, discussed in Chapter 1. If we start with two argon atoms at infinite separation, then their mutual potential energy at separation R tells us the energy change on bringing the atoms together to that distance from infinity.

Even for the simplest pair of molecules, the intermolecular mutual potential energy will depend on their relative orientations in addition to their separation. Perhaps you can now see why the study of intermolecular forces has taken so much effort by so many brilliant scientists, over very many years.

2.1 Pair Potential

So, to start with, we concern ourselves with two atomic or molecular species A and B, and ask how they interact. No chemical reaction is implied, and I should say straightaway that I am not going to be concerned with bond making and bond breaking in this chapter. That is

the subject of *valence theory*. In the (unusual) case that the two species A and B concerned are ions, you may think that the problem is more or less solved. We simply calculate their mutual Coulomb potential energy as discussed in Chapter 1:

$$U_{AB} = \frac{1}{4\pi\epsilon_0} \frac{Q_A Q_B}{R_{AB}} \quad (2.1)$$

You would certainly be on the right lines in this approach except that ions are not point charges and they can be polarized just like any other continuous charge distribution. But as I explained in Chapter 1, we rarely have to concern ourselves with ions, and to get started we will consider the very simple case of a pair of like atoms (such as two argon atoms). We know from experiment that the general form of their mutual potential energy must be that shown in Figure 2.1. This curve is meant to be schematic, and I have been careful not to quantify the axes. It has all the right features but we have yet to discover its precise form. We very often speak about the *pair potential*, and write it $U(R)$, where R is the separation between the two atoms. The zero of $U(R)$, denoted by the horizontal line, is commonly taken to refer to the two atoms at infinity. We often characterize the curve in terms of a small number of parameters; for example, the collision diameter σ being the distance at which $U(R) = 0$, the minimum R_{\min} and minus the value of $U(R)$ at R_{\min} (often written ϵ , and as defined is a positive quantity).

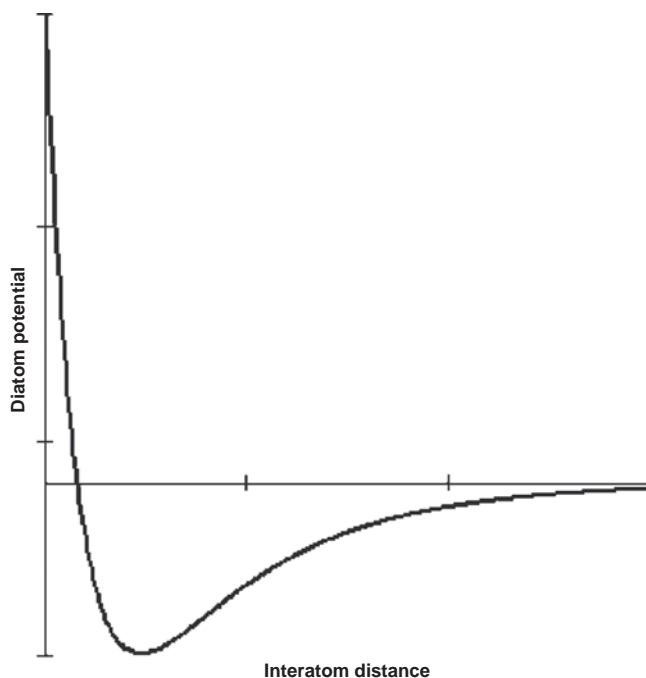


Figure 2.1 *Schematic argon–argon interaction*

We now need to investigate more closely the precise form of this pair potential. The potential comprises a repulsive part (important for small R) and an attractive part (important for large R). It turns out that there are three major contributions to the attractive part, as we will see below.

2.2 Multipole Expansion

Suppose that we have two molecules with centres a distance R apart (Figure 2.2). The distance R is taken to be large compared with a molecular dimension.

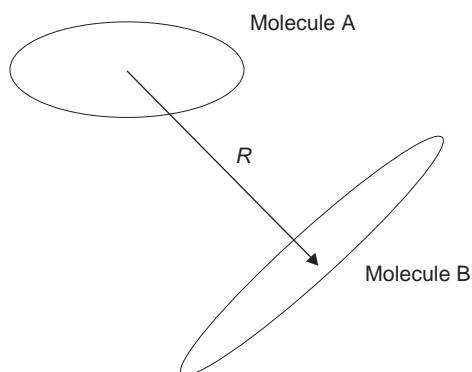


Figure 2.2 Two interacting molecules

Each molecule consists of a number of charged particles, and in principle we can write down an expression for the mutual potential energy of these two molecules in terms of the pair potentials between the various point charges. The basic physical idea of the *multipole expansion* is to make use of the fact that several of these particles go to form molecule A, and the remainder to form molecule B, each of which has a distinct chemical identity. We therefore seek to write the mutual potential energy of A and B in terms of properties of the two molecular charge distributions and their separation.

2.3 Charge–Dipole Interaction

I can illustrate the ideas by considering an elementary textbook problem, the mutual potential energy of a simple small electric dipole and a point charge (Figure 2.3).

Suppose then that we have a simple dipole consisting of a pair of charges Q_A and Q_B aligned along the horizontal axis and equally separated from the coordinate origin by distance d . We introduce a third charge Q as shown, with the scalar distance R from the origin. This point charge makes an angle θ with the electric dipole, as shown.

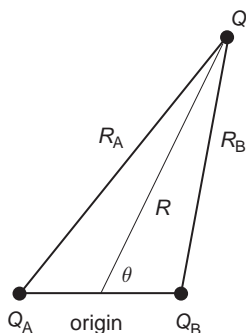


Figure 2.3 Charge-dipole interaction

The two point charges Q_A and Q_B have a mutual potential energy of

$$\frac{1}{4\pi\epsilon_0} \frac{Q_A Q_B}{2d} \quad (2.2)$$

but we are going to investigate what happens to the mutual potential energy of the system as we change the position vector of Q and so we ignore this term since it remains constant.

The mutual potential energy U (charge-dipole) of the point charge and the electric dipole is given exactly by

$$U(\text{charge-dipole}) = \frac{1}{4\pi\epsilon_0} Q \left(\frac{Q_A}{R_A} + \frac{Q_B}{R_B} \right) \quad (2.3)$$

This can also be written in terms of R and θ as

$$U(\text{charge-dipole}) = \frac{1}{4\pi\epsilon_0} Q \left(\frac{Q_A}{\sqrt{(R^2 + d^2 + 2dR \cos \theta)}} + \frac{Q_B}{\sqrt{(R^2 + d^2 - 2dR \cos \theta)}} \right) \quad (2.4)$$

and, once again, this is an exact expression.

In the case where the point charge Q gets progressively far away from the coordinate origin, we can usefully expand the two denominators using the binomial theorem to give

$$U(\text{charge-dipole}) = \frac{1}{4\pi\epsilon_0} Q \left(\frac{(Q_A + Q_B)}{R} + \frac{(Q_B - Q_A)d}{R^2} \cos \theta + \frac{(Q_A + Q_B)d^2}{2R^3} (3 \cos^2 \theta - 1) + \dots \right) \quad (2.5)$$

The first term on the right-hand side contains the sum of the two charges making up the dipole. Very often, we deal with simple dipoles that carry no overall charge, and this term is zero because $Q_A = -Q_B$.

The second term on the right-hand side obviously involves the electric dipole moment, whose magnitude is $(Q_B - Q_A)d$. The third term involves the electric second moment whose magnitude is $(Q_B + Q_A)d^2$, and so on. The mutual potential energy is therefore seen to be a sum of terms; each is a product of a moment of the electric charge, and a function of

the inverse distance. Hopefully, as R increases, the magnitude of the successive terms will become less and eventually the mutual potential energy will be dominated by the first few terms in the expansion.

In the more general case where we replace Q_A and Q_B with an arbitrary array of point charges Q_1, Q_2, \dots, Q_n whose position vectors are $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_n$ (or for that matter a continuous charge distribution), it turns out that we can always write the mutual interaction potential with Q as

$$U = \frac{Q}{4\pi\epsilon_0} \left\{ \left(\sum_{i=1}^n Q_i \right) \frac{1}{R} - \left(\sum_{i=1}^n Q_i \mathbf{R}_i \right) \cdot \text{grad} \left(\frac{1}{R} \right) + \text{higher order terms} \right\} \quad (2.6)$$

The first summation on the right-hand side gives the overall charge of the charge distribution, the second term involves the electric dipole moment, the third term involves the electric quadrupole moment, and so on.

2.4 Dipole–Dipole Interaction

Consider now a slightly more realistic model for the interaction of two simple (diatomic) molecules (Figure 2.4). Molecule A consists of two point charges, Q_{1A} and Q_{2A} . Molecule B consists of two point charges Q_{1B} and Q_{2B} . The overall charge on molecule A is therefore $Q_A = Q_{1A} + Q_{2A}$ with a similar expression for molecule B. The electric dipole moments of A and B are written $\mathbf{p}_A, \mathbf{p}_B$ in an obvious notation, and their scalar magnitudes are written p_A, p_B . The second moments of the two molecules are each determined by a scalar value q_A and q_B , simply because they are linear.

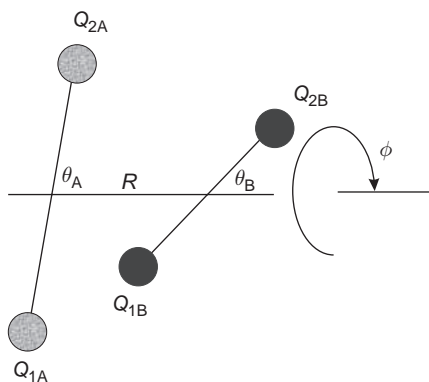


Figure 2.4 Multipole expansion for a pair of diatomics

Molecule A is centred at the origin, whilst molecule B has its centre a distance R away along the horizontal axis. The inclinations to the axis are θ_A and θ_B , and ϕ gives the relative orientation of the two molecules. The distance between the two molecules is much less than their separation, so we can make convenient approximations such as for the small dipole. After some standard analysis we find that the mutual potential energy of A and B is

$$\begin{aligned}
(4\pi\epsilon_0) U_{AB} &= \frac{Q_A Q_B}{R} + \frac{1}{R^2} (Q_B p_A \cos \theta_A - Q_A p_B \cos \theta_B) \\
&\quad - \frac{p_A p_B}{R^3} (2 \cos \theta_A \cos \theta_B - \sin \theta_A \sin \theta_B \cos \phi) \\
&\quad + \frac{1}{2R^3} (Q_A q_B (3 \cos^2 \theta_B - 1) + Q_B q_A (3 \cos^2 \theta_A - 1)) + \dots \quad (2.7)
\end{aligned}$$

The physical interpretation is as follows. The first term on the right-hand side gives the mutual potential energy of the two charged molecules A and B. The second term gives a contribution due to each charged molecule with the other dipole. The third term is a dipole–dipole contribution, and so on.

If A and B correspond to uncharged molecules, then the leading term is seen to be the dipole–dipole interaction

$$(4\pi\epsilon_0) U_{AB,\text{dip...dip}} = -\frac{p_A p_B}{R^3} (2 \cos \theta_A \cos \theta_B - \sin \theta_A \sin \theta_B \cos \phi) \quad (2.8)$$

The sign and magnitude of this term depend critically on the relative orientation of the two molecules. Table 2.1 shows three possible examples, all of which have $\phi = 0$.

Table 2.1 *Representative dipole–dipole terms for two diatomics*

θ_A	θ_B	Relative orientations	Expression for dipole–dipole U
0	0	Parallel	$-\frac{2p_A p_B}{4\pi\epsilon_0 R^3}$
0	π	Antiparallel	$+\frac{2p_A p_B}{4\pi\epsilon_0 R^3}$
0	$\pi/2$	Perpendicular	0

2.5 Taking Account of the Temperature

We now imagine that the two molecules undergo thermal motion; we keep their separation R constant but allow the angles to vary. The aim is to calculate the average dipole–dipole interaction. Some orientations of the two dipoles will be more energetically favoured than others and we allow for this by including a Boltzmann factor $\exp(-U/k_B T)$, where k_B is the Boltzmann constant and T the thermodynamic temperature. It is conventional to denote mean values by angle brackets $\langle \dots \rangle$ and the mean value of the dipole–dipole interaction is given formally by

$$\langle U_{AB} \rangle_{\text{dip...dip}} = \frac{\int U_{AB} \exp\left(-\frac{U_{AB}}{k_B T}\right) d\tau}{\int \exp\left(-\frac{U_{AB}}{k_B T}\right) d\tau} \quad (2.9)$$

The integral has to be done over all possible values of the angles, keeping R fixed. After some standard integration, we find

$$\langle U_{AB} \rangle_{\text{dip...dip}} = -\frac{2p_A^2 p_B^2}{3k_B T (4\pi \epsilon_0)^2 R^6} \quad (2.10)$$

The overall value is therefore negative, and the term is inversely dependent on the temperature. It also falls off as $1/R^6$.

2.6 Induction Energy

The next step is the case of two interacting molecules, one of which has a permanent dipole moment and one of which is polarizable but does not have a permanent electric dipole moment.

Figure 2.5 shows molecule A with a permanent dipole moment \mathbf{p}_A . I have indicated the direction of \mathbf{p}_A in the diagram, and an arbitrary point P in molecule B. The dipole \mathbf{p}_A is distant R from point P, and makes an angle θ as shown. The molecules are sufficiently far apart for the precise location of the point P inside the second molecule to be irrelevant.

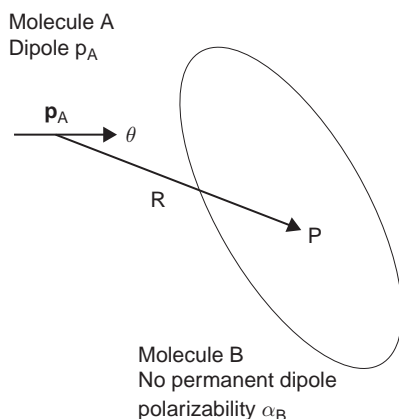


Figure 2.5 Dipole-induced dipole

The basic physical idea is that the electric dipole \mathbf{p}_A induces a dipole in molecule B, since B is polarizable. We evaluate the potential energy involved and finally average over all possible geometrical arrangements, for a fixed value of the intermolecular separation.

The steps involved are as follows. The electrostatic potential due to the small dipole \mathbf{p}_A is

$$\phi_A(\mathbf{R}) = \frac{1}{4\pi \epsilon_0} \frac{\mathbf{p}_A \mathbf{R}}{R^3}$$

This is related to the electrostatic field by the general formula

$$\mathbf{E}(\mathbf{R}) = -\text{grad } \phi(\mathbf{R})$$

and direct differentiation gives the following formula:

$$\mathbf{E}_A(\mathbf{R}) = -\frac{1}{4\pi\epsilon_0} \left\{ \frac{\mathbf{p}_A}{R^3} - 3\frac{\mathbf{p}_A\mathbf{R}}{R^5}\mathbf{R} \right\} \quad (2.11)$$

Molecule A therefore generates an electrostatic field in the region of molecule B, according to the above vector equation. The modulus of this vector at point P is

$$E_A = \frac{1}{4\pi\epsilon_0} \frac{p_A}{R^3} \sqrt{(1 + 3\cos^2\theta)} \quad (2.12)$$

This electrostatic field induces a dipole in molecule B. For the sake of argument, I will assume that the induced dipole is in the direction of the applied field (and so we need not worry about the fact that the polarizability is a tensor property).

A calculation of the resulting mutual potential energy U_{AB} gives

$$U_{AB} = -\frac{1}{(4\pi\epsilon_0)^2} \frac{\alpha_B p_A^2}{R^6} \frac{1}{2} (3\cos^2\theta + 1) \quad (2.13)$$

Polarizabilities are positive quantities and so U_{AB} is negative for all values of θ at a given intermolecular separation. This is quite different from the dipole–dipole interaction, where some alignments of the dipoles gave a positive contribution to the mutual potential energy and some gave a negative one.

Finally we have to average over all possible alignments keeping the internuclear separation fixed. This averaging has again to be done using the Boltzmann weightings, and we find eventually an expression for the *induction* contribution to the mutual potential energy of A and B:

$$\langle U_{AB} \rangle_{\text{ind}} = -\frac{1}{(4\pi\epsilon_0)^2} \frac{p_A^2 \alpha_B}{R^6} \quad (2.14)$$

Note that the interaction falls off as $1/R^6$ just as for the dipole–dipole interaction, but this time there is no temperature dependence. For two identical A molecules each with permanent electric dipole p_A and polarizability α_A the expression becomes

$$\langle U_{AA} \rangle_{\text{ind}} = -\frac{2}{(4\pi\epsilon_0)^2} \frac{p_A^2 \alpha_A}{R^6} \quad (2.15)$$

This of course has to be added to the dipole–dipole expression of the previous section.

2.7 Dispersion Energy

It is an experimental fact that the inert gases can be liquefied. Atoms do not have permanent electric moments, so the dipole–dipole and induction contributions to the mutual potential energy of an array of inert gas atoms must both be zero. There is clearly a third interaction mechanism (referred to as *dispersion*), and this was first identified by Fritz W. London in 1930.

The two contributions to the mutual potential energy discussed in previous sections can be described by classical electromagnetism. There is no need to invoke the concepts of quantum mechanics. Dispersion interactions can only be correctly described using the

language of quantum mechanics. Nevertheless, the following qualitative discussion is to be found in all elementary texts:

The electrons in an atom or molecule are in continual motion, even in the ground state. So, although on average the dipole moment of a spherically symmetrical system is zero, at any instant a temporary dipole moment can occur. This temporary dipole can induce a further temporary dipole in a neighbouring atom or molecule and, as in the case of the inductive interaction, the net effect will be attractive.

Paul K. L. Drude gave a simple quantum mechanical description of the effect, and his theory suggests that the dispersion contribution can be written

$$\langle U \rangle_{\text{disp}} = - \left(\frac{D_6}{R^6} + \frac{D_8}{R^8} + \frac{D_{10}}{R^{10}} + \dots \right) \quad (2.16)$$

The first term (that I have written D_6) is to be identified with the instantaneous dipole-induced dipole mechanism. The higher terms are caused by instantaneous quadrupole-induced quadrupoles, etc. According to Drude's theory

$$D_6 = - \frac{3\alpha^2 \varepsilon_1}{4 (4\pi \varepsilon_0)^2} \quad (2.17)$$

In this expression, ε_1 is the first excitation energy of the atomic or molecular species concerned. Once again I apologise for the clash of symbols, for ε_0 is of course the permittivity of free space.

The dispersion energy is again seen to be attractive and to fall off as $1/R^6$.

2.8 Repulsive Contributions

When two molecular species approach so closely that their electron clouds overlap, the positively charged nuclei become less well shielded by the negative electrons and so the two species repel each other. The repulsive term is sometimes written

$$U_{\text{rep}} = A \exp(-BR) \quad (2.18)$$

where A and B are specific to the particular molecular pair and have to be determined from experiment. The precise form of the repulsive term is not well understood; all that is certain is that it must fall off quickly with distance, and the exponential function is therefore a possible suitable candidate.

The total interaction is $U = U_{\text{rep}} + U_{\text{dip...dip}} + U_{\text{ind}} + U_{\text{disp}}$, which we can write

$$U = A \exp(-BR) - \frac{C}{R^6} \quad (2.19)$$

since all the attractive forces fall off as $1/R^6$. This is known as the exp-6 potential. In the *Lennard-Jones 12-6* potential, we take a repulsive term proportional to $1/R^{12}$ and so

$$U_{L-J} = \frac{C_{12}}{R^{12}} - \frac{C_6}{R^6} \quad (2.20)$$

Once again the coefficients C_{12} and C_6 have to be determined from experiments on the species under study. The Lennard-Jones (L-J) potential is usually written in terms of the well depth ε and the distance of closest approach σ as follows:

$$U_{L-J} = 4\varepsilon \left(\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right) \quad (2.21)$$

The two L-J parameters σ and ε have been deduced for a range of atoms. The quantity ε/k_B (which has dimensions of temperature) is usually recorded in the literature rather than ε . Sample atomic parameters are listed in Table 2.2.

Table 2.2 *Representative L-J atomic parameters*

Atom	$(\varepsilon/k_B)/K$	σ/pm
He	10.22	258
Ne	35.7	279
Ar	124	342
Xe	229	406

Over the years, people have extended these ideas to the interaction of simple molecules. Some caution is needed; the interaction between two molecules will generally depend on the precise details of their orientation, and the values given in Table 2.3 must be interpreted as some kind of geometrical average.

Figure 2.6 shows a L-J 12-6 plot for argon.

Table 2.3 *L-J parameters for simple molecules*

Molecule	$(\varepsilon/k_B)/K$	σ/pm
H ₂	33.3	297
O ₂	113	343
N ₂	91.5	368
Cl ₂	357	412
Br ₂	520	427
CO ₂	190	400
CH ₄	137	382
CCl ₄	327	588
C ₂ H ₄	205	423
C ₆ H ₆	440	527

Source: Hirschfelder *et al.* (1954).

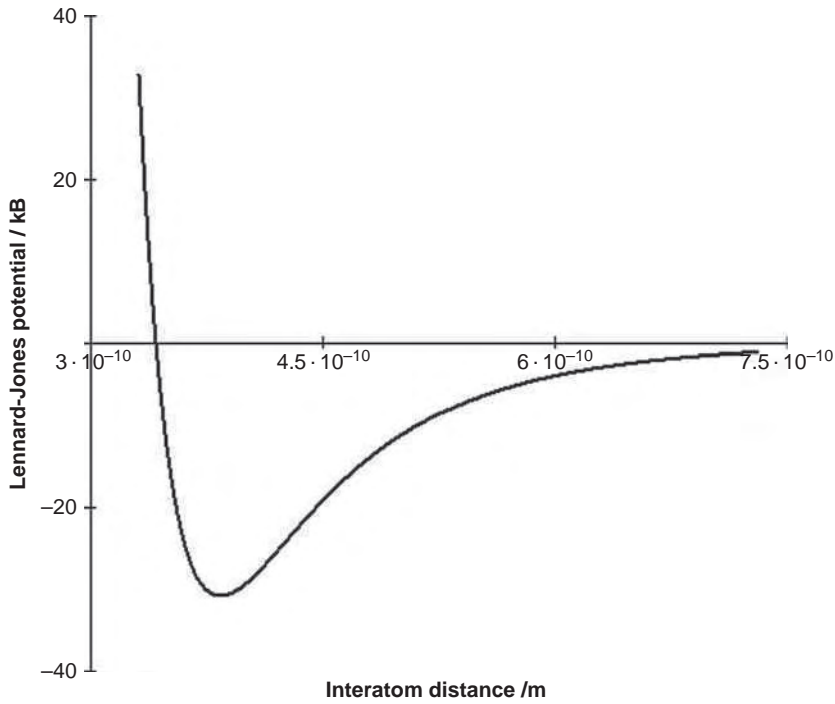


Figure 2.6 Lennard-Jones 12-6 potential for Ar–Ar. The plot is of ϵ/k_B against R .

2.9 Combination Rules

A large number of L-J parameters have been deduced over the years, but they relate to pairs of like atoms. Rather than try to deduce corresponding parameters for unlike pairs, it is usual to use so-called *combination rules*, which enable us to relate the C_{12} and C_6 parameters for an unlike-atom pair A–B to those of A–A and B–B. The use of such combination rules is common in subjects like chemical engineering, and is widely applied to many physical properties.

There are three common combination rules in the literature. The first is

$$C_{12,ij} = \left(\frac{R_i^*}{2} + \frac{R_j^*}{2} \right)^{12} \sqrt{\epsilon_i \epsilon_j}$$

$$C_{6,ij} = 2 \left(\frac{R_i^*}{2} + \frac{R_j^*}{2} \right)^6 \sqrt{\epsilon_i \epsilon_j} \quad (2.22)$$

where R_i^* is the minimum energy separation for two atoms of type i and ϵ_i the well depth. The second is

$$\begin{aligned}
 C_{12,ij} &= 4 (\sigma_i \sigma_j)^6 \sqrt{\varepsilon_i \varepsilon_j} \\
 C_{6,ij} &= 4 (\sigma_i \sigma_j)^3 \sqrt{\varepsilon_i \varepsilon_j}
 \end{aligned}
 \tag{2.23}$$

Finally

$$\begin{aligned}
 C_{6,ij} &= C \frac{\alpha_i \alpha_j}{\sqrt{\frac{\alpha_i}{N_i}} + \sqrt{\frac{\alpha_j}{N_j}}} \\
 C_{12,ij} &= \frac{1}{2} C_{6,ij} (R_i + R_j)^6
 \end{aligned}
 \tag{2.24}$$

where α_i is the dipole polarizability of atom i , N_i the number of valence electrons and R_i the van der Waals radius.

2.10 Comparison with Experiment

You will have got the idea by now that we have to determine the parameters in any pair potential by appeal to experiment. There are two kinds of experiment to consider. First there are those that are essentially in the gas phase, where pairs of atoms genuinely interact with each other unencumbered by other species. This means that the total mutual potential energy is given by the sum of the interacting pairs. Second there are experiments that essentially relate to condensed phases, where the interacting particles are sufficiently close to raise doubts about the credibility of the pair wise additivity assumption.

2.10.1 Gas Imperfections

The deviation of gases from perfect behaviour can be expressed in the form of a virial equation of state

$$\frac{pV}{nRT} = 1 + \frac{nB(T)}{V} + \frac{n^2C(T)}{V^2} + \dots
 \tag{2.25}$$

where the virial coefficients $B(T)$, $C(T)$, . . . depend on the temperature and on the characteristics of the species under study. Here, n is the amount of substance, p the pressure, V the volume, R the gas constant and T the thermodynamic temperature. $B(T)$ is called the *second virial coefficient*, $C(T)$ is called the *third virial coefficient*, and so on. They have to be determined experimentally by fitting the pVT data of the gas under study.

The virial equation of state has a special significance in that the virial coefficients can be related directly to molecular properties. $B(T)$ depends on the pair potential $U(R)$ in the following way:

$$B(T) = 2\pi \int_0^\infty \left(1 - \exp\left(-\frac{U(R)}{k_B T}\right) \right) R^2 dR
 \tag{2.26}$$

2.10.2 Molecular Beams

In a molecular beam experiment, a beam of mono-energetic molecules is produced and allowed to collide either with other molecules in a scattering chamber, or with a similar beam travelling at right angles to the original beam. Measurements of the amount by which

the incident beam is reduced in intensity, or the number of molecules scattered in a particular direction allow determination of the parameters in the pair potential.

2.11 Improved Pair Potentials

The L-J 12-6 potential for a pair of interacting atoms

$$\begin{aligned} U_{\text{L-J}}(R) &= \frac{C_{12}}{R^{12}} - \frac{C_6}{R^6} \\ &= 4\varepsilon \left(\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right) \end{aligned}$$

contains two parameters (ε and σ , or C_{12} and C_6) that have to be determined by experimental observation. The exp-6 model

$$U(R) = A \exp(-BR) - \frac{C}{R^6}$$

contains three parameters, which allows for a little more flexibility.

The Born–Mayer–Huggins potential

$$U(R) = A \exp(-BR) - \frac{C}{R^6} - \frac{D}{R^8} \quad (2.27)$$

contains four parameters.

More recent investigations have concentrated on pair potentials having many more disposable parameters, for example

$$\begin{aligned} U(r) &= \exp\left(A \left(1 - \frac{R}{\sigma}\right)\right) \sum_{i=0}^n B_i \left(\frac{R}{\sigma} - 1\right)^i \\ &+ \frac{C_6}{\left(D + \left(\frac{R}{\sigma}\right)^6\right)} + \frac{C_8}{\left(D + \left(\frac{R}{\sigma}\right)^8\right)} + \frac{C_{10}}{\left(D + \left(\frac{R}{\sigma}\right)^{10}\right)} \end{aligned} \quad (2.28)$$

There is no obvious relationship between the various parameters in these different models; they all have to be determined by fitting experimental data. Roughly speaking, the more parameters there are the better.

2.12 A Numerical Potential

An alternative approach is to abandon all attempts at a precise functional form and try to find a numerically tabulated potential by direct comparison between theory and experiment. In their famous 1969 paper, which is also a good reference for historically important pair potentials, Dymond and Alder (1969) say

A study of the effect of variations in the interatomic potential of argon on the calculated properties of the dilute gas has led to a numerically tabulated potential which satisfactorily agrees with all available equilibrium and transport data. This potential differs significantly from previously used pair potentials.

Their approach makes use of the second virial coefficient mentioned above.

2.13 Site–Site Potentials

The L-J potential has played an important role in the history of molecular modelling. Early work focused on atoms, but, as I explained, there were many ambitious attempts to model simple molecules as if they were in some way L-J atoms, and the parameters have to be interpreted as applying to some kind of average over molecular rotations (and presumably vibrations).

Suppose now that we want to try and understand the interaction between two dinitrogen molecules in more depth. In view of our discussion above, the instantaneous interaction energy will clearly depend on the separation of the two diatoms, together with their mutual angular arrangement in space.

Figure 2.7 shows two such dinitrogens, oriented arbitrarily in space with respect to each other. Nitrogen A and nitrogen B make up a stable diatom, as do atoms C and D. We ignore the fact that the molecules have vibrational energy, and the two diatoms are taken to be rigid. As a first approximation, the mutual potential energy of the pair of diatoms could be calculated by adding together the appropriate L-J parameters. If I write as shorthand the L-J interaction between atoms A and C as

$$U_{L-J}(A, C) = 4\epsilon \left(\left(\frac{\sigma}{R_{AC}} \right)^{12} - \left(\frac{\sigma}{R_{AC}} \right)^6 \right)$$

then the total interaction between the two diatomics is taken as

$$U_{L-J} = U_{L-J}(A, C) + U_{L-J}(A, D) + U_{L-J}(B, C) + U_{L-J}(B, D) \quad (2.29)$$

Such a potential is called a *site–site potential*. We will meet such potentials later in the text.

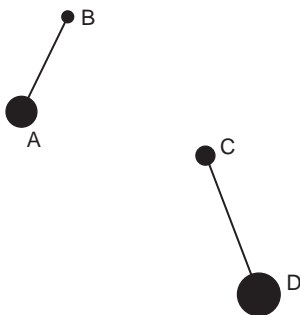


Figure 2.7 Two dinitrogen molecules, arbitrarily oriented

2.14 Problem Set

A problem set, with solutions, is available on the associated website.

References

- Dymond, J.H. and Alder, B.J. (1969) *J. Chem. Phys.*, **51**, 309–320.
Hirschfelder, J.O., Curtiss, C.F. and Bird, R.B. (1954) *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York.

3

Balls on Springs

The theory of intermolecular forces relates to atomic and/or molecular species that are some distance apart (say, a few bond lengths). We saw in Chapter 2 that progress can be made in such a theory without normally invoking the concepts of quantum mechanics. If we want to truly understand why two atoms combine together to give a chemical bond, and how bonds get broken and reformed in chemical reactions, then we enter the realms of *valence theory*. Quantum mechanics plays a dominant part in such discussions.

These are simple-minded comments and my arbitrary division of molecular interactions is subjective. At first sight, the stability of an NaCl ion pair can be explained in terms of elementary electrostatics, and we can usefully model argon liquid without recourse to quantum mechanics (apart from the London dispersion potential, which is a 'pure' quantum mechanical effect). A C–C bond in ethane is at first sight a quantum mechanical animal, and we will certainly have to invoke quantum mechanical ideas to explain the reaction of ethene with dichlorine. But there are grey areas that I can bring to your attention by considering the phenomenon of hydrogen bonding. The hydrogen bond is an attractive interaction between a proton donor X–H and a proton acceptor Y in the same or a different molecule



The bond is usually symbolized by three dots as shown above, in order to reconcile the existence of compounds such as



with the trivalence of nitrogen, the divalence of oxygen in oxonium salts and other compounds that apparently break the classical valence rules. Hydrogen bonds typically have strengths of 10–100 kJ mol⁻¹. The lone pairs of oxygen and nitrogen and the partially charged character of the proton were eventually recognized as the sources of this bond. The first reference to this 'weak bond' was made by Latimer and Rodebush (1920).

The individual monomers X–H and Y retain their chemical identity to a large extent on hydrogen bond formation; in other words no new covalent bond gets made. A great deal of evidence suggests that simple electrostatic models of the hydrogen bond give perfectly acceptable quantitative descriptions of the structure, vibrations and electric dipole moments of such hydrogen bonded species.

The hydrogen-bonded species



is well known and has been well studied, but it cannot be written



because the proton is equally shared between the two fluorine atoms. Such a species is best thought of as covalently bound, and has to be treated by the methods of molecular quantum theory.

Having warned about bond breaking and bond making, I should tell you that a great deal of molecular modelling is concerned with the prediction and rationalization of molecular bond lengths and bond angles. Here we usually deal with isolated molecules in the gas phase and the theoretical treatments often refer to 0 K. A surprising amount of progress can be made by treating molecules as structureless balls (atoms) held together with springs (bonds). The array of balls and springs is then treated according to the laws of classical mechanics. Such calculations are remarkably accurate, and are taken very seriously.

3.1 Vibrational Motion

To get started, consider a particle of mass m lying on a frictionless horizontal table, and attached to the wall by a spring as shown in Figure 3.1.

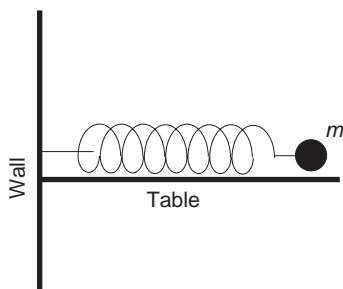


Figure 3.1 *Ball attached to wall by spring*

The particle is initially at rest, when the length of the spring is R_e (where the subscript 'e' stands for equilibrium). If we stretch the spring, it exerts a restoring force on the particle whilst if we compress the spring there is also a force that acts to restore the particle to its equilibrium position. If R denotes the length of the spring, then the extension is $R - R_e$ and

if F_s is the restoring force due to the spring, it is often observed experimentally that the force is directly proportional to the extension:

$$F_s = -k_s (R - R_e) \quad (3.1)$$

The constant of proportionality k_s is called the force constant and it tells us the strength of the spring. This law is known as *Hooke's law* and it applies to very many springs made from many different materials. It invariably fails for large values of the extension, but is good for small deviations from equilibrium.

Suppose that we now set the particle in motion, so that it oscillates about R_e . According to Newton's second law we have

$$m \frac{d^2 R}{dt^2} = -k_s (R - R_e) \quad (3.2)$$

This second-order differential equation has the general solution

$$R = R_e + A \sin \left(\sqrt{\frac{k_s}{m}} t \right) + B \cos \left(\sqrt{\frac{k_s}{m}} t \right) \quad (3.3)$$

where A and B are constants of integration. These constants have to be fixed by taking account of the *boundary conditions*. For example, if the particle starts its motion at time $t = 0$ from $R = R_e$ then we have

$$R_e = R_e + A \sin \left(\sqrt{\frac{k_s}{m}} 0 \right) + B \cos \left(\sqrt{\frac{k_s}{m}} 0 \right)$$

from which we deduce that $B = 0$ for this particular case. Normally we have to find A and B by a similar procedure.

The trigonometric functions sine and cosine repeat every 2π and a little manipulation shows that the general solution of Equation (3.3) can also be written

$$R = R_e + A \sin \left(\sqrt{\frac{k_s}{m}} \left(t + 2\pi \sqrt{\frac{m}{k_s}} \right) \right) + B \cos \left(\sqrt{\frac{k_s}{m}} \left(t + 2\pi \sqrt{\frac{m}{k_s}} \right) \right)$$

The quantity $\sqrt{k_s/m}$ has the dimension of inverse time and it is obviously an important quantity. We therefore give it a special symbol (ω) and name (the *angular vibration frequency*). We often write the general solution as

$$R = R_e + A \sin (\omega t) + B \cos (\omega t) \quad (3.4)$$

A typical solution is shown as Figure 3.2 (for which I took $A = 1$ m, $B = 0$, $m = 1$ kg and $k_s = 1$ N m⁻¹). Such motions are called *simple harmonic*. At any given time, the displacement of the particle from its equilibrium position may be nonzero, but it should be clear from Figure 3.2 that the average value of the displacement $R - R_e$ is zero. As noted earlier, it is usual to denote average values by angle brackets $\langle \dots \rangle$ and so we write

$$\langle R - R_e \rangle = 0$$

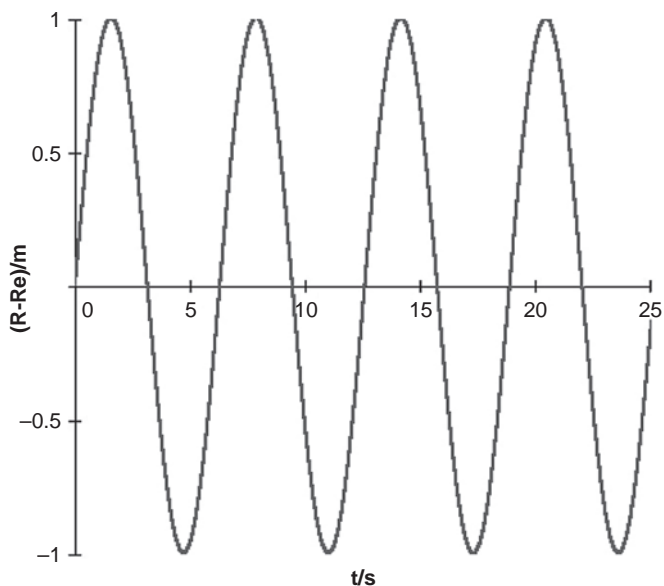


Figure 3.2 Simple harmonic motion

It should also be clear from Figure 3.3 that the average value of $(R - R_e)^2$ is not zero. A direct calculation using $\langle \sin^2(\omega t) \rangle = 1/2$ gives

$$\langle (R - R_e)^2 \rangle = \frac{1}{2} (A^2 + B^2) \quad (3.5)$$

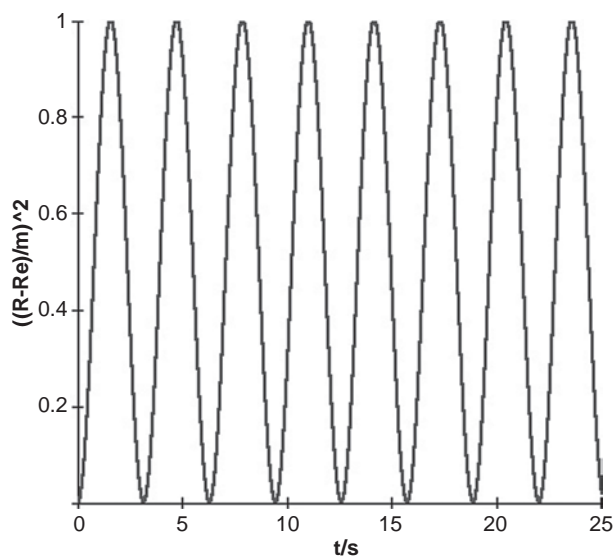


Figure 3.3 Variation of $(R - R_e)^2$ with time

3.2 The Force Law

If we use the general one-dimensional result that links force and mutual potential energy

$$U(x) = - \int F(x) dx$$

we find

$$U(R) = U(R_e) + \frac{1}{2}k_s(R - R_e)^2 \quad (3.6)$$

where $U(R_e)$ is the value of $U(R)$ at the equilibrium position. As mentioned earlier, we often set this constant of integration arbitrarily to zero and so

$$U(R) = \frac{1}{2}k_s(R - R_e)^2 \quad (3.7)$$

Because the motion is simple harmonic, we refer to this potential as a *harmonic potential*. The potential energy varies over each cycle, in the same way as shown in Figure 3.3. The average value of the potential energy over a cycle is

$$\begin{aligned} \langle U(R) \rangle &= \frac{1}{2}k_s \langle (R - R_e)^2 \rangle \\ &= \frac{1}{4}(A^2 + B^2) \end{aligned} \quad (3.8)$$

Finally, the kinetic energy T is given by

$$T(R) = \frac{1}{2}m \left(\frac{dR}{dt} \right)^2$$

The average value is

$$\begin{aligned} \langle T(R) \rangle &= \frac{1}{4}m\omega^2(A^2 + B^2) \\ &= \frac{1}{4}k_s(A^2 + B^2) \end{aligned} \quad (3.9)$$

The kinetic and potential energies vary with time, but the total energy $U(R) + T(R)$ is constant: it does not vary with time. The average value of the kinetic energy over a cycle is equal to the average value of the potential energy, each of which is one half of the total energy. I am going to use the symbol ε for energy when referring to a single atom or molecule, throughout the text. The total energy ε can take any value with no restrictions.

3.3 A Simple Diatomic

Consider next the Hooke's law model of a diatomic molecule (Figure 3.4). The atoms have masses m_1 and m_2 , and the spring has a harmonic force constant k_s .

I am going to consider the motion along the horizontal axis, which I will refer to as the x axis. The x coordinates of the two atoms are x_1 and x_2 relative to an arbitrary axis,

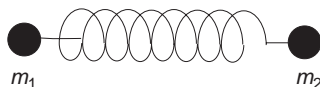


Figure 3.4 Diatomic molecule

the equilibrium length of the spring is R_e and the length of the extended spring at some given time is

$$R = x_2 - x_1$$

The spring extension is therefore

$$x_2 - x_1 - R_e$$

Considering atom 1, the spring exerts a force of $k_s(x_2 - x_1 - R_e)$ and so, according to Newton's second law

$$\begin{aligned} m_1 \frac{d^2 x_1}{dt^2} &= k_s (x_2 - x_1 - R_e) \\ &= k_s (R - R_e) \end{aligned} \quad (3.10)$$

As far as atom 2 is concerned, the extended spring exerts a force of magnitude $k_s(x_2 - x_1 - R_e)$ in the direction of decreasing x_2 and so

$$\begin{aligned} m_2 \frac{d^2 x_2}{dt^2} &= -k_s (x_2 - x_1 - R_e) \\ &= -k_s (R - R_e) \end{aligned} \quad (3.11)$$

After a little rearrangement we find

$$\begin{aligned} \frac{d^2 R}{dt^2} &= -\frac{k_s}{m_2} (R - R_e) - \frac{k_s}{m_1} (R - R_e) \\ &= -k_s \left(\frac{1}{m_1} + \frac{1}{m_2} \right) (R - R_e) \end{aligned} \quad (3.12)$$

We now define a quantity μ , called the *reduced mass*, by

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}$$

and so we have

$$\mu \frac{d^2 R}{dt^2} = -k_s (R - R_e) \quad (3.13)$$

which is identical to Equation (3.2) already derived for a single particle of mass μ on a spring.

The general solution is therefore

$$R = R_e + A \sin \left(\sqrt{\frac{k_s}{\mu}} t \right) + B \cos \left(\sqrt{\frac{k_s}{\mu}} t \right) \quad (3.14)$$

and the angular frequency is $\omega = \sqrt{k_s/\mu}$.

It is easy to demonstrate that the potential energy is

$$U = \frac{1}{2}k_s(x_2 - x_1 - R_e)^2$$

and the total energy ε_{vib} of the harmonically vibrating diatomic is therefore

$$\varepsilon_{\text{vib}} = \frac{1}{2}m_1\left(\frac{dx_1}{dt}\right)^2 + \frac{1}{2}m_2\left(\frac{dx_2}{dt}\right)^2 + \frac{1}{2}k_s(x_2 - x_1 - R_e)^2 \quad (3.15)$$

3.4 Three Problems

This simple treatment suggests three problems. First, how do we determine the spring constant for a simple molecule such as $^1\text{H}^{35}\text{Cl}$ or $^{12}\text{C}^{16}\text{O}$? Second, how good is the harmonic approximation? Third, have we missed anything by trying to treat a molecular species as if it obeyed the laws of classical mechanics rather than quantum mechanics?

The three questions are interlinked, but let me start with the third one. The experimental evidence suggests that we have made a serious error in neglecting the quantum mechanical details. If we irradiate a gaseous sample of $^1\text{H}^{35}\text{Cl}$ with infrared radiation, it is observed that the molecules strongly absorb radiation of wavenumber 2886 cm^{-1} . With hindsight we would of course explain the observation by saying that the molecular vibrational energies are *quantized*. A major flaw of the classical treatment is that the total vibrational energy is completely unrestricted and quantization does not arise.

The quantum mechanical treatment of a harmonically vibrating diatomic molecule is given in all the elementary chemistry texts. The results are quite different from the classical ones.

1. The vibrational energy cannot take arbitrary values, it is *quantized*.
2. There is a single quantum number ν which takes values $0, 1, 2, \dots$. This is called the *vibrational quantum number*.
3. Vibrational energies ε_{vib} are given by

$$\varepsilon_{\text{vib}} = \frac{h}{2\pi} \sqrt{\frac{k_s}{\mu}} \left(\nu + \frac{1}{2} \right)$$

where h is Planck's constant.

The results are usually summarized on an energy level diagram, such as that shown in Figure 3.5. I have just drawn the first four vibrational energy levels, but there are an infinite number of them. According to the harmonic model, the spacing between the levels is constant.

A careful investigation into the mechanism by which electromagnetic radiation interacts with matter suggests that transitions between these vibrational energy levels are allowed, provided the vibrational quantum number changes by just 1 unit. So molecules with $\nu = 0$ can absorb radiation of exactly the right energy for promotion to $\nu = 1$. Molecules with $\nu = 1$ can either absorb radiation with exactly the right energy for promotion to $\nu = 2$ or they can emit radiation and fall to $\nu = 0$, and so on.

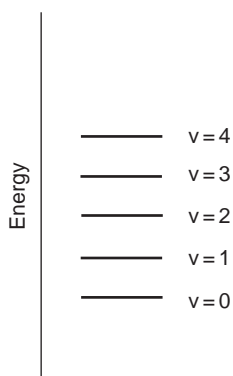


Figure 3.5 Vibrational energy levels

According to the quantum model, then, molecules can only have certain vibrational energies and this behaviour is totally at variance with the classical treatment.

Also, the quantum treatment differs from the classical treatment in that the lowest energy is that with $v = 0$, where the energy is nonzero. This is called the *zero-point* energy. According to the classical treatment, a molecule can be completely at rest. According to the quantum treatment, the lowest vibrational energy allowed is the zero-point energy.

How do we measure the spring constant? According to the harmonic quantum model, the energy difference between any consecutive pair of energy levels is given by

$$\Delta\varepsilon = \frac{h}{2\pi} \sqrt{\frac{k_s}{\mu}} \quad (3.16)$$

so all we need to do is to measure this energy difference experimentally. The reduced mass μ of the $^1\text{H}^{35}\text{Cl}$ isotopic species is 1.6267×10^{-27} kg and substitution of the experimental value (2886 cm^{-1}) into the energy difference gives the harmonic force constant as 480.7 N m^{-1} .

In fact, there is more to the experiment than I have told you. Spectroscopic experiments are done at finite temperatures and a given sample of N molecules may have many energy levels populated. Relative populations N_v are given by the Boltzmann formula

$$N_v \propto \exp\left(-\frac{\varepsilon_v}{k_{\text{B}}T}\right) \quad (3.17)$$

Substitution of values into the formula shows that for many everyday diatomic molecules at everyday temperatures, the only vibrational level populated is that with $v = 0$. So an infrared absorption spectrum should just show a single absorption, corresponding to the transition $v = 0$ to $v = 1$.

A closer examination of the $^1\text{H}^{35}\text{Cl}$ spectrum shows weak absorptions at 5668 , 8347 , $\dots \text{ cm}^{-1}$, which are nearly (but not exactly) two and three times the fundamental vibration frequency.

The existence of these lines in the spectrum shows that our assumption of Hooke's law is not completely correct. Figure 3.6 shows the 'experimental' energy level diagram compared to the harmonic one.

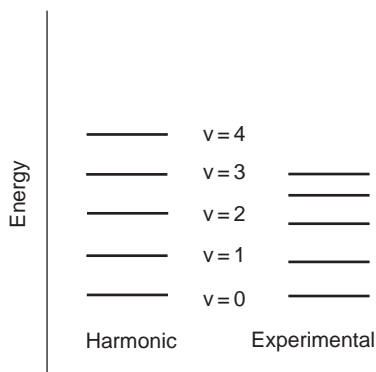


Figure 3.6 Harmonic versus experimental energy levels

Our conclusion from the experimental data is that vibrational energy levels get progressively closer together as the quantum number increases. This suggests that whilst the harmonic model is a reasonable one, we need to look more carefully at the form of the potential in order to get better agreement with experiment.

3.5 The Morse Potential

Professional spectroscopists would be unhappy with the idea of using Hooke's law as a model for the vibrational motion. They would be more concerned with matching their experimental energy levels to a more accurate potential. Many such potentials have been used over the years, with that due to Morse being widely quoted in elementary chemistry texts. The Morse potential is as follows:

$$U = D_e \{1 - \exp[-\beta(R - R_e)]\}^2 \quad (3.18)$$

where D_e is the depth of the potential well, i.e. the thermodynamic dissociation energy, and

$$\begin{aligned} \beta &= \sqrt{\frac{\left(\frac{d^2U}{dR^2}\right)}{2D_e}} \\ &= \frac{\omega_e}{2} \sqrt{\frac{2\mu}{D_e}} \end{aligned} \quad (3.19)$$

This potential contains three parameters D_e , ω_e and R_e , and so should be capable of giving a better representation to the potential energy curve than the simple harmonic, which contains just the two parameters k_s and R_e .

In the case of $^1\text{H}^{35}\text{Cl}$, a simple calculation shows that the dissociation energy

$$D_e = D_0 + \frac{1}{2}h(2\pi\omega_e) \quad (3.20)$$

is $4.430 + 0.186 \text{ eV} = 4.616 \text{ eV}$. The Morse potential for $^1\text{H}^{35}\text{Cl}$ is shown in Figure 3.7, compared to the simple harmonic model. The full curve is the simple harmonic potential, the dashed curve the Morse potential.

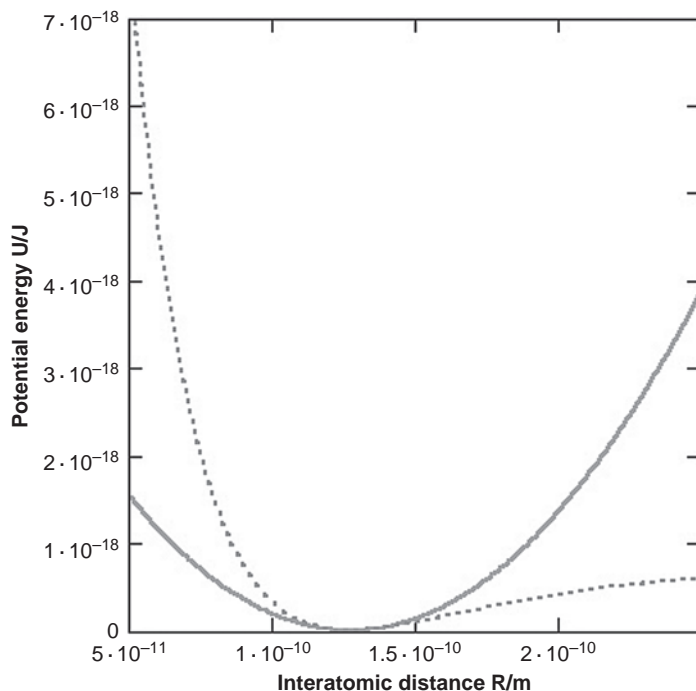


Figure 3.7 Simple harmonic and Morse curves for HCl

3.6 More Advanced Potentials

More often than not, the following spectroscopic constants are available for a diatomic molecule:

- R_e the equilibrium internuclear separation;
- D_e the dissociation energy;
- k_s the force constant;
- $\omega_e x_e$ the anharmonicity constant (sometimes written x_e only);
- α_e the vibration–rotation coupling constant.

Usually these five constants can be found to good experimental accuracy.

There are a number of three- to five-parameter potential functions in the literature, of which the Morse potential is the most popular. A typical five-parameter potential is the Linnert function (Linnert 1940, 1942)

$$U(x) = \frac{K}{x^m} - b \exp(-tx) [1 + g_1(x - 1) + g_2(x - 1)^2]$$

The five parameters K , b , t , g_1 and g_2 can be related to the molecular constants. At this point I should explain how we recover the force constant from such a complicated expression, and to do this I use a Taylor expansion of the potential about the equilibrium bond length:

$$U(R) = U(R_e) + (R - R_e) \left(\frac{dU}{dR} \right)_{R=R_e} + \frac{1}{2} (R - R_e)^2 \left(\frac{d^2U}{dR^2} \right)_{R=R_e} + \dots \quad (3.21)$$

$U(R)$ is obviously equal to $U(R_e)$ when $R = R_e$, and this fixes the constant of integration. The equation is sometimes written as

$$U(R) - U(R_e) = (R - R_e) \left(\frac{dU}{dR} \right)_{R=R_e} + \frac{1}{2} (R - R_e)^2 \left(\frac{d^2U}{dR^2} \right)_{R=R_e} + \dots$$

or even

$$U(R) = (R - R_e) \left(\frac{dU}{dR} \right)_{R=R_e} + \frac{1}{2} (R - R_e)^2 \left(\frac{d^2U}{dR^2} \right)_{R=R_e} + \dots$$

where it is understood that $U(R)$ is measured relative to the potential energy minimum (that is to say, we take the zero as $U(R_e)$).

The quantity dU/dR is of course the gradient of U . The second derivative evaluated at the minimum where $R = R_e$ is called the (harmonic) force constant. To find R_e we solve the equation $dU/dR = 0$ and substitute this value into the second derivative to evaluate the force constant. In the special case of a harmonic potential, the second derivative is a constant and is equal to the force constant.

References

- Latimer, W.M. and Rodebush, W.H. (1920) *J. Am. Chem. Soc.*, **42**, 1419.
 Linnett, J.W. (1940) *Trans. Faraday Soc.*, **36**, 1123.
 Linnett, J.W. (1942) *Trans. Faraday Soc.*, **38**, 1.

4

Molecular Mechanics

In Chapter 3, I showed you how to use classical mechanics to model the vibrational motion of a diatomic molecule. I also explained the shortcomings of this treatment, and hinted at applications where a quantum mechanical model would be more appropriate. We will deal specifically with quantum mechanical models in later chapters.

4.1 More About Balls on Springs

It is time to move on to more complicated molecules. I want to start the discussion by considering the arrangement of balls on springs shown in Figure 4.1.

We assume that the springs each satisfy Hooke's law. I will call the spring constant of the left-hand spring k_1 and the spring constant of the right-hand spring k_2 . The equilibrium position corresponds to the two masses having x coordinates $R_{1,e}$ and $R_{2,e}$, and we constrain the motion so that the springs can only move along the x axis. The particle masses are shown in Figure 4.1.

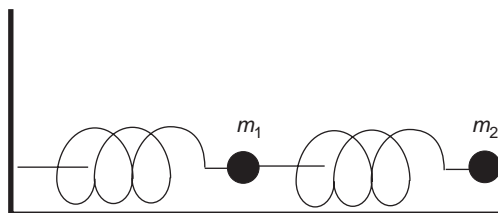


Figure 4.1 Two balls, two springs

We then stretch the system so extending the two springs, and I will call the instantaneous positions of the two masses x_1 and x_2 . The extensions of the springs from their equilibrium positions are

$$\begin{aligned}\xi_1 &= x_1 - R_{1,e} \\ \xi_2 &= x_2 - R_{2,e}\end{aligned}$$

Consider the left-hand spring; it exerts a restoring force on particle 1 of $-k_1\xi_1$. Now consider the right-hand spring. This spring is stretched by an amount $(\xi_2 - \xi_1)$, and so it exerts a force of $k_2(\xi_2 - \xi_1)$; this force acts to the left on particle 2 and to the right on particle 1. Application of Newton's second law gives

$$\begin{aligned}k_2(\xi_2 - \xi_1) - k_1\xi_1 &= m_1 \frac{d^2\xi_1}{dt^2} \\ -k_2(\xi_2 - \xi_1) &= m_2 \frac{d^2\xi_2}{dt^2}\end{aligned}\quad (4.1)$$

There are many different solutions to these simultaneous differential equations, but it proves possible to find two particularly simple ones called *normal modes of vibration*. These have the property that both particles execute simple harmonic motion at the same angular frequency. Not only that, every possible vibrational motion of the two particles can be described as linear combinations of the normal modes.

Having said that it proves possible to find such solutions where both particles vibrate with the same angular frequency ω , let me assume that there exist such solutions to the equations of motion such that

$$\begin{aligned}\xi_1(t) &= A \sin(\omega t + \varphi_1) \\ \xi_2(t) &= B \sin(\omega t + \varphi_2)\end{aligned}$$

A , B , φ_1 and φ_2 are constants that have to be determined from the boundary conditions.

Differentiating these two equations with respect to time gives

$$\begin{aligned}\frac{d^2\xi_1(t)}{dt^2} &= -\omega^2 A \sin(\omega t + \varphi_1) \\ \frac{d^2\xi_2(t)}{dt^2} &= -\omega^2 B \sin(\omega t + \varphi_2)\end{aligned}$$

and substituting these expressions into the equations of motion gives

$$\begin{aligned}-\frac{(k_1 + k_2)}{m_1}\xi_1 + \frac{k_2}{m_1}\xi_2 &= -\omega^2\xi_1 \\ \frac{k_2}{m_2}\xi_1 - \frac{k_2}{m_2}\xi_2 &= -\omega^2\xi_2\end{aligned}\quad (4.2)$$

These two equations are simultaneously valid only when ω has one of two possible values called the normal mode angular frequencies. In either case, both particles oscillate with the same angular frequency.

In order to investigate these normal modes of vibration, I write the above equations in matrix form, and then find the eigenvalues and eigenvectors as follows:

$$\begin{pmatrix} -\frac{(k_1 + k_2)}{m_1} & \frac{k_2}{m_1} \\ \frac{k_2}{m_2} & -\frac{k_2}{m_2} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = -\omega^2 \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \quad (4.3)$$

Matrix diagonalization gives the allowed values of ω^2 (the eigenvalues), and for each value of $-\omega^2$ we calculate the relevant combinations of the ξ values (the eigenvectors). The eigenvectors of the matrix are called the *normal coordinates*.

4.2 Larger Systems of Balls on Springs

For a molecule comprising N atoms, there are $3N$ Cartesian coordinates. Of these, three can be associated with the position of the centre of mass of the whole molecule and three for the orientation of the molecule at the centre of mass (two for linear molecules). This leaves $3N - 6$ vibrational degrees of freedom ($3N - 5$ if the molecule is linear), and it is appropriate to generalize some concepts at this point. I am going to use matrix notation, in order to make the equations look friendlier.

The molecular potential energy U will depend on $p = 3N - 6$ (independent) variables. For the minute, let me call them q_1, q_2, \dots, q_p , and let me also write $q_{1,e}, q_{2,e}, \dots, q_{p,e}$ for their ‘equilibrium’ values. These coordinates are often referred to as *internal coordinates*, and they will be linear combinations of the Cartesian coordinates.

First of all, for the sake of neatness, I will collect all the q ’s into a column matrix \mathbf{q} . I will also collect together the ‘equilibrium’ values into a column matrix \mathbf{q}_e and the extensions into a column matrix ξ :

$$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_p \end{pmatrix}, \quad \mathbf{q}_e = \begin{pmatrix} q_{1,e} \\ q_{2,e} \\ \vdots \\ q_{p,e} \end{pmatrix}, \quad \xi = \begin{pmatrix} q_1 - q_{1,e} \\ q_2 - q_{2,e} \\ \vdots \\ q_p - q_{p,e} \end{pmatrix} \quad (4.4)$$

I will now write $U(\mathbf{q})$ to indicate the dependence of U on these variables. If I use Taylor’s theorem to expand $U(\mathbf{q})$ about point \mathbf{q}_e then the one-dimensional equation

$$U(R) - U(R_e) = (R - R_e) \left(\frac{dU}{dR} \right)_{R=R_e} + \frac{1}{2} (R - R_e)^2 \left(\frac{d^2U}{dR^2} \right)_{R=R_e} + \dots$$

(given previously in Chapter 3) has to be modified to take account of the larger number of variables. First derivatives become partial first derivatives, and we have to take account of the ‘mixed’ second-order derivatives:

$$U(\mathbf{q}) - U(\mathbf{q}_e) = \sum_{i=1}^p \xi_i \left(\frac{\partial U}{\partial q_i} \right)_{\xi_i=0} + \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \xi_i \xi_j \left(\frac{\partial^2 U}{\partial q_i \partial q_j} \right)_{\xi_i=0, \xi_j=0} + \dots \quad (4.5)$$

In ordinary vector differentiation, we meet the gradient of a scalar field f , defined in Cartesian coordinates as

$$\text{grad } f = \frac{\partial f}{\partial x} \mathbf{e}_x + \frac{\partial f}{\partial y} \mathbf{e}_y + \frac{\partial f}{\partial z} \mathbf{e}_z$$

where \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z are Cartesian unit vectors. When dealing with functions of many variables it proves useful to make a generalization and write the gradient of (for example) U as

$$\text{grad } U = \begin{pmatrix} \frac{\partial U}{\partial q_1} \\ \frac{\partial U}{\partial q_2} \\ \vdots \\ \frac{\partial U}{\partial q_p} \end{pmatrix} \quad (4.6)$$

so $\text{grad } U$ is a column matrix that stores all the partial derivatives. This ‘vector’ will occur many times through the text, and I am going to give it the symbol \mathbf{g} (for gradient).

The second derivatives can be collected into a symmetric $p \times p$ matrix that is called the *Hessian* of U and I will give this the symbol \mathbf{H} . In the case where $p = 3$, we have

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 U}{\partial q_1^2} & \frac{\partial^2 U}{\partial q_1 \partial q_2} & \frac{\partial^2 U}{\partial q_1 \partial q_3} \\ \frac{\partial^2 U}{\partial q_2 \partial q_1} & \frac{\partial^2 U}{\partial q_2^2} & \frac{\partial^2 U}{\partial q_2 \partial q_3} \\ \frac{\partial^2 U}{\partial q_3 \partial q_1} & \frac{\partial^2 U}{\partial q_3 \partial q_2} & \frac{\partial^2 U}{\partial q_3^2} \end{pmatrix} \quad (4.7)$$

The Taylor expansion then becomes

$$U(\mathbf{q}) - U(\mathbf{q}_e) = \xi^T \mathbf{g} + \frac{1}{2} \xi^T \mathbf{H} \xi + \dots \quad (4.8)$$

Both the gradient and the Hessian have to be evaluated at the point \mathbf{q}_e , and so you will sometimes see the equation written with an ‘e’ subscript:

$$U(\mathbf{q}) - U(\mathbf{q}_e) = \xi^T \mathbf{g}_e + \frac{1}{2} \xi^T \mathbf{H}_e \xi + \dots$$

The superscript T as in ξ^T indicates the transpose of a matrix; the transpose of a column matrix is a row matrix. The Hessian is often referred to as the force constant matrix.

Finally, if I denote the $3N$ Cartesian coordinates X_1, X_2, \dots, X_{3N} , we usually write the transformation from Cartesian coordinates to internal coordinates as

$$\mathbf{q} = \mathbf{B}\mathbf{X} \quad (4.9)$$

where the rectangular matrix \mathbf{B} is called the Wilson B-matrix. The \mathbf{B} matrix has p rows and $3N$ columns.

4.3 Force Fields

I have been vague so far about which variables are the ‘correct’ ones to take. Chemists visualize molecules in terms of bond lengths, bond angles and dihedral angles, yet this information is also contained in the set of Cartesian coordinates for the constituent atoms. Both are therefore ‘correct’; it is largely a matter of personal choice and professional training. I should mention that there are only $3N - 6$ vibrational coordinates, and so one has to treat the $3N$ Cartesian coordinates with a little care; they contain three translational and three rotational degrees of freedom. More of this technical point will be discussed later.

Spectroscopists are usually interested in finding a set of equilibrium geometric parameters and force constants that give an exact fit with their experimental data. This is harder than it sounds, because for a molecule comprising N atoms and hence $p = 3N - 6$ vibrational degrees of freedom, there are $\frac{1}{2}p(p - 1)$ force constants (diagonal and off-diagonal). In order to measure the individual force constants, the spectroscopist usually has to make experimental measurements on all possible isotopically labelled species. It usually turns out that there are many more unknowns than pieces of experimental information. Spectroscopists usually want a *force field* (comprising force constants, equilibrium quantities and every other included parameter) that is specific for a given molecule. They want to match up ‘theory’ with their incredibly accurate measurements.

Many of the ‘off-diagonal’ force constants turn out to be small, and spectroscopists have developed systematic simplifications to the force fields in order to make as many as possible of the small terms vanish. If the force field contains only ‘chemical’ terms such as bond lengths, bond angles and dihedral angles, it is referred to as a *valence force field* (VFF). There are other types of force field in the literature, intermediate between the VFF and the general force field discussed above.

4.4 Molecular Mechanics (MM)

Molecular modellers usually have a quite different objective; they want a force field that can be transferred from molecule to molecule, in order to predict (for example) the geometry of a new molecule by using data derived from other related molecules. They make use of the bond concept, and appeal to traditional chemists’ ideas that a molecule comprises a sum of bonded atoms; a large molecule consists of the same features we know about in small molecules, but combined together in different ways.

The term *molecular mechanics* was coined in the 1970s to describe the application of classical mechanics to determinations of molecular equilibrium structures. The method was previously known by at least two different names: the *Westheimer* method and the *force-field* method. The name and acronym MM are now firmly established.

The idea of treating molecules as balls joined by springs can be traced back to the work of Andrews (1930). A key study on the development of MM was that by Snyder and Schachtschneider (1965), who showed that transferable force constants could be obtained for alkanes provided that a few off-diagonal terms were retained. These authors found that off-diagonal terms are usually largest when neighbouring atoms are involved, and so we have to take account of nonbonded interactions but only between next nearest neighbours.

A final point for consideration is that we must also take account of the chemical environment of a given atom. An sp carbon atom is different from an sp^2 carbon atom and so on. It is traditional to speak of *atom types* in MM.

Our idea is to treat the force field as a set of constants that have to be fixed by appeal to experiment or more rigorous calculation. In MM we take account of nonbonded interactions, and also the chemical sense of each atom. A valence force field that contains nonbonded interactions is often referred to as a *Urey–Bradley force field*.

4.4.1 Bond Stretching

If we consider phenylalanine (the structure of which is shown in Figure 4.2) we can identify a variety of bond types including $C(sp^2) - C(sp^2)$, $C(sp^2) - C(sp^3)$, $O - H$, $C = O$, and so on. If we assume that Hooke's law is adequate, then each bond stretch between atom types A and B makes a contribution to the total molecular potential energy of

$$U_{AB} = \frac{1}{2}k_{AB} (R_{AB} - R_{e,AB})^2 \quad (4.10)$$

in an obvious notation. Parameter k_{AB} is the force constant, R_{AB} the instantaneous bond length and $R_{e,AB}$ the equilibrium bond length.

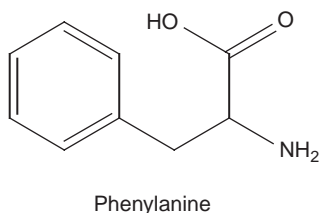


Figure 4.2 Structure of phenylalanine

Other scientists recommend the Morse potential:

$$U_{AB} = D \{1 - \exp[-\alpha(R_{AB} - R_{AB,e})]\}^2$$

whilst some recommend the addition of extra terms to the simple harmonic expression

$$U_{AB} = k_1 (R_{AB} - R_{e,AB})^2 + k_2 (R_{AB} - R_{e,AB})^4 \quad (4.11)$$

4.4.2 Bond Bending

Next we have to consider the bond bending vibrations. It is usual to write these as harmonic ones, typically for the connected atoms A–B–C

$$U_{ABC} = \frac{1}{2}k_{ABC} (\theta_{ABC} - \theta_{e,ABC})^2 \quad (4.12)$$

where k is the force constant, and the subscript 'e' refers to the equilibrium value where the molecule is at rest. A variation on the theme is given by

$$U_{ABC} = \frac{k_{ABC}}{2 \sin^2 \theta_{ABC,e}} (\cos \theta_{ABC} - \cos \theta_{ABC,e})^2 \quad (4.13)$$

4.4.3 Dihedral Motions

Next we must consider the dihedral angle ABCD between the four bonded atoms A, B, C and D (Figure 4.3).

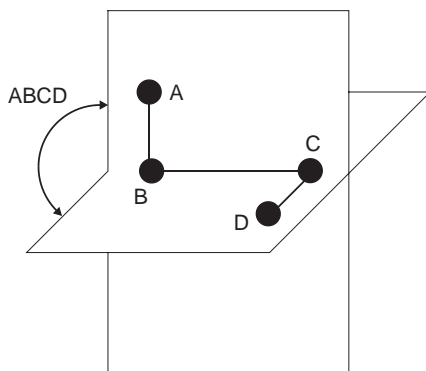


Figure 4.3 Dihedral angle

Some authors divide these into *proper* dihedrals, where we might expect full rotation about the connecting bond B–C, and *improper* dihedrals where the rotation is limited. For example, if C–D were a C–H fragment of a methyl group we would expect full rotation about B–C and a three-fold symmetry in the potential energy term. A –CH–CH– linkage in a benzene ring would only show a moderate flexing from its planar value (angle zero).

If we use χ to denote the ABCD angle, then a popular dihedral potential is given by

$$U = \frac{U_0}{2} (1 - \cos (n (\chi - \chi_e))) \quad (4.14)$$

Here n is the periodicity parameter, which would be 3 for a methyl group. Angle χ_e is the equilibrium torsional angle. A more complicated example is given by

$$U = \frac{V_1}{N_d} (1 + \cos (n_1 \chi - g_1)) + \frac{V_2}{N_d} (1 + \cos (n_2 \chi - g_2)) + \frac{V_3}{N_d} (1 + \cos (n_3 \chi - g_3))$$

The V 's are energy terms, the n 's are periodicity parameters, the g 's are phase parameters and N_d is a constant that depends on the number of bonds.

Some authors treat improper dihedrals in the same way as bond bending, and take a contribution to the molecular potential energy as

$$U_{ABCD} = \frac{1}{2} k_{ABCD} (\chi_{ABCD} - \chi_{e,ABCD})^2 \quad (4.15)$$

where χ is the dihedral angle, as above.

4.4.4 Out-of-Plane Angle Potential (Inversion)

Next we consider the out-of-plane potential terms. Imagine molecule ABCD in Figure 4.4 to be ammonia, a molecule with a very low barrier to inversion; as ψ changes from positive to negative, the molecule inverts. We can write the inversion potential either in terms of the angle indicated (ψ) or in terms of the height (h) of atom D above the plane of atoms A, B and C. Four examples are as follows:

$$\begin{aligned}
 U &= \frac{k_1}{2 \sin^2 \psi_e} (\cos \psi - \cos \psi_e)^2 \\
 U &= k_2 h^2 \\
 U &= k_3 (1 + k_4 \cos(n\psi)) \\
 U &= k_5 (1 + \cos(n\psi - k_6))
 \end{aligned}
 \tag{4.16}$$

where n is a periodicity parameter, the k 's are constants that have to be fixed against experiment and ψ_e is the equilibrium value.

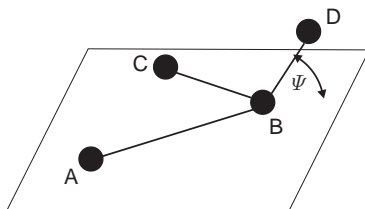


Figure 4.4 Out-of-plane (inversion) potential

4.4.5 Nonbonded Interactions

I mentioned earlier that molecular mechanics force fields have to be transferable from molecule to molecule, and explained the necessity for nonbonded interactions. These are usually taken to be Lennard-Jones 12-6 type, and they are included between all nonbonded pairs of atoms:

$$U_{L-J} = \frac{C_{12}}{R^{12}} - \frac{C_6}{R^6}$$

The Born–Mayer–Huggins potential

$$U_{BMH} = A \exp(-BR) - \frac{C'_6}{R^6} - \frac{C_8}{R^8}$$

is sometimes used when dealing with polar species. B is a parameter determined by the size and ‘softness’ of an ion, C'_6 (not the same as C_6 in U_{L-J}) has to do with dipole–dipole interactions whilst C_8 is determined by dipole–quadrupole interactions.

Some force fields make special provision for hydrogen-bonded atoms; they treat them as nonbonded interactions but soften the Lennard-Jones 12-6 potential for A–H...B to a 12-10 version:

$$U_{\text{HB}} = \frac{C_{12}}{R_{\text{HB}}^{12}} - \frac{C_{10}}{R_{\text{HB}}^{10}} \quad (4.17)$$

Other authors take the view that hydrogen bonds are perfectly respectable chemical bonds that should be treated just like any other bond. They are therefore given a force constant and so on.

4.4.6 Coulomb Interactions

Many force fields take account of electronegativity differences between atoms and add electrostatic terms. Atomic charges Q_A and Q_B are assigned to atoms A and B according to the rules of the particular force field, and we write

$$U_{\text{AB}} = \frac{1}{4\pi\epsilon_0} \frac{Q_A Q_B}{R_{\text{AB}}}$$

4.5 Modelling the Solvent

I should remind you that the electrostatic expression above relates only to point charges in free space. In the presence of a dielectric material (such as water), the force between point charges is reduced by a factor ϵ_r called the relative permittivity. Many force fields were developed at a time when it was not feasible to include a solvent explicitly in such calculations.

Various attempts were made to allow for the effect of a solvent; the most obvious thing to do is to alter the relative permittivity even though no solvent molecules are actually taken into account. There is no agreement between authors as to the correct value of ϵ_r and values ranging between 1 and 80 have been used for water. Some force fields take ϵ_r proportional to the distance between the point charges. I will explain some more up-to-date ways of modelling the solvent in later chapters.

4.6 Time- and Money-Saving Tricks

All the contributions to the molecular potential energy U given above can be done on a pocket calculator. The larger the molecular system, the larger the number of individual contributions to U and the relationship between molecular size and computational effort is roughly dependent on the square of the number of atoms. Over the years, people have tried to reduce the computational time for a given problem by the use of various tricks of the trade. Two such methods are as follows.

4.6.1 United Atoms

Some professional force fields use the so-called *united atom* approach. Here, we regard (for example) a CH_3 group as a pseudo-atom, X, and develop parameters for a $\text{C}(\text{sp}^2)\text{-X}$ stretch, and so on. It is customary to treat methyl, methylene and methane groups as united atoms, especially when dealing with large biological systems.

4.6.2 Cut-Offs

For a large molecule, there are many more nonbonded interactions than bonded interactions. Molecular mechanics force fields very often cut these off to zero at some finite distance, in order to save computer time. This can sometimes lead to mathematical difficulties because of the discontinuity, and various ingenious methods have been proposed to circumvent the problem (other than actually retaining the terms). I will show you in a later chapter that there are other problems associated with this cut-off procedure; it is a real problem, not just one that I have mentioned out of historical interest.

Figure 4.5 shows a Lennard-Jones 12-6 potential with a cut-off (after a certain value of R , the potential is set to zero).

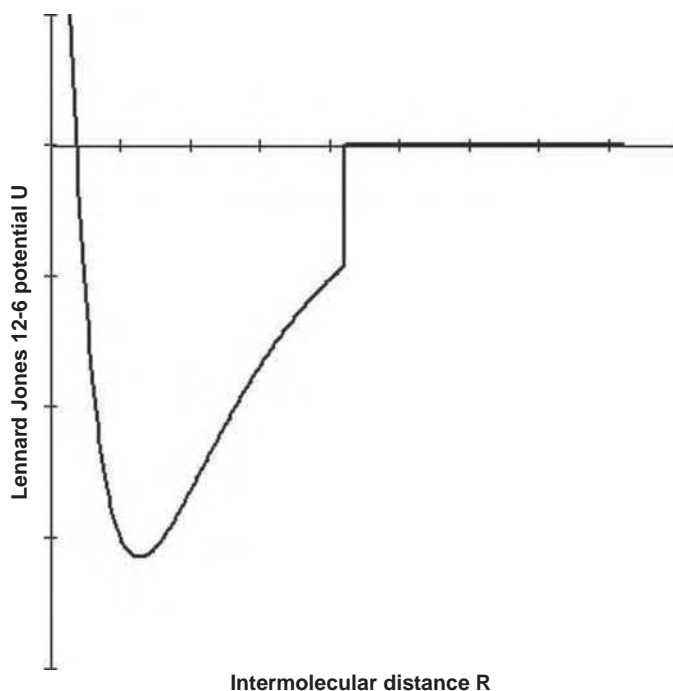


Figure 4.5 *Schematic cut-off of Lennard-Jones potential*

4.7 Modern Force Fields

A 'standard' modern MM force field can be written

$$U = \sum_{\text{stretch}} U_{AB} + \sum_{\text{bend}} U_{ABC} + \sum_{\text{dihedral}} U_{ABCD} + \sum_{\text{out-of-plane}} U_{ABCD} + \sum_{\text{nonbonded}} U_{AB} + \sum_{\text{Coulomb}} U_{AB} \quad (4.18)$$

or written explicitly in terms of the contributions discussed above

$$\begin{aligned}
 U = & \sum_{\text{bonds}} \frac{1}{2} k_{\text{AB}} (R_{\text{AB}} - R_{\text{e,AB}})^2 + \sum_{\text{bends}} \frac{1}{2} k_{\text{ABC}} (\theta_{\text{ABC}} - \theta_{\text{e,ABC}})^2 \\
 & + \sum_{\text{dihedrals}} \frac{U_0}{2} (1 - \cos(n(\chi - \chi_0))) + \sum_{\text{out-of-plane}} \frac{k}{2 \sin^2 \psi_e} (\cos \psi - \cos \psi_e)^2 \quad (4.19) \\
 & + \sum_{\text{nonbonded}} \left(\frac{C_{\text{AB}}^{12}}{R_{\text{AB}}^{12}} - \frac{C_{\text{AB}}^6}{R_{\text{AB}}^6} \right) + \frac{1}{4\pi \epsilon_0} \sum_{\text{charges}} \frac{Q_A Q_B}{R_{\text{AB}}}
 \end{aligned}$$

4.7.1 Variations on a Theme

There are a number of variants of this expression in the literature. Some force fields contain mixed terms like

$$\frac{k}{2} (R - R_e) (\theta - \theta_e)$$

which couple together the bond stretching modes with angle bending. Others use more complicated expressions for the individual bending and stretching terms. Some force fields allow interactions between lone pairs, which are often referred to as nonatomic interaction centres. In addition, there are specialist force fields that are appropriate for restricted ranges of compounds such as ions, liquid metals and salts.

Force fields are determined by one of two routes. First of all, in an ideal world, one might calibrate their parameters against accurate quantum mechanical calculations on clusters of small molecules. The alternative is to calibrate against experimental data such as crystal structure, infrared absorption, X-ray measurements and liquid properties such as density, enthalpy of vaporization, Gibbs energies of solvation and the like. To date, almost all modern force fields have been obtained by the latter approach.

The choice of a particular force field for a given application should depend on the type of system for which the force field was designed. For example, some force fields have been calibrated against the solution properties of amino acids. These are obviously the ones to choose when it comes to modelling proteins in solution.

Finally, I must emphasize the importance of the *atom type* (i.e. the chemical environment). The chemical environment of an atom can be distinguished by

- its hybridization
- its formal atomic charge
- its nearest neighbours.

For example, one well-known force field distinguishes five types of oxygen atom:

1. A carbonyl oxygen
2. A hydroxyl oxygen
3. A carboxylic or phosphate oxygen
4. An ester or ether oxygen
5. An oxygen in water.

The interactions are calculated according to atom type, not the 'raw' elements.

4.8 Some Commercial Force Fields

With these principles in mind, it is time to examine some of the common force fields found in professional molecular modelling programs.

4.8.1 Dreiding

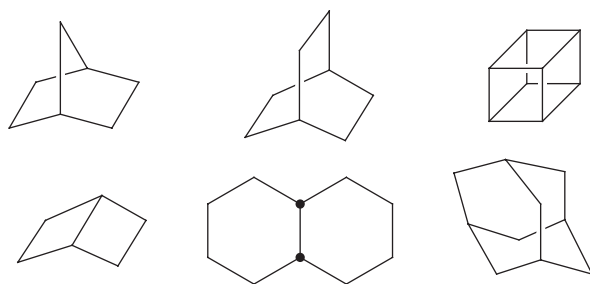
This force field (Mayo *et al.* 1990) is parameterized for all atom types that any chemist would expect for the elements H, C, N, O, P, S, F, Cl, Br and I. In terms of the ‘standard’ expression we write

$$\begin{aligned}
 U = & \sum_{\text{bonds}} \frac{1}{2} k_{AB} (R_{AB} - R_{e,AB})^2 + \sum_{\text{bends}} \frac{1}{2} k_{ABC} (\cos \theta_{ABC} - \cos \theta_{e,ABC})^2 \\
 & + \sum_{\text{dihedrals}} \frac{U_0}{2} (1 - \cos(n(\chi - \chi_0))) + \sum_{\text{out-of-plane}} \frac{k}{2} (\psi - \psi_e)^2 \quad (4.20) \\
 & + \sum_{\text{nonbonded}} \left(\frac{C_{AB}^{12}}{R_{AB}^{12}} - \frac{C_{AB}^6}{R_{AB}^6} \right)
 \end{aligned}$$

4.8.2 MM1

In his review, Allinger (1976) essentially defined what we now call the MM1 force field. He treated hydrocarbons only, ignored the Coulomb terms and used an exp-6 Lennard-Jones potential.

Once all the standard cases have been successfully treated, one naturally looks at the difficult ones. In the case of hydrocarbons, these difficult cases comprise strained rings such as cyclobutane. The problem with cyclobutane is that, whilst having all carbon atoms planar can minimize the angular deformation from tetrahedral, the molecule is actually puckered by a substantial angle from planarity. In addition, the C–C bond lengths are unusually large. The obvious solution is to say that a four-membered ring is different from any other hydrocarbon and that the bond angle does not have a natural tetrahedral value,



Bicyclic and related molecules

Figure 4.6 *Bicyclic and related molecules*

but one then goes down the undesirable path where all difficult cases have their own set of parameters.

Allinger and others introduced a variety of ‘mixed’ terms into the standard MM potential; for example, a bond length–bond angle term and a torsion–bend interaction.

Figure 4.6 shows typical bicyclic and related hydrocarbons described in Allinger (1976).

4.8.3 MM2 (Improved Hydrocarbon Force Field)

Allinger (1977) subsequently introduced MM2. At the time there was a deal of discussion in the literature about how different force fields should represent hydrogen atoms: as ‘hard’ or ‘soft’ atoms. A hard atom was said to be one whose plot of force versus distance (in the diatom) showed a steep slope. A soft atom was one where the slope was gentler. It all boiled down to the repulsive part of the nonbonded interactions but eventually Allinger decided to retain his exp-6 intermolecular potential.

MM2 differs from MM1 in three main respects.

1. The dihedral term was extended to

$$U = \frac{U_1}{2} (1 + \cos \omega) + \frac{U_2}{2} (1 + \cos 2\omega) + \frac{U_3}{2} (1 + \cos 3\omega) \quad (4.21)$$

where each of the U 's was found by calibration against experiment.

2. The bending term was extended to

$$U_{ABC} = \frac{1}{2} k_{ABC} (\theta_{ABC} - \theta_{e,ABC})^2 + \frac{1}{2} k'_{ABC} (\theta_{ABC} - \theta_{e,ABC})^6 \quad (4.22)$$

3. All mention of cross terms between bond stretches and bends were finally dropped.

A great deal of importance was attached to the calculation of enthalpies of formation, and 42 hydrocarbons were treated. The author claimed that his enthalpy results were comparable to experiment in terms of experimental error.

4.8.4 AMBER

AMBER (an acronym for assisted model building and energy refinement; Weiner *et al.* 1984) is a force field for the simulation of nucleic acids and proteins. It was calibrated against experimental bond lengths and angles obtained from microwave, neutron diffraction and accurate quantum chemical studies. The parameters were then refined with MM studies on model compounds such as tetrahydrofuran, deoxyadenosine, dimethyl phosphate, 9-methyladenine-1-methylthymine hydrogen bonded complex and others.

The model differs from our standard expression (Equation (4.19)) in four ways.

1. Hydrogen bonds were included explicitly with a 12-10 potential

$$U_{\text{H-bonds}} = \sum_{\text{H-bonds}} \left(\frac{C_{12}}{R^{12}} - \frac{C_{10}}{R^{10}} \right)$$

2. An attempt was made to include solvent effects by inclusion of the Coulomb term with a distance-dependent relative permittivity.
3. The AMBER force field is a ‘united atom’ one, and hydrogen atoms bonded to carbons are not explicitly included. They are absorbed into the atom type parameters for neighbouring atoms.
4. Lone pairs were explicitly included for sulfur hydrogen bonding.

There are a number of different versions of AMBER; the original united atom version was later extended to include all atoms. Just to give you a flavour, one modern software package has the following choices.

1. Amber 2
2. Amber 3
3. Amber for saccharides
4. Amber 94
5. Amber 96.

4.8.5 OPLS (Optimized Potentials for Liquid Simulations)

Like AMBER, OPLS is designed for calculations on amino acids and proteins. The easiest thing is for me to quote part of the abstract to the keynote paper (Jorgensen and Tirado-Rives 1988).

A complete set of inter molecular potential functions has been developed for use in computer simulations of proteins in their native environment. Parameters have been reported for 25 peptide residues as well as the common neutral and charged terminal groups. The potential functions have the simple Coulomb plus Lennard-Jones form and are compatible with the widely used models for water, TIP4P, TIP3P and SPC. The parameters were obtained and tested primarily in conjunction with Monte Carlo statistical mechanics simulations of 36 pure organic liquids and numerous aqueous solutions of organic ions representative of subunits in the side chains and backbones of proteins . . .

Improvement is apparent over the AMBER united-atom force field which has previously been demonstrated to be superior to many alternatives.

I will explain about TIP and Monte Carlo in later chapters. Each atomic nucleus is an interaction site, except that CH_n groups are treated as united atoms centred on the carbon. Hydrogen bonds are not given any special treatment, and no special account is taken of lone pairs.

4.8.6 Johnson

I mentioned earlier the existence of a number of specialist force fields. To take an even more extreme example, the Johnson force field (Johnson 1964) is specific to solid-state models involving the pure elements Fe, W and V. The pair potential terms are written

$$\begin{aligned}
 U &= a_1 (R - b_1)^3 + c_1 R + d_1 \text{ if } \varepsilon_1 < R < \varepsilon_2 \\
 &= a_2 (R - b_2)^3 + c_2 R + d_2 \text{ if } \varepsilon_2 < R < \varepsilon_3 \\
 &= a_3 (R - b_3)^3 + c_3 R + d_3 \text{ if } \varepsilon_3 < R < \varepsilon_4 \\
 &= 0 \text{ if } \varepsilon_4 < R
 \end{aligned}
 \tag{4.23}$$

where R is the distance between a pair of atoms, ε are characteristic distances and a , b , c and d are parameters.

4.8.7 Unified Force Field (UFF)

I have discussed so far a number of force fields of increasing degrees of speciality and application. Let me finish this chapter by drawing your attention to the UFF; in the words of the authors Rappé *et al.* (1992), 'a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations'. Molecular dynamics is the subject of a later chapter in this book. I commend this article to you as a model of logic and lucidity and it illustrates beautifully all the principles I have discussed so far. Here is the abstract (Rappé *et al.* 1992):

A new molecular mechanics force field, the universal force field (UFF), is described wherein the force field parameters are estimated using general rules based only on the element, its hybridization and its connectivity. The force field functional form, parameters, and generating formulas for the full periodic table are presented.

In particular, the parameters used to generate the UFF include a set of hybridization-dependent atomic bond radii, a set of hybridization angles, van der Waals parameters, torsional and inversion barriers and a set of effective nuclear charges.

Thus, the force field is represented by Equation (4.20); there is a slight clash of notation, for my 'out-of-plane' read 'inversion'. If I focus on the bond stretch, UFF allows both the simple harmonic and Morse pair potentials. Both involve an 'equilibrium' value for the bond length, and this is one of the 'parameters' that has to be fixed. The authors took this to be the sum of atom type specific single bond data, plus a bond order correction, plus an electronegativity correction.

The authors describe many simple and more difficult organics, together with a good number of metal-containing molecules, and conclude that their predicted bond lengths are good to 10 pm, and bond angles to about 7°.

References

- Allinger, N.L. (1976) *Advances in Physical Organic Chemistry 13* (eds V. Gold and D. Bethell), Academic Press, London.
- Allinger, N.L. (1977) *J. Am. Chem. Soc.*, **99**, 8127.
- Andrews, D.H. (1930) *Phys. Rev.*, **36**, 544.
- Johnson, R.A. (1964) *Phys. Rev.*, **134**, A1329.
- Jorgensen, W.L. and Tirado-Rives, J. (1988) *J. Am. Chem. Soc.*, **110**, 1657.
- Mayo, S.L., Olafson, B.D. and Goddard III, W.A. (1990) *J. Chem. Phys.*, **94**, 8897.
- Rappé, A.K., Casewit, C.J., Colwell, K.S. *et al.* (1992) *J. Am. Chem. Soc.*, **114**, 10024.
- Snyder, R.G. and Schachtschneider, J.H. (1965) *Spectrochim. Acta*, **21**, 169.
- Weiner, S.J., Kollman, P.A., Case, D.A. and Singh, V.C. (1984) *J. Am. Chem. Soc.*, **106**, 765.

5

The Molecular Potential Energy Surface

For one-dimensional problems, we speak about a molecular potential energy *curve*. The simple potential energy curves we have met so far have all shown a single minimum. From now on, life gets more complicated (or interesting, depending on your viewpoint).

5.1 Multiple Minima

The plot in Figure 5.1 shows how the ethane molecular potential varies with dihedral angle. The figure shows a full rotation of 360° ; all the remaining geometrical variables were kept constant. Note that there are three identical minima (and of course three identical maxima), and the differences between maxima and minima are all the same. The chemical interpretation is that these three minima correspond to conformers where the hydrogens are as far apart as possible (i.e. in the *trans* position). The maxima correspond to conformers where the C–H bonds eclipse each other. Multiple minima are common in potential energy surface studies, as we will see.

Consider now the substituted ethane $\text{CH}_2\text{Cl}-\text{CH}_2\text{Cl}$. A plot of the potential energy versus the ClC–CCl dihedral angle is shown in Figure 5.2. There are three minima, one lower than the other two. The three minima are referred to as *local minima* and the minimum at 180° is called the *global minimum*. The global minimum corresponds to a conformation with the two chlorines as far apart as possible. The two other minima correspond to conformers where each chlorine is *trans* to hydrogen.

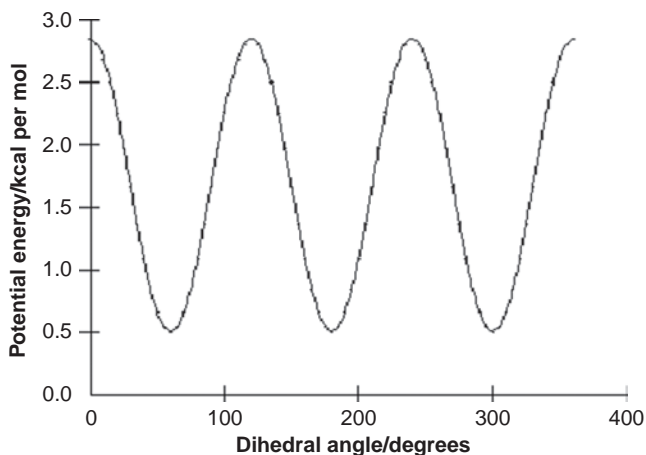


Figure 5.1 Ethane dihedral motion

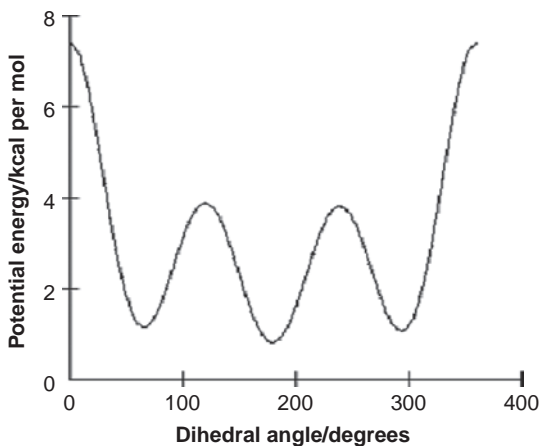


Figure 5.2 Plot of potential energy versus dihedral angle for $\text{CH}_2\text{Cl}-\text{CH}_2\text{Cl}$

5.2 Saddle Points

One-dimensional potential energy curves show minima, maxima and points of inflexion. Two- and higher-dimensional surfaces often show a new feature called a *saddle point*. Saddle points are a maximum in some variables and a minimum in the remainder. The one type of saddle point that interests chemists is where we have a minimum in all variables but one, and a maximum in this remaining one. This corresponds to a transition state in theories of reaction mechanisms. I just mention it here, despite the fact that molecular mechanics cannot properly treat bond breaking or bond making. I am not going to deal in detail with saddle points until I have given you a firm grounding in quantum chemistry, but you should be aware of their existence.

There are also points where the molecular potential energy is a maximum. We are rarely interested in these.

Figure 5.3 shows a saddle point (for the function $f(x, y) = x^2 - y^2$). The vertical axis shows the function values, and the two horizontal axes are x and y . The apocryphal reason why such points are called saddle points is that the figure above resembles a horse's saddle. Molecular potential energy surfaces generally contain all possible features and you might like to consider the example shown in Figure 5.4, which shows two (equal) minima and a saddle point in between.

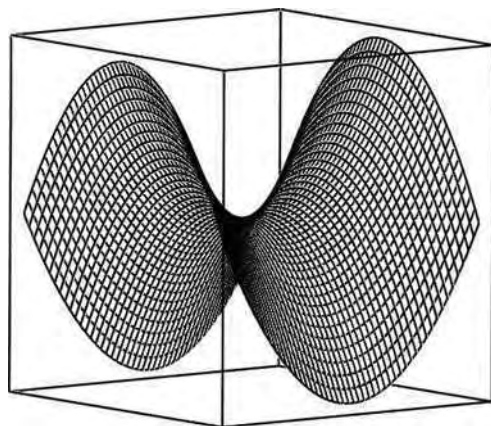


Figure 5.3 Saddle point

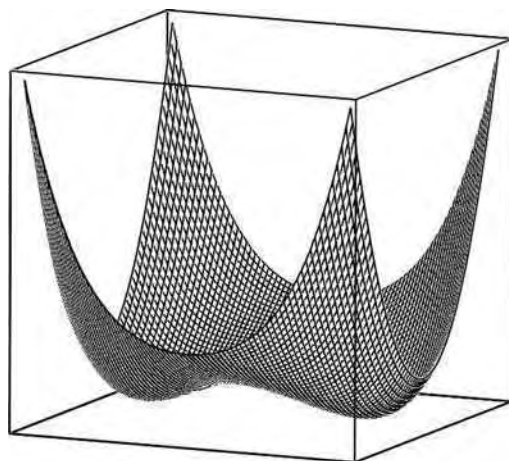


Figure 5.4 Two minima and a saddle point

In this chapter I want to show you how we locate stationary points on the molecular potential energy surface, and how we characterize them. I will start with the second problem, because it is straightforward.

5.3 Characterization

For a linear molecule with N atoms, there are in total $3N$ degrees of freedom. Three of these are translational degrees of freedom, three are rotational (two for a linear molecule) and the remaining $p = 3N - 6$ ($3N - 5$ if linear) are vibrational degrees of freedom. We therefore need p independent variables to describe them, and these are known as the *internal coordinates*. It is traditional to denote such variables as q ; if I write them q_1, q_2, \dots, q_p then the molecular potential energy will depend on these q 's and the normal coordinates will be linear combinations of them. For the minute, do not worry about the nature of these variables. In Chapter 4, I showed how to collect such variables into the column vector \mathbf{q} (a matrix of dimension $p \times 1$):

$$\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_p \end{pmatrix}$$

and I also defined the gradient of U , $\mathbf{g} = \text{grad } U$, as a column vector, and its Hessian \mathbf{H} as a symmetric square matrix:

$$\mathbf{g} = \begin{pmatrix} \frac{\partial U}{\partial q_1} \\ \frac{\partial U}{\partial q_2} \\ \vdots \\ \frac{\partial U}{\partial q_p} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \frac{\partial^2 U}{\partial q_1^2} & \cdots & \frac{\partial^2 U}{\partial q_1 \partial q_p} \\ \vdots & \cdots & \vdots \\ \frac{\partial^2 U}{\partial q_p \partial q_1} & \cdots & \frac{\partial^2 U}{\partial q_p^2} \end{pmatrix}$$

The molecular potential energy depends on the q 's, and I will write it as $U(\mathbf{q})$. At a stationary point, the gradient is zero: it is as easy as that. In order to characterize the stationary point, we have to find the eigenvalues of the Hessian calculated at that point. If the eigenvalues are all positive, the point is a minimum. If the eigenvalues are all negative, the point is a maximum. Otherwise the point is a saddle point. Saddle points of interest to chemists are those that are a minimum in all degrees of freedom except one, and it transpires that the Hessian has just one negative eigenvalues at such a point. They are sometimes called *first-order saddle points*.

5.4 Finding Minima

MM tends to be concerned with searching the molecular potential energy surfaces of very large molecules for minima (equilibrium structures). Many mathematicians, scientists and

engineers are concerned with related problems. The field of study is now an established part of applied mathematics, and is referred to as *optimization theory*. An optimization problem involves minimizing a function of several variables, possibly subject to restrictions on the values of the variables. There is no single universal algorithm for optimization, and it is necessary for us to spend a little time dealing with the different methods of interest.

I should tell you that one speaks of *optimization* and less commonly of minimization or maximization; the latter two are equivalent, for maximization of function f is the same as minimization of the function $-f$.

The first MM optimization seems to have been carried out by Westheimer (1956), who did the calculations ‘by hand’. The first computer calculations seem to have been done by Hendrickson (1961). Neither of their methods was generally applicable to any molecule.

Many algorithms have been developed over a number of years for the location of stationary points, some of which are suitable for MM calculations, some of which are suitable for quantum mechanical calculations and many of which are not particularly suited to either. In MM we tend to deal with large molecules and consequently the molecular potential energy function will depend on hundreds if not thousands of variables. On the other hand, evaluation of the potential energy at each point on the hypersurface is simple.

Different considerations apply to quantum chemical calculations, to be dealt with later in the text. Calculation of the energy at points on the surface is far from easy but the number of variables tends to be smaller.

As stated above, transition states are special cases of saddle points; they are stationary points on the surface where the Hessian has just one negative eigenvalue. Transition state searching is a hot topic in chemistry, and a number of specialist algorithms have been proposed. I will deal with transition states in Chapter 22.

5.5 Multivariate Grid Search

This is the oldest method for finding minima, and it has a long history. What we do is the following.

1. Choose a suitable grid for the variables.
2. Choose a starting point A on the grid.
3. For each variable q_1, q_2, \dots, q_p evaluate the molecular potential energy U at the two points surrounding A (as determined by the grid size).
4. Select the new point for which U is a minimum, and repeat steps 3 and 4 until the local minimum is identified.

The method is intuitive, and it is apparent that a local minimum will eventually be found.

5.5.1 Univariate Search

This is sometimes called the cyclic search, for we perform successive one-dimensional searches for each of the variables in turn.

1. Choose a starting point A.
2. Minimize $U(\mathbf{q})$ for each variable q_1, q_2, \dots, q_p in turn.
3. Repeat the cycle as necessary.

The second step is the crucial one, and a favourite strategy is quadratic interpolation. Suppose that x_1, x_2 and x_3 are three successive values of one of the variables q_i (say), chosen such that they bracket a minimum in that particular variable, as shown in Figure 5.5. We fit the three points to a quadratic, and a little algebra shows that the minimum α is given by

$$\alpha = \frac{1}{2} \frac{U(x_1)(x_3^2 - x_2^2) + U(x_2)(x_1^2 - x_3^2) + U(x_3)(x_2^2 - x_1^2)}{U(x_1)(x_3 - x_2) + U(x_2)(x_1 - x_3) + U(x_3)(x_2 - x_1)} \quad (5.1)$$

The method is also known as a *line search*.

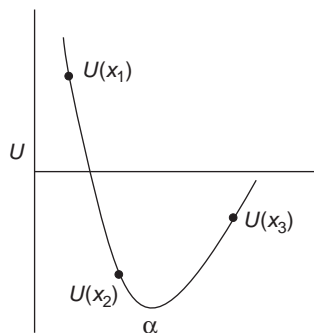


Figure 5.5 Univariate minimization

5.6 Derivative Methods

It is conventional to divide optimization algorithms into those methods that make use of derivatives (gradients and Hessians) and those that do not. Neither of the two methods discussed so far make use of derivatives. Some authors use a further subdivision into *first-order* derivative methods (where we make use of the gradient) and *second-order* derivative methods (where we use both the gradient and the Hessian).

There are many algorithms in the literature; I can only give you a flavour.

5.7 First-Order Methods

5.7.1 Steepest Descent

This first-order derivative scheme for locating minima on molecular potential energy surfaces was put forward by Wiberg (1965). His basic algorithm can be summarized as follows.

1. Calculate U for the initial structure.
2. Calculate U for structures where each atom is moved along the x , y and z axes by a small increment. Movement of some atoms will lead to a small change in U , whilst movements of other atoms will lead to a large change in U . (The important quantity is clearly the gradient.)

3. Move the atoms to new positions such that the energy U decreases by the maximum possible amount.
4. Repeat the relevant steps above until a local minimum is found.

Wiberg's method would today be classified as 'steepest descent'; the idea is that we start from point A on a molecular potential energy surface and identify the fastest way down the surface to the local minimum. Figure 5.6 shows a two-dimensional potential energy surface. The negative of $\text{grad } U$ at point A gives the maximum rate of decrease of U , and also the direction in which to proceed (the gradient vector is perpendicular to the contour at A). The only problem is to choose a step length; this can be done by a line search along the vector indicated, and it is clear that U will pass through a linear minimum around point B.

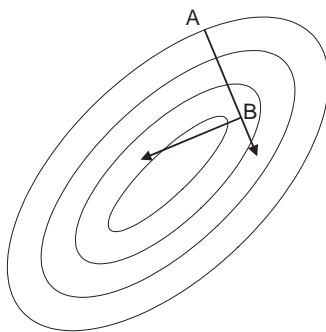


Figure 5.6 Steepest descents

We repeat the process from point B until the local minimum is (eventually) reached. It can be proved that every step is necessarily at right angles to the one before it, and this leads to one of the infuriating characteristics of the steepest descent method: it takes very many tiny steps when proceeding down a long narrow valley, as it is forced to make a right-angled turn at every point, even though that might not be the best route to the minimum.

At this point, I should introduce some notation that is universally understood by workers in the field of optimization. Like it or not, you will come across it when you read the technical literature.

All of the algorithms that I will describe are *iterative*; that is, we start from some initial point A on a surface, and then move on in cycles, according to an algorithm, hopefully toward a stationary point. Each cycle of the calculation is called an iteration, and people often use the symbol k to keep track of the iteration count. I am going to follow common practice and add a superscript ' k ' to variables in order to indicate this iteration count. Sometimes I will just write it as ' k ', when there is no possibility that you will regard the iteration count as meaning 'to the power of'.

5.7.2 Conjugate Gradients

Starting from point $\mathbf{q}^{(k)}$ (where k is the iteration count), we move in a direction given by the vector

$$\mathbf{V}^{(k)} = -\mathbf{g}^{(k)} + \gamma^{(k)}\mathbf{V}^{(k-1)} \quad (5.2)$$

where $\mathbf{g}^{(k)}$ is the gradient vector at point $\mathbf{q}^{(k)}$ and $\gamma^{(k)}$ is a scalar given by

$$\gamma^{(k)} = \frac{(\mathbf{g}^{(k)})^T \mathbf{g}^{(k)}}{(\mathbf{g}^{(k-1)})^T \mathbf{g}^{(k-1)}} \quad (5.3)$$

I have used the superscript T to denote the transpose of a matrix. The line search method is then used to determine the distance to be moved along this vector. This algorithm was first proposed by Fletcher and Reeves (1964).

Conjugate gradients methods produce a set of directions that overcome the oscillatory behaviour of steepest descents in narrow valleys. Successive directions are not at right angles to each other. Such methods are also referred to as *conjugate direction* methods.

Polak and Ribière (1969) proposed an alternative form for the scalar $\gamma^{(k)}$:

$$\gamma^{(k)} = \frac{(\mathbf{g}^{(k)} - \mathbf{g}^{(k-1)})^T \mathbf{g}^{(k)}}{(\mathbf{g}^{(k-1)})^T \mathbf{g}^{(k-1)}} \quad (5.4)$$

For a purely quadratic function, their method is identical to the original Fletcher–Reeves algorithm. The authors pointed out that most U 's encountered in chemistry are only approximately quadratic, and that their method was therefore superior.

5.8 Second-Order Methods

5.8.1 Newton–Raphson

Second-order methods use not only the gradient but also the Hessian to locate a minimum. Before launching into any detail about second-order methods, it is worth spending a little time discussing the well-known Newton–Raphson method for finding the roots of an equation of a single variable. This is a simple illustration of a second-order method. For illustration, consider the function

$$f(x) = x \exp(-x^2)$$

shown as the full curve in Figure 5.7. It is obvious by inspection that our chosen function has roots at $x = 0$ and at $+\infty$ and $-\infty$, but suppose for the sake of argument that we do not know that the function had a root at $x = 0$. We might make a guess that the function has a root at (say) $x = 0.4$. This is our first guess, so I set the iteration count = 1 and write the guess $x^{(1)}$ (also, the function has a value of $f^{(1)} = 0.3409$). Thus we have

$$x^{(1)} = 0.4, \quad f^{(1)} = 0.3409$$

The function can be differentiated to give

$$\frac{df}{dx} = (1 - 2x^2) \exp(-x^2)$$

and the gradient $g^{(1)}$ at $x^{(1)} = 0.4$ has a value 0.5795. The tangent line at $x^{(1)} = 0.4$ is therefore

$$y = 0.5795x + 0.1091$$

which I have shown as a full straight line on Figure 5.7. A little algebra shows that this line crosses the x axis at

$$x^{(2)} = x^{(1)} - \frac{f^{(1)}}{g^{(1)}}$$

This crosses the x axis at -0.1883 and this value is taken as the next best estimate $x^{(2)}$ of the root. I then draw the tangent line at $x^{(2)}$, which is shown dotted in Figure 5.7, and so on. Table 5.1 summarizes the iterations.

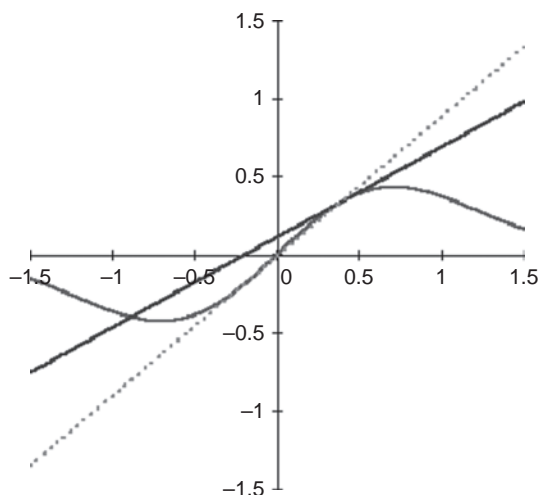


Figure 5.7 Newton–Raphson

Table 5.1 Newton–Raphson iterations

k	$x^{(k)}$	$f^{(k)}$	$g^{(k)}$
1	0.4	0.3409	0.5795
2	-0.1882	-0.1817	0.8968
3	1.4357×10^{-2}	1.4354×10^{-2}	0.9994
4	-5.9204×10^{-6}	-5.9204×10^{-6}	1.0000
5	0.0000	0.0000	1.0000

There are a few points worth noticing. First, the convergence is rapid. In fact, for a quadratic function the Newton–Raphson method will locate the nearest root in just one step.

Second, the choice of starting point is crucial. If I start with $x = \pm 0.5$, the successive estimates simply oscillate between $+0.5$ and -0.5 . If I start with $|x| < 0.5$, then the method converges to the root $x = 0$. If I start with $|x| > 0.5$ then we find the infinite roots.

The method can be easily modified to search for stationary points, which are characterized by a zero gradient. The iteration formula

$$x^{(k+1)} = x^{(k)} - f^{(k)}/g^{(k)} \quad (5.5)$$

becomes

$$x^{(k+1)} = x^{(k)} - g^{(k)}/H^{(k)} \quad (5.6)$$

where $H^{(k)}$ is the value of the second derivative at the current point $x^{(k)}$.

For this particular function, we have

$$H(x) = (4x^3 - 6x) \exp(-x^2)$$

and I can now repeat the argument given above, but this time searching for stationary points rather than roots of the equation.

Figure 5.8 shows the gradient (the full curve), and the first two iterations of the Newton–Raphson algorithm for the stationary point, starting at $x^{(1)} = 0.2$. The gradient is zero at infinity and at $x = \pm 0.7071$.

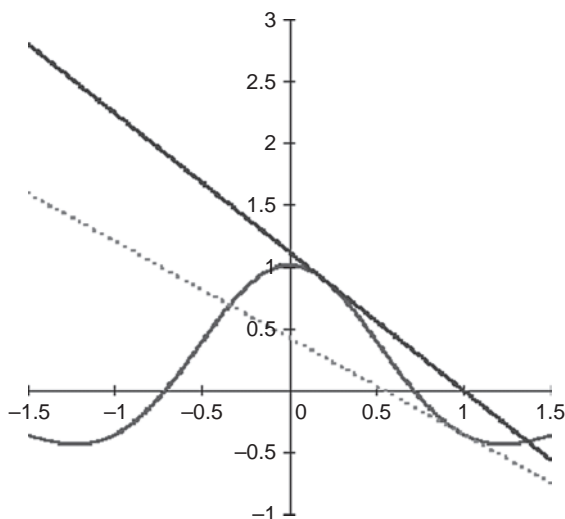


Figure 5.8 Location of stationary point

In the case of a function of many variables x_1, x_2, \dots, x_n , the Newton–Raphson optimization algorithm can be written

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)} \quad (5.7)$$

where I have collected the x 's into a column vector \mathbf{X} and so on. The formula can be easily derived from the Taylor expansion.

We therefore have to calculate the gradient and the Hessian at each successive point in the iterations. Normally, the algorithm needs an infinite number of steps but it will find the minimum of any quadratic function in just one step, starting at an arbitrary point on the

surface. For example, if we consider the function $f(x, y) = x^2 + y^2$ which has a minimum at $x = 0, y = 0$ we have

$$\mathbf{g} = \begin{pmatrix} 2x \\ 2y \end{pmatrix} \mathbf{H} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

$$\mathbf{H}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$$

If we start from the point (2, 2) we calculate

$$\mathbf{X}^{(2)} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 4 \\ 4 \end{pmatrix}$$

which is a null vector.

In practice molecular potential energy surfaces are rarely quadratic and so an infinite number of steps will be required. We simply stop the calculation once a given number of decimal places have been achieved.

5.8.2 Block Diagonal Newton–Raphson

There are a number of variations on the Newton–Raphson method, many of which aim to eliminate the need to calculate the full Hessian. A widely used algorithm is the *block diagonal Newton–Raphson* method, where just one atom is moved at each iteration. This means that all the elements of the Hessian are zero except for a block along the diagonal describing the atom in question.

5.8.3 Quasi-Newton–Raphson

In addition, there are a number of so-called *quasi-Newton–Raphson* methods that gradually build up the inverse Hessian in successive iterations. At each iteration, the vector $\mathbf{X}^{(k)}$ is updated to $\mathbf{X}^{(k+1)}$

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)}$$

using the gradient and the current estimate of the inverse Hessian. Having made the move to the new position, \mathbf{H}^{-1} is updated from its value at the previous step by an approximate procedure depending on the algorithm employed. The methods of Davidon–Fletcher–Powell, Broyden–Fletcher–Goldfarb–Shanno and Murtaugh–Sargent are often encountered, but there are many others.

5.8.4 Fletcher–Powell Algorithm

The Fletcher–Powell algorithm (Fletcher and Powell 1963) is a derivative method where elements of the gradient and the Hessian are estimated numerically each cycle. Essentially, it assumes that the surface is quadratic around each chosen point, and finds the minimum for the approximate quadratic surface. Steps in the algorithm are as follows.

1. Calculate the energy $U^{(1)}$ for an initial geometry $\mathbf{X}^{(1)}$, and at positive and negative displacements for each of the coordinates.
2. Fit a quadratic for each of the coordinates according to the formula

$$U(\mathbf{X}) = U^k + \sum_{i=1}^p \left(g_i^k (X_i - X_i^k) + \frac{1}{2} H_{ii}^k (X_i - X_i^k)^2 \right) \quad (5.8)$$

(this essentially gives numerical estimates of the gradient and the Hessian; I have dropped the brackets round the iteration count for clarity).

3. Find a minimum of this expression; we have

$$\begin{aligned} \frac{\partial U}{\partial X_i} &= 0 \\ g_i^k + H_{ii}^k (X_i - X_i^k) &= 0 \\ c_i^k = X_i - X_i^k &= -\frac{g_i^k}{H_{ii}^k} \end{aligned} \quad (5.9)$$

The last term gives the correction to coordinate x_i ; if these are small enough then stop.

4. Calculate the energy at points \mathbf{X}^k , $\mathbf{X}^k + \mathbf{c}^k$ and $\mathbf{X}^k + 2\mathbf{c}^k$.
5. Fit a quadratic to these three points as above.
6. Find the energy minimum, as above. This gives point \mathbf{X}^{k+1} on the surface.
7. Calculate the gradient \mathbf{g}^{k+1} at this point, increase the iteration count and go back to step 3.

5.9 Choice of Method

The choice of algorithm is dictated by a number of factors, including the storage and computing requirements, the relative speeds at which the various parts of the calculation can be performed and the availability of an analytical gradient and Hessian. Analytic first and second derivatives are easily evaluated for MM force fields; the only problem might be the physical size of the Hessian. For this reason, MM calculations on large systems are often performed using steepest descent and conjugate gradients. The Newton–Raphson method is popular for smaller systems, although the method can have problems with structures that are far from a minimum. For this reason, it is usual to perform a few iterations using (for example) steepest descent before switching to Newton–Raphson. The terms ‘large’ and ‘small’ when applied to the size of a molecule are of course completely relative and are dependent on the computer power available when you do the calculation.

In cases where the differentiation cannot be done analytically, it is always possible to *estimate* a gradient numerically; for example in the case of a function of one variable

$$\left(\frac{df}{dx} \right)_{x=x_1} \approx \frac{f(x_1 + \Delta) - f(x_1)}{\Delta}$$

where Δ is small. Algorithms that rely on numerical estimates of the derivatives need more function evaluations than would otherwise be the case, so there is a delicate trade-off in computer time. It is generally supposed that gradient methods are superior to nongradient methods, and it is also generally thought to be advantageous to have an analytical expression for the gradient and Hessian.

5.10 The Z-Matrix

For a nonlinear molecule of N atoms, there are $p = 3N - 6$ vibrational degrees of freedom that should be described by p independent coordinates q_1, q_2, \dots, q_p and at first sight all should be varied in a geometry optimization. These coordinates are often defined to modelling packages using the so-called Z-matrix, which gives a way of building up a molecular geometry in terms of valence descriptors such as bond lengths, bond angles and dihedral angles. Two examples will help to make things clear. First of all water, for which I will take a bond length of 95.6 pm (0.956 Å) and an H–O–H bond angle of 104° (Figure 5.9).



Z-Matrix

```
Oxy
H1 Oxy ROH
H2 Oxy ROH H1 Ang

ROH 0.956
Ang 104.
```

Figure 5.9 Z-matrix for water

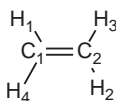
I can start my Z-matrix at any atom, and that atom occupies a line by itself. I choose oxygen, which I have called Oxy. H1 is joined to oxygen with a bond length of ROH (0.956 Å). I have assumed that the bond lengths are equal. H2 is also joined to oxygen with a bond length of 0.956 Å, and the H2–oxygen–H1 bond angle is Ang (104°).

Notice that I have subconsciously added a *symmetry constraint* by requiring that the two bond lengths are equal. This means that I am effectively reducing the number of degrees of freedom, for the purposes of geometry optimization. We will see later that this is not necessary.

I had no particular software in mind when I wrote this example (although it would be fine for Gaussian 03, as we will see in later chapters). There are many Z-matrix pits for the unwary to fall into, some simple and some subtle. Two simple ones: I did not call the first atom ‘Oxygen’ because many packages limit you to a small number of characters for atomic descriptors; also, I was careful to include a decimal point for the starting values of ROH and Ang, since many packages still expect floating-point numbers rather than integers.

For molecules having more than three atoms, we also have to define the dihedral angles in addition to bond lengths and bond angles. Figure 5.10 shows a Z-matrix for ethene, where I have made a lot of assumptions about the molecular geometry. All the C–H bond lengths are equal, and all the HCC bond angles are also taken to be equal. I have also assumed that the molecule is planar (as indeed it is, in its electronic ground state).

I should admit that Z-matrices are not easy things to write for large molecules, and in any case downloads of molecular geometries (for example, from the Protein Data Bank) are usually in Cartesian coordinates. There is also a problem to be addressed, namely that



Z-Matrix

```

C1
C2  C1  RCC
H1  C1  RCH  C2  AngleHCC
H2  C2  RCH  C1  AngleHCC  H1  180.
H3  C2  RCH  C1  AngleHCC  H1  0.
H4  C1  RCH  C2  AngleHCC  H3  180.

RCC  1.34
RCH  1.08
AngleHCC  120.

```

Figure 5.10 *Ethene Z-matrix*

there are at most $3N - 6$ vibrational degrees of freedom for a nonlinear molecule that is described by $3N$ Cartesian (or other) coordinates. The Z-matrix has to be correctly written in order to describe these $3N - 6$ internal coordinates, and a frequent cause of problems in optimizations specified by a Z-matrix was (and still is) the specification of too many or too few internal coordinates. Too few coordinates correspond to a constraint on the geometry and means that the full surface is not searched. Too many coordinates result in a redundancy leading to zero eigenvalues in the Hessian (which cannot then be inverted).

In the early days of geometry optimizations it was usual to take all kinds of shortcuts in order to save on computer resources. For example, we would routinely take all the C–H bond lengths equal in substituted benzenes, yet there are subtle, systematic differences and these should come out of a respectable calculation. Computer resources are much cheaper than was once the case, and the assumptions of constant bond length, etc. are no longer needed.

The Z-matrix method is often claimed to be intuitive to chemists because it uses the everyday concepts of bond lengths, bond angles and so on. It was always an acquired taste and in any case many databases give molecular geometries in terms of Cartesian coordinates, not internal ones.

5.11 The End of the Z-Matrix

Geometry optimization is of major importance in modern molecular modelling. Most of the early packages used internal coordinates as input by the Z-matrix. Practically all modern (gradient) optimization procedures require calculation of the Hessian \mathbf{H} and/or its inverse. In practice, it is usual to make an estimate and update these estimates at every iteration. Sometimes, the initial Hessian is taken to be the unit matrix, sometimes not. A great strength of the internal coordinate method is that construction of the initial Hessian can be based on chemical ideas; the individual diagonal elements of \mathbf{H} are identified as bond stretching and

bond bending force constants, etc. Also, the redundant translational and rotational degrees of freedom have already been eliminated.

Baker and Hehre (1991) investigated the possibility of performing gradient geometry optimization directly in terms of Cartesian coordinates. Their key finding concerned the identification of a suitable initial Hessian, together with a strategy for updates. They reported on a test set of 20 molecules, and argued that optimization in Cartesian coordinates can be just as efficient as optimization in internal coordinates, provided due care is taken of the redundancies.

There are actually two mathematical problems here, both to do with the Newton–Raphson formula. Modern optimizations use gradient techniques, and the formula

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)}$$

demonstrates that we need to know the gradient vector and the Hessian in order to progress iterations. In the discussion above, I was careful to stress the use of p independent variables q_1, q_2, \dots, q_p .

As stressed many times, there are $p = 3N - 6$ independent internal coordinates for a nonlinear molecule but $3N$ Cartesian coordinates. The familiar Wilson \mathbf{B} matrix relates these

$$\mathbf{q} = \mathbf{B}\mathbf{X} \quad (5.10)$$

\mathbf{B} has $3N$ rows and p columns, and the rows of \mathbf{B} are linearly dependent. The molecular potential energy depends on the q 's, and also on the X 's but we need to be careful to distinguish between dependent and independent variables. We can certainly write gradient vectors in terms of the two sets of variables:

$$\text{grad } U = \begin{pmatrix} \frac{\partial U}{\partial q_1} \\ \frac{\partial U}{\partial q_2} \\ \vdots \\ \frac{\partial U}{\partial q_p} \end{pmatrix}, \quad (\text{grad } U)_C = \begin{pmatrix} \frac{\partial U}{\partial X_1} \\ \frac{\partial U}{\partial X_2} \\ \vdots \\ \frac{\partial U}{\partial X_{3N}} \end{pmatrix}$$

where I have added a subscript 'C' to show that the differentiations are with respect to the $3N$ Cartesian coordinates, but it is $\text{grad } U$ that appears in optimization expressions and not $(\text{grad } U)_C$. We therefore have to relate the two.

Starting from $\mathbf{q} = \mathbf{B}\mathbf{X}$ we can show that the gradients are related by

$$\mathbf{B}^T \text{grad } U = (\text{grad } U)_C$$

Remember that we want an expression for $\text{grad } U$; if \mathbf{B} were square and invertible we would just write

$$\text{grad } U = (\mathbf{B}^T)^{-1} (\text{grad } U)_C$$

but unfortunately \mathbf{B} is rectangular. We therefore appeal to the mathematical concept of a *generalized inverse*. Consider the $3N \times 3N$ matrix $\mathbf{G} = \mathbf{B}\mathbf{u}\mathbf{B}^T$, where \mathbf{u} is an arbitrary nonsingular $p \times p$ matrix. The generalized inverse of \mathbf{G} , written \mathbf{G}^{-1} , is a matrix with the property that $\mathbf{G}\mathbf{G}^{-1}$ is a diagonal matrix with a certain number of 1's and a certain number

of 0's along the diagonal. In our case, \mathbf{GG}^{-1} will have p 1's along the diagonal and $3N - p$ zeros. There are many algorithms for calculating such a generalized inverse. The zeros might actually be tiny numbers, depending on the accuracy of the numerical algorithm used.

It can be shown that the gradient vector referred to the independent variables is related to the Coordinate gradient vector by

$$\text{grad } U = \mathbf{G}^{-1}\mathbf{Bu}(\text{grad } U)_C \quad (5.11)$$

Similar considerations apply to the Hessian, and formulae are available in the literature.

5.12 Redundant Internal Coordinates

All coordinate systems are equal in principle, but I have stressed above that:

- they should ideally give transferable force constants from molecule to molecule;
- they should ideally be capable of being represented as harmonic terms in U , i.e. cubic and higher corrections should not be needed.

These requirements can be best satisfied by local internal valence coordinates such as bond lengths, bond angles and dihedral angles. The expression 'local' in this context means that the coordinates should extend to only a few atoms.

Pulay *et al.* (1979) and Schlegel and co-workers (see Peng *et al.* 1994, 1996) both investigated the use of redundant internal coordinates for gradient optimizations. Pulay *et al.* defined an internal coordinate system similar to that used by vibrational spectroscopists. It minimizes the number of redundancies by using local symmetry coordinates about each atom and special coordinates for ring deformations, ring fusions, etc. Schlegel and co-workers used a simpler set of internal coordinates composed of all bond lengths, valence angles and dihedral angles.

The mathematical considerations outlined above also apply here.

Packages like Gaussian 03 offer a choice between Z-matrix and Cartesian coordinates for input of the molecular geometry. Gaussian 03 then routinely uses redundant internal coordinates for the optimizations rather than the Z-matrix.

References

- Baker, J. and Hehre, W.J. (1991) *J. Comput. Chem.*, **12**, 606.
 Fletcher, R and Powell, M.J.D. (1963) *Comput. J.*, **6**, 163.
 Fletcher, R and Reeves, C.M. (1964) *Comput. J.*, **7**, 149.
 Hendrickson, J.B. (1961) *J. Am. Chem. Soc.*, **83**, 4537.
 Peng, C and Schlegel, H.B. (1994) *Israel J. Chem.*, **33**, 449.
 Peng, C., Ayala, P.Y. and Schlegel, H.B. (1996) *J. Comput. Chem.*, **17**, 49.
 Polak, E. and Ribière, G. (1969) *Rev. Fr. Inform. Rech. Operation*, **16-R1**, 35.
 Pulay, P., Fogarasi, G., Pang, F. and Boggs, J.E. (1979) *J. Am. Chem. Soc.*, **101**, 2550.
 Westheimer, F.H. (1956) in *Steric Effects in Organic Chemistry* (ed. M.S. Newman), John Wiley & Sons, Inc., New York.
 Wiberg, K. (1965) *J. Am. Chem. Soc.*, **87**, 1070.

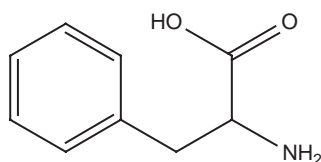
6

Molecular Mechanics Examples

As a rough guide, molecular mechanics (MM) is used these days to study possible equilibrium conformations of large molecules. In this chapter I will give an example of what can be routinely done at this level of theory.

There are many packages on the market, and everyone has their favourite (or has it imposed on them by their employer). My two favourite packages are Gaussian 03 (<http://www.gaussian.com>) and HyperChem v7 (<http://www.hyper.com>). The latest release of HyperChem is actually v8. In my opinion, everything boils down to the graphical user interface (GUI) for MM applications on large molecules and HyperChem has a first-class integrated GUI. You do not need a third party package to input and visualize the calculations.

So for this chapter I am going to use HyperChem and treat the very simple molecule phenylalanine (Figure 6.1) discussed earlier in the book.



Phenylalanine

Figure 6.1 Structure of phenylalanine

6.1 Geometry Optimization

Geometry optimization requires a starting structure. I mentioned the Z-matrix in Chapter 5, and that is how molecules were input before the days of the GUI. We have moved along

since then, and to get started I simply built the molecule onscreen in HyperChem using the ‘build’ option (Figure 6.2).

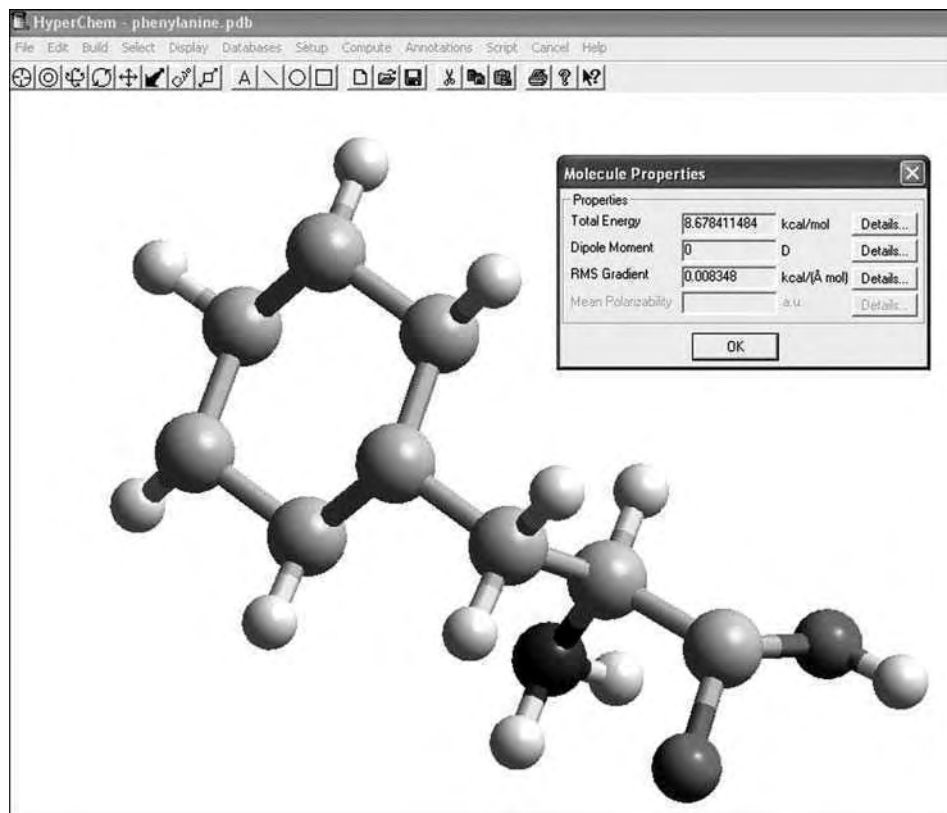


Figure 6.2 *HyperChem screen grab*

You probably noticed the single chiral centre; most packages have an automatic inversion feature in case you want to switch between D and L forms.

Most packages also give the so-called ‘rendering’ as an option. I chose ‘balls and cylinders’ but you can experiment to get whatever pleases you best. There is a choice of optimization method and a choice of force field. AMBER seemed the appropriate choice for an amino acid, and I easily located a stationary point with energy 8.678411 kcal mol⁻¹.

This energy is not for direct comparison with experiment; only differences are of any significance in MM calculations. In this example, I am going to use the energy as a relative marker in our search for the global minimum.

6.2 Conformation Searches

The stationary point located above is of course a local minimum. Figure 6.3 shows the results we might expect to find from a conformational search on a (hypothetical) molecule.

Repeating the optimization from different starting points has identified many local minima and the conformation corresponding to energy ε_1 appears to be the global minimum.

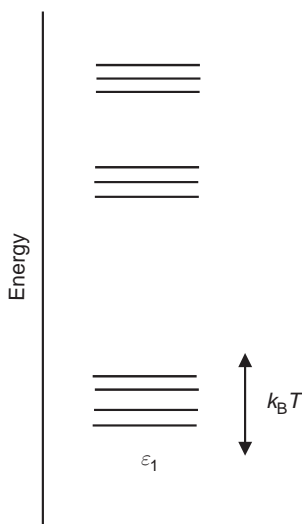


Figure 6.3 Conformation search

Our MM calculations refer to the molecule at 0 K, whilst we would normally concern ourselves with molecules at room temperature. At 0 K, we would find all molecules in the lowest energy conformation but at other temperatures we will find certain fractions of the molecules in higher energy conformations. These fractions are determined by the Boltzmann factor

$$\exp\left(-\frac{\varepsilon}{k_B T}\right)$$

where k_B is the Boltzmann constant. I have therefore drawn a vertical line on the figure of size $k_B T$, and we see that the first four conformations lie close together (compared to the yardstick $k_B T$), and so we expect that all four would make a significant contribution to measured properties of the molecule at temperature T .

The remaining groups of conformations are far away from the first four in terms of the yardstick $k_B T$, and so can be discarded.

Several approaches have been applied to the problem of automatically identifying the low-energy conformations of a given molecule. Molecular flexibility is usually associated with rotation of unhindered bond dihedral angles, with little change in bond lengths and bond angles. A common theme is therefore to limit the exploration to those parameters that have the *least* effect on the energy. Another strategy is to automatically scan a possible structure for unfavourably close nonbonded atoms and unlikely ring closure bond lengths before optimization.

A choice then has to be made to decide whether the starting structure should be the same for every optimization, or each optimization should start with the structure found by the

previous successful optimization. The latter choice is often favoured, on the grounds that one low-energy conformation is pretty much like another, and starting from one successful geometry will tend to keep the search in a low-energy region of the potential surface.

Variations in the parameters can be either random or systematic; in the case of systematic variations, it is usual to scan the surface first at ‘low resolution’, then to refine likely looking regions at high resolution. In the case of random variations, a common problem is that as the calculation progresses, more and more of the trial structures lead to conformations that have already been examined. Many modern software packages have sophisticated algorithms that can recognize identical conformations, and so reject them from the global search. They also automatically reject conformations that are too high in energy, compared to $k_B T$.

Figure 6.4 shows a run using the ‘Conformational Search’ option of HyperChem. I had to decide which torsional (dihedral) angles to vary, and I chose just two for this illustrative calculation: these are called ‘4-7-8-10’ and ‘5-4-7-8’, with the atom numbering shown.

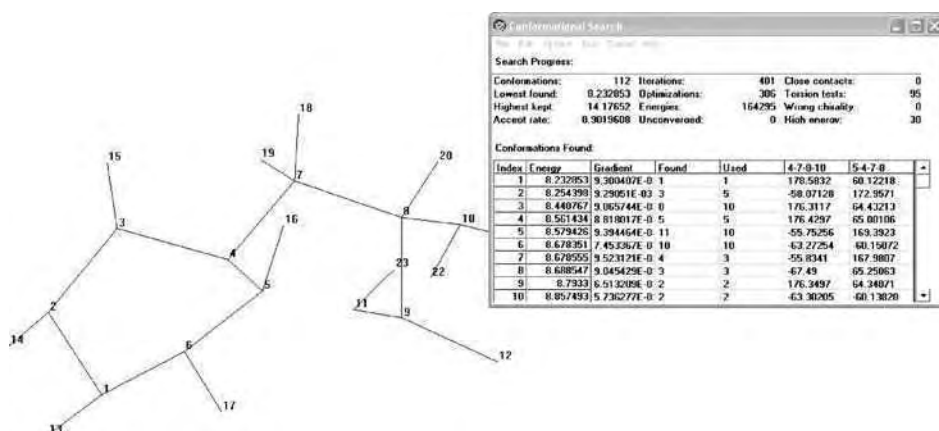


Figure 6.4 Random walk conformation search

The ‘Found’ table entry shows the number of times that each particular conformation has been found, and ‘Used’ shows the number of times it has been used as a starting structure. Energies are recorded as kcal mol^{-1} .

6.3 Amino Acids

Whilst on the subject of amino acids, Hyperchem has a database of the amino acid residues that can be quickly joined together to construct a chain. The pK of most amino acids is such that the molecules exist in zwitterion form in aqueous solution, and there is a ‘Make Zwitterion’ option. It can be argued that one should include any ‘formal’ atom charges such as the zwitterion charges explicitly, and in the case of phenylalanine I would modify the formal charge on each oxygen atom to be $-1/2$ electron and add a charge of $+1$ electron on the nitrogen.

6.3.1 Molecular Structure Databases

Molecular geometries can be determined for gas-phase molecules by microwave spectroscopy and by electron diffraction. In the solid state, the field of structure determination is dominated by X-ray and neutron diffraction and very many crystal structures are known. Nuclear magnetic resonance also has a role to play, especially for proteins. All of these topics are well discussed in every university-level general chemistry text.

Over the years, a vast number of molecular structures have been determined and there are several well-known structural databases. One is the Cambridge Structural Database (CSD), which is supported by the Cambridge Crystallographic Data Centre (CCDC). The CCDC was established in 1965 to undertake the compilation of a computerized database containing comprehensive data for organic and metal–organic compounds studied by X-ray and neutron diffraction. It was originally funded as part of the UK contribution to international data compilation. According to its mission statement, the CCDC serves the scientific community through the acquisition, evaluation, dissemination and use of the world's output of small-molecule crystal structures.

For each entry in the CSD, three types of information are stored. First is the bibliographic information: who reported the crystal structure, where they reported it, and so on. Next comes the *connectivity* data; this is a list showing which atom is bonded to which in the molecule. Finally the molecular geometry and the crystal structure are given. The molecular geometry consists of Cartesian coordinates. The database can be easily reached through the Internet, but individual records can only be accessed on a fee-paying basis.

The Brookhaven Protein Data Bank (PDB) is the single worldwide repository for the processing and distribution of three-dimensional biological macromolecular structural data. It is operated by the Research Collaboratory for Structural Bioinformatics. At the time of writing, there were 49 048 structures in the databank, relating to proteins, nucleic acids, protein–nucleic acid complexes and viruses. The databank is available free of charge.

Information can be retrieved from the main website. A four-character alphanumeric identifier such as 1PCN represents each structure. The PDB database can be searched using a number of techniques, all of which are described in detail at the homepage.

6.3.2 The .pdb File Format

The PDB file format (.pdb) is widely used to report and distribute molecular structure data. A typical .pdb file for phenylalanine would start with bibliographic data, then move on to the Cartesian coordinates (expressed in angstroms and relative to an arbitrary reference frame) and connectivity data as shown below.

```
HETATM 1 C 1 -1.576 -0.433 0.004
HETATM 2 C 2 -1.301 0.777 0.643
HETATM 3 C 3 -0.072 1.410 0.444
HETATM 4 C 4 0.898 0.848 -0.394
HETATM 5 C 5 0.609 -0.365 -1.029
HETATM 6 C 6 -0.617 -1.004 -0.834
HETATM 7 C 7 2.227 1.537 -0.618
HETATM 8 C 8 3.352 0.904 0.217
```

```
HETATM 9 C 9 4.694 1.557 -0.049
HETATM 10 N 10 3.025 1.002 1.616
```

etc.etc.

```
CONNECT 1 2 6 13
CONNECT 2 1 3 14
CONNECT 3 2 4 15
CONNECT 4 3 5 7
```

etc. etc.

END

Records 1 through 23 identify the atoms; so, for example, atom 1 is a carbon. The ‘CONNECT 1 2 6 13’ record tells us that atom 1 (C) is joined to atoms 2, 6 and 13.

6.4 QSAR

Quantitative structure–activity relationships (QSARs) are attempts to correlate molecular structure with chemical or biochemical activity. I will use the remainder of this chapter to give you a flavour for the QSAR properties that are often routinely calculated in MM studies.

6.4.1 Atomic Partial Charges

Atomic charges can be calculated using the (costly) techniques of molecular quantum mechanics, as we will discover in later chapters. They are often calculated within the spirit of QSAR analyses using the partial equalization of orbital electronegativity (PEOE) approach of Gasteiger and Marsili (1980). The key concept of their method is atom *electronegativity* (often given the symbol χ), and an attraction is that their calculation is rapid. Pauling and Yost (1932) first introduced this concept, and they tried to give numerical values to atomic electronegativities based on bond energy values. Mulliken (1934) put the concept on a firmer theoretical footing by relating an atomic electronegativity to the ionization energy I and the electron affinity E of the atom concerned:

$$\chi = \frac{1}{2}(I + E) \quad (6.1)$$

The only problem with Mulliken’s definition is that both E and I relate to some mythical state of an atom called the *valence state*. There have been many learned discussions relating to the concept of valence state, and such discussions still occasionally appear in the literature.

Whatever the valence state of an atom is (depending on the atom), it rarely corresponds to a spectroscopic state and so cannot be studied experimentally. Electronegativities and valence states go hand in hand; they are part of our folklore and both are widely quoted in even the most elementary chemistry texts.

Sanderson (1951) proposed that, on bond formation, atoms change their electron density until their electronegativities are equal. These ideas were further elaborated by Hinze and co-workers (1963), who introduced the concept of *orbital electronegativity*. This is the

electronegativity of a specific orbital in a given valence state. You will understand that I am introducing quantum mechanical concepts before we have finished our discussion of techniques based on classical mechanics; just bear with me.

In the abstract of their paper, Gasteiger and Marsili (1980) state:

A method is presented for the rapid calculation of atomic charges in σ -bonded and nonconjugated π systems. Atoms are characterized by their orbital electronegativities. In the calculation, only the connectivities of the atoms are considered. Thus only the topology of a molecule is of importance. Through an iterative procedure partial equalization of orbital electronegativity is obtained. Excellent correlations of the atomic charges with core electron binding energies and with acidity constants are observed.

These authors decided to relate the electronegativity of orbital ν on atom i ($\chi_{i,\nu}$) to the total charge Q_i on atom i as a quadratic

$$\chi_{i,\nu} = a_{i,\nu} + b_{i,\nu}Q_i + c_{i,\nu}Q_i^2 \quad (6.2)$$

The three unknowns a , b and c were determined from values used at the time for orbital ionization energies and affinities. Typical values are listed in Table 6.1.

Table 6.1 Parameters needed to calculate electronegativity

Atom	Hybridization	a	b	c
H		7.17	6.24	-0.56
C	sp^3	7.98	9.18	1.88
	sp^2	8.79	9.32	1.51
	sp^1	10.39	9.45	0.73

The calculation is iterative; starting from a reference structure where all the atoms carry zero charge, electric charge is permitted to flow from the less to the more electronegative atoms. The electronegativity of an atom decreases as it accumulates electric charge, according to the formula, and in the next iteration less charge will flow until eventually the electronegativity of each (charged) atom is equalized. At this point the flow of charge stops.

Molecular mechanics is of course a classical treatment, and orbitals do not appear. The key equation is therefore written in terms of the atoms overall rather than in terms of orbital contributions:

$$\chi_i = a_i + b_iQ_i + c_iQ_i^2 \quad (6.3)$$

Atom electronegativities are adjusted at each iteration using the calculated charges, until self-consistency is achieved. Figure 6.5 shows typical results for phenylalanine. Notice the negative charges on oxygen and the positive charge on nitrogen. Charges calculated in this way depend only on the atomic connectivity and not on the molecular geometry.

6.4.2 Polarizabilities

I introduced you to polarizabilities in Chapter 2, when we discussed the theory of intermolecular forces. The experimental determination of a molecular polarizability is far from straightforward, especially if the molecule has little or no symmetry. The classical experimental route to the mean polarizability $\langle\alpha\rangle$ is via the refractive index or relative

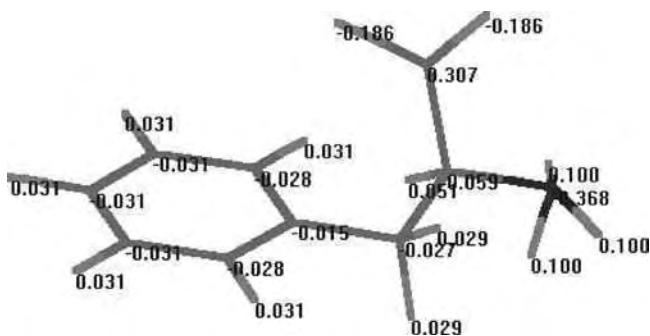


Figure 6.5 Atomic partial charges for phenylalanine zwitterion

permittivity of a gas. Polarizabilities can be routinely calculated by quantum mechanical techniques, but the calculations are compute-intensive and this route is of no help if your aim in life is the high-throughput screening of very many potential pharmaceutical molecules.

An insight into the order of magnitude of polarizabilities can be obtained by considering Figure 6.6, which represents a (mythical) atom. The nucleus of charge Q is at the coordinate origin, and the nucleus is surrounded by an electron cloud of radius a . The electrons are assumed to have a uniform density in space. The total electron charge is $-Q$ and so the atom is electrically neutral.

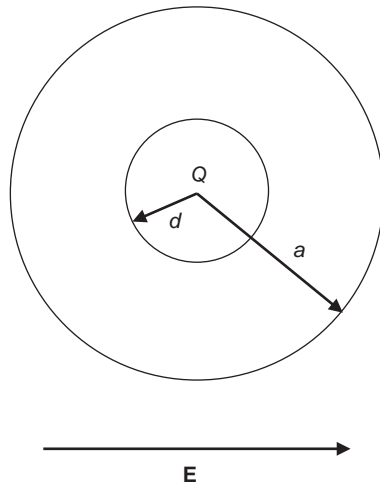


Figure 6.6 Construct needed to discuss polarizability

When an external uniform electric field is applied, the nucleus is displaced a relative distance d in the direction of the field. At this point, the force on the nucleus QE is exactly balanced by the force exerted on the nucleus by the electron cloud.

According to Gauss' electrostatic theorem, this is the same force as would be exerted if all the charge within a sphere of radius d were concentrated at the centre. This charge is $-Q$ times the ratio of the volumes of spheres of radius a and d and so is $-Qd^3/a^3$. This gives a force of magnitude

$$Q^2 \frac{d^3}{a^3} \frac{1}{4\pi\epsilon_0} \frac{1}{d^2} = \frac{Q^2 d}{4\pi\epsilon_0 a^3}$$

Hence the displacement d satisfies

$$\frac{Q^2 d}{4\pi\epsilon_0 a^3} = QE$$

The induced dipole is Qd and the polarizability is Qd/E and so

$$\alpha = 4\pi\epsilon_0 a^3 \quad (6.4)$$

Apart from the dimensional factor, the volume of the sphere determines the polarizability.

Table 6.2 shows the experimental polarizabilities for three inert gas atoms, which illustrates the dependence on volume.

Table 6.2 Polarizabilities of inert gases

Inert gas	$\alpha/10^{-40} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$
He	0.23
Ne	0.44
Ar	1.83

Table 6.3 shows a comparison between typical $\langle\alpha\rangle$ QSAR model calculations for a number of alkanes, together with the result of a serious quantum mechanical calculation. The latter are recorded in the 'Accurate' column. At the chosen level of quantum mechanical theory (BLYP/6-311++G(3d,2p) with optimized geometries, to be discussed in later chapters), the results are at least as reliable as those that can be deduced from spectroscopic experiments. The downside is that the butane calculation took approximately four hours on my home PC; the corresponding parameterized calculation took essentially no time at all.

Table 6.3 Comparison of accurate quantum mechanical calculation with QSAR

Alkane	Accurate $\langle\alpha\rangle/10^{-40} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$	QSAR $\langle\alpha\rangle/10^{-40} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$
Methane	2.831	2.90
Ethane	4.904	4.94
Propane	7.056	6.99
Butane	9.214	9.02

It looks at first sight that there is an incremental effect. Adding a CH_2 increases the mean polarizability by about $2.13 \times 10^{-40} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$ and there has been much speculation over the

years as to whether a molecular polarizability can generally be written as a sum of atomic contributions.

Many atomic parameter schemes have been proposed, yet there were always exceptions that needed a complicated sub-rule in order to get agreement with experiment. Silberstein (1917) wrote in his famous paper of 1917 of

... the ever growing hierarchy of rules indicating how to treat the exceptions to the law of additivity, although helpful to the chemist, is the clearest confession of non-additivity.

The concept of bond polarizability was then introduced, in order to try and circumvent the additivity problem. For example, the C–H bond polarizability α_{CH} is taken to be a quarter of the methane mean polarizability $\langle\alpha\rangle$, and for the alkanes of formula $\text{C}_n\text{H}_{2n+2}$ we have

$$\langle\alpha(\text{C}_n\text{H}_{2n+2})\rangle = (n-1)\langle\alpha_{\text{CC}}\rangle + (2n+2)\langle\alpha_{\text{CH}}\rangle \quad (6.5)$$

from which it is possible to deduce a value for α_{CC} . This method actually reproduces experimental data for hydrocarbons to within a few percent. But once again, such additivity schemes fail when applied to molecules containing heteroatoms, and to hydrocarbons containing atoms in different states of hybridization.

Miller and Savchik (1979) proposed a method whereby $\langle\alpha\rangle$ was written as the square of a sum of atomic hybrid components. If they had known about 4-quantity electromagnetic formulae they would have written their formula

$$\langle\alpha\rangle = 4\pi\epsilon_0\frac{4}{N}\left[\sum_A\tau_A\right]^2 \quad (6.6)$$

but in the event, the $4\pi\epsilon_0$ is missing from their paper.

The number of electrons in the molecule is N , and τ are parameters for each atom in its particular hybrid configuration. The summation is over all atoms in the molecule, and the τ 's depend on the chemical hybridization but not on the chemical environment; they are constant for (for example) an sp^2 carbon atom.

The authors based their algorithm on an old quantum mechanical treatment of dipole polarizability discussed in the classic book by Hirschfelder *et al.* (1954). The Miller–Savchik treatment is widely used in QSAR studies, and to give you a flavour I have recalculated the mean polarizabilities and included them in Table 6.3 under the QSAR heading. Given that each calculation took just a few seconds, you can see why QSAR practitioners think so highly of the Miller–Savchik technique. To put things in perspective, we should not be surprised to find such excellent agreement between the two sets of results in Table 6.3; it is heterosubstituted molecules that pose the severe test.

6.4.3 Molecular Volume and Surface Area

Molecular volumes are often calculated by a numerical integration grid technique that I can illustrate by considering the trivial problem of finding the volume of an atom whose van der Waals radius is R (the volume is of course $\frac{4}{3}\pi R^3$).

Figure 6.7 shows a two-dimensional representation of the atom whose van der Waals radius is R , surrounded by a three-dimensional grid of equally spaced points. The grid has its centre at the atom centre, and the edges of the grid correspond to the van der Waals radius.

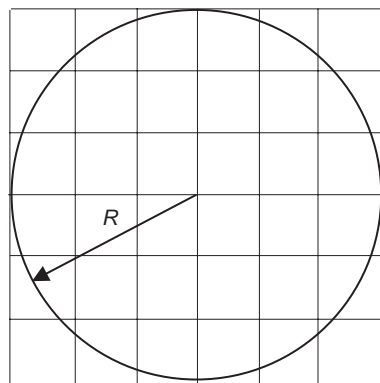


Figure 6.7 Grid around an atom

For each grid point in turn, we calculate its distance from the centre and determine whether the grid point lies inside or outside the atom. If n is the total number of grid points and n_a the number that lie within the atom whose volume is V , we have

$$\frac{V}{8R^3} = \frac{n_a}{n}$$

For a polyatomic, we have to give special consideration to grid points that lie in the overlap region. Figure 6.8 shows two atoms A and B, with radii R_A and R_B . The overlap region is labelled X.

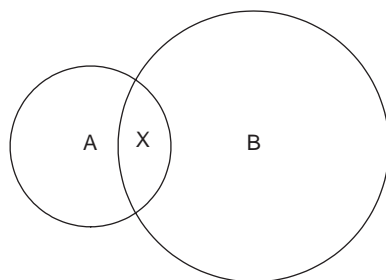


Figure 6.8 Atoms A, B and overlap region X

For atom A, we know that the volume is $\frac{4}{3}\pi R_A^3$. We now surround atom B with a grid as described above, and test each grid point in turn. If the grid point lies within sphere B, we test to see if it lies in region X and so has already been counted as part of the volume of atom A. The algorithm proceeds until all atoms have been considered in turn. The molecular volume is found by adding all atomic contributions.

There are similar methods for the estimation of molecular surface area.

6.4.4 Log(*P*)

When an organic solute is in equilibrium with water and a polar solvent such as octan-1-ol, it is observed that the ratio of the concentrations

$$P = \frac{[\text{Solute in octan-1-ol}]}{[\text{Solute in water}]}$$

is roughly constant. *P* is called the *partition coefficient*, and *P* can be used in predicting transmembrane transport properties, protein binding, receptor affinity and pharmacological activity. It is easy to determine *P* experimentally, but in the process of molecular design we have to deal with a high throughput of possible molecular structures and so numbers of attempts have been made to give simple models for predicting *P*.

In studying the effect of structural variations on *P*, it was suggested that it had an additive-constitutive character. In the case of a π -substituent, people made use of Hammett's ideas and wrote

$$\pi_X = \log_{10}(P_X) - \log_{10}(P_H) \quad (6.7)$$

where P_H is the partition coefficient for the parent compound and P_X the partition coefficient for a molecule where X has been substituted for H. Workers in the field refer to 'log(*P*)' rather than '*P*'.

It was originally envisaged that specific substituents would have the same contribution in different molecules. It has been demonstrated, however, that this hoped-for additivity does not even hold for many disubstituted benzenes. There are two classical methods for estimating log(*P*), both based on the assumed additivity: the *f*-constant method of Rekker (1976) and the fragment approach of Leo *et al.* (1975).

Rekker defined an arbitrary set of terminal fragments using a database of some 1000 compounds with known log(*P*). Linear regression was performed, and the regression coefficients designated *group contributions*. Deviations from the straight lines were corrected by the introduction of multiples of a so-called 'magic factor' that described special structural effects such as polar groups, etc. Log(*P*) is calculated from the fragmental contributions and the correction factors.

Leo *et al.* derived their own set of terminal fragments, together with a great number of correction factors.

Klopman and Iroff (1981) seem to be the first authors to make use of quantum mechanical molecular structure calculations. They performed calculations at the quantum mechanical MINDO/3 level of theory (to be discussed in Chapter 18) in order to calculate the atomic charge densities of a set of 61 simple organic molecules. They then developed a linear regression model that included the number of C, N, H and O atoms in the given molecule, the atomic charges on C, N and O, and certain 'indicator' variables n_A , n_T and n_M designed to allow for the presence of acid/ester, nitrile and amide functionalities. They found

$$\log_{10}(P) = 0.344 + 0.2078n_H + 0.093n_C - 2.119n_N - 1.937n_O \\ - 1.389q_C^2 - 17.28q_N^2 + 0.7316q_O^2 + 2.844n_A + 0.910n_T + 1.709n_M \quad (6.8)$$

The terms involving q^2 represent the interaction of the solute and solvent.

The method of Klopman and Iroff was a great step forward; there are many fewer parameters, it does not produce ambiguous results depending on an arbitrary choice of fragment scheme and it does not have a complicated correction scheme, apart from the

indicator variables. As we will see in Chapter 18, MINDO/3 calculations can be done very quickly even for large molecules.

Bodor *et al.* (1989) decided to extend and enhance Klopman's work by including the molecular volume, weight and surface area. In addition, they included the molecular electric dipole moment in the correlation, all possible products of pairs of charges and a further 57 compounds in the analysis. They used a more up-to-date quantum mechanical model, AM1 (again to be discussed in Chapter 18). The molecular volume and surface area were calculated by a grid method. Their equation is

$$\begin{aligned} \log_{10}(P) = & -1.167 \times 10^{-4} S^2 - 6.106 \times 10^{-2} S + 14.87 O^2 - 43.67 O + 0.9986 I_{\text{alkane}} \\ & + 9.57 \times 10^{-3} M_w - 0.1300 D - 4.929 Q_{\text{ON}} - 12.17 Q_{\text{N}}^4 + 26.81 Q_{\text{N}}^2 \quad (6.9) \\ & - 7.416 Q_{\text{N}} - 4.551 Q_{\text{O}}^4 + 17.92 Q_{\text{O}}^2 - 4.03 Q_{\text{O}} + 27.273 \end{aligned}$$

where S (cm²) is the molecular surface area, O the molecular ovality, I_{alkane} the indicator variable for alkanes (it is 1 if the molecule is an alkane, 0 otherwise), M_w the relative molar mass, D the calculated electric dipole moment and Q_{ON} the sum of absolute values of the atomic charges on nitrogen and oxygen. The remaining Q 's differ from those used by Klopman and Iroff in that they are the square root of the sum of charges on the nitrogen or oxygen atoms. The most significant parameters are the volume and the surface area; the authors claim that this demonstrates that the most important contribution to $\log(P)$ is the creation of a hole in the structure of water.

6.5 Problem Set

A set of problems, together with solutions, is available on the associated website.

References

- Bodor, N., Gabanyi, Z. and Wong, C.-K. (1989) *J. Am. Chem. Soc.*, **111**, 3783.
 Gasteiger, J. and Marsili, M. (1980) *Tetrahedron*, **36**, 3219.
 Hinze, J., Whitehead, M.A. and Jaffé, H.H. (1963) *J. Am. Chem. Soc.*, **85**, 148.
 Hirschfelder, J.O., Curtiss, C.F. and Bird, R.B. (1954) *Molecular Theory of Gases and Liquids*, John Wiley & Sons, Inc., New York.
 Klopman, G. and Iroff, L.D. (1981) *J. Comput. Chem.*, **2**, 157.
 Leo, A., Jow, P.Y.C., Silipo, C. and Hansch, C. (1975) *J. Med. Chem.*, **18**, 865.
 Miller, K.J. and Savchik, J.A. (1979) *J. Am. Chem. Soc.*, **101**, 7206.
 Mulliken, R.S. (1934) *J. Chem. Phys.*, **2**, 782.
 Pauling, L. and Yost, D.M. (1932) *Proc. Natl Acad. Sci. USA*, **14**, 414.
 Rekker, R.E. (1976) *The Hydrophobic Fragment Constant*, Elsevier, Amsterdam.
 Sanderson, R.T. (1951) *Science*, **144**, 670.
 Silberstein, L. (1917) *Philos. Mag.*, **33**, 92.

7

Sharing Out the Energy

The structure of matter can be thought about on two different levels. On the one hand, there is the macroscopic view where we focus on equations such as the perfect gas equation of state

$$pV = nRT \quad (7.1)$$

This approach deals with large chunks of matter and involves ideas like temperature, heat capacity and pressure that are familiar from everyday life. It is based on measurements taken with thermometers, pressure gauges and other instruments that are familiar in an ordinary science laboratory.

On the other hand there is the microscopic approach where we focus on individual atoms and molecules. This approach involves rather abstract ideas and tends to be based on theory rather than experiment. Atoms and molecules can be studied directly by spectroscopic techniques, but they bring their own problems because we usually have to understand quantum mechanical principles before we can interpret the experimental findings.

Another problem is that there are typically 10^{23} particles in a macroscopic sample, a number that beggars understanding.

Then again, the concept of temperature does not appear from Newton's equations, nor does it appear in the Schrödinger equation. It is actually a measure of the way that energy is shared out between available quantum states, but more of this shortly.

There must be a link between the two approaches, but what can it be? To be positive, things are not as bad as they might seem. What matters in a macroscopic sample are not the individual properties of the constituent atoms and molecules, rather their statistical properties. It does not matter whether we know which individual particle of 10^{23} collides with the wall of a container when we want to understand the pressure; what does matter is the average behaviour of the 10^{23} particles present.

The two great pioneers of this subject were Ludwig Boltzmann (1844–1906) and Josiah Willard Gibbs (1839–1903). In fact they worked independently. Boltzmann's tomb carries his famous equation relating a thermodynamic property (entropy, S) to a statistical

quantity W . You might have come across Gibbs in a different context as the pioneer of modern vector calculus.

The link between microscopic and macroscopic behaviour is provided by *statistics*, and in this case a special branch of mechanics called *statistical mechanics*. Statistics deals with the collection, organization and analysis of experimental data and it is important that you should have the basic concepts at your fingertips before progressing.

The idea of a random number will appear again and again through the text; a sequence of random numbers is one where there is no discernible trend from one number to another. Your pocket calculator has a key labelled RAN#; press the key 10 times and make a note of the numbers generated. My results are given in Table 7.1; yours will be different. This is a series of random numbers; there is no discernible trend over the series and in this case they are equally distributed over the range of values 0 to 1. If you switch off your pocket calculator, switch it back on again and repeat the experiment, you will almost certainly find exactly the same sequence because they are generated numerically from a *seed*. For this reason (amongst others), mathematicians refer to such series as *pseudo-random*. Such numbers are good enough for experiments where random numbers are needed, and the sequence can be reproduced time and time again. If you choose a different seed, you will get a different sequence.

Table 7.1 Ten pseudo-random numbers from a calculator

Key press	1	2	3	4	5	6	7	8	9	10
Number	0.861	0.478	0.127	0.427	0.648	0.275	0.571	0.813	0.471	0.063

The first thing to do is to calculate the mean, which is given by the sum of the set of numbers divided by the number of values. If I write x_1, x_2, \dots, x_n for the numbers then we usually write the mean in brackets $\langle x \rangle$, given by

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^n x_i \quad (7.2)$$

The next thing we might want to know is the spread of points about the average; sometimes the points will be widely spread about the average, sometimes they will be closely clustered. For this experiment, the points are chosen at random with values between 0 and 1, and so the spread is wide. To get a measure of the spread, we calculate the standard deviation

$$\sigma_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \langle x \rangle)^2} \quad (7.3)$$

For reasons based on statistical theory, you will often find the standard deviation calculated with a factor of $n - 1$ instead of n

$$\sigma_{n-1} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \langle x \rangle)^2} \quad (7.4)$$

For large values of n , the two standard deviations are pretty much the same. Using my data I found the following:

$$\langle x \rangle = \frac{1}{10} (0.861 + 0.478 + 0.127 + 0.427 + 0.648 + 0.275 + 0.571 + 0.813 + 0.471 + 0.063)$$

$$\langle x \rangle = 0.473 \quad \sigma_n = 0.252 \quad \sigma_{n-1} = 0.266$$

There are other ways of reporting a spread, for example the *median* is the value for which there are as many values greater than the median as there are values less than it. So for example, a company with three employees and a CEO whose hourly rates of pay are £4.20, £4.50, £5.00 and £1000 could truthfully quote in their annual report a mean salary of £253.43 per hour. The median is £4.75, which tells a different story.

7.1 Games of Chance

Let me start the discussion with a very simple experiment. I take a six-faced die whose faces have the usual scores $S = 1, 2, \dots, 6$ and throw the die 10 times. My results are given in Table 7.2. Note that not all the possible scores have appeared. It is easier to work with the *frequency* of a given score rather than keep track of every single throw; these are shown in Figure 7.1.

Table 7.2 Throwing a single die 10 times

Throw	1	2	3	4	5	6	7	8	9	10
Score, S	6	1	2	3	3	4	4	1	3	4

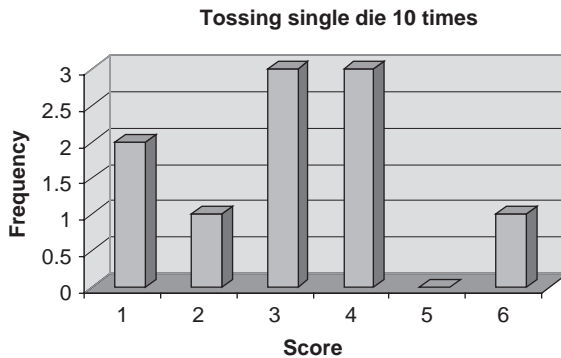


Figure 7.1 Tossing a single die

Throwing dice can be very exciting and profitable in the right surroundings, but we would like a more automated way to generate results. I gave you the hint with RAN#; and I can convert the random numbers between 0 and 1 into scores between 1 and 6 by:

- multiplying the random number by 6;
- adding 1; and then
- taking the integral part of the number

In MS/EXCEL, you would write `= INT(6*RAND() + 1)` as the cell formula and copy the formula into as many cells as necessary.

Table 7.3 shows how things change as we increase the number of throws.

Table 7.3 *Frequencies for throwing a single die n times*

Score, S	$n = 10$	$n = 100$	$n = 1000$	$n = 1\,000\,000$
1	2	19	158	166 689
2	1	17	178	166 346
3	3	21	179	167 326
4	3	16	170	165 934
5	0	16	161	166 923
6	1	11	154	166 782
$\langle S \rangle$	3.1000	3.2600	3.4600	3.5004
σ	1.4457	1.6286	1.6728	1.7080

Again, I can illustrate this as a histogram (shown in Figure 7.2, where I have plotted the fractional frequencies). The front set of results refer to $n = 100$, and so on. You can see that the larger the number of throws, the closer the number of occurrences of each score.



Figure 7.2 *Throwing a single die n times*

Advanced texts differentiate between the *sample mean* (samples of 10, 100, 1000 and 1 000 000 in this case) and the mean of an infinite number of readings called the *population mean*; it should be intuitively obvious that the population mean is $3\frac{1}{2}$ and so sample means give approximations to population means. For a large number of experiments, the fractional frequencies are found to level off at one-sixth the total number of throws, and the *probability* of each score is defined as the limiting value of the fractional frequency for a large number of experiments. In this case the probabilities are all equal to $1/6$.

7.1.1 Two Dice

Table 7.4 and Figure 7.3 show what happens when we repeat the experiment with two dice. In this case the possible scores are 2, 3, . . . , 12 and this is certainly not the same as having a single 12-sided die. To model the experiment, use your random number generator for each die and add the result; obviously your values will not be exactly the same as mine.

Table 7.4 Throwing two dice n times

	$n = 100$	$n = 1000$	$n = 1\,000\,000$
$\langle S \rangle$	7.010	6.980	7.003
σ	2.500	2.357	2.415

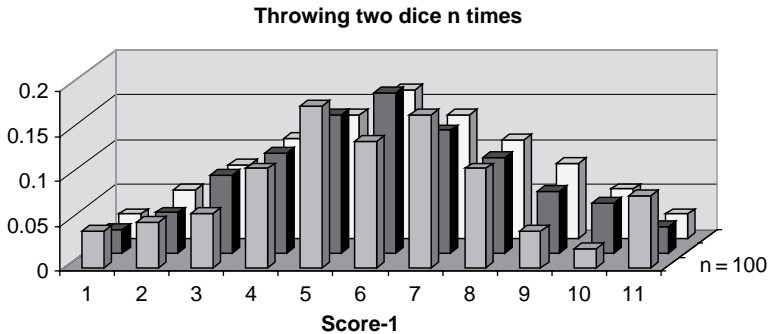


Figure 7.3 Throwing two dice

I will tell you that if we toss p dice then the theoretical mean score is p times the mean for a single die, and the standard deviation is \sqrt{p} times. The distribution has obviously settled down by the time we get to a million throws, but what has it settled down to?

7.2 Enumeration

It should be clear that we can deduce the theoretical probabilities by asking how many ways a given event can occur from all allowed possibilities. The probability is given by the ratio of the number of times an event can happen to the total number of ways. Table 7.5 shows the number of ways the two dice can fall in order to make up a given score. The theoretical probabilities are $1/36$, $2/36$, . . . , $1/36$ whilst the mean and standard deviation work out as 7 and $2\sqrt{2}$. These probabilities agree completely with the ‘experimental’ values shown.

Table 7.5 Enumeration for two dice A and B

Total score, S	Possible individual scores of A and B	$W =$ number of ways	Probability
2	1 + 1	1	1/36
3	1 + 2, 2 + 1	2	2/36
4	1 + 3, 3 + 1, 2 + 2	3	3/36
5	1 + 4, 4 + 1, 2 + 3, 3 + 2	4	4/36
6	1 + 5, 5 + 1, 2 + 4, 4 + 2, 3 + 3	5	5/36
7	1 + 6, 6 + 1, 2 + 5, 5 + 2, 3 + 4, 4 + 3	6	6/36
8	2 + 6, 6 + 2, 3 + 5, 5 + 3, 4 + 4	5	5/36
9	3 + 6, 6 + 3, 4 + 5, 5 + 4	4	4/36
10	4 + 6, 6 + 4, 5 + 5	3	3/36
11	5 + 6, 6 + 5	2	2/36
12	6 + 6	1	1/36
Total		36	1

7.3 Boltzmann Probability

I have mentioned the Boltzmann distribution several times already, so let us now consider a very simple macroscopic system comprising a gas of N inert particles at low pressure in a thermally insulated constant container of volume V .

I am going to look ahead and assume that you know about energy quantization and therefore assume that the particles have the very simple energy level diagram shown in Figure 7.4; each energy level corresponds to a single quantum state, and the allowed energies are

$$\varepsilon_n = nD \quad n = 1, 2, 3 \dots \quad (7.5)$$

I should mention that not all molecular systems show quantized energy; translational motion is not quantized either in quantum mechanics or in classical mechanics. It just makes my discussion simpler if we take the energy to be quantized.

Once again, I am assuming that you have come across such diagrams if you have studied elementary atomic structure theory, and you might have been expecting me to add electrons as up and down arrows (to denote the spin states). The meaning of these diagrams here is a little different; the diagram shows the allowed quantum states for each particle and I add particle labels to indicate their quantum states as in Figure 7.5. The left-hand arrangement of the three particles A, B and C has particles A and B in the lowest quantum state, none in the second and C in the third. I have chosen a different arrangement on the right-hand side of the figure.

I labelled the particles A, B and C to distinguish one from another, but there is more to this business of distinguishability than meets the eye. Keep this comment in mind as the topic will reappear later in the text.

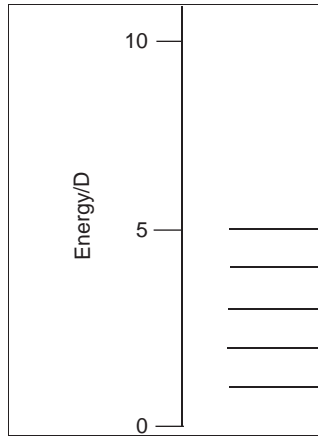


Figure 7.4 Simple energy level diagram

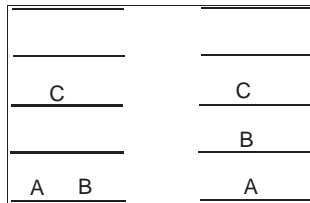


Figure 7.5 Quantum states

In a macroscopic sample of matter, each particle can have many microscopic quantum states to choose from and a key question is, for an equilibrium temperature, how do the particles distribute themselves amongst their allowed quantum states? Let me illustrate the problem with just three particles.

I am going to make another big assumption in order to keep the treatment simple. I assume that the particles form a perfect gas and whilst they may collide with each other and with the walls of the container, their energies do not change and indeed I can find the total energy of the system just by summing over the energies of the individual particles.

For the sake of argument, I will take the internal energy U as 9 D and the number of particles $N = 3$ as in Figure 7.5. We have to enumerate the ways in which three particles with these simple quantum state energies can achieve a total energy of 9 D.

The rows of Table 7.6 indicate every possible combination of integers that make up 9. To save space, I have only included one of each possible arrangements after the first; so for example, particle A could have 7 D with particles B and C having 1 D, particle B could have 7 D with particles A and C having 1 D, or particle C could have 7 D with particles A and B having 1 D. This is the origin of the integer 3 in the final column; it is the number of

Table 7.6 Arrangements and probabilities

A	B	C	$W = \text{no. of ways}$	Probability
7 D	1 D	1 D		
1	7	1	3	3/28
1	1	7		
6	2	1	6	6/28
5	2	2	3	3/28
5	3	1	6	6/28
4	4	1	3	3/28
4	3	2	6	6/28
3	3	3	1	1/28

ways we can achieve the total energy with each individual distinguishable particle having a certain amount of energy. For clarity, I have omitted the ‘D’ from all entries except the first.

The numbers in the fourth column are obviously related to factorials; if each particle has a different energy then there are $3!$ ways in which we can allocate the energies amongst the three particles. If two energies are the same then we have to divide the $3!$ by $2!$ and so on. If we were to make a spot check of the quantum state occupancy, we would find on average a situation where one particle had 7 D and the other two 1 D, for 3/28 of our measurements (on the other hand, since this is a quantum system, I had better be a bit more precise and say that if we were to prepare a very large number of such systems and measure the energy distribution in each copy, we would find 7 D, 1 D, 1 D in 3/28 of the systems).

Each arrangement is known as a *configuration*.

Consider now a macroscopic system comprising N identical but distinguishable particles, each of which has a similar quantum state diagram to that shown in Figure 7.5. The particles are noninteracting. We can fix the total energy U and the number of particles, but we cannot fix how the energy is distributed amongst the available energy levels (and it does not make much sense to ask about the energy distribution amongst N individual particles, given that N might be 10^{23}). The most sensible thing is to say that there are N_1 particles with energy ε_1 , N_2 with energy ε_2 and so on. Suppose also that the highest possible quantum state p has energy ε_p .

Consideration of our simple example above shows that the number of ways in which we can allocate the particles to the quantum states is

$$W = \frac{N!}{N_1!N_2! \dots N_p!} \quad (7.6)$$

The fact that a particular quantum state may have no particles (or, in other words, it has a zero probability) is accounted for by the fact that $0! = 1$.

It turns out that for large N there is one particular configuration that dominates, in which case the values of N_i are those that maximize W . At first sight they can be found by setting the gradient of W to zero, but N_i cannot vary freely; they are constrained by the fact that the number of particles and the total energy U are both constant

$$\begin{aligned} N &= N_1 + N_2 + \dots + N_p \\ U &= N_1\varepsilon_1 + N_2\varepsilon_2 + \dots + N_p\varepsilon_p \end{aligned} \quad (7.7)$$

A standard method for tackling *constrained variational* problems of this kind is *Lagrange's method of undetermined multipliers*, which is discussed in detail in all the advanced texts. I will just outline the solution without justification.

The first step in our derivation is to take logarithms of either side of expression (7.6):

$$\ln W = \ln N! - \sum_{i=1}^p \ln N_i! \quad (7.8)$$

For large values of n it is appropriate to make use of the Stirling formula for factorials

$$\ln n! = \frac{1}{2} \ln 2\pi + \left(n + \frac{1}{2}\right) \ln n - n + \frac{1}{12n} + O\left(\frac{1}{n^2}\right) + \dots \quad (7.9)$$

In practice all terms except $n \ln n$ are usually omitted and we have

$$\ln W = \ln N! - \sum_{i=1}^p (N_i \ln N_i! - N_i) \quad (7.10)$$

We therefore look for a maximum of $\ln W$. We know from elementary differential calculus that

$$d \ln W = \sum_{i=1}^p \frac{\partial \ln W}{\partial N_i} dN_i = 0 \quad (7.11)$$

at a stationary point. We cater for the constraints by differentiating Equations (7.7)

$$0 = dN_1 + dN_2 + \dots + dN_p$$

$$0 = dN_1 \varepsilon_1 + dN_2 \varepsilon_2 + \dots + dN_p \varepsilon_p$$

multiplying by arbitrary undetermined multipliers that I will call α and $-\beta$ and then adding:

$$d \ln W = \sum_{i=1}^p \frac{\partial \ln W}{\partial N_i} dN_i + \alpha \sum_{i=1}^p dN_i - \beta \sum_{i=1}^p \varepsilon_i dN_i \quad (7.12)$$

All the dN_i can now be treated as independent variables and at a stationary point

$$\frac{\partial \ln W}{\partial N_i} + \alpha - \beta \varepsilon_i = 0 \quad (7.13)$$

All that remains is to use the Stirling formula, and the final result turns out as

$$N_i = \exp(\alpha) \exp(-\beta \varepsilon_i) \quad (7.14)$$

Next we have to evaluate the undetermined multipliers α and β ; if we sum over all the allowed quantum states then we have

$$N = \sum_{i=1}^p N_i = \exp(\alpha) \sum_{i=1}^p \exp(-\beta \varepsilon_i)$$

which gives

$$\exp(\alpha) = \frac{N}{\sum_{i=1}^p \exp(-\beta \varepsilon_i)} \quad (7.15)$$

The value of β is not quite as straightforward. What we need to do is to use the expression to calculate a quantity such as the heat capacity of a perfect monatomic gas which can be compared with experiment. When this is done we find

$$N_i = \frac{N \exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{\sum_{i=1}^p \exp\left(-\frac{\varepsilon_i}{k_B T}\right)} \quad (7.16)$$

The formula is correct even when the quantum states are degenerate, as it refers to the individual quantum states. It also holds in an equivalent integral form when the energies form a continuum.

7.4 Safety in Numbers

I should now make a few points about Boltzmann's law. First of all, the basic result

$$\frac{N_2}{N_1} = \exp\left(-\frac{(\varepsilon_2 - \varepsilon_1)}{k_B T}\right)$$

can only be an approximation, for if I take a system of 10 000 two-state particles in equilibrium at a temperature $T = (\varepsilon_2 - \varepsilon_1)/k_B$ then we find

$$\frac{N_2}{N_1} = 0.36788$$

which corresponds to $N_2 = 7310.6$ and $N_1 = 2689.4$. These estimates cannot be true, since the numbers must be integers. The particles inevitably collide with one another and interact with the surroundings (technically referred to as the heat bath), and the picture that emerges is one of ceaseless activity. The number of particles in each state fluctuates from one moment to the next as energy flows into and out of the system via the heat bath. We therefore have to understand that the numbers appearing are averages, which I ought to write as

$$\frac{\langle N_2 \rangle}{\langle N_1 \rangle} = \exp\left(-\frac{(\varepsilon_2 - \varepsilon_1)}{k_B T}\right) \quad (7.17)$$

In order for Boltzmann's law to apply, the system must be in thermal equilibrium. The particles must be distinct and independent, which means that it must be possible in principle to distinguish one particle from another (by colouring, numbering or whatever means we want to use). With all of that in mind, I will drop the '<...>' notation unless it matters.

The average number of particles with energy ε fluctuates at random; the average population is given correctly by Boltzmann's law but the actual population fluctuates. Do these fluctuations matter when we want to predict macroscopic properties? I want to show you in this section that they are negligible.

First of all, let me establish that there is truly safety in numbers. Consider tossing a number of coins a certain number of times. It is common in the UK to call one side of the coin the 'head' side (H) and the other side the 'tail' side (T). In this experiment, I am going to throw the coin(s) in the air and record whether the coin lands H or T uppermost. I then repeat the experiment (say) 10 times. There is no particular reason for the coin to land either

H or T uppermost, and so the theoretical probability of either event is $\frac{1}{2}$. Table 7.7 shows a typical result.

Table 7.7 Heads and tails

Experiment	1	2	3	4	5	6	7	8	9	10
Outcome	T	T	T	H	H	T	H	T	T	T

If I plot out the results (Figure 7.6) we see that the result is of course unpredictable. We cannot say what each result will be.

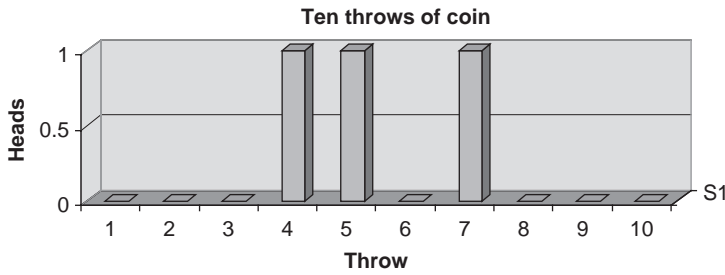


Figure 7.6 Single coin tossed 10 times

I now increase the number of coins to 10 and repeat the experiment again 10 times. That is to say, I take 10 coins and throw them into the air simultaneously. I record the result and then repeat the experiment a further 9 times. The results of my experiments are shown in Figure 7.7 (which of course I simulated with a random number generator).

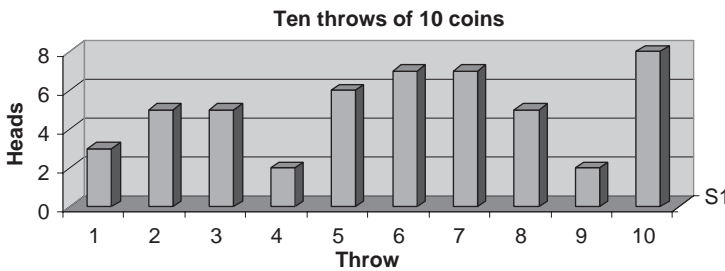


Figure 7.7 Ten coins tossed 10 times

We would expect an average of 5 heads per throw but we see a significant amount of scatter. I now repeat the ten experiments with 100 coins. The results are shown in Figure 7.8, and you can see that there is much less scatter about the mean. The point is that there is

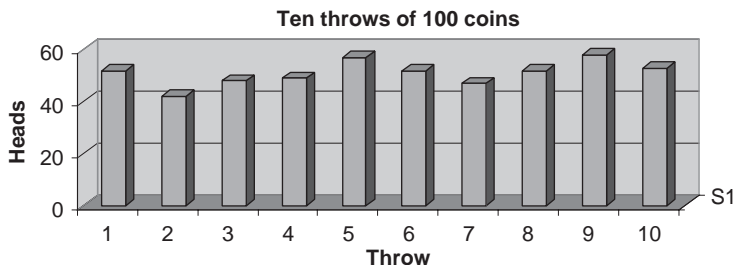


Figure 7.8 One hundred coins tossed 10 times

safety in numbers; if we use 100 coins we can be pretty sure of getting a result between 40 and 60 heads.

To return to the energy problem, because of the very large number of particles that go to make up a macroscopic sample, we can be very certain that fluctuations about the mean are negligible.

7.5 Partition Function

As discussed above, I can write Boltzmann's law as

$$\frac{N_i}{N} = \frac{\exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{\sum_{i=1}^p \exp\left(-\frac{\varepsilon_i}{k_B T}\right)} \quad (7.18)$$

The denominator in the Boltzmann expression turns out to be a useful quantity and so it is given a special name and symbol:

$$Q = \sum_{i=1}^p \exp\left(-\frac{\varepsilon_i}{k_B T}\right) \quad (7.19)$$

Q is the partition function. Some authors use Z rather than Q , and lowercase symbols z and q are often encountered when the emphasis is on individual particles rather than the macroscopic collection of particles.

These formulae are correct even when the energy levels are degenerate (i.e. each energy level corresponds to a number of quantum states), but of course the sums run over the quantum states and not the energy levels. In the above formulae, N_i is the number of particles in one of the quantum states and many authors choose to write the sums over the energy levels. For this purpose, they introduce a degeneracy factor g_i for each energy level, and so

$$\frac{N_i}{N} = \frac{g_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{\sum_{i=1}^n g_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right)} \quad (7.20)$$

It is understood that the sum runs over distinct energy levels and not the individual quantum states. I have used the symbol n in the summation to indicate the number of available energy levels.

7.6 Two-Level Quantum System

You should be familiar with the phenomenon of spin and know that charged particles with spin behave essentially as small bar magnets. According to the laws of (classical) electromagnetism, any particle with charge Q , mass M and angular momentum \mathbf{l} has a magnetic dipole moment, a vector quantity that is usually written \mathbf{p}_m :

$$\mathbf{p}_m = \frac{Q}{2M} \mathbf{l} \quad (7.21)$$

The energy of a magnetic dipole in an external field \mathbf{B} (Figure 7.9) is given by

$$U = -\mathbf{p}_m \mathbf{B} \quad (7.22)$$

and according to the rules of classical electromagnetism, all orientations of the dipole in the field are allowed but each orientation corresponds to a different energy.

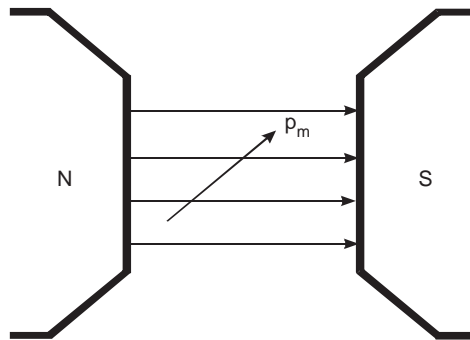


Figure 7.9 Magnetic dipole in a magnetic field

The quantum mechanical picture is quite different; only certain orientations are allowed, and the energy is quantized. The Stern–Gerlach experiment demonstrated that electrons have an ‘internal’ magnetic moment called *electron spin*, and only two orientations are possible. These are visualized as being (roughly) parallel to the field and (roughly) antiparallel to the field and the energies are given by

$$\varepsilon_{m_s} = m_s \frac{h}{2\pi} g_e \frac{e}{2m_e} B$$

where m_s is the spin quantum number, which takes values of $\pm 1/2$ for electron spin. The lower energy state ε_l is therefore that with $m_s = -1/2$, the higher energy state ε_u that corresponding to $m_s = +1/2$.

Consider then a sample of N such two-state particles. (It is hard to do experiments on free electrons, so we had better imagine a sample of hydrogen atoms; the only fly in the ointment is that the nuclei also have a spin of $1/2$ which I am going to ignore.) We keep the temperature constant by surrounding the sample with a heat bath and keep the magnetic field fixed. Figure 7.10 shows a few of the hydrogen atoms in the heat bath.

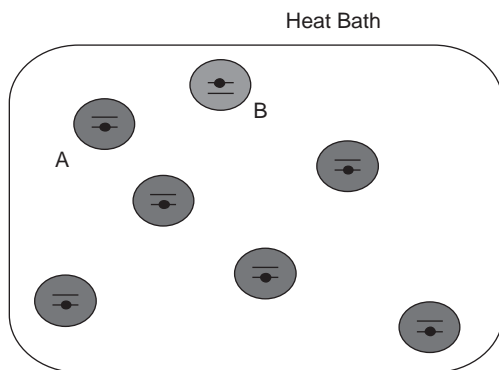


Figure 7.10 *The heat bath*

The diagram should be pretty obvious: atom A has its electron in the lower spin state, atom B has its electron in the higher spin state and so on. The heat bath keeps the temperature of our system constant.

Do not get confused with electronic energy levels, though. If N_u and N_l are the number of electrons in the upper and lower spin energy levels and N the total then

$$N = N_u + N_l \quad (7.23)$$

and the magnetic internal energy of the system is

$$U = N_u \varepsilon_u + N_l \varepsilon_l \quad (7.24)$$

We also know from the Boltzmann distribution law

$$N_i = \frac{N \exp\left(-\frac{\varepsilon_i}{k_B T}\right)}{Q} \quad (7.25)$$

In order to simplify the algebra, I will take the magnetic energy zero to be ε_l and the energy difference $\varepsilon_u - \varepsilon_l$ to be D . With this notation, explicit expressions for N_u and N_l are

$$N_l = \frac{N}{1 + \exp\left(-\frac{D}{k_B T}\right)}; \quad N_u = \frac{N \exp\left(-\frac{D}{k_B T}\right)}{1 + \exp\left(-\frac{D}{k_B T}\right)} \quad (7.26)$$

Figure 7.11 shows how the populations vary with temperature.

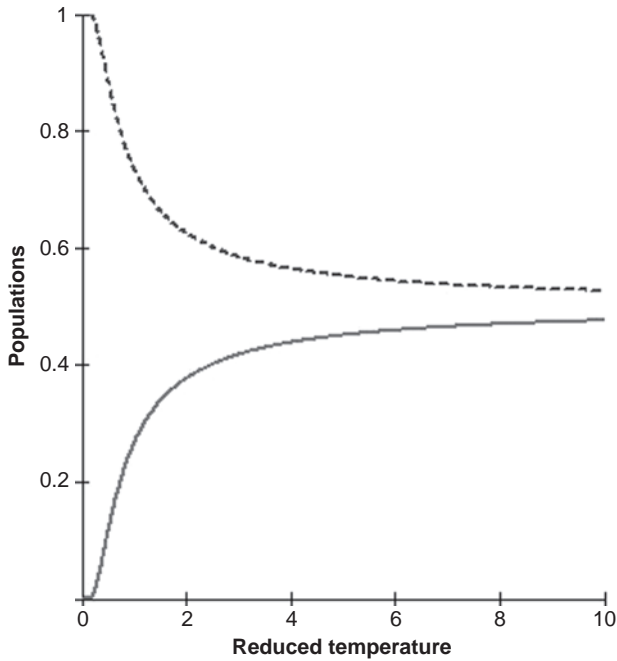


Figure 7.11 Populations as a function of temperature

7.7 Lindemann's Theory of Melting

We are going to catch our breath here, and draw together several strands already discussed. Figure 7.12 shows a two-dimensional cut through a simple solid. Think of a crystal where all the particles are the same, such as argon or possibly a metal. Metals actually get us into deep water (so to speak) because they have a sea of free electrons, which is why many conduct electricity, but do not worry about this problem for the minute. The important thing is that the particles do not appear to interact with each other and they are all the same. (Actually, if the particles did not interact they would not form a crystal, they would always exist as a perfect gas. That is a matter for discussion later in the text.) The bottom right-hand particle is constrained to sit in its equilibrium position by the other particles, and we will allow it to make simple harmonic excursions from its equilibrium position depending on the temperature. The higher the temperature, the higher are the excursions from the equilibrium position. So when does this solid melt?

The treatment I am going to give is a very simple one that relies on the *equipartition of energy principle*. You will probably know that a molecule comprising N atoms has $3N$ degrees of freedom, and that we usually think of these as 3 translational, 3 rotational (or 2 if the molecule is linear) and $3N - 6$ vibrational ($3N - 5$ if the molecule is linear) degrees of freedom. According to the laws of classical physics, each of these degrees of freedom has energy of $\frac{1}{2}k_B T$.

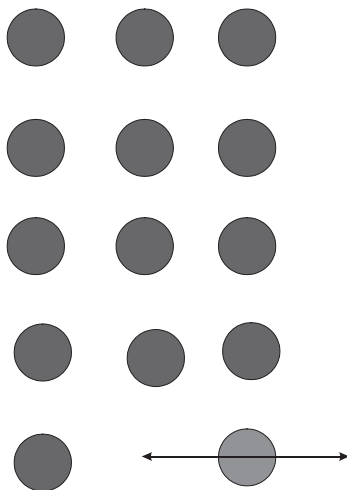


Figure 7.12 Part of a three-dimensional crystal

Now back to Figure 7.12. For the minute, assume that the particles only vibrate simple harmonically along the horizontal axis, which I will call the x axis. The energy of each particle is

$$\varepsilon_{\text{tot}} = \varepsilon_{\text{kinetic}} + \varepsilon_{\text{potential}} = \frac{1}{2}mv_x^2 + \frac{1}{2}k_s(x - x_e)^2 \quad (7.27)$$

According to classical physics, the average translational energy of any atom or molecule is always $\frac{1}{2}k_B T$ per degree of freedom. But, because the particle is vibrating simply harmonically, the average value of the potential energy is also equal to $\frac{1}{2}k_B T$ per degree of freedom. In other words

$$\frac{1}{2}k_s \langle (x - x_e)^2 \rangle = \frac{1}{2}k_B T \quad (7.28)$$

Taking the square root of both sides gives

$$\sqrt{\langle (x - x_e)^2 \rangle} = \frac{k_B T}{k_s} \quad (7.29)$$

The expression on the left-hand side is the root mean square of x and so

$$x_{\text{rms}} = \frac{k_B T}{k_s} \quad (7.30)$$

Most authors agree that a solid will melt when x_{rms} achieves a substantial percentage of the natural spacing d between the particles in a crystal. Many take 10%, and so

$$T_{\text{melt}} = \frac{k_s d^2}{100k_B} \quad (7.31)$$

Table 7.8 gives data for five metals, and the agreement between Lindemann's model and experiment is very reasonable for such a simple model.

Table 7.8 The Lindemann model for melting

Solid	d/pm	$k_s/\text{N m}^{-1}$	$T_{\text{melt}}(\text{theory})/\text{K}$	$T_{\text{melt}}(\text{expt})/\text{K}$
Copper	230	30	1163	1356
Gold	260	20	991	1336
Lead	310	5	352	600
Silver	260	21	1041	1234
Tungsten	250	100	4582	3650

7.8 Problem Set

A set of problems, together with solutions, is available on the associated website.

8

Introduction to Statistical Thermodynamics

I assume that you are familiar with the material presented in Chapter 7.

Suppose then that we have a macroscopic pure liquid sample, which might consist of 10^{23} particles, and we want to try and model some simple thermodynamic properties like the pressure p , the internal energy U or the Gibbs energy G . At room temperature, the individual particles making up the sample will be in motion so at first sight we ought to try and solve the equations of motion for these particles. In view of the large number of particles present, such an approach would be folly. Just to try and specify the initial positions and momenta of so many particles would not be possible, and in any case such a calculation would give too much information. You will know from Chapter 7 that the correct approach has to be a statistical one.

For the sake of argument, suppose that the container is a cube. I show a two-dimensional slice through the cube on the left-hand side of Figure 8.1, and I have exaggerated the size of the particles by a factor of approximately 10^{10} .

The pressure exerted by a gas on a container wall depends on the rate at which particles collide with the wall. It is not necessary, or even helpful, to know which particle underwent a particular collision. What we need to know about are the root mean square speed of the

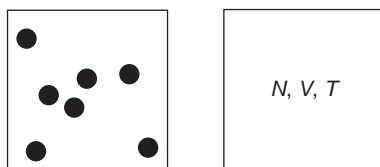


Figure 8.1 Box of particles

particles, their standard deviation about the mean, the temperature and so on. In statistical thermodynamics, we do not enquire about the behaviour of the individual particles that make up a macroscopic sample; we just enquire about their average properties.

If we were to measure the pressure exerted on the walls at time intervals t_1, t_2, \dots, t_n then we might record results $p(t_1), p(t_2), \dots, p(t_n)$. We could calculate a sample mean $\langle p \rangle$ and a sample standard deviation using these results:

$$\langle p \rangle = \frac{1}{n} \sum_{i=1}^n p(t_i)$$

$$\sigma_p = \sqrt{\frac{1}{n} \sum_{i=1}^n (p(t_i) - \langle p \rangle)^2} \quad (8.1)$$

We might expect that the greater the number of measurements, the closer the sample mean would be to the true mean, and the smaller the sample deviation would become.

8.1 The Ensemble

When we considered Figure 8.1, I was careful to draw your attention to the difference between particle properties and bulk properties. I also mentioned that classical thermodynamics is essentially particle-free; all that really matters to such a thermodynamicist are bulk properties such as the number of particles N , the temperature T and the volume of the container V . I have represented this information in the right-hand box in Figure 8.1.

Rather than worry about the time development of the particles in the left-hand box in Figure 8.1, what we do is to make a very large number of copies of the system on the right-hand side. We then calculate average values over this large number of replications and, according to the *ergodic theorem*, the average value we calculate is exactly the same as the time average we would calculate by studying the time evolution of the original system. The two are the same.

I am not suggesting that all the *cells* in the ensemble are exact replicas at the molecular level: all we do is to ensure that each cell has a certain number of thermodynamic properties that are the same. There is no mention of molecular properties at this stage of the game.

Figure 8.2 shows an ensemble of cells all with the same values of N , V and T .

This array of cells is said to form a *canonical ensemble*. There are three other important ensembles in the theory of statistical thermodynamics, and they are classified according to what is kept constant in each cell. Apart from the canonical ensemble, where N , V and T are kept constant, we have the following.

- The *microcanonical ensemble* where N , the total energy E and V are kept constant in each cell. In fact, this is a very simple ensemble because energy cannot flow from one cell to another.
- In an *isothermal-isobaric* ensemble, N , T and the pressure p are kept constant.

- Finally we have the *grand canonical ensemble*, where V , T and the chemical potential are kept constant. The grand canonical ensemble is a fascinating one because the number of particles is allowed to fluctuate.

N, V, T	N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T	N, V, T
N, V, T	N, V, T	N, V, T	N, V, T

Figure 8.2 Canonical ensemble

Supposing then that we consider a canonical ensemble of N^* cells (not to be confused with the N particles) comprising the original cell together with $N^* - 1$ replications. Energy may flow between the cells, but the total energy of the ensemble is constant. Suppose that the possible total energies of the N particles contained in each cell are E_1^* , E_2^* , and so on. We take an energy snapshot, and find a distribution of energies amongst the cells as follows:

N_1^* cells have energy E_1^*
 N_2^* cells have energy E_2^* , etc.

According to Boltzmann, E_i^* and N_i^* are related by

$$\frac{N_i^*}{N^*} = \frac{\exp\left(-\frac{E_i^*}{k_B T}\right)}{\sum_i \exp\left(-\frac{E_i^*}{k_B T}\right)} \quad (8.2)$$

Be careful to understand that the energies E_i^* are not molecular energies; they are the total energies of the collection of the N particles contained in each cell. Also note that N^* is the number of cells in the ensemble, and that the energies are taken as relative to a common arbitrary zero.

The denominator in the expression above plays an important role in our theory, and so it is given a special name and symbol:

$$Q = \sum_i \exp\left(-\frac{E_i^*}{k_B T}\right) \quad (8.3)$$

Q is (in this case) the *canonical partition function*, and it can be used to calculate the usual chemical thermodynamic functions as follows.

8.2 Internal Energy, U_{th}

The IUPAC recommended symbol for thermodynamic internal energy is U , but I have added a subscript ‘th’ for ‘thermodynamic’ so that there is no confusion with the total potential energy of a system U (often written Φ). Internal energy is obviously related to the ensemble energy average but we have to exercise caution. Chemical measurements only give changes in the internal energy, not absolute values. I will therefore write the internal energy as $U_{\text{th}} - U_0$, where U_0 is an arbitrary constant. For most purposes we can take U_0 to be zero.

We have, for the ensemble of N^* members

$$U_{\text{th}} - U_0 = \frac{\sum_i N_i^* E_i^*}{N^*} \quad (8.4)$$

and according to the ergodic theorem, this is equal to the time average of $U_{\text{th}} - U_0$ for any one cell. Using the Boltzmann expression we have

$$U_{\text{th}} - U_0 = \frac{\sum_i E_i^* \exp\left(-\frac{E_i^*}{k_{\text{B}}T}\right)}{Q} \quad (8.5)$$

I can tidy up Equation (8.5) by noting

$$\left(\frac{\partial Q}{\partial T}\right)_{v,N} = \frac{1}{k_{\text{B}}T^2} \sum_i E_i^* \exp\left(-\frac{E_i^*}{k_{\text{B}}T}\right) \quad (8.6)$$

and so on substitution

$$U_{\text{th}} - U_0 = \frac{k_{\text{B}}T^2}{Q} \left(\frac{\partial Q}{\partial T}\right)_v \quad (8.7)$$

8.3 Helmholtz Energy, A

From the definition of A we have the following:

$$\begin{aligned} A &= U - TS \\ S &= -\left(\frac{\partial A}{\partial T}\right)_{v,N} \\ A &= U + T\left(\frac{\partial A}{\partial T}\right)_{v,N} \end{aligned} \quad (8.8)$$

A little manipulation gives

$$\begin{aligned} \left(\frac{\partial}{\partial T} \left(\frac{A}{T}\right)\right)_{v,N} &= -k_{\text{B}} \left(\frac{\partial \ln Q}{\partial T}\right)_{v,N} \\ A - A_0 &= -k_{\text{B}}T \ln Q \end{aligned} \quad (8.9)$$

Again, the arbitrary constant A_0 can be taken as zero, since only changes in A are ever measured.

8.4 Entropy S

Finally, since

$$S = \frac{U - A}{T}$$

we have

$$S = k_B T \left(\frac{\partial \ln Q}{\partial T} \right)_{V,N} + k_B \ln Q \quad (8.10)$$

8.5 Equation of State and Pressure

The pressure is related to the Helmholtz energy by

$$p = - \left(\frac{\partial A}{\partial V} \right)_{T,N}$$

and so we find

$$p = k_B T \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \quad (8.11)$$

This equation is sometimes called the *equation of state*. The enthalpy and the Gibbs energy can be derived using similar arguments. They turn out to be

$$\begin{aligned} H &= k_B T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_{V,N} + k_B T V \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \\ G &= -k_B T \ln Q + k_B T V \left(\frac{\partial \ln Q}{\partial V} \right)_{T,N} \end{aligned} \quad (8.12)$$

8.6 Phase Space

Sophisticated methods such as those due to Hamilton and to Lagrange exist for the systematic treatment of problems in particle dynamics. Such techniques make use of *generalized coordinates* (written q_1, q_2, \dots, q_n) and *generalized momenta* (written p_1, p_2, \dots, p_n); in Hamilton's method we write the total energy as the *Hamiltonian* H . H is the sum of the kinetic energy and the potential energy, and it is a constant provided that the potentials are time independent. H has to be written in terms of the p 's and the q 's in a certain way, and systematic application of Hamilton's equations gives a set of differential equations for the system.

To fix our ideas, consider the particle of mass m undergoing simple harmonic motion as discussed in Chapter 4. In this one-dimensional problem I wrote the potential as

$$U = \frac{1}{2} k_s (R - R_e)^2$$

so that the total energy is

$$\varepsilon = \frac{1}{2}m \left(\frac{dR}{dt} \right)^2 + \frac{1}{2}k_s (R - R_e)^2$$

If I put $q = R - R_e$ then the momentum p is mdR/dt and I can write the Hamiltonian as

$$H = \frac{p^2}{2m} + \frac{k_s q^2}{2} \quad (8.13)$$

We say that the particle moves through *phase space* and in this example the trajectory through phase space is an ellipse (Figure 8.3), which can be easily seen by rewriting Equation (8.13) as

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} = 1$$

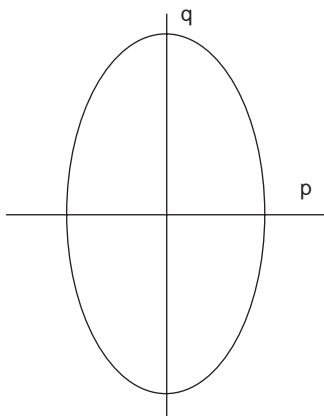


Figure 8.3 Phase space

Hamilton's equations of motion are

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}; \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \quad (8.14)$$

so for a general problem with N atoms, we have to solve $6N$ first-order differential equations rather than the $3N$ second-order differential equations we would get from straightforward application of Newton's law.

In the case of a one-particle three-dimensional system, the Hamiltonian will be a function of the three coordinates \mathbf{q} and the three momenta \mathbf{p} , and for a more general problem involving N particles the Hamiltonian will be a function of the $3N$ \mathbf{q} 's and the $3N$ \mathbf{p} 's. We say that the \mathbf{p} 's and the \mathbf{q} 's together determine a point in $6N$ -dimensional phase space, and this point is often denoted Γ .

8.7 Configurational Integral

Returning now to the canonical partition function, Equation (8.3)

$$Q = \sum_i \exp\left(-\frac{E_i^*}{k_B T}\right)$$

the first thing to note is that all points in phase space contribute to the sum, and the summation has to be replaced by an integral. For an ideal monatomic gas the expression becomes

$$Q = \frac{1}{N!} \frac{1}{h^{3N}} \int \int \exp\left(-\frac{E}{k_B T}\right) d\mathbf{p} d\mathbf{q} \quad (8.15)$$

The equation is often written with the Hamiltonian H replacing E , for the reasons discussed above.

The $N!$ term is needed in situations where the particles are completely indistinguishable from one another; for particles that can be distinguished there is no $N!$ term. The integrals have to be done over the spatial variables of all the N particles, and also the momentum variables of the N particles. The integral is therefore a $6N$ -dimensional one.

The energy (the Hamiltonian) is always expressible as a sum of kinetic and potential energies, and I have written the mass of each particle m :

$$E = \sum_{i=1}^N \frac{p_i^2}{2m} + \Phi(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N) \quad (8.16)$$

Kinetic energies depend on the momentum coordinates \mathbf{p} . All the potential energies we will meet depend on the spatial coordinates \mathbf{q} but not on the momenta and so the partition function can be factorized into a product of a kinetic part and a potential part:

$$Q = \frac{1}{N!} \frac{1}{h^{3N}} \int \exp\left(-\frac{1}{k_B T} \sum_{i=1}^N \frac{p_i^2}{2m}\right) d\mathbf{p} \int \exp\left(-\frac{\Phi}{k_B T}\right) d\mathbf{q} \quad (8.17)$$

The kinetic integral has to be done over the momentum coordinates of all N particles, and it can be seen to be a product of N identical three-dimensional integrals of the type

$$\int \exp\left(-\frac{1}{k_B T} \frac{p_1^2}{2m}\right) d\mathbf{p}_1$$

Each of these is a product of three identical standard integrals of the type

$$\int \exp\left(-\frac{1}{k_B T} \frac{p_x^2}{2m}\right) dp_x$$

and the final result is

$$Q = \frac{1}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2} \int \exp\left(-\frac{\Phi}{k_B T}\right) d\mathbf{q} \quad (8.18)$$

The $3N$ -dimensional integral over the position variables is often referred to as the *configurational integral*. For an ideal gas $\Phi = 0$ and so the configurational integral is V^N , where V is the volume of the container. Some authors include $N!$ in the definition of the configurational integral.

The canonical partition function for an ideal gas is therefore

$$Q = \frac{V^N}{N!} \left(\frac{2\pi mk_B T}{h^2} \right)^{3N/2} \quad (8.19)$$

The partition function for a real system is often written as the product of an ideal part and an excess part due to nonideal behaviour:

$$Q = Q^{\text{ideal}} Q^{\text{excess}} \quad \text{where}$$

$$Q^{\text{excess}} = \frac{1}{V^N} \int \exp\left(-\frac{\Phi}{k_B T}\right) d\mathbf{q} \quad (8.20)$$

The point of doing this is that thermodynamic properties such as A are often measured experimentally as an ideal and an excess part:

$$A = A^{\text{ideal}} + A^{\text{excess}}$$

The ideal part can be related to Q^{ideal} and the excess part to Q^{excess} .

8.8 Virial of Clausius

Let me focus attention on one particular particle i moving in the box shown in Figure 8.1. As this particle moves it will be subject to some varying force \mathbf{F}_i and

$$\mathbf{F}_i = m \frac{d\mathbf{v}_i}{dt} \quad (8.21)$$

Taking the scalar product of both side of this equation with \mathbf{r}_i we get

$$\mathbf{r}_i \cdot \mathbf{F}_i = m \mathbf{r}_i \cdot \left(\frac{d\mathbf{v}_i}{dt} \right) \quad (8.22)$$

Consider now the vector identity

$$\frac{d}{dt} (\mathbf{r}_i \cdot \mathbf{v}_i) = \mathbf{r}_i \cdot \frac{d\mathbf{v}_i}{dt} + \frac{d\mathbf{r}_i}{dt} \cdot \mathbf{v}_i \quad (8.23)$$

which can also be written

$$\frac{d}{dt} (\mathbf{r}_i \cdot \mathbf{v}_i) = \mathbf{r}_i \cdot \frac{d\mathbf{v}_i}{dt} + v_i^2 \quad (8.24)$$

On comparison of Equations (8.22) and (8.24), we have

$$\mathbf{r}_i \cdot \mathbf{F}_i = m \left(\frac{d}{dt} (\mathbf{r}_i \cdot \mathbf{v}_i) - v_i^2 \right) \quad (8.25)$$

or

$$-\frac{1}{2} \mathbf{r}_i \cdot \mathbf{F}_i = -\frac{1}{2} m \frac{d}{dt} \mathbf{r}_i \cdot \mathbf{v}_i + \frac{1}{2} m v_i^2 \quad (8.26)$$

The next step is to sum corresponding terms on both sides of the equation for each particle in the box. For N particles each of mass m , this gives

$$-\frac{1}{2} \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i = -\frac{1}{2} m \frac{d}{dt} \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{v}_i + \frac{1}{2} m \sum_{i=1}^N v_i^2 \quad (8.27)$$

Finally, we take a time average over all the particles in the box, which is assumed to be in an equilibrium state:

$$-\frac{1}{2} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle = -\frac{m}{2} \frac{d}{dt} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{v}_i \right\rangle + \frac{1}{2} m \left\langle \sum_{i=1}^N v_i^2 \right\rangle \quad (8.28)$$

The second term on the right-hand side is obviously the mean kinetic energy of all the particles in the box. This must be $\frac{3}{2} N k_B T$, according to the equipartition of energy principle.

Whatever the value of the first average quantity in brackets on the right-hand side it cannot vary with time because we are dealing with an equilibrium state and so the first time derivative must vanish:

$$-\frac{m}{2} \frac{d}{dt} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{v}_i \right\rangle = 0$$

Thus we have

$$-\frac{1}{2} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle = \frac{1}{2} m \left\langle \sum_{i=1}^N v_i^2 \right\rangle \quad (8.29)$$

The summation term on the left hand side

$$-\frac{1}{2} \left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle$$

involving the forces and coordinates is often referred to as the *virial of Clausius*. According to the equipartition of energy principle, it is equal to $\frac{3}{2} N k_B T$.

9

Monte Carlo Simulations

When discussing QSAR in Chapter 6, I mentioned that the molecular volume was thought to be a useful index and explained how it could be calculated within the framework of a MM study.

To recap, Figure 9.1 shows a diatomic molecule, where the atomic radii are R_A and R_B . The problem is to find the volume of the molecule AB. The volume of atom A is $\frac{4}{3}\pi R_A^3$ and the volume of atom B is $\frac{4}{3}\pi R_B^3$. The volume of AB is of course less than or equal to the sum of these atomic volumes depending on the extent to which the atoms overlap. We have to be careful not to count the shared volume X twice in the calculation.

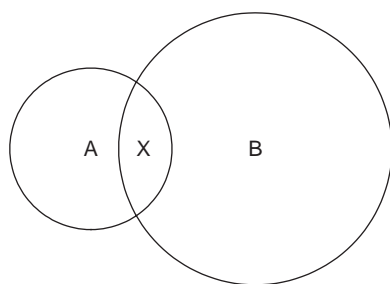


Figure 9.1 A diatomic molecule AB

To expand on the discussion in Chapter 6, I suggest that the total can be calculated using the following algorithm.

1. Start with A, for which the volume is $\frac{4}{3}\pi R_A^3$.
2. Surround B with a regular grid of side $2R_B$, and examine which grid points lie inside B but exclude those that lie in region X. To do this, test if they are within distance R_A of nucleus A.
3. Also exclude points which lie outside atom B.
4. A simple ratio of points then gives the contribution to the molecular volume from atom B.

The process can be easily extended to any polyatomic system. This is a pretty simple problem; suppose there are N atoms and we take a grid of m one-dimensional points for each atom. This means that we will have to consider $(N - 1) \times m^3$ points in space. If $N = 101$ and $m = 50$ the number is 12.5×10^6 ; a large number of thankfully very simple calculations.

Rather than take a regularly spaced cubic array of points around atom B, we can surround atom B by an (imaginary) cube of side $2R_B$ and choose points at random inside this cube. Here is the revised algorithm.

1. Choose a point at random and examine whether it lies inside atom A, in which case we accept it.
2. If the point lies inside atom B and not in region X, we accept it.
3. Increment the number of points and return to 1.

A simple proportionality between the number of successful points and the total number tried gives the required volume.

This calculation is an example of the *Monte Carlo* technique (denoted MC in this text). MC is a generic term applied to calculations that involve the use of random numbers for sampling; it became widely used towards the end of the Second World War by physicists trying to study the diffusion of neutrons in fissionable material.

In fact, the MC method is usually attributed to a French eighteenth-century naturalist named Búffon, who discovered that if a needle of length l were dropped at random onto a floor consisting of wooden planks with uniform separation $d > l$, then the probability of the needle crossing the join between two planks is $2l/\pi d$. Búffon's subsequent experiments enabled him to make an accurate estimation of π .

In the case of the molecular volume calculation, the problem scales as the number of atoms. That is good news. We can easily identify worse problems; for example, the trapezium rule enables us to approximate a definite integral

$$\int_a^b f(x) \, dx = \frac{1}{2}d (f(a) + 2f(a+d) + \dots + f(b))$$

as the sum of trapezia shown in Figure 9.2. If we choose n points along the x axis then we have to evaluate the function $(n + 1)$ times in order to evaluate the sum.

For the corresponding two-dimensional integral

$$\int_a^b \int_c^d f(x, y) \, dx dy$$

we would have to make $(n+1)^2$ function evaluations and the number of function evaluations is seen to rise as the power of the dimension of the integral. That is bad news.

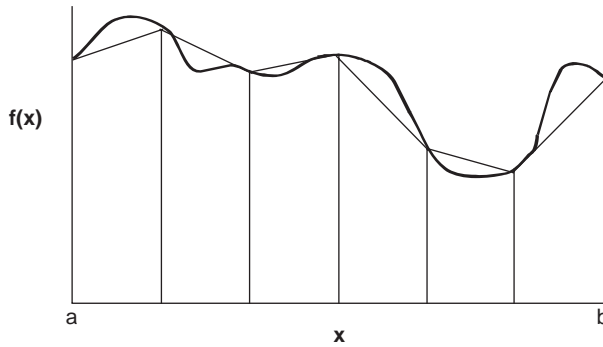


Figure 9.2 Trapezium rule for a definite integral

9.1 An Early Paper

Metropolis and Ulam’s (1949) keynote paper has an interesting abstract.

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

and there is much to be gained by studying the opening sentences. . .

Already in the nineteenth century a sharp distinction began to appear between two different mathematical methods for treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of sets of particles.

Straightforward thermodynamic quantities such as the pressure, the internal energy and the Gibbs energy turn out to be impossible to calculate directly for a macroscopic system, simply because of the large number of particles involved. In fact, it is not even sensible to contemplate recording an initial starting point for 10^{23} particles, let alone devising methods for solving the equations of motion.

Suppose for the sake of argument that we have a system of N simple atomic particles (such as argon atoms) and we are interested in the total mutual potential energy (conventionally written Φ rather than U). If the pair potential is $U_{ij}(R)$ and the potentials are pair wise additive, we have

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^N U_{ij}(R_{ij})$$

If I denote the position vectors of the N particles measured at time t by $\mathbf{R}_A(t), \mathbf{R}_B(t), \dots, \mathbf{R}_N(t)$, then these position vectors will depend on time and the instantaneous value of Φ will depend on the values of the variables at that time. If we make enough measurements

$\Phi(t_1), \Phi(t_2), \dots, \Phi(t_n)$ at times t_1, t_2, \dots, t_n then the average of these measurements will approach the mean value of Φ :

$$\langle \Phi \rangle = \frac{1}{n} \sum_{i=1}^n \Phi(t_i)$$

Other properties such as the self-diffusion depend on fluctuations about mean values.

In statistical thermodynamics, rather than calculating such a time average we consider the average over a large number of replications of the system (an ensemble). The *ergodic theorem* tells us that the two are the same.

In Chapter 8 I considered the special case of a canonical ensemble, which consists of a large number of replications (or cells), each of which are identical in the sense that the number of particles N , the volume V and the temperature T are the same. The cells are not identical at the atomic level, all that matters is N , V and T . Energy can flow from one cell to another, but the total energy of the ensemble is constant.

Under these conditions, the chance of finding a cell with energy E is proportional to the Boltzmann factor $\exp(-E/k_B T)$. Mean values of quantities such as the energy $\langle E \rangle$ can be found by averaging over the replications

$$\langle E \rangle = \frac{\sum E_i \exp\left(-\frac{E_i}{k_B T}\right)}{\sum \exp\left(-\frac{E_i}{k_B T}\right)} \quad (9.1)$$

or for an infinite number of replications we would replace the sums by integrals

$$\langle E \rangle = \frac{\int E \exp\left(-\frac{E}{k_B T}\right) dE}{\int \exp\left(-\frac{E}{k_B T}\right) dE}$$

Provided the potential energies do not depend on momenta, it is possible to simplify such integrals and I explained earlier the importance of the configurational integral, which depends only on exponential terms such as

$$\int \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q} \quad (9.2)$$

involving the total mutual potential energy Φ . I also showed how it could be related to the so-called excess thermodynamic functions.

I have written the configurational integral in a simplified way; if there are N particles, it is actually a $3N$ -dimensional integral over the positions of the N particles. We might want to take $N = 10\,000$ particles for a very simple simulation but the considerations above show that it is impractical to carry out such a multidimensional integral by usual techniques of numerical analysis such as the trapezoid rule. Instead we resort to the Monte Carlo method and generate a representative number of points in conformation space. For each point we calculate a value of Φ . The integral is approximated by a finite sum.

The crucial point is that there is no momentum contribution to the partition function in MC. The momentum contribution is essentially what we would expect for a system of noninteracting particles.

9.2 The First ‘Chemical’ Monte Carlo Simulation

The first serious Monte Carlo calculation appears to be that of Metropolis *et al.* (1953) (although the method was proposed independently by J.E. Meyer and by S. Ulam). In fact, the Metropolis *et al.* paper marked the birth of computer simulation as a statistical mechanical technique, and it preceded the molecular dynamics studies to be discussed in Chapter 10. I cannot do better than repeat the words of the authors (i.e. quote their abstract):

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC, and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

The rigid-sphere potential shown in Figure 9.3 is the simplest one imaginable; the spheres cannot overlap, but their mutual potential energy is zero everywhere else.

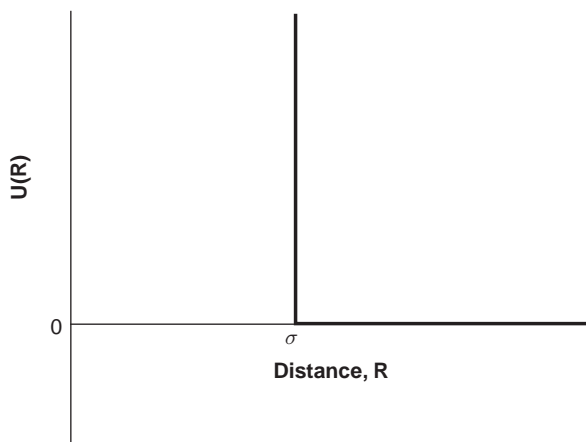


Figure 9.3 Rigid sphere potential

The authors took a two-dimensional square array of 224 hard spheres (i.e. hard *discs*) each with fixed radius σ . Here is the algorithm.

1. A starting configuration is chosen. Metropolis placed the 224 discs in a triangular lattice of 14 particles per row by 16 particles per column, alternate rows being displaced by 0.5σ .
2. Each disc is moved in succession in the XY plane according to the formula

$$\begin{aligned} X &\rightarrow X + \alpha\xi_1 \\ Y &\rightarrow Y + \alpha\xi_2 \end{aligned}$$

where α is some (arbitrary) maximum allowed distance and ξ_i are random numbers between -1 and $+1$. This means that after we move a particle, it is equally likely to be anywhere within a square of side 2σ centred about its original position.

3. At the end of the moves, the discs are examined for overlap. In the case that the move did indeed put one disc on top of another disc, the particles are returned to their original positions. This is a primitive form of the technique known as *importance sampling* that is now a key feature of modern MC simulations.
4. We then calculate any useful data for averaging. In this model, $\Phi = 0$ for all allowed conformations, but we might be interested for example in the disc–disc separations in order to investigate the radial distribution functions to be discussed in later chapters.
5. After a large number of allowed configurations have been examined (and the calculation has lost all knowledge of its starting configuration), we stop and do the final averaging. The choice of α is actually important here. If α is large then discs invariably overlap and if α is small then the conformations do not change much and in either case we have to perform a very large number of iterations to get a meaningful average. It is normally chosen so that 50% of the trial moves are accepted.
6. Finally the acid test: compare with experiment. It is rather difficult to think of a real-life system that could be realistically represented by the rigid-sphere model but the authors did their best; they used the virial theorem of Clausius to derive an equation of state for the rigid discs

$$pA = Nk_B T \left(1 + \frac{\pi\sigma^2 \langle n \rangle}{2} \right)$$

where $\langle n \rangle$ is the average density of other particles at the surface of a particle and A the area. They then calculated $\langle n \rangle$ from the MC simulation and compared with the equation of state.

9.3 Importance Sampling

In the case of a more complicated problem (even a modest Lennard-Jones potential), a more sophisticated treatment of importance sampling is needed for the following reason. The average potential energy can indeed be obtained from the configuration integral

$$\langle \Phi \rangle = \frac{\int \Phi(\mathbf{q}) \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q}}{\int \exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right) d\mathbf{q}} \quad (9.3)$$

and each integral can indeed be approximated by a finite sum of M terms in MC

$$\langle \Phi \rangle \approx \frac{\sum_{i=1}^M \Phi_i(\mathbf{q}) \exp\left(-\frac{\Phi_i(\mathbf{q})}{k_B T}\right)}{\sum_{i=1}^M \exp\left(-\frac{\Phi_i(\mathbf{q})}{k_B T}\right)} \quad (9.4)$$

If we calculate the mutual potential energy $\Phi_i(\mathbf{q})$ for a random array of particles, then the array will have a Boltzmann weight of

$$\exp\left(-\frac{\Phi_i(\mathbf{q})}{k_B T}\right)$$

and if some particles are close together Φ_i will be large and the Boltzmann factor small. Most of phase space corresponds to nonphysical configurations with very high energies and only for a very small proportion of phase space does the Boltzmann factor have an appreciable value. This means that most of the computing time will be spent on configurations that give negligible contributions to properties of interest.

Metropolis *et al.* (1953) therefore proposed a modified MC scheme where, instead of choosing configurations randomly then weighting them with $\exp(-\Phi/k_B T)$, configurations were chosen with a probability distribution of $\exp(-\Phi/k_B T)$ and then weighted equally. This adds a branch to the algorithm as follows.

1. Before making a move, we calculate the energy change of the system caused by the move. If the energy change is negative, the move is allowed.
2. If the energy change is positive, we allow the move but with a probability of $\exp(-\Phi/k_B T)$; to do this, we generate a random number ξ_3 between 0 and 1, and if

$$\xi_3 < \exp\left(-\frac{\Phi}{k_B T}\right)$$

we allow the move to take place.

3. Otherwise we leave the particle in its old position and move on to the next particle.
4. Whether the move is allowed or not, we consider that we have generated a new configuration for the purposes of taking averages.

This summarizes the *Metropolis Monte Carlo* (MMC) technique. The crucial feature of MMC is that it biases the generation of configurations towards those that make significant contributions to the configuration integral. It generates configurations with a probability

$$\exp\left(-\frac{\Phi(\mathbf{q})}{k_B T}\right)$$

and then counts each of them equally. By contrast the simple MC integration method generates states with equal probability and then assigns them this weight.

9.4 Periodic Box

Figure 9.4 shows a box of argon atoms, and examination reveals two problems.

1. Atoms near the edges of the box will experience quite different resultant forces from those atoms near the centre of the box.
2. Second, a MC move may well place one of the atoms outside the box and this will alter the density of the system under study.

The periodic box concept, illustrated in Figure 9.5, gives a solution to these problems. We appeal to the ensemble concept of statistical thermodynamics, and surround our system

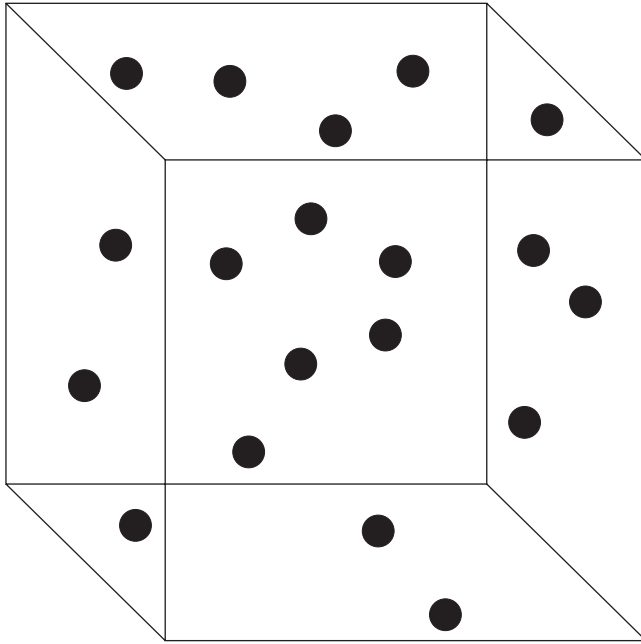


Figure 9.4 Box of argon atoms

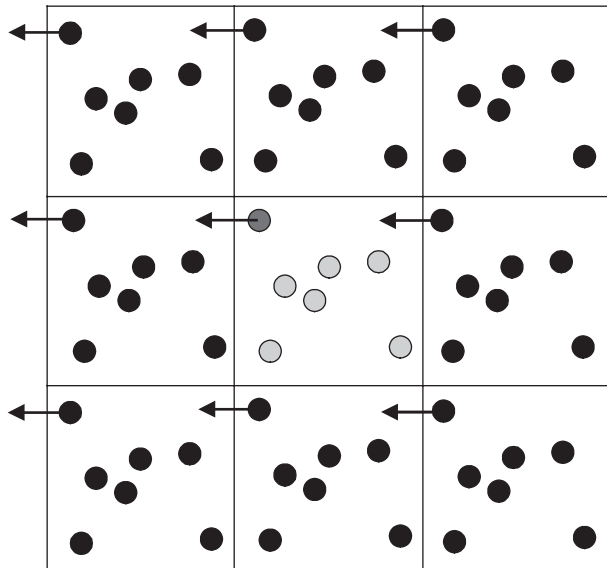


Figure 9.5 Small part of the molecular ensemble

with a large number of *identical* copies. In this case the boxes are truly identical at the atomic level rather than in the usual thermodynamic sense of having N , V and T in common.

Figure 9.5 shows a two-dimensional slice through a small portion of the system (the central box where the atoms are shown grey) and the copies (where the atoms are shown black). Each copy is identical at the atomic level, and each corresponding atom undergoes the same MC change as its image in every other copy.

If a MC move causes the dark grey atom (top left, central cell) to leave the central cell, its image enters from an adjoining copy, shown by the vector displacements in the figure. This keeps the density constant.

There are no effects due to the walls because each atom in the central cell is under the influence of every other atom in the central cell and in all other cells.

9.5 Cut-Offs

This means we have to be able to calculate the mutual potential energy of (for example) the dark grey atom with all the others shown in Figure 9.5. This would give an infinite sum.

In the case of a short-range potential such as the Lennard-Jones, we simply decide on a cut-off distance R_C beyond which the pair potential will be negligible; this defines a sphere of radius R_C centred on each atom. In order to treat the dark grey atom, we just include contributions from all other atoms in the sphere as shown in two dimensions in Figure 9.6.

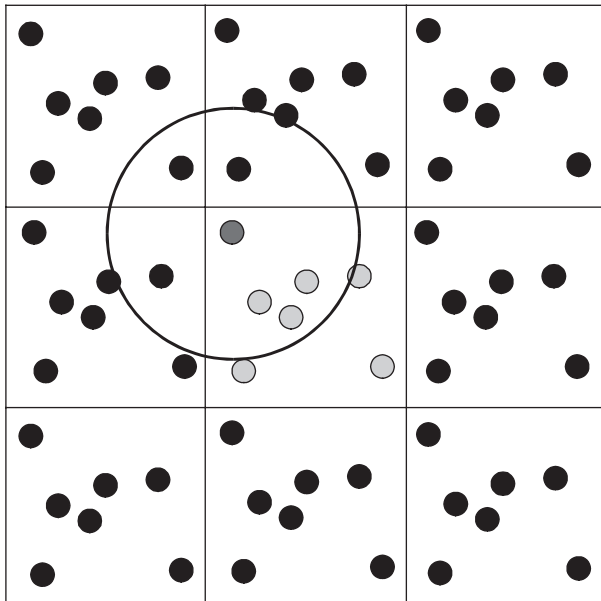


Figure 9.6 *The cut-off distance*

Truncation of the intermolecular potential at a cut-off distance actually introduces two technical difficulties. Firstly, the pair potential has a discontinuity at the cut-off distance R_C , and, secondly, whenever a pair of particles A and B has separation greater than R_C the total energy is not conserved. The first problem is solved by shifting the potential function by an amount $U(R_C)$; that is, we take

$$U^s(R_{AB}) = \begin{cases} U(R_{AB}) - U(R_C) & \text{if } R_{AB} \leq R_C \\ 0 & \text{if } R_{AB} > R_C \end{cases} \quad (9.5)$$

The second problem can be solved by adding a small linear term to the potential, chosen so that its derivative is zero at the cut-off distance

$$U^{\text{lin}}(R_{AB}) = \begin{cases} U(R_{AB}) - U(R_C) - \left(\frac{dU}{dR_{AB}} \right)_{R_C} (R_{AB} - R_C) & \text{if } R_{AB} \leq R_C \\ 0 & \text{if } R_{AB} > R_C \end{cases}$$

9.6 MC Simulation of Rigid Molecules

For rigid nonspherical molecules, it is necessary to vary their orientation as well as the position in space; usually each molecule is translated and rotated by random amounts once per cycle. The simplest approach is to rotate about each Cartesian axis in turn by an angle chosen at random but subject to a suitable maximum allowed variation. For example, if the position vector of an atom is the column vector \mathbf{R} , the position vector after rotation by α about the x axis is given by

$$\mathbf{R}' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{pmatrix} \mathbf{R} \quad (9.6)$$

The three *Euler angles* ϕ , θ and ψ are often used when discussing rotational motion. First of all we rotate by ϕ about the Cartesian z axis. This changes the x and y axes to x' and y' as shown in Figure 9.7. Then we rotate by θ about the new x axis which changes

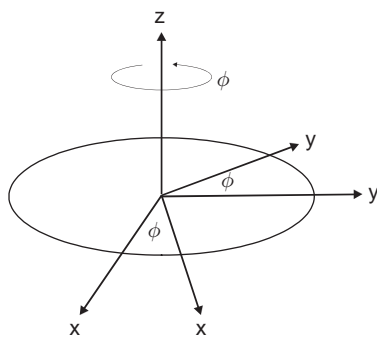


Figure 9.7 First Euler angle

the y and z axes. Finally we rotate by ψ about the new z axis. The relationship between a position vector before and after the three rotations is

$$\mathbf{R}' = \begin{pmatrix} \cos \phi \cos \psi - \sin \phi \cos \theta \sin \psi & \sin \phi \cos \psi + \cos \phi \cos \theta \sin \psi & \sin \theta \sin \psi \\ -\cos \phi \sin \psi - \sin \phi \cos \theta \cos \psi & -\sin \phi \sin \psi + \cos \phi \cos \theta \cos \psi & \sin \theta \cos \psi \\ \sin \phi \sin \theta & -\cos \phi \sin \theta & \cos \theta \end{pmatrix} \mathbf{R} \quad (9.7)$$

There are two technical problems with the use of Euler angles. First, sampling the angles at random does not give a uniform distribution; it is necessary to sample ψ , $\cos \theta$ and ψ . Second, there are a total of six trigonometric function evaluations per rotation.

A preferred alternative makes use of *quaternions*, which are four-dimensional unit vectors. A quaternion \mathbf{q} is usually written in terms of the scalar quantities q_0, q_1, q_2, q_3 as

$$\mathbf{q} = (q_0 \quad q_1 \quad q_2 \quad q_3)^T$$

and the components satisfy

$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$$

They can be related to the Euler angles as

$$\begin{aligned} q_0 &= \cos \frac{1}{2}\theta \cos \frac{1}{2}(\phi + \psi) \\ q_1 &= \sin \frac{1}{2}\theta \cos \frac{1}{2}(\phi - \psi) \\ q_2 &= \sin \frac{1}{2}\theta \sin \frac{1}{2}(\phi - \psi) \\ q_3 &= \cos \frac{1}{2}\theta \sin \frac{1}{2}(\phi + \psi) \end{aligned} \quad (9.8)$$

The rotation matrix can be written as

$$\mathbf{R}' = \begin{pmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{pmatrix} \mathbf{R} \quad (9.9)$$

and all that is necessary is to generate four suitable random numbers.

9.7 Flexible Molecules

Monte Carlo simulations of flexible molecules are difficult to perform unless the system is small or some of the internal degrees of freedom are kept fixed. The simplest way to generate a new configuration is to perform random changes to the Cartesian coordinates of individual atoms in the molecule but it is a common experience that very small changes are needed in order to produce an acceptable Boltzmann factor (in the MMC sense).

References

- Metropolis, N. and Ulam, S. (1949) *J. Am. Stat. Assoc.*, **44**, 335.
 Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. *et al.* (1953) *J. Chem. Phys.*, **21**, 1087.

10

Molecular Dynamics

Molecular mechanics (MM) tends these days to be concerned only with prediction of local minima on molecular potential energy surfaces. QSAR properties are often calculated in order to assist high-volume screening studies in pharmaceuticals applications. Should we want to study the motions of the molecule, all that would be needed would be to investigate the normal modes of vibration (which can be obtained from the Hessian). MM does not take account of zero-point vibrations and the calculations refer to a molecule at 0 K, when it is completely at rest. Workers in the modelling field often refer to MM as *energy minimization*.

As we saw in Chapter 9, Monte Carlo modelling tends to be concerned with thermodynamic properties. It is now time to consider molecular dynamics which gives a direct way to study time evolution.

I have stressed in previous chapters the intermolecular potential energy, which I chose to write as Φ in Chapter 9. Assuming pair wise additivity, Φ can be found by summing over all distinct pairs of particles

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^N U_{ij} \quad (10.1)$$

If the assumption of pair wise additivity is not valid, then one way forward is to include all possible triples and so on

$$\Phi = \sum_{i=1}^{N-1} \sum_{j=i+1}^N U_{ij} + \sum_{i=1}^{N-2} \sum_{j=i+1}^{N-1} \sum_{k=j+1}^N U_{ijk} + \dots \quad (10.2)$$

In this text, we will generally be concerned with situations where the potentials (and so the forces) are pair wise additive. If we focus on one particular particle A, then the mutual potential energy of A with all the other particles U_A is found from

$$U_A = \sum_{j \neq A} U_{Aj}$$

and we can find the force on particle A, \mathbf{F}_A , by differentiating U_A with respect to the coordinates of particle A:

$$\mathbf{F}_A = -\text{grad } U_A$$

For example, if we consider a pair of Lennard-Jones particles A and B where

$$U_{AB} = 4\varepsilon \left(\left(\frac{\sigma}{R_{AB}} \right)^{12} - \left(\frac{\sigma}{R_{AB}} \right)^6 \right)$$

we note that the potential only depends on the distance between the particles. The expression for $\text{grad } U$ is particularly simple in this case so that

$$\mathbf{F}_A = - \frac{\partial U}{\partial R_A} \frac{\mathbf{R}_{BA}}{R_{BA}}$$

which gives

$$\mathbf{F}_A = 24\varepsilon \left(2 \left(\frac{\sigma}{R_{AB}} \right)^{12} - \left(\frac{\sigma}{R_{AB}} \right)^6 \right) \frac{\mathbf{R}_{AB}}{R_{AB}^2}$$

Newton's second law connects force and acceleration by

$$\mathbf{F}_A = m_A \frac{d^2 \mathbf{R}_A}{dt^2}$$

and in principle we could study the time development of a system by solving this second-order differential equation, one such equation for each of the particles in our system. Calculating the trajectories of N particles therefore appears to involve the solution of a set of $3N$ second-order ordinary differential equations. Alternatively, we could use an advanced method such as Hamilton's to solve $6N$ first-order differential equations. For any set of N particles it is always possible to find three coordinates that correspond to translation of the centre of mass of the system, and, if the particles have 'shape', three coordinates that correspond to rotations about three axes that pass through the centre of mass.

Most of the early molecular dynamics studies were directed at the problem of liquid structure, so that is where we will begin our discussion.

10.1 Radial Distribution Function

Of the three states of matter, gases are the easiest to model because the constituent particles are so far apart on average that we can ignore intermolecular interactions apart from during their brief collisions. This is why researchers were able to bring the kinetic theory of gases to such an advanced stage by the end of the nineteenth century (and before the existence of a satisfactory theory of molecular structure).

The atoms or molecules in a crystalline solid are arranged in a regular order, and for this reason we usually start a discussion of the solid state from the properties of regular solids. Once such patterns were truly understood at the beginning of the twentieth century, the theory of the solid state made rapid progress.

Liquids are much harder to model and to study experimentally than solids and gases; elementary textbooks usually state that liquids show neither complete order nor complete disorder. The basis of this remark concerns a property called the *radial distribution function* $g(r)$. Consider Figure 10.1, which is a snapshot of the particles in a simple atomic liquid.

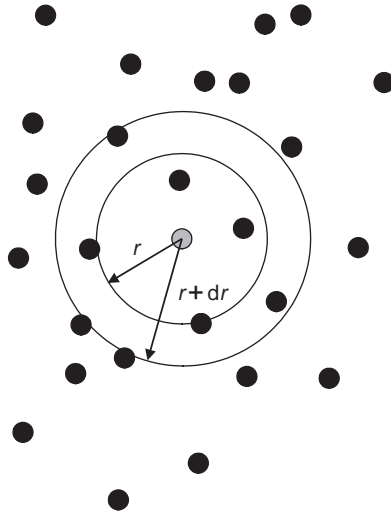


Figure 10.1 Radial distribution function

We take a typical atom (the grey one, designated i) and draw two spheres of radii r and $r + dr$. We then count the number of atoms whose centres lie between these two spheres, and repeat the process for a large number N of atoms. If the result for atom i is $g_i(r)dr$ then the radial distribution function is defined as

$$g(r) dr = \frac{1}{N} \sum_{i=1}^N g_i(r) dr \quad (10.3)$$

This process then has to be repeated for many complete shells over the range of values of r thought to be significant.

In the case of an ideal gas, we would expect to find the number of particles to be proportional to the volume enclosed by the two spheres, which is $4\pi r^2 dr$. This gives $g(r) = 4\pi r^2$, a simple quadratic curve.

Consider now the simple cubic solid shown in Figure 10.2 whose nearest neighbour distance is a . Each atom is surrounded by 6 nearest neighbours at a distance a , 12 at a distance $\sqrt{2}a$, 8 next-next nearest neighbours at a distance $\sqrt{3}a$, 6 at a further distance $2a$ and so on. We would therefore expect to find a radial distribution function similar to the one shown in Figure 10.3. The height of each peak is proportional to the number of atoms distant r from any given atom.

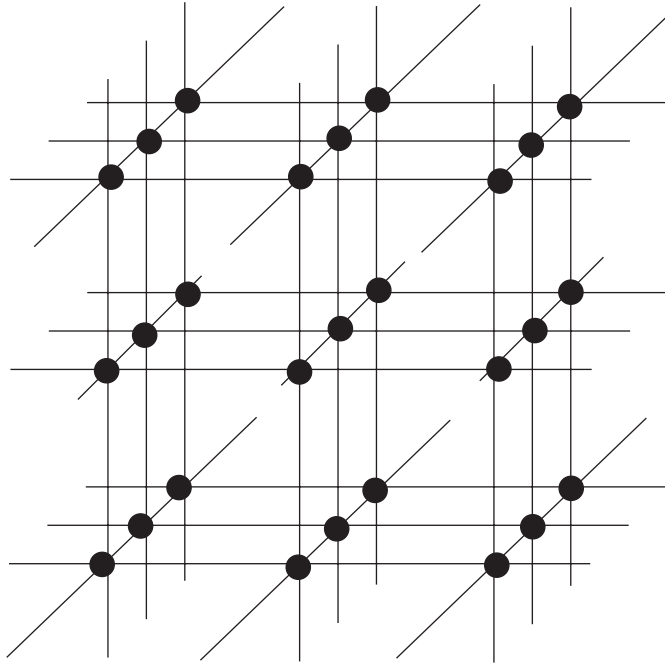


Figure 10.2 Simple cubic lattice

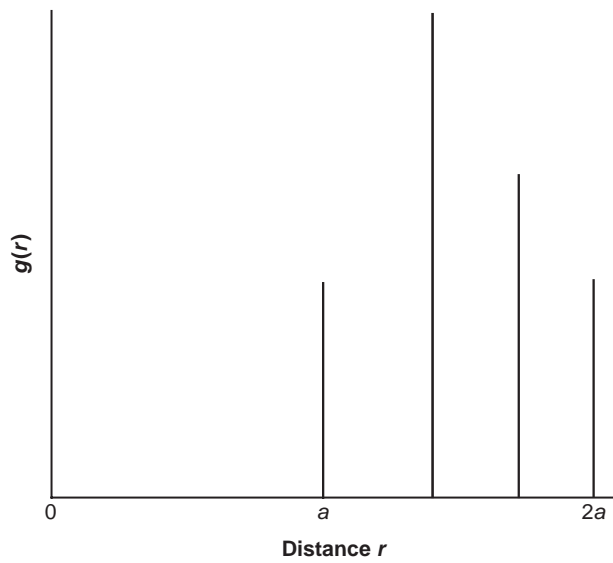


Figure 10.3 First part of the radial distribution function for a simple solid

Radial distribution functions can be deduced experimentally from diffraction studies. In the case of a liquid (Figure 10.4), the curve resembles that expected for a solid at low temperatures, and at high temperatures it resembles the quadratic expected for an ideal gas. At intermediate temperatures, the two features can be clearly seen; essentially a solid pattern is superimposed on the gas pattern. This gives the experimental basis for the well-known remark about liquid structure quoted above.

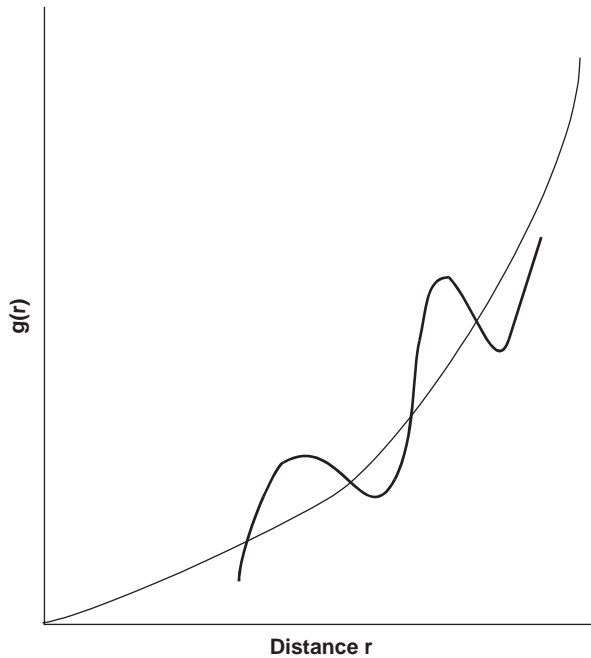


Figure 10.4 Radial distribution function for a liquid superimposed on an ideal gas

10.2 Pair Correlation Functions

The radial distribution function for a gas varies as $4\pi r^2$ and so tends to infinity as r tends to infinity. It is usual to remove the $4\pi r^2$ dependence by defining a related quantity called the *pair correlation function* $g_{AB}(r)$, which gives information about the probability of finding two particles A and B separated by distance r . If the volume of a system is V and it contains N_A species of type A and N_B species of type B, then the number densities are N_A/V and N_B/V . The fraction of time that the differential volume elements $d\tau_1$ and $d\tau_2$, which are separated by distance r , simultaneously contain species of type A and B is given by

$$\frac{N_A}{V} \frac{N_B}{V} g_{AB}(r) d\tau_1 d\tau_2$$

In a mixture of A and B we would be interested in the three distinct pair correlation functions $g_{AA}(r)$, $g_{BB}(r)$ and $g_{AB}(r)$. These pair correlation functions have a limiting value of 1 for a fluid.

10.3 Molecular Dynamics Methodology

In an ideal gas, the particles do not interact with each other and so the potential Φ is zero. Deviations from ideality are due to the interparticle potential, and most of the early studies were made on just three types of particle: the hard sphere model discussed in Chapter 9, the finite square well and the Lennard-Jones model.

10.3.1 Hard Sphere Potential

Just to remind you, the hard sphere potential is the simplest one imaginable; the system consists of spheres of radii σ and $U_{AB}(R_{AB})$ is zero everywhere except when the two spheres A and B touch, when it becomes infinite.

The hard sphere potential is of great theoretical interest not because it represents the intermolecular potential of any known substance, rather because any calculations based on the potential are simple.

Alder and Wainwright (1957) introduced the modelling technique now known as molecular dynamics to the world in a short article in 1957. They reported a study of hard discs, the two-dimensional equivalent of hard spheres.

10.3.2 Finite Square Well

A subsequent paper by Alder and Wainwright (1959) is usually regarded as the keynote paper in the field, and I quote the abstract:

A method is outlined by which it is possible to calculate exactly the behaviour of several hundred interacting classical particles. The study of this many-body problem is carried out by an electronic computer that solves numerically the simultaneous equations of motion. The limitations of this numerical scheme are enumerated and the important steps in making the program efficient on computers are indicated. The applicability of this method to the solution of many problems in both equilibrium and nonequilibrium statistical thermodynamics is discussed.

In this second paper they chose a three-dimensional system of particles and the finite square well potential shown in Figure 10.5. This potential is especially simple because a given particle does not experience any change in velocity except when it is separated from another particle by σ_2 (when it undergoes an attractive collision) or σ_1 (when it undergoes a repulsive collision).

On collision, the velocities are adjusted and the calculation restarts. Statistical data is collected every collision.

In their dynamical calculation, all the particles were given initial velocities and positions. Typically, equal kinetic energies with the three direction cosines of the velocity vector were chosen at random and the initial positions corresponding to a face-centred cubic lattice. Once the initial configuration was set up, they calculated exactly the time at which the first collision occurs. The collision time can be found by evaluating for every pair in the system, the time taken for the projected paths to reach a separation of σ_1 or σ_2 .

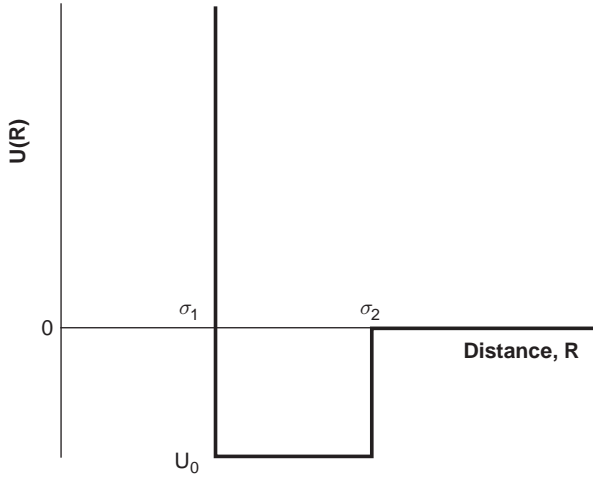


Figure 10.5 Finite square well

If two particles A and B have initial positions $\mathbf{r}_{A,0}$ and $\mathbf{r}_{B,0}$ and velocities \mathbf{v}_A and \mathbf{v}_B then the instantaneous positions at time t will be

$$\begin{aligned}\mathbf{r}_A &= \mathbf{r}_{A,0} + \mathbf{u}_A t \\ \mathbf{r}_B &= \mathbf{r}_{B,0} + \mathbf{u}_B t\end{aligned}$$

giving

$$\mathbf{r}_A - \mathbf{r}_B = \mathbf{r}_{A,0} - \mathbf{r}_{B,0} + (\mathbf{u}_A - \mathbf{u}_B) t$$

and so

$$(\mathbf{r}_A - \mathbf{r}_B)^2 = (\mathbf{r}_{A,0} - \mathbf{r}_{B,0})^2 + 2t (\mathbf{r}_{A,0} - \mathbf{r}_{B,0}) \cdot (\mathbf{u}_A - \mathbf{u}_B) + t^2 (\mathbf{u}_A - \mathbf{u}_B)^2$$

If we rewrite the last equation as a quadratic in t as

$$u_{AB}^2 t^2 + 2b_{AB} t + r_{AB,0}^2 = \sigma_\alpha^2$$

where α takes values 1 or 2, we see that the time required for a repulsive or attractive collision is

$$t_{AB}^{(\alpha)} = \frac{-b_{AB} \pm (b_{AB}^2 - u_{AB}^2 (r_{AB,0}^2 - \sigma_\alpha^2))^{1/2}}{u_{AB}^2} \quad (10.4)$$

In order to find the first collision time, all pairs have to be analysed. All the particles are then allowed to move for such time, and the velocities of the colliding pair are adjusted according to the equations of motion.

The finite square well occupies an important place in the history of molecular modelling. Real atomic and molecular systems have much more complicated mutual potential energy functions, but the finite square well does at least show a minimum. On the other hand, because of the finite square well potential, the equations of motion are particularly simple and no complicated numerical techniques are needed. There are no accelerations until two particles collide.

10.3.3 Lennardjonesium

The first simulation of a ‘real’ chemical system was Rahman’s (1964) study of liquid argon. He studied a system comprising 864 Lennard-Jones (L-J) particles under conditions appropriate to liquid argon at 84.4 K and a density of 1.374 g cm^{-3} . Once again, there is much to be gained by studying the abstract, so here is the first part of it.

A system of 864 particles interacting with a Lennard-Jones potential and obeying classical equations of motion has been studied on a digital computer (CDC 3600) to simulate molecular dynamics in liquid argon at 94.4 K and a density of 1.374 g cm^{-3} . The pair correlation function and the constant of self-diffusion are found to agree well with experiment; the latter is 15% lower than the experimental value. The spectrum of the velocity autocorrelation function shows a broad maximum in the frequency range $\omega = 0.25(2\pi k_B T/h)$. The shape of the Van Hove function $G_s(r, t)$ attains a maximum departure from a Gaussian at about $t = 0.3 \times 10^{-2} \text{ s}$ and becomes a Gaussian again at about $10^{-11} \text{ s} \dots$

There are several interrelated problems. A sample size has to be chosen; this is usually determined by the available computer resource and the complexity of the potential function, because the potential function has to be calculated very many times during the simulation. The number of particles and the density determine the size of the container. At the same time we need to decide on a potential function; the natural choice for the inert gases is the L-J potential, and we should note that the L-J potential is essentially short range.

So many early papers used the L-J potential that the noun *Lennardjonesium* was coined to describe a nonexistent element whose atoms interacted via the L-J potential.

10.4 Algorithms for Time Dependence

Once we have calculated the potential and hence the force by differentiation, we have to solve Newton’s equation of motion. If \mathbf{F}_A is the force on particle A, whose position vector is \mathbf{r}_A and whose mass is m_A then

$$\begin{aligned}\mathbf{F}_A &= m_A \frac{d^2 \mathbf{r}_A}{dt^2} \\ &= m_A \mathbf{a}_A\end{aligned}\quad (10.5)$$

This is a second-order differential equation that I can write equivalently as two first-order differential equations for the particle position \mathbf{r}_A and the velocity \mathbf{v}_A

$$\begin{aligned}\mathbf{F}_A &= m_A \frac{d\mathbf{v}_A}{dt} \\ \mathbf{v}_A &= \frac{d\mathbf{r}_A}{dt}\end{aligned}\quad (10.6)$$

10.4.1 Leapfrog Algorithm

A simple algorithm for integration of these two equations numerically in small time steps Δt can be found by considering the Taylor expansion for $\mathbf{v}(t)$:

$$\begin{aligned}\mathbf{v}_A \left(t + \frac{\Delta t}{2} \right) &= \mathbf{v}_A(t) + \left(\frac{d\mathbf{v}_A}{dt} \right)_t \frac{\Delta t}{2} + \frac{1}{2} \left(\frac{d^2 \mathbf{v}_A}{dt^2} \right)_t \left(\frac{\Delta t}{2} \right)^2 + \dots \\ \mathbf{v}_A \left(t - \frac{\Delta t}{2} \right) &= \mathbf{v}_A(t) - \left(\frac{d\mathbf{v}_A}{dt} \right)_t \frac{\Delta t}{2} + \frac{1}{2} \left(\frac{d^2 \mathbf{v}_A}{dt^2} \right)_t \left(\frac{\Delta t}{2} \right)^2 + \dots\end{aligned}\quad (10.7)$$

Subtracting and rearranging we get

$$\mathbf{v}_A \left(t + \frac{\Delta t}{2} \right) = \mathbf{v}_A \left(t - \frac{\Delta t}{2} \right) + \mathbf{a}_A(t) \Delta t + \dots \quad (10.8)$$

I will switch back and forth between (for example) \mathbf{v} and $d\mathbf{r}/dt$ in order to try and improve the readability of the equations. Also, I could have written $\mathbf{v}_A(t)$ or $(\mathbf{v}_A)_t$ to mean the instantaneous velocity of particle A at time t . The acceleration \mathbf{a} is calculated from the force.

Using the same procedure for the Taylor expansion of \mathbf{r}_A at the time point $t + \Delta t/2$ we get

$$\mathbf{r}_A(t + \Delta t) = \mathbf{r}_A(t) + \mathbf{v}_A \left(t + \frac{\Delta t}{2} \right) \Delta t + \dots \quad (10.9)$$

Equations (10.8) and (10.9) form the so-called *leapfrog algorithm*, which is reputed to be one of the most accurate and stable techniques for use in molecular dynamics.

A suitable time increment Δt for molecular dynamics is a femtosecond (10^{-15} s). In the leapfrog scheme, the velocities are first calculated at time $t + \Delta t/2$. These are used to calculate the positions of the particles at time $t + \Delta t$ and so on. In this way the velocities leap over the positions and then the positions leap over the velocities.

10.4.2 Verlet Algorithm

If instead we start from the Taylor expansion of $\mathbf{r}_A(t)$ we have

$$\begin{aligned} \mathbf{r}_A(t + \Delta t) &= \mathbf{r}_A(t) + \left(\frac{d\mathbf{r}_A}{dt} \right)_t \Delta t + \frac{1}{2} \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 + \dots \\ \mathbf{r}_A(t - \Delta t) &= \mathbf{r}_A(t) - \left(\frac{d\mathbf{r}_A}{dt} \right)_t \Delta t + \frac{1}{2} \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 + \dots \end{aligned} \quad (10.10)$$

which gives (assuming that third-order and higher terms are negligible)

$$\mathbf{r}_A(t + \Delta t) = 2\mathbf{r}_A(t) - \mathbf{r}_A(t - \Delta t) + \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 \quad (10.11)$$

This is known as the *Verlet algorithm*. The acceleration is obtained from the force experienced by atom A at time t . The velocity does not appear in the expression, but it can be obtained from the finite difference formula

$$\mathbf{v}_A(t) = \frac{\mathbf{r}_A(t + \Delta t) - \mathbf{r}_A(t - \Delta t)}{2\Delta t} \quad (10.12)$$

The Verlet algorithm uses positions and accelerations at time t and the position at time $t - \Delta t$ to calculate a new position at time $t + \Delta t$. All these have to be stored at every iteration.

A variant is the *velocity Verlet* algorithm, which requires only the storage of positions, velocities and accelerations that all correspond to the same time step. It takes the form

$$\begin{aligned} \mathbf{r}_A(t + \Delta t) &= \mathbf{r}_A(t) + \left(\frac{d\mathbf{r}_A}{dt} \right)_t \Delta t + \frac{1}{2} \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 \\ \mathbf{v}_A(t + \Delta t) &= \left(\frac{d\mathbf{r}_A}{dt} \right)_t + \frac{1}{2} \left(\left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t + \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_{t+\Delta t} \right) \Delta t \end{aligned} \quad (10.13)$$

There are many other algorithms in the literature, each with their own strengths and weaknesses.

10.5 Molten Salts

Molten salts (such as the alkali halides) are of great technological interest in the field of metal extraction. The first simulations were done in by Woodcock (1971). Molten salts bring a new problem because the potential energy terms are long range. Consider a (hypothetical) one-dimensional infinite crystal, part of which is shown in Figure 10.6.

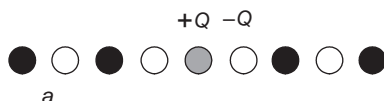


Figure 10.6 Part of a one-dimensional crystal; separation between ions is a

The unshaded ions have charge $-Q$, the shaded ions have charge $+Q$ and the spacing between the ions is a . Suppose we start with the central (grey) ion at infinity, and all the other ions at their lattice positions as shown. The work done W in bringing the grey ion from infinity to its place in the lattice is

$$W = -\frac{2Q^2}{4\pi\epsilon_0 a} \left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots \right)$$

and the term in brackets converges very slowly to its limiting value of $\ln(2)$:

No. of terms	Sum
1	1.0000
2	0.5000
3	0.8333
4	0.5833
100	0.6882

Such series have to be summed when calculating the force on a given ion in a periodic box; in the case of neutral species, the sum is very quickly convergent because of the short-range nature of the forces. For neutral systems, a cut-off radius is usually taken beyond which the interactions are set to zero.

10.6 Liquid Water

Water plays a prominent place amongst solvents for obvious reasons. Two-thirds of our planet is covered in water, chemical reactions tend to be done in aqueous solution, and so on. The relatively strong hydrogen bonding in liquid water causes many of its physical properties to be ‘anomalous’, and the structure of ice has long interested both theoreticians and experimentalists.

Neutron diffraction studies on heavy ice, solid D_2O , have shown that water molecules retain their identity in condensed phases with very little distortion of their molecular geometry. This means that water molecules may be treated as rigid asymmetric rotors (with six

degrees of freedom) rather than explicitly treating the three nuclei separately (nine degrees of freedom). The classical energy for a collection of N rigid rotor molecules consists of the kinetic energy for translation and rotation, together with the intermolecular potential. Each water molecule is described by six coordinates: three specify the centre of mass and three angles that fix the spatial orientation about the centre of mass. In this section, I will denote these coordinates by the six-dimensional vector \mathbf{X} . In terms of the linear velocities \mathbf{v}_i , the angular velocity vectors $\boldsymbol{\omega}_i$, the moments of inertia \mathbf{I}_i and the coordinates \mathbf{X}_i , the energy turns out to be

$$\varepsilon = \frac{1}{2} \sum_{i=1}^N (m v_i^2 + \boldsymbol{\omega}_i^T \mathbf{I}_i \boldsymbol{\omega}_i) + U(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N)$$

The question then is the extent to which the intermolecular potential is pair wise additive; such functions may always be resolved into contributions from pairs, triples and higher contributions:

$$U(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N) = \sum_{i < j}^N U^{(2)}(\mathbf{X}_i, \mathbf{X}_j) + \sum_{i < j < k}^N U^{(3)}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_k) + \dots + U^{(N)}(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N) \quad (10.14)$$

In the case of simple fluids such as liquid argon it is widely believed that U is practically pair wise additive. In other words, the contributions $U^{(p)}$ for $p > 2$ are negligible. However, the local structure in liquid water is thought to depend on just these higher order contributions and so it is unrealistic in principle to terminate the expansion with the pair contributions. It is legitimate to write the potential as a sum of pair potentials, provided one understands that they are effective pair potentials that somehow take higher terms into account.

The classic molecular dynamics study of liquid water is that of Rahman and Stillinger (1971). They wrote the effective pair potential as a sum of two contributions: a L-J 12-6 potential $U_{\text{L-J}}$ for the two oxygen atoms

$$U_{\text{L-J}}(R) = 4\varepsilon \left(\left(\frac{\sigma}{R} \right)^{12} - \left(\frac{\sigma}{R} \right)^6 \right)$$

and a function U_{el} , modulated by a function $S(R_{ij})$ that depends sensitively on the molecular orientations about the oxygen atoms

$$U^{(2)}(\mathbf{X}_i, \mathbf{X}_j) = U_{\text{L-J}}(R_{ij}) + S(R_{ij}) U_{\text{el}}(\mathbf{X}_i, \mathbf{X}_j) \quad (10.15)$$

The L-J parameters were taken to be those appropriate to neon, $\sigma = 282$ pm and $\varepsilon = 5.01 \times 10^{-22}$ J, on the grounds that neon is isoelectronic with water.

Four point charges Q each of magnitude $0.19e$ and each 100 pm from the oxygen nucleus are embedded in the water molecule in order to give U_{el} . Two charges are positive, to simulate the hydrogen atoms, whilst the other two are negative and simulate the lone pairs. The four charges are arranged tetrahedrally about the oxygen.

The set of 16 Coulomb interactions between two water molecules gives U_{el} . Figure 10.7 shows the minimum energy conformation, with the two 'molecules' 76 pm apart (i.e. an oxygen–oxygen distance of 276 pm).

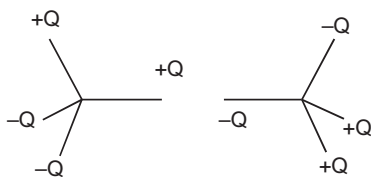


Figure 10.7 Two water molecules

The so-called switching function prevents the charges overlapping, and takes values between 0 and 1 depending on the distance between the oxygen atoms.

The authors took $N = 216$ water molecules in a cube of side 1862 pm, which corresponds to a density of 1 g cm^{-3} . They took periodic boundary conditions in order to eliminate undesirable wall effects, and a cut-off of distance of 916.5 pm beyond which the intermolecular interactions were set to zero.

The first step in a molecular dynamics calculation is to generate a suitable initial state. In general, random structures created using density values consistent with experiment may give atoms that are very close together. This causes undesirable large repulsive forces in small localized regions of space. One approach is to start from a low-density system with the atoms or molecules placed on simple cell lattices. The density is then increased gradually.

Starting velocities then have to be assigned. One technique is to calculate a value of the speed from equipartition of energy, and then assign random directions to each particle, with the constraint that the system has zero overall momentum.

The next step is to let the system progress for a sufficient number of time steps in order to eliminate any undesirable effects due to the choice of initial conditions.

After that, collection of statistics begins. The temperature is inferred from the average values of the translational and rotational kinetic energies, both of which should equal $\frac{3}{2}Nk_B T$.

It is possible in principle to follow a molecular dynamics simulation by displaying the atom coordinates at a series of time points. It proves more convenient to calculate certain statistical quantities. I mentioned the pair correlation function $g^{(2)}(r)$ earlier; in water the three distinct types of nuclear pairs lead to three corresponding functions $g_{\text{OO}}^{(2)}(r)$, $g_{\text{OH}}^{(2)}(r)$ and $g_{\text{HH}}^{(2)}(r)$, which give information about the fraction of time that differential volume elements separated by distance r simultaneously contain pairs of the nuclei given by the subscripts. These functions all have a limiting value of 1 as $r \rightarrow \infty$. The $g_{\text{OO}}^{(2)}(r)$ curve is shown schematically in Figure 10.8. The horizontal axis is actually a reduced distance r/σ .

Two important features of the curve are as follows.

1. The narrow first peak at 275 pm corresponds to 5.5 neighbouring oxygen atoms.
2. The second peak is low and broad, with a maximum at about 465 pm. The ratio of the second peak to the first, $465 \text{ pm}/275 \text{ pm} = 1.69$ is close to that observed for an ideal ice structure ($2\sqrt{2}/\sqrt{3} = 1.633$).

The *mean square displacement* is often calculated. This indicates the average displacement of an atom during a fixed time period.

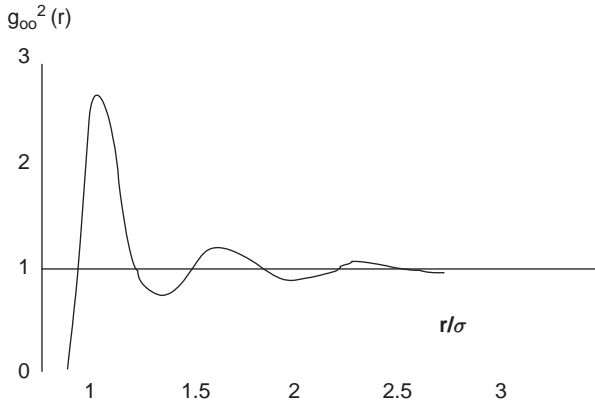


Figure 10.8 O–O pair distribution function

10.6.1 Other Water Potentials

Over the years, a number of authors have tackled the problem of finding a suitable pair potential for liquid water. All involve a rigid water monomer with a number of interaction sites. The original TIP3 model proposed by Jorgensen (1981) has positive charges on the hydrogens and a negative charge ($q_O = -2q_H$) on the oxygen. The interaction potential is taken as the sum of all intermolecular Coulomb terms together with a single L-J 12-6 term between the oxygens:

$$U_{AB} = \frac{1}{4\pi\epsilon_0} \sum_{i \text{ on A}} \sum_{j \text{ on B}} \frac{q_i q_j}{R_{ij}} + \frac{A}{R_{OO}^{12}} - \frac{B}{R_{OO}^6} \quad (10.16)$$

The parameters q_i , A and B were chosen to give reasonable energetic results for gas-phase complexes of water and alcohols.

The author subsequently reoptimized the parameters to give TIP3P, an improved three-site model.

Four-site models have also been used. The oldest is that due to Bernal and Fowler (1933), with a more modern version referred to as TIP4P by Jorgensen *et al.* (1983). In this case, the negative charge is moved off the oxygen and towards the hydrogens at a point along the bisector of the HOH angle. The pair potential is still calculated according to the equation above but more distances have to be evaluated.

Stillinger and Rahman's ST2 potential is widely used; this is an improved version of the one discussed above, but adds a L-J 12-6 term between the oxygen atoms. In all, 17 distances have to be evaluated.

10.7 Different Types of Molecular Dynamics

When Newton's equations of motion are integrated, the energy is conserved. Since the energy and temperature are related by the equipartition of energy principle, the temperature should also be constant. Slow temperature drifts do occur as a result of the numerical

integration and also because of the truncation of the forces. Several methods for performing molecular dynamics at constant temperature have been proposed, of which the simplest is known as *rescaling*. If the current temperature at time t is $T(t)$ then for a system of N particles and $3N$ degrees of freedom we have

$$\langle \varepsilon_{\text{kin}} \rangle = \left\langle \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \right\rangle = \frac{3}{2} N k_B T(t)$$

To adjust the temperature to exactly a reference temperature T_{ref} we simply rescale the velocities by a factor

$$\left(\frac{T_{\text{ref}}}{T(t)} \right)^{1/2}$$

Scaling the velocities at appropriate intervals can therefore control the temperature of the system. We speak about *constant-temperature molecular dynamics*.

Nose (1984) proposed an alternative method that adds an extra degree of freedom referred to as a heat bath to the atomic degrees of freedom. Extra kinetic energy and potential energy terms are added to the total energy, and the molecular dynamics is carried out with this one extra degree of freedom. Energy can flow back and forth between the heat bath and the system, and an equation of motion for the extra degree of freedom is solved.

Similar considerations apply to the pressure; in order to control the pressure in a system, any change in volume must be monitored and adjusted. Methods are available, and such calculations are referred to as *constant-pressure molecular dynamics*.

10.8 Uses in Conformational Studies

Molecular Dynamics is generally used to study the structure and behaviour of materials, where one is concerned with the intermolecular potential. It is of course also perfectly possible to study single molecules and it is sometimes used as an aid in conformational analysis. As noted previously, there are very many local minima on the potential energy surface and it may be that small variations in torsional angles do not facilitate searching of all regions.

The idea is to start with a MM local minimum, which corresponds to a temperature of 0 K. We use molecular dynamics to raise the temperature, let the molecule vibrate for a while at an elevated temperature and cool it back to 0 K. Hopefully the molecule will then have crossed over any potential barriers and a reoptimization will give a new structure.

To give a simple example, consider the straight-chain hydrocarbon $\text{C}_{20}\text{H}_{42}$, shown in Figure 10.9. A MM+ optimization starting from the fully stretched structure gave the structure shown. There is restricted rotation about each C–C bond and in a later chapter I will show how theories of polymeric molecules focus attention on the end-to-end distance; the simplest theories predict relationships between this and the number of C–C links in such a molecule. $\text{C}_{20}\text{H}_{42}$ is barely a polymer, but the principles are the same. There are very many conformers, all similar in energy. I therefore set up a 15 ps molecular dynamics calculation, as shown in Figure 10.9.

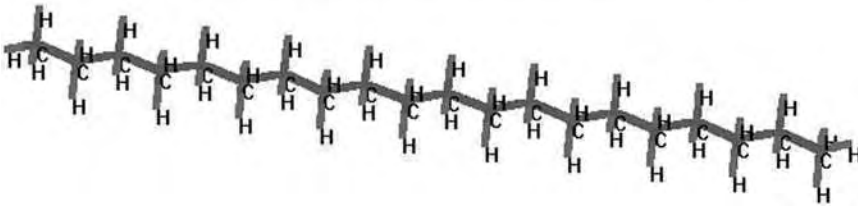
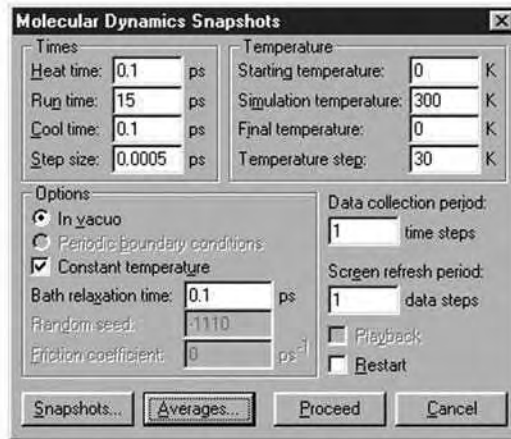


Figure 10.9 Molecular dynamics calculation on a hydrocarbon

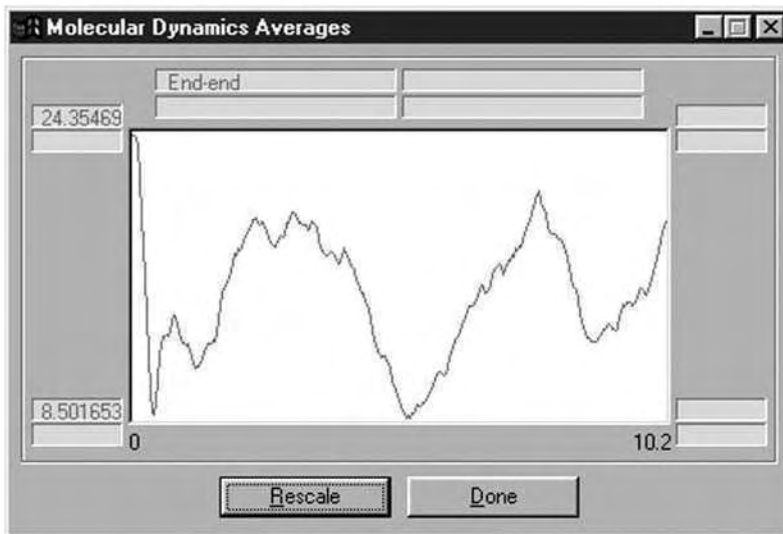


Figure 10.10 End-to-end distance

The options are shown, and statistics were collected in an MS/EXCEL file. The end-to-end distance varied dramatically over the experiment, from its initial value of 2435 pm through 850 pm. The average end-to-end distance was 1546 pm, as shown in Figures 10.10 and 10.11.

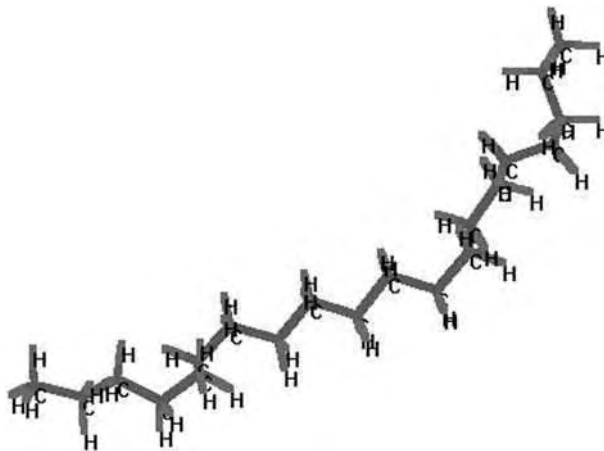


Figure 10.11 C_{20} hydrocarbon at end of molecular dynamics experiment

The resulting structure was then optimized. One strategy in such calculations is to take snapshots of the structure at regular time intervals and then optimize from each snapshot.

References

- Alder, B.J. and Wainwright, T.E. (1957) *J. Chem. Phys.*, **27**, 1208.
Alder, B.J. and Wainwright, T.E. (1959) *J. Chem. Phys.*, **31**, 459.
Bernal, J.D. and Fowler, R.H. (1933) *J. Chem. Phys.*, **1**, 515.
Jorgensen, W.L. (1981) *J. Am. Chem. Soc.*, **103**, 335.
Jorgensen, W.L., Chandrasekhar, J., Madura, J.D. *et al.* (1983) *J. Chem. Phys.*, **79**, 926.
Nose, S. (1984) *Mol. Phys.*, **52**, 255.
Rahman, A. (1964) *Phys. Rev.*, **1136**, A405.
Rahman, A. and Stillinger, F.H. (1971) *J. Chem. Phys.*, **55**, 3336.
Woodcock, L.V. (1971) *Chem. Phys. Lett.*, **10**, 257.

11

Introduction to Quantum Modelling

By the early days of the twentieth century, scientists had successfully developed three of the four great cornerstones of physics: Sir Isaac Newton's mechanics, James Clerk Maxwell's electromagnetic theory and Albert Einstein's theory of special relativity. They had a picture of the physical world where matter was made from point particles and radiation consisted of electromagnetic waves, and this picture seemed to explain all known physical phenomena with just a few untidy exceptions.

These untidy exceptions comprised phenomena such as the *theory of black body radiation*, the *photoelectric effect*, *Compton scattering*, *atomic structure and spectra* and a few other apparently unrelated experimental findings. I do not have space in this text to go into the historical detail; I will simply say that a thorough study of such phenomena led to the fourth cornerstone, quantum theory. Every textbook that has to deal with quantum theory has its own particular treatment and starting point, and this one is no exception; I am going to assume that you have heard of Erwin Schrödinger and his famous equations, and start the discussion at that point. Perhaps I can reassure you by saying that most professional scientists perceive quantum theory as a hard subject (along with electromagnetism). Even Schrödinger did not fully understand the physical meaning that we now attach to his wavefunctions when he first wrote down his famous equation and solved it for the hydrogen atom.

11.1 The Schrödinger Equation

Consider then the simple case of a particle of mass m constrained to the x axis by a potential $U(x, t)$ that depends on x and time t . I have allowed the potential to be time dependent in order to cover the possibility of (for example) an electron being influenced by an external electromagnetic wave. Schrödinger's time-dependent equation states that

$$\left(-\frac{\hbar^2}{8\pi^2 m} \frac{\partial^2}{\partial x^2} + U(x, t)\right) \Psi(x, t) = j \frac{\hbar}{2\pi} \frac{\partial \Psi(x, t)}{\partial t} \quad (11.1)$$

where j is the square root of -1 ($j^2 = -1$). $\Psi(x, t)$ is the *wavefunction* and in general it is a complex valued quantity. The square of the modulus of Ψ can be found from

$$|\Psi(x, t)|^2 = \Psi^*(x, t) \Psi(x, t)$$

where Ψ^* is the complex conjugate of Ψ . The fact that complex wavefunctions are allowed and that the Schrödinger equation mixes up real and imaginary quantities caused Schrödinger and his contemporaries serious difficulties in trying to give a physical meaning to Ψ . According to Max Born, the physical interpretation attaching to the wavefunction $\Psi(x, t)$ is that $\Psi^*(x, t)\Psi(x, t)dx$ gives the probability of finding the particle between x and $x + dx$ at time t . It is therefore the square of the modulus of the wavefunction rather than the wavefunction itself that is related to a physical measurement.

Since $\Psi^*(x, t)\Psi(x, t)dx$ is the probability of finding the particle (at time t) between x and $x + dx$, it follows that the probability of finding the particle between any two values a and b is found by summing the probabilities

$$P = \int_a^b \Psi^*(x, t) \Psi(x, t) dx$$

In three dimensions the equation is more complicated because both the potential U and the wavefunction Ψ might depend on the three dimensions x , y and z (written as vector \mathbf{r}) as well as time:

$$\left(-\frac{\hbar^2}{8\pi^2 m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r}, t)\right) \Psi(\mathbf{r}, t) = j \frac{\hbar}{2\pi} \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} \quad (11.2)$$

You are probably aware that the operator in the outer left-hand bracket is the *Hamiltonian operator* and that we normally write the equation

$$H\Psi(\mathbf{r}, t) = j \frac{\hbar}{2\pi} \frac{\partial \Psi(\mathbf{r}, t)}{\partial t} \quad (11.3)$$

The Born interpretation is then extended as follows:

$$\Psi^*(\mathbf{r}, t) \Psi(\mathbf{r}, t) dx dy dz \quad (11.4)$$

which gives the probability that the particle will be found at time t in the volume element $dx dy dz$. Such volume elements are often written $d\tau$.

I have not made any mention of time-dependent potentials in this text, but they certainly exist. You will have met examples if you have studied electromagnetism where the *retarded potentials* account for the generation of electromagnetic radiation.

For the applications we will meet in this book, the potentials are time independent; in such cases the wavefunction still depends on time, but in a straightforward way as I will now demonstrate.

11.2 The Time-Independent Schrödinger Equation

I am going to use a standard technique from the theory of differential equations called *separation of variables* to simplify the time-dependent equation; I want to show you that when the potential does not depend on time, we can ‘factor out’ the time dependence from the time-dependent Schrödinger equation and just concentrate on the spatial problem. I will therefore investigate the possibility that we can write $\Psi(\mathbf{r}, t)$ as a product of functions of the spatial (\mathbf{r}) and time (t) variables, given that the potential only depends on \mathbf{r} . We write

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r})T(t) \quad (11.5)$$

and substitute into the time-dependent Schrödinger equation to get

$$\left(-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r}, t)\right) \psi(\mathbf{r})T(t) = j\frac{\hbar}{2\pi} \frac{\partial \psi(\mathbf{r})T(t)}{\partial t}$$

Notice that the differential operators are partial ones and they only operate on functions that contain them. So we rearrange the above equation

$$T(t) \left(-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r}, t)\right) \psi(\mathbf{r}) = j\psi(\mathbf{r}) \frac{\hbar}{2\pi} \frac{dT(t)}{dt}$$

divide by $\Psi(\mathbf{r}, t)$ and simplify to get

$$\frac{1}{\psi(\mathbf{r})} \left(-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r}, t)\right) \psi(\mathbf{r}) = j\frac{1}{T(t)} \frac{\hbar}{2\pi} \frac{dT(t)}{dt}$$

The left hand side of the equation depends on the spatial variable \mathbf{r} and the time variable t . Suppose now that the potential depends only on \mathbf{r} and not on t . That is to say

$$\frac{1}{\psi(\mathbf{r})} \left(-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r})\right) \psi(\mathbf{r}) = j\frac{1}{T(t)} \frac{\hbar}{2\pi} \frac{dT(t)}{dt} \quad (11.6)$$

Under these circumstances, the left-hand side of the equation depends only on \mathbf{r} and the right-hand side only on t . The spatial and the time variables are completely independent and so the right-hand side and the left-hand side must both equal a constant. In the theory of differential equations, this constant is called the *separation constant*. In this case it is equal to the total energy so I will write it as ε :

$$\begin{aligned} \frac{1}{\psi(\mathbf{r})} \left(-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) + U(\mathbf{r})\right) \psi(\mathbf{r}) &= \varepsilon \\ j\frac{1}{T(t)} \frac{\hbar}{2\pi} \frac{dT(t)}{dt} &= \varepsilon \end{aligned}$$

The second equation can be solved to give

$$T(t) = A \exp\left(-\frac{2\pi j\varepsilon t}{\hbar}\right) \quad (11.7)$$

and so the full solution to the time-dependent equation in the special case that the potential is time independent can be written

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) \exp\left(-\frac{2\pi j\varepsilon t}{\hbar}\right) \quad (11.8)$$

Probabilities then become time independent since

$$\begin{aligned}\Psi^*(\mathbf{r}, t) \Psi(\mathbf{r}, t) &= \psi^*(\mathbf{r}) \exp\left(+\frac{2\pi j \varepsilon t}{h}\right) \psi(\mathbf{r}) \exp\left(-\frac{2\pi j \varepsilon t}{h}\right) \\ &= \psi^*(\mathbf{r}) \psi(\mathbf{r})\end{aligned}$$

11.3 Particles in Potential Wells

I have mentioned a number of model potentials in earlier chapters when dealing with molecular dynamics and Monte Carlo. A model potential of historical interest in quantum mechanical studies is the ‘infinite well’.

11.3.1 One-Dimensional Infinite Well

A particle of mass m is constrained to a region $0 \leq x \leq L$ by a potential that is infinite everywhere apart from this region, where it is equal to a constant U_0 (Figure 11.1). This might be a very crude model for the conjugated electrons in a polyene, or for the conduction electrons in a metallic conductor. (In fact there is a technical problem with this particular model potential caused by the infinities, as I will explain shortly.)

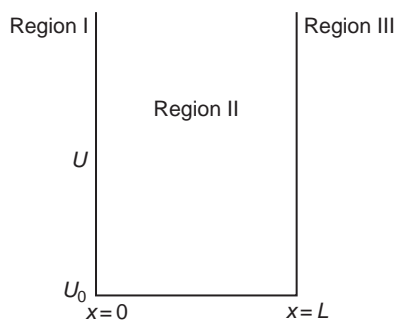


Figure 11.1 *One-dimensional infinite well*

To solve the problem, we examine the Schrödinger equation in each of the three regions and then match up the wavefunction at the boundaries. The wavefunction must be continuous at any boundary.

In regions I and III the potential is infinite and so the wavefunction must be zero, in order to keep the Schrödinger equation finite. In region II the potential is a finite constant U_0 and we have

$$\left(-\frac{\hbar^2}{8\pi^2 m} \frac{d^2}{dx^2} + U_0\right) \psi(x) = \varepsilon \psi(x)$$

which can be rewritten

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2m(\varepsilon - U_0)}{h^2}\psi = 0 \quad (11.9)$$

The standard solution to this second-order differential equation is

$$\psi = A \sin \sqrt{\frac{8\pi^2m(\varepsilon - U_0)}{h^2}}x + B \cos \sqrt{\frac{8\pi^2m(\varepsilon - U_0)}{h^2}}x$$

where A and B are constants of integration. We know that $\psi = 0$ in regions I and III so the wavefunction must also be zero at $x = 0$ and at $x = L$ in order to be continuous at these points. Substituting $\psi = 0$ at $x = 0$ shows that $B = 0$. If we substitute $\psi = 0$ at $x = L$ we get

$$0 = A \sin \sqrt{\frac{8\pi^2m(\varepsilon - U_0)}{h^2}}L$$

This means that either $A = 0$ or the argument of \sin must be an integral multiple of π (written n). A cannot be 0, for otherwise the wavefunction would be 0 everywhere, and so, on rearranging

$$\begin{aligned} \varepsilon_n &= U_0 + \frac{n^2h^2}{8mL^2} \\ \psi_n &= A \sin\left(\frac{n\pi}{L}x\right) \end{aligned} \quad (11.10)$$

The energy is quantized, and so I have added a subscript to the ε for emphasis. It generally happens that the imposition of boundary conditions leads to energy quantization. It is often necessary when tackling such problems to match up both the wavefunction and its gradient at the boundaries, but in this particular case the solution appeared just by considering the wavefunction. (In fact, the infinities in this particular potential mean that the first derivative is not continuous at the boundaries; 'real' potentials do not contain such sudden jumps to infinity.)

Parameter n is the *quantum number* and at first sight the allowed values are $n = \dots -2, -1, 0, 1, 2, \dots$. The value $n = 0$ is not allowed for the same reason that A cannot be 0. We can deduce the constant of integration A by recourse to the Born interpretation of quantum mechanics; the probability of finding the particle between x and $x + dx$ is

$$\psi_n^2(x) dx$$

Probabilities have to sum to 1 and so we require

$$\int_0^L \psi_n^2(x) dx = 1$$

from which we deduce

$$\psi_n = \pm \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right)$$

We notice also that the values of $\pm n$ give wavefunctions that differ only by multiplication by -1 , since

$$\sin\left(-\frac{n\pi x}{L}\right) = -\sin\left(\frac{n\pi x}{L}\right)$$

According to the Born interpretation, both give the same probability density and so we only need one set. It is conventional to take the quantum numbers as positive and the value of A as the positive square root so the results can be collected together as

$$\begin{aligned}\varepsilon_n &= U_0 + \frac{n^2 h^2}{8mL^2} \\ \psi_n &= \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \\ n &= 1, 2, 3, \dots\end{aligned}\tag{11.11}$$

Plots of the squares of the wavefunctions for $n = 1$ and $n = 3$ are shown in Figure 11.2. For a particle with $n = 1$ there is zero probability of finding the particle at the end points and a maximum probability of finding the particle at the centre point. For $n = 3$ there are three equal maxima of probability.

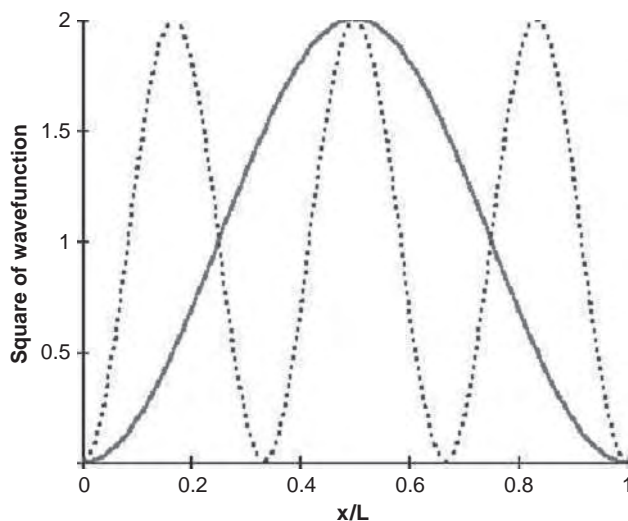


Figure 11.2 Squares of wavefunction for $n = 1$ (full curve) and $n = 3$ (dotted curve)

11.4 Correspondence Principle

In 1923, Neils Bohr formulated a principle that he had already used in his treatment of the hydrogen atom, called the *correspondence principle*; this states that quantum and classical predictions have to agree with each other in the limit, as the quantum number(s) get large. In other words, the results of classical physics may be considered to be the limiting cases of

quantum mechanical calculations with large quantum numbers. A plot of ψ_{30}^2 (Figure 11.3) shows the operation of the correspondence principle; the probabilities appear much more evenly spread out over the region. In the case of a ‘classical’ particle in such a well, we would expect an equal probability of finding the particle at any point along the line.

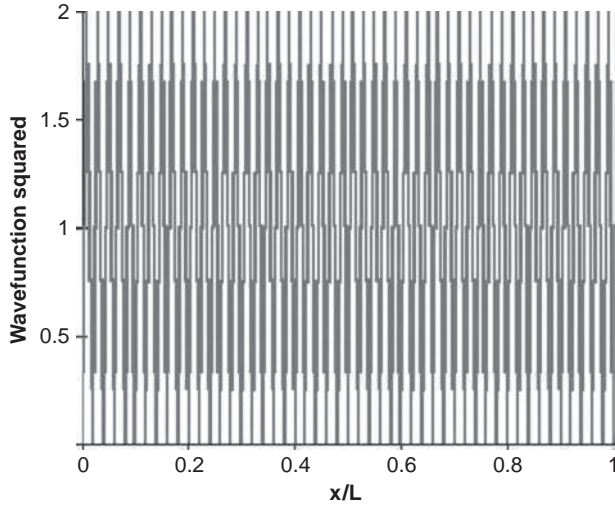


Figure 11.3 Square of wavefunction $n = 30$

11.5 Two-Dimensional Infinite Well

In the two-dimensional version of this problem, we consider a single particle of mass m constrained to a region of the xy plane by a potential that is infinite everywhere except for a region where $0 \leq x \leq L$, $0 \leq y \leq L$ where it is equal to a constant U_0 , as shown in Figure 11.4. I have taken the region to be square, but this is not necessary.

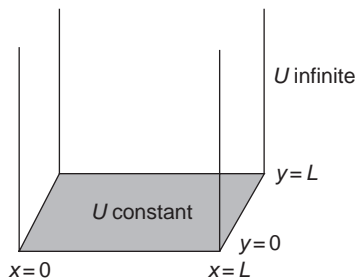


Figure 11.4 Two-dimensional infinite well

The (time-independent) Schrödinger equation is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{8\pi^2 m (\varepsilon - U_0)}{h^2} \psi = 0 \quad (11.12)$$

and the wavefunction now depends on x and y . Straightforward application of the ‘separation of variables’ technique discussed above gives the following wavefunctions and energies:

$$\begin{aligned} \varepsilon_{n,k} &= U_0 + \frac{n^2 h^2}{8mL^2} + \frac{k^2 h^2}{8mL^2} \\ \psi_{n,k} &= \frac{2}{L} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{k\pi y}{L}\right) \end{aligned} \quad (11.13)$$

$n, k = 1, 2, 3, \dots$

There are now two quantum numbers that I have decided to call n and k . Each quantum number can equal a positive integer. The wavefunction is a product of sine functions, and

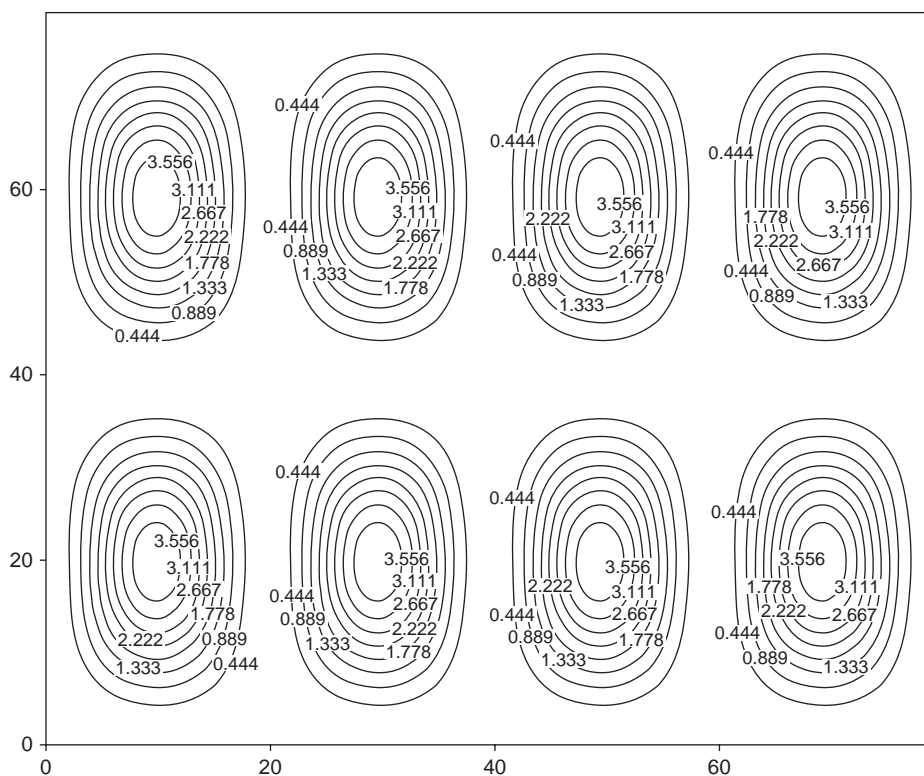


Figure 11.5 Square of wavefunction, contour view

we now have to think about visualization. Two popular ways to visualize such objects are as contour diagrams or as surface plots. The two representative Figures 11.5 and 11.6 refer to the solution $n = 4$ and $k = 2$, and in keeping with the importance of the Born interpretation, I have plotted the squares of the wavefunctions. The horizontal axis is x , and the vertical axis is y .

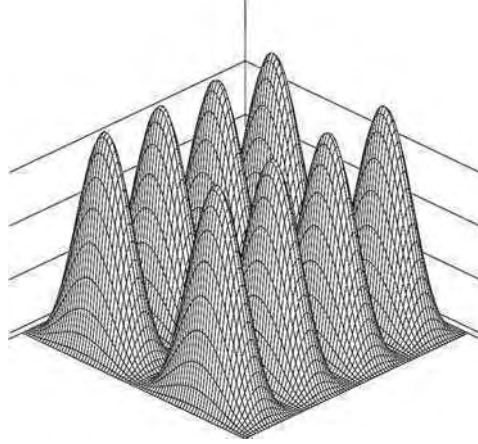


Figure 11.6 Two-dimensional box, surface plot

11.6 Three-Dimensional Infinite Well

It should be obvious by now how we proceed to three dimensions. The relevant Schrödinger equation is

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{8\pi^2 m (\varepsilon - U_0)}{h^2} \psi = 0 \quad (11.14)$$

which can be solved by the standard techniques discussed above. Application of the boundary conditions gives three quantum numbers that I will write n , k and l . The solutions are

$$\begin{aligned} \varepsilon_{n,k,l} &= U_0 + \frac{n^2 h^2}{8mL^2} + \frac{k^2 h^2}{8mL^2} + \frac{l^2 h^2}{8mL^2} \\ \psi_{n,k,l} &= \left(\frac{2}{L}\right)^{3/2} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{k\pi y}{L}\right) \sin\left(\frac{l\pi z}{L}\right) \end{aligned} \quad (11.15)$$

$n, k, l = 1, 2, 3, \dots$

Figure 11.7 summarizes the one-dimensional and the cubic three-dimensional infinite well problems. The energies are plotted in units of $h^2/8mL^2$. The one-dimensional energy levels are shown as the left-hand 'column' of the figure. They simply increase as n^2 .

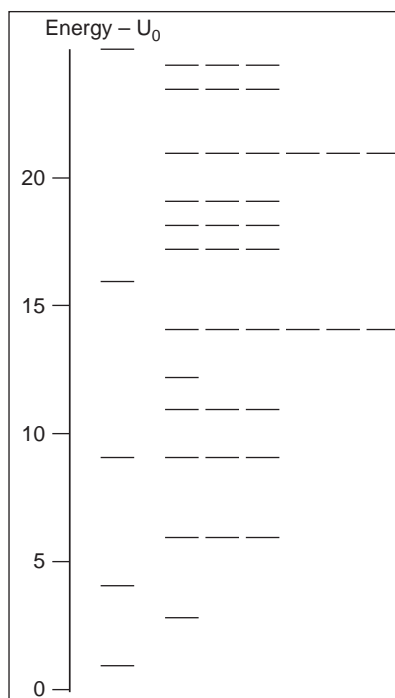


Figure 11.7 One-dimensional and cubic three-dimensional wells

The three-dimensional problem shows a much more interesting pattern. Firstly, many of the energy levels are *degenerate* (that is to say, there are many *quantum states* with the same energy). Some of the degeneracies are *natural*, for example $\psi_{1,2,3}$, $\psi_{1,3,2}$, $\psi_{2,1,3}$, $\psi_{2,3,1}$, $\psi_{3,1,2}$ and $\psi_{3,2,1}$ all have energy $14h^2/(8mL^2)$. There are also many *accidental* degeneracies; for example there are 9 degenerate quantum states with energy $41h^2/(8mL^2)$ because $1^2+2^2+6^2$ and $4^2+4^2+3^2$ are both equal to 41. Secondly, as the energy increases, the energy levels crowd together and the number of quantum states of near equal energy is seen to increase. This crowding together does not happen in the one-dimensional case.

The square two-dimensional case lies somewhere in between, with fewer degeneracies.

The degeneracies do not occur if we take arbitrary sides for the two- and three-dimensional regions. Nevertheless, their quantum states still crowd together as the quantum numbers increase.

11.7 Two Noninteracting Particles

Consider now the case of two noninteracting particles A and B in a one-dimensional well. Writing x_A and x_B for their position coordinates along the x axis, the time-independent Schrödinger equation is (taking the constant of integration to be zero, for simplicity)

$$-\frac{\hbar^2}{8\pi^2m} \left(\frac{\partial^2}{\partial x_A^2} + \frac{\partial^2}{\partial x_B^2} \right) \Psi(x_A, x_B) = E\Psi(x_A, x_B) \quad (11.16)$$

I have written capital Ψ and E to emphasize that the equation refers to a system of two particles, and I will use this convention from now on. The equation looks like a candidate for separation of variables, so we investigate the conditions under which the total wavefunction can be written as a product of one-particle wavefunctions

$$\Psi(x_A, x_B) = \psi_A(x_A) \psi_B(x_B)$$

Substitution and rearrangement gives

$$-\frac{\hbar^2}{8\pi^2m} \frac{1}{\psi_A} \frac{d^2\psi_A}{dx_A^2} - \frac{\hbar^2}{8\pi^2m} \frac{1}{\psi_B} \frac{d^2\psi_B}{dx_B^2} = E$$

Applying the now familiar argument, we equate each term on the left-hand side to a (different) separation constant ε_A and ε_B , which have to sum to E . The total wavefunction is therefore a product of wavefunctions for the individual particles, and the total energy is the sum of the one-particle energies. There are two quantum numbers that I will call n_A and n_B and we find

$$\begin{aligned} \Psi_{n_A, n_B} &= \frac{2}{L} \sin\left(\frac{n_A\pi x_A}{L}\right) \sin\left(\frac{n_B\pi x_B}{L}\right) \\ E_{n_A, n_B} &= (n_A^2 + n_B^2) \frac{\hbar^2}{8mL^2} \\ n_A, n_B &= 1, 2, 3, \dots \end{aligned} \quad (11.17)$$

We often summarize the results of such calculations, where the solution turns out to be a product of individual one-particle solutions, by a simple extension of Figure 11.7. Figure 11.8 shows one possible solution, where particle A has quantum number 2 and particle B has quantum number 1. There is more to this simple problem than meets the eye; you might be wondering how it is possible to ‘label’ one particle A and one B. If A and B had been two motorcars travelling along a motorway then the labelling could be done by the vehicle number plate. But how can we possibly label electrons, or any other atomic particles for that matter? This is not an easy question; I will come back to it later in the book.

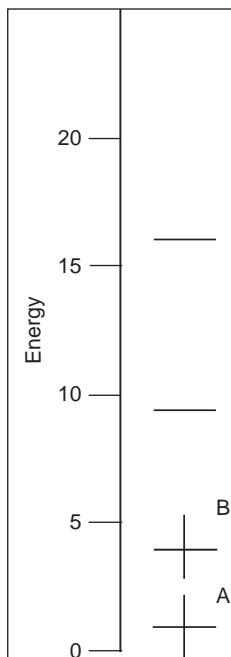


Figure 11.8 Two particles in a one-dimensional box

11.8 Finite Well

The infinite well occupies an important place in the history of molecular quantum mechanics, and we will meet it again in Chapter 12. We discussed the finite well in Chapters 9 and 10, but from the viewpoint of classical mechanics. The finite well potential gives a rather crude model for a diatomic molecule, and it is shown again in Figure 11.9. It keeps an infinite potential for the close approach of the two nuclei, but allows for a ‘valence region’ where the potential is constant.

The potential is infinite for $-\infty \leq x \leq 0$, zero (or a constant) for the valence region where $0 \leq x \leq L$ and a constant D for $L \leq x \leq \infty$. As usual, we divide up the x axis into three regions and then match up ψ at the two boundaries. It is also necessary to match up the first derivative of ψ at the $x = L$ boundary.

In region I, the potential is infinite and so $\psi = 0$. This means that $\psi = 0$ at the boundary when $x = 0$.

In region II, the potential is constant (zero) and so we have

$$-\frac{\hbar^2}{8\pi^2m} \frac{d^2\psi(x)}{dx^2} = \varepsilon\psi(x)$$

This can be solved to give

$$\psi = A \sin \sqrt{\frac{8\pi^2m\varepsilon}{\hbar^2}}x + B \cos \sqrt{\frac{8\pi^2m\varepsilon}{\hbar^2}}x$$

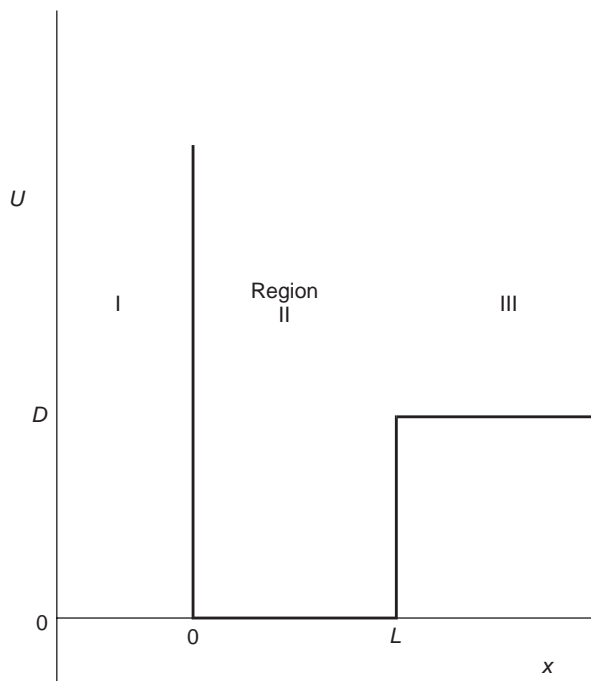


Figure 11.9 The finite box

where A and B are constants of integration. We can eliminate the B term straight away by the argument of Section 11.3.

In region III we have

$$\left(-\frac{\hbar^2}{8\pi^2 m} \frac{d^2}{dx^2} + D \right) \psi(x) = \varepsilon \psi(x)$$

For now, let us interest ourselves only in those solutions where the energy is less than or equal to D . Such solutions are called *bound states*. We will shortly see that solutions exist where the energy is greater than D and these are called *unbound states*. I can rewrite the Schrödinger equation for region III as

$$\frac{d^2 \psi}{dx^2} - \frac{\hbar^2}{8\pi^2 m} (D - \varepsilon) \psi = 0$$

The standard solution to this second-order differential equation is

$$\psi = E \exp\left(\sqrt{\frac{8\pi^2 m (D - \varepsilon)}{\hbar^2}} x\right) + F \exp\left(-\sqrt{\frac{8\pi^2 m (D - \varepsilon)}{\hbar^2}} x\right)$$

where E and F are constants of integration.

We now have to apply boundary conditions for the bound state solutions where ε is less than D . If we want the wavefunction to be zero at infinity, then E must be zero. The wavefunction and its first derivative must also be equal at the boundary $x = L$. Thus we have

$$A \sin \left(\sqrt{\frac{8\pi^2 m \varepsilon}{h^2}} L \right) = F \exp \left(-\sqrt{\frac{8\pi^2 m (D - \varepsilon)}{h^2}} L \right)$$

$$\sqrt{\frac{8\pi^2 m \varepsilon}{h^2}} A \cos \left(\sqrt{\frac{8\pi^2 m \varepsilon}{h^2}} L \right) = -\sqrt{\frac{8\pi^2 m (D - \varepsilon)}{h^2}} F \exp \left(-\sqrt{\frac{8\pi^2 m (D - \varepsilon)}{h^2}} L \right)$$

Dividing one equation by the other and rearranging gives a formula involving the energy ε :

$$\tan \left(\sqrt{\frac{8\pi^2 m \varepsilon}{h^2}} L \right) = -\sqrt{\frac{\varepsilon}{D - \varepsilon}} \quad (11.18)$$

The energy is quantized and the precise values of ε depend on D and L .

11.9 Unbound States

I mentioned in the discussion above that we would be concerned only with solutions where the energy ε was less than the well depth D . Such states are called *bound states*, by which one means that the particle is somehow localized to the potential well in the sense that the probability of finding the particle between x and $x + dx$ should decrease the further we move away from the potential well (you probably noticed that I *assumed* the wavefunction should fall off exponentially, and so I deliberately excluded the possibility that the probability could get larger with distance).

Depending on L and D , a one-dimensional finite well can have one, many or an infinite number of bound state solutions. Finite wells can also have *unbound* solutions that are perfectly acceptable in so far as both ψ and its first derivative are finite and continuous for all values of x , but they do not necessarily have the property that they tend to zero for large values of x .

Taken individually these solutions cannot represent the state of a particle because they do not vanish for large values of x . This causes a real problem with the Born interpretation because they cannot be made to satisfy the normalization condition. The solution to this problem is to combine waves together into *wave packets*, and this topic is discussed in more advanced textbooks.

11.10 Free Particles

The simplest example of unbound states is afforded by studying free particles; consider again the one-dimensional problem of a particle (of mass m) and a potential that is a constant which I will take to be zero. The time-independent Schrödinger equation is

$$\left(-\frac{\hbar^2}{8\pi^2 m} \frac{d^2}{dx^2} \right) \psi(x) = \varepsilon \psi(x)$$

which has solutions that I can write

$$\psi(x) = A \exp(jkx) \quad \text{where } k = \pm \frac{\sqrt{8\pi^2 m \varepsilon}}{h}$$

Here k can take any value; the energy ε is positive, but now it has a continuum of possible values. The time-dependent solution is therefore

$$\Psi(x, t) = A \exp(jkx) \exp\left(-\frac{2\pi j\varepsilon t}{h}\right)$$

which can be written

$$\Psi(x, t) = A \exp(j(kx - \omega t))$$

where

$$\omega = \frac{2\pi\varepsilon}{h}$$

This is a classic one-dimensional wave equation, and it corresponds to a de Broglie wave. In three dimensions we write

$$\Psi(\mathbf{r}, t) = A \exp(j(\mathbf{k} \cdot \mathbf{r} - \omega t)) \quad (11.19)$$

where \mathbf{k} is the wavevector. You will have encountered the wavevector if you have studied electromagnetism beyond elementary level.

We notice then that for each value of the energy, there are two possible wavefunctions

$$\begin{aligned} \psi(x) &= A \exp\left(+j\frac{\sqrt{8\pi^2 m\varepsilon}}{h}x\right) \\ \psi(x) &= A \exp\left(-j\frac{\sqrt{8\pi^2 m\varepsilon}}{h}x\right) \end{aligned} \quad (11.20)$$

That is, there is a double degeneracy. A feature of the free particle wavefunctions is that they cannot be made to satisfy the Born normalization condition, because there is no value of A that will satisfy the ‘equation’

$$|A|^2 \int_{-\infty}^{+\infty} |\psi(x)|^2 dx = 1$$

This is not a trivial problem, and workers in the field often interpret solutions of the form of Equation (11.20) as an *infinite beam* of noninteracting particles moving to the right or left (according to the $+$ or $-$ sign) with a constant number of particles per length.

11.11 Vibrational Motion

I gave you the classical treatment of a diatomic molecule in Chapters 4 and 5, where I assumed that the relevant potential was Hooke’s law. I demonstrated that the vibrations of a diatomic molecule (atomic masses m_1 and m_2) were exactly equivalent to the vibrations of a single atom of reduced mass μ where

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

If I write the ‘spring’ extension $R - R_e$ as x , then the vibrational Schrödinger equation can be written

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2\mu}{h^2} \left(\varepsilon - \frac{1}{2}k_s x^2 \right) = 0 \quad (11.21)$$

where k_s is the force constant. The energy ε and wavefunction ψ now describe the nuclear vibrations. This is a more difficult differential equation than the ones discussed earlier in this chapter, but fortunately it is well known in many areas of engineering, physics and applied mathematics. Detailed solution of the vibrational Schrödinger equation is given in all the advanced texts and to cut a long story short we interest ourselves in bound states and so impose a condition $\psi \rightarrow 0$ as $x \rightarrow \pm\infty$. This gives rise to energy quantization described by a single quantum number that is written ν . The general formula for the vibrational energies is

$$\varepsilon_\nu = \frac{h}{2\pi} \sqrt{\frac{k_s}{\mu}} \left(\nu + \frac{1}{2} \right) \quad (11.22)$$

$$\nu = 0, 1, 2, \dots$$

The harmonic oscillator is not allowed to have zero vibrational energy. The smallest allowed energy is when $\nu = 0$, and this is referred to as the *zero point energy*. This finding is in agreement with Heisenberg's uncertainty principle, for if the oscillating particle had zero energy it would have also zero momentum and would be located exactly at the position of the minimum in the potential energy curve.

The normalized vibrational wavefunctions are given by the general expression

$$\psi_\nu(\xi) = \left(\frac{1}{2^\nu \nu!} \sqrt{\frac{\beta}{\pi}} \right)^{1/2} H_\nu(\xi) \exp\left(-\frac{\xi^2}{2}\right) \quad (11.23)$$

where

$$\beta = \frac{2\pi\sqrt{k_s\mu}}{h} \quad \text{and} \quad \xi = \sqrt{\beta}x$$

The polynomials H_ν are called the *Hermite polynomials*; they are solutions of the second-order differential equation

$$\frac{d^2H}{d\xi^2} - 2\xi \frac{dH}{d\xi} + 2\nu H = 0 \quad (11.24)$$

and are most easily found from the recursion formula

$$H_{\nu+1}(\xi) = 2\xi H_\nu(\xi) - 2\nu H_{\nu-1}(\xi)$$

The first few are shown in Table 11.1.

Squares of the vibrational wavefunctions for $\nu = 0$ and $\nu = 5$ (with numerical values appropriate to $^{14}\text{N}_2$) are shown as Figures 11.10 and 11.11.

Table 11.1 The first few Hermite polynomials $H_\nu(\xi)$

ν	$H_\nu(\xi)$
0	1
1	2ξ
2	$4\xi^2 - 2$
3	$8\xi^3 - 12\xi$
4	$16\xi^4 - 48\xi^2 + 12$
5	$32\xi^5 - 160\xi^3 + 120\xi$

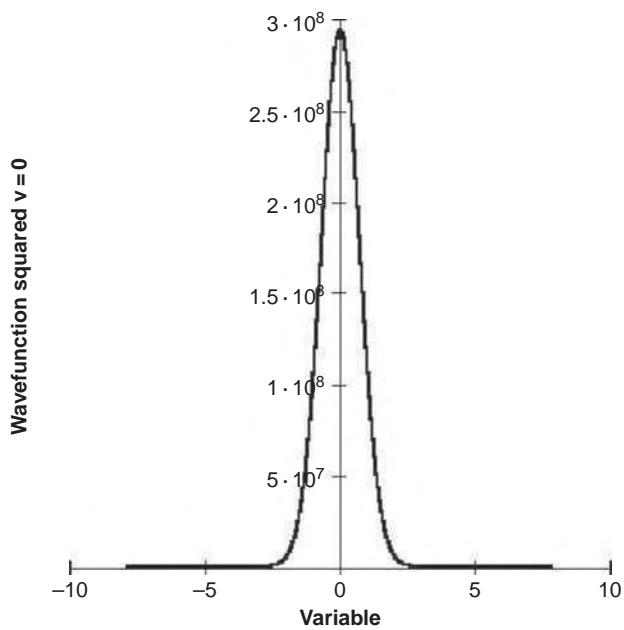


Figure 11.10 Square of vibrational wavefunction $v = 0$ for dinitrogen

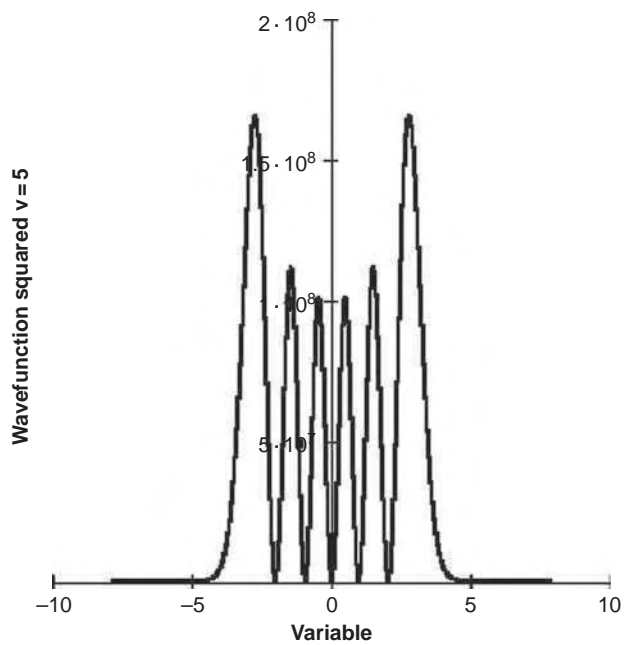


Figure 11.11 Square of vibrational wavefunction $v = 5$ for dinitrogen

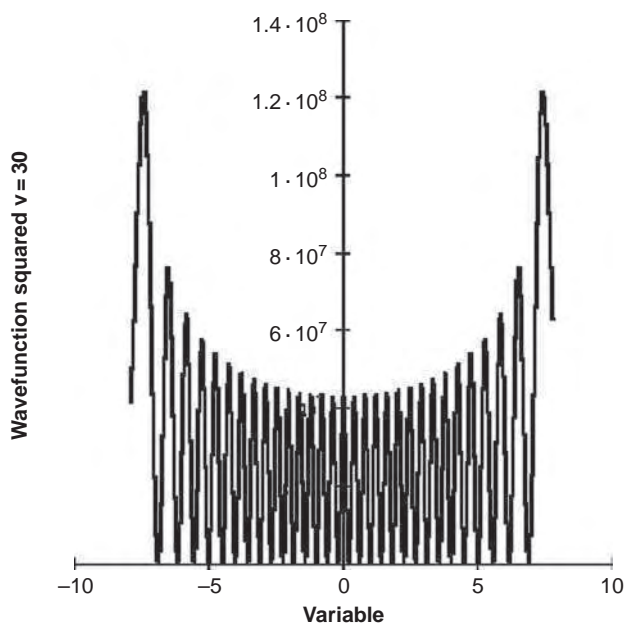


Figure 11.12 Square of vibrational wavefunction $v = 30$ for dinitrogen

The classical model predicts that the maximum of probability is when the particle is at the ends of its vibrations (when it is stationary). The quantum mechanical model seems quite different, since the $v = 0$ solution has a probability maximum half way along the bond. The operation of the correspondence principle should be clear once we consider higher quantum numbers. Even by $v = 5$ it is becoming clear that the maxima are moving toward the ends of the vibration. Examination of $v = 30$ as shown in Figure 11.12 confirms the trend, which in any case can be deduced analytically.

12

Quantum Gases

In Chapter 11, I discussed the quantum mechanical description of a particle in a selection of potential wells, including the three-dimensional infinite cubic box of side L . For a single particle of mass m , I showed that the energy is quantized and that each quantum state is described by three quantum numbers n , k and l as follows:

$$\begin{aligned}\varepsilon_{n,k,l} &= U_0 + \frac{n^2 h^2}{8mL^2} + \frac{k^2 h^2}{8mL^2} + \frac{l^2 h^2}{8mL^2} \\ \psi_{n,k,l} &= \left(\frac{2}{L}\right)^{3/2} \sin\left(\frac{n\pi x}{L}\right) \sin\left(\frac{k\pi y}{L}\right) \sin\left(\frac{l\pi z}{L}\right) \\ n, k, l &= 1, 2, 3, \dots\end{aligned}\tag{12.1}$$

We often set the constant of integration U_0 to zero, without any loss of generality. We will make this simplification for the rest of the chapter.

Consider now an ideal gas comprising N such particles, constrained to the cubic three-dimensional infinite potential well. Ideal gas particles do not interact with each other and the total wavefunction is therefore a product of one-particle wavefunctions (orbitals), as discussed in Chapter 11. Also, the total energy is the sum of the one-particle energies.

Suppose we arrange that the system is thermally isolated and then make a measurement on the system to find how the energy is distributed amongst the particles at (say) 300 K. We will find some particles with low quantum numbers and some with high. The kinetic energy of particle i is given by

$$\varepsilon_i = (n_i^2 + k_i^2 + l_i^2) \frac{h^2}{8mL^2}$$

and the total kinetic energy is therefore

$$E_{\text{kin}} = \frac{h^2}{8mL^2} \sum_{i=1}^N (n_i^2 + k_i^2 + l_i^2)\tag{12.2}$$

This can be identified with the thermodynamic internal energy U_{th}

$$U_{\text{th}} = \frac{h^2}{8mL^2} \sum_{i=1}^N (n_i^2 + k_i^2 + l_i^2) \quad (12.3)$$

We do not yet know how the kinetic energy will be divided out amongst the particles, all we know is that the summation will be a complicated function of the quantum numbers. For the sake of argument, call the summation A . Since the box is a cube of side L , the volume V of the box is L^3 and so we can write

$$U_{\text{th}} = \frac{h^2}{8m} AV^{-2/3} \quad (12.4)$$

We know from elementary chemical thermodynamics that for a thermally isolated system

$$p = - \left(\frac{\partial U_{\text{th}}}{\partial V} \right)_N$$

and so the particle in a box model gives the following relationship between pressure and volume:

$$p = \frac{2}{3} \frac{h^2}{8m} AV^{-5/3} \quad (12.5)$$

which can be rearranged together with the expression for U_{th} to give

$$pV = \frac{2}{3} U_{\text{th}} \quad (12.6)$$

That is to say, the pressure times the volume is a constant at a given temperature. Not only that, but if the internal energy of the sample (of amount n) is given by the equipartition of energy expression $\frac{3}{2}nRT$, we recover the ideal gas law.

12.1 Sharing Out the Energy

Statistical thermodynamics teaches that the properties of a system in equilibrium can be determined in principle by counting the number of states accessible to the system under different conditions of temperature and pressure. There is as ever some interesting small print to this procedure, which I will illustrate by considering three collections of different particles in the same cubic infinite potential well. First of all, I consider distinguishable particles like the ones discussed above that have mass and quantized energy. I raised the question in Chapter 11 as to whether or not we could truly distinguish one particle from another, and there is a further problem in that some particles have an intrinsic property called *spin* just as they have a mass and a charge.

We will then extend the ideas to photons (which have momentum but no mass), and finally consider the case where the particles are electrons in a metal.

We noted in Chapter 11 two differences between a one-dimensional and a three-dimensional infinite well: firstly, the three-dimensional well has degeneracies both natural and accidental; secondly, the three-dimensional quantum states tend to crowd together as the quantum numbers increase. If we consider a box of 1 dm^3 containing (for example) ^{20}Ne

atoms at 300 K, then each atom will have $\frac{3}{2}k_B T$ kinetic energy, according to the equipartition of energy principle. A short calculation suggests that we will find typically quantum numbers as high as 10^9 ; energy is certainly quantized, but the separation between the energy levels is minute in comparison to $k_B T$. Under these circumstances, it is possible to treat the energy levels and the quantum numbers as if they were continuous rather than discrete. This is called the *continuum approximation*.

Direct calculation shows how the number of quantum states lying in a given small interval increases with energy (Table 12.1).

Table 12.1 Properties of quantum states. Here $E = h^2/8mL^2$

Number of quantum states with energy lying between	
(100 and 110) E	85
(1000 and 1010) E	246
(10 000 and 10 010) E	1029
(100 000 and 100 010) E	2925
(1 000 000 and 1 000 010) E	8820

What matters in many experiments is not the precise details of an energy level diagram, rather a quantity called the *density of states* $D(\varepsilon)$ defined such that the number of states between ε and $\varepsilon + d\varepsilon$ is $D(\varepsilon)d\varepsilon$.

Figure 12.1 shows a plot of the number of states versus the square root of the energy, and it seems that there is a very good linear relationship.

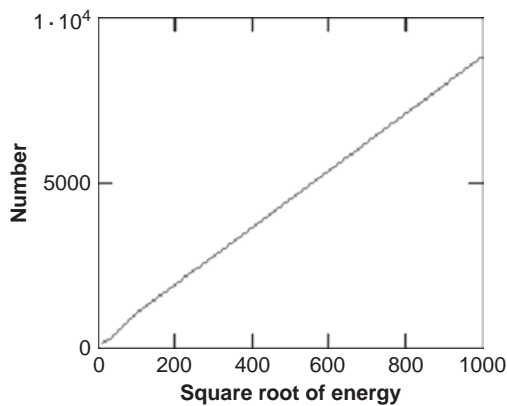


Figure 12.1 Square root relationship

I want now to establish that this is indeed the general case, and I can do this using a process called *Rayleigh counting*.

12.2 Rayleigh Counting

First of all I plot out the quantum numbers n , k and l along the x , y and z Cartesian axes as shown in Figure 12.2.

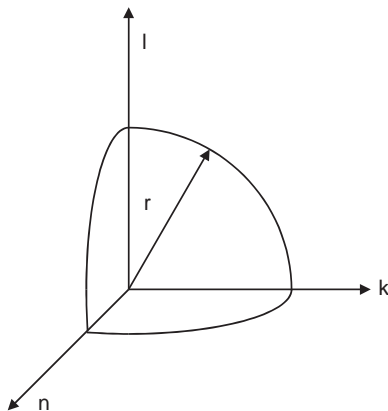


Figure 12.2 Rayleigh counting

For large quantum numbers the points will be very close together and there will be very many possible combinations of the three particular quantum numbers n , k and l that correspond to the same energy

$$\varepsilon_{n,k,l} = (n^2 + k^2 + l^2) \frac{h^2}{8mL^2}$$

According to Pythagoras' theorem, each combination will lie on the surface of a sphere of radius r where

$$r^2 = n^2 + k^2 + l^2$$

We therefore draw the positive octant of a sphere of radius r , as shown. The volume of the octant also contains all other combinations of the three quantum numbers that have energy less than or equal to the value in question. We only need consider this octant, since all quantum numbers have to be positive. The number of states with energy less than or equal to ε is

$$\frac{1}{8} \frac{4\pi}{3} (n^2 + k^2 + l^2)^{3/2}$$

or in terms of the energy

$$\frac{1}{8} \frac{4\pi}{3} \frac{8mL^2}{h^2} \varepsilon^{3/2}$$

I now draw a second octant of radius $r + dr$, shown in Figure 12.3, which contains all quantum states having energy less than or equal to $\varepsilon + d\varepsilon$. The volume of the outer sphere is

$$\frac{1}{8} \frac{4\pi}{3} \frac{8mL^2}{h^2} (\varepsilon + d\varepsilon)^{3/2}$$

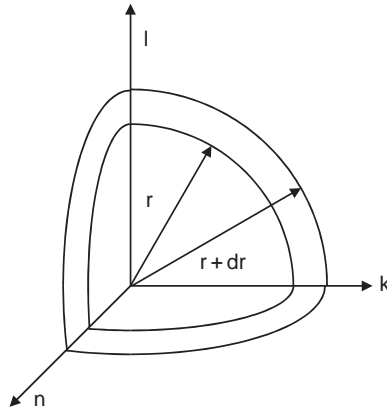


Figure 12.3 Rayleigh counting continued

The number of states having energy between ε and $\varepsilon + d\varepsilon$ is given by the volume contained between the two spheres

$$\frac{1}{8} \frac{4\pi}{3} \frac{8mL^2}{h^2} (\varepsilon + d\varepsilon)^{3/2} - \frac{1}{8} \frac{4\pi}{3} \frac{8mL^2}{h^2} \varepsilon^{3/2} \quad (12.7)$$

which is by definition $D(\varepsilon)d\varepsilon$. Use of the binomial theorem on Equation (12.7) gives

$$\begin{aligned} D(\varepsilon) d\varepsilon &= \frac{1}{8} 2\pi \frac{8mL^2}{h^2} \varepsilon^{1/2} d\varepsilon \\ &= \frac{2\pi mL^2}{h^2} \varepsilon^{1/2} d\varepsilon \end{aligned} \quad (12.8)$$

I will write the result as

$$D(\varepsilon) = B\varepsilon^{1/2} \quad (12.9)$$

12.3 Maxwell–Boltzmann Distribution of Atomic Kinetic Energies

It is possible to perform experiments that measure the kinetic energies of atoms in gaseous samples. Once again we note that there are so many particles present that the quantities of interest are the mean and standard deviation, rather than the properties of every individual particle. We normally study variation of a quantity

$$f = \frac{1}{N} \frac{dN}{d\varepsilon} \quad (12.10)$$

with energy ε at a fixed thermodynamic temperature. Here, N is the number of atoms in the sample. The related quantity

$$g = \frac{dN}{d\varepsilon}$$

is referred to as the *distribution function* of the kinetic energy. When the experiments are carried out, we find a characteristic spread of kinetic energies as shown in Figure 12.4. The distribution is independent of the type of atom, the solid curve refers to a temperature of 500 K and the dashed curve refers to a temperature of 1500 K. The peak moves to higher energy as the temperature increases, but the area under the curve remains constant.

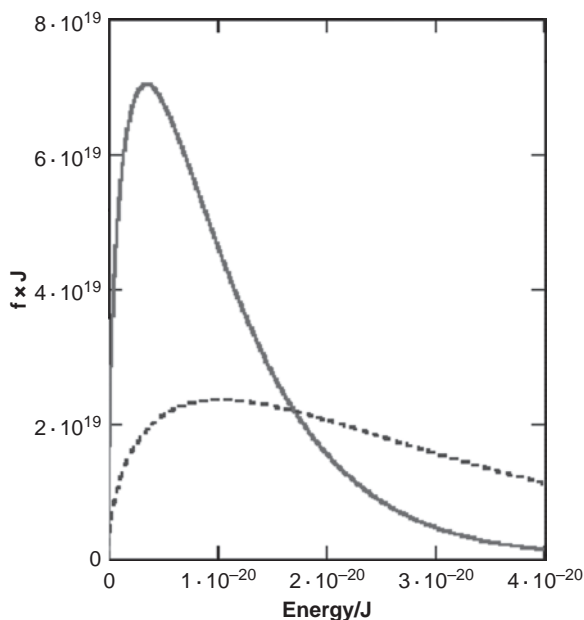


Figure 12.4 Distribution of atomic kinetic energies

The quantity $g(\varepsilon)$ gives the number of atoms whose energies lie between ε and $\varepsilon + d\varepsilon$. The number of atoms whose energies lie between ε_A and ε_B is therefore

$$\int_{\varepsilon_A}^{\varepsilon_B} g(\varepsilon) d\varepsilon$$

and the total number of atoms is given by

$$N = \int_0^{\infty} g(\varepsilon) d\varepsilon$$

I can give a very simple explanation of Figure 12.4, on the assumption that the sample of particles is constrained to a cubic three-dimensional infinite well. My explanation depends on the continuum approximation and so will only work for ordinary temperatures, for at low temperatures the spacing between quantum states will be comparable to $k_B T$ and so the continuum approximation will not apply. Also, it will not work for anything but an atom

since molecules have rotational and vibrational energies in addition to translational kinetic energy.

Apart from the density of states mentioned above, $D(\varepsilon) = B\varepsilon^{1/2}$, we have to consider the average occupancy of each quantum state. This is given by the Boltzmann factor

$$AN \exp\left(-\frac{\varepsilon}{k_B T}\right)$$

where A is a constant and N the number of atoms. Combining the two expressions we get

$$\frac{1}{N} \frac{dN}{d\varepsilon} = A \exp\left(-\frac{\varepsilon}{k_B T}\right) B\varepsilon^{1/2}$$

We can eliminate the constants A and B in terms of N by use of the equations above and we find after rearrangement

$$\frac{1}{N} \frac{dN}{d\varepsilon} = \frac{2}{\sqrt{\pi} (k_B T)^{3/2}} \exp\left(-\frac{\varepsilon}{k_B T}\right) \varepsilon^{1/2} \quad (12.11)$$

which gives exact agreement with the experimental curves. It can be established by differentiation of Equation (12.11) that the peak occurs at $\varepsilon = \frac{1}{2}k_B T$, and it can also be established by integration that the average energy per particle $\langle \varepsilon \rangle$ is $\frac{3}{2}k_B T$, in accord with the equipartition of energy principle.

12.4 Black Body Radiation

A study of black body radiation was a milestone in the path to our modern theory of quantum mechanics. The term black body radiation refers to the electromagnetic radiation emitted by a 'perfect emitter' (usually a heated cavity).

The curves in Figure 12.5 relate to the energy U emitted by a black body of volume V per wavelength λ at temperatures of 1000 K (bottom curve), 1500 K (middle curve) and 2000 K (top curve). The quantity plotted on the y axis is

$$\frac{1}{V} \frac{dU}{d\lambda}$$

and it is observed that:

- each of the curves has a peak at a certain maximum wavelength λ_{\max} ;
- λ_{\max} moves to shorter wavelength as the temperature increases.

These curves have a quite different functional form to the atomic kinetic energy curves shown in Figure 12.4. Max Planck studied the problem and was able to deduce the following expression that gives a very close fit to the experimental data:

$$\frac{1}{V} \frac{dU}{d\lambda} = \left(\frac{8\pi hc_0}{\lambda^5}\right) \frac{1}{\exp\left(\frac{hc_0}{\lambda k_B T}\right) - 1} \quad (12.12)$$

For present purposes, it proves profitable to think of the problem in the following way. We know that the heated cavity contains photons, and that the energy of a photon of

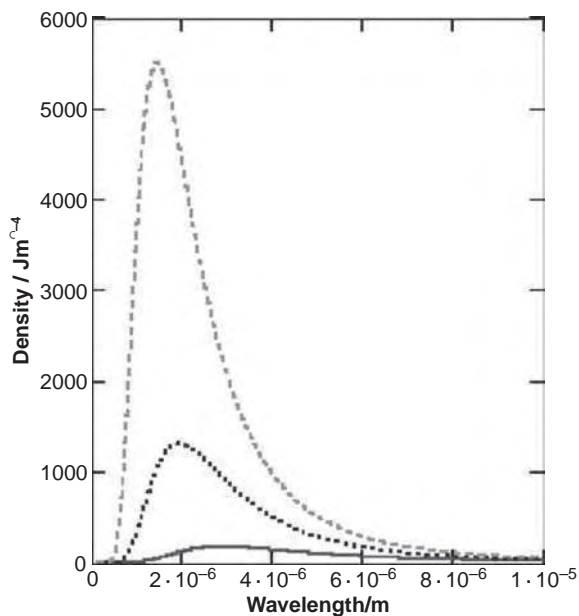


Figure 12.5 Black body radiation

wavelength λ is hc_0/λ . We also know that $dU = \varepsilon dN = \varepsilon g(\varepsilon)d\varepsilon$. The experiment actually gives a distribution function $g(\varepsilon)$ such that the number of photons with energies between ε and $\varepsilon + d\varepsilon$ is $g(\varepsilon)d\varepsilon$. Equation (12.12) can be developed to give

$$g(\varepsilon) = \frac{8\pi V}{h^3 c_0^3} \varepsilon^2 \frac{1}{\exp\left(\frac{\varepsilon}{k_B T}\right) - 1} \quad (12.13)$$

If we interpret this expression as

$$g(\varepsilon) = (\text{density of states}) \times (\text{average occupancy of a state})$$

we see that the density of states for photons is proportional to ε^2 whilst the average occupancy of a state is proportional to

$$\frac{1}{\exp\left(\frac{\varepsilon}{k_B T}\right) - 1}$$

rather than the Boltzmann factor. This factor is referred to as the *Bose factor* after S.N. Bose.

I can explain the density of states quite easily; photons have zero mass and so the energy formula (Equation (12.1)) for a particle in a three-dimensional infinite well

$$\varepsilon_{n,k,l} = \frac{h^2}{8mL^2} (n^2 + k^2 + l^2)$$

cannot apply. The correct formula turns out to be

$$\varepsilon_{n,k,l} = \frac{hc_0}{2L} (n^2 + k^2 + l^2)^{1/2} \quad (12.14)$$

There are still three quantum numbers, but the formula does not include a mass. As the energy increases, the quantum states become numerous and cram together. We can evaluate the density of states just as before by taking a narrow range of energies ε and $\varepsilon + d\varepsilon$, and counting the number of quantum states. We find

$$D(\varepsilon) \propto \varepsilon^2$$

that is to say, the photon states crowd together more quickly than for a classical particle in such a well. A detailed calculation gives the proportionality constant

$$D(\varepsilon) = \frac{4\pi V}{h^3 c_0^3} \varepsilon^2 \quad (12.15)$$

where V is the volume of the container. This explains the density of states, but not why the Bose factor is different from the Boltzmann factor.

12.5 Modelling Metals

We usually divide solids into *conductors* (materials like copper that conduct electricity), *insulators* (materials like glass that do not conduct electricity) and semiconductors that have properties in between. We visualize metallic conductors as comprising a rigid array of cations surrounded by a sea of mobile electrons, as shown in Figure 12.6.

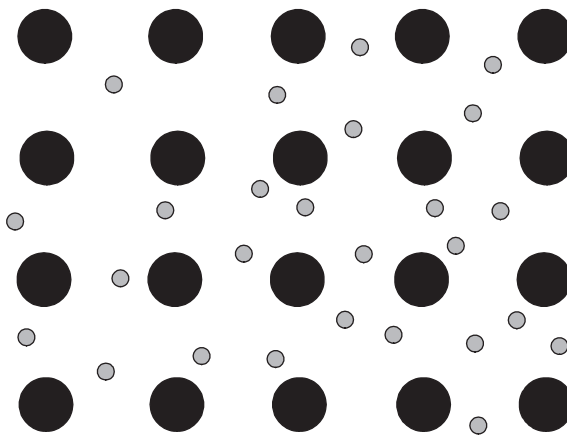


Figure 12.6 Simple model of a metallic conductor

12.5.1 The Drude Model

The first attempt to model metals was the *free electron* model developed independently by P. Drude and by Lorentz. According to this model, each metallic atom in the lattice donates its valence electrons to form a sea of electrons. So a sodium atom would lose one electron and an aluminium atom three; these conduction electrons move freely about the cation lattice points, rather like the atoms in an ideal gas.

The model had some early successes; electrical conduction can be explained because the application of an electric field to the metal will cause the conduction electrons to move from positions of high to low electrical potential and so an electric current can be established. In fact, Ohm's law can be recovered from the Drude model with a little analysis.

Eventually, the Drude model gradually fell into disrepute. An obvious failing was its inability to explain the heat capacity of metals. If the model of solid aluminium is correct and each Al atom loses three valence electrons, we might expect that these valence electrons would make a contribution to the internal energy of $3 \times (3/2)RT$, with a corresponding molar heat capacity of $(9/2)R$ (i.e. $32 \text{ J K}^{-1} \text{ mol}^{-1}$). In fact, Al obeys the Dulong–Petit rule with a molar heat capacity of about $25 \text{ J K}^{-1} \text{ mol}^{-1}$, so the equipartition of energy principle seems to overestimate the contribution from the conduction electrons.

The electrons are of course in motion, and we should ask about their energy distribution just as we did in earlier sections. There is no experimental technique that can measure the spread with the same accuracy, as can be done with photons in the black body experiment, or with particles in the kinetic theory experiments. A variation on the theme of the photoelectric effect (called high-energy photoemission) can however be used to probe electron kinetic energies in a metal. If I write the familiar expression for the distribution function $g(\varepsilon)$

$$g(\varepsilon) = (\text{density of states, } D(\varepsilon)) \times (\text{average occupancy of a state, } p(\varepsilon))$$

then it turns out that the density of states is proportional to $\varepsilon^{1/2}$, just as it was for the spread of kinetic energies in an ideal gas. The average occupancy factor is quite different from anything we have met so far:

$$p(\varepsilon) = \frac{1}{\exp\left(\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right) + 1}$$

Parameter ε_F is called the *Fermi energy*, and I have illustrated the factor for the case of a Fermi energy of 2.5 eV at 600 and 6000 K in Figure 12.7. The lower temperature shows a factor of 1 for all energies until we reach the Fermi level, when the factor becomes zero. The lower the temperature, the sharper is the cut-off. The corresponding Boltzmann factors are shown in Figure 12.8, from which it is seen that the two sets of results are radically different.

The experimental distribution function is therefore

$$g(\varepsilon) = B\varepsilon^{1/2} \frac{1}{\exp\left(\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right) + 1} \quad (12.16)$$

and the average occupancy of states factor cannot be explained by the Drude model.

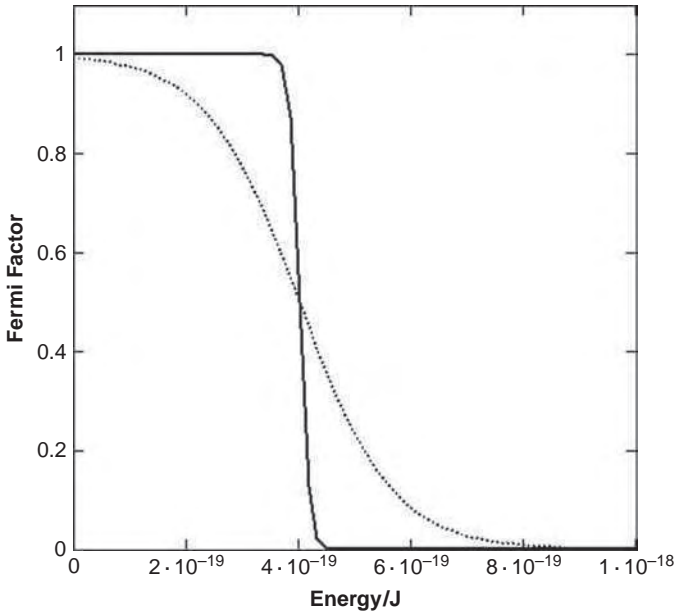


Figure 12.7 The Fermi factor

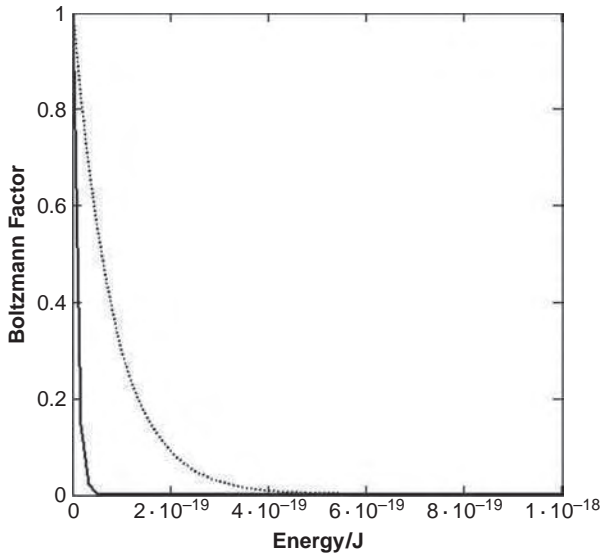


Figure 12.8 Corresponding Boltzmann factors

On the other hand, it could be argued that the ideal gas model is inappropriate because charged particles exert tremendous long-range forces on each other. Not only that, there are frequent collisions between all the charged particles. However, analysis shows that the problem lies with the *classical* Drude treatment, not with the ideal gas model.

12.5.2 The Pauli Treatment

It was W. Pauli who devised a quantum mechanical treatment of the electron gas; in Pauli's model, the conduction electrons are still treated as an ideal gas of noninteracting particles, but the analysis is done according to the rules of quantum mechanics. My starting point is the energy level diagram for a cubic three-dimensional infinite potential well (Figure 12.9).

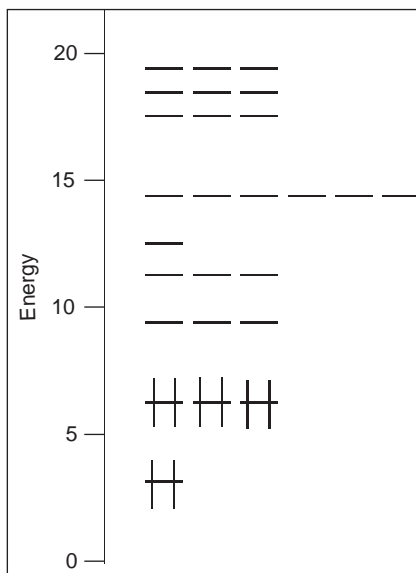


Figure 12.9 *Electrons in a metal*

Pauli assumed that the metal could be modelled as an infinite potential well, and that the electrons did not interact with each other. He also assumed that each quantum state could hold no more than two electrons. At 0 K, the N conduction electrons fill up the $N/2$ lowest energy quantum states. I have shown 8 conduction electrons in Figure 12.9, for illustration. The highest occupied quantum state corresponds to the Fermi level, and the Fermi factors are therefore

$$p(\varepsilon) = \begin{cases} 1 & \text{for } \varepsilon \leq \varepsilon_F \\ 0 & \text{for } \varepsilon > \varepsilon_F \end{cases}$$

At higher temperatures, the electrons will be spread amongst the available quantum states, but subject still to the restriction that each quantum state can hold a maximum of two electrons. For electron energies much above the Fermi level so that $\varepsilon - \varepsilon_F \gg k_B T$, the Fermi factor can be approximated by ignoring the 1 in the nominator to give

$$p(\varepsilon) \approx \exp\left(-\frac{\varepsilon - \varepsilon_F}{k_B T}\right)$$

which is very similar to the Boltzmann factor. For sufficiently large electron energies, the Fermi factor reduces to the Boltzmann factor. For electron energies below the Fermi level, the factor can be written

$$p(\varepsilon) \approx 1 - \exp\left(\frac{\varepsilon - \varepsilon_F}{k_B T}\right)$$

which simply tells us that nearly all the states are fully occupied.

We have therefore reached a point where three different types of particles in the same three-dimensional infinite potential well show very different behaviour (Table 12.2).

Table 12.2 Density of states and average occupancies

Particle	Density of states proportional to	Average occupancy proportional to
'Classical' ideal gas	$\varepsilon^{1/2}$	$\exp\left(-\frac{\varepsilon}{k_B T}\right)$
Photons in cavity	ε^2	$\frac{1}{\exp\left(\frac{\varepsilon}{k_B T}\right) - 1}$
Electrons in metal	$\varepsilon^{1/2}$	$\frac{1}{\exp\left(\frac{\varepsilon - \varepsilon_F}{k_B T}\right) + 1}$

12.6 Indistinguishability

Let me now return to the infinite one-dimensional well, with just two noninteracting particles present. I labelled the particles A and B, and showed in Chapter 11 that the solutions were

$$\begin{aligned}\Psi_{n_A, n_B} &= \frac{2}{L} \sin\left(\frac{n_A \pi x_A}{L}\right) \sin\left(\frac{n_B \pi x_B}{L}\right) \\ E_{n_A, n_B} &= (n_A^2 + n_B^2) \frac{h^2}{8mL^2} \\ n_A, n_B &= 1, 2, 3, \dots\end{aligned}\tag{12.17}$$

Suppose one particle has a quantum number of 1, whilst the other has a quantum number of 2. There are two possibilities

$$\begin{aligned}\Psi_{1,2} &= \frac{2}{L} \sin\left(\frac{1\pi x_A}{L}\right) \sin\left(\frac{2\pi x_B}{L}\right) \\ \Psi_{2,1} &= \frac{2}{L} \sin\left(\frac{2\pi x_A}{L}\right) \sin\left(\frac{1\pi x_B}{L}\right)\end{aligned}\tag{12.18}$$

and they are degenerate (they have the same energy). In such situations, it turns out that any linear combination of degenerate solutions is also an acceptable solution. This is easily seen by rewriting the problem in Hamiltonian operator form

$$\begin{aligned}\hat{H}\Psi_{1,2} &= E\Psi_{1,2} \\ \hat{H}\Psi_{2,1} &= E\Psi_{2,1} \\ \hat{H}(a\Psi_{1,2} + b\Psi_{2,1}) &= E(a\Psi_{1,2} + b\Psi_{2,1})\end{aligned}$$

In the above equation, a and b are nonzero scalars that could be complex.

We live in a world of mass production where items that come off a production line appear to be identical. A shopping trip to the supermarket might yield four tins of dog food, all with the same label, the same contents and the same price, but on close examination it usually turns out that there are differences. One tin might have a scratch, one might be slightly dented, the label might be peeling from the third and so on. In the everyday world, objects that seem at first sight to be identical may very well turn out to be distinguishable. Even if the cans were exactly identical, we could label them A, B, C and D with an indelible marker and so make them distinguishable.

Things are very different in the world of atoms and molecules. A visit to MolMart® for four hydrogen atoms would reveal that all hydrogen atoms are exactly the same; there is no question of one atom having a scratch or a dent, and it is certainly not possible to label a hydrogen atom with an indelible marker. We say that the four hydrogen atoms are *indistinguishable*. This simple observation has far-reaching consequences.

I mentioned in earlier sections that wavefunctions themselves do not have any physical interpretation, but that the modulus squared of a wavefunction is an experimentally observable probability density. So if $\psi(x)$ describes a single particle, then

$$|\psi(x)|^2 dx$$

gives the chance of finding the particle between x and $x + dx$ and this can be measured in experiments such as X-ray and neutron diffraction.

The wavefunctions $\Psi_{1,2}$ and $\Psi_{2,1}$ above involve the coordinates of two particles and in this case the physical interpretation is that, for example,

$$|\Psi(x_A, x_B)|^2 dx_A dx_B$$

represents the probability of simultaneously finding particle A between x_A and $x_A + dx_A$ with particle B between x_B and $x_B + dx_B$. If we now add the condition that the particles are truly indistinguishable then we should get exactly the same probability on renaming the particles B and A. So if we consider $(\Psi_{1,2})^2$

$$(\Psi_{1,2})^2 = \left(\frac{2}{L} \sin\left(\frac{1\pi x_A}{L}\right) \sin\left(\frac{2\pi x_B}{L}\right) \right)^2$$

it is clear that the probability density treats the two particles on a different footing and so implies that they are distinguishable. The acid test is to write down the probability density and then interchange the names of the particles. The probability density should stay the same.

The two wavefunctions $\Psi_{1,2}$ and $\Psi_{2,1}$ do not satisfy this requirement and so they are not consistent with our ideas about indistinguishability. I can illustrate the problem by plotting the squares of the wavefunctions as density maps, as in Figures 12.10 and 12.11.

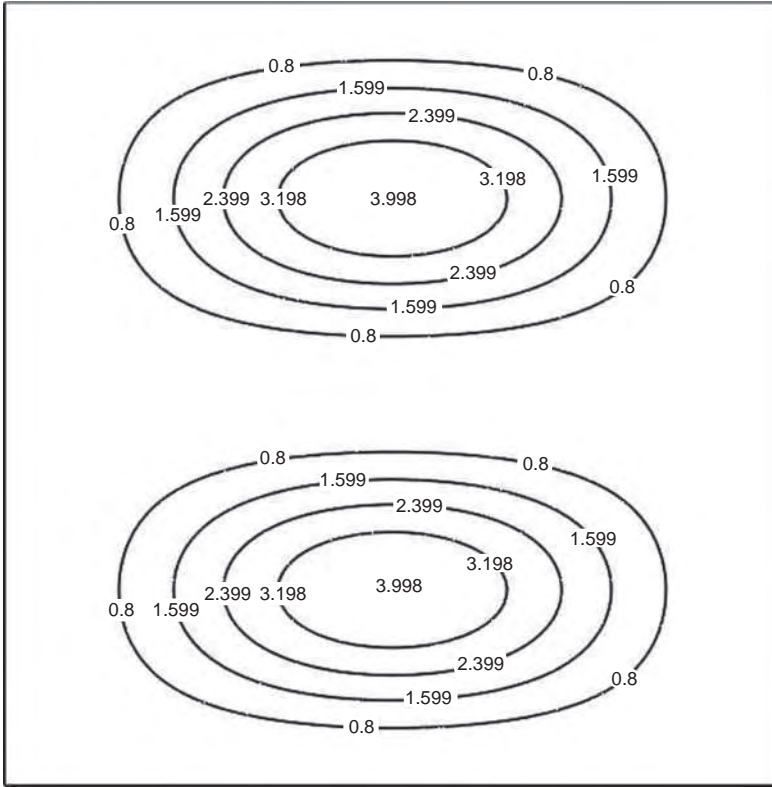


Figure 12.10 $(\Psi_{1,2})^2$ versus x_A and x_B

The plots are of the square of the wavefunction (the ‘out-of-plane’ axis) versus the x coordinates of the two electrons along the in-plane axes. The diagrams imply quite different probability densities for an arbitrary choice of the pair of coordinates; the probability contours ought to be symmetrical about their diagonals. The problem can be removed if we make use of the degeneracy mentioned above and construct a wavefunction that is either the sum or the difference of the product functions with x_A and x_B interchanged. Each of these two combinations gives a probability density that is unaffected by an interchange of the names of the two particles. Allowing for the correct normalization, these combinations are the following.

Symmetric:

$$\Psi_s(x_A, x_B) = \sqrt{\frac{1}{2}} (\Psi_{1,2}(x_A, x_B) + \Psi_{2,1}(x_A, x_B)) \quad (12.19)$$

Antisymmetric:

$$\Psi_a(x_A, x_B) = \sqrt{\frac{1}{2}} (\Psi_{1,2}(x_A, x_B) - \Psi_{2,1}(x_A, x_B)) \quad (12.20)$$

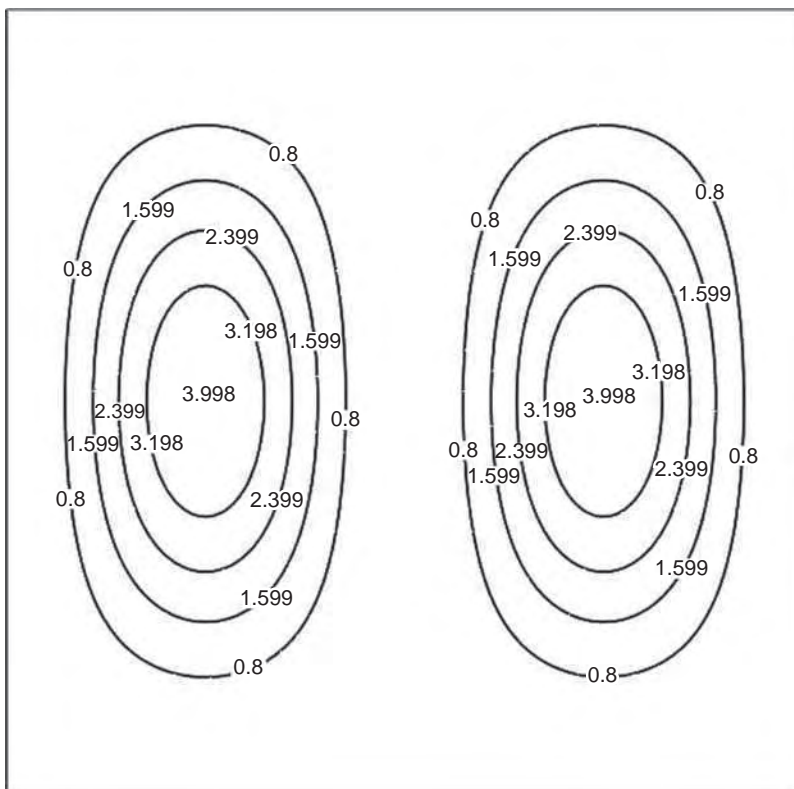


Figure 12.11 $(\Psi_{2,1})^2$ versus x_A and x_B

The labels *symmetric* and *antisymmetric* arise from the fact that interchange of the particle names leaves the first combination unchanged but leads to a reversal of sign in the second combination.

The resulting probability maps are shown in Figures 12.12 and 12.13.

In the symmetric state the maximum probability occurs at the points for which $x_A = x_B = L/4$ or $3L/4$ and quite generally the probability is large only if the difference between the particles is small.

The antisymmetric state is quite different; the probability is zero everywhere along the diagonal (which corresponds to $x_A = x_B$) and the probability maxima occur at positions where the two particles are far apart. I must emphasize that this apparent attraction and repulsion between particles has nothing whatever to do with the forces between them; indeed, we assumed at the start that the particles did not interact with each other. It certainly is not an additional force in nature; it is simply to do with the indistinguishability of particles and the symmetry of the wavefunction describing them. Like the zero-point vibrational energy, this is a purely quantum mechanical effect and has no 'explanation' in classical physics.

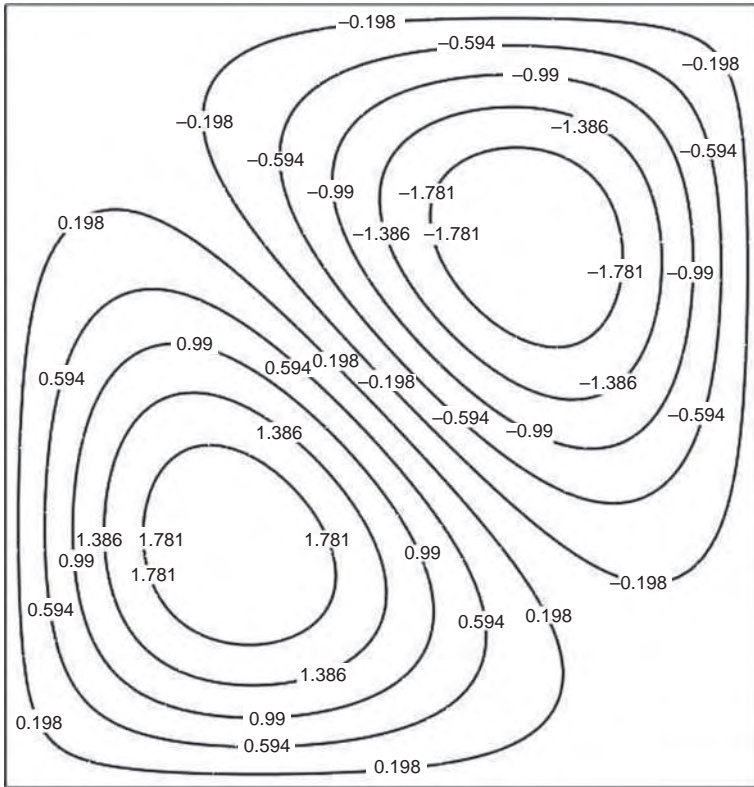


Figure 12.12 Symmetric wavefunction squared

12.7 Spin

It turns out as an experimental fact that many particles have an intrinsic property called *spin* just as they have a mass, a charge and so on. If the particles concerned are electrons, then we refer to the phenomenon as *electron spin*. Electron spin is expressed in two possible configurations usually referred to as *spin up* and *spin down* with respect to a given axis of quantization. I will have more to say about electron spin in a later chapter; it does not appear from the Schrödinger treatment, just accept it as an experimental fact for the minute.

The two states are written conventionally α and β and authors often use the symbol s for the spin variable. So in the case of electrons, we might write $\alpha(s_A)$ to mean that particle A was in the spin up state and $\beta(s_A)$ to mean that particle A was in the spin down state. In the case of two electrons we have four possibilities and we label the quantum states individually $\alpha(s_A)$, $\beta(s_A)$, $\alpha(s_B)$ and $\beta(s_B)$. The combination $\alpha(s_A)\beta(s_B)$ describes a situation where electron A has spin up and electron B has spin down and there are four possible combination states

$$\alpha(s_A)\alpha(s_B); \quad \alpha(s_A)\beta(s_B); \quad \beta(s_A)\alpha(s_B); \quad \beta(s_A)\beta(s_B)$$

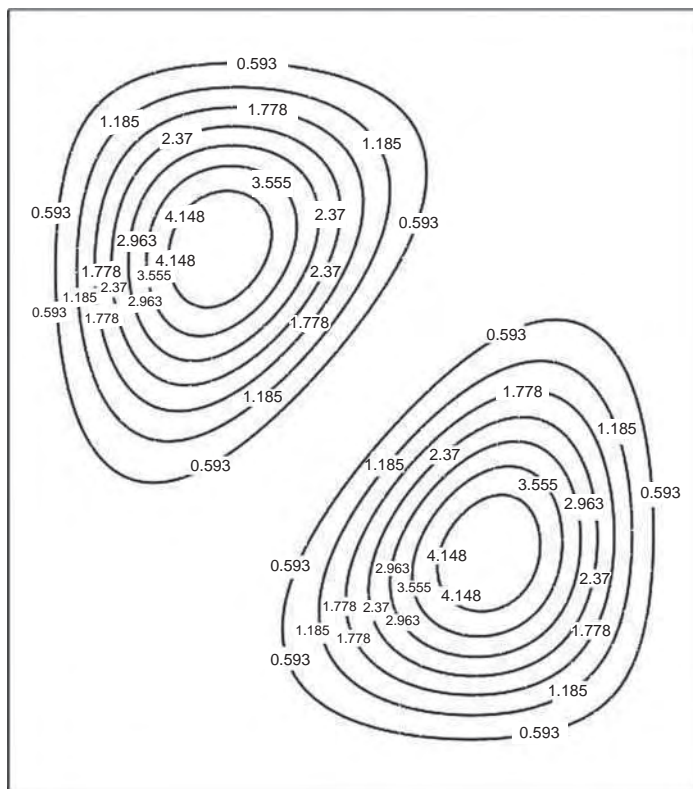


Figure 12.13 *Antisymmetric wavefunction squared*

The first and the last are automatically symmetric with respect to interchange of the two electron names, but the middle two are not satisfactory because they imply that the two electrons can somehow be treated differently. Bearing the discussion of Section 12.6 in mind, we construct two other spin functions that have definite spin symmetry properties.

Symmetric:

$$\frac{1}{\sqrt{2}} (\alpha(s_A)\beta(s_B) + \beta(s_A)\alpha(s_B))$$

Antisymmetric:

$$\frac{1}{\sqrt{2}} (\alpha(s_A)\beta(s_B) - \beta(s_A)\alpha(s_B))$$

If we had been considering two spinless particles, for example two alpha particles, there would be no spin states and the above complications would not have arisen. We now have to combine together the space and the spin wavefunctions. We do this by multiplying

together a space part and a spin part, and it should be clear that there are two possible results (Table 12.3) and to decide between them we have to resort to experiment. For many-electron systems, the only states that are ever observed are those that correspond to antisymmetric total wavefunctions. Heisenberg (1926) recognized this fact and this led to a detailed understanding of the electronic spectrum and structure of the helium atom. The condition of antisymmetry eliminates half of the possible states that would otherwise be possible.

Table 12.3 *Combining space and spin functions*

Space part	Spin part	Overall wavefunction
Symmetric	Symmetric	Symmetric
Antisymmetric	Symmetric	Antisymmetric
Symmetric	Antisymmetric	Antisymmetric
Antisymmetric	Antisymmetric	Symmetric

(When considering more than two particles, we have to consider all possible permutations amongst the space and spin functions.)

12.8 Fermions and Bosons

Not all wavefunctions for identical particles are antisymmetric, but detailed studies show that, for a given kind of particle, states of only one overall symmetry exist; the states are *either* all symmetric *or* all antisymmetric with respect to interchange of particle names. Particles whose overall wavefunction is antisymmetric are called *fermions* (after E. Fermi), and particles whose overall wavefunction is symmetric are called *bosons* (after S.N. Bose). Typical fermions and bosons are shown in Table 12.4.

Table 12.4 *Fermions and bosons*

Fermion	Boson
Electron	Photon
Proton	Deuteron
Neutron	Alpha particle
³ He atom	⁴ He atom

12.9 Pauli Exclusion Principle

The famous *Pauli exclusion principle* is a statement of our discussion above; in its general form it says that fermion wavefunctions must be antisymmetric to exchange of the names of two fermions, whilst boson wavefunctions must be symmetric. The principle is stated in different ways in different books. For example, the world's most popular physical chemistry

text (Atkins 2006) says: ‘No more than two electrons may occupy any orbital, and if two do occupy it their spin directions must be opposite.’

In principle we must invoke the symmetry or antisymmetry of a wavefunction when dealing with systems of two or more identical particles. For example, if we wish to describe two ground-state hydrogen atoms, the total wavefunction must be antisymmetric to exchange of the names of the two electrons, even if an infinite distance separates the two atoms. Common sense suggests that this is not necessary, and the criterion is to ask by how much the two wavefunctions overlap each other. If the overlap is appreciable, then we must worry about the exclusion principle. If the overlap is negligible, we can disregard the symmetry or otherwise of the total wavefunction.

12.10 Boltzmann’s Counting Rule

We must now return to the three probabilities discussed in Section 12.5. Consider a simple case of three particles A, B and C, each of which has a different energy.

Figure 12.14 shows three of the six possible arrangements of the three *distinguishable* particles, and, according to the Boltzmann counting procedure, we take account of all six and they are all weighted equally when we calculate the number of ways W .

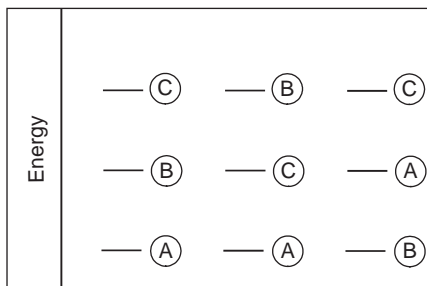


Figure 12.14 Boltzmann’s counting method

This procedure is not consistent with the principles of quantum mechanics discussed above. We certainly cannot label particles and make them distinguishable, and it is not meaningful to enquire which particle is in one particular state. All we can do is to say how many particles are in each state, and instead of six possibilities there is just one (with one particle in each state).

We then have to consider whether the particles are fermions or bosons; if they are fermions with spin $1/2$ (like electrons), then each quantum state can hold no more than two particles, one of α and one of β spin. If they are bosons, then the exclusion principle does not apply and all the particles can crowd together in any one state, should they so wish. This is the basis of the three different probabilities discussed in Section 12.5.

That leaves just one loose end: if all particles are either fermions or bosons, why does the Boltzmann distribution work so well in explaining the properties of ‘ordinary’ gases?

The overwhelming experimental evidence is that the particles in an ordinary gas can be treated as if they were distinguishable particles, subject to the Boltzmann probability law. The answer is quite simply that under many normal circumstances the quantum states are very sparsely occupied and so fermions do not need to be aware of the exclusion principle, whilst bosons cannot display their ability to all crowd together into the lowest state.

References

- Atkins, P.W. (2006) *Atkins' Physical Chemistry*, 8th edn, Oxford University Press, Oxford.
Heisenberg, W. (1926) *Z. Physik*, **39**, 499.

13

One-Electron Atoms

The hydrogen atom occupies an important place in the history of molecular modelling and quantum theory. It is the simplest possible atom, with superficially simple properties, yet it was a thorn in the side of late nineteenth and early twentieth century science. Rutherford had discovered the nucleus by 1911, and had a rough idea of its size from his scattering studies and so people pictured the hydrogen atom as a negative charge (the electron) making a circular orbit with constant angular velocity round the nucleus (a positive charge). This is shown in Figure 13.1.

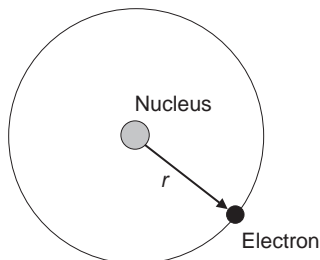


Figure 13.1 A one-electron atom

I should mention a small technical point: if we want to eliminate the overall translational motion of the atom from our discussion, the atomic system, nucleus plus electron, actually rotates round the centre of gravity. This rotational motion of the electron (mass m_e) and the nucleus (mass M) is equivalent to rotation of a single particle with reduced mass μ about the origin. Here

$$\frac{1}{\mu} = \frac{1}{M} + \frac{1}{m_e}$$

On account of its circular motion, the electron is accelerating. According to Maxwell's classical electromagnetic theory, accelerating electric charges emit energy as electromagnetic radiation and so the atom ought not to be stable, since it should lose all its energy as it orbits the nucleus and emits radiation. Also, the frequency of this emitted radiation should equal the frequency of revolution.

13.1 Atomic Spectra

The problems were that atoms are certainly stable, and that a hydrogen atom can emit or absorb electromagnetic radiation having only certain specific wavelengths. For example, Balmer had discovered in 1885 that a series of five spectral lines in the visible region with wavelengths $\lambda = 656, 486, 434, 410$ and 397 nm fitted the formula

$$\lambda_n(\text{nm}) = 364.6 \frac{n^2}{n^2 - 4} \quad (13.1)$$

It was eventually realized that the reciprocal of the wavelength was the more fundamental quantity and that the equation was better written

$$\frac{1}{\lambda_n} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right) \quad (13.2)$$

(where R_H is the *Rydberg constant for hydrogen*). Similar series of spectral lines were observed for He^+ where the proportionality constant is (very nearly) $\frac{1}{4}R_H$, for Li^{2+} (with a proportionality constant of almost $\frac{1}{9}R_H$) and so on. The important point is that the value of $1/\lambda$ for every spectral line is a difference between two terms that we now know correspond directly to particular energy levels in the atom.

13.1.1 Bohr's Theory

Neils Bohr is reputed to have said that once he saw the Balmer formula, the whole problem of the hydrogen atom became clear to him. In any event, in 1913 he was able to devise a spectacularly successful theory that could account for the spectral lines. He equated the centrifugal force for circular motion with the attractive electrostatic force between the nucleus (charge Ze) and the electron (charge $-e$):

$$\frac{Ze^2}{4\pi\epsilon_0 r^2} = \frac{\mu v^2}{r} \quad (13.3)$$

This gives two unknowns, r and the speed v . He assumed that the orbital angular momentum was quantized in units of $h/2\pi$:

$$\mu v r = n \frac{h}{2\pi} \quad (13.4)$$

where the quantum number n could take values $1, 2, 3, \dots$. This assumption is exactly equivalent to requiring that the circumference of the orbit is just the right value to permit the electron to behave as a standing wave (i.e. a de Broglie wave). In other words,

the circumference has to be an integral number n of de Broglie wavelengths, that is to say

$$2\pi r = n \frac{h}{p} \quad (13.5)$$

where \mathbf{p} is the momentum ($= \mu\mathbf{v}$). Simultaneous solution of Equations (13.3) and (13.4) gives allowed values for the radius and the speed

$$\begin{aligned} r_n &= \frac{\varepsilon_0 h^2 n^2}{\pi \mu Z e^2} \\ v_n &= \frac{Z e^2}{2 \varepsilon_0 h n} \end{aligned} \quad (13.6)$$

and also quantized energies given by

$$\begin{aligned} \varepsilon_n &= -\frac{Z e^2}{4\pi \varepsilon_0 r_n} + \frac{1}{2} \mu v_n^2 \\ &= -\frac{\mu Z^2 e^4}{8 h^2 \varepsilon_0^2} \frac{1}{n^2} \end{aligned} \quad (13.7)$$

Bohr referred to these orbits as *stationary states* because according to his theory the electron did not radiate when it was in any one such state. When the electron passed from one state (the initial state) to another (the final state), radiation was emitted or absorbed according to the Planck–Einstein formula

$$|\varepsilon_f - \varepsilon_i| = \frac{hc_0}{\lambda} \quad (13.8)$$

Bohr's theory predicted the existence of many series of spectral lines for the hydrogen atom (for which $Z = 1$) whose wavelengths fitted the generic formula

$$\begin{aligned} \frac{1}{\lambda_{if}} &= R_H \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right) \\ n_f &= 1, 2, 3, \dots \\ n_i &> n_f \end{aligned}$$

The Balmer series corresponds to $n_f = 2$, and the series corresponding to $n_f = 3$ had already been observed in the infrared region of the electromagnetic spectrum by Paschen. Soon after Bohr published his theory, Lyman identified the series with $n_f = 1$ in the ultraviolet. The $n_f = 4$ (Brackett) and $n_f = 5$ (Pfund) series were also identified. Bohr's theory also predicted a first ionization energy (given by $R_H hc_0$, approximately 2.16806×10^{-18} J or 13.5 eV) that agreed well with experiment, and a physically reasonable value (52.9 pm) for the radius of the lowest energy orbit.

Many series with high values of n have since been observed. For example, radio astronomers are very interested in the $n_i = 167$ to $n_f = 166$ emission which occurs at a wavelength of 21.04 cm.

Bohr's theory gives the following expression for R_H , in agreement with experiment:

$$R_H = \frac{\mu e^4}{8 \varepsilon_0^2 h^3 c_0}$$

13.2 Correspondence Principle

Bohr made much use of the correspondence principle discussed in Chapter 11, which says that quantum mechanical results must tend to those obtained from classical physics, in the limit of large quantum numbers. For example, if we consider the transition from level n to level $n - 1$, the emitted frequency is

$$\nu = \frac{\mu e^4 Z^2}{8\varepsilon_0^2 h^3} \left(\frac{1}{(n-1)^2} - \frac{1}{n^2} \right)$$

For large n this approximates to

$$\nu = \frac{\mu e^4 Z^2}{8\varepsilon_0^2 h^3} \frac{2}{n^3} \quad (13.9)$$

According to Maxwell's electromagnetic theory, an electron moving in a circular orbit should emit radiation with a frequency equal to its frequency of revolution $\nu/2\pi r$. Using the Bohr expressions for ν and r we deduce an expression in exact agreement with the frequency given by Equation (13.9).

Despite its early successes, Bohr's theory had many failings. For example, it could not explain the structure and properties of a helium atom. Many ingenious attempts were made to improve the model, for example by permitting elliptic orbits rather than circular ones, but the theory has gradually faded into history. It is sometimes referred to as the *old quantum theory*.

13.3 Infinite Nucleus Approximation

I started the chapter by correctly considering the motion of a one-electron atom about the centre of mass, and pointed out that this was equivalent to the motion of a single particle of reduced mass

$$\frac{1}{\mu} = \frac{1}{M} + \frac{1}{m_e}$$

Because the mass of the nucleus is so much greater than that of the electron, the reduced mass of the atom is almost equal to (and slightly less than) the mass of the electron. For that reason, workers in the field often treat the nucleus as the centre of coordinates and the electron as rotating round the nucleus, which is taken to have infinite mass; this is called the *infinite nucleus approximation*. It is just a small correction, the two masses μ and m_e are equal to 1 part in 10^4 which is usually good enough, but as we have seen, atomic spectroscopic data are known to incredible accuracy and it is sometimes necessary to take account of this difference. We write, for an infinite mass nucleus,

$$\varepsilon_n = -\frac{m_e e^4}{8h^2 \varepsilon_0^2} \frac{1}{n^2} \quad (13.10)$$

$$R_\infty = \frac{m_e e^4}{8\varepsilon_0^2 h^3 c_0} \quad \text{and} \quad a_0 = \frac{\varepsilon_0 h^2}{\pi m_e e^2}$$

where a_0 is called the *first Bohr radius* and R_∞ is 'the' Rydberg constant, with value $R_\infty = 10\,973\,731.5685458 \text{ m}^{-1}$.

13.4 Hartree's Atomic Units

One problem we encounter in dealing with atomic and molecular properties is that large powers of 10 invariably appear in equations, together with complicated combinations of physical constants such as those in Equations (13.10). For simplicity, we tend to work in a system of units called the Hartree units, or atomic units (we will soon come across the two Hartrees in Chapter 14). The 'atomic unit of length' is the *bohr* and is equal to a_0 above. The 'atomic unit of energy' is the *hartree* and is equal to $2R_\infty hc_0$. It is written E_h and it is also equal to the mutual potential energy of a pair of electrons separated by distance a_0 :

$$E_h = \frac{1}{4\pi\epsilon_0} \frac{e^2}{a_0}$$

Other atomic units can be defined, as shown in Table 13.1.

Table 13.1 Hartree's atomic units

Physical quantity	Symbol	X	Value of X
Length	l, x, y, z, r	a_0	$5.2918 \times 10^{-11} \text{ m}$
Mass	m	m_e	$9.1094 \times 10^{-31} \text{ kg}$
Energy	ϵ	E_h	$4.3598 \times 10^{-18} \text{ J}$
Charge	Q	e	$1.6022 \times 10^{-19} \text{ C}$
Electric dipole moment	p_e	ea_0	$8.4784 \times 10^{-30} \text{ C m}$
Electric quadrupole moment	θ_e	ea_0^2	$4.4866 \times 10^{-40} \text{ C m}^2$
Electric field	E	$E_h e^{-1} a_0^{-1}$	$5.1422 \times 10^{11} \text{ V m}^{-1}$
Electric field gradient	$-V_{zz}$	$E_h e^{-1} a_0^{-2}$	$9.7174 \times 10^{21} \text{ V m}^{-2}$
Magnetic induction	B	$(h/2\pi)e^{-1} a_0^{-2}$	$2.3505 \times 10^5 \text{ T}$
Electric dipole polarizability	α	$e^2 a_0^2 E_h^{-1}$	$1.6488 \times 10^{-41} \text{ C}^2 \text{ m}^2 \text{ J}^{-1}$
Magnetizability	ξ	$e^2 a_0^3 m_e^{-1}$	$7.8910 \times 10^{-29} \text{ JT}^{-2}$

The Born interpretation of quantum mechanics tells us that $\Psi^*(\mathbf{r})\Psi(\mathbf{r})d\tau$ represents the probability that we will find the particle whose spatial coordinate is \mathbf{r} within the volume element $d\tau$. Probabilities are real numbers and so the dimensions of the wavefunction must be (length) $^{-3/2}$. In the atomic system of units, the unit of wavefunction is therefore $a_0^{-3/2}$ (or (bohr) $^{-3/2}$).

13.5 Schrödinger Treatment of the Hydrogen Atom

The wavefunction of a hydrogen atom Ψ_{tot} depends on the coordinates of the electron x, y, z and the nucleus X, Y, Z . The time-independent Schrödinger equation is

$$-\frac{\hbar^2}{8\pi^2 m_e} \left(\frac{\partial^2 \Psi_{\text{tot}}}{\partial x^2} + \frac{\partial^2 \Psi_{\text{tot}}}{\partial y^2} + \frac{\partial^2 \Psi_{\text{tot}}}{\partial z^2} \right) - \frac{\hbar^2}{8\pi^2 M} \left(\frac{\partial^2 \Psi}{\partial X^2} + \frac{\partial^2 \Psi}{\partial Y^2} + \frac{\partial^2 \Psi}{\partial Z^2} \right) + U\Psi_{\text{tot}} = \epsilon_{\text{tot}}\Psi_{\text{tot}} \quad (13.11)$$

where m_e is the electron mass, M the nuclear mass and U the mutual electrostatic potential energy of the nucleus and the electron, given by

$$U = \frac{-Ze^2}{4\pi\epsilon_0} \frac{1}{\sqrt{(x-X)^2 + (y-Y)^2 + (z-Z)^2}} \quad (13.12)$$

I have temporarily added a subscript 'tot' to show that we are dealing with the total atom, nucleus plus electron, at this stage. The equation is not at all simple in this coordinate system and it proves profitable to make a change of variable. We take the coordinate origin as the centre of gravity of the atom (coordinates x_c, y_c, z_c) and use spherical polar coordinates for the nucleus and the electron. If the spherical polar coordinates of the electron are a, θ, ϕ and those of the nucleus b, θ, ϕ where

$$a = \frac{M}{M + m_e} r$$

$$b = -\frac{m_e}{M + m_e} r$$

then, if μ is the reduced mass

$$\mu = \frac{m_e M}{m_e + M}$$

$$X = x_c - \frac{\mu}{M} r \sin \theta \cos \phi \quad x = x_c + \frac{\mu}{m_e} r \sin \theta \cos \phi$$

$$Y = y_c - \frac{\mu}{M} r \sin \theta \sin \phi \quad y = y_c + \frac{\mu}{m_e} r \sin \theta \sin \phi$$

$$Z = z_c - \frac{\mu}{M} r \cos \theta \quad z = z_c + \frac{\mu}{m_e} r \cos \theta$$
(13.13)

With this change of variable we find

$$-\frac{\hbar^2}{8\pi^2(m_e + M)} \left(\frac{\partial^2 \Psi_{\text{tot}}}{\partial x_c^2} + \frac{\partial^2 \Psi_{\text{tot}}}{\partial y_c^2} + \frac{\partial^2 \Psi_{\text{tot}}}{\partial z_c^2} \right) \quad (13.14)$$

$$-\frac{\hbar^2}{8\pi^2\mu} \left(\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \Psi_{\text{tot}}}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \Psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \Psi_{\text{tot}}}{\partial \phi^2} \right) + U \Psi_{\text{tot}} = \epsilon_{\text{tot}} \Psi_{\text{tot}}$$
(13.15)

and also in this coordinate system

$$U = -\frac{1}{4\pi\epsilon_0} \frac{Ze^2}{r}$$

It is apparent that we can separate the wave equation into two equations, one referring to the centre of mass and involving the total mass of the atom, and the other containing r, θ and ϕ . We put $\Psi_{\text{tot}} = \psi_e \psi_c$ and follow through the separation of variables argument to obtain

$$-\frac{\hbar^2}{8\pi^2(m_e + M)} \left(\frac{\partial^2 \psi_c}{\partial x_c^2} + \frac{\partial^2 \psi_c}{\partial y_c^2} + \frac{\partial^2 \psi_c}{\partial z_c^2} \right) = (\varepsilon_{\text{tot}} - \varepsilon_e) \psi_c \quad (13.16)$$

$$-\frac{\hbar^2}{8\pi^2\mu} \left(\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi_c}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi_c}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi_c}{\partial \phi^2} \right) + U \psi_c = \varepsilon_e \psi_c \quad (13.17)$$

The first equation relates to the translation of the atom as a whole, and I have dealt with such equations in earlier chapters. The second equation is usually called the *electronic equation*.

It should be clear from this discussion that in the treatment of any atomic or molecular system the translational degree(s) of freedom may always be separated from the internal degrees of freedom and so need not be considered in general. Also, from now on, I will drop the subscript 'e' from the electronic equation.

In the special case of a one-electron atom, the electronic wavefunction depends only on the coordinates of (the) one electron and so it is technically described as an *atomic orbital*. In view of the spherical symmetry, we might expect that it would prove profitable to consider a further separation of variables

$$\psi(r, \theta, \phi) = R(r) Y(\theta, \phi) \quad (13.18)$$

This proves to be the case and $R(r)$ is usually called the *radial function* (although I should warn you that many authors call $P(r) = rR(r)$ the radial function; we will see why this proves a useful choice shortly). The Y 's turn out to be eigenfunctions of the orbital angular momentum, and again we will see why in due course.

It turns out that most of the wavefunction solutions of Equation (13.18) are not spherically symmetrical, and you may find this odd given that the mutual potential energy U depends only on the scalar distance r (we call such fields *central fields*). The fact is that $U(r)$ being spherically symmetrical does not imply that the solutions are also spherical; the gravitational field is also a central field, and you probably know that planetary motion is not spherically symmetrical.

13.6 Radial Solutions

Before launching into a discussion of the full solutions, I want to spend a little time considering just those solutions that depend only on r . That is, those solutions that have a constant angular part; this is not the same thing as finding $R(r)$ for the general case. I will also make the infinite nucleus approximation from this point on.

A little rearrangement and manipulation gives, for functions that depend only on r ,

$$-\frac{\hbar^2}{8\pi^2 m_e} \frac{d^2(rR)}{dr^2} - \frac{Ze^2}{4\pi \varepsilon_0 r} (rR) = \varepsilon (rR) \quad (13.19)$$

The form of this equation suggests that we make the substitution $P = rR$ which explains why some authors focus on $P(r)$ rather than $R(r)$ and refer to $P(r)$ as the radial function. This substitution gives

$$-\frac{h^2}{8\pi^2 m_e} \frac{d^2 P}{dr^2} - \frac{Ze^2}{4\pi \epsilon_0 r} P = \epsilon P$$

It is usual in many scientific and engineering applications to try and simplify such complicated equations by use of ‘reduced variables’. In this case we divide each variable by its atomic unit, so for example

$$\begin{aligned} r_{\text{red}} &= r/a_0 \\ \epsilon_{\text{red}} &= \epsilon/E_h \\ \frac{d^2}{dr_{\text{red}}^2} &= a_0^2 \frac{d^2}{dr^2} \end{aligned}$$

and simplify to give

$$-\frac{1}{2} \frac{d^2 P_{\text{red}}}{dr_{\text{red}}^2} - \frac{Z}{r_{\text{red}}} P_{\text{red}} = \epsilon_{\text{red}} P_{\text{red}}$$

It gets a bit tedious writing the subscript ‘red’, so most authors choose to state their equations in dimensionless form such as

$$-\frac{1}{2} \frac{d^2 P}{dr^2} - \frac{Z}{r} P = \epsilon P \quad (13.20)$$

with the tacit understanding that the variables are reduced ones and therefore dimensionless. Sometimes authors say (incorrectly) that the equation is written ‘in atomic units’, and occasionally we come across statements of the impossible (e.g. $m_e = 1$ and $h/2\pi = 1$).

We therefore have to solve the differential Equation (13.20) and we follow standard mathematical practice by first examining the limiting behaviour for large r :

$$\frac{d^2 P}{dr^2} = -2\epsilon P$$

If we limit the discussion to bound states (for which ϵ is negative), we note that

$$P(r) \approx \exp(-\beta r) \quad (13.21)$$

where β is a positive real constant. This suggests that we look for a wavefunction of the form

$$P(r) = \Phi(r) \exp(-\beta r) \quad (13.22)$$

where $\Phi(r)$ is a polynomial in r . Substitution of Equation (13.22) into the radial Equation (13.20) then gives

$$\frac{d^2 \Phi}{dr^2} - 2\beta \frac{d\Phi}{dr} + \left(\beta^2 + 2\epsilon + \frac{2Z}{r} \right) \Phi = 0 \quad (13.23)$$

Suppose for the sake of argument we choose

$$\Phi(r) = \alpha_2 r^2 + \alpha_1 r + \alpha_0 \quad (13.24)$$

where the coefficients α have to be determined. Substitution of Equation (13.24) into the differential Equation (13.23) gives

$$2\alpha_2 - 2\beta(2\alpha_2 r + \alpha_1) + \left(\beta^2 + 2\epsilon + \frac{2Z}{r} \right) \times (\alpha_2 r^2 + \alpha_1 r + \alpha_0) = 0$$

For this equation to hold for all values of the variable r , the coefficients of all powers of r must separately vanish; thus

$$\begin{aligned}(\beta^2 + 2\varepsilon)\alpha_2 &= 0 \\(\beta^2 + 2\varepsilon)\alpha_1 + 2(Z - 2\beta)\alpha_2 &= 0 \\(\beta^2 + 2\varepsilon)\alpha_0 + 2(Z - \beta)\alpha_1 + 2\alpha_2 &= 0 \\2Z\alpha_0 &= 0\end{aligned}$$

It follows that

$$\begin{aligned}\varepsilon &= -\frac{\beta^2}{2} \\ \beta &= \frac{Z}{2} \\ Z\alpha_1 + 2\alpha_2 &= 0 \\ \alpha_0 &= 0\end{aligned}$$

The first two equations give an energy of $-Z^2/8$ and because ε is actually the energy divided by an atomic unit of energy E_h we have

$$\begin{aligned}\varepsilon &= -\frac{Z^2}{8}E_h \\ &= -\frac{1}{2}\frac{Z^2}{2^2}E_h\end{aligned}$$

The third and fourth equations determine all the coefficients but one, which we are free to choose. If we take $\alpha_1 = 1$ then $\alpha_2 = -Z/2$ and we have

$$\Phi(r) = -\frac{Z}{2}r^2 + r \quad (13.25)$$

13.7 Atomic Orbitals

We must now examine the general case where the wavefunction depends on three variables. Substitution of

$$\psi(r, \theta, \phi) = R(r) Y(\theta, \phi) \quad (13.26)$$

into the electronic Schrödinger equation (Equation (13.17)) gives

$$\frac{1}{R} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + \frac{8\pi^2 m_e}{h^2} (\varepsilon - U) r^2 = -\frac{1}{Y \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) - \frac{1}{Y \sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} \quad (13.27)$$

By our usual argument, both sides of this equation must be equal to a constant that we will call λ . We then have two equations

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \left(\frac{8\pi^2 m_e}{h^2} (\varepsilon - U) - \frac{\lambda}{r^2} \right) R = 0$$

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 Y}{\partial \phi^2} + \lambda Y = 0$$
(13.28)

I have discussed the second equation in Appendix A when considering angular momentum; the allowed solutions are the spherical harmonics $Y_{l,m_l}(\theta, \phi)$, where $\lambda = l(l+1)$ and l and m_l are integers. Introducing the value for λ into the first equation of (13.28) and expanding the first term gives

$$\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \left(\frac{8\pi^2 m_e}{h^2} (\varepsilon - U) - \frac{l(l+1)}{r^2} \right) R = 0$$
(13.29)

or, in terms of $P(r)$ introduced above (where $P(r) = rR(r)$)

$$\frac{d^2 P}{dr^2} + \left(\frac{8\pi^2 m_e}{h^2} (\varepsilon - U) - \frac{l(l+1)}{r^2} \right) P = 0$$
(13.30)

The term in $l(l+1)$ is called the *centrifugal potential*; it adds to the Coulomb term to give an *effective potential*.

The radial equation (13.30) is more complicated than Equation (13.20) because of the $l(l+1)/r^2$ term but in essence the same method of solution can be used. The details are given in standard traditional quantum chemistry texts such as Eyring, Walter and Kimball (EWK) (Eyring *et al.* 1944). The radial solutions are a set of functions from mathematical physics called the *associated Laguerre polynomials* (the Laguerre polynomial $L_\alpha(x)$ of degree α in x is defined as

$$L_\alpha(x) = \exp(x) \frac{d^\alpha}{dx^\alpha} (x^\alpha \exp(-x))$$

and the β th derivative of $L_\alpha(x)$ is called an associated Laguerre polynomial).

13.7.1 $l = 0$ (s-orbitals)

The atomic orbitals are named depending on the three quantum numbers. The first three s-orbitals are given in Table 13.2; they are conventionally written in terms of the variable $\rho = Zr/a_0$ (where $Z = 1$ for hydrogen).

Table 13.2 The first few s-orbitals

n, l, m	Symbol	Normalized wavefunction
1, 0, 0	1s	$\frac{1}{\sqrt{\pi}} \left(\frac{Z}{a_0} \right)^{3/2} \exp(-\rho)$
2, 0, 0	2s	$\frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a_0} \right)^{3/2} (2 - \rho) \exp\left(-\frac{\rho}{2}\right)$
3, 0, 0	3s	$\frac{2}{81\sqrt{3\pi}} \left(\frac{Z}{a_0} \right)^{3/2} (27 - 18\rho + 2\rho^2) \exp\left(-\frac{\rho}{3}\right)$

They are often presented as plots of ψ versus ρ or of $4\pi r^2\psi^2$ versus ρ . The latter is known as the radial distribution function and it gives the probability of finding the electron between two shells of radii r and $r + dr$ surrounding the nucleus. Representative plots for hydrogen are shown in Figures 13.2 and 13.3. The dimension of ψ is $(\text{length})^{-3/2}$ and so the quantities plotted are respectively $a_0^{3/2}\psi$ and $a_0(4\pi r^2\psi^2)$ against r/a_0 .

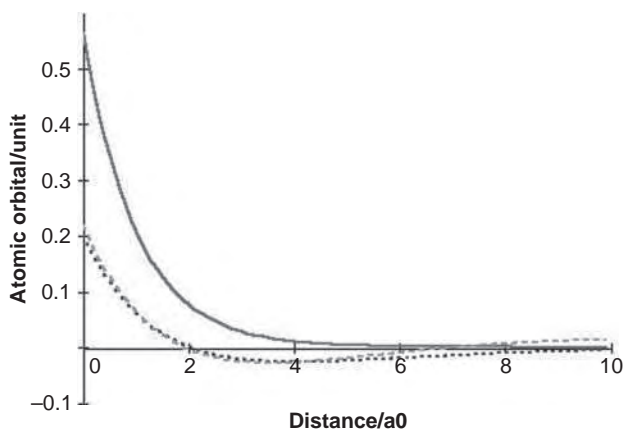


Figure 13.2 Hydrogen 1s, 2s and 3s orbital plots

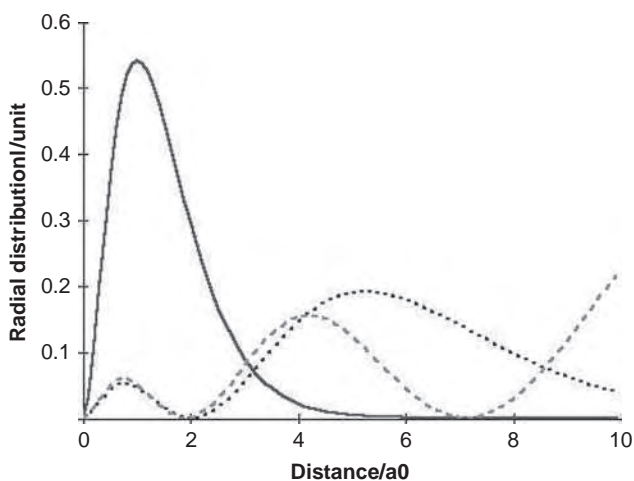


Figure 13.3 Radial distribution curves for the hydrogen 1s, 2s and 3s orbitals

All the curves are asymptotic to the horizontal axis at infinity. The 2s curve crosses the axis once (it has a *single radial node*), the 3s curve crosses the axis twice and so on. There is

nothing else particularly remarkable about such plots of decreasing exponential functions. The radial distribution curves are a little more interesting.

The radial distribution curve shows a maximum at the first Bohr radius for the 1s orbital. Bohr's theory states that the electron would be in a fixed circular orbit round the centre of mass with exactly this radius! As the principal quantum number increases, so does the average value of the radial distribution function. Elementary chemistry texts attach much importance to diagrams of these kinds.

There are other ways to visualize the atomic orbitals. Figures 13.4 and 13.5 show contour diagrams for the 1s and 2s hydrogen atomic orbitals. The particular software package I used marks the axes with somewhat arbitrary units depending on the range of points chosen, although the contour values themselves are correct and the nucleus is at the centre of the plane.

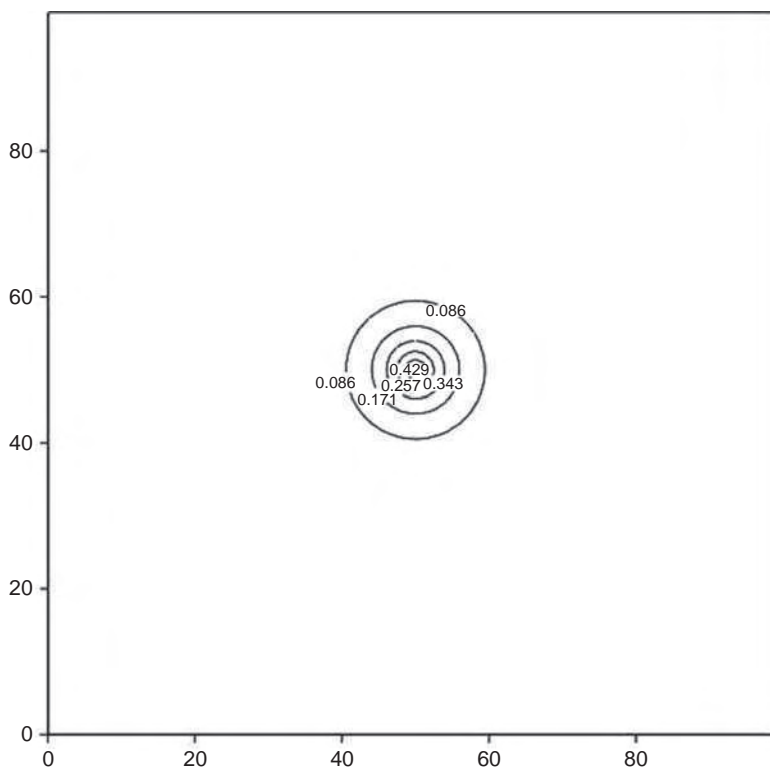


Figure 13.4 *Hydrogen 1s contour diagram (nucleus at centre)*

13.7.2 The p-orbitals

For angular momentum quantum number $l = 1$, there are three possible values of m_l , namely -1 , 0 and $+1$. The angular factors for $m_l = \pm 1$ are complex (they involve the

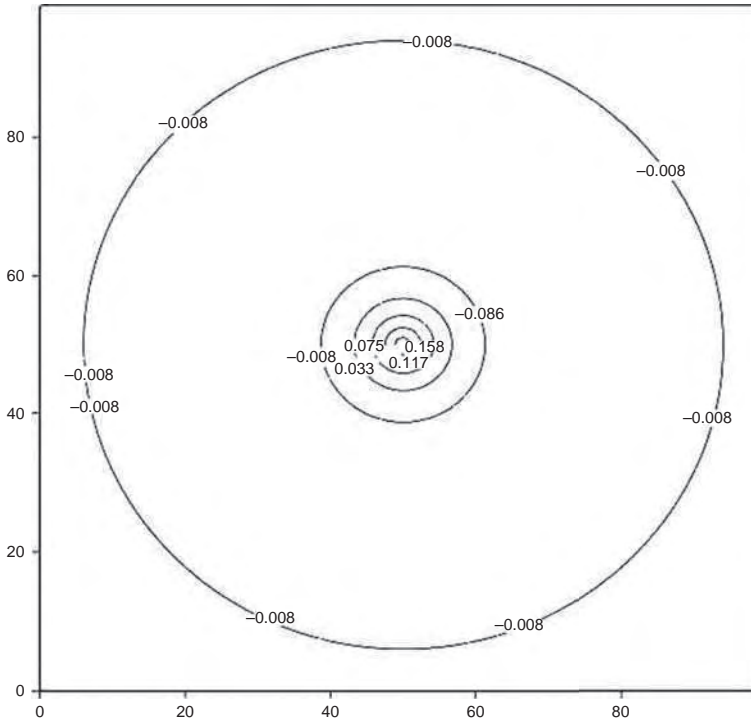


Figure 13.5 Hydrogen $2s$ contour diagram

square root of -1 in an exponential factor) and it is usual to make use of the de Moivre theorem in order to visualize the orbitals:

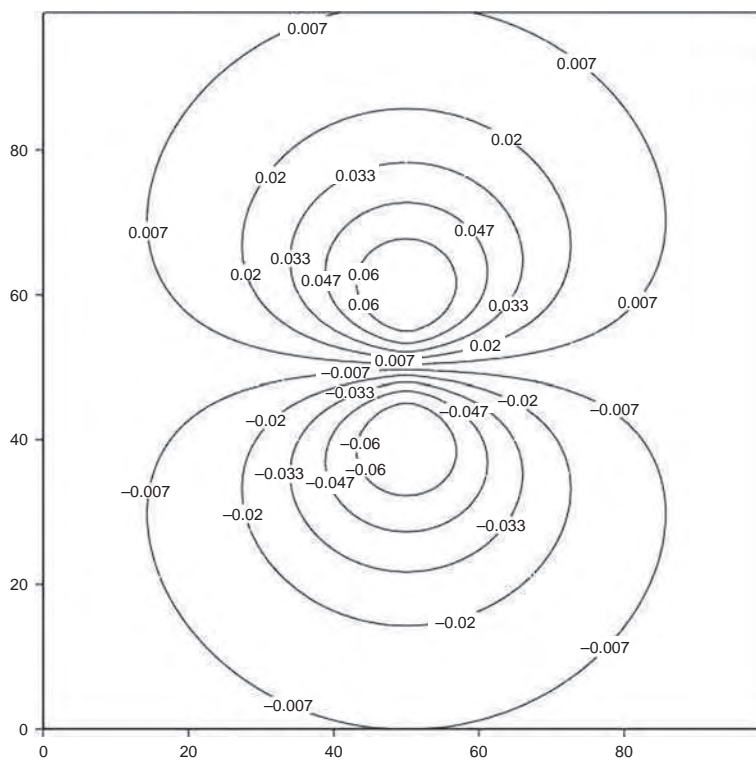
$$\begin{aligned}\cos \phi &= \frac{1}{2} (\exp(j\phi) + \exp(-j\phi)) \\ \sin \phi &= \frac{1}{2j} (\exp(j\phi) - \exp(-j\phi))\end{aligned}\tag{13.31}$$

The *real equivalent* p-orbitals are therefore taken as linear combinations of $m_l = +1$ and -1 . Despite the name, they are not completely equivalent to the complex orbitals as only the latter are eigenfunctions of the z component of the angular momentum. Luckily, this only matters in the presence of an external magnetic field. The real equivalent $2p$ and $3p$ orbitals are shown in Table 13.3.

These are usually represented as contour diagrams; the orbital shapes are memorable because they are equivalent and point along the appropriate axes. Figure 13.6 shows a hydrogenic $2p_z$ orbital, which points along the vertical z axis.

Table 13.3 The first few p -orbitals

n, l, m	Symbol	Normalized wavefunction
2, 1, ± 1	$2p_x$	$\frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho \sin(\theta) \cos(\phi) \exp\left(-\frac{\rho}{2}\right)$
	$2p_y$	$\frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho \sin(\theta) \sin(\phi) \exp\left(-\frac{\rho}{2}\right)$
2, 1, 0	$2p_z$	$\frac{1}{4\sqrt{2\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho \cos(\theta) \exp\left(-\frac{\rho}{2}\right)$
3, 1, ± 1	$3p_x$	$\frac{2}{81\sqrt{\pi}} \left(\frac{Z}{a_0}\right)^{3/2} (6\rho - \rho^2) \sin(\theta) \cos(\phi) \exp\left(-\frac{\rho}{3}\right)$
	$3p_y$	$\frac{2}{81\sqrt{\pi}} \left(\frac{Z}{a_0}\right)^{3/2} (6\rho - \rho^2) \sin(\theta) \sin(\phi) \exp\left(-\frac{\rho}{3}\right)$
3, 1, 0	$3p_z$	$\frac{2}{81\sqrt{\pi}} \left(\frac{Z}{a_0}\right)^{3/2} (6\rho - \rho^2) \cos(\theta) \exp\left(-\frac{\rho}{3}\right)$

**Figure 13.6** Hydrogen atom $2p_z$ orbital

13.7.3 The d-orbitals

Similar comments apply to the $l = 2$ orbitals. There are five in all, four of which come in complex pairs. We combine together the corresponding values of m_l and $-m_l$ just as for the p-orbitals to give real equivalents. Table 13.4 records the 3d orbitals.

Table 13.4 The 3d orbitals

n, l, m_l	Symbol	Normalized wavefunction
3, 2, 0	$3d_{zz}$	$\frac{1}{81\sqrt{6\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho^2 (3 \cos^2(\theta) - 1) \exp\left(-\frac{\rho}{3}\right)$
3, 2, ± 1	$3d_{xz}$	$\frac{\sqrt{2}}{81\sqrt{\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho^2 \sin(\theta) \cos(\theta) \cos(\phi) \exp\left(-\frac{\rho}{3}\right)$
	$3d_{yz}$	$\frac{\sqrt{2}}{81\sqrt{\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho^2 \sin(\theta) \cos(\theta) \sin(\phi) \exp\left(-\frac{\rho}{3}\right)$
3, 2, ± 2	$3d_{x^2-y^2}$	$\frac{1}{81\sqrt{2\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho^2 \sin^2(\theta) \cos(2\phi) \exp\left(-\frac{\rho}{3}\right)$
	$3d_{xy}$	$\frac{1}{81\sqrt{2\pi}} \left(\frac{Z}{a_0}\right)^{3/2} \rho^2 \sin^2(\theta) \sin(2\phi) \exp\left(-\frac{\rho}{3}\right)$

Once again these are usually represented as contour diagrams (Figures 13.7 and 13.8).

13.8 The Stern–Gerlach Experiment

If we suspend a compass needle in the Earth's magnetic field, it aligns itself along the magnetic field lines. A compass needle is an example of a magnetic dipole, and the strength of the interaction between the compass needle and this external magnetic field is determined by the *magnetic dipole moment* \mathbf{p}_m . This is a vector quantity pointing by convention from the south pole of the needle to the north pole. The interaction between the dipole and the field is determined by the magnetic potential energy

$$U_m = -\mathbf{p}_m \cdot \mathbf{B}(\mathbf{r}) \quad (13.32)$$

where the position vector \mathbf{r} is the position of the dipole and \mathbf{B} the magnetic induction. If \mathbf{B} is uniform and so does not depend on \mathbf{r} , the gradient is zero and so the force is also zero, in accord with experiment.

Around 1820, Oersted discovered experimentally that electric currents could exert forces similar to those exerted by permanent magnets, for example on a compass needle. Figure 13.9 shows a simple current loop located in the xz plane and carrying steady current I . If the loop is flat, the dipole is perpendicular to the plane and if we consider points on the axis far away from the loop it turns out that the magnitude of \mathbf{p}_m is

$$p_m = IA$$

where A is the area of the loop (which need not be circular).

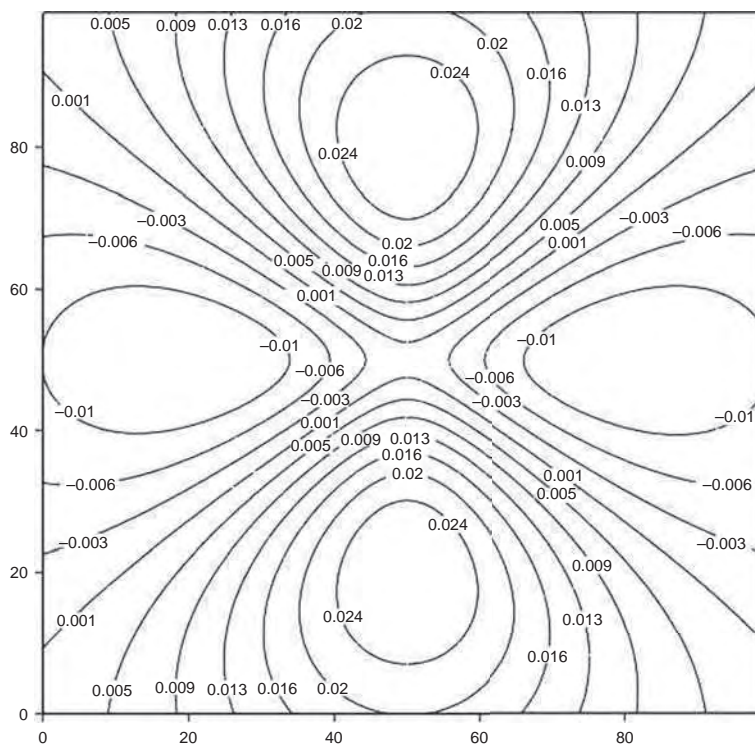


Figure 13.7 Hydrogen $3d_{zz}$ orbital

It is easy to demonstrate from classical physics that a particle with mass M , charge Q and angular momentum \mathbf{l} is a magnetic dipole:

$$\mathbf{p}_m = \frac{Q}{2M} \mathbf{l} \quad (13.33)$$

According to Bohr's model of the hydrogen atom, the allowed electron orbits each had an angular momentum that was an integral multiple of $h/2\pi$. Since magnetic dipoles are linked to angular momentum, the possibility arises that if we could measure the magnetic dipole moments of individual atoms we could investigate the quantization of angular momentum. In the presence of an external magnetic field, the bar magnet will take up one of its possible alignments with the axis of quantization (the direction of the magnetic field lines).

Stern and Gerlach (1921) performed the first and most famous experiment designed to investigate the quantization of angular momentum. Whilst the force on a magnetic dipole is zero in a uniform magnetic field, the force is not zero for a *nonuniform* field. It is difficult to hang a single atom between the poles of a magnet, so a beam of atoms was passed through such a field and the deflection of the beam measured.

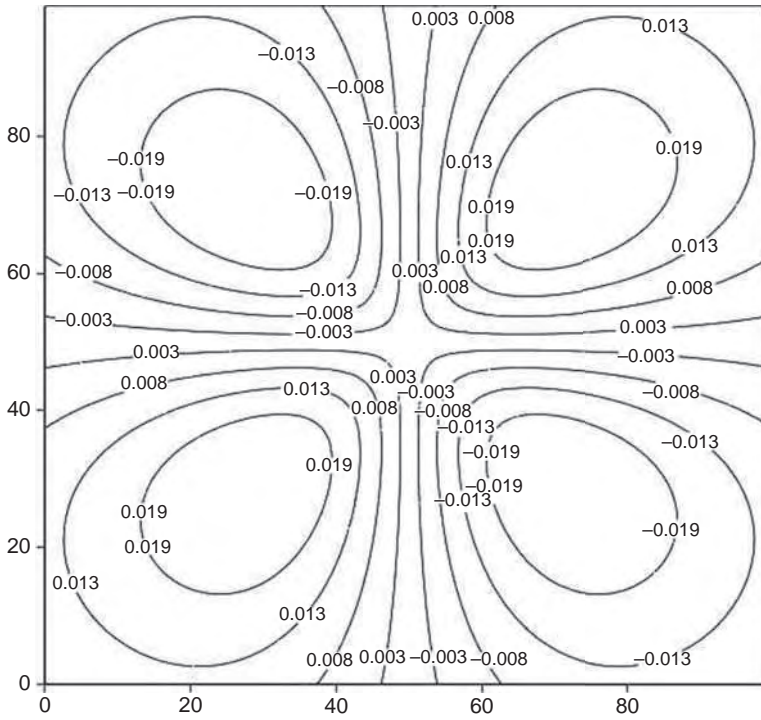


Figure 13.8 Hydrogen $3d_{z^2}$ orbital

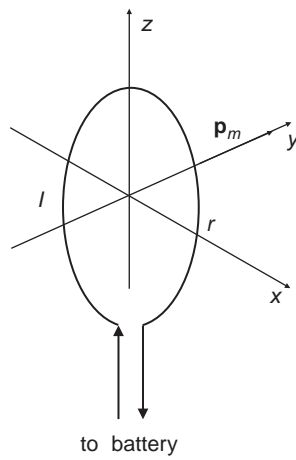


Figure 13.9 Current loop

A schematic diagram of the Stern–Gerlach equipment is shown in Figure 13.10. A beam of atoms is produced in an oven, and they pass through a collimating slit and often a velocity selector in order to form a mono-energetic beam. They then travel in the x direction through a long, powerful electromagnet whose pole pieces are deliberately shaped to give a nonuniform magnetic field in the z direction. Detailed analysis shows that the force on the magnetic dipoles in the z direction is given by a force proportional to the z component of the magnetic dipole moment (and hence the z component of the angular momentum vector):

$$F_z = p_{m,z} \frac{\partial B_z}{\partial z} \quad (13.34)$$

It is a difficult experiment to perform satisfactorily, for the atoms collide with each other in the beam and so distort the result. An essential part of such an *atomic beam* apparatus is that there should be a good vacuum system, and very low beam intensities have to be used. Despite the experimental difficulties, the findings were positive; the atom beam was shown to split on passage through the magnetic field and this confirmed the quantization of angular momentum.

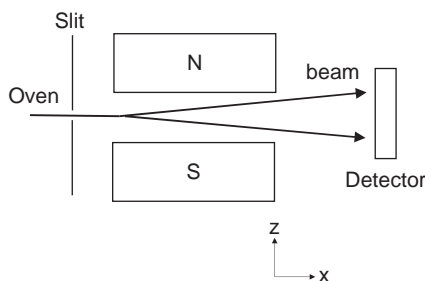


Figure 13.10 Stern–Gerlach experiment

13.9 Electron Spin

The first experiments were made with silver atoms. The surprising result of the Stern–Gerlach experiment was not that the beam split, but that it split into two components, with a different separation than that expected. Similar results were obtained for copper and gold, and in later work for the alkali metals and for hydrogen atoms. The point is that an atomic dipole moment characterized by quantum number l should show $2l + 1$ orientations with the magnetic field. Since the quantum number takes values $0, 1, 2, \dots$ there should always be an odd number of orientations and hence the atom beam should split into an odd number of components.

The explanation came in 1925, when Goudsmit and Uhlenbeck (1925) analysed the splittings of spectral lines occurring when atoms are subject to an external magnetic field (the *Zeeman effect*). They showed that the measured splittings could be explained if electrons were postulated to have an intrinsic magnetic moment, in addition to the one they acquired by motion about the atomic nucleus. In order to explain the experimental results, they

assumed the existence of a spin angular momentum vector \mathbf{s} , which had similar properties to \mathbf{l} above. In order to account for the splitting into two beams, they postulated a spin quantum number for electrons s of $1/2$ and so a spin magnetic quantum number m_s of $\pm 1/2$. The spin wavefunctions are usually written α (corresponding to the $+1/2$ spin quantum number) and β (corresponding to the $-1/2$ quantum number). So we write, just as for ordinary angular momentum

$$\begin{aligned}\hat{s}^2\alpha &= \frac{1}{2}\left(\frac{1}{2} + 1\right)\frac{h^2}{4\pi^2}\alpha \\ \hat{s}^2\beta &= \frac{1}{2}\left(\frac{1}{2} + 1\right)\frac{h^2}{4\pi^2}\beta \\ \hat{s}_z\alpha &= \frac{1}{2}\frac{h}{2\pi}\alpha \\ \hat{s}_z\beta &= -\frac{1}{2}\frac{h}{2\pi}\beta\end{aligned}\tag{13.35}$$

A problem appeared once the splittings were analysed. According to classical theory, the spin magnetic moment of an electron should be

$$\mathbf{p}_{\text{spin}} = \frac{-e}{2m_e}\mathbf{s}\tag{13.36}$$

whilst to get agreement with the splittings, an extra factor of (almost) 2 was needed. The solution to the problem was the introduction of the g -factor, an experimentally determined quantity that made Equation (13.36) correct. Thus we write

$$\mathbf{p}_{\text{spin}} = -g_e\frac{e}{2m_e}\mathbf{s}$$

The electronic g_e factor is another of physical science's very accurately known constants, having a value of $g_e = 2.0023193043787 \pm (82 \times 10^{-12})$

13.10 Total Angular Momentum

When orbital angular momentum and spin angular momentum both exist in the same atom, the magnetic moments that result from these two angular momenta interact to cause a splitting of the energy level. The interaction that leads to this splitting is called *spin-orbit coupling* and it couples the two into a resultant *total angular momentum*. A simple vector model that is very similar to the model used to describe orbital angular momentum can describe this. According to this model, the total angular momentum of an electron is characterized by a quantum number j . For any given nonzero value of l the possible values of j are given by

$$j = l \pm s$$

(The use of j as an atomic quantum number is not to be confused with the use of j for the square root of -1). The rule is that the quantum number j must always be positive, so if $l = 1$ then $j = 3/2$ and $1/2$, but if $l = 0$ then we only have $j = 1/2$).

I have emphasized earlier the interplay between spectroscopic data and theoretical developments; even before Schrödinger's time many highly precise spectroscopic data had been obtained experimentally. The traditional reference sources for such data are the three volumes of tables in the *Atomic Energy Levels* series (Moore 1949, 1952, 1958). In the past few years, the National Institute of Standards and Technology (NIST) atomic spectroscopists have made available a unified comprehensive Atomic Spectra Database on the World Wide Web, which contains spectral reference data for 91 000 wavelengths (<http://www.nist.gov>). The sample in Table 13.5 is taken from the database. Note that the data contains *spectroscopic term symbols* for each level, which are discussed in all elementary physical chemistry undergraduate texts. Spectroscopists traditionally deal with *term values* rather than energies; these are just ε/hc_0 .

Table 13.5 *Internet search for atomic term values*

Configuration	Term symbol	j	Term value/cm ⁻¹
1s ¹	² S	1/2	0
2p ¹	² P	1/2	82 258.9206
		3/2	82 259.2865
2s ¹	² S	1/2	82 258.9559
		3/2	82 259.2865
3p ¹	² P	1/2	97 492.2130
		3/2	97 492.3214
3s ¹	² S	1/2	97 492.2235
		3/2	97 492.3212
3d ¹	² D	5/2	97 492.3574
		3/2	97 492.3212
4p ¹	² P	1/2	102 823.8505
		3/2	102 823.8962
4s ¹	² S	1/2	102 823.8549
		3/2	102 823.8549
4d ¹	² D	3/2	102 823.8961
		5/2	102 823.9114
4f ¹	² F	5/2	102 823.9113
		7/2	102 823.9190

13.11 Dirac Theory of the Electron

There is no mention of electron spin from the Schrödinger equation, and certainly no clue as to why the classical equation for magnetic moments is in error by (roughly) a factor of 2 when applied to electron spin (Equation (13.36)). If we consider the time-dependent Schrödinger equation for a free electron

$$-\frac{\hbar^2}{8\pi^2m_e} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \Psi(\mathbf{r}, t) = j \frac{\hbar}{2\pi} \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) \quad (13.37)$$

(where j is the square root of -1), it is seen to be a second-order partial differential equation with respect to the spatial coordinates and a first-order partial differential equation with respect to time. It therefore is not consistent with the special theory of relativity, which requires that time and space should enter such equations on an equal footing. If the equation is second order in space, it should also be second order in time.

Erwin Schrödinger decided that the way ahead was to abandon the classical energy expression and start again with the relativistically correct equation

$$\varepsilon^2 = m_e^2 c_0^4 + c_0^2 \mathbf{p}^2 \quad (13.38)$$

He then made the operator substitutions

$$\varepsilon \rightarrow j \frac{h}{2\pi} \frac{\partial}{\partial t} \quad p_x \rightarrow -j \frac{h}{2\pi} \frac{\partial}{\partial x} \quad \text{etc.}$$

to arrive at the Klein–Gordan equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c_0^2} \frac{\partial^2}{\partial t^2} - \frac{4\pi^2 m_e^2 c_0^2}{h^2} \right) \Psi(\mathbf{r}, t) = 0 \quad (13.39)$$

In discussing relativistic matters, it is usual to write equations such as this in *four-vector notation*. We define a four-vector with components

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \\ jc_0 t \end{pmatrix}$$

so that Equation (13.39) becomes

$$\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + \frac{\partial^2}{\partial x_4^2} - \frac{4\pi^2 m_e^2 c_0^2}{h^2} \right) \Psi(x_1, x_2, x_3, x_4) = 0 \quad (13.40)$$

The Klein–Gordan equation is more satisfactory in that it has the desirable symmetry but it turns out that it cannot describe electron spin. In the limit of low energy, it is equivalent to the familiar Schrödinger equation.

Paul Dirac had the ingenious idea of working with a relativistic equation that was linear in the vector components. He wrote

$$\left(\gamma_1 \frac{\partial}{\partial x_1} + \gamma_2 \frac{\partial}{\partial x_2} + \gamma_3 \frac{\partial}{\partial x_3} + \gamma_4 \frac{\partial}{\partial x_4} - \frac{2\pi m_e c_0}{h} \right) \Psi(x_1, x_2, x_3, x_4) = 0 \quad (13.41)$$

where the multipliers γ_i have to be determined. This equation is called the *Dirac equation*. Both the Schrödinger and the Klein–Gordan equation are second order, and it is usual to manipulate the Dirac equation in order to give a corresponding second-order equation. This can be done by operating on Equation (13.41) with the operator

$$\gamma_1 \frac{\partial}{\partial x_1} + \gamma_2 \frac{\partial}{\partial x_2} + \gamma_3 \frac{\partial}{\partial x_3} + \gamma_4 \frac{\partial}{\partial x_4} + \frac{2\pi m_e c_0}{h}$$

A little operator algebra shows that the multipliers have to satisfy

$$\gamma_i \gamma_j + \gamma_j \gamma_i = \begin{cases} 2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

It can be shown that any particle whose wavefunction satisfies the Dirac equation must be a spin $-1/2$ particle. Not only that, the Dirac treatment gives the correct value for the magnetic moment in that it gives

$$\mathbf{p}_{\text{spin}} = -2 \frac{e}{2m_e} \mathbf{s}$$

13.12 Measurement in the Quantum World

The process of measurement in quantum mechanics is subtler than in classical mechanics. As discussed in Appendix A, the possible results of measurements depend on the eigenvalues of the appropriate operators. Also, if we wish to make simultaneous measurements of two observables (such as the linear momentum and position, or two components of an angular momentum vector), we have to take account of Heisenberg's uncertainty principle. Certain pairs of observables can be measured simultaneously to arbitrary precision whilst other pairs cannot.

There is worse to come, for the word *measurement* gets used in two different ways in quantum mechanics. Suppose we have a one-electron atom as discussed above; we know that the energies are found from Schrödinger's equation, which I will write in Hamiltonian form as

$$\hat{H}\psi_i = \varepsilon_i\psi_i$$

If the atom is in state ψ_i then repeated energy measurements on the same atom will always yield the same result, ε_i (we say that the system is in a *pure state*).

If on the other hand we pass a beam of electrons through a Stern–Gerlach apparatus, the magnet separates the beam into two components that correspond to the two spin eigenvalues $m_s = +1/2$ and $m_s = -1/2$. This kind of measurement is referred to as *state preparation* for if we pass the $m_s = +1/2$ beam through a further Stern–Gerlach apparatus oriented in the same way as the first, we simply observe the one beam. I have shown this schematically in Figure 13.11.

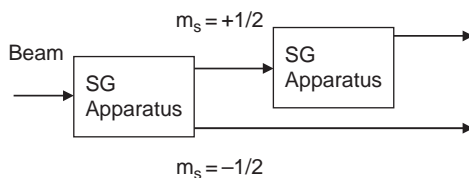


Figure 13.11 Repeated Stern–Gerlach (SG) measurements

However, electrons in the incident beam are not in pure states, and their spin wavefunction can be written as linear combinations of the two spin functions α and β :

$$\psi = a\alpha + b\beta$$

They are said to be in *mixed states*. Here, a and b are scalar (complex) constants. The question is how do we interpret the measurement process? For example, does a given

electron in some sense know which spin state it should be in, before passage through the apparatus? Similar questions puzzled Schrödinger and his contemporaries, and led Schrödinger to state the famous *cat paradox*. In a version of this thought experiment, a cat is placed in a closed box together with (for example) a radioactive atom that releases a deadly poison when it decays. If the wavefunction of the living cat is ψ_L and that of a dead cat ψ_D then the state of the cat at time t can be described by

$$\psi_{\text{Cat}}(t) = c_L(t) \psi_L + c_D(t) \psi_D$$

The coefficients are time dependent, as is the wavefunction. Before the box is closed, the cat is alive. Once the box is closed, the cat is apparently neither dead nor alive; it is in a mixed state. On opening the box at a certain time, the cat is either dead or alive, but is it the act of measurement that has forced the cat into one state or the other?

There are whole rafts of similar questions that can be asked. Most of the difficulties can be resolved by recognizing that we are asking a statistical question, not one about a specific cat. If we prepare a large number of identical experiments, with identical cats and amounts of radioactive isotopes, close the lids on the boxes and then examine a statistically significant number of boxes, we will find that a fraction $|c_D(t)|^2$ will have died and a fraction $|c_L(t)|^2$ will be alive, but no prediction can be made as to the state of a given cat in a given box.

References

- Eyring, H., Walter, J. and Kimball, G.E. (1944) *Quantum Chemistry*, John Wiley & Sons, Inc., New York.
- Goudsmit, S. and Uhlenbeck, G.E. (1925) *Naturwissenschaften*, **13**, 953.
- Moore, C.E. (1949) *Atomic Energy Levels Volume I (Hydrogen through Vanadium)*, Circular of the National Bureau of Standards, US Government Printing Office, Washington, DC.
- Moore, C.E. (1952) *Atomic Energy Levels Volume II (Chromium through Niobium)*, Circular of the National Bureau of Standards, US Government Printing Office, Washington, DC.
- Moore, C.E. (1958) *Atomic Energy Levels Volume III (Molybdenum through Lanthanum and Hafnium through Actinium)*, Circular of the National Bureau of Standards, US Government Printing Office, Washington, DC.
- Stern, O. and Gerlach, W. (1921) *Z. Physik*, **8**, 110.

14

The Orbital Model

Our next step is to consider a many-electron atom, such as that illustrated in Figure 14.1.

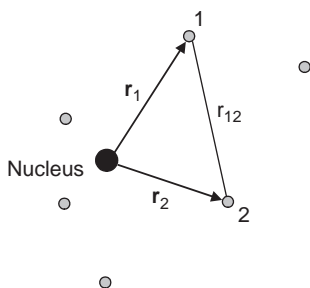


Figure 14.1 Many-electron atom

I am going to make the ‘infinite nuclear mass’ approximation, and only consider the electronic problem. The atom is therefore fixed in space with the nucleus at the centre of the coordinate system. The n electrons are at position vectors $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ and the scalar distance between (say) electrons 1 and 2 is r_{12} in an obvious notation. The electronic wavefunction will depend on the coordinates of all the electrons, and I will write it $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$. If the nuclear charge is Ze , we have to consider three contributions to the electronic energy: the kinetic energy of each electron, the Coulomb attraction between the nucleus and the electrons and finally the Coulomb repulsion between pairs of electrons. We therefore write the Hamiltonian operator as

$$\hat{H} = \sum_{i=1}^n \left(-\frac{\hbar^2}{8\pi^2 m_e} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) - \frac{Ze^2}{4\pi \epsilon_0 r_i} \right) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{e^2}{4\pi \epsilon_0 r_{ij}} \quad (14.1)$$

14.1 One- and Two-Electron Operators

In order to stop the notation becoming unwieldy, I will group these terms as follows. Each term in the first bracket refers to the coordinates of the individual electrons, and gives the kinetic energy of each electron together with its attraction to the nucleus. I will call such terms *one-electron operators* and authors normally write them as

$$\hat{h}^{(1)}(\mathbf{r}_i) = -\frac{\hbar^2}{8\pi^2 m_e} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) - \frac{Ze^2}{4\pi \epsilon_0 r_i} \quad (14.2)$$

Each term in the second double summation gives the Coulomb repulsion of a pair of electrons; I will refer to them as *two-electron operators* and write

$$\hat{g}(\mathbf{r}_i, \mathbf{r}_j) = \frac{e^2}{4\pi \epsilon_0 r_{ij}} \quad (14.3)$$

The Hamiltonian is then, in our compact notation

$$\hat{H} = \sum_{i=1}^n \hat{h}^{(1)}(\mathbf{r}_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \hat{g}(\mathbf{r}_i, \mathbf{r}_j) \quad (14.4)$$

and we wish to investigate the solutions of the Schrödinger equation

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \epsilon\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \quad (14.5)$$

14.2 Many-Body Problem

The many-electron atom is an example of a so-called *many-body problem*. These are not unique to quantum theory; a familiar example is planetary motion. Newton's equations of motion can be solved exactly for the motion of any one of the planets round the sun individually, but the planets also attract each other. During the eighteenth and nineteenth centuries a great deal of effort was expended trying to find an exact solution to planetary motion, but all efforts failed and it is generally accepted that exact solutions do not exist, even for just three bodies. Astronomers are lucky in the sense that the gravitational force depends on the product of two masses and the mass of the sun is much greater than the masses of the individual planets. Ingenious techniques were developed to treat the interplanetary attractions as small perturbations, and the planetary problem can be solved numerically to any accuracy required.

To continue with the planetary motion analogy, chemists are less lucky in one sense because the electrostatic force depends on the product of two charges and in the case of an electron and a proton these forces are roughly equal in magnitude. It will therefore come

as no surprise when I tell you that exact solution of the many-electron atomic Schrödinger problem seems to be impossible, because of the electron–electron repulsions, and that we are apparently in a weaker position than the astronomers. On the positive side, I have stressed the Born interpretation of quantum mechanics. Here we do not focus attention on the trajectories of the individual particles, rather we ask about the probability that a region of space is occupied by any one of the particles. The quantum mechanical problem therefore seems to be more hopeful than the astronomical one.

14.3 Orbital Model

Let me now investigate the Schrödinger equation that would result from an atom consisting of electrons that did not repel each other. We can think of this as some ‘zero-order’ approximation to a true many-electron atom, just as the astronomers might have investigated their simple model of planetary motion.

We therefore write

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \varepsilon\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$$

$$\left(\sum_{i=1}^n \hat{h}^{(i)}(\mathbf{r}_i)\right)\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) = \varepsilon\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$$

and this appears to be a candidate for separation of variables. I write

$$\Psi(\mathbf{r}_1, \mathbf{r}_1, \dots, \mathbf{r}_1) = \psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)\dots\psi_n(\mathbf{r}_n) \quad (14.6)$$

Substitution and separation gives n identical one-electron atom Schrödinger equations, and so the total wavefunction is a product of the familiar 1s, 2s, 2p atomic orbitals discussed in Chapter 13 (with nuclear charge Ze). The energy is given by the sum of the orbital energies.

We then have to take account of electron spin and the Pauli principle, as discussed in Chapter 13. I can remind you of the principles by writing down some of the lowest energy solutions for helium, in particular those that formally involve the 1s and 2s orbitals. I will label the atomic orbitals 1s and 2s for obvious reasons, and I will adopt the habit of writing s for the spin variable (sorry about the double use of the same symbol) as in Chapter 13.

The allowed wavefunctions are given in Table 14.1.

Ψ_1 describes the spectroscopic ground state, giving energy $2\varepsilon_{1s}$. The remaining wavefunctions describe excited states. Ψ_2 is the first excited singlet state whilst Ψ_3 , Ψ_4 and Ψ_5 are the three components of the first triplet state. We refer to Ψ_2 through Ψ_5 as *singly excited wavefunctions* because they have been formally produced from the ground-state wavefunction by exciting a single electron. Due to my neglect of electron repulsion, the energies of Ψ_2 through Ψ_5 are the same. Ψ_6 is a *doubly excited wavefunction* and so on.

Our zero-order model predicts that the singlet and triplet excited states derived from a $1s^1 2s^1$ orbital configuration will have energy

$$\varepsilon_{2s} - \varepsilon_{1s} = \frac{Z^2 m_e e^4}{8h^2 \varepsilon_0^2} \left(\frac{1}{1^2} - \frac{1}{2^2} \right)$$

Table 14.1 Some electronic states for helium

State	Spatial part	Spin part	Energy, ε
ψ_1	$1s(\mathbf{r}_1)1s(\mathbf{r}_2)$	$\frac{1}{\sqrt{2}}(\alpha(s_1)\beta(s_2) - \alpha(s_2)\beta(s_1))$	$2\varepsilon_{1s}$
ψ_2	$\frac{1}{\sqrt{2}}(1s(\mathbf{r}_1)2s(\mathbf{r}_2) + 1s(\mathbf{r}_2)2s(\mathbf{r}_1))$	$\frac{1}{\sqrt{2}}(\alpha(s_1)\beta(s_2) - \alpha(s_2)\beta(s_1))$	$\varepsilon_{1s} + \varepsilon_{2s}$
ψ_3		$\alpha(s_1)\alpha(s_2)$	$\varepsilon_{1s} + \varepsilon_{2s}$
ψ_4	$\frac{1}{\sqrt{2}}(1s(\mathbf{r}_1)2s(\mathbf{r}_2) - 1s(\mathbf{r}_2)2s(\mathbf{r}_1))$	$\frac{1}{\sqrt{2}}(\alpha(s_1)\beta(s_2) + \alpha(s_2)\beta(s_1))$	$\varepsilon_{1s} + \varepsilon_{2s}$
ψ_5		$\beta(s_1)\beta(s_2)$	$\varepsilon_{1s} + \varepsilon_{2s}$
ψ_6	$2s(\mathbf{r}_1)2s(\mathbf{r}_2)$	$\frac{1}{\sqrt{2}}(\alpha(s_1)\beta(s_2) - \alpha(s_2)\beta(s_1))$	$2\varepsilon_{2s}$

above the ground state. Since $Z = 2$ for helium we calculate a wavenumber of $329\,212\text{ cm}^{-1}$. Experimental data can be found at the NBS/NIST website <http://www.nist.gov> as given in Table 14.2.

Table 14.2 Experimental data for helium

Configuration	Term	J	Term value/ cm^{-1}
$1s^2$	1S	0	0
$1s^1 2s^1$	3S	1	159 856.07760
$1s^1 2s^1$	1S	0	166 277.542

I should explain that the J quantum number is similar to the j quantum number we met in our study of one-electron atoms in Chapter 13; for light atoms such as helium it is determined by combining together the individual electron orbital and spin angular momentum quantum numbers, according to a set of well-known rules called the *Russell–Saunders* scheme. We combine together the l quantum numbers for the two electrons; since $l = 0$ for an s electron, the allowed resultant L is also 0. The electronic states are therefore S states. We also combine together the spin quantum numbers s . Since $s = 1/2$ for an electron, the allowed values of the resultant S are $1/2 + 1/2$ and $1/2 - 1/2$ and the spin multiplicities $2S + 1$ are 1 and 3. We then combine the L 's and the S 's in the same way to get J .

The zero-order model is not even qualitatively correct; it overestimates the energy difference between the ground state and the excited states, and has nothing at all to say about the experimentally interesting difference between the singlet and the triplet excited states. This poor agreement with experiment is mostly due to our neglect of electron repulsion but is in part due to the fact that each electron shields the other electron from the Coulomb attraction due to the nucleus.

14.4 Perturbation Theory

There are very few physically interesting problems that we can solve exactly by the standard methods of quantum theory. The great majority of problems, including those of atomic and molecular structure, must therefore be tackled by approximate methods.

Suppose that our problem is to solve, for example, the helium atom electronic Schrödinger equation

$$\hat{H}_i \Psi_i(\mathbf{r}_1, \mathbf{r}_2) = \varepsilon_i \Psi_i(\mathbf{r}_1, \mathbf{r}_2) \quad (14.7)$$

We might suspect that this problem is similar to that of two superimposed hydrogen atoms, for which we can find exact solutions to the *zero-order* Schrödinger equation

$$\hat{H}^{(0)} \Psi_i^{(0)}(\mathbf{r}_1, \mathbf{r}_2) = \varepsilon_i^{(0)} \Psi_i^{(0)}(\mathbf{r}_1, \mathbf{r}_2) \quad (14.8)$$

The aim of perturbation theory is to relate the solutions of problem (14.7) to the exact zero-order solutions of (14.8). To simplify the notation, I will drop all references to two electrons; perturbation theory is a general technique, not one that is specific to helium. It is also general in that it can be applied to every solution not just the lowest energy one. There are two technical points. First, I am going to assume that the state of interest is not degenerate. There is a special version of perturbation theory that is applicable to degenerate states, and I refer the interested reader to the classic texts such as Eyring *et al.* (1944). Second, I am going to assume that the wavefunctions are real rather than complex. It makes the equations look a bit easier on the eye.

We proceed as follows. First we write the Hamiltonian as

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}^{(1)} \quad (14.9)$$

where λ is called the *perturbation parameter*. The second term is called the *perturbation*. We assume that the energies and wavefunctions for our problem can be expanded in terms of the zero-order problem as

$$\begin{aligned} \Psi_i &= \Psi_i^{(0)} + \lambda \Psi_i^{(1)} + \lambda^2 \Psi_i^{(2)} + \dots \\ \varepsilon_i &= \varepsilon_i^{(0)} + \lambda \varepsilon_i^{(1)} + \lambda^2 \varepsilon_i^{(2)} + \dots \end{aligned} \quad (14.10)$$

The superscript ' k ' refers to the order of perturbation theory, and the equation should demonstrate why the perturbation parameter λ is added; it is a formal device used to keep track of the 'orders' of the perturbation. It might physically correspond to an applied electric field (as in the Stark effect) or an applied magnetic induction (as in the Zeeman effect, in which case we need to use complex wavefunctions). If we substitute the expansions into our problem we find

$$\begin{aligned} \hat{H}^{(0)} \Psi_i^{(0)} + \lambda \left(\hat{H}^{(1)} \Psi_i^{(0)} + \hat{H}^{(0)} \Psi_i^{(1)} \right) + \lambda^2 \left(\hat{H}^{(1)} \Psi_i^{(1)} + \hat{H}^{(0)} \Psi_i^{(2)} \right) + \dots \\ = \varepsilon_i^{(0)} \Psi_i^{(0)} + \lambda \left(\varepsilon_i^{(1)} \Psi_i^{(0)} + \varepsilon_i^{(0)} \Psi_i^{(1)} \right) + \lambda^2 \left(\varepsilon_i^{(2)} \Psi_i^{(0)} + \varepsilon_i^{(1)} \Psi_i^{(1)} + \varepsilon_i^{(0)} \Psi_i^{(2)} \right) + \dots \end{aligned} \quad (14.11)$$

In order that Equation (14.11) may be true for all values of λ , the coefficients of λ on either side of the equation must be equal. Equating and rearranging we find

$$\begin{aligned}\hat{H}^{(0)}\Psi_i^{(0)} &= \varepsilon_i^{(0)}\Psi_i^{(0)} \\ (\hat{H}^{(0)} - \varepsilon_i^{(0)})\Psi_i^{(1)} &= (\varepsilon_i^{(1)} - \hat{H}^{(1)})\Psi_i^{(0)} \\ (\hat{H}^{(0)} - \varepsilon_i^{(0)})\Psi_i^{(2)} &= \varepsilon_i^{(2)}\Psi_i^{(0)} + \varepsilon_i^{(1)}\Psi_i^{(1)} - \hat{H}^{(1)}\Psi_i^{(1)}\end{aligned}\quad (14.12)$$

Solution of these equations gives

$$\begin{aligned}\varepsilon_i &= \varepsilon_i^{(0)} + \lambda H_{ii}^{(1)} + \lambda^2 \sum_{j \neq i} \frac{H_{ij}^{(1)} H_{ji}^{(1)}}{\varepsilon_i^{(0)} - \varepsilon_j^{(0)}} + \dots \\ \Psi_i &= \Psi_i^{(0)} + \lambda \sum_{j \neq i} \frac{H_{ji}^{(1)}}{\varepsilon_i^{(0)} - \varepsilon_j^{(0)}} \Psi_j^{(0)} + \dots\end{aligned}\quad (14.13)$$

I have used the following shorthand for the integrals:

$$H_{ij}^{(1)} = \int \Psi_i^{(0)} \hat{H}^{(1)} \Psi_j^{(0)} d\tau \quad (14.14)$$

The first-order correction to the energy can therefore be calculated from the unperturbed wavefunction and the perturbing Hamiltonian:

$$\begin{aligned}\varepsilon_i^{(1)} &= H_{ii}^{(1)} \\ &= \int \Psi_i^{(0)} \hat{H}^{(1)} \Psi_i^{(0)} d\tau\end{aligned}$$

On the other hand, the second-order correction to the energy requires knowledge of the remaining states

$$\varepsilon_i^{(2)} = \sum_{j \neq i} \frac{\left(\int \Psi_i^{(0)} \hat{H}^{(1)} \Psi_j^{(0)} d\tau \right)^2}{\varepsilon_i^{(0)} - \varepsilon_j^{(0)}} \quad (14.15)$$

If we return to the helium atom where the zero-order problem is two noninteracting atomic electrons, the zero-order wavefunctions and energies are shown in Table 14.1. The perturbation is the Coulomb repulsion between the two electrons

$$\hat{H}^{(1)} = \frac{e^2}{4\pi \varepsilon_0 r_{12}}$$

and the first-order correction to the ground-state energy is

$$\varepsilon^{(1)} = \int \Psi_1 \left(\frac{e^2}{4\pi \varepsilon_0 r_{12}} \right) \Psi_1 d\tau$$

where Ψ_1 is given in Table 14.1.

Evaluation of the integral is far from easy, since it involves the coordinates of both electrons (it is therefore a six-dimensional integral), and it has a singularity (it tends to

infinity as the two electrons approach each other). It is shown in the classic texts such as Eyring *et al.* (1944) that

$$\begin{aligned}\varepsilon^{(1)} &= \frac{5Z}{8} \frac{e^2}{4\pi\varepsilon_0 a_0} \\ \frac{\varepsilon^{(1)}}{hc_0} &= \frac{5Z}{8hc_0} \frac{e^2}{4\pi\varepsilon_0 a_0} \\ &= 274\,343\text{ cm}^{-1}\end{aligned}$$

and my revised estimate of the first ionization energy is now $164\,606\text{ cm}^{-1}$, in better agreement with experiment.

14.5 Variation Method

Another, completely different method of finding approximate solutions to the Schrödinger equation is based on the following theorem.

Theorem

If the lowest eigenvalue of a Hamiltonian \hat{H} is ε and Φ is a function with the correct boundary conditions, then

$$\frac{\int \Phi^* \hat{H} \Phi \, d\tau}{\int \Phi^* \Phi \, d\tau} \geq \varepsilon$$

Once again, the proof is given in all the classic texts such as Eyring *et al.* (1944). I can illustrate the use of this technique by reference to the helium atom. The spatial part of the zero-order wavefunction is

$$\Psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{Z^3}{\pi a_0^3} \exp\left(-Z \frac{(r_1 + r_2)}{a_0}\right)$$

where Z is the atomic number (2 in this case). We now allow for the possibility that one electron partially screens the nucleus and so the second electron sees a reduced nuclear charge Z' . This suggests that we use a trial wavefunction that comprises a product of modified helium 1s orbitals with effective nuclear charge $Z'e$

$$\Phi(\mathbf{r}_1, \mathbf{r}_2) = \frac{(Z')^3}{\pi a_0^3} \exp\left(-Z' \frac{(r_1 + r_2)}{a_0}\right)$$

We would expect Z' to lie between 1 and 2 and we look for the value of Z' that makes the *variational integral*

$$\varepsilon' = \frac{\int \Phi(\mathbf{r}_1, \mathbf{r}_2) \left(\hat{h}^{(1)}(\mathbf{r}_1) + \hat{h}^{(1)}(\mathbf{r}_2) + \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \right) \Phi(\mathbf{r}_1, \mathbf{r}_2) \, d\tau_1 d\tau_2}{\int \Phi^2(\mathbf{r}_1, \mathbf{r}_2) \, d\tau_1 d\tau_2} \quad (14.16)$$

a minimum. I have dropped the complex conjugate signs $*$ because of the real nature of the wavefunction. The denominator is

$$\begin{aligned}\int \Phi^2(\mathbf{r}_1, \mathbf{r}_2) \, d\tau_1 d\tau_2 &= \int 1s^2(\mathbf{r}_1) d\tau_1 \int 1s^2(\mathbf{r}_2) d\tau_2 \\ &= 1\end{aligned}$$

The numerator is

$$\begin{aligned}\varepsilon' &= \int \Phi(\mathbf{r}_1, \mathbf{r}_2) \hat{h}^{(1)}(\mathbf{r}_1) \Phi(\mathbf{r}_1, \mathbf{r}_2) d\tau_1 d\tau_2 + \int \Phi(\mathbf{r}_1, \mathbf{r}_2) \hat{h}^{(1)}(\mathbf{r}_2) \Phi(\mathbf{r}_1, \mathbf{r}_2) d\tau_1 d\tau_2 \\ &\quad + \int \Phi(\mathbf{r}_1, \mathbf{r}_2) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \Phi(\mathbf{r}_1, \mathbf{r}_2) d\tau_1 d\tau_2 \\ \varepsilon' &= \int 1s(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) 1s(\mathbf{r}_1) d\tau_1 \int 1s^2(\mathbf{r}_2) d\tau_2 + \int 1s(\mathbf{r}_2) \hat{h}^{(1)}(\mathbf{r}_2) 1s(\mathbf{r}_2) d\tau_2 \\ &\quad \times \int 1s^2(\mathbf{r}_1) d\tau_1 + \frac{e^2}{4\pi\varepsilon_0} \int 1s(\mathbf{r}_1) 1s(\mathbf{r}_2) \frac{1}{r_{12}} 1s(\mathbf{r}_1) 1s(\mathbf{r}_2) d\tau_1 d\tau_2\end{aligned}\quad (14.17)$$

The first two integrals are related to the energy of a 1s orbital of He^+ ; each integral is equal because of the indistinguishability of the two electrons, and I can write each one as

$$\begin{aligned}&\int 1s(\mathbf{r}_1) \left(-\frac{\hbar^2}{8\pi^2 m_e} \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial y_1^2} + \frac{\partial^2}{\partial z_1^2} \right) - \frac{Z'e^2}{4\pi\varepsilon_0 r_1} \right) 1s(\mathbf{r}_1) d\tau_1 \\ &\quad + \int 1s(\mathbf{r}_1) \left(\frac{(Z' - Z)e^2}{4\pi\varepsilon_0 r_1} \right) 1s(\mathbf{r}_1) d\tau_1 \\ &= -\frac{1}{2} (Z')^2 E_h + \int 1s(\mathbf{r}_1) \left(\frac{(Z' - Z)e^2}{4\pi\varepsilon_0 r_1} \right) 1s(\mathbf{r}_1) d\tau_1 \\ &= -\frac{1}{2} (Z')^2 E_h + Z'(Z' - Z)E_h\end{aligned}$$

Adding the two-electron integral from above we find a variational energy of

$$\varepsilon = \left((Z')^2 - 2ZZ' + \frac{5}{8}Z' \right) E_h \quad (14.18)$$

We obtain the best approximation to the true energy by giving Z' the value which will make the energy a minimum. This means

$$\begin{aligned}\frac{d\varepsilon}{dZ'} &= \left(2Z' - 2Z + \frac{5}{8} \right) E_h \\ &= 0\end{aligned}$$

and so the best energy results when we take

$$Z' = Z - \frac{5}{16}$$

This leads to an improved ground-state energy of

$$\varepsilon = -\left(\frac{27}{16} \right)^2 E_h$$

By introducing more and more parameters like Z' into the wavefunction Φ we can approach more and more closely the experimental result. Hylleraas (1930) experimented

with wavefunctions that were explicit functions of the interelectron coordinate r_{12} . His first attempt was to write

$$\Phi = A (\exp(-Z'(r_1 + r_2)) (1 + cr_{12})) \quad (14.19)$$

where A is the normalization constant and Z' and c are adjustable parameters. In later work he made use of expansions such as

$$\Phi = A (\exp(-Z'(r_1 + r_2)) (\text{polynomial in } r_1, r_2, r_{12})) \quad (14.20)$$

and was able to demonstrate impressive agreement with experiment. We refer to such wavefunctions as *correlated* wavefunctions; as we will soon see, the most sophisticated orbital models average over the electron interactions whilst correlated wavefunctions allow for the 'instantaneous' interaction between electrons.

Some of Hylleraas's results are shown in Table 14.3; the most accurate theoretical value known is that of Frankowski and Pekeris (1966).

Table 14.3 Hylleraas's variational results for helium

Approximate wavefunction	ε/E_h
$\exp(-k(r_1 + r_2))$	-2.8478
$\exp(-k(r_1 + r_2)) \exp(-c_1 r_{12})$	-2.8896
$\exp(-k(r_1 + r_2)) \exp(-c_1 r_{12}) \cosh(c(r_1 - r_2))$	-2.8994
$\exp\left(-\left(\frac{r_1 + r_2}{2}\right)\right) (c_0 + c_1 r_2 + c_2 (r_1 - r_2)^2 + c_3 (r_1 + r_2) + c_4 (r_1 + r_2)^2 + c_5 r_{12}^2)$	-2.9032
'Exact'	-2.90372437703

Hylleraas's approach whereby we write the interelectron distances explicitly in the wavefunction (and so abandon the orbital model) gives by far the most accurate treatment of atomic systems, but like most attractive propositions there is a catch. I glossed over calculation of the two-electron integral in the discussion above, but this causes a major problem for molecular systems. Any attempt to use hydrogenic orbitals for a molecular system leads to two-electron integrals that are impossibly hard to evaluate. Despite many attempts, Hylleraas's method has never been successfully applied to a large polyatomic molecule.

14.6 Linear Variation Method

The variation principle as stated above applies only to the lowest energy solution of any given symmetry (spatial and spin). A special kind of variation function widely used in molecular structure theory is the *linear variation function*. For the sake of argument, suppose we try to improve the helium atom ground-state wavefunction Ψ_1 by adding Ψ_2 through Ψ_6 (given in Table 14.1) and so write

$$\Phi = c_1\Psi_1 + c_2\Psi_2 + \cdots + c_6\Psi_6$$

We keep the effective nuclear charge Z' constant throughout the calculation (since it is a nonlinear parameter), but seek the values of the linear parameters c_1 through c_6 that minimize the variational integral

$$\varepsilon = \frac{\int \Phi^* \hat{H} \Phi \, d\tau}{\int \Phi^* \Phi \, d\tau}$$

For neatness, I will recast the problem in matrix notation and treat the general case with n rather than six Ψ 's. We write

$$\Phi = (\Psi_1 \quad \Psi_2 \quad \dots \quad \Psi_n) \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \Psi \mathbf{c}$$

I will also collect integrals such as

$$H_{ij} = \int \Psi_i^* \hat{H} \Psi_j \, d\tau \quad \text{and} \quad S_{ij} = \int \Psi_i^* \Psi_j \, d\tau$$

into the n by n matrices \mathbf{H} and \mathbf{S} . I will generally work with real rather than complex wavefunctions, so we can drop the complex conjugate sign $*$ and the variational integral becomes

$$\varepsilon = \frac{\mathbf{c}^T \mathbf{H} \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c}}$$

We now let \mathbf{c} change by a small amount $\delta \mathbf{c}$ and find the change in ε (which we will then set to zero for a stationary point). We have

$$\begin{aligned} \varepsilon + \delta \varepsilon &= \frac{(\mathbf{c} + \delta \mathbf{c})^T \mathbf{H} (\mathbf{c} + \delta \mathbf{c})}{(\mathbf{c} + \delta \mathbf{c})^T \mathbf{S} (\mathbf{c} + \delta \mathbf{c})} \\ &= \frac{\mathbf{c}^T \mathbf{H} \mathbf{c} + (\delta \mathbf{c})^T \mathbf{H} \mathbf{c} + \mathbf{c}^T \mathbf{H} \delta \mathbf{c} + (\delta \mathbf{c})^T \mathbf{H} \delta \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c} + (\delta \mathbf{c})^T \mathbf{S} \mathbf{c} + \mathbf{c}^T \mathbf{S} \delta \mathbf{c} + (\delta \mathbf{c})^T \mathbf{S} \delta \mathbf{c}} \quad (14.21) \\ &= \frac{\mathbf{c}^T \mathbf{H} \mathbf{c} + (\delta \mathbf{c})^T \mathbf{H} \mathbf{c} + \mathbf{c}^T \mathbf{H} \delta \mathbf{c} + (\delta \mathbf{c})^T \mathbf{H} \delta \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c} \left(1 + \frac{(\delta \mathbf{c})^T \mathbf{S} \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c}} + \frac{\mathbf{c}^T \mathbf{S} \delta \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c}} + \frac{(\delta \mathbf{c})^T \mathbf{S} \delta \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c}} \right)} \end{aligned}$$

Expanding the denominator by the binomial theorem and retaining only first order in $\delta \mathbf{c}$ we find

$$\varepsilon + \delta \varepsilon = \frac{(\mathbf{c}^T \mathbf{H} \mathbf{c} + (\delta \mathbf{c})^T \mathbf{H} \mathbf{c} + \mathbf{c}^T \mathbf{H} \delta \mathbf{c} + \dots)}{\mathbf{c}^T \mathbf{S} \mathbf{c}} \left(1 - \frac{(\delta \mathbf{c})^T \mathbf{S} \mathbf{c} + \mathbf{c}^T \mathbf{S} \delta \mathbf{c}}{\mathbf{c}^T \mathbf{S} \mathbf{c}} + \dots \right) \quad (14.22)$$

and after multiplying out

$$(\varepsilon + \delta \varepsilon) \mathbf{c}^T \mathbf{S} \mathbf{c} = (\delta \mathbf{c})^T (\mathbf{H} \mathbf{c} - \varepsilon \mathbf{S} \mathbf{c}) + (\mathbf{c}^T \mathbf{H} - \varepsilon \mathbf{c}^T \mathbf{S}) \delta \mathbf{c} + \dots \quad (14.23)$$

The two terms on the right-hand side are simply matrix transposes of each other; they carry the same information. I did not place any requirements on $\delta \mathbf{c}$: it is quite arbitrary and so for a stationary point we must have

$$\mathbf{H} \mathbf{c} = \varepsilon \mathbf{S} \mathbf{c}$$

This is called a *generalized matrix eigenvalue problem* and there are exactly n possible solutions $\mathbf{c}_1, \varepsilon_1; \mathbf{c}_2, \varepsilon_2; \dots \mathbf{c}_n, \varepsilon_n$. Each of the n solutions is an upper bound to n electronic states (including the ground state), and so the linear variation method has the added bonus of giving approximations to each of n states simultaneously. Not only that, but if we add further Ψ 's and repeat the calculation, each of the energies will at worst stay the same or possibly get closer to the true energy. I have illustrated *MacDonald's theorem* for the case $n = 6$ in Figure 14.2.

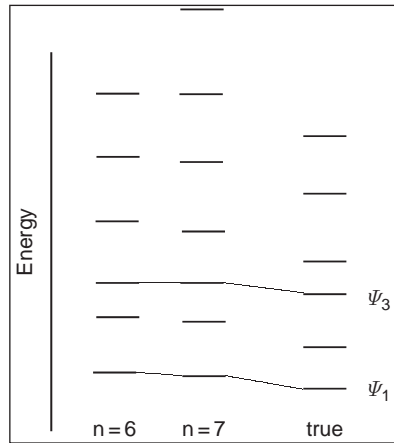


Figure 14.2 MacDonald's theorem

Addition of Ψ_7 has had no effect on the $n = 6$ approximation for Ψ_3 but has (for example) lowered the energy of the Ψ_1 approximation.

I have stressed the special property of atoms; they are spherical and their electronic Hamiltonian commutes with the square of the angular momentum operator, together with the z component. The Schrödinger equation makes no mention of electron spin and so the Hamiltonian also commutes with the square of the spin angular momentum operator and the z component. In Table 14.4, I have summarized the relevant quantum numbers for the six states of helium considered.

Table 14.4 Angular momentum quantum numbers for helium states

	L	M_L	S	M_S
Ψ_1	0	0	0	0
Ψ_2	0	0	0	0
Ψ_3	0	0	1	1
Ψ_4	0	0	1	0
Ψ_5	0	0	1	-1
Ψ_6	0	0	0	0

It can be shown (e.g. Eyring *et al.* 1944) that if Ψ_i and Ψ_j are eigenfunctions of an operator \hat{A} that commutes with the Hamiltonian \hat{H} , and that the eigenvalues are different

$$\hat{A}\Psi_i = a_i\Psi_i$$

$$\hat{A}\Psi_j = a_j\Psi_j$$

$$a_i \neq a_j$$

then

$$\int \Psi_i \hat{H} \Psi_j \, d\tau = 0$$

This means that the Hamiltonian matrix will have a simple form

$$\begin{pmatrix} H_{11} & H_{12} & 0 & 0 & 0 & H_{16} \\ H_{21} & H_{22} & 0 & 0 & 0 & H_{26} \\ 0 & 0 & H_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & H_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & H_{55} & 0 \\ H_{61} & H_{62} & 0 & 0 & 0 & H_{66} \end{pmatrix}$$

so instead of a 6×6 matrix eigenvalue problem, we have a 3×3 and $3 @ 1 \times 1$. There is nothing particularly hard about solving a 6×6 matrix eigenvalue problem, but this simple example demonstrates how angular momentum can be used to help with atomic calculations. In the case of polyatomic molecules, things are not so easy. Angular momentum operators do not generally commute with molecular Hamiltonians and so molecular problems are much harder than atomic ones. Spin operators commute with molecular Hamiltonians, as do symmetry operators.

14.7 Slater Determinants

Solution of the appropriate electronic Schrödinger equation is only one aspect of a calculation; we have to also take account of electron spin and because electrons are fermions the electronic wavefunction has to satisfy the Pauli principle. Neither electron spin nor the Pauli principle appears from the Schrödinger treatment. As I mentioned in Chapters 12 and 13, Pauli's principle can be stated in a number of different ways; I am going to restate it as

Electronic wavefunctions must be antisymmetric to exchange of electron names.

I produced the helium orbital wavefunctions in Table 14.1 by a somewhat *ad hoc* method; I constructed suitable spatial parts and spin parts, which I combined together in such a way that the Pauli principle was satisfied. A more systematic method for constructing antisymmetric orbital wavefunctions is needed.

Electron spin can be conveniently treated by combining spatial orbitals with the spin functions α and β ; for a given spatial orbital $\psi(\mathbf{r})$ we work with two space and spin wavefunctions that we write $\psi(\mathbf{r})\alpha(s)$ and $\psi(\mathbf{r})\beta(s)$; these are usually called *spinorbitals*

and electrons are allocated to spinorbitals. A spinorbital can hold a maximum of one electron. Suppose then that we have four electrons that we wish to allocate to the four spinorbitals $\psi_A(\mathbf{r})\alpha(s)$, $\psi_B(\mathbf{r})\alpha(s)$, $\psi_A(\mathbf{r})\beta(s)$ and $\psi_B(\mathbf{r})\beta(s)$. One possible allocation is

$$\psi_A(\mathbf{r}_1)\alpha(s_1)\psi_A(\mathbf{r}_2)\beta(s_2)\psi_B(\mathbf{r}_3)\alpha(s_3)\psi_B(\mathbf{r}_4)\beta(s_4)$$

but we must now allow for the indistinguishability of the electrons and take account of all the remaining $4! - 1$ permutations of electrons through the spinorbitals, for example

$$\psi_A(\mathbf{r}_2)\alpha(s_2)\psi_A(\mathbf{r}_1)\beta(s_1)\psi_B(\mathbf{r}_3)\alpha(s_3)\psi_B(\mathbf{r}_4)\beta(s_4)$$

which has been obtained from the original allocation by permuting the names of electrons 1 and 2. All the $4!$ permutations have to appear with equal weight in the total wavefunction, and in order to satisfy the Pauli principle we multiply each of them by the sign of the permutation. This is -1 if we permute an even number of electrons, and $+1$ if we permute an odd number. A total orbital wavefunction that satisfies the Pauli principle will therefore be

$$\begin{aligned} \Psi = & \psi_A(\mathbf{r}_1)\alpha(s_1)\psi_A(\mathbf{r}_2)\beta(s_2)\psi_B(\mathbf{r}_3)\alpha(s_3)\psi_B(\mathbf{r}_4)\beta(s_4) \\ & - \psi_A(\mathbf{r}_2)\alpha(s_2)\psi_A(\mathbf{r}_1)\beta(s_1)\psi_B(\mathbf{r}_3)\alpha(s_3)\psi_B(\mathbf{r}_4)\beta(s_4) + \dots \end{aligned}$$

John C. Slater is credited with having noticed that these terms could be written as a determinant (of order 4, in this case), which we construct as

$$\Psi = \begin{vmatrix} \psi_A(\mathbf{r}_1)\alpha(s_1) & \psi_A(\mathbf{r}_1)\beta(s_1) & \psi_B(\mathbf{r}_1)\alpha(s_1) & \psi_B(\mathbf{r}_1)\beta(s_1) \\ \psi_A(\mathbf{r}_2)\alpha(s_2) & \psi_A(\mathbf{r}_2)\beta(s_2) & \psi_B(\mathbf{r}_2)\alpha(s_2) & \psi_B(\mathbf{r}_2)\beta(s_2) \\ \psi_A(\mathbf{r}_3)\alpha(s_3) & \psi_A(\mathbf{r}_3)\beta(s_3) & \psi_B(\mathbf{r}_3)\alpha(s_3) & \psi_B(\mathbf{r}_3)\beta(s_3) \\ \psi_A(\mathbf{r}_4)\alpha(s_4) & \psi_A(\mathbf{r}_4)\beta(s_4) & \psi_B(\mathbf{r}_4)\alpha(s_4) & \psi_B(\mathbf{r}_4)\beta(s_4) \end{vmatrix} \quad (14.24)$$

Some authors write the determinants with rows and columns interchanged, which of course leaves the value of the determinant unchanged:

$$\Psi = \begin{vmatrix} \psi_A(\mathbf{r}_1)\alpha(s_1) & \psi_A(\mathbf{r}_2)\alpha(s_2) & \psi_A(\mathbf{r}_3)\alpha(s_3) & \psi_A(\mathbf{r}_4)\alpha(s_4) \\ \psi_A(\mathbf{r}_1)\beta(s_1) & \psi_A(\mathbf{r}_2)\beta(s_2) & \psi_A(\mathbf{r}_3)\beta(s_2) & \psi_A(\mathbf{r}_4)\beta(s_2) \\ \psi_B(\mathbf{r}_1)\alpha(s_1) & \psi_B(\mathbf{r}_2)\alpha(s_2) & \psi_B(\mathbf{r}_3)\alpha(s_3) & \psi_B(\mathbf{r}_3)\alpha(s_3) \\ \psi_B(\mathbf{r}_1)\beta(s_1) & \psi_B(\mathbf{r}_2)\beta(s_2) & \psi_B(\mathbf{r}_3)\beta(s_3) & \psi_B(\mathbf{r}_4)\beta(s_4) \end{vmatrix}$$

It is an attractive property of determinants that they change sign if we interchange two rows (or columns), and this is formally equivalent to interchanging the name of two of the electrons. Also, if two columns are the same then the determinant is zero, which is formally equivalent to letting two electrons occupy the same spinorbital.

Not every electronic state of every atom or molecule can be written as a single Slater determinant and linear combinations are then needed. For example, of the wavefunctions shown in Table 14.1, we see by inspection that Ψ_1 , Ψ_3 , Ψ_5 and Ψ_6 can be written as single Slater determinants, for example

$$\Psi_1 = \sqrt{\frac{1}{2}} \begin{vmatrix} 1s(\mathbf{r}_1)\alpha(s_1) & 1s(\mathbf{r}_1)\beta(s_1) \\ 1s(\mathbf{r}_2)\alpha(s_2) & 1s(\mathbf{r}_2)\beta(s_2) \end{vmatrix} \quad \Psi_3 = \sqrt{\frac{1}{2}} \begin{vmatrix} 1s(\mathbf{r}_1)\alpha(s_1) & 2s(\mathbf{r}_1)\alpha(s_1) \\ 1s(\mathbf{r}_2)\alpha(s_2) & 2s(\mathbf{r}_2)\alpha(s_2) \end{vmatrix}$$

but Ψ_2 and Ψ_4 have to be written as a sum of two determinants, for example

$$\Psi_2 = \frac{1}{2} \left\{ \begin{vmatrix} 1s(\mathbf{r}_1)\alpha(s_1) & 2s(\mathbf{r}_1)\beta(s_1) \\ 1s(\mathbf{r}_2)\alpha(s_2) & 2s(\mathbf{r}_2)\beta(s_2) \end{vmatrix} - \begin{vmatrix} 1s(\mathbf{r}_1)\beta(s_1) & 2s(\mathbf{r}_1)\alpha(s_1) \\ 1s(\mathbf{r}_2)\beta(s_2) & 2s(\mathbf{r}_2)\alpha(s_2) \end{vmatrix} \right\}$$

14.8 Slater–Condon–Shortley Rules

Slater determinants are compact summaries of all possible permutations of electrons and spin orbitals but the way I have written them down is unwieldy, and many authors adopt simplified notations. Suppose for example we have a many-electron system whose electronic configuration can be written

$$(\psi_A)^2(\psi_B)^2 \dots (\psi_M)^2$$

This is chemical shorthand for $2M$ spinorbitals $\psi_A\alpha\psi_A\beta\psi_B\alpha\psi_B\beta \dots \psi_M\alpha\psi_M\beta$ occupied by $2M$ electrons. One convention is to write A for $\psi_A\alpha$ and \bar{A} for $\psi_A\beta$ with the Slater determinant represented

$$D = |A\bar{A}B\bar{B} \dots M\bar{M}|$$

From time to time we need to know the expectation values of sums of certain one-electron operators and certain two-electron operators. Suppose that there are n electrons; these are indistinguishable and so any sum of operators must include all of them on an equal footing. Expectation values are typically the electronic contribution to the molecular electric dipole moment

$$-e \int \Psi \left(\sum_{i=1}^n \mathbf{r}_i \right) \Psi \, d\tau$$

and the electron repulsion in a polyatomic system with n electrons is

$$\frac{e^2}{4\pi\epsilon_0} \int \int \Psi \left(\sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{r_{ij}} \right) \Psi \, d\tau_1 \, d\tau_2$$

Ψ has to be a linear combination of Slater determinants D_1, D_2, \dots so we need a systematic set of rules for working out such expectation values between single Slater determinants that I will call D_1 and D_2 which are, in the simplified notation

$$D_1 = |UVW \dots Z|$$

$$D_2 = |U'V'W' \dots Z'|$$

The first step is to rearrange one of the determinants to make as many of the spinorbitals equal as possible. This may introduce a sign change. The algebra is easier if we assume that the individual orbitals are normalized and orthogonal. Consider the overlap integral

$$\int D_1 D_2 \, d\tau$$

where the integration is over the space and spin coordinates of all the electrons. Each determinant expands into $n!$ terms, and the product has $(n!)^2$ terms. On integration, orthonormality of the individual spinorbitals means that there will be just $n!$ nonzero terms each equal to

1 with all remaining terms zero. If at least one of the spinorbitals are different, say $U \neq U'$, then the complete overlap integral is zero. Thus

$$\int D_1 D_2 \, d\tau = \begin{cases} n! & \text{if } D_1 = D_2 \\ 0 & \text{otherwise} \end{cases}$$

This gives a Slater determinant normalizing factor of $1/\sqrt{n!}$.

The rules for sums of one- and two-electron operators can be found in more advanced texts such as Eyring *et al.* (1944); all we need to note are the following results.

1. If two or more spinorbitals are zero, then the expectation value of a sum of one-electron operators is zero.
2. If three or more spinorbitals are different, then the expectation value of a sum of two-electron operators is zero.

14.9 Hartree Model

We now return to the problem of the helium atom. We have established that the electronic wavefunction would be exactly a product of hydrogenic orbitals in the absence of electron repulsion. We have also seen that neglect of electron repulsion leads to impossibly poor agreement with experiment. The orbital model is an extremely attractive one, so the question is how can we allow for electron repulsion in some average way, whilst retaining the orbital picture? Douglas R. Hartree's solution to the problem was to allow each electron to come under the influence of an *average* potential due to the other electron and the nucleus. Suppose for the sake of argument that electrons 1 and 2 occupy orbitals ψ_A and ψ_B (which might be the same, but are to be determined). We could have a guess that the ψ 's might be hydrogenic, to get the calculation started.

According to the familiar Born interpretation, $\psi_B^2 d\tau$ is a probability and so we can regard the second electron as a charge distribution, with density $-e\psi_B^2$. Electron 1 therefore sees a potential due to the nucleus and the smeared-out electron 2, as shown in Figure 14.3.

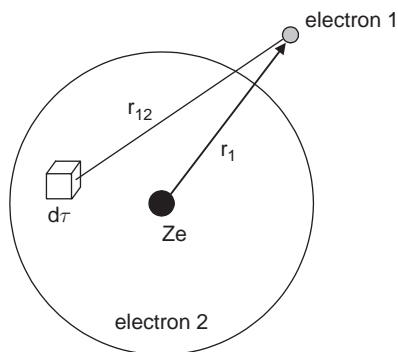


Figure 14.3 The Hartree model

The electronic Schrödinger equation for electron 1 is

$$\left(-\frac{\hbar^2}{8\pi^2 m_e} \nabla^2 - \frac{Ze^2}{4\pi \epsilon_0 r_1} + \frac{e^2}{4\pi \epsilon_0} \int \frac{\psi_B^2}{r_{12}} d\tau_B \right) \psi_A = \epsilon_A \psi_A \quad (14.25)$$

That gives us ψ_A . We now focus on electron 2, for which

$$\left(-\frac{\hbar^2}{8\pi^2 m_e} \nabla^2 - \frac{Ze^2}{4\pi \epsilon_0 r_2} + \frac{e^2}{4\pi \epsilon_0} \int \frac{\psi_A^2}{r_{12}} d\tau_A \right) \psi_B = \epsilon_B \psi_B \quad (14.26)$$

and calculate ψ_B then back to electron 1 so on. The calculation is an iterative one and we stop once the change between iterations is sufficiently small. Each electron experiences a field due to the remaining electrons, and at the end of the calculation the average electron density derived from the field must be the same as the field and so Hartree (1927) coined the phrase *self-consistent field* (SCF for short). William Hartree and his son Douglas did much of the early work and so we speak of the *Hartree self-consistent field method*.

The theory of atomic structure is dominated by angular momentum considerations, since the square of the orbital angular momentum operator and its z component commute both with each other and the electronic Hamiltonian. This simplifies the problem considerably and the Hartrees wrote each atomic orbital as

$$\psi(\mathbf{r}) = \frac{1}{r} P(nl; r) Y_{l,m_l}(\theta, \phi) \quad (14.27)$$

where $P(nl; r)$ is a radial function and Y_{lm} a spherical harmonic. Their notation for P should be clear; each shell has a different radial function. From now on I am going to make the notation more consistent with previous chapters and write

$$\psi(\mathbf{r}) = \frac{1}{r} P_{nl}(r) Y_{l,m_l}(\theta, \phi)$$

Thus for fluorine we would expect three different radial functions, $P_{1s}(r)$, $P_{2s}(r)$ and $P_{2p}(r)$. Because of the spherical symmetry of atoms, all the 2p solutions have the same radial part.

Details of the method are given in the book *The Calculation of Atomic Structures* (Hartree 1957) and essentially the radial functions are determined from the variation principle. That is to say, they are chosen so as to minimize the variational energy

$$\frac{\int \Phi^* \hat{H} \Phi d\tau}{\int \Phi^* \Phi d\tau}$$

14.10 Hartree–Fock Model

Hartree's calculations were done numerically. It soon became apparent that these early calculations gave energies that were in poor agreement with experiment; Fock (1930a,b) pointed out that Hartree had not included the Pauli principle in his method. Essentially, the *Hartree model* considered a simple orbital product such as

$$\Psi_{\text{Hartree}} = \psi_A(\mathbf{r}_1) \alpha(s_1) \psi_A(\mathbf{r}_2) \beta(s_2) \psi_B(\mathbf{r}_3) \alpha(s_3) \psi_B(\mathbf{r}_4) \beta(s_4) \quad (14.28)$$

whilst the *Hartree–Fock (HF)* model uses a fully antisymmetrized wavefunction such as

$$\Psi_{\text{Hartree–Fock}} = \begin{vmatrix} \psi_A(\mathbf{r}_1)\alpha(s_1) & \psi_A(\mathbf{r}_1)\beta(s_1) & \psi_B(\mathbf{r}_1)\alpha(s_1) & \psi_B(\mathbf{r}_1)\beta(s_1) \\ \psi_A(\mathbf{r}_2)\alpha(s_2) & \psi_A(\mathbf{r}_2)\beta(s_2) & \psi_B(\mathbf{r}_2)\alpha(s_2) & \psi_B(\mathbf{r}_2)\beta(s_2) \\ \psi_A(\mathbf{r}_3)\alpha(s_3) & \psi_A(\mathbf{r}_3)\beta(s_3) & \psi_B(\mathbf{r}_3)\alpha(s_3) & \psi_B(\mathbf{r}_3)\beta(s_3) \\ \psi_A(\mathbf{r}_4)\alpha(s_4) & \psi_A(\mathbf{r}_4)\beta(s_4) & \psi_B(\mathbf{r}_4)\alpha(s_4) & \psi_B(\mathbf{r}_4)\beta(s_4) \end{vmatrix} \quad (14.29)$$

In the simplest version, HF theory concentrates on electronic states that can be represented as a single Slater determinant. We find an extra term in the energy expression called the *exchange energy*, discussed earlier in Chapter 12. Evaluation of the energy needs a more complicated numerical procedure than the simpler Hartree theory. Inclusion of electron exchange by the numerical methods used in their day proved more and more difficult for atoms towards the bottom right-hand corner of the Periodic Table. The Hartrees were able to study a wide range of atoms in different electronic states, together with their ions. To give a flavour of systems studied, I have reproduced in Table 14.5 the systems that Hartree (1957) reports in his book, which cover the period 1948 through 1957.

Table 14.5 Selection of atoms treated in Hartree (1957)

Atom	Atom	Atom	Atom
H ⁻	Ne ²⁺	Cl ⁻	Fe ¹⁶⁺
He ⁺	Na	Cl	Zn ²⁺
Li ⁺	Na ⁺	Ca	Zr ⁴⁺
Be	Mg	Ti ²⁺	Mo ⁺
B	Al ²⁺	V ⁺	In ³⁺
C	Al ³⁺	Mn	Sb ³⁺
O ⁶⁺	S ⁻	Mn ⁺	Au ⁺
F ⁻	S	Fe ¹³⁺	Tl ⁺
			Tl ²⁺

The author mentions that all up to Mn²⁺ and also Fe¹⁶⁺ and Zn²⁺ have exchange terms included in the calculations.

14.11 Atomic Shielding Constants

Hartree–Fock wavefunctions are the best wavefunctions that are possible within the orbital model, as shown in Figure 14.4. Wavefunctions A and B are simple orbital wavefunctions, perhaps hydrogenic or improved hydrogenic. The difference between the Hartree–Fock energy and experiment is called the *correlation energy*.

Output from an atomic HF program consists of the radial function, together with data for each shell such as those shown in Table 14.6.

It is usual to work with normalized radial functions, and this determines the functions apart from their sign (for if $P_{nl}(r)$ is a solution, so is $-P_{nl}(r)$). Hartree used a convention that the radial function should be positive near the nucleus, whilst other authors use a convention that it should be positive for large r .

Several authors pointed out that it was often desirable to have simple approximations to the wavefunctions and energy levels of atoms and ions. For an atom of n electrons, there are

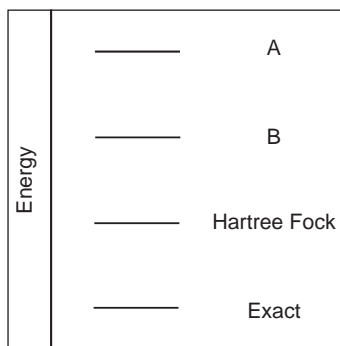


Figure 14.4 Hartree–Fock energy

Table 14.6 Properties output from an atomic HF study

ε	Orbital energy
A	Initial slope $P_{nl}(r)/r^{l+1}$, $r \rightarrow 0$
s	Screening parameter
$\langle 1/R^3 \rangle$	Expectation value of $1/R^3$
$\langle 1/R \rangle$	Expectation value of $1/R$
$\langle R \rangle$	Expectation value of R
$\langle R^2 \rangle$	Expectation value of R^2
Virial ratio	$\langle \text{Potential} \rangle / \langle \text{Kinetic} \rangle$
Spin–orbit coupling	ζ_{nl}
Orbit–orbit coupling	$M^k(nl, nl)$

$3n$ independent variables in the Schrödinger equation (plus spin). Hartree (1957) expressed the need most dramatically as follows:

One way of representing a solution quantitatively would be by a table of its numerical values, but an example will illustrate that such a table would be far too large ever to evaluate, or to use if it were evaluated. Consider, for example, the tabulation of a solution for one stationary state of Fe. Tabulation has to be at discrete values of the variables, and 10 values of each variable would provide only a very coarse tabulation; but even this would require 10^{78} entries to cover the whole field; and even though this might be reduced to, say, $5^{78} \cong 10^{53}$ by use of the symmetry properties of the solution, the whole solar system does not contain enough matter to print such a table. And, even if it could be printed, such a table would be far too bulky to use. And all this is for a single stage of ionization of a single atom.

14.11.1 Zener's Wavefunctions

In Section 14.5, we addressed ways of improving the analytical 1s orbital for helium, especially by treating the effective nuclear charge Z' as a variational parameter. In his keynote paper Zener (1930) extended this simple treatment to first-row atoms. As usual, I will let the author tell the story in his own words through the abstract.

The wave functions for the atoms Be, B, C, N, O, F and Ne are written as simple analytic expressions with several parameters. The best values of these parameters are then determined by the variation method. In the final wave functions the effective quantum number is very nearly two, the radial node is so small as to have little effect upon the charge distribution, the coefficient in the exponential is related to an empirical 'mean effective charge'.

14.11.2 Slater's Rules

Finally we return to the work of Slater (1930), and once again you might like to read his abstract.

In analogy with the method of Zener for the atoms Li to F, simple rules are set up giving approximate analytical atomic wavefunctions for all the atoms, in any stage of ionization. These are applied to X-ray levels, sizes of atoms and ions, diamagnetic susceptibility etc. In connection with ferromagnetism, it is shown that if this really depends on the existence of incomplete shells within the atoms, rather far apart in the crystal, then the metals most likely to show it would be Fe, Co, Ni and alloys of Mn and Cu (Heuser alloys).

Slater extended Zener's shielding constants for Li to F to the other atoms by adjusting the values until he got agreement with experimental results of stripped atom and X-ray levels, atom sizes and the other quantities mentioned. He noticed that Zener's wavefunctions had radial nodes but argued that they were unimportant since they come much closer to the nucleus than for hydrogen. Consequently he decided to ignore them altogether and wrote a radial part as

$$r^{n^*-1} \exp\left(-\left(\frac{Z-s}{n^*}\right)r\right) \quad (14.30)$$

where n^* is an effective quantum number and s the shielding constant. Parameters n^* and s are found by simple rules (that have become known as Slater's rules) as shown in Table 14.7 and discussed below.

Table 14.7 Slater n^* values

n	n^*
1	1
2	2
3	3
4	3.7
5	4.0
6	4.2

For determining $Z - s$, the electrons are divided into the following groups, each having a different shielding constant: 1s; 2s, 2p; 3s, 3p; 3d; 4s, 4p; 4d; 4f; 5s, 5p; etc. That is, the s and p of a given n are grouped together but the d and f are separated. The shells are considered to be arranged from inside out in the order named.

The shielding constant s is formed, for any group of electrons, from the following contributions.

1. Nothing from any shell outside the one considered.
2. An amount 0.35 from each other electron in the group considered (except the 1s group, where 0.30 is used instead).
3. If the shell considered is an s, p shell, an amount 0.85 from each electron with total quantum number less by one, and an amount 1.00 from every electron still further in. But if the shell is a d or f, an amount 1.00 from every electron inside it.

Slater gives three examples, reproduced in Table 14.8, which are worth quoting; C, Fe and Fe cation lacking a K-electron so that there is only 1 @ 1s electron.

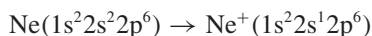
Slater's rules are still widely quoted in atomic and molecular structure theory.

Table 14.8 Slater $Z - s$ values

	C	Fe	Fe ⁺ (1s ¹)
1s	5.70 = 6 - 0.30	25.70 = 26 - 0.30	26.00
2s, 2p	3.25 = 6 - 3(0.35) - 2(0.85)	21.85 = 26 - 7(0.35) - 2(0.85)	22.70
3s, 3p		14.75 = 26 - 7(0.35) - 8(0.85) - 2(1.00)	15.75
3d		6.25 = 26 - 5(0.35) - 18(1.00)	7.25
4s		3.75 = 26 - 1(0.35) - 14(0.85) - 18(0.85) - 10(1.00)	4.75

14.12 Koopmans' Theorem

Consider the ionization process



where I have ionized a 2s electron from neon (Figure 14.5). Suppose that the energy level diagram represents a HF calculation on neon before ionization, and that the orbitals do not relax in any way after ionization. That is, the neutral atom and the ionized cation have the same HF orbitals. According to Koopmans' theorem, the ionization energy for the process is the negative of the HF orbital energy. The theorem holds for all HF orbitals.

Koopmans' Theorem
Ionization from HF orbital ψ_i
Ionization energy = -orbital energy

Ionization energies can be measured accurately by modern-day versions of the photoelectric effect. For valence shell ionization we use ultraviolet photoelectron spectroscopy (UVPES) and, as the name suggests, ultraviolet photons are the ionizing source. Inner shell electrons need X-rays for ionization.

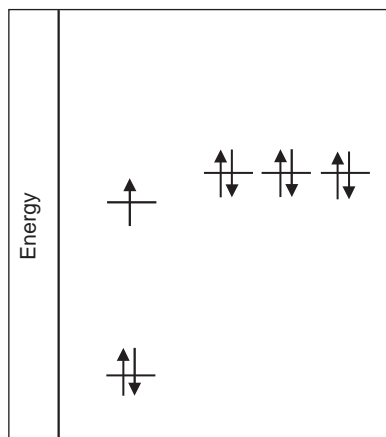


Figure 14.5 Ionized neon

References

- Eyring, H., Walter, J. and Kimball, G.E. (1944) *Quantum Chemistry*, John Wiley & Sons, Inc., New York.
- Fock, V. (1930a) *Z. Physik*, **61**, 126.
- Fock, V. (1930b) *Z. Physik*, **62**, 795.
- Frankowski, K. and Pekeris, C.L. (1966) *Phys. Rev.*, **146**, 46.
- Hartree, D.R. (1927) *Proc. Cambridge Phil. Soc.*, **24**, 89.
- Hartree, D.R. (1957) *The Calculation of Atomic Structures*, John Wiley & Sons, Inc., New York.
- Hylleraas, E. (1930) *Z. Physik*, **65**, 209.
- Slater, J.C. (1930) *Phys. Rev.*, **36**, 57.
- Zener, C. (1930) *Phys. Rev.*, **36**, 51.

15

Simple Molecules

It is time to progress to the quantum theory of molecules, and it should come as no surprise when I tell you that the orbital model is a good starting point for many of the calculations we professionals do today.

Molecules are more complicated than atoms for the following reasons.

- Whilst angular momentum considerations dominate the theory of atomic structure (it is amazing what can be done with pencil and paper before any nasty integrals have to be evaluated) the \hat{L}^2 and \hat{L}_z operators do not commute with molecular Hamiltonians (except for linear molecules; if the molecular axis is the z axis, then L_z does commute). To a certain extent, molecular symmetry operators help us since they commute with molecular Hamiltonians, but most molecules of any real chemical interest have no symmetry apart from the identity operation. So we are stuck with the Hamiltonian, and nothing to help us simplify the eigenvalue problem apart from electron spin. The Schrödinger equation for a free molecule in the absence of an applied field does not contain spin and so both \hat{S}^2 and \hat{S}_z inevitably commute with molecular Hamiltonians.
- Molecules are not spherical, and the great simplifying feature of atomic HF theory

$$\psi(\mathbf{r}) = \frac{1}{r} P_{nl}(r) Y_{l,m_l}(\theta, \phi)$$

is no longer appropriate. This statement follows from the one above.

- The Hartree–Fock (HF) limit is attainable for atoms by direct numerical integration of the HF equations. This limit is unattainable for molecules except in a few simple cases of high symmetry, and numerical integration techniques that are fine for atoms are inapplicable for molecules. In any case, we have to concern ourselves with calculations that are beyond the HF limit if we want to study chemical reactions.

15.1 Hydrogen Molecule Ion, H_2^+

The simplest molecule is the hydrogen molecule ion, H_2^+ , shown in Figure 15.1. It is formed by passing an electric discharge through dihydrogen, and it has been well studied experimentally. Its dissociation energy is known to be $D_e = 269.6 \text{ kJ mol}^{-1}$ and it has an equilibrium bond length of 106 pm. There is only one well-established electronic state, namely the ground state.

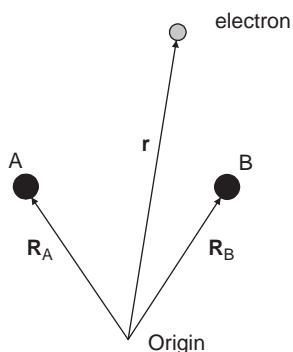


Figure 15.1 Hydrogen molecule ion

First of all, in quantum mechanics just as in classical mechanics, we can rigorously separate off the translational motion of the molecule. That leaves us to concentrate on the two nuclei (each of mass $m_p = 1.673 \times 10^{-27} \text{ kg}$) and the electron (of mass $m_e = 9.109 \times 10^{-31} \text{ kg}$) about the centre of mass. The wavefunction therefore depends on the coordinates of the electron (\mathbf{r}) and the two nuclei (\mathbf{R}_A and \mathbf{R}_B):

$$\Psi_{\text{tot}} = \Psi_{\text{tot}}(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r})$$

If we regard such a system from the viewpoint of classical mechanics, we would be tempted to try and separate the motions of the nuclei and the electron, because of the great difference in their masses (a factor of 1:1836). This does not mean that the motions are truly separable like the molecular translational motion of the centre of mass and the relative motion of the nuclei and electrons within the molecule, more that the separation can be done with only a small error.

Born and Oppenheimer (1927) first investigated this possibility and wrote

$$\Psi_{\text{tot}}(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r}) = \Psi_{\text{nuc}}(\mathbf{R}_A, \mathbf{R}_B) \psi_e(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r})$$

I am using a convention that lower case ψ refers to one electron, whilst upper case Ψ refers to many electrons. In this case there is only one electron and so the electronic wavefunction is technically an *orbital*; in this case a *molecular orbital* (MO). Born and Oppenheimer showed that the approximation was good to almost (but not exactly) the ratio of the particle masses and so we normally glue the nuclei to fixed positions in space and concentrate on the electron(s) when investigating many problems in molecular electronic structure theory.

We treat the nuclei separately; if we want to know about vibrational and rotational motion we have to solve the relevant nuclear vibrational Schrödinger equation and the relevant nuclear rotational Schrödinger equation. This is not the same as doing a MM calculation for the nuclei: there are proper quantum mechanical vibrational and rotational Schrödinger equations as we have seen in Chapter 11.

The separation of electronic and nuclear motions allows us to specify a molecular geometry; the electrons experience an electrostatic potential energy due to the nuclei, which are fixed at positions in space for the purpose of calculating the electronic wavefunction. Just as for MM, molecular geometries can be found by investigating stationary points on the potential energy surface.

The electronic wavefunction is given as the solution of an electronic Schrödinger equation

$$-\frac{\hbar^2}{8\pi^2 m_e} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi_e(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r}) + U \psi_e(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r}) = \varepsilon_e \psi_e(\mathbf{R}_A, \mathbf{R}_B, \mathbf{r}) \quad (15.1)$$

where the electrostatic potential energy is

$$U = \frac{-e^2}{4\pi\epsilon_0} \left(\frac{1}{|\mathbf{r} - \mathbf{R}_A|} + \frac{1}{|\mathbf{r} - \mathbf{R}_B|} \right)$$

We solve the electronic equation and the total energy is given by adding on the fixed nuclear repulsion

$$\varepsilon_{\text{tot}} = \varepsilon_e + \frac{e^2}{4\pi\epsilon_0 |\mathbf{R}_A - \mathbf{R}_B|} \quad (15.2)$$

The hydrogen molecule ion is unique amongst molecules in that we can solve the electronic Schrödinger equation exactly (by numerical methods) to any required accuracy, within the Born–Oppenheimer approximation. The first step is to make a change of variable to so-called *elliptic coordinates* that are defined with reference to Figure 15.2:

$$\mu = \frac{r_A + r_B}{R_{AB}} \quad \nu = \frac{r_A - r_B}{R_{AB}} \quad \phi$$

In this system of coordinates the electronic Schrödinger equation is

$$-\frac{\hbar^2}{8\pi^2 m_e} \frac{4}{R_{AB}^2 (\mu^2 - \nu^2)} \left(\frac{\partial}{\partial \mu} \left[(\mu^2 - 1) \frac{\partial}{\partial \mu} \right] + \frac{\partial}{\partial \nu} \left[(1 - \nu^2) \frac{\partial}{\partial \nu} \right] + \frac{\mu^2 - \nu^2}{(\mu^2 - 1)(1 - \nu^2)} \frac{\partial^2}{\partial \phi^2} \right) \psi_e + U \psi_e = \varepsilon_e \psi_e \quad (15.3)$$

and we look for a separation of variables solution

$$\psi_e(\mu, \nu, \phi) = M(\mu) N(\nu) \Phi(\phi)$$

Since ϕ only enters the equation in the $\partial^2/\partial\phi^2$ term, it is at once apparent that we can factor off the ϕ term. If we call the separation parameter $-\lambda^2$ we have

$$\frac{d^2\Phi}{d\phi^2} = -\lambda^2\Phi$$

and so

$$\Phi(\phi) = \frac{1}{\sqrt{2\pi}} \exp(j\lambda\phi) \quad (15.4)$$

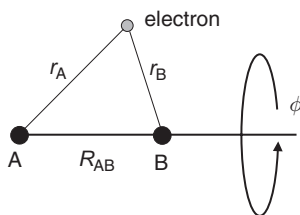


Figure 15.2 Elliptic coordinates

where λ can take on positive and negative values, and j is the square root of -1 . In the limit as $R_{AB} \rightarrow 0$, the quantum number λ becomes equivalent to the atomic quantum number m_l .

Separation into a μ and a ν equation also proves possible, and the equations have been solved by Teller (1930), by Burrau (1927) and by others leading to results in complete agreement with experiment. Solution of the differential equations is far from easy and the best references are Bates *et al.* (1953), Wind (1965) and of course Eyring *et al.* (1944).

15.2 LCAO Model

The hydrogen molecule ion is unusual amongst molecules in that we can solve the electronic problem exactly (by numerical methods). Once we consider polyelectron systems, we have to seek approximate methods. Any chemist would argue that molecules are built from atoms and so we should capitalize on this chemical knowledge by attempting to build molecular wavefunctions from atomic ones.

Suppose we build the hydrogen molecular ion starting from a hydrogen atom and a proton initially separated by a large distance. The electronic wavefunction will be a hydrogen $1s$ orbital until the proton is close enough to make any significant perturbation and so we might guess that the molecular wavefunction should resemble an atomic $1s$ orbital, at least near the appropriate nucleus.

We therefore guess that the low-energy molecular orbitals of H_2^+ might be represented as

$$\psi = c_A 1s_A + c_B 1s_B$$

where the coefficients c_A and c_B have to be determined. This technique is called the *linear combination of atomic orbitals* (LCAO) and I have used the shorthand that $1s_A$ is a hydrogen $1s$ orbital centred on nucleus A. In this particular case we can deduce the coefficients from symmetry. According to the Born interpretation, $\psi^2 d\tau$ gives the chance that an electron can be found in the volume element $d\tau$. We have

$$\psi^2 d\tau = (c_A^2 1s_A^2 + 2c_A c_B 1s_A 1s_B + c_B^2 1s_B^2) d\tau$$

Electron densities around the two hydrogen atoms have to be the same by symmetry, and the nuclear labels A and B can be used interchangeably, which means that

$$\begin{aligned}c_A^2 &= c_B^2 \\c_A &= \pm c_B\end{aligned}$$

This gives two possibilities, which I will label ψ_+ and ψ_- , and we usually write them in normalized form as

$$\begin{aligned}\psi_+ &= \frac{1}{\sqrt{2(1+S)}} (1s_A + 1s_B) \\ \psi_- &= \frac{1}{\sqrt{2(1-S)}} (1s_A - 1s_B)\end{aligned}\tag{15.5}$$

$$\begin{aligned}\text{where } S &= \int 1s_A 1s_B d\tau \\ &= \exp\left(\frac{-R_{AB}}{a_0}\right) \left(1 + \frac{R_{AB}}{a_0} + \frac{R_{AB}^2}{3a_0^2}\right)\end{aligned}$$

We can test these approximate wavefunctions by calculating the variational energies, which can be written

$$\begin{aligned}\varepsilon_{\pm} &= \varepsilon_H(1s) + \frac{e^2}{4\pi\varepsilon_0 R_{AB}} - \frac{\varepsilon_{AA} \pm \varepsilon_{AB}}{1 \pm S} \\ \varepsilon_{AA} &= \frac{e^2}{4\pi\varepsilon_0} \int \frac{1s_A^2}{r_A} d\tau \\ &= \frac{e^2}{4\pi\varepsilon_0 a_0} \left(1 - \left(1 + \frac{R_{AB}}{a_0}\right) \exp\left(-2\frac{R_{AB}}{a_0}\right)\right) \\ \varepsilon_{AB} &= \frac{e^2}{4\pi\varepsilon_0} \int \frac{1s_A^2}{r_B} d\tau \\ &= \frac{e^2}{4\pi\varepsilon_0 a_0} \left(1 + \frac{R_{AB}}{a_0}\right) \exp\left(-\frac{R_{AB}}{a_0}\right)\end{aligned}\tag{15.6}$$

The analytical solution of these integrals together with the overlap integral S (in Equation (15.5)) is covered by the classic texts such as Eyring *et al.* (1944); I have just quoted the results. This gives us the two potential energy curves shown in many elementary quantum chemistry textbooks (Figure 15.3). There is a subtle point: they are often called molecular potential energy curves because the nuclei experience just this potential.

The ground-state energy of a hydrogen atom is $-1/2E_h$ and the curves tend asymptotically to the correct limit. The upper curve describes an excited state whilst the lower curve describes the ground state. The calculated binding energy is in poor agreement with experiment (Table 15.1), whilst the equilibrium bond length is in modest agreement with experiment.

A hydrogenic 1s orbital has the form

$$1s = \sqrt{\frac{\zeta^3}{\pi a_0^3}} \exp\left(-\frac{\zeta r}{a_0}\right)\tag{15.7}$$

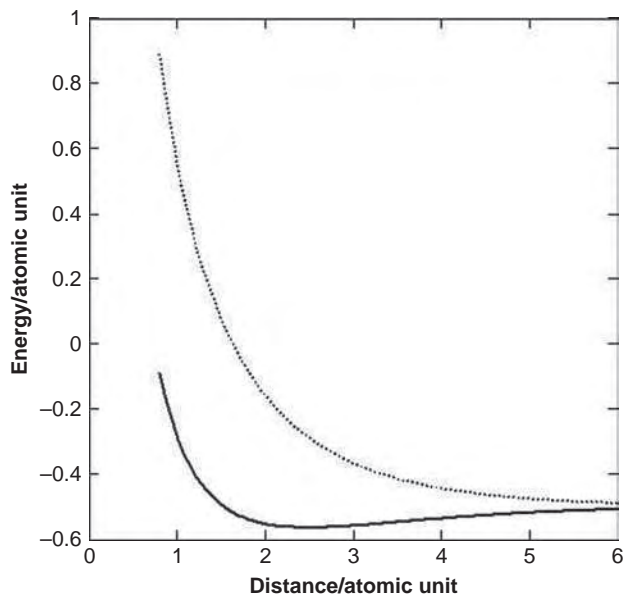


Figure 15.3 Potential energy curves for lowest states of H_2^+

Table 15.1 Hydrogen molecule ion results

	D_e/eV	R_e/pm
Experiment	2.791	106
Simple LCAO with $\zeta = 1$	1.76	132.3
Simple LCAO, best $\zeta = 1.238$	2.25	106
James, elliptic	2.772	106

where the orbital exponent $\zeta = 1$ for a hydrogen atom, 2 for He^+ and so on. The next step is to find the best value of ζ that is appropriate for a hydrogen atom within a molecule. The best value turns out to be 1.238 showing that the 1s orbital contracts a little on molecule formation. Table 15.1 shows better agreement with experiment.

15.3 Elliptic Orbitals

The best orbitals that have been obtained for H_2^+ were found by a different approach; the natural coordinates to use are the elliptic coordinates μ , ν , ϕ and James (1935) found that a good approximation to the lowest orbital is

$$\psi = \exp(-\delta\mu) (1 + c\nu^2) \quad (15.8)$$

His results are shown in Table 15.1.

The hydrogen molecular ion is interesting in that it does not have an electron pair yet it is stable. The obvious question is, can we give a hand-waving explanation for this stability, one that does not rely on the presence of electron pairs?

The obvious place to look is the electron density. The electron density plots by themselves are not particularly informative; the ground-state wavefunction shows a region of apparently enhanced electron density between the nuclei, whilst the excited-state wavefunction has less electron density in this region. It is more interesting to look at the difference between the electron density in the molecule and two ground-state hydrogen atoms, each of which contains half an electron. Such plots are called *density differences*. Positive contours correspond to a gain of electron density (in units of e) compared to two overlapped half hydrogens; negative contours correspond to a loss. The molecule lies along the horizontal axis.

The density difference plot for the ground state (Figure 15.4) illustrates an electron density enhancement in the region between the nuclei and so a stable molecule. Electron density has to be redistributed, it cannot appear from nowhere and we see that the bond region gains at the expense of regions beyond the nuclei. The excited-state wavefunction shows the opposite effect: there is a depletion of electron density in the bond region, and a gain of electron density beyond the nuclei (Figure 15.5). This explains the instability of this state, so we can give a simple electrostatic explanation for the stability of the molecule in its ground state, and its instability in the excited state.

It is a general principle that, once we have calculated an electron density from the laws of quantum mechanics, we can analyse it by the laws of classical electrostatics.

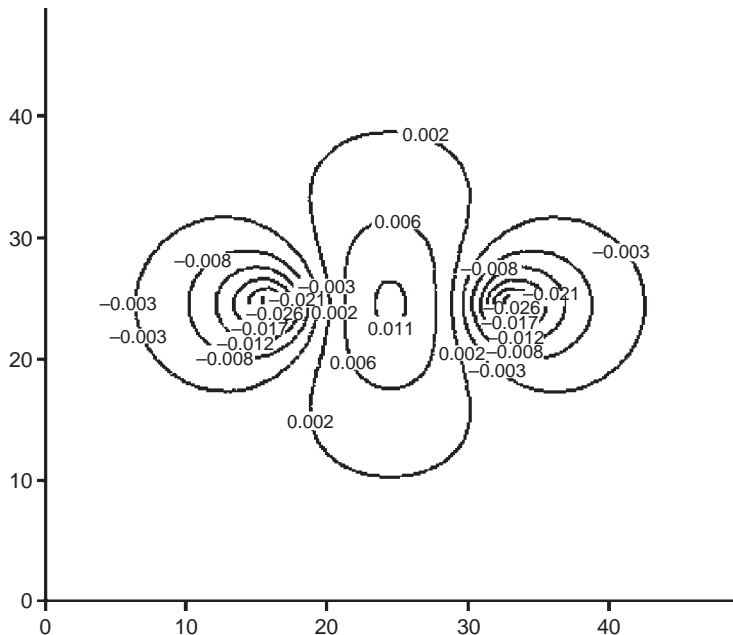


Figure 15.4 Density difference for ground state

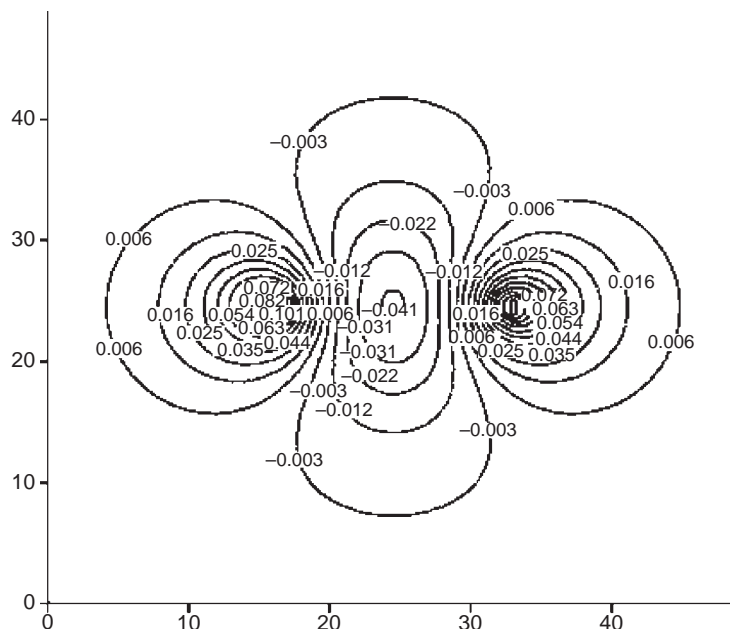


Figure 15.5 Density difference for excited state

15.4 Heitler–London Treatment of Dihydrogen

The simplest molecule of any true chemical importance is dihydrogen. The methods discussed above for the hydrogen molecule ion are not applicable to dihydrogen, because of the extra electron and the electron–electron repulsion. The first successful treatment of dihydrogen was that of Heitler and London (1927). They argued as follows. Consider a ‘reaction’



where the left-hand-side hydrogen atoms labelled A and B are initially at infinite separation. Each hydrogen atom is in its lowest electronic state and there are two spin orbitals per atom giving four in total: $1s_A\alpha$, $1s_A\beta$, $1s_B\alpha$ and $1s_B\beta$. As we bring the atoms closer and closer together, eventually their orbitals will overlap significantly and we have to take account of the Pauli principle. We need to make sure that the total wavefunction is antisymmetric to exchange of electron names. There are $(4!/2!) = 12$ possible ways of distributing two electrons amongst four atomic spin orbitals but not all such distributions satisfy the Pauli principle. Analysis along the lines given for helium in Chapter 14 gives the following unnormalized building blocks (Slater determinants):

$$D_1 = \begin{vmatrix} 1s_A(\mathbf{r}_1)\alpha(s_1) & 1s_B(\mathbf{r}_1)\beta(s_1) \\ 1s_A(\mathbf{r}_2)\alpha(s_2) & 1s_B(\mathbf{r}_2)\beta(s_2) \end{vmatrix} \quad D_2 = \begin{vmatrix} 1s_A(\mathbf{r}_1)\beta(s_1) & 1s_B(\mathbf{r}_1)\alpha(s_1) \\ 1s_A(\mathbf{r}_2)\beta(s_2) & 1s_B(\mathbf{r}_2)\alpha(s_2) \end{vmatrix}$$

$$D_3 = \begin{vmatrix} 1s_A(\mathbf{r}_1)\alpha(s_1) & 1s_B(\mathbf{r}_1)\alpha(s_1) \\ 1s_A(\mathbf{r}_2)\alpha(s_2) & 1s_B(\mathbf{r}_2)\alpha(s_2) \end{vmatrix} \quad D_4 = \begin{vmatrix} 1s_A(\mathbf{r}_1)\beta(s_1) & 1s_B(\mathbf{r}_1)\beta(s_1) \\ 1s_A(\mathbf{r}_2)\beta(s_2) & 1s_B(\mathbf{r}_2)\beta(s_2) \end{vmatrix}$$

$$D_5 = \begin{vmatrix} 1s_A(\mathbf{r}_1)\alpha(s_1) & 1s_A(\mathbf{r}_1)\beta(s_1) \\ 1s_A(\mathbf{r}_2)\alpha(s_2) & 1s_A(\mathbf{r}_2)\beta(s_2) \end{vmatrix} \quad D_6 = \begin{vmatrix} 1s_B(\mathbf{r}_1)\alpha(s_1) & 1s_B(\mathbf{r}_1)\beta(s_1) \\ 1s_B(\mathbf{r}_2)\alpha(s_2) & 1s_B(\mathbf{r}_2)\beta(s_2) \end{vmatrix}$$

It is useful to combine the D 's into spin eigenfunctions, since the spin operators commute with the molecular Hamiltonian. The advantage is that we only need take combinations of those wavefunctions having the same spinquantum numbers when seeking to improve our description of the electronic states (Table 15.2). The combination $\Psi_1 = D_1 - D_2$ is a singlet spin state and is said to represent the covalent bond, since it gives an equal sharing to the two equivalent $1s$ orbitals by the two electrons. Ψ_2 through Ψ_4 correspond to the first excited state, which is a triplet spin state. They have the same energy in the absence of an external magnetic field. Ψ_5 and Ψ_6 are called ionic terms, because they represent an electron density distribution in which both electrons are associated with the same nucleus. Heitler and London included Ψ_1 (for the electronic ground state) and Ψ_2 through Ψ_4 (for the excited triplet state) in their original calculation. The necessary integrals needed for a variational calculation are given in Heitler and London's paper, and in the paper by Sugiura (1937). We often refer to the Heitler–London approach as the *valence bond* (VB) method and I will use the two descriptors interchangeably. It was the first successful treatment of an electron pair bond.

Table 15.2 Dihydrogen elementary valence bond (Heitler–London) calculation

Ψ	Combination	S	M_S	Comment
Ψ_1	$D_1 - D_2$	0	0	Covalent ground state
Ψ_2	$D_1 + D_2$	1	0	Excited triplet state
Ψ_3	D_3	1	1	Excited triplet state
Ψ_4	D_4	1	-1	Excited triplet state
Ψ_5	D_5	0	0	Ionic term
Ψ_6	D_6	0	0	Ionic term

The energy corresponding to Ψ_1 is sometimes written

$$\begin{aligned} \varepsilon &= \frac{J + K}{1 + S^2} \\ J &= \int 1s_A(\mathbf{r}_1) 1s_B(\mathbf{r}_2) \hat{H} 1s_A(\mathbf{r}_1) 1s_B(\mathbf{r}_2) d\tau_1 d\tau_2 \\ K &= \int 1s_A(\mathbf{r}_1) 1s_B(\mathbf{r}_2) \hat{H} 1s_A(\mathbf{r}_2) 1s_B(\mathbf{r}_1) d\tau_1 d\tau_2 \\ S &= \int 1s_A(\mathbf{r}_1) 1s_B(\mathbf{r}_1) d\tau \end{aligned} \quad (15.9)$$

and early 'explanations' of chemical bonding focused on the *Coulomb* ($= J/1 + S^2$) and *exchange* ($= K/1 + S^2$) contributions to molecular energies. Several improvements were made to the simple VB treatment of dihydrogen, for example treating the orbital exponent as a variational parameter, and inclusion of the ionic terms once again correctly weighted by use of the variation principle. This latter procedure is referred to as *configuration interaction* (CI).

15.5 Dihydrogen MO Treatment

The MO treatment was given by Hellmann (1937) amongst others. He took the lowest energy MO as a combination of hydrogen $1s$ orbitals, as for H_2^+ ,

$$\psi_+ = \frac{1}{\sqrt{2(1+S)}} (1s_A + 1s_B)$$

and wrote

$$D_7 = \sqrt{\frac{1}{2!}} \begin{vmatrix} \psi_+(\mathbf{r}_1)\alpha(s_1) & \psi_+(\mathbf{r}_1)\beta(s_1) \\ \psi_+(\mathbf{r}_2)\alpha(s_2) & \psi_+(\mathbf{r}_2)\beta(s_2) \end{vmatrix}$$

The calculated bond length and dissociation energy are in poorer agreement with experiment than those obtained from the simple VB treatment (Table 15.3), and this puzzled many people at the time. It also led them to believe that the VB method was the correct way forward for the description of molecular structure; in the event, advances in computer technology and numerical linear algebra have proved them wrong, but that is for a later chapter.

Table 15.3 *Dihydrogen calculations*

Comment	D_e/eV	R_e/a_0
Experiment	4.72	1.40
Simple VB	3.14	1.64
Simple MO	2.65	1.60
James and Coolidge	4.698	1.40

But why is the simple MO treatment of dihydrogen so poor? If we expand D_7 in terms of the atomic orbitals $1s_A$ and $1s_B$ we find (apart from the normalizing constant)

$$D_7 = (1s_A(\mathbf{r}_1)1s_B(\mathbf{r}_2) + 1s_B(\mathbf{r}_1)1s_A(\mathbf{r}_2) + 1s_A(\mathbf{r}_1)1s_A(\mathbf{r}_2) + 1s_B(\mathbf{r}_1)1s_B(\mathbf{r}_2)) \\ \times (\alpha(s_1)\beta(s_2) - \alpha(s_2)\beta(s_1))$$

which is equal to the simple VB wavefunction but with ionic terms included and weighted equally to the covalent terms. The solution to the problem is to include excited states in the wavefunction, just as we did for helium, but with a variable weight. This process is configuration interaction (CI). In the limit, once all the refinements are made, the two refined treatments (VB with CI and MO with CI) give exactly the same results and so there is no particular reason to prefer one rather than the other. The starting points are different, but not the refined end results.

To summarize, Figure 15.6 shows the potential energy curves for the VB, LCAO and CI treatments. The full curve is the VB calculation, the dotted curve the MO and the dashed curve is a CI treatment.

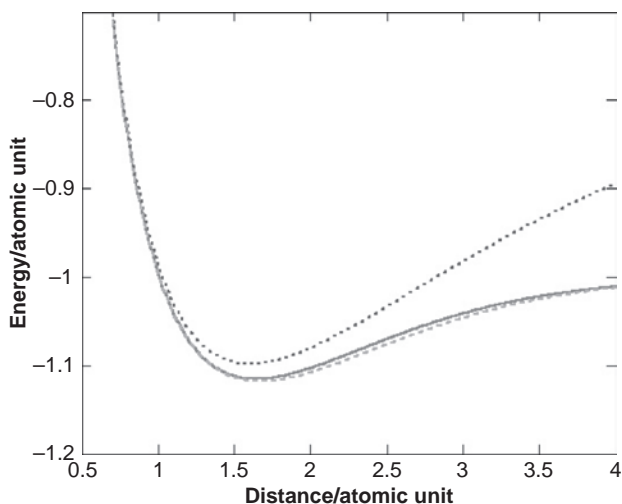


Figure 15.6 Simple dihydrogen calculations

15.6 James and Coolidge Treatment

Just as for H_2^+ , it is again found that an accurate value of the binding energy can be obtained by writing the wavefunction in terms of elliptic coordinates. James and Coolidge (1933) wrote such a wavefunction that included the interelectron distance r_{12} explicitly:

$$\psi = \exp(-\delta(\mu_1 + \mu_2)) \sum_{klmp} c_{klmp} (\mu_1^k \mu_2^l v_1^m v_2^n u^p + \mu_1^l \mu_2^k v_1^n v_2^m u^p) \quad (15.10)$$

$$\mu_1 = \frac{r_{A,1} + r_{B,1}}{R_{AB}} \quad v_1 = \frac{r_{A,1} - r_{B,1}}{R_{AB}} \quad u = \frac{2r_{12}}{R_{AB}}$$

where k , l , m , n and p are integers and the form of the function is that it is symmetric to interchange of electron names. Parameter δ is the orbital exponent. In order to make the wavefunction symmetric in the nuclear coordinates, the authors included only those terms having $(m + n)$ as an even integer. They found that a 13-term function gave essentially complete agreement with experiment.

15.7 Population Analysis

Our simple treatment of the hydrogen molecule ion was based on the bonding orbital

$$\psi_+ = \frac{1}{\sqrt{2(1+S)}} (1s_A + 1s_B)$$

which corresponds to a charge distribution of density $-e\psi_+^2$. We normally use the symbol ρ for volume charge densities and to emphasize that it depends on positions in space we write

$$\rho(\mathbf{r}) = -e(\psi_+(\mathbf{r}))^2$$

Substituting and expanding we have

$$\rho(\mathbf{r}) = -\frac{e}{2(1+S)} \left((1s_A(\mathbf{r}))^2 + (1s_B(\mathbf{r}))^2 + 2 \times 1s_A(\mathbf{r}) \times 1s_B(\mathbf{r}) \right)$$

which we can write formally in matrix language as

$$\rho(\mathbf{r}) = -e \begin{pmatrix} 1s_A(\mathbf{r}) & 1s_B(\mathbf{r}) \end{pmatrix} \begin{pmatrix} \frac{1}{2(1+S)} & \frac{1}{2(1+S)} \\ \frac{1}{2(1+S)} & \frac{1}{2(1+S)} \end{pmatrix} \begin{pmatrix} 1s_A(\mathbf{r}) \\ 1s_B(\mathbf{r}) \end{pmatrix} \quad (15.11)$$

Chemists often forget about the $-e$ since they tend to think in positive numbers of electrons, and they write the middle symmetric matrix \mathbf{P} . It is referred to as the *charge density matrix*, or the *matrix of charges and bond orders*. In this case we have

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2(1+S)} & \frac{1}{2(1+S)} \\ \frac{1}{2(1+S)} & \frac{1}{2(1+S)} \end{pmatrix}$$

From the early days of molecular quantum theory, authors have tried to divide up the electron density in chemically appealing ways, by allocating parts of the density to atom regions and parts to bond regions. If we integrate the electron density

$$\int \rho(\mathbf{r}) \, d\tau = -\frac{e}{2(1+S)} \left(\int (1s_A(\mathbf{r}))^2 \, d\tau + \int (1s_B(\mathbf{r}))^2 \, d\tau + 2 \int 1s_A(\mathbf{r}) \times 1s_B(\mathbf{r}) \, d\tau \right)$$

which must come to the number of electrons (1 in the case of hydrogen molecule ion) times $-e$. We therefore have, on doing the integrals and keeping the terms in order

$$\int \rho(\mathbf{r}) \, d\tau = -\frac{e}{2(1+S)} (1 + 1 + 2S)$$

We interpret this by saying that the electron charge fractions

$$\frac{-e}{2(1+S)}, \quad \frac{-e}{2(1+S)} \quad \text{and} \quad \frac{-2Se}{2(1+S)}$$

are associated with atom A, with atom B and with the bond region between atoms A and B. These fractions are referred to as the *net atom and bond (or overlap) populations*. Mulliken (1955) introduced the idea of population analysis, but the ideas had already been around for a very long time. He had the idea to divide each overlap population into equal parts and allocate part to each atom formally contributing to the bond. Here is the abstract of his famous 1955 paper:

With increasing availability of good all-electron LCAO-MO wavefunctions for molecules, a systematic procedure for obtaining maximum insight from such data has become desirable. An analysis in quantitative form is given here in terms of breakdowns of the electronic population

into partial and total ‘gross atomic populations’ or into partial and total ‘net atomic populations’ together with ‘overlap populations’. ‘Gross atomic populations’ distribute the electrons almost perfectly among the various AOs of the various atoms in the molecule. From these numbers, a definite figure is obtained for the amount of promotion (e.g., from 2s to 2p) in each atom; and also for the gross charge Q on each atom if the bonds are polar. The total overlap population for any pair of atoms in a molecule is in general made up of positive and negative contributions.

Mulliken’s aim was to give rough-and-ready indices that characterize a molecular charge distribution and that can be used for comparisons between molecules. The Mulliken gross atom populations in our case are

$$\frac{1}{2(1+S)} + \frac{S}{2(1+S)} = \frac{1}{2}$$

15.7.1 Extension to Many-Electron Systems

A more careful analysis is needed when dealing with many-electron wavefunctions. Suppose Ψ is an m -electron wavefunction; then it will depend on spatial coordinates and spin coordinates

$$\Psi = \Psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \dots, \mathbf{r}_m, s_m)$$

If I use the convention that $d\tau$ is a spatial differential volume element (equal to $dx dy dz$ in Cartesian coordinates) and ds refers to the spin variable, then the Born interpretation is that

$$|\Psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \dots, \mathbf{r}_m, s_m)|^2 d\tau_1 ds_1 d\tau_2 ds_2 \dots d\tau_m ds_m$$

gives the probability of finding simultaneously electron 1 in $d\tau_1 ds_1$, electron 2 in $d\tau_2 ds_2, \dots$ electron m in $d\tau_m ds_m$. Many simple physical properties such as the electric dipole moment depend only on the occupancy of an arbitrary space–spin differential element $d\tau ds$ by any electron. We can find this by averaging over all electrons except electron number 1. Indeed, the electric dipole moment does not depend on electron spin and so we can also average over the electron spin. I can illustrate with reference to dihydrogen LCAO

$$\begin{aligned} D_7 &= \sqrt{\frac{1}{2!}} \begin{vmatrix} \psi_+(\mathbf{r}_1)\alpha(s_1) & \psi_+(\mathbf{r}_1)\beta(s_1) \\ \psi_+(\mathbf{r}_2)\alpha(s_2) & \psi_+(\mathbf{r}_2)\beta(s_2) \end{vmatrix} \\ &= \sqrt{\frac{1}{2}} \psi_+(\mathbf{r}_1)\psi_+(\mathbf{r}_2) \{\alpha(s_1)\beta(s_2) - \beta(s_1)\alpha(s_2)\} \\ D_7^2 &= \frac{1}{2} \{\psi_+(\mathbf{r}_1)\psi_+(\mathbf{r}_2)\}^2 \{\alpha^2(s_1)\beta^2(s_2) + \beta^2(s_1)\alpha^2(s_2) - 2\alpha(s_1)\alpha(s_2)\beta(s_1)\beta(s_2)\} \end{aligned}$$

If we average over electron 2 (i.e. integrate with respect to its space and spin coordinates) we get

$$\int D_7^2 d\tau_2 ds_2 = \frac{1}{2} \psi_+(\mathbf{r}_1)^2 \{\alpha^2(s_1) + \beta^2(s_1)\}$$

and if we are not particularly interested in the spin variable we find, on integration over the spin of electron 1

$$\int D_7^2 d\tau_2 ds_2 ds_1 = \{\psi_+(\mathbf{r}_1)\}^2 \quad (15.12)$$

I have of course assumed that all the wavefunctions are real rather than complex; it would not have made any difference to my argument.

This latter quantity Equation (15.12) times $d\tau_1$ gives the probability of finding electron 1 in $d\tau_1$ with either spin, and the other electrons anywhere and again with either spin. Since there are two indistinguishable electrons in dihydrogen, the total electron density must be twice my result:

$$\rho(\mathbf{r}) = -2e \{\psi_+(\mathbf{r})\}^2$$

The charges and bond orders matrix is therefore just twice what we found for the hydrogen molecular ion:

$$\mathbf{P}^{\text{LCAO}} = \begin{pmatrix} \frac{1}{1+S} & \frac{1}{1+S} \\ \frac{1}{1+S} & \frac{1}{1+S} \end{pmatrix} \quad (15.13)$$

A corresponding analysis for the VB function gives

$$\mathbf{P}^{\text{VB}} = \begin{pmatrix} \frac{1}{1+S^2} & \frac{S}{1+S^2} \\ \frac{S}{1+S^2} & \frac{1}{1+S^2} \end{pmatrix} \quad (15.14)$$

and the gross Mulliken population for each hydrogen nucleus is once again 1/2, just as it should be.

There is a nice way to remember the result, provided you are happy with matrices. If we write an *overlap matrix* \mathbf{S} for the two atomic orbitals $1s_A$ and $1s_B$ as

$$\mathbf{S} = \begin{pmatrix} \int 1s_A 1s_A d\tau & \int 1s_A 1s_B d\tau \\ \int 1s_B 1s_A d\tau & \int 1s_B 1s_B d\tau \end{pmatrix}$$

then a little analysis shows that

$$\sum \sum \mathbf{P}_{ij} \mathbf{S}_{ij} = \text{number of electrons}$$

This can be restated in terms of the trace of the matrix product as

$$\text{Tr}(\mathbf{PS}) = \text{number of electrons} \quad (15.15)$$

References

- Bates, D.R., Ledsham, K. and Stewart, A.L. (1953) *Phil. Trans. R. Soc. A*, **246**, 215.
 Born, M. and Oppenheimer, J.R. (1927) *Ann. Phys.*, **84**, 457.
 Burrau, O. (1927) *Kgl Danske Videnskab Selskab*, **7**, 1.
 Eyring, H., Walter, J. and Kimball, G.E. (1944) *Quantum Chemistry*, John Wiley & Sons, Inc., New York.
 Heitler, W. and London, F. (1927) *Z. Physik*, **44**, 455.

- Hellmann, H. (1937) *Einführung in die Quantenchemie*, Franz Deuticke, p. 133.
James, H.M. (1935) *J. Chem. Phys.*, **3**, 7.
James, H. and Coolidge, M. (1933) *J. Chem. Phys.*, **1**, 825.
Mulliken, R.S. (1955) *J. Chem. Phys.*, **23**, 1833.
Sugira, Y. (1937) *Z. Physik*, **45**, 484.
Teller, E. (1930) *Z. Physik*, **61**, 458.
Wind, H. (1965) *J. Chem. Phys.*, **42**, 2371.

16

The HF-LCAO Model

In this chapter I want to explain how we progress the Hartree–Fock (HF) orbital model to more complicated molecules. The HF model encompasses the Born–Oppenheimer approximation and so we glue the nuclei into place before trying to calculate the electronic wavefunction. Each electron experiences an average field due to the remaining electrons, together with the field due to the fixed nuclei. If we are interested in geometry optimization, we have to perform HF calculations at each point on the molecular potential energy surface.

We can solve the relevant HF equation numerically for atoms to whatever accuracy we require, because of their high symmetry. In the case of a molecule, we have to resort to the LCAO approximation as discussed in Chapter 15.

In the case of dihydrogen and the hydrogen molecule ion, we were able to deduce simple forms for the lowest energy molecular orbitals

$$\psi_{\pm} = \frac{1}{\sqrt{2(1 \pm S)}} (1s_A \pm 1s_B)$$

just on the basis of symmetry considerations, assuming that only the hydrogenic 1s orbitals had a role to play. We have to use very many atomic orbitals (*basis functions*, to use the correct jargon) for polyatomic molecules and it is not possible to deduce the LCAO coefficients by hand-waving symmetry considerations. The ‘shape’ of the resulting orbitals, by which I mean the LCAO coefficients, is determined by the variation principle. We therefore have to find a variational expression for the electronic energy.

There is a third issue, namely which basis functions should we use for a polyatomic molecule; we have already met hydrogenic and Slater orbitals, but it turns out that neither of these is suitable by themselves for molecular calculations. The reason is one of pragmatism, and has to do with the difficulty of calculating certain two-electron integrals that appear in our orbital model. More of this later in the chapter.

16.1 Roothaan's 1951 Landmark Paper

Our next landmark is the paper by Roothaan (1951). In those days papers did not always have a synopsis. We can learn a great deal from the first paragraph of the introduction, as follows.

For dealing with the problems of molecular quantum mechanics, two methods of approximation have been developed which are capable of handling many-electron systems. The Heitler–London–Pauling–Slater or valence bond (VB) method originated from a chemical point of view. The atoms are considered as the material from which the molecule is built; accordingly, the molecular wave function is constructed from the wave functions of the individual atoms. The Hund–Mulliken or molecular orbital (MO) method is an extension of the Bohr theory of electron configurations from atoms to molecules. Each electron is assigned to a one-electron wave function or molecular orbital, which is the quantum mechanical analog of an electron orbit . . .

It is the purpose of this paper to build a rigorous mathematical framework for the MO method.

Within the Born–Oppenheimer approximation we consider a molecule as a set of N point charges of magnitudes eZ_1, eZ_2, \dots, eZ_N at fixed position vectors $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$. Their mutual potential energy is

$$U_{nuc} = \frac{e^2}{4\pi\epsilon_0} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Z_i Z_j}{R_{ij}} \quad (16.1)$$

Assume for the moment that we have chosen n basis functions, which could be the Slater orbitals from Chapter 14. Basis functions are usually real quantities in the mathematical sense, but complex basis functions have to be used in difficult cases such as when we have to deal with molecular magnetic properties. I will assume that we have chosen a set of real basis functions written $\chi_1(\mathbf{r}), \chi_2(\mathbf{r}), \dots, \chi_n(\mathbf{r})$. The HF-LCAO method seeks to express each orbital $\psi_1(\mathbf{r}), \psi_2(\mathbf{r}), \dots, \psi_M(\mathbf{r})$ as a linear combination of the basis functions

$$\psi_i = c_{i,1}\chi_1 + c_{i,2}\chi_2 + \dots + c_{i,n}\chi_n \quad (16.2)$$

and the process gives n LCAO-MO orbitals in total.

Roothaan's original treatment only applies to molecular electronic states where each HF-LCAO is doubly occupied (so-called 'closed shell states'). This covers the case of the vast majority of organic molecules in their electronic ground state, and we think of the electronic configuration as

$$(\psi_1)^2 (\psi_2)^2 \dots (\psi_M)^2$$

as shown in Figure 16.1. Also, the treatment is restricted to the lowest energy state of each allowed symmetry type; since most organic molecules have no symmetry, this means that only the electronic ground state can be treated. The HF-LCAO wavefunction is a single Slater determinant, and we can assume without any loss of generality that the HF-LCAO orbitals are normalized and orthogonal. Each orbital is doubly occupied and there are therefore $m = 2M$ electrons.

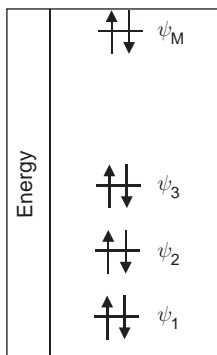


Figure 16.1 Electronic closed shell with 2M electrons

We now need to find the variational energy, which for a real wavefunction is

$$\epsilon_{el} = \frac{\int \Psi \hat{H} \Psi \, d\tau}{\int \Psi^2 \, d\tau}$$

It is neatest if I express the Hamiltonian as a sum of one-electron and two-electron operators as in Chapter 14:

$$\hat{H} = \sum_{i=1}^m \hat{h}^{(1)}(\mathbf{r}_i) + \sum_{i=1}^{m-1} \sum_{j=i+1}^m \hat{g}(\mathbf{r}_i, \mathbf{r}_j) \tag{16.3}$$

I also discussed the Slater–Condon–Shortley rules in Chapter 14. Application of these rules to our wavefunction gives

$$\begin{aligned} \epsilon_{el} = & 2 \sum_{i=1}^M \int \psi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \, d\tau_1 \\ & + \sum_{i=1}^M \sum_{j=1}^M 2 \iint \psi_i(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_j(\mathbf{r}_2) \psi_j(\mathbf{r}_2) \, d\tau_1 d\tau_2 \\ & - \sum_{i=1}^M \sum_{j=1}^M \iint \psi_i(\mathbf{r}_1) \psi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i(\mathbf{r}_2) \psi_j(\mathbf{r}_2) \, d\tau_1 d\tau_2 \end{aligned} \tag{16.4}$$

The one-electron term represents the kinetic energy of the 2M electrons and the mutual electrostatic potential energy of the electrons and the nuclei. The first of the two-electron terms is the Coulomb term; it represents the mutual electrostatic repulsion of a pair of electrons whose charge densities are $-e\psi_R^2$ and $-e\psi_S^2$. The second of the two-electron terms is the exchange term, which arises because of the fermion nature of the electrons. These equations apply for any single determinant closed shell; they do not yet contain any reference to the LCAO procedure.

16.2 The \hat{J} and \hat{K} Operators

It is sometimes useful to recast the equation as the expectation value of a sum of one-electron and certain *pseudo one-electron operators*

$$\varepsilon_e = 2 \int \sum_{i=1}^M \psi_i(\mathbf{r}_1) \left(\hat{h}^{(1)}(\mathbf{r}_1) + \hat{J}(\mathbf{r}_1) - \frac{1}{2} \hat{K}(\mathbf{r}_1) \right) \psi_i(\mathbf{r}_1) d\tau_1 \quad (16.5)$$

The operator $\hat{h}^{(1)}$ represents the kinetic energy of an electron and the nuclear attraction. The operators \hat{J} and \hat{K} are called the *Coulomb* and the *exchange operators*. They can be defined through their expectation values as follows:

$$\int \psi_R(\mathbf{r}_1) \hat{J}(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau_1 = \sum_{i=1}^M \iint \psi_R^2(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i^2(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (16.6)$$

and

$$\int \psi_R(\mathbf{r}_1) \hat{K}(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau_1 = \sum_{i=1}^M \iint \psi_R(\mathbf{r}_1) \psi_i(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_R(\mathbf{r}_2) \psi_i(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (16.7)$$

The HF Hamiltonian is a one-electron operator, defined by

$$\hat{h}^F(\mathbf{r}) = \hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) - \frac{1}{2} \hat{K}(\mathbf{r}) \quad (16.8)$$

where the coordinates \mathbf{r} refer to an arbitrary electron. HF orbitals are solutions of the eigenvalue equation

$$\hat{h}^F(\mathbf{r}) \psi(\mathbf{r}) = \varepsilon \psi(\mathbf{r})$$

16.3 HF-LCAO Equations

I am going to make use of matrices and matrix algebra for many of the derived equations, for the simple reason that they look neater than they would otherwise do.

The n basis functions are usually real and usually overlap each other and so they are not necessarily orthogonal. Following the arguments of Chapters 13 and 14, we collect together the basis function overlap integrals into an $n \times n$ real symmetric matrix \mathbf{S} that has typically an i, j element

$$S_{ij} = \int \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau_1$$

It is convenient to store the LCAO coefficients in an $n \times n$ matrix

$$\mathbf{U} = \begin{pmatrix} c_{1,1} & c_{2,1} & \dots & c_{n,1} \\ c_{1,2} & c_{2,2} & \dots & c_{n,2} \\ \vdots & \vdots & \dots & \vdots \\ c_{1,n} & c_{2,n} & \dots & c_{n,n} \end{pmatrix} \quad (16.9)$$

so that the first column collects the coefficient of the first occupied HF-LCAO orbital and so on. If we collect together the m occupied LCAO orbitals into an $n \times m$ matrix \mathbf{U}^{occ}

$$\mathbf{U}^{\text{occ}} = \begin{pmatrix} c_{1,1} & c_{2,1} & \cdots & c_{m,1} \\ c_{1,2} & c_{2,2} & \cdots & c_{m,2} \\ \vdots & \vdots & \cdots & \vdots \\ c_{1,n} & c_{2,n} & \cdots & c_{m,n} \end{pmatrix} \quad (16.10)$$

then the matrix $2\mathbf{U}^{\text{occ}}(\mathbf{U}^{\text{occ}})^T$ gives the charges and bond orders matrix \mathbf{P} discussed earlier. The next step is to express the energy in terms of the basis functions and the matrix \mathbf{P} .

The one-electron contribution is

$$2 \sum_{R=1}^M \int \psi_R(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau = 2 \sum_{R=1}^M \sum_{i=1}^n \sum_{j=1}^n c_{R,i} c_{R,j} \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau$$

If we switch the summation signs on the right-hand side we recognize elements of the charges and bond orders matrix \mathbf{P}

$$\begin{aligned} 2 \sum_{R=1}^M \sum_{i=1}^n \sum_{j=1}^n c_{R,i} c_{R,j} \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau &= \sum_{i=1}^n \sum_{j=1}^n \left\{ 2 \sum_{R=1}^M c_{R,i} c_{R,j} \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^n \left\{ P_{ij} \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \right\} \end{aligned}$$

Finally, on collecting together the one-electron integrals over the basis functions into an $n \times n$ matrix \mathbf{h}_1 whose i, j th element is

$$(\mathbf{h}_1)_{ij} = \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau_1$$

then the one-electron term emerges as the trace of the matrix product $\mathbf{P}\mathbf{h}_1$

$$\sum_{i=1}^n \sum_{j=1}^n P_{ij} (h^{(1)})_{ij} = \text{Tr}(\mathbf{P}\mathbf{h}_1) \quad (16.11)$$

A corresponding analysis shows that the two-electron terms can be written as

$$\frac{1}{2} \text{Tr} \left(\mathbf{P}(\mathbf{J} - \frac{1}{2}\mathbf{K}) \right)$$

where the elements of the matrices \mathbf{J} and \mathbf{K} depend on those of \mathbf{P} in a more complicated way:

$$\begin{aligned} J_{ij} &= \sum_{k=1}^n \sum_{l=1}^n P_{kl} \iint \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \\ K_{ij} &= \sum_{k=1}^n \sum_{l=1}^n P_{kl} \iint \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \end{aligned} \quad (16.12)$$

Many authors collect together these *Coulomb and exchange matrices* into a composite called the *electron repulsion matrix*:

$$\mathbf{G} = \mathbf{J} - \frac{1}{2}\mathbf{K}$$

The electronic energy therefore comes out as

$$\begin{aligned}\varepsilon_{\text{el}} &= \text{Tr}(\mathbf{P}\mathbf{h}_1) + \frac{1}{2}\text{Tr}(\mathbf{P}\mathbf{J}) - \frac{1}{4}\text{Tr}(\mathbf{P}\mathbf{K}) \\ &= \text{Tr}(\mathbf{P}\mathbf{h}_1) + \frac{1}{2}\text{Tr}(\mathbf{P}\mathbf{G})\end{aligned}\quad (16.13)$$

We now examine how the electronic energy changes when \mathbf{P} changes by a small amount $\delta\mathbf{P}$ (equivalent to asking how ε_{el} changes as the LCAO coefficients change). We let

$$\mathbf{P} \rightarrow \mathbf{P} + \delta\mathbf{P} \quad (16.14)$$

and after a little manipulation find the first-order change in the electronic energy

$$\begin{aligned}\delta\varepsilon_{\text{el}} &= \text{Tr}(\mathbf{h}_1\delta\mathbf{P}) + \text{Tr}(\mathbf{G}\delta\mathbf{P}) \\ &= \text{Tr}(\mathbf{h}^F\delta\mathbf{P})\end{aligned}\quad (16.15)$$

where I have defined the *Hartree–Fock Hamiltonian matrix*

$$\mathbf{h}^F = \mathbf{h}_1 + \mathbf{G} \quad (16.16)$$

We want to find $\delta\mathbf{P}$ so that $\delta\varepsilon_{\text{el}}$ is zero; the trivial solution is when $\delta\mathbf{P}$ equals the zero matrix, but there obviously have to be restrictions on the allowed form of $\delta\mathbf{P}$. I noted above the requirement that the HF-LCAO orbitals were normalized and orthogonal. If we collect together all the overlap integrals over basis functions into the matrix \mathbf{S} then a little matrix manipulation will establish that

$$\begin{aligned}\mathbf{U}^T\mathbf{S}\mathbf{U} &= \mathbf{1} \\ \mathbf{P}\mathbf{S}\mathbf{P} &= 4\mathbf{P}\end{aligned}$$

The condition that the HF-LCAO orbitals are normalized and orthogonal is equivalent to the matrix equation $\mathbf{P}\mathbf{S}\mathbf{P} = 4\mathbf{P}$ and the modified \mathbf{P} must also satisfy this condition

$$(\mathbf{P} + \delta\mathbf{P})\mathbf{S}(\mathbf{P} + \delta\mathbf{P}) = 4(\mathbf{P} + \delta\mathbf{P})$$

A little manipulation shows that, at the energy minimum

$$\mathbf{h}^F\mathbf{P} = \mathbf{P}\mathbf{h}^F \quad (16.17)$$

and whilst this does not actually help find the electron density at the minimum, it gives a condition that has to be satisfied.

Roothaan solved the problem in a different but equivalent way; he let the HF-LCAO coefficients vary subject only to the condition that the HF-LCAO orbitals remained normalized and orthogonal. He demonstrated that the coefficients \mathbf{c}_i are given by the generalized matrix eigenvalue problem

$$\mathbf{h}^F\mathbf{c}_i = \varepsilon_i\mathbf{S}\mathbf{c}_i \quad (16.18)$$

The HF matrix is $n \times n$ and there are exactly n solutions to the eigenvalue problem. The lowest energy m solutions correspond to the occupied HF-LCAO orbitals. The energy ε is called the HF orbital energy, and for each value of ε there is a column vector of HF-LCAO coefficients. Once again, this does not help us to find the coefficients, because they are contained within \mathbf{h}^F so that an iterative procedure is necessary.

16.3.1 The HF-LCAO Equations

For the record, since I will need to refer to the HF Hamiltonian many times, here they are (for the closed shell system as shown in Figure 16.1, and assuming real basis functions):

$$\begin{aligned}
 h_{ij}^F &= \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \\
 &+ \sum_{k=1}^n \sum_{l=1}^n P_{kl} \iint \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \\
 &- \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n P_{kl} \iint \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2
 \end{aligned} \quad (16.19)$$

Over the years, many workers have devised iterative procedures for solving the problem. The simplest procedure is as follows.

- Choose a molecular geometry.
- Choose a suitable basis set.
- Calculate all integrals over the basis functions and store them.
- Make an educated guess at the HF-LCAO coefficients \mathbf{U}^{occ} .
- Calculate \mathbf{P} and \mathbf{h}^F (the time-consuming step).
- Solve the matrix eigenvalue problem to give the new \mathbf{U}^{occ} .
- Check for convergence (test ε_{el} and/or \mathbf{P}).
- Exit or go back three steps.

There are other procedures. I will give you a numerical example later, and explain some of the methods that people use in order to speed up their calculation. Naturally, having progressed this far in the text, you will know that the HF-LCAO calculation simply gives one point on a molecular potential energy surface, as defined by the Born–Oppenheimer approximation. If your interest in life is molecular geometry optimization then you will have to follow the same kind of procedures as with MM in Chapter 5; there is a very big difference in that MM energies can be calculated in almost no time at all, whilst HF-LCAO energies consume considerable resource.

16.4 Electronic Energy

As noted in Equation (16.4), the HF-LCAO electronic energy is given by

$$\begin{aligned}
 \varepsilon_{\text{el}} &= \text{Tr}(\mathbf{P}\mathbf{h}_1) + \frac{1}{2} \text{Tr}(\mathbf{P}\mathbf{G}) \\
 &= \text{Tr}(\mathbf{P}\mathbf{h}^F) - \frac{1}{2} \text{Tr}(\mathbf{P}\mathbf{G})
 \end{aligned}$$

The HF-LCAO matrix eigenvalue equation is $\mathbf{h}^F \mathbf{c} = \varepsilon \mathbf{S} \mathbf{c}$, and the lowest energy m solutions determine the electronic ground state of a closed shell molecule. The sum of orbital energies ε_{orb} is therefore

$$\varepsilon_{\text{orb}} = 2(\varepsilon_A + \varepsilon_B + \cdots + \varepsilon_M)$$

$$\begin{aligned} \mathbf{h}^F \mathbf{c}_A &= \varepsilon_A \mathbf{S} \mathbf{c}_A \\ (\mathbf{c}_A)^T \mathbf{h}^F \mathbf{c}_A &= \varepsilon_A (\mathbf{c}_A)^T \mathbf{S} \mathbf{c}_A \\ \varepsilon_A &= \frac{(\mathbf{c}_A)^T \mathbf{h}^F \mathbf{c}_A}{(\mathbf{c}_A)^T \mathbf{S} \mathbf{c}_A} \\ &= (\mathbf{c}_A)^T \mathbf{h}^F \mathbf{c}_A \end{aligned}$$

and a little manipulation shows that

$$\varepsilon_{\text{el}} = \varepsilon_{\text{orb}} - \frac{1}{2} \text{Tr}(\mathbf{P}\mathbf{G}) \quad (16.20)$$

Orbital energies do not sum to give the total electronic energy. I have written the formula in matrix terms for convenience when working in the LCAO approximation. The result is true for all HF wavefunctions, whether they are numerical atomic ones (and therefore at the HF limit) or just rough and ready ones that have been through the HF treatment but are nowhere near the HF limit.

16.5 Koopmans' Theorem

Koopmans' theorem relates ionization energies to HF orbital energies, as discussed already for the atomic HF case. It is valid for any closed shell HF wavefunction, no matter how good or bad. It is subject to the following small print: the orbitals of the parent molecule and cation must remain unchanged on ionization. The fact is that the electron density invariably reorganizes on ionization, but Koopmans' theorem calculations have been widely used to interpret photoelectron spectra.

16.6 Open Shell Systems

The vast majority of organic molecules have electronic configurations of the type

$$(\psi_1)^2 (\psi_2)^2 \dots (\psi_M)^2$$

that can be described by the closed shell version of HF-LCAO procedure detailed above. A good deal of chemistry is concerned with excited triplet states, free radicals, cations and anions where there is at least one unpaired electron, and the procedure outlined above is not appropriate because the energy formula is incorrect.

The HF-LCAO procedure is a chemically attractive one since it gives a clear orbital picture and it turns out that several classes of open shell molecules can still be treated using modified forms of the basic theory. The simplest case is shown in Figure 16.2. We have a number $n_1 = M$ of doubly occupied orbitals and a number $n_2 = P - (M + 1)N$ of singly occupied orbitals, all with parallel spin electrons.

We refer to this as the *restricted open shell (ROHF)* model. The closed shell energy formula of Equation (16.4) has to be modified, and for the sake of neatness I will introduce the *occupation numbers* $v_1 (= 2)$ and $v_2 (= 1)$ for the two shells. If I use i and j as indices for the doubly occupied orbitals and u and v for the singly occupied orbitals we find

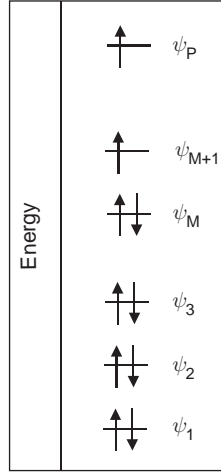


Figure 16.2 Restricted open shell HF

$$\begin{aligned}
 \varepsilon_{\text{el}} = & \nu_1 \left(\sum_{i=1}^M \int \psi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_i(\mathbf{r}_1) d\tau_1 + \frac{1}{2} \nu_1 \sum_{i=1}^M \sum_{j=1}^M \left(\iint \psi_i^2(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_j^2(\mathbf{r}_2) d\tau_1 d\tau_2 \right. \right. \\
 & \left. \left. - \frac{1}{2} \iint \psi_i(\mathbf{r}_1) \psi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i(\mathbf{r}_2) \psi_j(\mathbf{r}_2) d\tau_1 d\tau_2 \right) \right) \\
 & + \nu_2 \left(\sum_{u=M+1}^P \int \psi_u(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_u(\mathbf{r}_1) d\tau_1 + \frac{1}{2} \nu_2 \sum_{u=M+1}^P \sum_{v=M+1}^P \left(\iint \psi_u^2(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_v^2(\mathbf{r}_2) d\tau_1 d\tau_2 \right. \right. \\
 & \left. \left. - \iint \psi_u(\mathbf{r}_1) \psi_v(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_u(\mathbf{r}_2) \psi_v(\mathbf{r}_2) d\tau_1 d\tau_2 \right) \right) \quad (16.21) \\
 & + \nu_1 \nu_2 \left(\sum_{i=1}^M \sum_{u=M+1}^P \left(\iint \psi_i^2(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_u^2(\mathbf{r}_2) d\tau_1 d\tau_2 \right. \right. \\
 & \left. \left. - \frac{1}{2} \iint \psi_i(\mathbf{r}_1) \psi_u(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i(\mathbf{r}_2) \psi_u(\mathbf{r}_2) d\tau_1 d\tau_2 \right) \right)
 \end{aligned}$$

In the LCAO variation, we introduce n basis functions. We can collect the HF-LCAO coefficients of the n_1 doubly occupied orbitals into columns of a matrix \mathbf{U}_1 ($n \times n_1$) and the coefficients of the n_2 singly occupied orbitals into columns of \mathbf{U}_2 ($n \times n_2$) and define the $n \times n$ density matrices

$$\begin{aligned}
 \mathbf{R}_1 &= \mathbf{U}_1 \mathbf{U}_1^T \\
 \mathbf{R}_2 &= \mathbf{U}_2 \mathbf{U}_2^T
 \end{aligned}$$

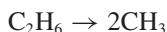
Repeating the analysis given for closed shell states gives

$$\varepsilon_{\text{el}} = \nu_1 \text{Tr} \left(\mathbf{R}_1 \left(\mathbf{h}_1 + \frac{1}{2} \mathbf{G}_1 \right) \right) + \nu_2 \text{Tr} \left(\mathbf{R}_2 \left(\mathbf{h}_1 + \frac{1}{2} \mathbf{G}_2 \right) \right) \quad (16.22)$$

where the two \mathbf{G} matrices contain Coulomb and exchange contributions similar to those defined by Equation (16.12) for the closed shell case. We then allow \mathbf{R}_1 and \mathbf{R}_2 to vary, subject to orthonormality, just as in the closed shell case. Once again, Roothaan (1960) showed how to write a Hamiltonian matrix whose eigenvectors give the columns \mathbf{U}_1 and \mathbf{U}_2 above.

16.7 Unrestricted Hartree–Fock (UHF) Model

The methyl radical CH_3 can be easily produced by breaking the C–C bond in ethane:



It has nine electrons and there are two ways we can progress, as shown in Figure 16.3.

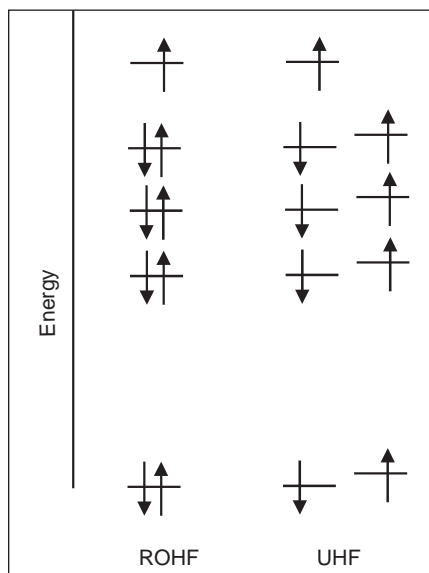


Figure 16.3 *Methyl radical*

First we can insist that the lowest energy molecular orbitals are doubly occupied, with the extra electron occupying the highest energy orbital (ROHF). On the other hand we can let the α and β spin electrons occupy different spatial orbitals (the *unrestricted Hartree–Fock* (UHF) model).

The treatment will be familiar to you by now. We write a single Slater determinant but allow the p α and the q β spin electrons to have different spatial orbitals. Application of the Slater–Condon–Shortley rules gives an energy expression in terms of the HF orbitals as

$$\begin{aligned}
\varepsilon_{\text{el}} = & \sum_{i=1}^P \int \psi_i^\alpha(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_i^\alpha(\mathbf{r}_1) d\tau_1 \sum_{j=1}^Q \int \psi_j^\beta(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \psi_j^\beta(\mathbf{r}_1) d\tau_1 \\
& + \sum_{i=1}^P \sum_{j=1}^P \iint \psi_i^\alpha(\mathbf{r}_1) \psi_i^\alpha(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_j^\alpha(\mathbf{r}_2) \psi_j^\alpha(\mathbf{r}_2) d\tau_1 d\tau_2 \\
& - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \iint \psi_i^\alpha(\mathbf{r}_1) \psi_j^\alpha(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i^\alpha(\mathbf{r}_2) \psi_j^\alpha(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (16.23) \\
& + \sum_{i=1}^Q \sum_{j=1}^Q \iint \psi_i^\beta(\mathbf{r}_1) \psi_i^\beta(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_j^\beta(\mathbf{r}_2) \psi_j^\beta(\mathbf{r}_2) d\tau_1 d\tau_2 \\
& - \frac{1}{2} \sum_{i=1}^P \sum_{j=1}^P \iint \psi_i^\beta(\mathbf{r}_1) \psi_j^\beta(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_i^\beta(\mathbf{r}_2) \psi_j^\beta(\mathbf{r}_2) d\tau_1 d\tau_2
\end{aligned}$$

There are no cross terms between the two sets of orbitals because of the orthogonality of the spin functions. We now introduce the LCAO concept; we expand each set of HF orbitals in terms of the same basis set $\chi_1, \chi_2, \dots, \chi_n$ and form two density matrices, one for the α -spin electrons P^α and one for the β -spin electrons P^β in the obvious way. We finally arrive at two linked HF-LCAO Hamiltonian matrices; the α -spin matrix has elements

$$\begin{aligned}
h_{ij}^{F,\alpha} = & \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \\
& + \sum_{k=1}^n \sum_{l=1}^n (P_{kl}^\alpha + P_{kl}^\beta) \iint \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (16.24) \\
& - \sum_{k=1}^n \sum_{l=1}^n P_{kl}^\alpha \iint \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2
\end{aligned}$$

with a similar expression for the β electrons. The UHF-LCAO orbitals are found by any of the standard techniques already discussed, for example, repeated construction of the density matrices, the two Hamiltonians, matrix diagonalization and so on until consistency is attained. The orbital energies can be associated with ionization energies using an extension of Koopmans' theorem.

16.7.1 Three Technical Points

There are three technical points of which you should be aware.

1. I chose the example CH_3 with some care, to make sure that the electronic ground state could be written as a single Slater determinant. There are some electronic states of very simple molecules where this is not possible, and more advanced considerations apply; for example, the first excited singlet state of dihydrogen cannot be represented as a single Slater determinant.
2. The UHF method does not generally give wavefunctions that are eigenfunctions of the spin operator \hat{S}^2 . The methyl radical UHF wavefunction will actually be a mixture of

various spin components ranging from the (correct) doublet state (spin quantum number $s = 1/2$) through in steps of 1 to the spin state with $s = 9/2$

$$\Psi_{\text{UHF}} = c_2\Psi_{s=1/2} + c_4\Psi_{s=3/2} + \dots \quad (16.25)$$

and very often it happens that the ‘correct’ spin state dominates ($c_2 \approx 1$ in this case, a doublet) with the largest impurity coming from the next highest spin state (c_4 in this case, a quartet). There are two strategies: the easier option is to run the UHF calculation and eliminate the largest contaminant after the UHF wavefunction has been found; the more correct but very much more complicated procedure is to eliminate all spin contaminants from the wavefunction before performing the UHF calculation. The second option leads to a linear combination of Slater determinants, such as that given in Equation (16.25), where we have to find the HF-LCAO coefficients and also the expansion coefficients c_2 , c_4 and so on. Such *extended Hartree–Fock* (EHF) calculations are very expensive in computer resource, and I will return to this idea in Chapter 19.

3. The ‘charges and bond orders matrix’ concept needs to be generalized to cover the case of open shells, because some orbitals are singly occupied (in the ROHF model) and because the spatial parts of the spinorbitals can be different (in the case of the UHF model). We collect the LCAO coefficients for the occupied α and β spin electrons separately into two matrices that I will call \mathbf{U}^α and \mathbf{U}^β . Each matrix product $\mathbf{U}^\alpha(\mathbf{U}^\alpha)^T$ and $\mathbf{U}^\beta(\mathbf{U}^\beta)^T$ gives us the charges and bond orders matrices for the individual α - and β -spin electrons. These matrices are sometimes referred to as \mathbf{P}^α and \mathbf{P}^β . The usual charge density \mathbf{P} is the sum of these, and the difference is called the spin density \mathbf{Q} .

16.8 Basis Sets

The stumbling block for molecular HF applications was the very large number of difficult integrals, especially those of the two-electron kind. A typical two-electron integral over basis functions

$$\frac{e^2}{4\pi\epsilon_0} \iint \chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2)\chi_l(\mathbf{r}_2) d\tau_1 d\tau_2$$

might involve four different basis functions, all of which might be on different atomic centres, or not even on atomic centres. Generally there will be a lot of them, there is no obvious choice of coordinate origin, the integral has a singularity (that is to say, it becomes infinite as the two electrons approach each other) and it is six dimensional. Authors therefore spoke of the *integral bottleneck*, and the bottleneck was not broken until the early 1950s.

16.8.1 Clementi and Raimondi

Slater’s atomic orbital rules were the first step to a correct treatment of basis functions that could be used in molecules. Clementi and Raimondi (1963) refined Slater’s empirical ideas by performing atomic HF-LCAO calculations on atoms with atomic number 2 through 36, in order to calculate the variationally correct orbital exponents. Their synopsis is worth reading.

The self-consistent field functions for atoms with 2 to 36 electrons are computed with a minimum basis set of Slater-type orbitals. The orbital exponents of the atomic orbitals are optimized as to ensure the energy minimum. The analysis of the optimized orbital exponents allows us to obtain simple and accurate rules for the 1s, 2s, 3s, 4s, 2p, 3p, 4p and 3d electronic screening constants. These rules are compared with those proposed by Slater and reveal the need for the screening due to the outside electrons. The analysis of the screening constants (and orbital exponents) is extended to the excited states of the ground state configuration and the positive ions.

What they did, starting from Slater's ideas of 1s, 2s, 2p, . . . atomic orbitals with modified orbital exponents (effective nuclear charges), was as follows for each atom in their set.

- Decide on the electronic ground-state configuration and choose starting values of the orbital exponents from Slater's rules.
- Optimize each orbital exponent individually by the HF-LCAO procedure. At the end of each optimization, the earlier optimized values will have changed and so . . .
- Check for self-consistency amongst the orbital exponents and either exit or go back one step.

There are better optimization procedures, as you will know from reading earlier chapters, but that is how the early workers did things.

Some of the results of Clementi and Raimondi are given in Table 16.1. We call such basis sets *single zeta* or *minimal* because they use exactly the same number of atomic orbitals as in descriptive chemistry. For each atom there is just one 1s orbital, one 2s, three 2p and so on.

Table 16.1 Comparison of Slater's exponents with those of Clementi and Raimondi (CR)

Atom	CR 1s exponent	Slater value for 1s	CR 2s exponent	CR 2p exponent	Slater value for 2s/2p
H	1	1			
He	1.6875	1.70			
Li	2.6906	2.70	0.6396		0.650
Be	3.6848	3.70	0.9560		0.975
B	4.6795	4.70	1.2881	1.2107	1.300
C	5.6727	5.70	1.6083	1.5679	1.625
N	6.6651	6.70	1.9237	1.9170	1.950
O	7.6579	7.70	2.2458	2.2266	2.275
F	8.6501	8.70	2.5638	2.5500	2.600
Ne	9.6241	9.70	2.8792	2.8792	2.925

16.8.2 Extension to Second-Row Atoms

Clementi (1964) extended this treatment to the second row in his 1964 paper, and he wrote the following in his abstract.

The self-consistent field functions for the ground state of the first and second row atoms (from He to Ar) are computed with a basis set in which two Slater-type orbitals (STO) are chosen for each atomic orbital. The reported STOs have carefully optimized orbital exponents. The total energy is not far from the accurate Hartree-Fock energy given by Clementi, Roothaan and Yoshimine for the first row atoms and unpublished data for the second-row atoms. The

obtained basis sets have sufficient flexibility to be a most useful starting set for molecular computations, as noted by Richardson. With the addition of 3d and 4f functions, the reported atomic basis sets provide a molecular basis set which duplicates quantitatively most of the chemical information derivable by the more extended basis sets needed to obtain accurate Hartree–Fock molecular functions.

Clementi thus doubled the number of atomic orbitals and so used two slightly different 1s orbitals, two slightly different 2s orbitals and so on. Once again, he optimized the orbital exponents by systematic variations and a small sample from his results is shown in Table 16.2. We refer to such a basis set as a *double zeta* basis set. Where the single zeta basis set for atomic lithium has a 1s exponent of 2.6906, the double zeta basis set has two 1s orbitals with exponents 2.4331 and 4.5177 (the inner and outer 1s orbitals).

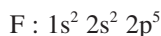
Table 16.2 *A selection from Clementi's double zeta basis set*

Atom	1s exponents	2s exponents	2p exponents
He	1.4461 2.8622		
Li	2.4331 4.5177	0.6714 1.9781	
Be	3.3370 5.5063	0.6040 1.0118	
B	4.3048 6.8469	0.8814 1.4070	1.0037 2.2086
C	5.2309 7.9690	1.1678 1.8203	1.2557 2.7263
N	6.1186 8.9384	1.3933 2.2216	1.5059 3.2674
O	7.0623 10.1085	1.6271 2.6216	1.6537 3.6813
F	7.9179 11.0110	1.9467 3.0960	1.8454 4.1710
Ne	8.9141 12.3454	2.1839 3.4921	2.0514 4.6748

The lowest energy HF-LCAO atomic orbital for a lithium atom will be a combination of the four s-type basis functions, and we call this combination ‘the’ atomic 1s orbital.

16.8.3 Polarization Functions

There is a second point to notice from Clementi's paper, where he speaks about ‘the addition of 3d and 4f functions . . .’ with reference to first- and second-row atoms respectively. Any chemist would write the fluorine electronic ground state configuration



and so the d and f orbitals are unoccupied. When atoms combine to form molecules the atomic charge density may well distort from spherical symmetry and such *polarization functions* are needed to accurately describe this distortion.

16.9 Gaussian Orbitals

None of these considerations actually broke the integrals bottleneck. There are two considerations. First is the sheer number of integrals that have to be processed. If we run a HF-LCAO calculation with n basis functions, we have to calculate $p = \frac{1}{2}n(n+1)$ one-electron integrals of each type (overlap, kinetic energy and electron–nuclear attraction). If $n = 100$ then we have to calculate 15 150 one-electron integrals. It is the two-electron integrals that cause the problem, for we have to calculate at most $q = \frac{1}{2}p(p+1)$ of them; if $n = 100$ then $q = 1.25 \times 10^7$. There are approximately $\frac{1}{8}n^4$ integrals (for large n), although this is to be seen as an upper limit. Many of the integrals turn out to be so tiny that they can be neglected for practical calculations on large molecules.

Second is the mathematical intractability of the integrals, as mentioned above. The real breakthrough came in the 1950s with the introduction of (*Cartesian*) *Gaussian basis functions*; these are similar to Slater orbitals but they have an exponential factor that goes as the *square* of the distance between the electron and the orbital centre:

$$\chi_G = N_G x^l y^m z^n \exp\left(-\alpha \frac{r^2}{a_0^2}\right) \quad (16.26)$$

Here l, m and n can have any integral values and the orbital exponent α is positive. If all of l, m and n are zero we talk about a 1s-type *Gaussian-type orbital* (GTO). If one of l, m or n is equal to 1 we have a 2p-type GTO. Expressions for normalized 1s and 2p_x GTOs are

$$G_{1s} = \left(\frac{2\alpha}{\pi a_0^2}\right)^{3/4} \exp\left(-\frac{\alpha r^2}{a_0^2}\right)$$

$$G_{2p_x} = \left(\frac{128\alpha^5}{\pi^3 a_0^{10}}\right)^{1/4} x \exp\left(-\frac{\alpha r^2}{a_0^2}\right)$$

When dealing with d-type (where $l+m+n=2$) we note that there are six possibilities, xx, xy, xz, yy, yz, zz , rather than the five combinations we normally encounter for Slater-type orbitals (STOs). The combination $(x^2 + y^2 + z^2)$ actually gives a 3s GTO, but we normally include all six Cartesian Gaussians in calculations. Similar considerations apply to the f, g, . . . orbitals.

Why use Gaussians? Let me illustrate the answer by considering the two one-dimensional normal distribution curves (which are one-dimensional Gaussians) shown in Figure 16.4:

$$G_A(x) = \exp(-\alpha_A(x-x_A)^2)$$

$$G_B(x) = \exp(-\alpha_B(x-x_B)^2)$$

I have taken $\alpha_A = 0.1$, $x_A = 1$ (the full curve) and $\alpha_B = 0.3$ and $x_B = -2$ (the dotted curve). It is easily shown that the product $G_A(x) \times G_B(x)$ is another Gaussian G_C whose centre lies between G_A and G_B :

$$G_C(x) = \exp\left(-\frac{\alpha_A\alpha_B}{\alpha_A + \alpha_B} [x_A - x_B]^2\right) \exp(-(\alpha_A + \alpha_B)(x - x_C)^2) \quad (16.27)$$

$$x_C = \frac{\alpha_A x_A + \alpha_B x_B}{\alpha_A + \alpha_B}$$

which is shown as the dashed curve in Figure 16.4. This is a general property of Gaussians.

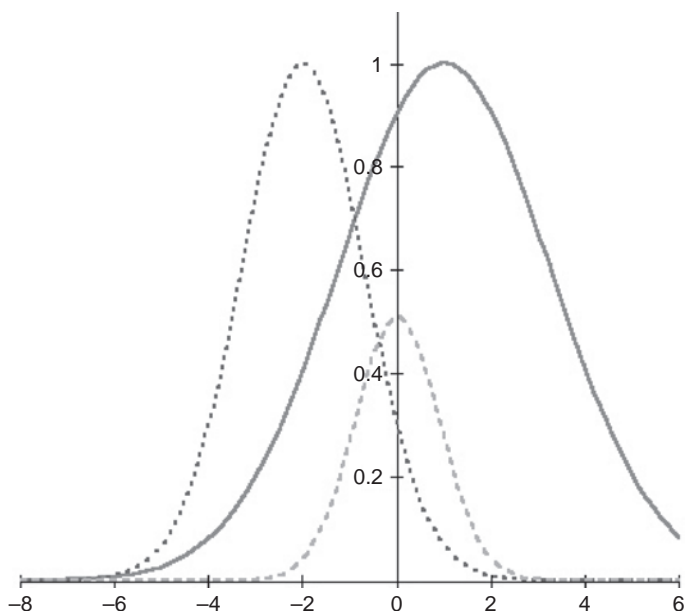


Figure 16.4 Product of two one-dimensional Gaussians

Credit for the introduction of Gaussian basis functions is usually given to S.F. Boys for his 1950 paper, and here is the famous synopsis (Boys 1950).

This communication deals with the general theory of obtaining numerical electronic wavefunctions for the stationary states of atoms and molecules. It is shown that by taking Gaussian functions, and functions derived from these by differentiation with respect to the parameters, complete systems of functions can be constructed appropriate to any molecular problem, and that all the necessary integrals can be explicitly evaluated. These can be used in connection with the molecular orbital treatment, or localized bond method, or the general method of linear combinations of many Slater determinants by the variation procedure. This general method of obtaining a sequence of solutions converging to the accurate solution is examined. It is shown that the only obstacle to the evaluation of wavefunctions of any required degree of accuracy is the labour of computation. A modification of the general method applicable to atoms is discussed and considered to be extremely practicable

GTOs have one great advantage over STOs; the nasty integrals (especially the two-electron integrals) we need for molecular quantum mechanics are relatively straightforward because they can be reduced from at most a four-centre integral to a one-centre integral by repeated application of the principle above. Consider for example an overlap integral between the two s-type GTOs shown in Figure 16.5. This is a three-dimensional extension to the one-dimensional problem discussed above.

Gaussian A has exponent α_A and is centred at \mathbf{r}_A ; Gaussian B has exponent α_B and is centred at \mathbf{r}_B . If the position vector of the electron is \mathbf{r} then its position vector relative to the centre of G_A is $\mathbf{r} - \mathbf{r}_A$, with a similar expression for G_B . The overlap integral (apart from the normalizing constants N and N') is

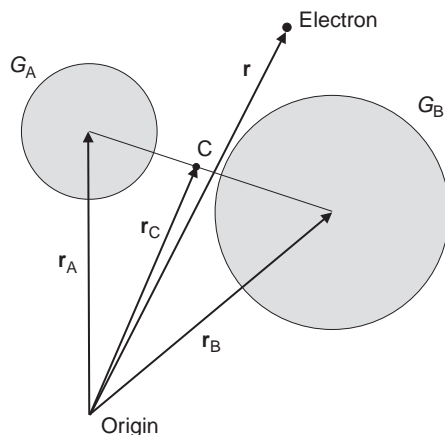


Figure 16.5 Overlap integral between two GTOs

$$\begin{aligned}
 S_{AB} &= N \int \exp\left(-\alpha_A \frac{|\mathbf{r} - \mathbf{r}_A|^2}{a_0^2}\right) \exp\left(-\alpha_B \frac{|\mathbf{r} - \mathbf{r}_B|^2}{a_0^2}\right) d\tau \\
 &= N' \int \exp\left(-\alpha_C \frac{|\mathbf{r} - \mathbf{r}_C|^2}{a_0^2}\right) d\tau
 \end{aligned} \tag{16.28}$$

The product GTO G_C has exponent α_C and centre \mathbf{r}_C given by

$$\begin{aligned}
 \alpha_C &= \alpha_A + \alpha_B \\
 \mathbf{r}_C &= \frac{1}{\alpha_A + \alpha_B} (\alpha_A \mathbf{r}_A + \alpha_B \mathbf{r}_B)
 \end{aligned}$$

The remaining integral is a product of three standard integrals

$$\begin{aligned}
 &\int \exp\left(-\alpha_C \frac{|\mathbf{r} - \mathbf{r}_C|^2}{a_0^2}\right) d\tau \\
 &= \int \exp\left(-\alpha_C \frac{(x - x_C)^2}{a_0^2}\right) dx \int \exp\left(-\alpha_C \frac{(y - y_C)^2}{a_0^2}\right) dy \int \exp\left(-\alpha_C \frac{(z - z_C)^2}{a_0^2}\right) dz
 \end{aligned}$$

One major problem that concerned early workers is that GTOs do not give terribly good energies. If we try a variational calculation on a hydrogen atom with a single s-type GTO

$$G(\alpha) = \left(\frac{\alpha}{\pi a_0^3}\right)^{3/4} \exp\left(-\alpha \frac{r^2}{a_0^2}\right)$$

and calculate the optimal Gaussian exponent, we find $\alpha_{\text{opt}} = 0.283$ with a variational energy of $-0.424E_h$. The experimental energy is $-0.5E_h$ and the error is some 15%. A second problem is that the shapes of GTOs and STOs are quite different. Figure 16.6 shows the dependences of the STO (with exponent $\zeta = 1$) and the GTO (with exponent $\alpha = 0.283$) on distance for a hydrogen atom. The plot is of wavefunction/ $a_0^{-3/2}$ versus distance from the nucleus, r/a_0 . The full curve is the STO, the dashed curve the best GTO.

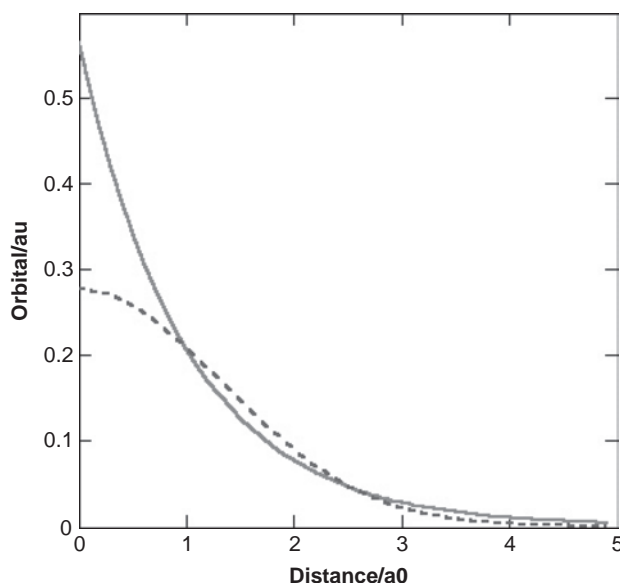


Figure 16.6 GTO versus STO for hydrogen atom

GTOs show the wrong behaviour at the nucleus, where they should have a cusp because the mutual potential energy of the electron and the nucleus becomes infinite as the distance becomes zero. GTOs also fall off far too quickly with the distance from the nucleus.

16.9.1 STO/ n G

The next step was to address the long-distance behaviour, and Hehre, Stewart and Pople proposed the idea of fitting a fixed linear combination of n GTOs to a given STO. The GTOs are not explicitly included in a HF-LCAO calculation, they are just used to give a good fit to an STO for integral evaluation. The resulting HF-LCAO orbitals can be thought of as minimal basis STOs. The GTOs are called *primitive* GTOs, and we say that the resulting atomic (STO) orbital is *contracted*. So, for example, we would use least squares fitting techniques to find the best three primitive GTO exponents α_i and contraction coefficients d_i in the STO/3G fit to a 1s STO orbital with exponent 1:

$$\text{STO}(\zeta = 1) = d_1\text{GTO}(\alpha_1) + d_2\text{GTO}(\alpha_2) + d_3\text{GTO}(\alpha_3) \quad (16.29)$$

The next keynote paper is therefore that of Hehre *et al.* (1969). As usual I will let the authors explain their ideas.

Least Squares representations of Slater-type atomic orbitals as a sum of Gaussian-type orbitals are presented. These have the special feature that common Gaussian exponents are shared between Slater-type 2s and 2p functions. Use of these atomic orbitals in self-consistent molecular-orbital calculations is shown to lead to values of atomization energies, atomic populations, and electric dipole moments which converge rapidly (with increasing size of the Gaussian expansion) to the values appropriate for pure Slater-type orbitals. The ζ exponents (or scale factors) for the atomic orbitals which are optimized for a number of molecules are

also shown to be nearly independent of the number of Gaussian functions. A standard set of ζ values for use in molecular calculations is suggested on the basis of this study and is shown to be adequate for the calculation of total and atomization energies, but less appropriate for studies of the charge distribution.

As we increase the number of primitive GTOs in the expansion, the resultant looks more and more like an STO, except at the nucleus where it can never attain the correct shape (the cusp). I show the comparison in Figure 16.7 for the STO/3G basis set.

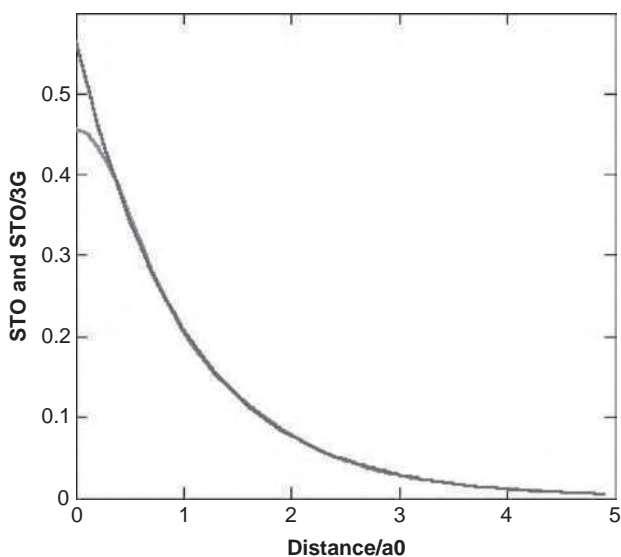


Figure 16.7 The STO/3G expansion

We therefore regard a minimal molecular basis set as comprised of STOs, except for integral evaluation where we use a linear expansion of n GTOs.

Many molecular properties depend on the valence electrons rather than the shape of the wavefunction at a nucleus, two exceptions being properties such as electron spin resonance and nuclear magnetic resonance parameters.

Tables of exponents and expansion coefficients are given in the original reference, and these all refer to an STO exponent of 1. These original GTO basis sets were 'universal' in that they applied to every atom irrespective of the atomic configuration; to convert from the STO exponent 1 to an exponent ζ you simply multiply the primitive exponents by $\zeta^{3/2}$. For reasons of computational efficiency, all basis functions in a given valence shell are taken to have the same primitive GTOs (but with different contraction coefficients).

16.9.2 STO/4-31G

Such STO/ n G calculations were common in the literature of the 1970s. It soon became apparent that they give poor results in a number of circumstances. There are particular problems for molecules containing atoms towards the end of the first period, such as oxygen

and fluorine, where they give poor bond lengths and just about every other property you can think of. Eventually it was realized that whilst most of the energy comes from the inner shell regions, some flexibility ought to be given to the valence regions. The valence orbitals are therefore split into $(n - 1)$ primitives and one primitive so we represent a hydrogen atom as two basis functions as shown in Table 16.3. We think of an inner (3 GTO) and an outer (1 GTO) basis function.

Table 16.3 *STO/4-31G hydrogen atom basis functions*

Orbital exponent	Contraction coefficient
13.00773	0.0334960
1.962079	0.22472720
0.4445290	0.8137573
0.1219492	1

For other atoms, the inner shell basis functions are left in an STO/ n G contraction. Again, you might like to read the synopsis of the keynote paper by Ditchfield *et al.* (1971).

An extended basis set of atomic functions expressed as fixed linear combinations of Gaussian functions is presented for hydrogen and the first row atoms carbon to fluorine. In this set, described as 4-31G, each inner shell is represented by a single basis function taken as a sum over four Gaussians and each valence orbital is split into inner and outer parts described by three and one Gaussian function respectively. The expansion coefficients and Gaussian exponents are determined by minimizing the total calculated energy of the electronic ground state. This basis set is then used in single-determinant molecular-orbital studies of a group of small polyatomic molecules. Optimization of valence-shell scaling factors shows that considerable rescaling of atomic functions occurs in molecules, the largest effects being observed for hydrogen and carbon. However, the range of optimum scale factors for each atom is small enough to allow the selection of a standard molecular set. The use of this standard basis gives theoretical equilibrium geometries in reasonable agreement with experiment.

16.9.3 Gaussian Polarization and Diffuse Functions

I mentioned polarization functions briefly in Section 16.8. The best thing is for me to quote the synopsis of a keynote paper by Collins *et al.* (1976) at this point.

Three basis sets (minimal s-p, extended s-p and minimal s-p with d functions on the second row atoms) are used to calculate geometries and binding energies of 24 molecules containing second row atoms. d functions are found to be essential in the description of both properties for hypervalent molecules and to be important in the calculations of two-heavy-atom bond lengths even for molecules of normal valence.

The addition of a single set of polarization functions to a heavy atom STO/4-31G basis set gives the so-called STO/4-31G*, and further addition of (p-type) polarization functions to hydrogen gives STO/4-31G** in an obvious notation. There are more explicit conventions when using more than one set of polarization functions per atom.

Polarization functions essentially allow spherical atomic charge distributions to distort on molecule formation or in the presence of external electric fields. In order to treat species that carry formal negative charges or deal with molecular properties that

depend on regions of the valence shells that are far from the nucleus, it is necessary to include *diffuse* basis functions (primitives with very small exponents), and they are denoted + and ++.

16.9.4 Extended Basis Sets

Huzinaga's set of uncontracted GTOs provides the classic example for our next topic. These large sets of primitive (uncontracted) GTOs comprise ten primitive s-type and six primitive p-type, for a first-row atom. The orbital exponents were carefully optimized for every first-row atom and for hydrogen. There is no reason in principle why we should not use them as they stand for molecular calculations but the contraction process is found to give great computational efficiency with little cost in energy. What is needed is a way of effectively grouping them together for a molecular calculation, i.e. a contraction scheme.

Many authors performed HF-LCAO calculations on atoms and small molecules, and looked for groupings of the primitive GTOs that did not change their relative weightings from orbital to orbital and from molecule to molecule. I can illustrate these ideas by mentioning Dunning's (1975) work, with the by-now inevitable synopsis.

Contracted [5s3p] and [5s4p] Gaussian basis sets for the first-row atoms are derived from the (10s6p) primitive basis sets of Huzinaga. Contracted [2s] and [3s] sets for the hydrogen atom obtained from primitive sets ranging from (4s) to (6s) are also examined. Calculations on the water and nitrogen molecules indicate that such basis sets when augmented with suitable polarization functions should yield wavefunctions near the Hartree–Fock limit.

Dunning's notation and ideas can be explained with the example in Table 16.4, for the oxygen atom.

Table 16.4 Dunning's [5s3p] contraction scheme for Huzinaga's (10s6p) GTO set

GTO type	Exponent	Contraction coefficient
s	1805.0	0.000757
	2660.0	0.006066
	585.7	0.032782
	160.9	0.132609
	51.16	0.396839
	17.90	0.542572
s	17.90	0.262490
	6.639	0.769828
s	2.077	1
s	0.7736	1
s	0.2558	1
p	49.83	0.016358
	11.49	0.106453
	3.609	0.349302
	1.321	0.657183
p	0.4821	1
p	0.1651	1

The first six primitive GTOs with exponents 1805.0 through 0.2558 contribute mostly to what we call 'the' atomic 1s orbital. The two most diffuse s functions (those with exponents 0.7736 and 0.2558) are the main components of what we call the 2s STO, and they are allowed to vary freely in molecular calculations. The 1s primitive with exponent 2.077 turns out to make substantial contributions to both the atomic 1s and 2s orbitals, so that one is left free as a separate basis function.

A typical package such as Gaussian98 will have very many basis sets as part of their database; you do not have to rediscover the wheel. On the other hand, some basis sets are good for one particular application and some are poor and there are implications of cost. The larger the basis set the higher the cost of the calculation, and the proportionality is far from linear. Choice of basis set is a specialist subject, just like many others in our subject; you simply have to take advice and look up recent literature citations.

References

- Boys, S.F. (1950) *Proc. R. Soc. A*, **200**, 542.
Clementi, E. (1964) *J. Chem. Phys.*, **40**, 1944.
Clementi, E. and Raimondi, D.L. (1963) *J. Chem. Phys.*, **38**, 2686.
Collins, J.B., Schleyer, P.v.R., Binkley, J.S. and Pople, J.A. (1976) *J. Chem. Phys.*, **64**, 5142.
Ditchfield, R., Hehre, W.J. and Pople, J.A. (1971) *J. Chem. Phys.*, **54**, 724.
Dunning Jr, T.H., (1975) *J. Chem. Phys.*, **55**, 716.
Hehre, W.J., Stewart R.F. and Pople, J.A. (1969) *J. Chem. Phys.*, **51**, 2657.
Roothaan, C.C.J. (1951) *Rev. Modern Phys.*, **23**, 161.
Roothaan, C.C.J. (1960) *Rev. Modern Phys.*, **32**, 179.

17

HF-LCAO Examples

I am going to start with L-phenylalanine, as in Chapters 4–6, to exemplify the principles discussed so far for the HF-LCAO model. There are a number of commercial packages that perform (amongst other things) HF-LCAO calculations, and everyone has their favourites. All the packages will do the simplest types of application whilst certain algorithms are only found in a subset of the available packages.

I am going to use mostly Gaussian 03 (<http://www.gaussian.com>) from this point on, but I stress that it is to a large extent a matter of personal choice and what is available on your local computer facilities.

The simplest way for me to run a job on my local Unix system, given that I already have Protein Data Bank (.pdb) files from earlier chapters, is to edit a .pdb file to produce a 'Gaussian job file' (these have extensions .gjf) and run the job. In a typical Unix environment I might type (after the \$ prompt)

```
-bash-3.00$ module load gaussian
-bash-3.00$ g03 <phenylalanine.gjf >phenylalanine.out
```

A suitable .gjf file is given below.

```
%chk=phenylalanine.chk
# HF/6-311G** Pop=Full SCF=Direct
L-phenylalanine
0 1
N 0.000 0.000 0.000
H 0.000 0.000 1.010
C 1.366 0.000 -0.483
```

etc. etc.

C 6.224 1.487 -1.290

H 7.262 1.551 -1.617

Calculations at this level of theory are iterative and resource intensive. The first statement sets up a *checkpoint file* that is used to store information about the calculation in hand, together with the results of the calculation. Calculations can be restarted from checkpoint files.

Next comes the ‘route’ through the package. I have chosen a closed-shell HF-LCAO calculation using the standard STO/6-311G** basis set. The major part of resource consumption for a HF-LCAO calculation is concerned with calculating and manipulating the two-electron integrals of which there a maximum of (about) $n^4/8$, where n is the number of basis functions. A number of strategies have been developed for dealing with them in an efficient manner. In the early days it was usual to calculate all the integrals once, before the HF iterations began, and store them on magnetic tape or exchangeable disk storage (EDS). At each iteration, the integrals were read back into computer memory, the HF matrix constructed and diagonalized. Matrix diagonalization is ‘only’ an n^2 process and even the old-fashioned Jacobi or Householder algorithms are still found in modern modelling packages.

It usually happens that many of the two-electron integrals are negligibly small, and it is possible to estimate whether an integral will fall into this category before actually calculating it. Only the significant ones then have to be calculated and stored. It is usual to store the two-electron basis function *label* i, j, k, l packed as bytes into a 32-bit word, together with the value of the integral (often multiplied by an appropriate factor of 10 and stored as an integer).

The maximum number of basis functions determines the amount of storage needed for the label; using one byte for each index limits us to $2^8 - 1 = 255$ basis functions. Raffanetti (1973) proposed an alternative way of storing the indices and the integral to make the generation of the HF-LCAO matrix easier.

For molecules with any substantial amount of symmetry, many of the integrals turn out to be plus or minus each other and many can be shown to be identically zero by symmetry arguments, without even having to calculate them. Early packages such as POLYATOM made extensive use of this symmetry feature, although it must be said that most organic molecules have no symmetry to speak of.

Such old-fashioned HF-LCAO calculations where the integrals are calculated once and then stored are usually *input-output bound* because of the finite time taken for disk or magnetic tape transfers compared to the time needed to set up the HF matrix; ideally one would like to calculate the integrals just once and hold them in RAM rather than store them externally, but early computers had very limited RAM. Technology has moved on, and many packages have a so-called *incore* option, where the integrals are calculated once at the start of the HF optimizations, and held in memory. In Gaussian 03 we would put `SCF = Incore` in the route (usage of the word ‘core’ is historical, and dates from the time when computer memory was made of magnetized ferrite rings and so people spoke about computer cores). Again, processor speed has also increased beyond belief, and it was eventually realized that a cost-effective alternative to calculation of the integrals once and

repeated read/write operations involving slow devices like disks and magnetic tape was to calculate afresh the integrals as and when needed on each cycle. That is the meaning of SCF = Direct in the route. The break-even point between Traditional and Direct calculations depends on many factors, but roughly speaking calculations involving more than 100 basis functions are best done with the Direct option. In fact this is the default option in Gaussian 03, I only wrote it out to draw attention to the issue. It is simply a historical trade-off between processor power and disk transfer rate.

The Pop = Full is simply an option to print out all the LCAO coefficients and everything else one might possibly need at the end of the calculation. The next two records are the title then the charge on the molecule (0 in this case) and the spin multiplicity (it is a singlet spin state). Then come the Cartesian coordinates in angstroms, in the usual manner. I could have also have input the molecular geometry as a Z-matrix, as I would have been obliged to do all those years ago.

17.1 Output

I actually ran this job on my laptop under MS/Windows. The output is straightforward and I will explain it one piece at a time.

```
Entering Link 1 = C:\G03W\l1.exe PID=2368.
```

```
Copyright (c) 1988,1990,1992,1993,1995,1998,2003, Gaussian, Inc.  
All Rights Reserved.
```

```
This is the Gaussian(R) 03 program. It is based on the  
the Gaussian(R) 98 system (copyright 1998, Gaussian, Inc.),  
the Gaussian(R) 94 system (copyright 1995, Gaussian, Inc.),  
the Gaussian 92(TM) system (copyright 1992, Gaussian, Inc.),  
the Gaussian 90(TM) system (copyright 1990, Gaussian, Inc.),  
the Gaussian 88(TM) system (copyright 1988, Gaussian, Inc.),  
the Gaussian 86(TM) system (copyright 1986, Carnegie Mellon  
University), and the Gaussian 82(TM) system (copyright 1983,  
Carnegie Mellon University). Gaussian is a federally registered  
trademark of Gaussian, Inc.
```

Cite this work as:

```
Gaussian 03, Revision B.01,  
M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,  
M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven,  
K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi,  
V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega,  
G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota,  
R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao,  
H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross,  
C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev,  
A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala,  
K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg,  
V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain,  
O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari,
```

J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2003.

```
*****
Gaussian 03: x86-Win32-G03RevB.01 3-Mar-2003
      25-Sep-2007
*****
```

The first part reminds us that packages such as Gaussian 03 have been developed over a number of years by a large team of people. There was a time in the 1970s when scientists freely exchanged their software, but large packages now tend to be commercial in nature and generally subject to the laws of copyright. The text box also serves the purpose of giving the appropriate citation for my calculations in this book.

The next piece of relevant output summarizes the iterative calculations

```
Rotational constants (GHZ): 2.0153582 0.6105112 0.5071802
Standard basis: 6-311G(d,p) (5D, 7F)
There are 282 symmetry adapted basis functions of A symmetry.
Integral buffers will be 262144 words long.
Raffenetti 1 integral format.
Two-electron integral symmetry is turned on.
  282 basis functions,  472 primitive gaussians,  294 cartesian basis
    functions
  44 alpha electrons  44 beta electrons
    nuclear repulsion energy 695.8408871868 Hartrees.
NAtoms=23 NActive=23 NUniq=23 SFac=1.00D+00 NATFMM=60 Big=F
One-electron integrals computed using PRISM.
NBasis=282 RedAO=T NBF=282
NBsUse=282 1.00D-06 NBFU=282
Harris functional with IExCor=205 diagonalized for initial guess.
ExpMin=1.03D-01 ExpMax=8.59D+03 ExpMxC=1.30D+03 IAcc=1 IRadAn=1
AccDes= 1.00D-06
HarFok: IExCor=205 AccDes=1.00D-06 IRadAn=1 IDoV=1
ScaDFX=1.000000 1.000000 1.000000 1.000000
Initial guess orbital symmetries:
.....
.....
The electronic state of the initial guess is 1-A.
Warning! Cutoffs for single-point calculations used.
Requested convergence on RMS density matrix=1.00D-04 within 128 cycles.
Requested convergence on MAX density matrix=1.00D-02.
Requested convergence on energy=5.00D-05.
No special actions if energy rises.
SCF Done: E(RHF)=-551.433305084 A.U. after 7 cycles
  Convg=0.7930D-05          -V/T=1.9998
  S**2=0.0000
```

You should by now know about most things in the text above apart from INDO. I will explain INDO in a later chapter; just accept for the minute that it is a cheap and cheerful way of setting off the HF-LCAO iterative calculations. The S**2 item is the expectation value of the spin operator and it shows that we are indeed dealing with a pure singlet spin state. This should not come as a surprise, since we required pairs of electrons to occupy the same spatial orbital. The virial ratio $-V/T$ is the ratio of the expectation values of the potential and kinetic energies, and for an exact wavefunction it would equal 2. It is a quantity of some historical interest.

A sensible choice of the initial electron density can make all the difference between success and failure; one strategy is to allow a few of the highest occupied and the lowest unoccupied orbitals to be partially occupied, in order to try and sort out the ground state from the excited states. Most packages make such choices as a matter of course.

The final energy shown is the total, that is to say electronic plus nuclear repulsion; the quantity quoted is technically the reduced energy, ε/E_h . Next come the orbital energies (called not surprisingly EIGENVALUES since they are eigenvalues of the generalized HF-LCAO matrix eigenvalue equation) and LCAO coefficients; a small subset is shown in the next excerpt.

Molecular Orbital Coefficients

	1	2	3	4	5
	(A) -- O	(A) -- O	(A) -- O	(A) -- O	(A) -- O
EIGENVALUES --	-20.62766	-20.55967	-15.54226	-11.39650	-11.28230
1 1 N 1S	0.00000	0.00000	0.55763	0.00000	-0.00009
2 2S	0.00000	0.00000	0.46824	0.00000	-0.00014
3 2PX	0.00000	0.00000	0.00007	0.00001	-0.00009
4 2PY	0.00000	0.00000	0.00140	-0.00001	0.00006
5 2PZ	0.00000	0.00000	0.00105	0.00000	0.00000
6 3S	-0.00002	-0.00003	0.00519	-0.00003	0.00025
7 3PX	0.00000	0.00001	-0.00004	-0.00003	-0.00042
8 3PY	0.00001	-0.00001	-0.00055	0.00002	0.00024
9 3PZ	-0.00001	0.00000	-0.00040	-0.00001	-0.00003
10 4S	0.00015	0.00032	-0.00177	0.00021	0.00048
11 4PX	-0.00002	-0.00006	0.00045	-0.00008	-0.00022
12 4PY	0.00008	0.00011	-0.00022	0.00003	0.00009
13 4PZ	0.00000	0.00002	-0.00001	0.00009	0.00003

The '(A)' refers to the point group symmetry (there is none), the '--O' means an occupied orbital. The first 12 MOs describe the heavy atom inner shells. Koopmans' theorem applies to all HF wavefunctions, no matter whether they are at the Hartree-Fock (HF) limit or just minimal basis set ones. The orbital energies can therefore be used to predict and rationalize ionization energies.

Mulliken atomic charges:

1	N	-0.467251
2	H	0.193008
3	C	-0.102938

etc. etc.

```
21 H 0.084857
22 C -0.103404
23 H 0.093449
```

Sum of Mulliken charges=0.00000

Above come Mulliken population analysis indices. As explained in earlier chapters, these give a rough-and-ready decomposition of the molecular charge density, but the Mulliken partitioning scheme is at best subjective and the numbers have to be treated with caution; the numbers should only be used when comparing similar molecules calculated with the same basis set.

Finally we have a number of molecular properties, thought for the day and most important of all, timing for the calculation.

Dipole moment (field-independent basis, Debye):

```
X=-0.5064 Y=1.2770 Z=0.0362 Tot=1.3743
```

Quadrupole moment (field-independent basis, Debye-Ang):

```
XX=-75.3438 YY=-64.7431 ZZ=-75.3306
XY=0.6908 XZ=-4.1359 YZ=-3.8869
```

Traceless Quadrupole moment (field-independent basis, Debye-Ang):

```
XX=-3.5380 YY=7.0627 ZZ=-3.5248
XY=0.6908 XZ=-4.1359 YZ=-3.8869
```

Octapole moment (field-independent basis, Debye-Ang**2):

```
XXX=-1.5522 YYY=32.4958 ZZZ=-1.2096 XYY=-0.4675
XXY=-1.0455 XXZ=-6.3241 XZZ=13.0673 YYZ=2.1498
YYZ=9.8991 XYZ=-6.7480
```

Hexadecapole moment (field-independent basis, Debye-Ang**3):

```
XXXX=-2014.3972 YYYY=-555.8933 ZZZZ=-188.9884 XXXY=-20.3528
XXXZ=-65.9431 YYYX=46.4166 YYYZ=-5.4711 ZZZX=-2.5108
ZZZY=-0.3561 XXYX=-464.9388 XXZZ=-423.3291 YYZZ=-138.1946
XXYZ=-52.7118 YYXZ=19.1827 ZZXY=2.9326
N-N=6.958408871868D+02 E-N=-2.683656283961D+03
KE=5.515438104516D+02
```

THE ONLY DIFFERENCE BETWEEN ECCENTRICS AND JOGGERS
IS THAT JOGGERS WEAR SWEATBANDS WHEN THEY RUN IN THE RAIN.

Job cpu time: 0 days 0 hours 6 minutes 2.0 seconds.

File lengths (MBytes): RWF=51 Int=0 D2E=0 Chk=12 Scr=1

17.1.1 GaussView, the Graphical User Interface for Gaussian 03

The early versions of Gaussian did not have a GUI, and geometry input by the tedious Z-matrix was normal. A number of third-party GUIs appeared, which usually permitted constructions of the molecule on-screen. Their 'output' was a .gjf, although this was often hidden from the user.

GaussView is *the* GUI for Gaussian 03. The latest version is GaussView 4, and the MS/Windows version is GaussViewW. It allows you to build molecules, submit jobs and visualize output. The screenshot shown in Figure 17.1 will give the flavour.

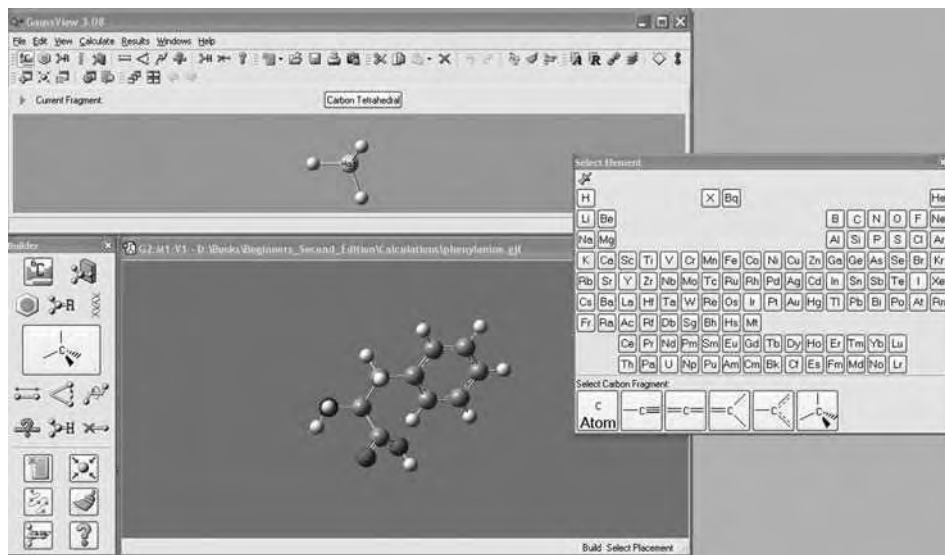


Figure 17.1 A GaussView session

GaussView sits on top of Gaussian 03; use it or not as you wish. It certainly adds to our quality of life.

17.2 Visualization

Visualization follows the path discussed in Chapter 15, but naturally there are very many more HF-LCAOs than in the case of dihydrogen. In any case, the HF-LCAOs are not uniquely determined by the HF procedure. This is because the electronic wavefunction has to be written as a Slater determinant; one of the properties of determinants is that we can add together multiples of rows and/or columns without changing their value. This is equivalent to mixing the HF-LCAOs, and at one time it was fashionable to use this idea in order to produce *localized* orbitals. Organic chemists place great emphasis on the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). I have used GaussViewW to produce the HOMO and the LUMO orbital plots for L-phenylalanine in Figures 17.2 and 17.3.

The HOMO comprises large contributions from the benzene ring. By contrast, the LUMO has a nodal plane through the benzene ring as shown, and a much larger contribution from the remainder of the molecule.

Plotting individual orbitals is all good fun, but there can be very many of them.

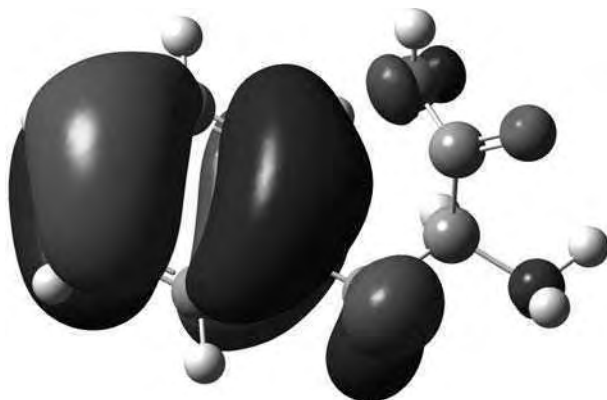


Figure 17.2 The HOMO of L-phenylalanine

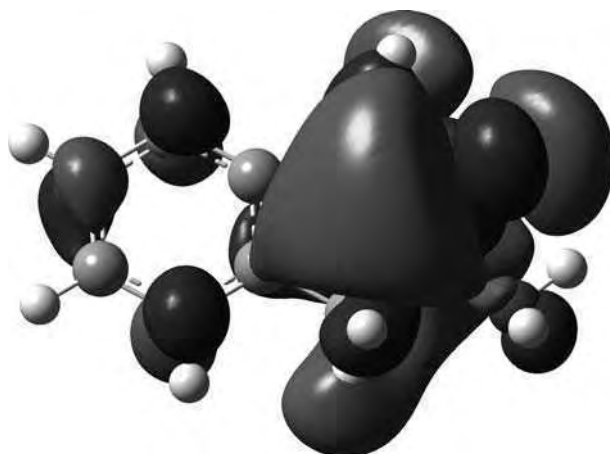


Figure 17.3 The LUMO of L-phenylalanine

17.3 Properties

Once the wavefunction has been determined, then various molecular properties can be calculated. My final section of standard output shows a selection of molecular electric moments. The so-called ‘quadrupole moment’ diagonal elements do not sum to zero which gives the game away; they are actually second moments.

Such properties are often called *primary* properties because their electronic part can be obtained directly from the wavefunction Ψ_{el}

$$\langle \hat{X} \rangle = \int \Psi_{\text{el}}^* \left(\sum_{i=1}^n \hat{X}_i \right) \Psi_{\text{el}} \, d\tau \quad (17.1)$$

where the operators refer to each of the n electrons. Usually it is necessary to add a corresponding nuclear contribution, since we work within the Born–Oppenheimer approximation. For example, the electric dipole moment operator is

$$\hat{\mathbf{p}}_e = e \sum_{\alpha=1}^N Z_{\alpha} \mathbf{R}_{\alpha} - e \sum_{i=1}^n \mathbf{r}_i$$

where the first sum runs over the N nuclei and the second sum over the n electrons. All electrons enter the sum on an equal footing, as they should, and the expectation value can be written in terms of the charge density $P(\mathbf{r})$ as

$$\langle \hat{\mathbf{p}}_e \rangle = e \sum_{\alpha=1}^N Z_{\alpha} \mathbf{R}_{\alpha} - e \int \mathbf{r} P(\mathbf{r}) d\tau \quad (17.2)$$

Such electric moments are often reported in non-SI units; the old-fashioned unit of length is the angstrom and the debye, itself a relic from the days of electrostatic units, is the dipole moment corresponding to a pair of equal and opposite charges of magnitude 10^{-10} electrostatic units (esu = $\text{g}^{-1/2} \text{cm}^{3/2} \text{s}^{-1}$) separated by 1 Å (= 10^{-10} m). There are 2.9989×10^9 esu per coulomb, and so 1 D = 10^{-10} esu $\times 10^{-10}$ m or 3.336×10^{-30} C m. The atomic unit of electric dipole is $e a_0 = 8.4784 \times 10^{-30}$ C m, which is 2.5418 D.

17.3.1 Electrostatic Potential

One of the fundamental objectives of chemistry is to predict and rationalize chemical reactivity. In principle, this involves searching a potential energy surface for saddle points (i.e. transition states) and minima (reactants and products), and this kind of detailed investigation has only become possible in the last decade. Most of the traditional theories of chemical reactivity have concerned themselves with organic molecules, and the simplest theories have attempted to extract useful information from the electronic properties of reactants. We can distinguish *static* theories, which in essence make use of the electronic wavefunction and/or electronic properties appropriate to an isolated molecule in the gas phase, and *dynamic* theories. Dynamic theories aim (for example) to predict the likely reaction sites for the approach of a charged reagent, usually modelled as a point charge.

The electrostatic potential gives an index that has been widely used since the 1970s for just this purpose (e.g. Scrocco and Tomasi 1978).

Figure 17.4 shows phenylalanine; within the Born–Oppenheimer approximation, molecules are thought of as point positive charges (the nuclei) surrounded by continuous distributions of electron charge. A small portion of the molecular electron charge resides in the volume element $d\tau$. I can now calculate the electrostatic potential at points in space around the molecule using the methods of classical electromagnetism. I have placed a point charge Q at the origin; the electrostatic potential ϕ at this point will contain contributions from the nuclei such as N:

$$\frac{1}{4\pi\epsilon_0} \frac{eZ_N}{R_N}$$

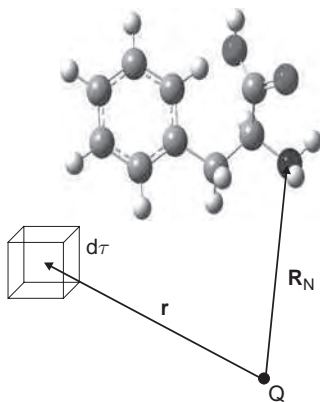


Figure 17.4 Electrostatic potential due to phenylalanine

where $Z_N = 7$, and a contribution from the electrons, which I can write in terms of the electron density $-eP(\mathbf{r})$ as

$$-\frac{e}{4\pi\epsilon_0} \int \frac{P(\mathbf{r})}{r} d\tau$$

It is conventional to record the *mutual* potential energy of the molecule and a unit positive charge Q , rather than the electrostatic potential but authors in the field still speak of potential energy maps without making the distinction. Again, it is usual to present the calculations as contour maps, which would make beautiful illustrations for textbooks like this one, if only it were printed in colour. The electrostatic potential maps are used to give a rough-and-ready survey of a molecule, and the spirit of the calculation is that one does not need a particularly sophisticated basis set. Figure 17.5 shows phenylalanine.



Figure 17.5 Phenylalanine electrostatic potential

Such objects are usually colour-coded. I have turned the molecule around a little from the previous plots. Interpretation is obviously a job for experts, and it is necessary to know the colour-coding.

17.4 Geometry Optimization

The next step might well be a geometry optimization. All the MM considerations of earlier chapters apply here, but there are two major differences. A typical MM energy minimization on a large biological molecule might involve many thousands of variables. Calculation of the energy at points on the molecular potential energy surface is trivial. A typical HF calculation will involve fewer variables but the energy calculation for each point on the molecular potential energy surface is now much more time consuming because a HF-LCAO calculation is involved.

I explained in Chapter 16 how the HF energy could be written in terms of the electron density and various one- and two-electron integrals over the basis functions χ . The HF-LCAO matrix is

$$\begin{aligned} H_{ij}^F &= \int \chi_i(\mathbf{r}_1) \hat{h}^{(1)}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \\ &+ \sum_{k=1}^n \sum_{l=1}^n P_{kl} \int \int \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \\ &- \frac{1}{2} \sum_{k=1}^n \sum_{l=1}^n P_{kl} \int \int \chi_i(\mathbf{r}_1) \chi_k(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_j(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \end{aligned}$$

and the energy gradient will therefore involve terms like

$$\frac{\partial P_{kl}}{\partial c}, \quad \int \frac{\partial \chi_i(\mathbf{r}_1)}{\partial c} \hat{h}(\mathbf{r}_1) \chi_j(\mathbf{r}_1) d\tau \quad \text{and} \quad \int \int \frac{\partial \chi_i(\mathbf{r}_1)}{\partial c} \chi_j(\mathbf{r}_2) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2$$

where c is a parameter to be varied. Advanced methods of geometry optimization usually require both the gradient and the Hessian, and these can either be calculated numerically or analytically. It is generally accepted that the analytical methods are to be preferred in terms of resource consumption, and over the years a great deal of effort has gone into the elucidation of analytical expressions for the energy gradient and the Hessian. All the major molecular structure packages now use analytical gradients wherever these are available; in the case of the HF-LCAO model, the analytical gradients do not involve the LCAO coefficients directly, just the electron density matrix \mathbf{P} . This speeds up the calculation by a large factor.

By default, Gaussian 03 performs geometry optimizations in redundant internal coordinates via the Bery algorithm. The choice of coordinate system for the starting structure is irrelevant and any efficiency factors concerned with (for example) molecular symmetry are of no consequence since all structures are converted to redundant internal coordinates.

Let me return to L-phenylalanine for my example. For illustration, I will run the job on a parallel machine. Because of the way my local Unix system is configured, I have specified 8 processors each with a maximum of 2 Gb memory. This gives me maximum theoretical performance; your system will be different.

```
%chk=phenopt.chk
%nproc=8
%mem=2Gb
# HF/6-311G** Opt Freq T
L-phenylalanine geometry optimization
  0  1
N   0.000  0.000  0.000
H   0.000  0.000  1.010
C   1.366  0.000 -0.483
.....
..... (same as before)
```

In the Gaussian 03 package, the default optimization method uses analytical expressions for the gradient and the Hessian. Here are some relevant parts of the output file. First of all, an estimate of the Hessian has to be made.

```
GradGradGradGradGradGradGradGradGradGradGradGradGradGradGradGrad
Berny optimization.
Initialization pass.
```

```

-----
!      Initial Parameters      !
! (Angstroms and Degrees)    !
-----
! Name      Definition      Value      Derivative Info.  !
-----
! R1        R(1,2)          1.01       estimate D2E/DX2  !
! R2        R(1,3)          1.4489     estimate D2E/DX2  !
! R3        R(1,4)          1.01       estimate D2E/DX2  !
! R4        R(3,5)          1.0902     estimate D2E/DX2  !
! R5        R(3,6)          1.5219     estimate D2E/DX2  !
! R6        R(3,7)          1.5259     estimate D2E/DX2  !
! R7        R(6,8)          1.2301     estimate D2E/DX2  !
. . .
. . .
```

The first estimate is made numerically. Next the geometry iterations begin. The calculation proceeds (hopefully) downwards on the molecular potential energy surface:

Variable	Old X	-DE/DX (Linear)	Delta X (Quad)	Delta X (Total)	Delta X	New X
R1	1.90188	-0.00653	-0.00956	0.00000	-0.00956	1.89232
R2	2.73522	-0.00228	-0.00391	0.00000	-0.00391	2.73131
R3	1.90126	-0.00697	-0.01045	0.00000	-0.01045	1.89081
R4	2.05499	-0.00382	-0.00747	0.00000	-0.00747	2.04752
R5	2.89023	0.00948	0.02027	0.00000	0.01759	2.90782
R6	2.90544	0.00539	0.03101	0.00000	0.02505	2.93049
R7	2.29763	-0.05460	-0.03832	0.00000	-0.03832	2.25931

until eventually a stationary point is reached (that is, a point on the molecular potential energy surface where the gradient is zero). This could be a minimum or a saddle point:

```

-----
! Optimized Parameters !
! (Angstroms and Degrees) !
-----
! Name          Definition          Value          Derivative Info. !
-----
! R1            R(1,2)              1.0012        -DE/DX = 0.0    !
! R2            R(1,3)              1.4507        -DE/DX = 0.0    !
! R3            R(1,4)              1.0003        -DE/DX = 0.0    !
! R4            R(3,5)              1.0824        -DE/DX = 0.0    !
! R5            R(3,6)              1.5232        -DE/DX = 0.0    !
! R6            R(3,7)              1.5386        -DE/DX = 0.0    !
! R7            R(6,8)              1.185         -DE/DX = 0.0    !
. . .
. . .

```

and so on.

17.5 Vibrational Analysis

It is wise to calculate the force constants at this point on the surface, in order to characterize the stationary point. Just to remind you, harmonic force constants correspond to the eigenvalues of the Hessian, calculated at the stationary point. A minimum on the molecular potential energy curve should have $3N - 6$ positive eigenvalues ($3N - 5$ for a linear molecule), a transition state of the chemical kind will have just one negative eigenvalue of the Hessian and so on. In this example there are 23 nuclei and so 69 vibrational coordinates in total. This gives 63 vibrational coordinates but because I have chosen to do the calculation in redundant internal coordinates rather than starting from a Z-matrix, there should be six redundant coordinates corresponding to the three translational coordinates of the centre of mass and three for the rotations. Each of these should have a force constant of zero, but there is a question as to how small a number needs to be before it is taken to be nonzero. This has to do with the optimization cut-off point, the convergence criterion for the HF-LCAO calculation and so on.

Here is what I found.

Full mass-weighted force constant matrix:

```
Low frequencies - - - -0.8888 -0.5279 -0.3445 0.0005 0.0006 0.0012
Low frequencies - - - 34.4264 43.3413 60.3827
```

The first six are essentially zero and are therefore taken to represent the redundant coordinates. The next piece of output gives the normal modes, as discussed in Chapters 3 and 4. Here is a snapshot.

		61			62			63		
...(Symmetry)...		A			A			A		
Frequencies	--	3722.9962			3800.9366			4108.2863		
Red. masses	--	1.0501			1.0942			1.0651		
Frc consts	--	8.5758			9.3135			10.5919		
IR Inten	--	2.7774			5.3387			106.2867		
Raman Activ	--	110.9373			67.9404			72.1007		
Depolar (P)	--	0.1335			0.6701			0.2986		
Depolar (U)	--	0.2356			0.8025			0.4599		
Atom AN		X	Y	Z	X	Y	Z	X	Y	Z
1	7	-0.05	0.02	-0.01	0.02	0.02	-0.08	0.00	0.00	0.00
2	1	0.23	-0.29	0.62	0.24	-0.30	0.58	0.00	0.00	0.00
3	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	1	0.48	-0.05	-0.50	-0.51	0.08	0.50	0.00	0.00	0.00
5	1	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
6	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.06	0.00
13	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	1	0.00	0.00	0.00	0.00	0.00	0.00	-0.07	1.00	-0.01

Output includes the vibration frequencies, reduced masses, force constants, infrared intensities, Raman activities and depolarization ratios, together with the normal modes expressed as linear combinations of the Cartesian coordinates of the atoms. Normal mode 63 for example, which is of A symmetry, comprises almost entirely the y component for atom 15, which is a hydrogen atom.

Most packages will simulate the infrared spectrum (Figure 17.6) and allow for on-screen animation of the normal modes of vibration. I must stress that geometries have to be optimized before calculating vibrational frequencies, since these are defined as the second derivative *calculated at the stationary point*.

I want to illustrate one final point by considering a much smaller molecule, carbon dioxide. It is a linear molecule and so has four vibrational degrees of freedom. It is also a very topical greenhouse gas.

Table 17.1 shows the calculated frequencies. The agreement with experiment is not particularly good. The intensities (not shown) are qualitatively correct; the symmetric stretch

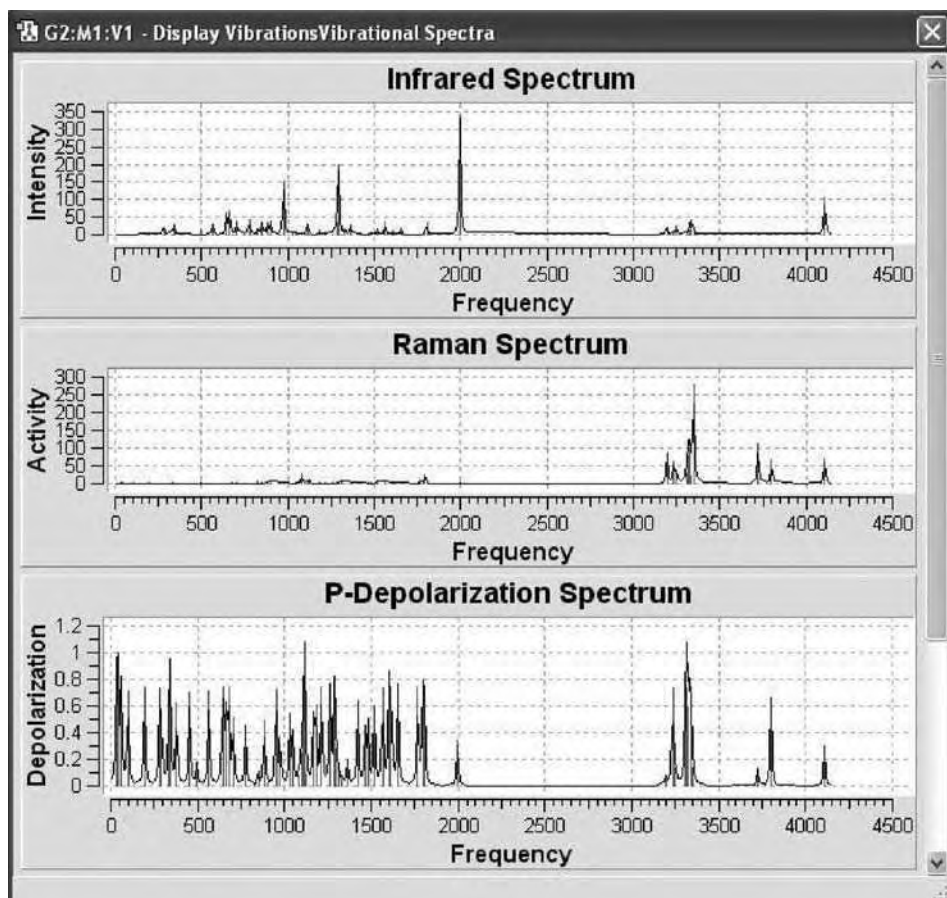


Figure 17.6 Vibrational spectra

Table 17.1 Calculated (HF-LCAO) and experimental vibration wavenumbers (in cm^{-1})

Mode	Expt	STO-3G	STO/6-31G*	STO/6-311G**
Symmetric stretch (σ_g)	1388.17	1435.2	1518.7	1522.8
Asymmetric stretch (σ_u)	2349.16	2535.7	2585.4	2593.5
Degenerate bend (π)	667.40	566.2	745.7	766.8

is not infrared active because the dipole moment does not change during the vibration, and the doubly degenerate bending vibration has a lower intensity than the asymmetric stretch.

One decision that has to be made with a HF-LCAO calculation is the choice of basis set, and this is usually a matter of experience. I ran the calculation with three fairly simple basis sets, and the results are also shown in Table 17.1. Apart from the minimal basis set calculation, it is a common experience that HF-LCAO calculations overestimate

force constants by about 10%. Provided the same basis set is used from molecule to molecule, it is possible to scale the calculation by a fixed factor to gain very reliable vibrational data.

17.5.1 Hellmann Feynman Theorem

Before I finish this subsection, I want to tell you about a very early (but unsuccessful) attempt to take a shortcut in energy minimizations.

Suppose that an electronic wavefunction Ψ depends on a single parameter c such as a single bond length or an orbital exponent. According to the variation principle, the best value of c is the one that minimizes the energy

$$\varepsilon(c) = \int \Psi^*(c) \hat{H} \Psi(c) d\tau \quad (17.3)$$

subject to the requirement that the wavefunction is normalized

$$\int \Psi^*(c) \Psi(c) d\tau = 1 \quad (17.4)$$

If we differentiate these two equations with respect to c we have

$$\frac{d\varepsilon(c)}{dc} = \int \frac{\partial \Psi^*(c)}{\partial c} \hat{H} \Psi(c) d\tau + \int \Psi^*(c) \frac{\partial \hat{H}}{\partial c} \Psi(c) d\tau + \int \Psi^*(c) \hat{H} \frac{\partial \Psi(c)}{\partial c} d\tau \quad (17.5)$$

$$\int \frac{\partial \Psi^*(c)}{\partial c} \Psi(c) d\tau + \int \Psi^*(c) \frac{\partial \Psi(c)}{\partial c} d\tau = 0$$

Normally $\Psi(c)$ will be an approximate wavefunction but suppose that $\Psi(c)$ happens to be an eigenfunction of the Hamiltonian. A little analysis shows that

$$\frac{d\varepsilon(c)}{dc} = \int \Psi^*(c) \frac{\partial \hat{H}}{\partial c} \Psi(c) d\tau \quad (17.6)$$

This result is known as the Hellmann Feynman theorem, and it was hoped in the early days that it could be used to optimize geometries. The first applications were to isoelectronic processes



such as the internal rotation in ethane, where it was believed that the only nonzero contributions to the right-hand side of Equation (17.6) would be due to changes in the nuclear positions and hence in the one-electron integrals. It was hoped that such calculations would be useful for geometry optimizations, where the changes also involve two-electron integrals but enthusiasm vanished once it became clear that approximate wavefunctions also depend on the parameter(s) contained either explicitly or implicitly in the wavefunction and so wavefunction gradient terms such as $\partial \Psi / \partial c$ cannot be neglected.

17.6 Thermodynamic Properties

In Chapter 8 I discussed the canonical ensemble and the canonical partition function

$$Q = \sum_i \exp\left(-\frac{E_i^*}{k_B T}\right) \quad (17.7)$$

I added a * superscript to focus attention on the fact that E^* refers to the collection of particles in each cell of the ensemble. There will be many possible allowed values for E^* , and the N particles in each cell will contribute in some way to make up the total. So far, I have made no assumptions whatever about these particles, neither have I made any assumption about the way in which they might interact with each other.

Suppose now that the N particles are identical molecules, but that they form essentially an ideal gas. One characteristic of an ideal gas is that there is no interaction between the particles, so the total energy E^* of the N particles will be a simple sum of the particle energies. If I label the molecules 1, 2, 3, ..., N then we can write

$$E^* = \varepsilon^{(1)} + \varepsilon^{(2)} + \dots + \varepsilon^{(N)} \quad (17.8)$$

Each molecular energy will contain a kinetic and a potential part, but there are no intermolecular interactions because of the ideal gas behaviour. So for each possible value of E^* we have

$$E_i^* = \varepsilon_i^{(1)} + \varepsilon_i^{(2)} + \dots + \varepsilon_i^{(N)} \quad (17.9)$$

Each allowed value of E^* will correspond to different values of the constituent ε 's and a simple rearrangement of Q shows that

$$Q = \sum_i \exp\left(-\frac{\varepsilon_i^{(1)}}{k_B T}\right) \sum_j \exp\left(-\frac{\varepsilon_j^{(2)}}{k_B T}\right) \dots \sum_n \exp\left(-\frac{\varepsilon_n^{(N)}}{k_B T}\right) \quad (17.10)$$

I have used a different dummy index i, j, \dots, n but the sums are all the same since they refer to identical particles. At first sight we should therefore write

$$Q = \left(\sum_i \exp\left(-\frac{\varepsilon_i^{(1)}}{k_B T}\right) \right)^N$$

This assumes that the N particles can somehow be distinguished one from another (we say that such particles are *distinguishable*). The laws of quantum mechanics tell us that seemingly identical particles are truly identical, they are *indistinguishable* and to allow for this we have to divide Q by the number of ways in which N things can be permuted:

$$Q = \frac{1}{N!} \left(\sum_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right) \right)^N \quad (17.11)$$

I have dropped all reference to particle '1', since all N are identical.

The summation

$$q = \sum_i \exp\left(-\frac{\varepsilon_i}{k_B T}\right) \quad (17.12)$$

refers only to the energy states of a representative particle, and it is a quantity of great interest in our theories. It is called the *molecular partition function* and given the symbol q (sometimes z).

17.6.1 Ideal Monatomic Gas

Consider now an even more restricted case, a monatomic ideal gas. Each atom has mass m and the gas is constrained to a cubical container whose sides are a , b and c (so that the volume V is abc). We will ignore the fact that each atom could well have different electronic energy, and concentrate on the translational energy. Elementary quantum mechanics texts (and Chapter 11) show that the available (translational) energy levels are characterized by three quantum numbers n_x , n_y and n_z which can each take integral values ranging from 1 to infinity. The translational energy is given by

$$\varepsilon_{n_x, n_y, n_z} = \frac{h^2}{8m} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right)$$

and so each particle will make a contribution

$$\begin{aligned} \sum_{n_x} \sum_{n_y} \sum_{n_z} \exp \left(-\frac{\varepsilon_{n_x, n_y, n_z}}{k_B T} \right) &= \sum_{n_x} \sum_{n_y} \sum_{n_z} \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \left(\frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right) \right) \\ &= \sum_{n_x} \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \frac{n_x^2}{a^2} \right) \sum_{n_y} \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \frac{n_y^2}{b^2} \right) \sum_{n_z} \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \frac{n_z^2}{c^2} \right) \end{aligned} \quad (17.13)$$

to the translational molecular partition function. Each of the three summations can be treated as follows. A simple calculation shows that, for an ordinary atom such as argon at room temperature constrained to such a macroscopic container, typical quantum numbers are of the order of 10^9 by which time the allowed energy states essentially form a continuum. Under these circumstances, we replace the summation by an integral and treat n_x as a continuous variable. So for example

$$\begin{aligned} \sum_{n_x} \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \frac{n_x^2}{a^2} \right) &\approx \int_0^\infty \exp \left(-\frac{1}{k_B T} \frac{h^2}{8m} \frac{n_x^2}{a^2} \right) dn_x \\ &= \sqrt{2\pi m k_B T} \frac{a}{h} \end{aligned}$$

The translational partition functions are then given by

$$\begin{aligned} q_{\text{trans}} &= \left(\frac{2\pi m k_B T}{h^2} \right)^{3/2} V \\ Q_{\text{trans}} &= \frac{1}{N!} \left(\left(\frac{2\pi m k_B T}{h^2} \right)^{3/2} V \right)^N \end{aligned} \quad (17.14)$$

The thermodynamic functions can be calculated using the equations given in Chapter 8. We have for example

$$U = k_{\text{B}}T^2 \left(\frac{\partial \ln Q}{\partial T} \right)_V$$

and so

$$U = \frac{3}{2}Nk_{\text{B}}T$$

For a monatomic ideal gas, the translational kinetic energy is the only energy that an atom will possess. The internal energy is therefore just what we would expect from the equipartition of energy principle. In a similar way we find that the entropy is

$$S = Nk_{\text{B}} \left(\frac{5}{2} + \ln \left(\left(\frac{2\pi mk_{\text{B}}T}{h^2} \right)^{3/2} \frac{V}{N} \right) \right) \quad (17.15)$$

and this is called the *Sackur–Tetrode equation*.

17.6.2 Ideal Diatomic Gas

Monatomic gas particles can only store translational kinetic energy, whilst, for a polyatomic molecule, we have to consider:

- rotational energy;
- vibrational energy.

We also need to consider electronic energies. In order to progress, we make the assumption that the different energy modes are independent, and that each set has its own Boltzmann distribution. We therefore write for a typical molecule

$$\varepsilon = \varepsilon_{\text{trans}} + \varepsilon_{\text{rot}} + \varepsilon_{\text{vib}} + \varepsilon_{\text{elec}} \quad (17.16)$$

and it is easily seen that the molecular partition function is the product of a translational, rotational, vibrational and electronic molecular partition function. These are related to the canonical partition function by

$$Q = \frac{1}{N!} (q_{\text{trans}}q_{\text{rot}}q_{\text{vib}}q_{\text{elec}})^N \quad (17.17)$$

17.6.3 q_{rot}

The rotational energy levels for a rigid diatomic rotator are given by

$$\begin{aligned} \varepsilon_J &= J(J+1) \frac{h^2}{8\pi^2 I} \\ &= Bhc_0 J(J+1) \end{aligned} \quad (17.18)$$

where the rotational quantum number J takes integral values $0, 1, 2, \dots$, I is the moment of inertia about an axis through the centre of mass of the molecule and c_0 is the speed of light

in *vacuo*. B is called the rotation constant. Each energy level is $(2J + 1)$ -fold degenerate, and so all $2J + 1$ individual quantum states must be counted in the Boltzmann formula:

$$q_{\text{rot}} = \sum_{J=0}^{\infty} (2J + 1) \exp\left(-\frac{Bhc_0J(J+1)}{k_{\text{B}}T}\right)$$

The sum cannot be found in a closed form, but we note that for molecules with small rotational constants, the rotational energy levels crowd together and the summation can be replaced by an integral, treating J as a continuous variable:

$$\begin{aligned} q_{\text{rot}} &= \int_0^{\infty} (2J + 1) \exp\left(-\frac{Bhc_0J(J+1)}{k_{\text{B}}T}\right) dJ \\ &= \frac{k_{\text{B}}T}{hc_0B} \end{aligned} \quad (17.19)$$

Equation (17.19) works with complete reliability for all heteronuclear diatomic molecules, subject to the accuracy of the rotational energy level formula and the applicability of the continuum approximation. For homonuclear diatomics we have to think more carefully about indistinguishability of the two nuclei which results in occupancy of either the odd J or the even J levels; this depends on what the nuclei are, and need not concern us in detail here. It is dealt with by introducing a *symmetry factor* σ that is 1 for a heteronuclear diatomic and 2 for a homonuclear diatomic. Parameter q_{rot} is written

$$q_{\text{rot}} = \frac{1}{\sigma} \frac{k_{\text{B}}T}{hc_0B} \quad (17.20)$$

The so-called *rotational temperature* θ_{rot} is often used in discussions of statistical mechanics; we rewrite Equation (17.20) as

$$q_{\text{rot}} = \frac{1}{\sigma} \frac{T}{\theta_{\text{rot}}} \quad (17.21)$$

17.6.4 q_{vib}

Vibrational energy levels have separations that are at least an order of magnitude greater than the rotational modes, which are in turn some 20 orders of magnitude greater than the translational modes. As a consequence, the spacing is comparable to $k_{\text{B}}T$ for everyday molecules and temperatures. If we consider a single harmonic vibrational mode for which

$$\begin{aligned} \varepsilon_v &= hc_0\omega_e \left(v + \frac{1}{2}\right) \\ \omega_e &= \frac{1}{2\pi c_0} \sqrt{\frac{k_s}{\mu}} \end{aligned}$$

and if we take the energy zero as that of the energy level $v = 0$ we have

$$q_{\text{vib}} = \sum_{v=0}^{\infty} \exp\left(-\frac{hc_0v\omega_e}{k_{\text{B}}T}\right)$$

This simple geometric series can be summed exactly to give

$$q_{\text{vib}} = \frac{1}{1 - \exp\left(-\frac{hc_0\omega_e}{k_B T}\right)} \quad (17.22)$$

and once again it is usually written in terms of the vibrational temperature θ_{vib} :

$$q_{\text{vib}} = \frac{1}{1 - \exp\left(-\frac{\theta_{\text{vib}}}{T}\right)} \quad (17.23)$$

Finally we need to know the electronic partition function. This is almost always 1, unless the electronic ground state is degenerate (in which case it is equal to the degeneracy) or unless there is a very low-lying excited state.

17.7 Back to L-Phenylalanine

Let me now return to L-phenylalanine. The Gaussian 03 ‘frequency’ calculation also gives thermodynamic properties, calculated along the lines discussed above. Here is an abbreviated output. First we have moments of inertia and then a calculation of the rotational and vibrational temperatures.

Principal axes and moments of inertia in atomic units:

	1	2	3
EIGENVALUES --	1026.19740	3017.84285	3333.43073
X	0.99982	0.01842	0.00491
Y	-0.01827	0.99942	-0.02889
Z	-0.00544	0.02879	0.99957

This molecule is an asymmetric top.

Rotational symmetry number 1.

Warning -- assumption of classical behavior for rotation
may cause significant error

Rotational temperatures (Kelvin) 0.08440 0.02870 0.02598

Rotational constants (GHZ): 1.75867 0.59802 0.54141

Zero-point vibrational energy 532871.0 (Joules/Mol)

127.35923 (Kcal/Mol)

Vibrational temperatures:	49.53	62.36	86.88	142.01	280.87
(Kelvin)	405.40	422.99	483.72	489.81	545.09
	653.24	713.44	813.05	927.86	952.80

Next come the various contributions to the internal energy U (here called E), the heat capacity C and the entropy S .

Zero-point correction=	0.202960 (Hartree/Particle)
Thermal correction to Energy=	0.213506
Thermal correction to Enthalpy=	0.214451

Thermal correction to Gibbs Free Energy=	0.164839		
Sum of electronic and zero-point Energies=	-551.349677		
Sum of electronic and thermal Energies=	-551.339131		
Sum of electronic and thermal Enthalpies=	-551.338187		
Sum of electronic and thermal Free Energies=	-551.387798		
	E (Thermal)	CV	S
	KCal/Mol	Cal/Mol-Kelvin	Cal/Mol-Kelvin
Total	133.977	38.971	104.416
Electronic	0.000	0.000	0.000
Translational	0.889	2.981	41.212
Rotational	0.889	2.981	30.713
Vibrational	132.200	33.010	32.490
Vibration 1	0.594	1.983	5.556
Vibration 2	0.595	1.980	5.100
Vibration 3	0.597	1.973	4.445
...			
... etc			

The partition functions are also printed.

17.8 Excited States

Electronic ground states are all well and good, but there is a wealth of experimental spectroscopic data that relates to the excited states of molecules. One of the simplest and most widely studied sequences of organic molecules to be investigated by spectrophotometry is afforded by the set of monocyclic azines, of which pyridine forms the first member. It is instructive to compare this series with benzene.

All the molecules have 42 electrons, 21 electron pairs, and I can write their ground state electronic configurations

$$\Psi_0 : (\psi_1)^2 (\psi_2)^2 \dots (\psi_{21})^2$$

We refer to Ψ_0 as the *reference state*. Single excitations can be produced by promoting an electron from one of the occupied orbitals to an unoccupied (virtual) one, just as we did in Chapter 15 for dihydrogen, so the first few single excitations have configurations

$$\begin{aligned} \Psi_1 &: \psi_1^2 \psi_2^2 \dots \psi_{21}^1 \psi_{22}^1 \\ \Psi_2 &: \psi_1^2 \psi_2^2 \dots \psi_{21}^1 \psi_{23}^1 \\ \Psi_3 &: \psi_1^2 \psi_2^2 \dots \psi_{20}^1 \psi_{21}^2 \psi_{22}^1 \\ \Psi_4 &: \psi_1^2 \psi_2^2 \dots \psi_{20}^1 \psi_{21}^2 \psi_{23}^1 \end{aligned}$$

Electron spin has to be taken into account, and proper Slater determinants constructed. The ground-state wavefunction will be a single Slater determinant but there are four possibilities (say, (a), (b), (c) and (d)) for each of the four excited states. In our dihydrogen calculation, I grouped the determinants together so that they were spin eigenfunctions. In modern configuration interaction (CI) calculations, we work directly with the

individual Slater determinants and classify the resulting wavefunctions at the end of the calculation.

We write the CI wavefunction as

$$\Psi_{\text{CI}} = c_0\Psi_0 + c_{1a}\Psi_{1a} + c_{1b}\Psi_{1b} + \dots$$

where the CI expansion coefficients have to be determined from a variational calculation. This involves finding the eigenvalues and eigenvectors of a matrix whose elements are typically

$$\int \Psi_i \Psi_j \, d\tau$$

The matrix elements can be determined from the Slater–Condon–Shortley rules, giving typically HF-LCAO orbital energies and various two-electron integrals such as

$$\frac{e^2}{4\pi\epsilon_0} \int \psi_{21}(\mathbf{r}_1)\psi_{22}(\mathbf{r}_2) \frac{1}{r_{12}} \psi_{21}(\mathbf{r}_1)\psi_{22}(\mathbf{r}_2) \, d\tau_1 d\tau_2$$

These have to be calculated from the two-electron integrals over the HF-LCAO basis functions, at first sight a four-dimensional sum known in the trade as the *four-index transformation*.

Two of each of the (a), (b), (c) and (d) Slater determinants correspond to spin eigenfunctions having spin quantum number $M_s = 0$, one to $M_s = +1$ and one to $M_s = -1$ and so need not be considered because states of different spin symmetry do not mix. In addition, Brillouin's theorem (Brillouin 1933) tells us that singly excited states constructed using HF wavefunctions do not mix with the ground state for a closed shell system, so we do not need to include the ground state Ψ_0 in the variational calculation.

If we take all possible excited Slater determinants and solve the variational problem we reach a level of theory known as *CI singles (CIS)*. A keynote paper is that by Foresman *et al.* (1992), and their synopsis puts everything into perspective.

This work reviews the methodological and computational considerations necessary for the determination of the ab initio energy, wavefunction and gradient of a molecule in an electronically excited state using molecular orbital theory. In particular, this paper re-examines a fundamental level of theory which was employed several years ago for the interpretation of the electronic spectra of simple organic molecules: configuration interaction (CI) among all singly substituted determinants using a Hartree Fock reference state. This investigation presents several new enhancements to this general theory. First, it is shown how the CI singles wavefunction can be used to compute efficiently the analytic first derivative of the energy . . . Second, a computer program is described which allows these computations to be done in a 'direct' fashion.

You should have picked up the words 'direct' and 'gradient'.

To return to my example of the azines; in the earliest and necessarily qualitative treatments, the first four transitions were identified with the single excitations shown in Figure 17.7. Benzene can be regarded as a special case of the series and the highest occupied orbitals are doubly degenerate, as are the lowest unoccupied ones. This leads to many degeneracies in the excited states that simply do not appear (other than by accident) in the substituted azines. The assignment of bands in the benzene spectrum led to considerable discussion in the primary literature.

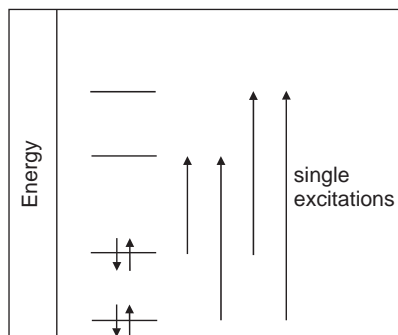


Figure 17.7 Simple CI treatment of azines

Excitation energies and oscillator strengths:

Excited State 1: Singlet-B1 6.2005 eV 199.96 nm f=0.0072

19 ->25	0.60076
19 ->29	0.16645
19 ->36	-0.24330
19 ->51	-0.17296

Excited State 2: Singlet-B2 6.2623 eV 197.99 nm f=0.0882

20 ->28	0.32051
20 ->39	-0.11996
21 ->25	0.57094
21 ->29	0.14308
21 ->36	-0.14253

Excited State 3: Singlet-A1 6.5730 eV 188.63 nm f=0.0191

20 ->25	-0.39008
20 ->29	-0.12134
20 ->36	0.10609
21 ->28	0.52009
21 ->39	-0.15078

Excited State 4: Singlet-A2 7.1312 eV 173.86 nm f=0.0000

21 ->22	0.65005
21 ->24	0.12283
21 ->30	-0.14334
21 ->44	-0.12506

Excited State 5: Singlet-A2 7.3500 eV 168.69 nm f=0.0000

19 ->28	0.65180
19 ->39	-0.24025

Excited State 6: Singlet-B1 7.6497 eV 162.08 nm f=0.0671

20 ->22	-0.29313
21 ->23	0.61432
21 ->45	0.11735

```
Excited State 7: Singlet-A2 8.1194 eV 152.70 nm f=0.0000
 21 ->22      -0.14317
 21 ->24       0.64628
 21 ->30      -0.13451
```

```
Excited State 8: Singlet-B1 8.1247 eV 152.60 nm f=0.0011
 20 ->22       0.57081
 20 ->30      -0.11602
 21 ->23       0.28223
 21 ->26      -0.19020
```

The text above shows sample output from a CIS singlet state calculation on pyridine, and I have only included the first eight excited states. To allow for the possibility that excited states are diffuse in nature, I added the '+' diffuse functions to my 6-311** basis set.

The MOs ψ_{20} and ψ_{21} are both π type, whilst ψ_{22} and ψ_{23} are antibonding π^* . The 19th occupied orbital corresponds to the N lone pair. The lowest singlet excited state is $n \rightarrow \pi^*$ as is the fourth. The remainder in my small sample are all $\pi \rightarrow \pi^*$. The notation (for example) 20- > 220.57081 gives the contribution made to that particular excited state from the singly excited state $\psi_{20} \rightarrow \psi_{22}$.

I have collected a small number of calculated data in Table 17.2. In each case I have recorded the first $n \rightarrow \pi^*$ and first two allowed $\pi \rightarrow \pi^*$ transitions, together with the oscillator strengths.

Table 17.2 Spectroscopic data for various cyclic molecules: λ (in nm) and oscillator strength f (d means doubly degenerate)

	Benzene	Pyridine	1,2-Diazine	1,3-Diazine	1,4-Diazine
$n \rightarrow \pi^*$	—	200.0	249.8	207.9	238.4
		0.0072	0.0126	0.0131	0.0110
$\pi \rightarrow \pi^*$	167.9	198.0	186.5	183.7	197.4
	0.0965	0.0882	0.0425	0.0747	0.1901
$\pi \rightarrow \pi^*$	156.7	188.6	145.3	173.5	180.2
	0.9169	0.0191	0.0073	0.0389	0.0777

A satisfactory theory of the spectra of aromatic hydrocarbons was not developed until there had been a detailed classification of their absorption bands. This important work was first undertaken by Clar (1952). There are three types of absorption bands that are classified mainly by their intensity but also by their vibrational fine structure and the frequency shifts within a family of compounds. The ground state of benzene (with D_{6h} symmetry) has closed shell configuration $a_{1g}^2 e_{1g}^4$ which gives symmetry A_{1g} . The first excited configuration $a_{1u}^2 e_{1g}^3 e_{2u}^1$ gives rise to states of symmetry B_{1u} , B_{2u} and E_{1u} which can be either singlet or triplet. The experimental absorption spectrum shows a weak band near 260 nm, and a strong band at 185.0 nm. The CIS calculation gives a weak band at 167.9 nm and a strong band at 156.7 nm and so our calculation accounts only qualitatively for the features of the benzene spectrum.

The lowest singlet excitation for the azines is $n \rightarrow \pi^*$ in character, and Mason's (1962) experimental values are 270, 340, 298 and 328 nm. The CIS calculation gives the trend correctly but the absolute agreement with experiment is poor. Similar conclusions apply to the $\pi \rightarrow \pi^*$ transitions. The general conclusion is that the CIS method gives a useful first approximation to the excited states.

17.9 Consequences of the Brillouin Theorem

I mentioned primary properties in Section 17.3; these are ones such as the electric dipole moment, which is given as the expectation value of a sum of one-electron operators (together with the corresponding nuclear contribution):

$$\langle X \rangle = \int \Psi_{\text{el}}^* \left(\sum_{i=1}^n \hat{X}_i \right) \Psi_{\text{el}} d\tau$$

Suppose that we have a singlet ground-state closed shell HF wavefunction Ψ_0 and we try to improve it by adding singly excited states; if I denote occupied orbitals A, B, . . . and virtual ones X, Y, . . . then

$$\Psi = c_0 \Psi_0 + \sum_{A,X} c_A^X \Psi_A^X \quad (17.24)$$

where c_A^X is the CI expansion coefficient and Ψ_A^X the appropriate combination of Slater determinants describing a singlet singly excited state. We find

$$\langle X \rangle = c_0^2 \int \Psi_0 \sum_{i=1}^n \hat{X}_i \Psi_0 d\tau + 2 \sum_{A,X} c_A^X c_0 \int \Psi_0 \sum_{i=1}^n \hat{X}_i \Psi_A^X d\tau + \sum_{A,X} \sum_{B,Y} c_A^X c_B^Y \int \Psi_B^Y \sum_{i=1}^n \hat{X}_i \Psi_A^X d\tau \quad (17.25)$$

and all terms on the right-hand side apart from the first are zero, since the c 's are zero by Brillouin's theorem. In addition the only nonzero integrals in the third sum are those where the excited states differ by no more than one spin orbital. This can be easily proved by application of the Slater–Condon–Shortley rules.

If we now repeat the argument but include doubly excited rather than singly excited states

$$\begin{aligned} \Psi &= c_0 \Psi_0 + \sum_{A,B,X,Y} c_{AB}^{XY} \Psi_{AB}^{XY} \\ \langle X \rangle &= c_0^2 \int \Psi_0 \sum_{i=1}^n \hat{X}_i \Psi_0 d\tau + 2 \sum_{A,B,X,Y} c_{AB}^{XY} c_0 \int \Psi_0 \sum_{i=1}^n \hat{X}_i \Psi_{AB}^{XY} d\tau \\ &\quad + \sum_{A,B,X,Y} \sum_{C,D,U,V} c_{AB}^{XY} c_{CD}^{UV} \int \Psi_{CD}^{UV} \sum_{i=1}^n \hat{X}_i \Psi_{AB}^{XY} d\tau \end{aligned} \quad (17.26)$$

then the CI expansion coefficients are no longer zero. However, the integrals in the second sum are all zero because the doubly excited states differ from Ψ_0 by two spin orbitals. All the integrals in the third sum except those where $\Psi_{AB}^{XY} = \Psi_{CD}^{UV}$ are zero for the same reason, and the final expression is

$$\langle X \rangle = c_0^2 \int \Psi_0 \sum_{i=1}^n \hat{X}_i \Psi_0 d\tau + \sum_{A,B,X,Y} \int \Psi_{AB}^{XY} \sum_{i=1}^n \hat{X}_i \Psi_{AB}^{XY} d\tau \quad (17.27)$$

At first sight, one-electron properties calculated from HF wavefunctions ought to be very accurate because the lowest ‘correction’ terms are certain doubly excited configurations.

The argument is correct as far as it goes, but if we include *both* the singly and doubly excited states together

$$\Psi = c_0 \Psi_0 + \sum_{A,X} c_A^X \Psi_A^X + \sum_{A,B,X,Y} c_{AB}^{XY} \Psi_{AB}^{XY} \quad (17.28)$$

the singly excited states couple with the doubly excited ones, which couple with the ground state and so the singly excited states do make a contribution to the ground state.

In any case, the key choice when performing HF-LCAO calculations is the choice of atomic orbital basis set, as can be seen from my representative HF-LCAO calculations on pyridine in Table 17.3. In each case I optimized the geometry.

Table 17.3 Electric dipole moment for pyridine; HF-LCAO model

Level of theory	$\rho_e/10^{-30}$ C m
HF/STO-3G	6.870
HF/6-31G	8.811
HF/6-31G*	7.720
HF/6-311G**	7.553
HF/cc-pvQz	10.846
Experiment	$7.31 \pm 2\%$

17.10 Electric Field Gradients

I have made mention at several points in the text of electron spin, the internal angular momentum possessed by electrons just because they are electrons. The total angular momentum of an atomic electron is found by summing the electron spin and the electron orbital angular momentum.

Many nuclei have a corresponding internal angular momentum, which is given the symbol **I** and the vector is characterized by two quantum numbers I and m_I . The vector obeys the laws of quantum mechanics appropriate to angular momenta; for each value of I , $m_I = -I, -I + 1, \dots, +I$, the size of the vector is

$$\sqrt{I(I+1)} \frac{h}{2\pi}$$

and we can only measure simultaneously the size of the vector and one component that we conventionally take as the z component. The z component has size

$$m_I \frac{h}{2\pi}$$

The nuclear spin quantum number is characteristic of a given nucleus and can have different values for different isotopic species. A sample is shown in Table 17.4.

Table 17.4 Representative nuclear spin quantum numbers

Isotopic species	I
^1H	1/2
^2D	1
^{12}C	0
^{13}C	1/2
^{17}O	5/2

Many nuclei with $I \geq 1$ also possess a nuclear quadrupole \mathbf{Q}_n . It is usually defined in terms of the nuclear charge distribution $\rho_n(\mathbf{r})$ as

$$\mathbf{Q}_n = \frac{1}{e} \begin{pmatrix} \int \rho_n (3x^2 - r^2) d\tau & 3 \int \rho_n xy d\tau & 3 \int \rho_n xz d\tau \\ 3 \int \rho_n yx d\tau & \int \rho_n (3y^2 - r^2) d\tau & 3 \int \rho_n yz d\tau \\ 3 \int \rho_n zx d\tau & 3 \int \rho_n zy d\tau & \int \rho_n (3z^2 - r^2) d\tau \end{pmatrix} \quad (17.29)$$

This definition gives a traceless tensor. Here the integration is over the nuclear charge distribution. Nuclear wavefunctions are pretty hard to come by, and we normally have to determine the components of \mathbf{Q}_n by experiment.

In a molecule, a given nucleus will generally experience an electric field gradient due to the surrounding electrons. Electric fields are vector quantities and so an electric field gradient is a tensor quantity. The electric field gradient at nucleus n is usually written \mathbf{q}_n . The energy of interaction U between the nuclear quadrupole and the electric field gradient is

$$U = -\frac{e}{6} \sum (Q_n)_{ij} (q_n)_{ij} \quad (17.30)$$

In principal axes, the interaction is determined by $Q_n = (Q_n)_{zz}$ and $q_n =$ the largest of the diagonal components of \mathbf{q}_n . The quantity $eQ_n q_n / h$ is referred to as the *quadrupole coupling constant*. According to Townes and Dailey (1949), since filled shells and s orbitals have spherical symmetry, and since d and f orbitals do not penetrate near the nucleus, the quadrupole coupling constant should be largely due to any p electrons present in the valence shell.

Molecular quadrupole coupling constants are usually determined from the hyperfine structure of pure rotational spectra or from electric beam and magnetic beam resonance spectroscopies. Nuclear magnetic resonance, electron spin resonance and Mossbauer spectroscopies are also routes to this property.

I can use the well-studied series HCN, FCN and ClCN to illustrate a couple of points. These molecules, including many isotopic species, have been exhaustively studied by spectroscopic techniques. An interesting feature of the experimental geometries is that the CN bond lengths are almost identical in length yet the ^{14}N quadrupole coupling constant (QCC) is quite different (Table 17.5).

Table 17.6 shows results calculated at the HF/6-311G** level of theory, including geometry optimization.

Table 17.5 Experimental results for XCN

Molecule	$R(\text{C}\equiv\text{N})/\text{pm}$	^{14}N QCC/MHz
HCN	115.5	-4.58
FCN	115.9	-2.67
ClCN	115.9	-3.63

Table 17.6 HF/6-311G** results for XCN

Molecule	$R(\text{C}\equiv\text{N})/\text{pm}$	^{14}N QCC/MHz
HCN	112.7	-4.53
FCN	112.5	-2.95
ClCN	112.8	-3.85

The agreement with experiment is good to within a few percent. What chemical insight these calculations give into the reason why the three coupling constants are so different is not clear!

17.11 Hyperfine Interactions

We considered a simple two-state system in Chapter 7; I took for granted you were familiar with electron and nuclear spin, and had come across nuclear magnetic resonance and electron spin resonance.

Atomic and molecular magnetic dipoles have to obey the quantum mechanical laws of angular momentum, and each dipole can make only a number of allowed orientations with an applied magnetic field. Each orientation corresponds to a different energy and absorption of a suitable photon may cause a change in orientation. This is the physical basis of the magnetic resonance experiment.

The energy level differences are tiny in comparison to electronic ones, and it is conventional to discuss such hyperfine effects as perturbations on the electronic energy levels. In the electron spin resonance experiment, there are two terms of interest. First there is a classical electron spin dipole–nuclear spin dipole term

$$g_e \beta_e \sum_i \sum_\alpha g_\alpha \beta_\alpha \left(\frac{3 (\mathbf{R}_{\alpha i} \cdot \mathbf{S}_i) (\mathbf{R}_{\alpha i} \cdot \mathbf{I}_\alpha) - R_{\alpha i}^2 (\mathbf{S}_i \cdot \mathbf{I}_\alpha)}{R_{\alpha i}^5} \right) \quad (17.31)$$

that averages to zero for tumbling species in the gas phase or in solution.

Second there is the *Fermi contact term*

$$\frac{4\pi}{3} g_e \beta_e \sum_i \sum_\alpha g_\alpha \beta_\alpha \mathbf{I}_\alpha \cdot \mathbf{S}_i \delta(R_{\alpha i}) \quad (17.32)$$

Here \mathbf{I} is a nuclear angular momentum operator, \mathbf{S} an electron angular momentum operator and $\mathbf{R}_{\alpha i}$ the vector joining nucleus α and electron i . The g 's and β 's are the corresponding g -factors and magnetons.

Experimentalists work with a *spin Hamiltonian* which in the latter case would be written

$$h \sum_{\alpha} a_{\alpha} \hat{S}_z \hat{I}_{\alpha z} \quad (17.33)$$

This spin Hamiltonian operates only on spin wavefunctions, and all electronic details are absorbed into the *coupling constant* a_{α} . These can be written very neatly in terms of the spin density $Q(\mathbf{r})$ as

$$a_{\alpha} = \frac{4\pi}{3} g_e \beta_e g_n \beta_n \frac{h}{\langle S_z \rangle} Q(\mathbf{R}_{\alpha}) \quad (17.34)$$

where the spin density has to be evaluated at the position of nucleus α . The spin density is the difference between the α - and the β -spin electrons, and unlike the total electron density it can have negative as well as positive regions.

To give an example, the vinyl radical $\text{H}_2\text{C}=\text{CH}$ (Figure 17.8) can be produced by cleavage of a C–H bond in ethane, and it has been well studied in the gas phase.

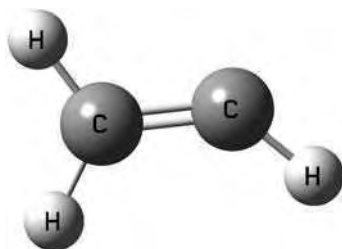


Figure 17.8 *The vinyl radical*

A ROHF-6-311G** calculation gives the HOMO shown in Figure 17.9. In the language of descriptive organic chemistry, this corresponds to the odd electron occupying a C sp^3 hybrid, but of course there are negative regions as well as positive. The spin density is just the square of the HOMO and so is positive everywhere. The experimental evidence is that coupling constants can be positive or negative and so the ROHF model is not satisfactory.

In the UHF model, the α - and β -spin electrons occupy spatially different orbitals and so the spin density can indeed be negative in regions of space. A UHF/6-311G** calculation on the vinyl radical gives regions of space where the spin density is negative and in particular it is negative at the position of the second carbon atom.

Some results of calculations and experiments are given in Table 17.7. I have put the experimental negative signs in brackets because the sign of the coupling constant cannot easily be deduced from studies in solution. Agreement with experiment is not particularly

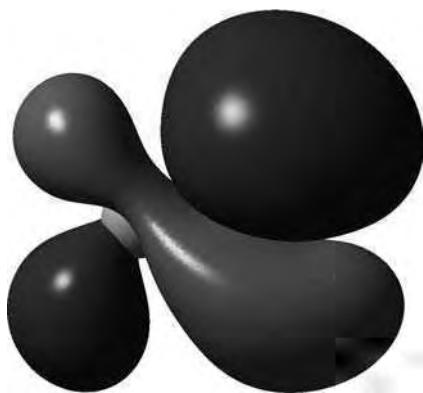


Figure 17.9 ROHF calculation on the vinyl radical

good in either case, and a further major problem with the UHF model is that it does not necessarily give a pure spin state

$$\hat{S}^2 \Psi_{\text{UHF}} \neq s(s+1) \frac{h^2}{4\pi^2} \Psi_{\text{UHF}}$$

In general, if there are p electrons of α spin and q electrons of β spin, the UHF wavefunction can contain spin components ranging from $s = \frac{1}{2}(p - q)$ to $\frac{1}{2}(p + q)$. The expectation value

$$\langle \hat{S}^2 \rangle = \int \Psi_{\text{UHF}} \hat{S}^2 \Psi_{\text{UHF}} d\tau ds$$

is taken as a measure of the spin impurities. In this calculation it came out as 0.9901 instead of the pure doublet value $3/4$.

Table 17.7 Hyperfine coupling constants (in mT) for vinyl radical

	C ₁	H ₁	C ₂	H ₂	H ₃
Experiment	10.76	(-)-1.33	(-)-0.86	6.81	3.41
ROHF	11.49	2.37	1.19	2.20	1.32
UHF	17.75	-1.33	-4.22	6.53	4.56

17.12 Problem Set

A set of problems, together with solutions, is available on the associated website.

References

- Brillouin, L., (1933) *Actual. Sci. Ind.*, 71.
- Clar, E. (1952) *Aromatische Kohlenwasserstoffe*, Springer Verlag, Berlin.
- Foresman, J.B., Head-Gordon, M., Pople, J.A. and Frisch, M.J. (1992) *J. Phys. Chem.*, **96**, 135.
- Mason, S.F. (1962) *J. Chem. Soc.*, 493.
- Raffeneti, R.C. (1973) *Chem. Phys. Lett.*, **20**, 335.
- Scrocco, E. and Tomasi, J. (1978) in *Topics in Current Chemistry*, 42, *New Concepts II*, Springer-Verlag, Berlin.
- Townes, C.H. and Dailey, B.P. (1949) *J. Chem. Phys.*, **17**, 782.

18

Semiempirical Models

I have spent a good deal of time discussing the current state of the art in quantum chemical calculations, and I have deliberately focused on techniques where we perform all of the calculations as exactly as possible. Over the years, it has become fashionable to refer to such models as *ab initio*. This Latin phrase translates into ‘from the beginning’; it certainly does not necessarily mean ‘correct’. Once we have stated the Hamiltonian and if necessary the atomic orbital basis set then all the integrals are evaluated exactly without any attempt at calibrating them against experiment. Thus the atomic Hartree–Fock model is an *ab initio* model, and the calculations I described for the hydrogen molecule ion and dihydrogen are *ab initio* calculations.

These early calculations were done in the 1920s and the 1930s, and you probably noticed that I then jumped to the 1970s to discuss modern HF-LCAO calculations with GTO. Scientists first got their hands on computers in a meaningful way in the 1970s, and one unfortunate characteristic of *ab initio* calculations is that they are compute-intensive. From the early days, by necessity rather than choice, workers tried to develop simple HF-LCAO-based models that applied to certain classes of molecule, and a characteristic of these models is that they avoid all the difficult integrals. There is usually a price to pay, and the models have to be calibrated against experiment; some workers regard this as a major advantage in order to achieve meaningful chemical predictions from cheap calculations. We often refer to such calibrated models as *semiempirical*. This is not a Latin phrase but it translates into ‘based on a correct model but calibrated and/or parameterized’. We can of course argue for ever about what constitutes a ‘correct’ model; Hartree–Fock theory is certainly not absolutely ‘correct’ because it averages over the electron interactions and so it usually runs into trouble when trying to describe bond breaking.

18.1 Hückel π -Electron Theory

Molecules such as benzene, pyridine and naphthalene that have extensive π -bonding systems are not well described by classical valence theory. Such molecules are often planar, or largely planar, and two types of bond contribute to their structure: σ and π bonds. The σ bonds are thought of as largely localized between atoms, whilst the π bonds are to be thought of as delocalized over large portions of the molecule. Much of the interesting chemistry of such compounds appears to relate to the so-called π -electrons, and we visualize these π -electrons as moving under the influence of the nuclear framework and the σ -electrons. The σ -electrons are thought to form an inert core.

We meet such compounds every day; for example, β -carotene is found in carrots and is known to be a biological precursor of vitamin A. In 1933, the compound benzo[a]pyrene was isolated from coal tar; this compound and 7,12-dimethylbenz[a]anthracene are two of the most potent carcinogens known. Such materials are found in soot and cigarette smoke. On the other hand, methyl orange is a prototype for a class of commercially important compounds known as *azo dyes*. The structures of β -carotene, benzo[a]pyrene and methyl orange are shown in Figure 18.1.

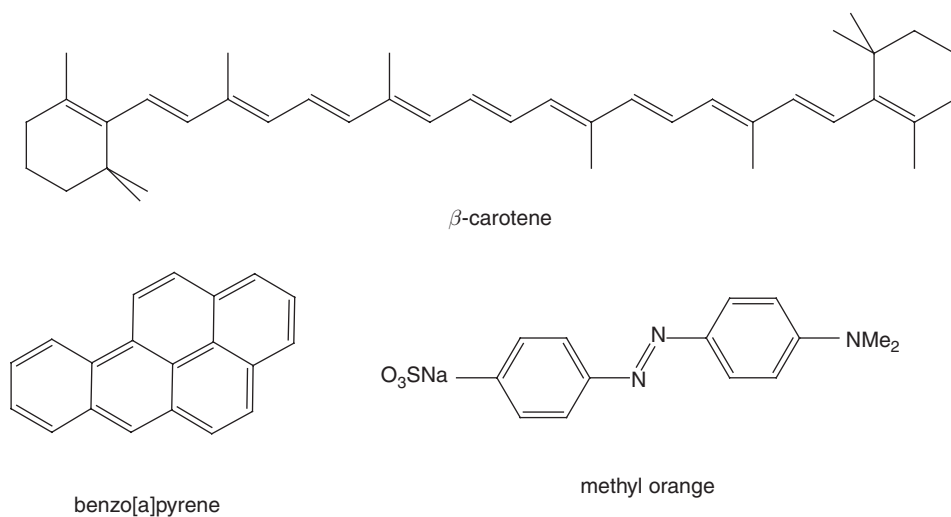


Figure 18.1 Some conjugated systems

Probably for all these reasons, people have striven from the early days of quantum chemistry to give simple models of such compounds in an effort to:

- explain their special structure, stability and properties;
- predict new properties of related but undiscovered molecules (without the labour of having to actually synthesize them).

One of the earliest models for π -electron molecules is the Hückel π -electron model, which dates from 1931 (Hückel 1931). The ideas are simple and chemically appealing, and

the model enjoyed many years of successful application to individual molecules, clusters and solids.

Imagine then a simple molecule such as ethene; there are 16 electrons in total ($14\sigma + 2\pi$) and as usual we concentrate on the electronic wavefunction. The total electronic wavefunction will be some complicated function of the space and spin variables of all 16 electrons, but a basic tenet of Hückel theory is that we can absorb the 14σ -electrons into the ‘core’ and just concentrate on the 2π -electrons.

This is a perfectly sound procedure, akin to the idea of effective core potentials to be discussed later.

In its simplest form, Hückel theory uses the HF-LCAO idea and seeks to write the π molecular orbitals as linear combinations of a suitable set of atomic basis functions; these are not actually defined but we can think of them as the carbon $2p_\pi$ atomic orbitals. Because the atomic $2p_\pi$ orbitals have a nodal plane in the molecular plane, we do not need to consider any other atomic orbitals (such as the C 2s and the H 1s). So, it is an extremely minimal basis set calculation.

If I label the basis functions χ_1 and χ_2 , then we write the HF-LCAO's ψ as

$$\psi = c_1\chi_1 + c_2\chi_2$$

and find the c 's from the standard HF-LCAO matrix eigenvalue problem

$$\mathbf{h}^F \mathbf{c} = \varepsilon \mathbf{S} \mathbf{c}.$$

At this point, Hückel theory introduces a number of simplifying assumptions.

- The basis functions are assumed not to overlap each other, and we also take them to be normalized. This means that the overlap matrix \mathbf{S} is the unit matrix and so the HF-LCAO matrix eigenvalue problem simplifies to $\mathbf{h}^F \mathbf{c} = \varepsilon \mathbf{c}$.
- The elements of the HF-LCAO Hamiltonian are taken to be constants that depend only on the atom orbital types involved. The ‘atom’ parameter α depends only on the type of atom and not the chemical environment; the ‘bond’ parameters depend only on the constituent atom types. Matrix elements between orbitals on nonbonded atoms are set to zero. This means there are no HF-LCAO-type iterations, all that is needed is a single matrix diagonalization. In Hückel's day this was a major problem.
- Atom and bond parameters are only given numerical values as a last resort.

Thus, for example, in the case of ethene we have to solve

$$\begin{pmatrix} \alpha_C & \beta_{CC} \\ \beta_{CC} & \alpha_C \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \varepsilon \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

No HF-SCF iterations are needed because the terms in the square matrix are constant; they do not depend on the electron density (through the c 's). At first sight, it looks like we might need numerical values for α_C and β_{CC} but the trick is to divide through by β_{CC} and put $x = (\alpha_C - \varepsilon)/\beta_{CC}$ giving a matrix eigenvalue problem

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = x \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

The script below shows how to tackle the problem using Mathcad 8.

<pre> ORIGIN := 1 H_{1,1} := 0 H_{2,2} := 0 H_{1,2} := 1 H_{2,1} := 1 H = [0 1] [1 0] x := eigenvals (H) v1 := eigenvec (H, x₁) v2 := eigenvec (H, x₂) </pre>	<p>Sample Huckel MO calculation using Mathcad 8.0 H is the Huckel hamiltonian the vector x contains the eigenvalues vectors v1, v2 contain the corresponding eigenvectors</p> <p style="text-align: center;">+</p> $x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ $v1 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$ $v2 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$
--	---

It turns out that β is a negative quantity, so in the electronic ground state the first π MO corresponds to eigenvector v_2 . I have therefore relabelled the results in Table 18.1.

Table 18.1 Hückel calculation on ethane

MO and energy	c_1	c_2
$\psi_1\alpha + \beta$	0.7071	0.7071
$\psi_2\alpha - \beta$	0.7071	-0.7071

MO ψ_1 is doubly occupied. In organic-speak, this is the HOMO whilst ψ_2 is the LUMO. The HOMO is $\psi_1 = 0.7071(\chi_1 + \chi_2)$ and so the electron density is twice the square of this times $-e$. A simple calculation shows that the Mulliken population indices are, when collected as the matrix of charges and bond orders

$$\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

Each C atom formally contributes a single electron to the π system leaving a C^+ and so the overall charge on each C atom is 0.

18.1.1 Treatment of Heteroatoms

How do we deal with a heteroatomic system such as pyridine? We need new parameters for each heteroatom and a new parameter for each new bond. For historical reasons these are written in terms of the carbon parameters as

$$\alpha_X = \alpha_C + h_X\beta_{CC}$$

$$\beta_{XY} = k_{XY}\beta_{CC}$$

where the h_X and the k_{XY} have to be determined by appeal to experiment. For example, the nitrogen atom in pyridine ought to carry a larger negative charge because nitrogen is more electronegative than carbon and a direct calculation shows that h_N should be positive. Also, the values are different for the N atom in pyridine (where the N atom formally donates one electron to the π system) and pyrrole (where the N atom donates two electrons) (Figure 18.2). Typical values are shown in Table 18.2. Notice that CH_3 is treated as a π group.

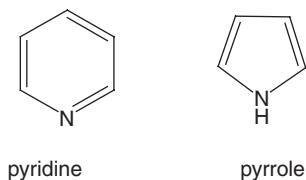


Figure 18.2 Pyridine and pyrrole

Table 18.2 Typical h and k parameters

Atom X	h_X	k_{CX}
C	0	1
N (one π -electron)	0.5	0.8
N (two π -electrons)	1.5	1.0
B	-1.0	0.7
O (two π -electrons)	1.0	0.8
CH_3	2.0	0.7

Thus for pyridine with the N atom labelled 1, the secular equations are

$$\begin{pmatrix} 0.5 & 0.8 & 0 & 0 & 0 & 0.8 \\ 0.8 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0.8 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix} = x \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix}$$

18.2 Extended Hückel Theory

Hückel's calculations on planar conjugated systems were welcomed by the scientific community of the 1930s and were heavily exploited for many years. You might like to glance at the classic text *Molecular Orbital Theory for Organic Chemists* (Streitwieser 1962) to get the flavour of those early days.

Despite these successes, the natives were restless. Theoreticians were unhappy because of the arbitrary division of electrons into 'core' and 'valence' or ' π -electron' with no apparent justification other than it worked; how did the π -electrons know the core electrons were there, if they were not formally included in the calculation? Organic chemists were unhappy

because most of their interesting molecules did not contain benzene rings and inorganic chemists were particularly aggrieved because there was no mention of inorganic molecules in Hückel's theory.

Hückel's calculations referred to the so-called π -electron system. What few calculations had been done on the core had used a method described as the 'linear combination of bond orbitals', where Hückel-type schemes were used to try and combine together (for example) sp^3 hybrid orbitals.

18.2.1 Wolfsberg and Helmholtz

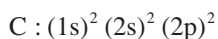
The pioneering work of Max Wolfsberg and Lindsay Helmholtz on the electronic structure of the ions CrO_4^- , MnO_4^- and CrO_4^{2-} is usually cited as the first serious attempt to apply the Hückel model to inorganic complexes. The synopsis is given below (Wolfsberg and Helmholtz 1952).

We have made use of a semi empirical treatment to calculate the energies of the molecular orbitals for the ground state and the first few excited states of permanganate, chromate and perchlorate ions. The calculations of the excitation energies is in agreement with the qualitative features of the observed spectra, i.e. absorption in the far ultraviolet for ClO_4^- with two strong maxima in the visible or near ultraviolet for MnO_4^- and CrO_4^{2-} with the chromate spectrum displaced towards higher energies. An approximate calculation of the relative f -values for the first two transitions in CrO_4^{2-} and MnO_4^- is also in agreement with experiment

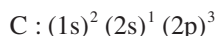
The authors point out that the tetrahedral ions PO_4^{2-} , SO_4^{2-} and ClO_4^- show no visible or near-ultraviolet absorption whilst the ions XO_4^- of the fourth row transition elements, which have the same number of valence electrons, show characteristic visible and near-ultraviolet absorption, namely two strong maxima with the corresponding peaks displaced towards shorter wavelengths (higher energies) as the atomic number of the central atom decreases. The oscillator strength (f) relates to the area under an absorption peak and is also a theoretical measure of the intensity of a given transition.

The extended Hückel model treats all valence electrons, within the spirit of the π -electron model. Each molecular orbital is written as an LCAO expansion of these valence orbitals, which can be thought of as being Slater-type orbitals; these orbitals are very similar to hydrogenic ones except that they do not have radial nodes. Once again we can understand the model best by considering the HF-LCAO equations.

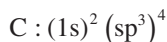
For H and He, the atomic basis set consists of a single 1s orbital. For Li through Ne, the inner shell electrons are treated as part of the core and the basis functions used are atomic 2s, 2p_x, 2p_y and 2p_z. For Na through Al, the inner shell is treated as part of the core and we consider only 3s, 3p_x, 3p_y and 3p_z orbitals. For Si through Cl we have to decide on whether or not to include the atomic 3d orbitals in addition, and practice varies. Most authors included them, especially for hypervalent molecules such as ClO_4^- . The diagonal elements of the HF-LCAO matrix are taken to be the negatives of the valence shell ionization energy (VSIE) for the orbital in question. You have probably met VSIE elsewhere; the orbital configuration for carbon ground state is written



which is arguably unhelpful when we want to describe the bonding in methane. We imagine a two-step process where one electron is promoted from the 2s to a 2p orbital giving



The four valence orbitals are then mixed to form sp^3 hybrids, giving four equivalent orbitals



which is said to form the valence state. Sadly, it does not correspond to a known spectroscopic state of the carbon atom and so you will anticipate that many learned discussions took place to try and assign an energy value to the diagonal matrix elements of extended Hückel theory. Things are far, far worse for transition metal ions with variable valencies. Wolfsberg and Helmholtz went to great pains to try and get reasonable values for the diagonal terms by comparison with electronic spectra. Modern thinking is that the parameters are simply parameters that should be chosen to give a good fit of some measurable physical property against experiment, and that they should take the same value for a given element irrespective of its chemical environment.

Wolfsberg and Helmholtz did their calculations in terms of ‘group’ orbitals. Such orbitals are simply linear combinations of atomic orbitals that mirror the symmetry of the molecule. In those days, it was essential to reduce the size of any matrix eigenvalue problem to one that could be done by hand. We have moved on and do not generally bother with group orbitals any more, although an understanding of molecular symmetry is vital in order to understand the spectroscopic transitions.

Once the diagonal elements have been assigned, the off-diagonal elements are calculated by the key formula

$$h_{ij}^F = kS_{ij} \frac{h_{ii}^F + h_{jj}^F}{2} \quad (18.1)$$

where S_{ij} is the overlap integral between the two basis functions χ_i and χ_j . Parameter k is a ‘constant’ that can be adjusted to give agreement with experiment; common experience suggests that 1.75 is a reasonable choice.

As a variation on the theme, some early authors chose to write

$$h_{ij}^F = k'S_{ij} \sqrt{h_{ii}^F \times h_{jj}^F} \quad (18.2)$$

To use the chlorate ion example cited by Wolfsberg and Helmholtz, we would these days use a basis set comprising 2s and 2p STOs on oxygen with 3s, 3p and 3d orbitals on chlorine giving a total of 25 basis functions. I have added d orbitals on Cl because it is hypervalent in the molecule under study and people argue that d orbitals are necessary in such circumstances. That gives 16 doubly occupied valence shell molecular orbitals (made up as follows: 6 valence electrons from each oxygen, 7 from chlorine and 1 from the excess negative charge). A standard modern extended Hückel calculation with a tetrahedral geometry and bond length of 147 pm gave me the orbital energies listed in Table 18.3. You might get slightly different results depending on your choice of Hückel parameters; I used the standard default values in Gaussian 03. Notice how close together the highest occupied orbitals are; the order obtained is a little different from that reported by Wolfsberg and Helmholtz, but that is not surprising. Unlike advanced theories where electron repulsion is taken explicitly into account, Hückel excitation energies are given as differences in orbital energies.

Table 18.3 *Extended Hückel treatment of ClO_4^- (tetrahedral, $R_{\text{Cl-O}} = 148 \text{ pm}$)*

Orbital	Energy/ E_h
$4t_2$	0.5234
$3a_1$	0.4938
$3t_2$	-0.5340
$1t_1$	-0.5361
$1e$	-0.5487
$2t_2$	-0.5890
$2a_1$	-0.7312
$1t_2$	-1.2324
$1a_1$	-1.3718

Some transitions are allowed (and possibly strong) and some are forbidden (but often are present as weak bands). Without going into the details and the group theory, my prediction is that absorption bands for ClO_4^- should be seen at about 1 hartree (1 hartree = 27.2114 eV), which is much too large compared to the experimental value of 6 eV. Perhaps you can see now why Wolfsberg and Helmholtz had to fiddle around so much in order to get decent agreement with experiment. ClO_4^- is of course colourless, so at least the prediction is on the right side of credibility.

18.2.2 Roald Hoffmann

Credit for extending the Hückel ideas to organic molecules belongs to Roald Hoffmann (Nobel prizewinner in 1981). The synopsis of his 1963 paper says it all (Hoffmann 1963).

The Hückel theory, with an extended basis set consisting of 2s and 2p carbon and 1s hydrogen orbitals, with inclusion of overlap and all interactions, yields a good qualitative solution of most hydrocarbon conformational problems. Calculations have been performed within the same parameterization for nearly all simple saturated and unsaturated compounds, testing a variety of geometries for each. Barriers to internal rotation, ring conformations, and geometrical isomerism are among the topics treated. Consistent σ and π charge distributions and overlap populations are obtained for aromatics and their relative roles discussed. For alkanes and alkenes charge distributions are also presented. Failures include overemphasis on steric factors, which leads to some incorrect isomerization energies; also the failure to predict strain energies. It is stressed that the geometry of a molecule appears to be its most predictable property.

Valence shell orbitals are used, and the same key formula (Equation (18.1)) is used to find the Hamiltonian.

18.3 Pariser, Parr and Pople

The next advances came in the 1950s, with a more systematic treatment of electron repulsion in π -electron molecules. These two keynote papers due to Rudolf Pariser and Robert Parr and to John Pople are so important that I will give the abstracts (almost) in full. Pople (1953) wrote:

An approximate form of the molecular orbital theory of unsaturated hydrocarbon molecules in their ground states is developed. The molecular orbital equations rigorously derived from the correct many-electron Hamiltonian are simplified by a series of systematic approximations and reduce to equations comparable with those used in the semi-empirical method based on an incompletely defined one-electron Hamiltonian. The two sets of equations differ, however, in that those of this paper include certain important terms representing electronic interactions. The theory is used to discuss the resonance energies, ionization potentials, charge densities, bond orders and bond lengths of some simple hydrocarbons. The electron interaction terms introduced in the theory are shown to play an important part in determining the ionization potentials. . .

You should have picked up many of the key phrases. He started from the HF-LCAO equations and made what is now known as the $\sigma - \pi$ separation approximation; the π -electrons are treated separately and the effect of the remaining σ -electrons is absorbed into the HF-LCAO Hamiltonian. The HF-LCAO equations have to be solved iteratively in order to get the HF-LCAO π -electron molecular orbitals, and many of the two-electron integrals (the 'electronic interaction terms') are retained. In order to take account of the effect of $\sigma - \pi$ separation, most integrals are calibrated by appeal to experiment. The 'charges and bond orders' are simply the Mulliken populations calculated with an overlap matrix equal to the unit matrix, and ionization energies are calculated according to Koopmans' theorem.

The second keynote paper by Pariser and Parr (1953) also gives a snapshot of the times, when there was a great deal of interest in the electronic spectra of conjugated molecules. They wrote:

A semi-empirical theory is outlined which is designed for the correlation and prediction of the wavelengths and intensities of the first main visible or ultraviolet bands and other properties of complex unsaturated molecules, and preliminary application of the theory is made to ethylene and benzene.

The theory is formulated in the language of the purely theoretical method of the antisymmetrized products of molecular orbitals (in LCAO approximation) including configuration interaction, but departs from this theory in several essential respects. First, atomic orbital integrals involving the core Hamiltonian are expressed in terms of quantities which may be regarded as semi-empirical. Second, an approximation of zero differential overlap is employed and an optionally uniformly charged sphere representation of atomic π -orbitals is introduced, which greatly simplify the evaluation of electronic repulsion integrals and make applications to complex molecules containing heteroatoms relatively simple. Finally, although the theory starts from the π -electron approximation, in which the unsaturated electrons are treated apart from the rest, provision is included for the adjustment of the σ -electrons to the π -electron distribution in a way which does not complicate the mathematics.

Once again you should have picked up many of the key phrases. We often speak of *the PPP method* in honour of its three originators.

18.4 Zero Differential Overlap

A key phrase that I have yet to explain is *zero differential overlap (ZDO)*; the atomic orbital basis set is not rigorously defined but we can imagine it to comprise the relevant STO with

one atomic orbital per heavy atom. Hydrogen atoms do not enter into π -electron models. The basis functions are taken to be orthonormal, so we have

$$\int \chi_i(\mathbf{r})\chi_j(\mathbf{r}) d\tau = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

ZDO extends this idea to the two-electron integrals; if we have a two-electron integral of the type

$$\frac{e^2}{4\pi\epsilon_0} \int \dots \frac{1}{r_{12}} \chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_1) d\tau_1$$

then the integral is set to zero unless $i = j$. So all two-electron integrals

$$\frac{e^2}{4\pi\epsilon_0} \int \chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (18.3)$$

are zero unless $i = j$ and $k = l$.

At first sight the basis functions are STO $2p_\pi$, and the remaining integrals can actually be calculated exactly. When Pariser and Parr first tried to calculate the ionization energies and spectra of simple conjugated hydrocarbons such as benzene with exact two-electron integrals, they got very poor agreement with experiment. They therefore proposed that the two-electron integrals should be treated as parameters to be calibrated against spectroscopic data. Once again, you have to remember that computers were still in their infancy and that any simple formula was good news. We had electromechanical calculators and log tables in those days, but nothing much else. One such popular formula was

$$\int \chi_i(\mathbf{r}_1)\chi_i(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2) \chi_k(\mathbf{r}_2) d\tau_1 d\tau_2 = \frac{1}{R_{ik} + \alpha_{ik}} \quad (18.4)$$

where R_{ik} is the separation between the orbital centres. All other two-electron integrals were taken as zero and the α_{ik} term has to be fixed by appeal to spectroscopic experiment. The expression is what you would get for the mutual potential energy of a pair of electrons separated by distance $R_{ik} + \alpha_{ik}$, and these repulsion integrals are traditionally given a symbol γ_{ik} .

Next we have to address the one-electron terms, and the treatment depends on whether the term is diagonal ($i = j$) or off-diagonal ($i \neq j$). In standard HF-LCAO theory the one-electron terms are given by

$$\int \chi_i \left(-\frac{h^2}{8\pi^2 m_e} \nabla^2 - \frac{e^2}{4\pi\epsilon_0} \sum_{I=1}^N \frac{Z_I}{R_I} \right) \chi_j d\tau \quad (18.5)$$

where the first term represents the kinetic energy of the electron and the second term gives the mutual potential energy between the electron and each nucleus.

In Pariser–Parr–Pople theory, the off-diagonal elements are taken as zero unless the atom pairs are directly bonded. If the atom pairs are directly bonded the matrix element is given a constant value β for each particular atom pair, which is certainly not the same value as the β value in ordinary Hückel theory. The value of either β is found by calibration against experiment. The one-electron diagonal terms are written so as to separate out the contribution from the nuclear centre I on which the atomic orbital is based (usually taken as minus the valence state ionization energy ω_I) and the other nuclear centres J (usually

taken as $-Z_j\gamma_{ij}$, where Z is the formal number of electrons that an atom contributes to the π system). There is no problem with hydrocarbons in that all carbon atoms are assigned the same valence state ionization energies, irrespective of their chemical environment, but many papers were written to justify one particular choice of ω_i against another for nitrogen.

For the record, the PPP HF-LCAO Hamiltonian \mathbf{h}^F for hydrocarbons and other first-row systems has dimension equal to the number of conjugated atoms and can be written

$$h_{ii}^F = \omega_i + \frac{1}{2}P_{ii}\gamma_{ii} + \sum_{j \neq i} (P_{ij} - Z_j)\gamma_{ij}$$

$$h_{ij}^F = \beta_{ij} - \frac{1}{2}P_{ij}\gamma_{ij}$$
(18.6)

Here \mathbf{P} is the usual matrix of charges and bond orders. The HF-LCAO equations are solved by the usual techniques.

In those early days, no one dreamed of geometry optimization; benzene rings were assumed to be regular hexagons with equal C–C bond length of 140 pm. PPP calculations were state of the art in the early 1970s and consumed many hours of early computer time. Once again there was an argument as to the role of d orbitals on sulfur, and a number of long-forgotten schemes and methodologies were advanced.

18.5 Which Basis Functions Are They?

I seem to have made two contradictory statements about the basis functions used in semiempirical work. On the one hand, I have said that they are orthonormal and so their overlap matrix is the unit matrix; on the other hand, I have used overlap integrals to calculate certain integrals.

Think of a Hückel π -electron treatment of ethene, and call the carbon $2p_\pi$ orbitals χ_1 and χ_2 . The matrix of overlap integrals is

$$\mathbf{S} = \begin{pmatrix} 1 & p \\ p & 1 \end{pmatrix}$$

where p is the overlap integral of the two atomic (STO) orbitals in question. The eigenvalues of this matrix are $1 \pm p$ and the normalized eigenvectors are

$$\mathbf{v}_1 = \sqrt{\frac{1}{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_2 = \sqrt{\frac{1}{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

A little matrix algebra will show that

$$\mathbf{S} = (1 + p) \mathbf{v}_1 \mathbf{v}_1^T + (1 - p) \mathbf{v}_2 \mathbf{v}_2^T$$

Mathematicians have a rather grand expression for this: they talk about the *spectral decomposition* of a matrix. We can make use of the expression to calculate powers of matrices, such as the negative square root

$$\mathbf{S}^{-1/2} = (1 + p)^{-1/2} \mathbf{v}_1 \mathbf{v}_1^T + (1 - p)^{-1/2} \mathbf{v}_2 \mathbf{v}_2^T$$

To cut a long story short, we regard the basis functions used in semiempirical calculations as related to ordinary STOs $\chi_1, \chi_2, \dots, \chi_n$ by the matrix transformation

$$\begin{pmatrix} \chi_1 & \chi_2 & \dots & \chi_n \end{pmatrix} \mathbf{S}^{-1/2} \quad (18.7)$$

These have the property that they are orthonormal, yet resemble the ordinary STOs as closely as possible, in a least-squares sense.

18.6 All Valence Electron ZDO Models

The early π -electron semiempirical models proved a great success, and they are still sometimes encountered in current publications. Attempts to extend them to the σ systems or to inorganic molecules met with mixed fortune for a variety of reasons, especially the following three.

- If we draw a molecule and then arbitrarily choose an axis system, physical properties such as the energy should not depend on the choice of axis system. We speak of *rotational invariance*; the answers should be the same if we rotate a local molecular axis system.
- Also, despite what one reads in elementary organic texts, calculated physical properties ought to be the same whether one works with ordinary atomic orbitals or the mixtures we call hybrids.
- Finally, we should get the same answers if we work with symmetry-adapted combinations of atomic orbitals rather than the ‘raw’ atomic orbitals. Whatever the outcome, we will not get a different energy.

18.7 CNDO

Pople and Segal (1965) seem to be the first authors to have addressed these problems. The most elementary theory that retains all the main features is the complete neglect of differential overlap (CNDO) model. The first paper dealt mainly with hydrocarbons, and only the valence electrons were treated. The inner shells contribute to the core that modifies the one-electron terms in the HF-LCAO Hamiltonian. The ZDO approximation is applied to all two-electron integrals so that

$$\iint \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (18.8)$$

is zero unless $i = j$ and $k = l$. Suppose now that atoms A and B are both carbon, and so we take $2s$, $2p_x$, $2p_y$ and $2p_z$ basis functions for either atom. In addition to the ZDO approximation, the CNDO model requires that all remaining two-electron integrals of type (18.8) involving basis functions on A and B are equal to a common value denoted γ_{AB} . Parameter γ_{AB} depends on the nature of the atoms A and B but not on the details of the atomic orbitals. This simplification guarantees rotational invariance.

Similar considerations apply to the one-electron terms

$$\int \chi_i \left(-\frac{\hbar^2}{8\pi^2 m_e} \nabla^2 - \frac{1}{4\pi \epsilon_0} \sum_{\alpha=1}^N \frac{Z_\alpha}{R_\alpha} \right) \chi_j d\tau$$

When $i = j$ and the basis function χ_i is centred on nucleus A we write the integral

$$U_{ii} = \int \chi_i \left(-\frac{\hbar^2}{8\pi^2 m_e} \nabla^2 - \frac{1}{4\pi \epsilon_0} \frac{Z_A}{R_A} \right) \chi_i d\tau$$

and determine U from atomic spectral data. The remaining terms in the sum are called *penetration integrals* and written V_{AB} .

If $i \neq j$ and the basis functions are on different atomic centres then all three-centre contributions are ignored. The remaining two-centre terms involving atoms A and B are written $\beta_{AB}^0 S_{ij}$, where S is the overlap integral and β a 'bonding' parameter. The bonding parameter is nonzero only for bonded pairs.

Collecting up terms and simplifying we find that a CNDO HF-LCAO Hamiltonian has elements

$$\begin{aligned} h_{ii}^F &= U_{ii} + \left(P_{AA} - \frac{1}{2} P_{ii} \right) \gamma_{AA} + \sum_{B \neq A} (P_{BB} \gamma_{AB} - V_{AB}) \\ h_{ij}^F &= \beta_{AB}^0 S_{ij} - \frac{1}{2} P_{ij} \gamma_{AB} \end{aligned} \quad (18.9)$$

A and B label atoms, i and j label the basis functions and P_{AA} is the sum of the diagonal charge density matrix for those basis functions centred on atom A.

The original parameter scheme was called CNDO/1. Electron repulsion integrals were calculated exactly, on the assumption that the basis functions were s-type STOs, and all overlap integrals were calculated exactly. The bonding parameters β_{AB}^0 were chosen by comparison with (crude) *ab initio* calculations on relevant small molecules, and a simple additivity scheme was employed:

$$\beta_{AB}^0 = \beta_A^0 + \beta_B^0$$

18.8 CNDO/2

It turned out that CNDO/1 calculations gave poor predictions of molecular geometries, and this failing was analysed as due to the approximations made for U_{ii} and the penetration term V_{AB} . These problems were corrected in CNDO/2; V_{AB} is no longer calculated exactly; rather it is taken as $-Z_B \gamma_{AB}$. The atomic terms become

$$U_{ii} = -\frac{1}{2} (I_i + E_i) - \left(Z_A - \frac{1}{2} \right) \gamma_{AA} \quad (18.10)$$

where I_i and E_i are valence state ionization energies and electron affinities. The HF-LCAO matrix elements are

$$h_{ii}^F = -\frac{1}{2}(I_i + E_i) + \left((P_{AA} - Z_A) - \frac{1}{2}(P_{ii} - 1) \right) \gamma_{AA} + \sum_{B \neq A} (P_{BB} \gamma_{AB} - V_{AB}) \quad (18.11)$$

$$h_{ij}^F = \beta_{AB}^0 S_{ij} - \frac{1}{2} P_{ij} \gamma_{AB}$$

18.9 CNDO/S

One failing of the π -electron models was their inability to treat $n \rightarrow \pi^*$ electronic transitions. The CNDO models gave a great breakthrough because they afforded for the first time a systematic treatment of both $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions in conjugated molecules. The treatment essentially followed the lines of CIS discussed in Chapter 17 but the first results were in poor agreement with experiment in that they gave excitation energies that were much too large. The most significant improvement to CNDO for use in spectroscopy was given by the work of Del Bene and Jaffé (1968). Their major contribution was to modify the bonding integrals; think of a pair of bonded carbon atoms, with one pair of 2p orbitals pointing along the bond (σ) and the remaining two pairs perpendicular to the bond (π). Del Bene and Jaffé introduced a new parameter κ such that the π -type interaction was reduced. They essentially replaced the term $\beta_{AB}^0 S_{ij}$ with

$$\begin{aligned} \beta_{AB}^\sigma S_{ij} &= \frac{1}{2} (\beta_A^0 + \beta_B^0) S_{ij} \\ \beta_{AB}^\pi S_{ij} &= \frac{1}{2} \kappa (\beta_A^0 + \beta_B^0) S_{ij} \end{aligned} \quad (18.12)$$

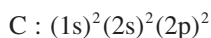
They recommend a value $\kappa = 0.585$, which gives best agreement with experiment for many conjugated molecules.

18.10 INDO

The CNDO models make an unnecessary and draconian simplification to the two-electron integrals

$$\iint \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2) \chi_l(\mathbf{r}_2) d\tau_1 d\tau_2 = \delta_{ij} \delta_{kl} \gamma_{AB}$$

That is to say, even if orbitals χ_i and χ_j are on the same centre but $i \neq j$ then the entire integral is taken to be zero. Direct calculation shows that the integrals can be far from zero. In fact, it is just one-centre integrals of this kind that give the correct ordering between atomic spectroscopic states. CNDO cannot distinguish between the 3P , 1D and 1S states arising from orbital configuration



Pople *et al.* (1967) introduced the intermediate neglect of differential overlap (INDO) scheme, which retained monocentric repulsion integrals. Some of these integrals are taken as parameters to be determined by comparison with experiment rather than by direct calculation.

18.11 NDDO (Neglect of Diatomic Differential Overlap)

The next level of approximation along these lines is the NDDO model, developed by Pople and Beveridge (1970). All two-electron integrals

$$\iint \chi_i(\mathbf{r}_1)\chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_k(\mathbf{r}_2)\chi_l(\mathbf{r}_2) d\tau_1 d\tau_2$$

are retained when χ_i and χ_j are on the same centre and both χ_k and χ_l are on a single different centre. NDDO never really caught on; it was overtaken by events as described in the following.

18.12 The MINDO Family

The CNDO/2, INDO and NDDO methods are now little used apart from providing initial estimates of the electron density for use in *ab initio* HF-LCAO iterations. Their significance is that they were the first family of semiempirical models which retained increasing numbers of smaller repulsion integrals in a systematic way. Even NDDO neglects the majority of two-electron integrals for real molecules, and so the models made HF-LCAO calculations on large systems feasible. They did not give very accurate results, perhaps because they were parameterized by comparing with low accuracy *ab initio* calculations. They also needed reliable geometries as input, because geometry optimization was still in its infancy. Pople and Beveridge's (1970) book contains a listing of a Fortran CNDO/INDO code, and it is interesting to note that there is no mention of geometry optimization.

The MINDO (modified INDO) models developed by Michael J.S. Dewar had a quite different philosophy; their objective was to develop a genuine reliable model for experimentalists. They realized that parameters had to be available for a wide range of atoms and the models had to deliver results to chemical accuracy in reasonable amounts of computer time. Attention was paid to reliable geometry optimization, so enabling the program to accept relatively crude initial geometries for unknown structures.

18.12.1 MINDO/3

The synopsis for the landmark MINDO/3 paper by Bingham *et al.* (1975) is terse:

The problems involved in attempts to develop quantitative treatments of organic chemistry are discussed. An improved version (MINDO/3) of the MINDO semiempirical SCF-MO treatment is described. Results obtained for a large number of molecules are summarised.

MINDO/3 uses an s, p minimal basis set of STOs and the elements of the HF-LCAO matrix are

$$\begin{aligned}
 h_{ii}^F &= U_{ii} + \sum_{j \text{ on A}} \left(P_{ij} \gamma_{ij} - \frac{1}{2} P_{ij} \lambda_{ij} \right) + \sum_{B \neq A} (P_{BB} - Z_B) \gamma_{AB} \\
 h_{ij}^F &= -\frac{1}{2} P_{ij} \lambda_{ij} \text{ if } \chi_i \text{ and } \chi_j \text{ on A} \\
 h_{ij}^F &= h_{ij}^{\text{core}} - \frac{1}{2} P_{ij} \gamma_{AB} \text{ otherwise}
 \end{aligned} \tag{18.13}$$

I have written the atomic Coulomb and exchange integrals

$$\begin{aligned}
 \gamma_{ij} &= \frac{e^2}{4\pi \epsilon_0} \int \chi_i(\mathbf{r}_1) \chi_i(\mathbf{r}_1) \frac{1}{r_{12}} \chi_j(\mathbf{r}_2) \chi_j(\mathbf{r}_2) d\tau_1 d\tau_2 \\
 \lambda_{ij} &= \frac{e^2}{4\pi \epsilon_0} \int \chi_i(\mathbf{r}_1) \chi_j(\mathbf{r}_1) \frac{1}{r_{12}} \chi_i(\mathbf{r}_2) \chi_j(\mathbf{r}_2) d\tau_1 d\tau_2
 \end{aligned}$$

for simplicity of notation. The parameters for MINDO/3 were obtained in an entirely different way from the CNDO/INDO/NDDO family; many quantities like the STO exponents were allowed to vary during the fitting procedure. The bonding parameter β_{AB}^0 was allowed to vary, and experimental data such as enthalpies of formation and accurate molecular geometries were also used to get the best fit.

An interesting feature was the treatment of core-core repulsions (the core in this case being identified with the nucleus plus any inner-shell atomic electrons). The simple Coulomb term

$$U_{AB} = \frac{e^2}{4\pi \epsilon_0} \frac{Z_A Z_B}{R_{AB}}$$

(where Z_A and Z_B are the 'effective' nuclear charges) was modified for various complicated reasons to make it a function of the electron repulsion integrals:

$$U_{AB} = Z_A Z_B \left(\gamma_{AB} + \left(\frac{e^2}{4\pi \epsilon_0} \frac{1}{R_{AB}} - \gamma_{AB} \right) \exp \left(-\frac{\alpha_{AB} R_{AB}}{m} \right) \right) \tag{18.14}$$

Here α_{AB} is a dimensionless constant and it depends on the natures of the atoms A and B. For O-H and N-H bonds a slightly different scheme was adopted:

$$U_{AH} = Z_A Z_H \left(\gamma_{AH} + \left(\frac{e^2}{4\pi \epsilon_0} \frac{1}{R_{AH}} - \gamma_{AB} \right) \alpha_{AH} \exp \left(-\frac{R_{AH}}{m} \right) \right) \tag{18.15}$$

18.13 MNDO

MINDO/3 proved very successful but it had a number of limitations; enthalpies of formation of conjugated molecules were generally too positive, bond angles were not well predicted, and so on. Dewar and Thiel (1977) introduced the modified neglect of differential overlap (MNDO) model, which they based on NDDO whilst retaining the philosophy of MINDO/3.

The core–core repulsions were further modified:

$$U_{AB} = \frac{e^2}{4\pi\epsilon_0} Z_A Z_B \int s_A(\mathbf{r}_1) s_A(\mathbf{r}_1) \frac{1}{r_{12}} s_B(\mathbf{r}_2) s_B(\mathbf{r}_2) d\tau_1 d\tau_2 \quad (18.16)$$

$$\left(1 + \exp\left(-\frac{\alpha_A R_{AB}}{m}\right) + \exp\left(-\frac{\alpha_B R_{AB}}{m}\right) \right)$$

with a different formula for OH and NH bonds.

18.14 Austin Model 1 (AM1)

Next came AM1, due to Dewar *et al.* (1985). AM1 was designed to eliminate the problems from MNDO caused by a tendency to overestimate repulsions between atoms separated by the sum of their van der Waals radii. The strategy adopted was to modify the core–core terms by multiplication of the Coulomb term with sums of Gaussian functions. In the original AM1 paper, there are four terms in the Gaussian expansion. Each Gaussian is characterized by its position along the A–B vector and by its width. This significantly increased the number of parameters for each atom.

18.15 PM3

PM3 is the third parameterization of MNDO, and the PM3 model contains essentially all the same terms as AM1. The parameters for PM3 were derived by Stewart (1989) in a more systematic way than those for AM1, many of which were derived by ‘chemical intuition’. As a consequence, some of the parameters are quite different from those of MNDO but the two models seem to predict physical properties to the same degree of accuracy.

18.16 SAM1

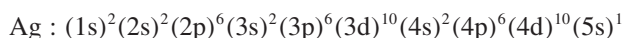
The final 1993 offering in Michael Dewar’s name was semi-*ab-initio* model 1 (SAM1) (Dewar *et al.* 1993). In SAM1, two-electron integrals are calculated using a standard STO-3G basis set (and hence the appearance of *ab initio* in the title). The resulting integrals were then scaled, and the Gaussian terms in the core–core repulsions were retained in order to fine-tune the calculations.

18.17 ZINDO/1 and ZINDO/S

Michael Zerner and co-workers developed these models for transition metal compounds (Zerner 1991). ZINDO/1 is a variant of INDO and ZINDO/S is an INDO-like method used to calculate electronic spectroscopic properties of molecules.

18.18 Effective Core Potentials

The ground state of the silver atom has electron configuration



and much of its chemistry can be explained in terms of the outer 5s electron. The remaining 46 electrons seem to form an inert core and this suggests that we might try to model the silver atom as a nucleus of charge $Z = 47$ and an inner shell of 46 electrons. This idea was tried in the very early days using potentials such as

$$U_{\text{core}} = -\frac{n_v}{4\pi\epsilon_0 r} + A \frac{\exp(-2kr)}{r} \quad (18.17)$$

where n_v is the number of valence electrons and A and k are constants that have to be determined by fitting an atomic property. Use of such a potential means that we have eliminated most of the two-electron integrals.

The idea of dividing electrons into groups is quite common in chemistry; in CNDO theory we treat the valence electrons only, without explicitly treating the core electrons. In ZDO π -electron models we treat the π -electrons explicitly and assume that the effect of the σ -electrons can be taken into account in the parameter choice. There is a serious problem that I can explain by reference to silver. The 5s orbital has to be orthogonal to all the inner orbitals, even though we do not use them (and do not necessarily want to calculate them).

Many *ab initio* packages allow the use of *effective core potentials* (ECPs), which replace the atomic cores in valence-only *ab initio* calculations; traditionally they were represented as linear combinations of functions of the type

$$r^{-n} \exp(-\alpha r^2)$$

with coefficients and exponents determined by fitting the potential generated from accurate HF-LCAO wavefunctions. In recent years it has become fashionable to represent the core potentials by one- and two-term Gaussians obtained directly from the appropriate atomic eigenvalue equation.

18.19 Problem Set

A set of problems, together with solutions, is available on the associated website.

References

- Bingham, R.C., Dewar, M.J.S. and Lo, D.H. (1975) *J. Am. Chem. Soc.*, **97**, 1285.
 Del Bene, J. and Jaffe, H.H. (1968) *J. Chem. Phys.*, **48**, 1807.
 Dewar, M.J.S. and Thiel, W. (1977) *J. Am. Chem. Soc.*, **99**, 4907.
 Dewar, M.J.S., Jie, C. and Yu, J. (1993) *Tetrahedron*, **49**, 5003.
 Dewar, M.J.S., Zoebisch, E.G., Healey, E.F. and Stewart, J.J.P. (1985) *J. Am. Chem. Soc.*, **107**, 3902.
 Hoffmann, R. (1963) *J. Chem. Phys.*, **39**, 1397.
 Hückel, E.P. (1931) *Z. Physik*, **70**, 204.

- Pariser, R. and Parr, R.G. (1953) *J. Chem. Phys.*, **21**, 466.
- Pople, J.A. (1953) *Trans. Faraday Soc.*, **49**, 1375.
- Pople, J.A. and Beveridge, D.L. (1970) *Approximate Molecular Orbital Theory*, McGraw-Hill, New York.
- Pople, J.A. and Segal, G.A. (1965) *J. Chem. Phys.*, **43**, S129.
- Pople, J.A., Beveridge, D. and Dobosh, P.A. (1967) *J. Chem. Phys.*, **47**, 2026.
- Stewart, J.J.P. (1989) *J. Comput. Chem.*, **10**, 209, 221.
- Streitwieser, A. (1962) *Molecular Orbital Theory for Organic Chemists*, John Wiley & Sons, Inc., New York.
- Wolfsberg, M. and Helmholz, L. (1952) *J. Chem. Phys.*, **20**, 837.
- Zerner, M. (1991) in *Reviews of Computational Chemistry* (eds K.B. Lipkowitz and D.B. Boyd), VCH Publishing, New York.

19

Electron Correlation

I have mentioned electron correlation at several points in the text, and I gave an operational definition of correlation energy as the difference between the exact HF energy and the true experimental energy. There is a great deal of small print; in general we cannot obtain exact HF wavefunctions for molecules, only LCAO approximations. The Schrödinger treatment takes no account of the theory of relativity, whilst we know from simple atomic spectra that relativistic effects are non-negligible. We have to be careful to treat zero-point energy in a consistent manner when dealing with vibrations, and so on.

19.1 Electron Density Functions

I have put a great deal of emphasis on the electron density. The wavefunction for a molecule with n electrons will depend on the $3n$ spatial electron coordinates $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ together with the n spin variables s_i (α or β). Many authors combine space and spin variables into a composite $\mathbf{x} = \mathbf{r}s$ and so we would write a wavefunction

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$$

According to the Born interpretation,

$$\Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) d\tau_1 ds_1 d\tau_2 ds_2 \dots d\tau_n ds_n$$

gives the probability of finding simultaneously electron 1 in spatial volume element $d\tau_1$ with spin between s_1 and $s_1 + ds_1$, electron 2 in spatial volume element $d\tau_2$ with spin between s_2 and $s_2 + ds_2, \dots$, electron n in spatial volume element $d\tau_n$ with spin between s_n and $s_n + ds_n$.

As we noted earlier, this expression contains too much information and we ask instead about the probability that any given element will be occupied by one electron with the other electrons anywhere. This is often called the *one-electron density function*

$$\rho_1(\mathbf{x}_1) = n \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) d\tau_2 ds_2 \dots d\tau_n ds_n \quad (19.1)$$

The \mathbf{x}_1 on the left refers to ‘point 1’ at which the density is to be evaluated rather than the coordinates of electron 1, the indistinguishability of the electrons being accounted for by the factor n .

If we want to know the probability of finding an electron of either spin in the spatial element $d\tau_1$ then we integrate over ds_1 to give the charge density discussed in Chapter 16 and measured by crystallographers:

$$P_1(\mathbf{r}_1) = n \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) ds_1 d\tau_2 ds_2 \dots d\tau_n ds_n \quad (19.2)$$

It is written either P or P_1 ; I have used P in earlier chapters. It also proves desirable to introduce probabilities for finding different configurations of any number of particles. Thus the *two-electron* (or *pair*) *density function*

$$\rho_2(\mathbf{x}_1, \mathbf{x}_2) = n(n-1) \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) d\tau_3 ds_3 \dots d\tau_n ds_n \quad (19.3)$$

determines the probability of two electrons being found simultaneously at points \mathbf{x}_1 and \mathbf{x}_2 , spins included, whilst

$$P_2(\mathbf{x}_1, \mathbf{x}_2) = n(n-1) \int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) ds_1 ds_2 d\tau_3 ds_3 \dots d\tau_n ds_n \quad (19.4)$$

determines the probability of finding them at points \mathbf{r}_1 and \mathbf{r}_2 in ordinary space, irrespective of spin.

Many common one-electron properties depend only on P_1 and since the Schrödinger equation only contains pair interactions we need not consider distributions higher than the pair functions. For a state of definite spin, as distinct from a mixed state, the one-electron density function has the form

$$\rho_1(\mathbf{x}_1) = P_1^\alpha(\mathbf{r}_1) \alpha^2(s_1) + P_1^\beta(\mathbf{r}_1) \beta^2(s_1) \quad (19.5)$$

where the P 's are spatial functions. There are no cross terms involving both α and β . In orbital models, the P 's are just sums over the squares of the occupied orbitals. The two-electron density function is also given by

$$\begin{aligned} \rho_2(\mathbf{x}_1, \mathbf{x}_2) = & P_2^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) \alpha^2(s_1) \alpha^2(s_2) + P_2^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2) \alpha^2(s_1) \beta^2(s_2) \\ & + P_2^{\beta\alpha}(\mathbf{r}_1, \mathbf{r}_2) \beta^2(s_1) \alpha^2(s_2) + P_2^{\beta\beta}(\mathbf{r}_1, \mathbf{r}_2) \beta^2(s_1) \beta^2(s_2) \end{aligned} \quad (19.6)$$

19.1.1 Fermi Correlation

The two-electron functions tell us how the two electrons are correlated. In the one-determinant HF model, we find that the P 's are related as follows:

$$\begin{aligned} P_2^{\alpha\alpha}(\mathbf{r}_1, \mathbf{r}_2) &= P_1^\alpha(\mathbf{r}_1) P_1^\alpha(\mathbf{r}_2) - X(\mathbf{r}_1, \mathbf{r}_2) \\ P_2^{\alpha\beta}(\mathbf{r}_1, \mathbf{r}_2) &= P_1^\alpha(\mathbf{r}_1) P_1^\beta(\mathbf{r}_2) \end{aligned} \quad (19.7)$$

where X is a function whose form need not concern us, except to note that it exactly cancels the first term when $\mathbf{r}_1 = \mathbf{r}_2$. From these results we get a clear picture of the electron correlation shown by standard closed-shell HF theory. The form of $P_2^{\alpha\beta}$ shows that there is no correlation between electrons of opposite spin, since the simultaneous probability is just the product of the individual ones. This is a defect of HF theory. Electrons of like spin are clearly correlated and they are never found at the same point in space, and HF theory is satisfactory here. This type of correlation arises from antisymmetry and applies to all fermions. It is usually called *Fermi correlation*.

19.2 Configuration Interaction

I have mentioned configuration interaction (CI) at various points in the text, in the dihydrogen discussion of Chapter 15 and in the CIS treatment of excited states in Chapter 17. The idea of modern CI is to start with a reference wavefunction that could be a HF-LCAO closed shell wavefunction, and systematically replace the occupied spinorbitals with virtual ones. So if A, B, \dots represent occupied orbitals and X, Y, \dots represent virtual ones we would seek to write

$$\Psi^{\text{CI}} = \Psi^{\text{HF}} + \sum_{A,X} c_A^X \Psi_A^X + \sum_{A,B,X,Y} c_{AB}^{XY} \Psi_{AB}^{XY} + \dots \quad (19.8)$$

The expansion coefficients can be found by solving the variational problem, in the usual way. This involves finding the eigenvalues and eigenvectors of the Hamiltonian matrix; usually one is only interested in a few of the electronic states, and methods are available for finding just a small number of eigenvalues and eigenvectors of very large matrices.

In a complete CI calculation we would include every possible excited configuration formed by promoting electrons from the occupied spinorbitals to the virtual ones; for a closed shell singlet state molecule with m electron pairs and n basis functions there are

$$\frac{n!(n+1)!}{m!(m+1)!(n-m)!(n-m+1)!}$$

possible Slater determinants. If we consider a HF/6-311G** wavefunction for benzene, we have $m = 21$ and $n = 144$, giving approximately 5×10^{50} possibilities. We therefore must truncate the CI expansion.

In the CIS model we just take singly excited states. If the reference wavefunction is a closed-shell HF one, the single excitations do not mix with the ground state and CIS tends to be used as a low-level method for studying excited states.

The next logical step is to include the double excitations. If this is done together with the single excitations we have the CISD (*CI singles and doubles*) model. If the double

excitations alone are included we have CID (*CI doubles*). Such truncations can lead to a problem called *size consistency* that I can illustrate by considering a very simple problem, that of two neon atoms. Table 19.1 shows a CISD/6-311G* calculation for dineon at 5000 pm separation, together with the free atom energy.

Table 19.1 Dineon calculations, 6-311G* basis set

Method	Dineon energy/ E_h at 5000 pm	Atom energy/ E_h
HF-LCAO	-257.0451061	-128.5225305
CISD	-257.4466147	-128.7283956

The HF-LCAO energy of a pair of neon atoms at large separation is exactly twice the free atom value, but this is not the case for the CISD calculation. If we have an ensemble of n particles and their energy is n times the energy of a single particle, we say that the theory *scales correctly* (or that the method is *size consistent*). Full CI calculations scale correctly, but truncated CI expansions do not.

After double excitations, quadruple excitations are next in order of importance. If singles, doubles, triples and quadruples are included the acronym becomes CISDTQ.

19.3 Coupled Cluster Method

The coupled cluster (CC) method was first used by physicists studying nuclear structure. Bartlett's (1989) review is fairly recent. The fundamental equation relates a HF wavefunction Ψ^{HF} to the best possible wavefunction Ψ by

$$\Psi = \exp(\hat{T}) \Psi^{\text{HF}} \quad (19.9)$$

The exponential operator is defined by a Taylor series expansion

$$\exp(\hat{T}) = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!} \quad (19.10)$$

and the *cluster operator* is defined as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_n \quad (19.11)$$

where n is the number of electrons in the molecule. The operators have the effect of replacing occupied spin orbitals in Ψ^{HF} with virtual spin orbitals. \hat{T}_1 performs all singly excited substitutions, \hat{T}_2 performs all doubly excited configurations and so on until all n electrons have been promoted from filled to virtual spinorbitals. The effect of the exponential operator

on the HF wavefunction is to express it as a linear combination that contains Ψ^{HF} and all possible excitations of electrons from occupied to virtual spinorbitals. We write

$$\hat{T}_1 \Psi^{\text{HF}} = \sum_{A,X} c_A^X \Psi_A^X$$

$$\hat{T}_2 \Psi^{\text{HF}} = \sum_{A,B,X,Y} c_{AB}^{XY} \Psi_{AB}^{XY}$$

and so on, and the aim of the theory is to determine the numerical coefficients. It is usual to approximate the \hat{T} operator by including only some of the terms, and it is generally accepted that the most important contribution is \hat{T}_2 . This gives the coupled cluster doubles (CCD) method. Since we take

$$\exp(\hat{T}) \approx \exp(\hat{T}_2) = 1 + \hat{T}_2 + \frac{1}{2!} \hat{T}_2^2 + \dots \quad (19.12)$$

the resulting wavefunction contains the HF wavefunction together with double quadruple, hextuple, etc. substitutions. The CCD method is size consistent.

The next step is to include the T_1 operator and so take

$$\exp(\hat{T}) \approx \exp(\hat{T}_1 + \hat{T}_2) = 1 + (\hat{T}_1 + \hat{T}_2) + \frac{1}{2!} (\hat{T}_1 + \hat{T}_2)^2 + \dots \quad (19.13)$$

which gives CCSD. Inclusion of T_3 gives CCSDT in an obvious notation.

Pople and co-workers have developed a quadratic CI method, which lies between CI and CC. This QCI method exists in size-consistent forms QCISD and QCISDT (Pople *et al.* 1987).

19.4 Møller–Plesset Perturbation Theory

The HF-LCAO model is thought to be satisfactory for treating many molecular properties near minima on the molecular potential energy surface. It cannot generally treat covalent bond breaking and making, but forms a useful starting point for the more advanced theories. The HF model averages over electron repulsions; each electron experiences an average field due to the remaining electrons plus the field due to the nuclei. In real life there are ‘instantaneous’ electron repulsions that go as

$$\frac{e^2}{4\pi\epsilon_0 r_{ij}}$$

A great advantage of the HF-LCAO model is that the equations can be solved to arbitrary precision, once a basis set has been defined. We therefore have a ‘zero-order’ problem with an exact solution, together with a perturbation. We discussed perturbation theory in Chapter 14 when dealing with the helium atom; the idea is to write the true Hamiltonian as that for a soluble, zero-order problem plus a perturbation

$$\hat{H} = \hat{H}^{(0)} + \lambda \hat{H}^{(1)}$$

The difference from Chapter 14 is that we now use a HF-LCAO wave function as the zero-order wave function and treat the difference between the true electron–electron repulsions

and the averaged ones from HF theory as the perturbation. This gives us Møller–Plesset perturbation theory. The method had been in the literature since 1934, but its potential for molecular calculations was not appreciated until the 1970s. It is of immense historical interest because it gave us the first realistic route for a treatment of small-molecule correlation energy, and you ought to read the original paper. The synopsis is reproduced below (Møller and Plesset 1934).

A Perturbation Theory is developed for treating a system of n electrons in which the Hartree Fock solution appears as the zero-order approximation. It is shown by this development that the first order correction for the energy and the charge density of the system is zero. The expression for the second order correction for the energy greatly simplifies because of the special property of the zero order solution. It is pointed out that the development of the higher order approximation involves only calculations based on a definite one-body problem.

The HF model averages over electron repulsions, and the HF pseudo-one-electron operator for each electron has the form (from Chapters 14 and 16)

$$\hat{h}^F(\mathbf{r}_i) = \hat{h}^{(1)}(\mathbf{r}_i) + \hat{J}(\mathbf{r}_i) - \frac{1}{2}\hat{K}(\mathbf{r}_i)$$

The unperturbed Hamiltonian is taken as the sum of the HF operators for each of the n electrons

$$\begin{aligned}\hat{H}^{(0)} &= \sum_{i=1}^n \hat{h}^F(\mathbf{r}_i) \\ &= \sum_{i=1}^n \left(\hat{h}^{(1)}(\mathbf{r}_i) + \hat{J}(\mathbf{r}_i) - \frac{1}{2}\hat{K}(\mathbf{r}_i) \right)\end{aligned}\quad (19.14)$$

whilst the true Hamiltonian makes no reference to averaging:

$$\hat{H} = \sum_{i=1}^n \hat{h}^{(1)}(\mathbf{r}_i) + \frac{e^2}{4\pi\epsilon_0} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{r_{ij}}$$

The first few key equations of the perturbation expansion, taken from Chapter 14 but simplified for our present discussion, are shown below. The electronic state of interest is the ground state, denoted $\Psi^{(0)}$, with energy $\varepsilon^{(0)}$. I have therefore dropped the (i) subscripts, since we are dealing with just the one electronic state. Promoting electrons from the occupied to the virtual HF spinorbitals gives the excited states, and we consider single, double, triple, etc. excitations in the usual fashion of CI:

$$\begin{aligned}\hat{H}^{(0)}\Psi^{(0)} &= \varepsilon^{(0)}\Psi^{(0)} \\ \left(\hat{H}^{(0)} - \varepsilon^{(0)}\right)\Psi^{(1)} &= \left(\varepsilon^{(1)} - \hat{H}^{(1)}\right)\Psi^{(0)} \\ \left(\hat{H}^{(0)} - \varepsilon^{(0)}\right)\Psi^{(2)} &= \varepsilon^{(2)}\Psi^{(0)} + \varepsilon^{(1)}\Psi^{(1)} - \hat{H}^{(1)}\Psi^{(1)}\end{aligned}\quad (19.15)$$

Since the zero-order Hamiltonian is a sum of HF operators, the zero-order energy is a sum of orbital energies

$$\varepsilon^{(0)} = \sum_i \varepsilon_i \quad (19.16)$$

where the sum runs over the occupied spinorbitals. The first-order energy is

$$\varepsilon^{(1)} = \int \Psi^{(0)} \hat{H}^{(1)} \Psi^{(0)} d\tau \quad (19.17)$$

and adding $\varepsilon^{(1)}$ to $\varepsilon^{(0)}$ gives the full HF energy ε^{HF} . We therefore need to progress to second-order theory in order to give any treatment of electron correlation. The levels of theory are denoted MP2, MP3, \dots , MP n , where n is the order of perturbation theory.

The second-order energy is given by

$$\varepsilon^{(2)} = \sum_j \frac{\left| \int \Psi^{(0)} \hat{H}^{(1)} \Psi_j d\tau \right|^2}{\varepsilon^{(0)} - \varepsilon_j} \quad (19.18)$$

and the first-order correction to the wavefunction is

$$\Psi^{(1)} = \sum_j \frac{\int \Psi^{(0)} \hat{H}^{(1)} \Psi_j d\tau}{\varepsilon^{(0)} - \varepsilon_j} \Psi_j \quad (19.19)$$

where the sum runs over all excited states, written Ψ_j , energies ε_j . The ground state $\Psi^{(0)}$ is a HF wavefunction and so the integral vanishes for all singly excited states because of the Brillouin theorem. It is also zero when the excited state differs from $\Psi^{(0)}$ by more than two spin orbitals, by the Slater–Condon–Shortley rules. Hence we only need consider doubly excited states in order to find the second order energy. The $\varepsilon^{(2)}$ numerator is nonnegative since it is the square of a modulus. The denominator is negative because $\Psi^{(0)}$ refers to the ground state and Ψ_j to excited states. The second-order energy correction is therefore always negative or zero. Higher orders of perturbation theory may give corrections of either sign.

The MP1 energy is therefore identical to the HF energy and MP2 is the simplest practicable perturbation treatment for electron correlation. It includes only the effect of double excitations.

By similar considerations, the third, fourth and higher orders of perturbation theory can be determined. The terms rapidly become algebraically complicated and the higher orders are increasingly costly to apply. For n basis functions, HF-LCAO theory scales as n^4 , MP2 as n^5 , MP3 as n^6 and MP4 as n^7 . These should be seen as theoretical upper bounds, since sophisticated use of symmetry and integral cut-offs mean that practical calculations need considerably less resource.

A great simplifying feature of MP2 is that a full four-index transformation is not necessary, all we have to do is semitransform the two-electron integrals and this leads to an immense saving in time compared to conventional CI treatments. MP3 also includes only double excitations, whilst MP4 includes a description of single excitations together with some triples and quadruples. The triple contributions in MP4 are the most expensive. In the MP4(SDQ) variation we just include the least expensive singles, doubles and quadruple excitations

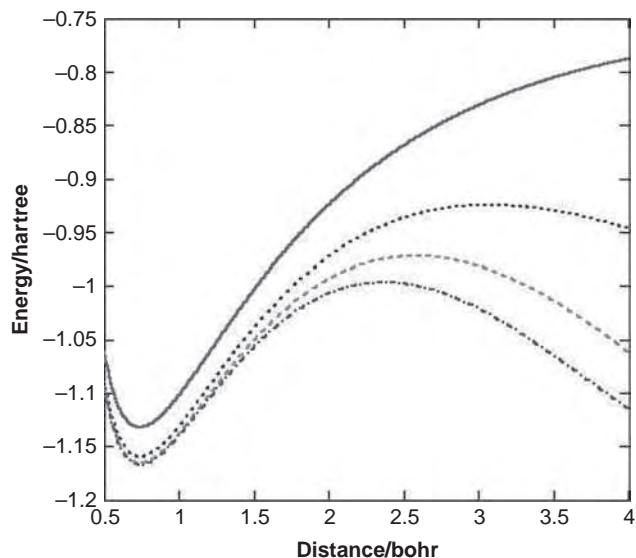
Analytical energy gradients have been determined for MP2, MP3 and MP4 making for very effective geometry searching. I can illustrate some points of interest by considering two examples. First is the geometry optimization of ethane, shown in Table 19.2. I ran all calculations from the same starting point on the molecular potential energy surface. The HF/STO-3G calculation would these days be regarded as woefully inadequate.

Table 19.2 Geometry optimization for ethene

	HF/STO-3G	HF/6-311G**	MP2/6-311G**	MP2/cc-pVTZ
$\varepsilon_{\text{HF}}/E_h$	-77.0789547	-78.0547244	-78.0539661	-78.0638977
$\varepsilon_{\text{MP2}}/E_h$			-78.3442916	-78.3993063
$\varepsilon_{\text{corr}}/E_h$			0.2903255	0.3354086
R_{CC}/pm	130.6	131.7	133.7	133.2
R_{CH}/pm	108.2	107.6	108.5	108.0
HCH/ $^\circ$	115.7	116.7	117.2	117.3
Cpu time/s	28	41	70	881

There are two sources of error in the calculations: the choice of a finite basis set, and truncation of the MP_n series at MP2. According to the literature, the basis set error is the larger of the two. One has to be careful to compare like with like and the results of the calculations should be compared with geometric parameters at the bottom of the potential energy well. Experimental values very often refer to the lowest vibrational energy level. The estimate of the correlation energy depends markedly on the quality of the basis set. Electron correlation studies demand basis sets that are capable of high accuracy, and a number of suitable basis sets are available in the literature. One popular family is the *correlation consistent* basis sets of Dunning, which go by acronyms such as cc-pV6Z. This contains seven s-type, six p-type, four d-type, two g-type and one h-type primitives.

We usually find that MP2 calculations give a small but significant correction to the corresponding HF-LCAO geometries. It is probably more worthwhile using MP_n models to investigate effects that depend crucially on electron correlation; bond breaking and bond making phenomena spring to mind. Figure 19.1 shows representative potential energy curves for dihydrogen, following the discussion of Chapter 15.

**Figure 19.1** Dihydrogen electron correlation

A 6-311G** basis set was used. The top curve is the HF-LCAO calculation, and shows the usual bad dissociation limit. The remaining three curves are (in descending order) MP2, MP3 and MP4(SDQ). The first point to notice is that MP n calculations are size consistent. The second point is that none of the curves appear to be approaching the correct dissociation limit of $-1E_h$. It is reported in the literature that MP n calculations do not work particularly well for geometries far from equilibrium. The MP n series has not been proved to converge to the true energy, but calculations using very high orders (MP48) have suggested that convergence will eventually occur.

Finally, MP n calculations need a HF-LCAO wavefunction to act as the zero-order approximation. It follows that MP n can only be used for those molecules and electronic states where it is possible to find such a wavefunction. This rules out most excited states.

19.5 Multiconfiguration SCF

One problem with traditional CI calculations is that excitations are to the virtual orbitals; the occupied orbitals are usually determined variationally by the HF-LCAO procedure whilst the virtual orbitals give poor representations of excited states. One solution to the problem is to write the ground-state wavefunction as a linear combination of Slater determinants and then use the variation procedure to optimize both the linear expansion coefficients and the HF-LCAO coefficients. That is, we write

$$\Psi = \Psi^{\text{HF}} + \sum c_j \Psi_j$$

where Ψ^{HF} is the Hartree–Fock wavefunction and Ψ_j an excited state, in the usual sense of CI calculations. In certain cases the ratio of some of the c_j will be fixed by symmetry. The MCSCF model is regarded as highly specialized, with a small number of adherents. A literature search revealed an average of 50 papers per annum in primary journals over the last 10 years.

References

- Bartlett, R.J. (1989) *J. Phys. Chem.*, **93**, 1697.
 Møller, Chr. and Plesset, M.S. (1934) *Phys. Rev.*, **46**, 618.
 Pople, J.A., Head-Gordon, M. and Raghavachari, K. (1987) *J. Chem. Phys.*, **87**, 5968.

20

Density Functional Theory and the Kohn–Sham LCAO Equations

The 1998 Nobel Prize for Chemistry was awarded jointly to John A. Pople and Walter Kohn, and the texts of their addresses on the occasion of the awards are reported in Pople (1999) and Kohn (1999). Pople’s name should be familiar to you if you have assiduously studied this text. Kohn opened his address with the following remarks.

The citation for my share of the 1998 Nobel Prize in Chemistry refers to the ‘Development of Density Functional Theory’. The initial work on Density Functional Theory (DFT) was reported in two publications: the first with Pierre Hohenberg (P Hohenberg and W Kohn, Phys Rev 136 (1964) B864) and the next with Lu J Sham (W Kohn and L J Sham, Phys Rev 140 (1965) A1133). This was almost 40 years after E Schrödinger (1926) published his first epoch-making paper marking the beginning of wave mechanics. The Thomas–Fermi theory, the most rudimentary form of DFT, was put forward shortly afterwards (E Fermi, Attl Accad Naz Lincei, Cl Sci Fis Mat Nat Rend 6 (1927) 602, L H Thomas, Proc Camb Phil Soc 23 (1927) 542) and received only modest attention.

It is worth spending a little time on the historical aspects before launching into modern DFT, and the *exchange potential* plays a key role in our discussion. If we consider a closed shell system with electron configuration

$$(\psi_1)^2 (\psi_2)^2 \dots (\psi_M)^2$$

then in Hartree’s original theory we would write the total wavefunction as a simple orbital product for which the electronic energy is

$$\begin{aligned} \varepsilon_{\text{el}} = & 2 \sum_{R=1}^M \int \psi_R(\mathbf{r}_1) \hat{h}(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau_1 \\ & + \sum_{R=1}^M \sum_{S=1}^M 2 \int \int \psi_R(\mathbf{r}_1) \psi_R(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_S(\mathbf{r}_2) \psi_S(\mathbf{r}_2) d\tau_1 d\tau_2 \end{aligned} \quad (20.1)$$

In Hartree–Fock (HF) theory we write the wavefunction as a Slater determinant and this gives an additional exchange term in the electronic energy

$$\begin{aligned} \varepsilon_{\text{el}} = & 2 \sum_{R=1}^M \int \psi_R(\mathbf{r}_1) \hat{h}(\mathbf{r}_1) \psi_R(\mathbf{r}_1) d\tau_1 \\ & + \sum_{R=1}^M \sum_{S=1}^M 2 \int \int \psi_R(\mathbf{r}_1) \psi_R(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_S(\mathbf{r}_2) \psi_S(\mathbf{r}_2) d\tau_1 d\tau_2 \\ & - \sum_{R=1}^M \sum_{S=1}^M \int \int \psi_R(\mathbf{r}_1) \psi_S(\mathbf{r}_1) \hat{g}(\mathbf{r}_1, \mathbf{r}_2) \psi_R(\mathbf{r}_2) \psi_S(\mathbf{r}_2) d\tau_1 d\tau_2 \end{aligned} \quad (20.2)$$

These expressions are general to Hartree and HF theory in that they do not depend on the LCAO approximation. In fact the Hartrees produced highly accurate numerical tables of radial functions with no LCAO coefficients in sight. The atomic problem is quite different from the molecular one because of the high symmetry of atoms and the HF limit can be easily reached by numerical integration. It is not necessary to invoke the LCAO approach. There has been a resurgence of interest in atomic HF calculations because astrophysicists want to study highly ionized species in the interstellar medium. They look to theory for their energy-level data because the species are hard to prepare and study in the laboratory.

I have also described the LCAO versions, where the energy expressions are written in terms of the ‘charges and bond orders’ matrix \mathbf{P} together with the matrices $\mathbf{h}^{(1)}$, \mathbf{J} and \mathbf{K} :

$$\varepsilon_{\text{el,H}} = \text{Tr}(\mathbf{P}\mathbf{h}_1) + \frac{1}{2} \text{Tr}(\mathbf{P}\mathbf{J}) \quad (20.3)$$

for the Hartree model and

$$\varepsilon_{\text{el,HF}} = \text{Tr}(\mathbf{P}\mathbf{h}_1) + \frac{1}{2} \text{Tr}(\mathbf{P}\mathbf{J}) - \frac{1}{4} \text{Tr}(\mathbf{P}\mathbf{K}) \quad (20.4)$$

for the HF version. The \mathbf{J} and \mathbf{K} matrices depend on \mathbf{P} in a complicated way, but the Hartree and HF electronic energies are completely determined from knowledge of the electron density \mathbf{P} .

There was initially a great deal of confusion about the extra term in the HF energy compared to the Hartree energy. The name ‘exchange term’ was coined and some authors tried to describe it in terms of a mysterious force called the ‘exchange potential’.

In the meantime, solid-state physics had been developing along a quite different direction. Wigner and Seitz (1934a,b) suggested what is now called the ‘cellular method’ for handling the problem of computing crystal orbitals. These orbitals have energies that form continuous bands and such models are therefore known as *energy band theories*.

20.1 Pauli and Thomas–Fermi Models

There is nothing sinister about the exchange term; it arises because of the fermion nature of electrons. Nevertheless, it caused considerable confusion among early workers in the field of molecular structure theory and much effort was spent in finding effective model potentials that could mimic electron exchange. I discussed the earliest models of metallic

conductors in Chapter 12. The Pauli model is the simplest one to take account of the quantum mechanical nature of the electrons: the electrons exist in a three-dimensional infinite potential well, the wavefunction obeys the Pauli principle and at 0 K the pairs of electrons occupy all orbitals having energy less than or equal to energy ε_F (which of course defines the Fermi energy). The number N of conduction electrons can be related to ε_F and we find

$$\varepsilon_F = \frac{h^2}{8m_e} \left(\frac{3N}{\pi L^3} \right)^{2/3} \quad (20.5)$$

Now N/L^3 is the number density of conduction electrons and so Pauli's model gives a simple relationship between the Fermi energy and the number density of electrons. Physicists very often use the symbol n for the number density; it can vary at points in space and so we write $n(\mathbf{r})$ but in this simple model, the number density is constant throughout the dimensions of the metallic box:

$$\begin{aligned} n &= \frac{N}{L^3} \\ &= \frac{8\pi}{3h^2} (2m_e)^{3/2} (\varepsilon_F)^{3/2} \end{aligned} \quad (20.6)$$

We now switch on an *external potential* $U_{\text{ext}}(\mathbf{r})$ that is slowly varying over the dimensions of the metallic conductor, making the number density inhomogeneous. This could be (for example) due to the set of metallic cations in such a conductor, or due to the nuclei in a molecule. A little analysis suggests that the number density at position \mathbf{r} should be written

$$n(\mathbf{r}) = \frac{8\pi}{3h^2} (2m_e)^{3/2} (\varepsilon_F - U_{\text{ext}}(\mathbf{r}))^{3/2} \quad (20.7)$$

This *Thomas–Fermi relation* therefore relates the number density at points in space to the potential at points in space. The number density of electrons at a point in space is just the charge density $P_1(\mathbf{r})$ discussed in several previous chapters and so we can write in more familiar chemical language

$$P_1(\mathbf{r}) = \frac{8\pi}{3h^2} (2m_e)^{3/2} (\varepsilon_F - U_{\text{ext}}(\mathbf{r}))^{3/2} \quad (20.8)$$

Thomas and Fermi suggested that such a statistical treatment would be appropriate for molecular systems where the number of electrons is 'large' and in the case of a molecule we formally identify the external potential as the electrostatic potential generated by the nuclei. The Thomas–Fermi approach replaces the problem of calculating an N -electron wavefunction by that of calculating the electron density in three-dimensional position space.

Dirac (1930) studied the effects of exchange interactions on the Thomas–Fermi model, and discovered that these could be modelled by an extra term

$$U_x(\mathbf{r}) = C (P_1(\mathbf{r}))^{1/3} \quad (20.9)$$

where C is a constant. R. Gáspar is also credited with this result, which Slater (1951) rediscovered, but with a slightly different numerical coefficient of $\frac{2}{3}C$. The disagreement

between Dirac's and Slater's numerical coefficient seems to have been first resolved by Gáspár (1954), and authors began to write the *local exchange potential* as

$$U_{X\alpha}(\mathbf{r}) = C\alpha (P_1(\mathbf{r}))^{1/3} \quad (20.10)$$

where α could take values between 2/3 and 1. This is the so-called *X α model*. The use of the symbol α here is not to be confused with the use of the same symbol for a spin eigenfunction.

Slater then had the ingenious idea of writing the atomic HF eigenvalue equation

$$\hat{h}^F(\mathbf{r})\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r})$$

$$\left(\hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) - \frac{1}{2}\hat{K}(\mathbf{r})\right)\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r})$$

as

$$\left(\hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) + \hat{U}_{X\alpha}(\mathbf{r})\right)\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r}) \quad (20.11)$$

and such calculations are usually referred to as *atomic X α -HF*. The resulting X α orbitals differ from conventional HF orbitals in one major way, namely that Koopmans' theorem is no longer valid for every orbital and so the orbital energies cannot generally be used to estimate ionization energies. Koopmans' theorem now applies only to the highest occupied orbital. A key difference between standard HF theory and density functional calculations is the way we conceive the *occupation number* ν of each orbital. In molecular HF theory, the occupation number is 2, 1 or 0 depending on whether a given spatial orbital is fully occupied by two electrons (one of either spin), singly occupied or a virtual orbital. For a system comprising very many electrons we focus on the statistical occupation of each orbital and the occupation number becomes a continuous variable having a value between 0 and 2. The relationship between the electronic energy ε_{el} and the occupation number of orbital i is

$$\frac{\partial \varepsilon_{\text{el}}}{\partial \nu_i} = \varepsilon_i \quad (20.12)$$

so that the ionization energy from a particular orbital ψ_i is given by

$$\varepsilon_{\text{el}}(\nu_i = 0) - \varepsilon_{\text{el}}(\nu_i = 1) = \int_1^0 \varepsilon_i \, d\nu_i$$

Notice, however, that the energy is determined by the charge density $P(\mathbf{r})$. When the calculations are performed, the resulting orbitals closely resemble those from standard HF theory and people use them in much the same way.

At the time, such calculations of ionization energies had an edge over the traditional HF variety because they take account of the electron relaxation that can accompany ionization.

20.2 Hohenberg–Kohn Theorems

The Thomas–Fermi and X α approaches were constructed as approximations to the quantum mechanical problem of calculating the electronic properties of a system of interest. The density functional theory of Hohenberg and Kohn, to be discussed, is in principle an exact

theory. The first keynote paper contains two fundamental theorems. The first theorem is what mathematicians call an *existence theorem*; it proves that the ground-state energy of any electronic system is determined by a *functional* of the electron density. The theorem means that in principle we only need to know the electron density in three-dimensional space and not the full wavefunction in order to calculate any ground-state property including the energy.

The term *functional* is mathematical jargon; a function is a mapping from one set of numbers to another (Figure 20.1) whilst a functional is a mapping from a set of functions to a set of numbers.

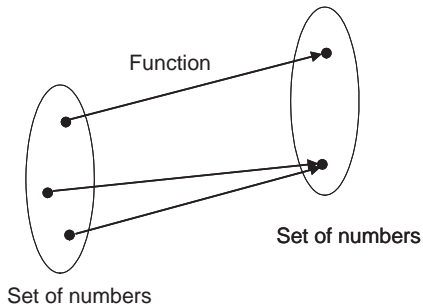


Figure 20.1 A function

The proof of the theorem is quite simple, but will not be reproduced here. The paper is so important that you should see at least the abstract (Hohenberg and Kohn 1964). The authors use square brackets [. . .] to denote functionals.

This paper deals with the ground state of an interacting electron gas in an external potential $v(\mathbf{r})$. It is proved that there exists a universal functional of the density, $F[n(\mathbf{r})]$, independent of $v(\mathbf{r})$, such that the expression $E \equiv \int v(\mathbf{r})n(\mathbf{r})d\mathbf{r} + F[n(\mathbf{r})]$ has as its minimum value the correct ground state energy associated with $v(\mathbf{r})$. The functional $F[n(\mathbf{r})]$ is then discussed for two situations: (1) $n(\mathbf{r}) = n_0 + n'(\mathbf{r})$, $n'/n_0 \ll 1$, and (2) $n(\mathbf{r}) = \varphi(\mathbf{r}/r_0)$ with φ arbitrary and $r_0 \rightarrow \infty$. In both cases F can be expressed entirely in terms of the correlation energy and linear and higher order electronic polarizabilities of a uniform electron gas. This approach also sheds some light on generalized Thomas–Fermi methods and their limitations. Some new extensions of these methods are presented.

The second theorem gives a variation principle for the density functionals; it states (in chemical language) that

$$\varepsilon_{\text{el}} [P(\mathbf{r})] \geq \varepsilon_{\text{el}} [P_0(\mathbf{r})] \quad (20.13)$$

where P_0 is the true density for the system and P any other density obeying

$$\int P(\mathbf{r}) d\tau = \int P_0(\mathbf{r}) d\tau = N \quad (20.14)$$

where N is the number of electrons. The difficulty is that the Hohenberg–Kohn theorems give us no clue as to the nature of the density functional, nor how to set about finding it.

20.3 Kohn–Sham (KS-LCAO) Equations

The Kohn–Sham paper gives us a practical solution to the problem, based on HF-LCAO theory; again, it is such an important paper that you should have sight of the abstract that in any case is self-explanatory (Kohn and Sham 1965).

From a theory of Hohenberg and Kohn, approximation methods for treating an inhomogeneous system of interacting electrons are developed. These methods are exact for systems of slowly varying or high density. For the ground state, they lead to self-consistent equations analogous to the Hartree and Hartree–Fock equations, respectively. In these equations the exchange and correlation portions of the chemical potential of a uniform electron gas appears as additional effective potentials. (The exchange portion of our effective potential differs from the due to Slater by a factor of 2/3.) Electronic systems at finite temperatures and in magnetic fields are also treated by similar methods. An appendix deals with a further correction for systems with short-wavelength density oscillations.

The Kohn–Sham equations are modifications of the standard HF equations and we write

$$\left(\hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) + \hat{U}_{\text{XC}}(\mathbf{r})\right) \psi(\mathbf{r}) = \varepsilon \psi(\mathbf{r}) \quad (20.15)$$

where $U_{\text{XC}}(\mathbf{r})$ is the local exchange–correlation term that accounts for the exchange phenomenon and the dynamic correlation in the motions of the individual electrons. We speak of the KS-LCAO procedure. It is usual to split $U_{\text{XC}}(\mathbf{r})$ into an exchange term and a correlation term and treat each separately.

$$\left(\hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) + \hat{U}_{\text{X}}(\mathbf{r}) + \hat{U}_{\text{C}}(\mathbf{r})\right) \psi(\mathbf{r}) = \varepsilon \psi(\mathbf{r}) \quad (20.16)$$

The electronic energy is usually written in DFT applications as

$$\varepsilon_{\text{el}}[P] = \varepsilon_{\text{I}}[P] + \varepsilon_{\text{J}}[P] + \varepsilon_{\text{X}}[P] + \varepsilon_{\text{C}}[P] \quad (20.17)$$

where the square brackets denote a functional of the one-electron density $P(\mathbf{r})$. The first term on the right-hand side gives the one-electron energy, the second term is the Coulomb contribution, the third term the exchange and the fourth term gives the correlation energy. We proceed along the usual HF-LCAO route; we choose a basis set and then all that is needed in principle is knowledge of the functional forms of $U_{\text{X}}(\mathbf{r})$ and $U_{\text{C}}(\mathbf{r})$. It is then a simple matter in principle to modify an existing standard HF-LCAO computer code to include these additional terms. Either component can be of two distinct types: *local functionals* that depend only on the electron density at a point in space and *gradient-corrected functionals* that depend both on the electron density and its gradient at that point.

In their most general form, the exchange and correlation functionals will depend on the density of the α and of the β electrons, and the gradients of these densities. Each energy term ε_{C} and ε_{X} will therefore be given by a volume integral

$$\begin{aligned} \varepsilon_{\text{X}} &= \int f_{\text{X}}(P^{\alpha}, P^{\beta}, \text{grad } P^{\alpha}, \text{grad } P^{\beta}) d\tau \\ \varepsilon_{\text{C}} &= \int f_{\text{C}}(P^{\alpha}, P^{\beta}, \text{grad } P^{\alpha}, \text{grad } P^{\beta}) d\tau \end{aligned} \quad (20.18)$$

where f is an energy density.

20.4 Numerical Integration (Quadrature)

Given the functionals and the electron density, we have to integrate over the space of the molecule to obtain the energy. These extra integrals cannot usually be done analytically and so a numerical integration has to be done each KS-LCAO cycle. What we do is to replace the integral by a sum over quadrature points, e.g.

$$\varepsilon_{X/C} = \sum_A \sum_i w_{Ai} f(P_1^\alpha, P_1^\beta, \text{grad}(P_1^\alpha), \text{grad}(P_1^\alpha); \mathbf{r}_{Ai}) \quad (20.19)$$

where the first summation is over the atoms and the second is over the numerical quadrature grid points. The w_{Ai} are the quadrature weights and the grid points are given by the sum of the position of nucleus A and a suitable one-centre grid:

$$\mathbf{r}_{Ai} = \mathbf{R}_A + \mathbf{r}_i$$

The keynote paper for numerical integration is that due to Becke (1988a), and once again the paper is so important that you should see the abstract.

We propose a simple scheme for decomposition of molecular functions into single-center components. The problem of three-dimensional integration in molecular systems thus reduces to a sum of one-center, atomic-like integrations, which are treated using standard numerical techniques in spherical polar coordinates. The resulting method is tested on representative diatomic and polyatomic systems for which we obtain five- or six-figure accuracy using a few thousand integration points per atom.

Becke's scheme rigorously separates the molecular integral into atomic contributions that may be treated by standard single-centre techniques. Numerical integration grids are usually specified as a specific number of radial shells around each atom, each of which contains a set number of integration points. For example, Gaussian 03 uses a grid designated (75,302) which has 75 radial shells per atom, each containing 302 points giving a total of 22 650 integration points.

Over the years, a large number of exchange and correlation functionals have been proposed.

20.5 Practical Details

There are two versions of DFT, one for closed shell systems analogous to the Roothaan closed shell treatment, and one for the open shell case analogous to the UHF technique and we implement the Kohn–Sham LCAO (KS-LCAO) equations by including the relevant exchange–correlation term(s) instead of the traditional HF exchange term in the HF-LCAO equations. That is, for the UHF case we write

$$\left(\hat{h}^{(1)}(\mathbf{r}) + \hat{J}(\mathbf{r}) + \hat{U}_X^\alpha(\mathbf{r}) + \hat{U}_C^\alpha(\mathbf{r}) \right) \psi^\alpha(\mathbf{r}) = \varepsilon^\alpha \psi^\alpha(\mathbf{r})$$

for the α -spin electrons and a similar expression for the β -spin electrons. The elements of the α -spin exchange–correlation matrices are related to the energy density f and the basis functions χ by

$$(U_{X/C}^\alpha)_{i,j} = \int \left\{ \frac{\partial f}{\partial P^\alpha} \chi_i \chi_j + \left[2 \frac{\partial f}{\partial \gamma^{\alpha\alpha}} \text{grad } P^\alpha + \frac{\partial f}{\partial \gamma^{\alpha\beta}} \text{grad } P^\beta \right] \cdot \text{grad } (\chi_i \chi_j) \right\} d\tau \quad (20.20)$$

where I have written

$$\gamma^{\alpha\beta} = \text{grad } (P^\alpha) \text{ grad } (P^\beta)$$

The β -spin electrons have a similar matrix.

So for example, the Slater $X\alpha$ exchange density is

$$f(\mathbf{r}) = -\frac{9}{4} \alpha^* \left(\frac{3}{4\pi} \right)^{1/3} \left((P^\alpha(\mathbf{r}))^{4/3} + (P^\beta(\mathbf{r}))^{4/3} \right) \quad (20.21)$$

where α^* is the $X\alpha$ parameter, not to be confused with the spin variable. The derivatives are simple:

$$\frac{\partial f}{\partial P_1^\alpha} = -3\alpha^* \left(\frac{3}{4\pi} \right)^{1/3} (P^\alpha)^{1/3} \quad (20.22)$$

with a similar expression for the β -spin term.

By itself, the Slater $X\alpha$ local term is not adequate for describing molecular systems. Becke (1988b) formulated the following gradient-corrected exchange functional:

$$f = (P^\alpha)^{4/3} g(x^\alpha) + (P^\beta)^{4/3} g(x^\beta)$$

where

$$g(x) = -\frac{3}{2} \left(\frac{3}{4\pi} \right)^{1/3} - \frac{bx^2}{1 + 6bx \sinh^{-1} x} \quad (20.23)$$

$$x^\alpha = \frac{\gamma_{\alpha\alpha}}{(P^\alpha)^{4/3}} \quad \text{and} \quad x^\beta = \frac{\gamma_{\beta\beta}}{(P^\beta)^{4/3}}$$

The derivatives are again straightforward.

In a similar way there are local and gradient-corrected correlation functionals. One of the oldest correlation functionals is that due to Lee, Yang and Parr:

$$f = -\frac{4a}{1+d(P)^{-1/3}} \frac{P^\alpha P^\beta}{P} - 2^{11/3} \frac{3}{10} (3\pi^2)^{2/3} ab\omega(P) P^\alpha P^\beta \left((P^\alpha)^{8/3} + (P^\beta)^{8/3} \right) + \frac{\partial f}{\partial \gamma_{\alpha\alpha}} \gamma_{\alpha\alpha} + \frac{\partial f}{\partial \gamma_{\alpha\beta}} \gamma_{\alpha\beta} + \frac{\partial f}{\partial \gamma_{\beta\beta}} \gamma_{\beta\beta} \quad (20.24)$$

where

$$\omega(P) = \frac{\exp(-cP^{-1/3})}{1+dP^{-1/3}}$$

where a , b , c and d are constants.

A modern reference that contains a wealth of detail for implementing DFT calculations is that of Johnson *et al.* (1993).

20.6 Custom and Hybrid Functionals

Hybrid functionals are those that contain mixtures of the HF exchange with DFT exchange–correlation, whilst *ab Initio* codes usually give the user a choice of constructing their own linear combination.

20.7 An Example

We worked through a typical HF-LCAO calculation on L-phenylalanine using the 6-311G** basis set in Chapter 16. The following output relates to a corresponding KS-LCAO single point calculation. The first part of the .gjf is:

```
# b3lyp/6-311G** Pop=full
L-phenylalanine single point
0 1
N 0.000 0.000 0.000
H 0.000 0.000 1.010
C 1.366 0.000 -0.483
. . .
```

Following current fashion, I chose the Becke three-parameter exchange and the Lee–Yang–Parr correlation functionals (B3LYP). The usual iterative procedure is followed and the output is all but identical to that shown in Chapters 16 and 17; a major difference from HF-LCAO theory is that the numerical integration has to be done at the end of each cycle, although this is not shown explicitly.

```
The electronic state of the initial guess is 1-A.
Warning! Cutoffs for single-point calculations used.
Requested convergence on RMS density matrix=1.00D-04 within 128 cycles.
Requested convergence on MAX density matrix=1.00D-02.
Requested convergence on energy=5.00D-05.
No special actions if energy rises.
SCF Done: E(RB+HF-LYP) = -554.845359196 A.U. after 8 cycles
Conv = 0.4690D-05 -V/T = 2.0027
S**2 = 0.0000
```

The energy reported is the equivalent of the usual HF-LCAO energy.

Next come the KS-LCAO orbital energies:

```
(A) (A) (A) (A) (A) (A) (A) (A) (A) (A)
The electronic state is 1-A.
Alpha occ. eigenvalues - -19.19197 -19.12622 -14.29542 -10.31653 -10.22146
Alpha occ. eigenvalues - -10.17833 -10.17422 -10.16835 -10.16719 -10.16615
Alpha occ. eigenvalues - -10.16444 -10.13667 -1.16057 -1.03817 -0.90705
Alpha occ. eigenvalues - -0.85012 -0.78882 -0.74037 -0.73113 -0.68902
Alpha occ. eigenvalues - -0.62748 -0.60433 -0.57744 -0.55764 -0.54022
Alpha occ. eigenvalues - -0.50460 -0.49534 -0.47506 -0.46924 -0.45013
```

```
Alpha occ. eigenvalues - -0.43607 -0.42978 -0.40963 -0.40051 -0.38414
Alpha occ. eigenvalues - -0.37210 -0.35197 -0.34396 -0.33353 -0.30417
Alpha occ. eigenvalues - -0.27739 -0.25593 -0.24403 -0.23759
```

and LCAO coefficients (not shown here). The LCAO orbitals tend to be used just like ordinary HF-LCAO orbitals, and many authors make use of the standard Mulliken population analysis indices:

Mulliken atomic charges:

```
1
  1 N -0.409568
  2 H  0.190855
  3 C -0.174658
etc. etc
 22 C -0.084815
 23 H  0.085758
```

Sum of Mulliken charges= 0.00000

Comparison with Chapter 17 shows that the Mulliken indices are somewhat different; there is nothing particularly sinister in this, we simply have to be careful to compare like with like when using such subjective quantities. Finally the properties and timing:

```
Electronic spatial extent (au): <R**2>= 2207.1383
Charge= 0.0000 electrons
Dipole moment (field-independent basis, Debye):
  X= -0.0080 Y= 1.3326 Z= 0.0814 Tot= 1.3351
Quadrupole moment (field-independent basis, Debye-Ang):
  XX= -73.9662 YY= -64.8347 ZZ= -74.2992
  XY= 1.0222 XZ= -3.5488 YZ= -3.4544
Traceless Quadrupole moment (field-independent basis, Debye-Ang):
  XX= -2.9328 YY= 6.1987 ZZ= -3.2658
  XY= 1.0222 XZ= -3.5488 YZ= -3.4544
Octapole moment (field-independent basis, Debye-Ang**2):
  XXX= 6.5417 YYY= 31.5703 ZZZ= -0.5142 XYY= 1.2511
  XXY= -0.4178 XXZ= -5.5372 XZZ= 12.1891 YZZ= 2.3353
  YYZ= 9.3170 XYZ= -6.7008
Hexadecapole moment (field-independent basis, Debye-Ang**3):
  XXXX= -2011.8342 YYYY= -566.5190 ZZZZ= -187.2276 XXXY= -15.7192
  XXXZ= -56.5913 YYYYX= 45.6503 YYYZ= -4.1547 ZZZX= -0.4688
  ZZZY= 0.4758 XXYY= -464.1778 XXZZ= -416.6522 YYZZ= -138.6556
  XXYZ= -47.1416 YYXZ= 18.4000 ZZXY= 3.7164
N-N= 6.958408871868D+02 E-N=-2.687191007826D+03 KE= 5.533590136629D+02

THERE'S NOTHING WRONG WITH FOLLOWING A TRAIN OF THOUGHT...
AS LONG AS IT ULTIMATELY PULLS INTO THE STATION.
Job cpu time: 0 days 0 hours 12 minutes 20.0 seconds.
File lengths (MBytes): RWF= 52 Int= 0 D2E= 0 Chk= 10 Scr= 1
```

The electric moments are similar to the HF values, and the increased cpu time comes about because of the numerical integrations.

References

- Becke, A.D. (1988a) *J. Chem. Phys.*, **88**, 2547.
Becke, A.D. (1988b) *Phys. Rev. A*, **38**, 3098.
Dirac, P.A.M. (1930) *Proc. Cambridge Phil. Soc.*, **26**, 376.
Gáspár, R. (1954) *Acta Phys. Acad. Hung.*, **3**, 263.
Hohenberg, P. and Kohn, W. (1964) *Phys. Rev.*, **136**, B864.
Johnson, B.G., Gill, P.M. and Pople, J.A. (1993) *J. Chem. Phys.*, **98**, 5612.
Kohn, W. (1999) *Rev. Modern Phys.*, **71**, 1267.
Kohn, W. and Sham, L.J. (1965) *Phys. Rev.*, **140**, A1133.
Pople, J.A. (1999) *Rev. Modern Phys.*, **71**, 1253.
Slater, J.C. (1951) *Phys. Rev.*, **81**, 385.
Wigner, E. and Seitz, E. (1934a) *Phys. Rev.*, **43**, 804.
Wigner, E. and Seitz, E. (1934b) *Phys. Rev.*, **46**, 509.

21

Accurate Thermodynamic Properties; the G_n Models

A major objective of quantum chemical modelling is the prediction of molecular structure and energy. The enthalpies of formation of many molecules have been determined very accurately by precise experimental techniques such as calorimetry and spectroscopy, but there is a wide range of species, for example unstable transients, for which the uncertainty is large. Molecular ionization energies, electron affinities and proton affinities have also been well studied experimentally. Realistic calculations of thermodynamic data have to be capable of giving ‘chemical’ accuracy and the target is usually set as $\pm 2 \text{ kcal mol}^{-1}$ for energy quantities.

It has been recognized for many years that bond dissociation energies are poorly predicted by standard *ab Initio* HF-LCAO calculations. The development of MP_n perturbation theory marked a turning point in the treatment of molecular correlation, and made the treatment of correlation energy at last possible for molecules of moderate size (but at a high cost in computer resource). The MP_n methodology is usually implemented up to MP_4 level, but the convergence of the perturbation series leaves something to be desired and for this reason a further quadratic CI correction is usually applied.

The effect of polarization functions is also important, and even f orbitals make a contribution to the total energies of first-row atoms.

21.1 G1 Theory

The *Gaussian-1 (G1) method* was introduced by Pople *et al.* (1989), in order to systematically investigate the shortcomings in the levels of theory known at that time. Here is the abstract of their paper:

A general procedure is developed for the computation of the total energies of molecules at their equilibrium geometries. Ab initio molecular orbital theory is used to calculate electronic energies by a composite method, utilizing large basis sets (including diffuse-sp, double-d and f-polarization functions) and treating electron correlation by Møller–Plesset perturbation theory and by quadratic configuration interaction. The theory is also used to compute zero-point vibrational energy corrections. Total atomization energies for a set of 31 molecules are found to agree with experimental thermochemical data to an accuracy greater than 2 kcal mol⁻¹ in most cases. Similar agreement is achieved for ionization energies, electron and proton affinities. Residual errors are assessed for the total energies of neutral atoms.

The original paper treated first-row atoms and molecules, but the extension to second-row species followed quickly. I will illustrate the calculation with ethene; here is a suitable .gif:

```
# G1
Ethene G1 calculation
0 1
C    -0.42918455  1.60944204  0.00000000
H     0.10397920  0.68173712  0.00000000
H    -1.49918455  1.60944204  0.00000000
C     0.24608976  2.78441933  0.00000000
H    -0.28707399  3.71212425  0.00000000
H     1.31608976  2.78441933  0.00000000
```

There are eight distinct stages in a G1 calculation; in the Gaussian 03 suite the steps are followed automatically once the G1 keyword is specified.

1. An initial structure is obtained at the HF/6-31G* level of theory. A frequency check is also run as a matter of course, and the harmonic frequencies are stored for later use.

```
1|1|UNPC-UNK|FOpt|RHF|6-31G(d)|C2H4|PCUSER|07-Jan-2008|0|# G1|Ethene
G1 calculation||0,1|C,-0.3292236947,-0.5703527276,0.|H,0.1799049547,
-1.5185022354,0.|H,-1.4048899659,-0.6041905569,0.|C,0.3292236947,0.5703
527276,0.|H,-0.1799049547,1.5185022354,0.|H,1.4048899659,0.6041905569,0.||
Version=x86-Win32-G03RevB.01|State=1-AG|HF=-78.0317178|RMSD=4.844e
-009|RMSF=1.122e-004|Dipole=0.,0.,0.|PG=C02H[SGH(C2H4)]||@
TRUST EVERYONE, BUT CUT THE CARDS.
Job cpu time: 0 days 0 hours 0 minutes 23.0 seconds.
```

2. The equilibrium structure is now revised to the MP2/6-31G* level of theory. All electrons are included in the correlation treatment; there are no frozen cores.

```
1|1|UNPC-UNK|FOpt|RMP2-FU|6-31G(d)|C2H4|PCUSER|07-Jan-2008|0|#N GEOM=
ALLCHECK GUESS=READ SCRF=CHECK MP2 (FULL)/6-31G(D) OPT=RCFC|Ethene G1
calculation||0,1|C,-0.3335853571,-0.5780899752,0.|H,0.1808762111,-1.53
31889839,0.|H,-1.4179417525,-0.6105921379,0.|C,0.3335853571,0.57808997
52,0.|H,-0.1808762111,1.5331889839,0.|H,1.4179417525,0.6105921379,0.||
Version=x86-Win32-G03RevB.01|State=1-AG|HF=-78.0310673|MP2=-78.2942862
|RMSD=8.954e-009|RMSF=3.758e-005|Dipole=0.,0.,0.|PG=C02H[SGH(C2H4)]||
```

3. The geometry from step 2 is now used in a number of single point calculations starting with MP4/6-311G**. This energy is improved in four distinct stages and these four improvements are assumed to be additive.

4. As a first correction we add diffuse s and p basis functions at the MP4 level of theory. These are known to be important for molecules with lone pairs. The correction is obtained by comparing the MP4/6-311+G** ($-78.38425E_h$) and the MP4/6-311G** ($-78.38239E_h$) energies. The correction is therefore $-1.86mE_h$.

5. As a second correction we take account of polarization functions on nonhydrogens. This correction is found by comparing MP4/6-311G(2df) and MP4/6-311G** energies. This turns out to be $-40.07mE_h$.

6. The third correction allows for the inadequacies of the MP4 treatment. We make an expensive QCISD(T)/6-311G** calculation. This gives $-1.75mE_h$.

7. If we add these three corrections to the MP4/6-311G** energy, we should get an approximation to a full calculation at the QCISD (T)/6-311G**(2df) level of theory. One such large calculation has therefore been replaced by four smaller and cheaper ones. This energy is still incomplete because of remaining deficiencies in the basis set. A major difficulty is that the correlation energy between two spin-paired electrons in a typical single bond converges only slowly as increasing angular momentum quantum numbers are included in the basis set. G1 theory allows for these effects in an empirical manner and so the fourth correction is to add terms devised to give agreement between theory and experiment for the hydrogen atom and dihydrogen. These are taken to be $6.14mE_h$ for each valence electron pair and $0.19mE_h$ for each unpaired electron, and are referred to as the *higher level correction* (HLC). In terms of the numbers of α - and β -spin electrons

$$\Delta\epsilon \text{ (HLC)} / mE_h = -0.19n_\alpha - 5.95n_\beta \quad (21.1)$$

which comes out as $-36.84 mE_h$ for ethene.

8. Finally, the harmonic frequencies obtained at the HF/6-31G* level of theory are scaled uniformly by the well-accepted factor of 0.8929 to give a zero-point correction of $+48.90mE_h$.

The total energy G1 (at 0 K) is therefore $-78.414021E_h$ and similar calculations for H and C give $-0.5E_h$ and $-37.78464E_h$. Combining these three values gives an ethene atomization energy of $530.1 \text{ kcal mol}^{-1}$ which compares well with the zero-kelvin JANAF thermodynamic value of $531.9 \text{ kcal mol}^{-1}$.

Final output contains extra useful thermodynamic data and is shown below. A simple restart gives rapid results for different temperatures and pressures.

Temperature=	298.150000	Pressure=	1.000000
E(ZPE)=	0.048894	E(Thermal)=	0.051951
E(QCISD(T))=	-78.384142	E(Empiric)=	-0.036840
DE(Plus)=	-0.001864	DE(2DF)=	-0.040069
G1(0 K)=	-78.414021	G1 Energy=	-78.410964
G1 Enthalpy=	-78.410020	G1 Free Energy=	-78.435526

21.2 G2 Theory

G1 theory was originally tested against experimental values for a range of simple first- and second-row molecules. It was observed that G1 theory did badly with ionic molecules,

with triplet state molecules and with hypervalent molecules. *Gaussian 2 (G2) theory* was introduced by Curtiss *et al.* (1991), and it eliminates some of the difficulties by making three improvements to G1.

1. G2 theory eliminates the assumption of additivity of the diffuse sp and the 2df basis functions used in G1 theory. This change gives a significant improvement for ionic species and some anions.
2. It adds a third d function to the nonhydrogen atoms and a second p function to the hydrogen atoms. The third d function is especially important for some hypervalent molecules such as SO₂, whilst the second p function significantly improves the atomization energy of some hydrogen-containing molecules.
3. The higher level correction is determined by a least squares fit for 55 molecules rather than just the hydrogen atom and dihydrogen. This also contributes to an improvement in calculated energies.

In fact, G2 only requires one extra calculation above G1, namely MP2/6-311+G(3df,2p).

A comparison was made for 79 well-established molecules, including 43 that were not included in the original G1 papers. The final total energies are essentially at the QCISD(T)/6-311+G(3df,2p) level of theory.

21.2.1 G2(MP2)

It was subsequently found that significant savings in computer resource could be obtained at little cost in accuracy by reducing the order of the MP4 calculation to MP2 (giving the G2 (MP2) variant).

21.3 G3 Theory

A recent reassessment by Curtiss *et al.* (1998) of G2 theory used 302 energies which included 148 enthalpies of formation, 88 ionization energies, 58 electron affinities and 8 proton affinities for larger and more diverse molecules. This revealed some interesting discrepancies. For example, the enthalpy of formation of CF₄ is too positive by 7.1 kcal mol⁻¹ whilst that of SiF₄ is too negative by 5.5 kcal mol⁻¹. The deviations were also much larger for unsaturated systems than for saturated ones. These considerations led Curtiss *et al.* (1999) to propose *Gaussian-3 (G3) theory*, which follows along the same lines as the earlier G1 and G2 theories in that it is a well-defined sequence of *ab Initio* steps to arrive at the total energy of a given molecule. As will be seen from the discussion below, G3 differs from G2 theory in several major ways.

1. An initial HF/6-31G(d) equilibrium structure is obtained using the RHF or UHF treatment (as in G2 theory).
2. HF/6-31G(d) structure is used to find vibration frequencies that are then scaled by 0.8929. The zero-point energy is calculated (as in G2 theory).
3. Equilibrium geometry is refined at MP2/6-31G(d) level, including all electrons. This geometry is then used in single point calculations (as in G2 theory).

4. The first higher level correction is a complete MP4/6-31G(d) single point, which is then modified to allow for diffuse functions, higher polarization functions, higher correlation terms using QCI, larger basis sets and the effect of nonadditivity.
5. The MP4/6-31G(d) energy and the four corrections are combined along with a spin-orbit correction.
6. The usual higher level correction is added to take account of remaining deficiencies:

$$\begin{aligned}\Delta\varepsilon(\text{HLC}) &= -An_\beta - B(n_\alpha - n_\beta) \quad \text{for molecules} \\ \Delta\varepsilon(\text{HLC}) &= -Cn_\beta - D(n_\alpha - n_\beta) \quad \text{for atoms}\end{aligned}\tag{21.2}$$

A through D are constants chosen to give the smallest average deviation from experiment for the G2 test set.

7. The total energy at 0 K is obtained by adding the zero-point energy from step 2.

The overall agreement with experiment for 299 test energies is 1.02 kcal mol⁻¹ compared to 1.48 kcal mol⁻¹ for G2 theory. Use of MP3 rather than MP4 gives a saving in computer resource but the agreement with experiment becomes 1.22 kcal mol⁻¹.

21.3.1 G3 Variations

In view of the discussion for G2, you will not be surprised to learn that there is a G3(MP2) version which aims to give a saving in computer resource by reducing the MP4 basis set extension calculations to MP2.

Given the steady rise in density functional theory, you might like to read the work of Baboul *et al.* (1999), who use B3LYP geometries and frequencies. The authors report that the use of density functional geometries and zero-point energies in G3 and G3(MP2) is a useful alternative to MP2 geometries and HF zero-point energies.

There have been a number of variations on similar themes over the years, and these are nicely summarized in Baboul *et al.* (1999).

References

- Baboul, A.G., Curtiss, L.A. and Redfern, P.C. (1999) *J. Chem. Phys.*, **110**, 7650.
 Curtiss, L.A., Raghavachari, K., Trucks, G.W. and Pople, J.A. (1991) *J. Chem. Phys.*, **94**, 7221.
 Curtiss, L.A., Raghavachari, K., Redfern, P.C. *et al.* (1998) *J. Chem. Phys.*, **109**, 7764.
 Curtiss, L.A., Redfern, P.C., Raghavachari, K. *et al.* (1999) *J. Chem. Phys.*, **110**, 4703.
 Pople, J.A., Head-Gordon, M., Fox, D.J. *et al.* (1989) *J. Chem. Phys.*, **90**, 5622.

22

Transition States

Geometry optimization plays a key role in modern molecular modelling. *Ab Initio* packages invariably contain powerful options for geometry optimization that can locate stationary points on a molecular potential energy surface starting from naive back-of-an-envelope molecular sketches. Whether that stationary point is the global minimum is of course open to question; you simply have to try a lot of possible starting geometries and use chemical intuition.

Most current methods for geometry optimization are based on a Taylor series expansion of the energy about a reference point \mathbf{X}_0 ; the energy at point \mathbf{X} is given in terms of the gradient \mathbf{g} and the Hessian \mathbf{H} by

$$\varepsilon = \varepsilon_0 + (\mathbf{X} - \mathbf{X}_0)^T \mathbf{g} + \frac{1}{2} (\mathbf{X} - \mathbf{X}_0)^T \mathbf{H} (\mathbf{X} - \mathbf{X}_0) + \dots \quad (22.1)$$

Newton–Raphson type methods truncate the energy expression at the quadratic term, and tell us that the best step to take in order to reach a stationary point is

$$\mathbf{X} - \mathbf{X}_0 = -\mathbf{H}^{-1} \mathbf{g} \quad (22.2)$$

A great deal of effort has gone into determining explicit expressions for the gradient and the Hessian at many levels of theory (such as HF and MP2). Calculations start with initial estimates of the gradient and the Hessian, and an iterative calculation follows. As noted earlier, many methods choose to increment the inverse \mathbf{H}^{-1} rather than \mathbf{H} as the optimization progresses.

The efficiency of a geometry optimization therefore depends on a number of factors: the initial geometry, the choice of coordinate system, the initial Hessian, the Hessian updating method and finally the search direction and step size.

Minima correspond to points on the molecular potential energy surface where the eigenvalues of the Hessian matrix are all positive. Transition states are characterized as stationary points having just one negative eigenvalues of \mathbf{H} . The requirement for a single negative

eigenvalue of \mathbf{H} means that one has to be much more careful with the step taken in an optimization. For example, it is always possible to take a steepest descent step in a minimum search, which will lower the energy or leave it unchanged. Such a step is not appropriate for transition state searching; it is therefore harder to find a transition state than a minimum and many Newton–Raphson type procedures need to start the optimization fairly close to the true transition state structure.

The key concept in our discussion is therefore the Hessian matrix \mathbf{H} . This is a real symmetric matrix and the eigenvalues and eigenvectors can be found by standard numerical methods. Simons *et al.* (1983) showed that each Newton–Raphson step is directed in the negative direction of the gradient for each eigenvector that has a positive Hessian eigenvalue, and along the positive direction of the gradient for each eigenvector that has a negative eigenvalue.

For a transition state search, if you are in a region of the molecular potential energy surface where the Hessian does indeed have one negative eigenvalue, the Newton–Raphson step is appropriate. If you have landed on some region of the surface where the Hessian does not have this desired property, then you must somehow get out of the region and back to a region where the Hessian has the correct structure of one negative eigenvalue.

For minima, qualitative theories of chemical structure are a valuable aid in choosing starting geometries whilst for transition states one only has a vague notion that the saddle point geometry must lie somewhere between the reactants and products. Molecular mechanics (MM) structures are often good as starting structures for geometry optimizations but MM cannot handle bond making and bond breaking.

One of the earliest algorithms that could take corrective action when the wrong region of the molecular potential energy surface was chosen was that due to Poppinger (1975). This author suggested that the lowest eigenvalue of the Hessian should be followed ‘uphill’. The technique has come to be known as *eigenvector following*. There is a good discussion of eigenvector following in Baker’s (1986) paper, which has the following abstract.

An algorithm for locating transition states designed for use in the ab initio program package GAUSSIAN-82 is presented. It is capable of locating transition states even if started in the wrong region of the energy surface, and, by incorporating the ideas on Hessian mode following due to Cerjan and Miller, can locate transition states for alternative rearrangements/dissociation reactions from the same initial starting point. It can also be used to locate minima.

Baker refers to eigenvector following as ‘Hessian mode following’. The big problem with eigenvector following is that the remaining directions along the potential energy surface are left in isolation until the initial fault is corrected.

If we think of a chemical reaction (either reactants or products) as minima on the molecular potential energy surface, then there is no unique way of moving uphill on the surface since all directions go uphill. The *linear synchronous transit* (LST) algorithm searches for a maximum along a path between reactants and products. It frequently yields a structure with two or more negative Hessian eigenvalues, and this is not a transition state. The *quadratic synchronous transit* (QST) method searches along a parabola for a minimum in all directions perpendicular to the parabola. The following is the abstract of Halgren and Lipscomb’s (1977) original reference for synchronous transit.

In the synchronous-transit method, a model linear synchronous transit pathway is first constructed and is then refined by optimizing one or more intermediate structures subject to the

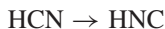
constraint that the optimized structure retain the same relative position along the path (orthogonal optimization). The method yields a series of energy estimates which progressively bound the energy of the transition state from above and from below. High computational efficiency is attainable, and sufficient flexibility is provided to deal with asynchronous processes. Comparisons are made to the alternative ‘reaction coordinate’ approach, which is shown to be subject to several serious deficiencies. The method is applied to a model two-dimensional energy surface and to the allowed electrocyclic interconversions of the cyclopropyl and allyl cations and of cyclobutadiene and *cis*-butadiene.

As we saw in earlier chapters, early work in geometry optimization using electronic structure methods used nonredundant internal coordinates (e.g. the *Z*-matrix). Pulay *et al.* (1979) demonstrated clearly that redundant internal coordinates are the best choice for minimizing many molecules. Peng *et al.* (1996) reported similar conclusions for transition state structures, especially when the initial steps of such calculations are guided by QST.

22.1 An Example

Gaussian 03 includes the *synchronous transit-guided quasi-Newton* method; that is, it uses QST to get close to the quadratic region of the transition state and then quasi-Newton–Raphson to complete the optimization. There are two flavours: Opt = QST2 requires reactants and products as input structures, whilst Opt = QST3 requires in addition the initial transition state.

The isomerization reaction



has been used frequently to test transition state optimizations and reaction path algorithms. Here is the QST3 input:

```
%chk=c:/hcn.chk
# mp2/6-311g(2d,1p) opt=qst3 freq
HCN (Reactant)
0 1
C 0.0 0.0 -0.506777
H 0.0 0.0 -1.574099
N 0.0 0.0 0.659252
HNC (Product)
0 1
C 0.0 0.0 -0.743084
H 0.0 0.0 1.432195
N 0.0 0.0 0.432330
HCN --> HNC Transition State H above CN
0 1
C 0.0 0.0 -0.506777
H 0.0 1.07 -0.506777
N 0.0 0.0 0.659252
```

It is wise to allow for electron correlation when dealing with reactions that involve bond breaking and bond making. I first optimized HCN and HNC and then made

a guess at the transition state (hydrogen above carbon). Here is a fragment of the output:

QST in optimization variable space.

```

Tangent TS vect // Eig F Eigenval
1 R1  0.47182  0.60987  0.00000  -0.07462
2 R2 -0.13266 -0.01687  0.00000  0.16000
3 R3 -0.61297 -0.48257 -0.00003  0.36068
4 A1  0.43494  0.44104  0.00000 1000.00000
5 A2  0.00000  0.00000  0.00000 1000.00000
6 A3 -0.44146 -0.44765  0.00000 1000.00000
7 A4  0.00000  0.00000  0.00000 1000.00000

```

RFO step: Lambda0=4.157509059D-11 Lambda=0.00000000D+00.

Linear search not attempted -- option 19 set.

Iteration 1 RMS(Cart)= 0.00001745 RMS(Int)= 0.00000000

Iteration 2 RMS(Cart)= 0.00000000 RMS(Int)= 0.00000000

Variable	Old X	-DE/DX (Linear)	Delta X (Quad)	Delta X (Total)	Delta X	New X
R1	2.47517	-0.00002	0.00000	-0.00003	-0.00003	2.47514
R2	2.21560	-0.00001	0.00000	-0.00004	-0.00004	2.21556
R3	2.61751	-0.00002	0.00000	-0.00006	-0.00006	2.61745
A1	5.10378	0.00000	0.00000	0.00001	0.00001	5.10379
A2	3.14159	0.00000	0.00000	0.00000	0.00000	3.14159
A3	5.21958	0.00000	0.00000	-0.00001	-0.00001	5.21957
A4	3.14159	0.00000	0.00000	0.00000	0.00000	3.14159

Item	Value	Threshold	Converged?
Maximum Force	0.000017	0.000450	YES
RMS Force	0.000010	0.000300	YES
Maximum Displacement	0.000030	0.001800	YES
RMS Displacement	0.000017	0.001200	YES

Predicted change in Energy=-9.906105D--10

Optimization completed.

-- Stationary point found.

```

-----
! Optimized Parameters !
! (Angstroms and Degrees) !
-----

```

! Name	Definition	TS	Reactant	Product	Derivative Info.
! R1	R(1,2)	1.3098	1.0673	2.1753	-DE/DX = 0.0
! R2	R(1,3)	1.1724	1.166	1.1754	-DE/DX = 0.0
! R3	R(2,3)	1.3851	2.2334	0.9999	-DE/DX = 0.0
! A1	L(2,1,3,-1,-1)	292.425	180.0	180.0	-DE/DX = 0.0
! A2	L(2,1,3,-2,-2)	180.0	180.0	180.0	-DE/DX = 0.0
! A3	L(1,3,2,-1,-1)	299.0599	180.0	180.0	-DE/DX = 0.0
! A4	L(1,3,2,-2,-2)	180.0	180.0	180.0	-DE/DX = 0.0

from which we note that a (triangular) stationary point has indeed been found. But is it a transition state? I included the freq control word to check, and sure enough

```
***** 1 imaginary frequencies (negative Signs) *****
Harmonic frequencies (cm**-1), IR intensities (KM/Mole), Raman scattering
activities (A**4/AMU), depolarization ratios for plane and unpolarized
incident light, reduced masses (AMU), force constants (mDyne/A),
and normal coordinates:
```

	1	2	3
	A'	A'	A'
Frequencies --	-1004.9940	2058.3451	2151.5200
Red. masses --	1.1934	1.0539	11.8266
Frc consts --	0.7102	2.6307	32.2554
IR Inten --	79.7952	218.6471	5.1097

Atom	AN	X	Y	Z	X	Y	Z	X	Y	Z
1	6	0.08	0.06	0.00	-0.06	0.00	0.00	-0.01	0.72	0.00
2	1	0.16	-0.98	0.00	0.99	0.13	0.00	-0.18	0.24	0.00
3	7	-0.08	0.02	0.00	-0.02	-0.01	0.00	0.02	-0.63	0.00

22.2 The Reaction Path

A transition structure is a maximum on the reaction pathway, but for some reactions the potential energy surface can be complicated. It may not be obvious whether the transition state connects the desired reactants and products and in such cases it is advisable to follow the reaction path from the transition state to the reactants and to the products. A reaction path can be defined for any coordinate system as the minimum energy path from the transition state down to the reactants and to the products; if the coordinates are mass-weighted Cartesians this path is the *intrinsic reaction coordinate (IRC)*. The key reference for IRC calculation is the paper by Gonzalez and Schlegel (1989), and this algorithm is implemented in Gaussian 03. The following is part of the abstract.

A new algorithm is presented for obtaining points on a steepest descent path from the transition state of the reactants and products. In mass-weighted coordinates, this path corresponds to the intrinsic reaction coordinate. Points on the reaction path are found by constrained optimizations involving all internal degrees of freedom of the molecule. The points are optimized so that the segment of the reaction path between any two adjacent points is given by an arc of a circle, and so that the gradient at each point is tangent to the path. Only the transition vector and the energy gradients are needed to construct this path. The resulting path is continuous, differentiable and piecewise quadratic. . .

We now take the HCN \rightarrow HNC example.

```
--Link1--
%chk=c:/hcn.chk
# mp2/6-311g(2d,1p) irc=rcfc geom=check
```

HCN TS

0 1

This `--Links--` continues our HCN isomerization reaction discussed above. By default, Gaussian 03 IRC calculations step six points in mass-weighted increments in the 'forward' direction and six in the 'reverse'. Force constants have to be provided (or recalculated), hence the `rcfc` option.

.....

Optimization completed.

-- Optimized point # 12 Found.

```

-----
! Optimized Parameters !
! (Angstroms and Degrees) !
-----
! Name           Value           Derivative information (Atomic Units) !
-----
! X1              0.0             -DE/DX = 0.0                          !
! Y1             -0.2539          -DE/DX = -0.0158                       !
! Z1             -0.603           -DE/DX = -0.007                        !
! X2              0.0             -DE/DX = 0.0                          !
! Y2              0.9864          -DE/DX = 0.0035                       !
! Z2             -0.1426          -DE/DX = 0.0149                       !
! X3              0.0             -DE/DX = 0.0                          !
! Y3              0.0715          -DE/DX = 0.0124                       !
! Z3              0.5549          -DE/DX = -0.0079                      !
-----

```

RADIUS OF CURVATURE = 1.09348

NET REACTION COORDINATE UP TO THIS POINT = 0.58614

OF POINTS ALONG THE PATH = 12

OF STEPS = 2

```

-----
Summary of reaction path following:
(Int. Coord: Angstroms, and Degrees)
-----

```

	ENERGY	RX.COORD	X1	Y1	Z1
1	-93.14253	-0.56420	0.00000	-0.24523	-0.60183
2	-93.14067	-0.46423	0.00000	-0.24115	-0.59957
3	-93.13882	-0.36429	0.00000	-0.23715	-0.59692
4	-93.13706	-0.26487	0.00000	-0.23350	-0.59337
5	-93.13522	-0.16762	0.00000	-0.23137	-0.58878
6	-93.13232	-0.06849	0.00000	-0.23122	-0.58392
7	-93.13097	0.00000	0.00000	-0.23507	-0.58223
8	-93.13388	0.08959	0.00000	-0.23876	-0.58527
9	-93.13711	0.18889	0.00000	-0.23961	-0.59000
10	-93.13936	0.28717	0.00000	-0.24193	-0.59446
11	-93.14130	0.38640	0.00000	-0.24558	-0.59813

12	-93.14313	0.48621	0.00000	-0.24970	-0.60087
13	-93.14485	0.58614	0.00000	-0.25393	-0.60301

The first part of this fragment shows the final geometry optimization, followed by all 13 points along the IRC path. Point 1 is seen to correspond to HCN, point 7 to the transition state and point 13 to HNC which demonstrates that the transition state really does join HCN and HNC.

References

- Baker, J. (1986) *J. Comput. Chem.*, **7**, 385.
Gonzalez, H. and Schlegel, H.B. (1989) *J. Chem. Phys.*, **90**, 2145.
Halgren, T.A. and Lipscomb, W.N. (1977) *Chem. Phys. Lett.*, **49**, 225.
Peng, C., Ayala, P.Y. and Schlegel, H.B. (1996) *J. Comput. Chem.*, **17**, 49.
Poppinger, D. (1975) *Chem. Phys. Lett.*, **35**, 550.
Pulay, P., Fogarasi, G., Pang, F. and Boggs, J.E. (1979) *J. Am. Chem. Soc.*, **101**, 2550.
Simons, J., Jorgensen, P., Taylor, H. and Szmert, J. (1983) *J. Phys. Chem.*, **87**, 2745.

23

Dealing with the Solvent

Chemical reactions generally take place in condensed media. I have had very little to say so far about the presence or absence of a solvent, and many of the applications discussed have referred to single, isolated molecules at 0 K and in free space. The benchmark molecular dynamics (MD) and Monte Carlo (MC) applications discussed in earlier chapters dealt with arrays of particles with particularly simple potentials, and the quantum mechanical models have made no mention of a solvent.

23.1 Solvent Models

The obvious way to take account of the solvent in a MM calculation is to physically add solvent molecules and then optimize the molecular geometry, but such calculations tend to be particularly compute-intensive and the necessary computer power simply was not available to the pioneers in the field. The same comment obviously holds for quantum mechanical studies.

In a medium of relative permittivity ϵ_r , the mutual potential energy of two point charges Q_A and Q_B is reduced by a factor of ϵ_r . Typical values are shown in Table 23.1.

Relative permittivities are temperature dependent, and they are usually recorded as a power series in the temperature:

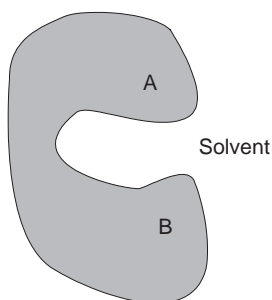
$$\epsilon_r = a + b \left(\frac{T}{K} \right) + c \left(\frac{T}{K} \right)^2 + \dots$$

Early workers in the MM field attempted to allow for the presence of a solvent by modifying any electrostatic contribution to their force field. The problem was that the chosen factor was at first sight quite arbitrary, and bore little relation to the values in Table 23.1. For example, a value of 2.5 was often used for liquid water. All kinds of convincing arguments were used to justify such choices, for example the two atoms A and

Table 23.1 Representative relative permittivities

Substance	ϵ_r
Free space	1
Air	1.0006
Glass	6
Water	81

B in the molecule of Figure 23.1 will not have much exposure to a solvent and so a pure solvent value of ϵ_r is not appropriate.

**Figure 23.1** Atoms A and B in a molecule surrounded by solvent

23.2 Langevin Dynamics

Molecules in solution undergo collisions with other molecules and the solvent, and they can also be thought to experience frictional forces as they move through the solvent. The Langevin technique allows for both these effects. The collisions are modelled by the addition of a random force \mathbf{R} , and the frictional effects are modelled by adding a frictional drag that is directly proportional to the velocity of a given particle. Solvent molecules are not explicitly included.

We saw in Chapter 10 that the key equation for a MD simulation is Newton's second law:

$$m \frac{d^2 \mathbf{r}}{dt^2} = \mathbf{F} \quad (23.1)$$

In Langevin dynamics, the force is modified according to the ideas discussed above:

$$m \frac{d^2 \mathbf{r}}{dt^2} = \mathbf{F} - m\gamma \frac{d\mathbf{r}}{dt} + \mathbf{R} \quad (23.2)$$

The frictional term introduces energy and the random force removes it as kinetic energy. The quantity γ is the collision frequency, and $1/\gamma$ is often called the *velocity relaxation time*; it can be thought of as the time taken for a particle to forget its initial velocity. There is an equation due to Einstein that relates γ to the diffusion constant D of the solvent

$$\gamma = \frac{k_B T}{mD}$$

Integration of the new equation of motion proceeds along the lines discussed in Chapter 10. A straightforward algorithm has been given by Ermak and Buckholtz (1980). The equations of motion are integrated over a time interval Δt that is sufficiently short so that the interparticle forces remain approximately constant. The algorithm for advancing the position \mathbf{r}_A and velocity \mathbf{v}_A of particle A is then similar to those discussed in Chapter 10; we have

$$\begin{aligned}\mathbf{r}_A(t + \Delta t) &= \mathbf{r}_A(t) + c_1 \left(\frac{d\mathbf{r}_A}{dt} \right)_t \Delta t + c_2 \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 + \Delta\mathbf{r}_A^G \\ \mathbf{v}_A(t + \Delta t) &= c_0 \left(\frac{d\mathbf{r}_A}{dt} \right)_t + c_1 \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t \Delta t + \Delta\mathbf{v}_A^G\end{aligned}\quad (23.3)$$

Here $\mathbf{r}_A(t)$, $(d\mathbf{r}_A/dt)_t$ and $(d^2\mathbf{r}_A/dt^2)_t$ are the instantaneous position, velocity and acceleration vector of particle A. The acceleration is calculated from the force. $\Delta\mathbf{r}^G$ and $\Delta\mathbf{v}^G$ are random vectors chosen from a Gaussian distribution with zero mean and standard deviations

$$\begin{aligned}\sigma_r^2 &= \Delta t^2 \left(\frac{k_B T}{m_A} \right) \frac{1}{\gamma \Delta t} \left(2 - \frac{1}{\gamma \Delta t} (3 - 4 \exp(-\gamma \Delta t) + \exp(-2\gamma \Delta t)) \right) \\ \sigma_v^2 &= \left(\frac{k_B T}{m_A} \right) (1 - \exp(-\gamma \Delta t))\end{aligned}\quad (23.4)$$

The numerical coefficients are given by

$$\begin{aligned}c_0 &= \exp(-\gamma \Delta t) \\ c_1 &= \frac{1 - c_0}{\gamma \Delta t} \\ c_2 &= \frac{1 - c_1}{\gamma \Delta t}\end{aligned}$$

At low values of the friction coefficient, the dynamical aspects dominate. If the interparticle forces are taken to vary linearly with time between each time step, the equations of motion can be rewritten in a form that is said to produce a more accurate simulation:

$$\begin{aligned}\mathbf{r}_A(t + \Delta t) &= \mathbf{r}_A(t) + c_1 \left(\frac{d\mathbf{r}_A}{dt} \right)_t \Delta t + c_2 \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t (\Delta t)^2 + \Delta\mathbf{r}_A^G \\ \mathbf{v}_A(t + \Delta t) &= c_0 \left(\frac{d\mathbf{r}_A}{dt} \right)_t + (c_1 - c_2) \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_t \Delta t + c_2 \left(\frac{d^2\mathbf{r}_A}{dt^2} \right)_{t+\Delta t} \Delta t + \Delta\mathbf{v}_A^G\end{aligned}\quad (23.5)$$

and as $\gamma \rightarrow 0$ we recover the *velocity Verlet* algorithm discussed in Chapter 10. For large values of γ , the random collisions dominate and the motion becomes diffusion-like.

The same considerations apply to Langevin dynamics as to standard MD; there are three stages to a simulation, the heating phase, the data collection stage and the optional cooling stage. For the sake of illustration, Figure 23.2 shows the end-to-end distance in the hydrocarbon $C_{20}H_{42}$, over a 10 ps experiment with a 4 ps^{-1} friction coefficient.

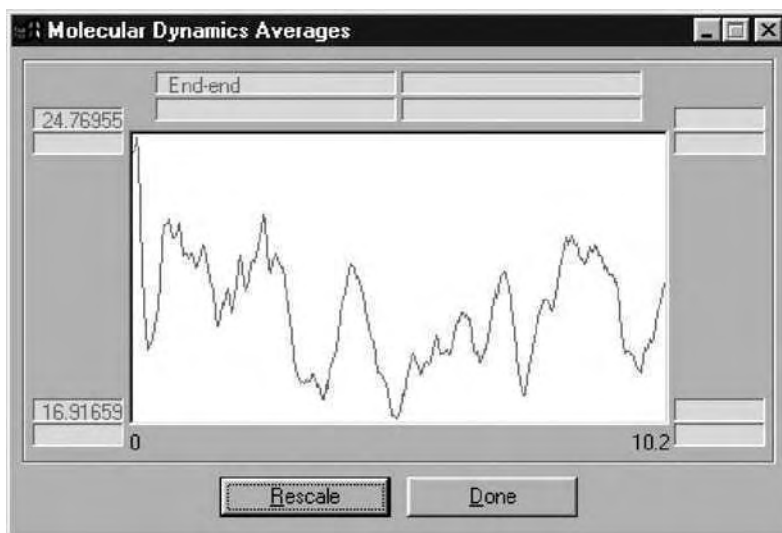


Figure 23.2 Langevin dynamics on C_{20} hydrocarbon

The end-to-end distance varied much less than a comparable gas-phase experiment, showing the damping effect of the random collisions.

23.3 Continuum Solvation Models

In our discussion of electric charges and their properties (Chapter 1), we met the concepts of polarization and polarizability. The dielectric polarization \mathbf{P} physically is the induced dipole moment per volume, and I mentioned that this experimentally determined quantity can depend on the applied field \mathbf{E} in all manner of complicated ways. For very simple media and for low field strengths, it turns out that \mathbf{P} is directly proportional to \mathbf{E} and so

$$\mathbf{P} = (\varepsilon_r - 1) \varepsilon_0 \mathbf{E}$$

$$\chi_e = \varepsilon_r - 1$$

where I have introduced the electric susceptibility χ_e . The relative permittivity and the refractive index n are also related by

$$\varepsilon_r = n^2$$

Suppose now that the material is a gas comprising N atoms each of polarizability α in a volume V . When an external field \mathbf{E} is applied, each atom acquires an induced dipole moment $\alpha\mathbf{E}$ and so the polarization is

$$\mathbf{P} = \frac{N\alpha}{V} \mathbf{E} \quad (23.6)$$

Comparison of the macroscopic and microscopic equations gives

$$n = \left(1 + \frac{N\alpha}{\epsilon_0 V} \right)^{1/2}$$

which can be expanded by the binomial theorem to give

$$n = 1 + \frac{N\alpha}{2\epsilon_0 V}$$

and if the gas is a perfect gas we have finally

$$n = 1 + \frac{p\alpha}{2\epsilon_0 k_B T} \quad (23.7)$$

which gives an experimental route to the polarizability α .

In condensed phases, the separation between molecules is of the order of the molecular dimensions and the result is that each molecule is polarized not by just the ordinary applied electric field \mathbf{E} but by a *local field* \mathbf{F} comprising \mathbf{E} plus the field of all the other dipoles. Once the local field is known, we can use Equation (23.6) to find \mathbf{P} simply by substituting $\alpha\mathbf{F}$ for $\alpha\mathbf{E}$. The calculation of \mathbf{F} is in fact difficult because the dipoles that contribute to \mathbf{F} are themselves determined by \mathbf{F} and a self-consistent treatment is needed.

A little manipulation yields the well-known *Clausius–Mossotti* equation

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{N\alpha}{3\epsilon_0 V} \quad (23.8)$$

which can be used to calculate α from an experimental value of ϵ_r .

So far we have assumed that the molecules in the system have no permanent electric dipole moments. When this restriction is relaxed we have to allow for the *orientation polarization*; the interaction energy of a dipole moment vector with an electric field is

$$U = -\mathbf{p}_0 \cdot \mathbf{E}$$

so that in a fluid the molecules will tend to orient themselves parallel to the field. This tendency will be opposed by random thermal agitation, and a calculation by the methods of statistical mechanics yields Debye's (1912) correction to the Clausius–Mossotti equation

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{N}{3\epsilon_0 V} \left(\alpha + \frac{p_0^2}{3k_B T} \right) \quad (23.9)$$

Debye's dipole theory accounts quantitatively for the dielectric properties of gases and qualitatively for those of condensed phases. A more thorough analysis of the problem shows that one should not expect the dipole moment to remain constant because real molecules have a nonzero polarizability. The polarization of the dielectric in the electric field of the molecule itself gives rise to a *reaction field*. The classic literature reference is Onsager (1936).

Continuum models of the solvent have their origin in these physical considerations; one component of the system (the solute M) is treated microscopically and the remainder of the system (the solvent) is treated macroscopically.

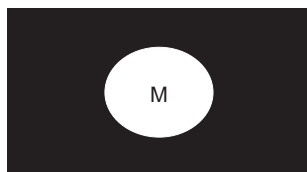


Figure 23.3 The Onsager model

In any continuum model, M resides in a cavity carved out of an infinite slab of continuous dielectric (the solvent), as shown in Figure 23.3. The cavity is a basic concept in all continuum models; the size and shape of the cavity are differently defined in the various versions of continuum models but, as a general rule, it should exclude the solvent and contain within its boundaries the largest possible part of the solute charge distribution. Much attention has been paid to the portion of the solute electronic charge outside the boundary of the cavity and the term *escaped charge* is often used in this context.

It is universally accepted that the cavity shape should ideally reproduce the molecular shape but as usual in modelling studies there is a trade-off between computer resource and the desire for better accuracy. Calculations are usually easiest when the cavity has a simple shape.

23.3.1 The Onsager Model

In the simplest such model, named after Onsager, the cavity is a sphere of fixed radius; the relative permittivity is 1 inside the cavity and equal to that of the solvent outside. M must have a nonzero dipole moment, which induces a polarization \mathbf{P} in the dielectric. \mathbf{P} then reacts with the solute dipole, leading to stabilization. Workers in the field speak about the *self-consistent reaction field (SCRF)*.

A suitable Onsager sphere radius for (for example) phenylalanine can be found from a Gaussian 03 run such as:

```
%chk=d:\phenopt.chk
# B3LYP/6-311G** Opt geom=check volume
L-phenylalanine geometry optimization then volume
0 1
```

which produced the following MC calculation of the molecular volume, and estimate of the cavity sphere radius.

```
Using the total density.
Monte-Carlo method of calculating molar volume:
based on 0.001 e/bohr**3 density envelope.
Number of points per bohr**3 = 20 CutOff= 1.00D-04
Using the SCF density.
There are 379 points. Will hold 379 in memory.
```

```

LenV= 6095246 MDV= 6291456.
Box volume = 6850.641 fraction occupied=0.214
Integrated density= 8.0982806152148228D+01
  error=-7.0171938478517717D+00
Molar volume = 1464.121 bohr**3/mol (130.657 cm**3/mol)
Recommended a0 for SCRF calculation = 4.60 angstrom ( 8.70 bohr)

```

Finally, an Onsager SCRF calculation using dichloromethane (arbitrarily) as solvent:

```

%chk=d:\phenopt.chk
# B3LYP/6-311G** scrf=(dipole,a0=4.60,
  solvent=dichloromethane)
# geom=check
L-phenylalanine Onsager model
  0 1

```

gives the following:

```

SCRF input: a0= 4.60 angstrom, dielectric constant= 8.93
Requested convergence on RMS density matrix=1.00D-04 within 128 cycles.
Requested convergence on MAX density matrix=1.00D-02.
Requested convergence on energy=5.00D-05.
No special actions if energy rises.
SCF Done: E(RB+HF-LYP) = -554.947318515  A.U. after 7 cycles
      Conv = 0.4858D-05      -V/T = 2.0038
      S**2 = 0.0000
Final SCRF E-Field is:
Dipole : -0.00005916 -0.00025177 -0.00075469
Quadrupole : 0.00000000 0.00000000 0.00000000
              0.00000000 0.00000000 0.00000000
Octapole : 0.00000000 0.00000000 0.00000000 0.00000000
              0.00000000 0.00000000 0.00000000 0.00000000
              0.00000000 0.00000000
Hexadecapole: 0.00000000 0.00000000 0.00000000 0.00000000
                0.00000000 0.00000000 0.00000000 0.00000000
                0.00000000 0.00000000 0.00000000 0.00000000
                0.00000000 0.00000000 0.00000000
Polarization energy = -0.248405263057E-03
Total energy of solute = -554.947566921
Total energy (include solvent energy) = -554.947318515
Total energy (without reaction field) = -554.947070110

```


23.3.2 Polarizable Continuum Models

It is possible in principle to define and determine the cavity from a straightforward *ab initio*/DFT calculation by defining a molecular isodensity level outside which electron density is treated as negligible. Such calculations are possible within Gaussian 03 but literature applications are rare (presumably because of the expense).

The generally accepted compromise between computer resource and accuracy is based on the definition of the cavity as a superposition of interlocked atomic spheres with radii near the van der Waals radius. Our key reference is the review by Tomasi *et al.* (2005), which builds on the oft-quoted review by Tomasi and Persico (1994).

For example, we might want to investigate the geometry and molecular properties of phenylalanine in different solvents using the PCM model. Here is a suitable Gaussian 03 .gjf:

```
%chk=d:\phenopt.chk
# B3LYP/6-311G** scrf=(pcm,solvent=dichloromethane) opt freq
# geom=check
L-phenylalanine PCM model
0 1
```

I have not shown the output for lack of space, but if you care to run the job then everything will be instantly recognizable, including the optimized geometry and final energy.

23.4 Periodic Solvent Box

In Langevin dynamics and continuum models, the solvent is simulated; no solvent molecules are explicitly included in the calculation. Such calculations give little microscopic detail.

I mentioned the interest in water as solvent in earlier chapters, and it is worth reading the abstract to Jorgensen's (1983) landmark Monte Carlo paper:

Classical Monte Carlo simulations have been carried out for liquid water in the NPT ensemble at 25°C and 1 atm using six of the simpler intermolecular potential functions for the water dimer: Bernal-Fowler (BF), SPC, ST2, TIP2, TIP3P and TIP4P. Comparisons are made with experimental thermodynamic and structural data including the recent neutron diffraction data of Thiessen and Narten. The computed densities and potential energies are in reasonable accord with experiment except for the original BF model, which yields an 18% overestimate of the density and poor structural results. The TIP2 and TIP4P potentials yield oxygen-oxygen partial structure functions in good agreement with the neutron diffraction results. The accord with the experimental OH and HH partial structure functions is poorer; however, the computed results for these functions are similar for all the potential functions. Consequently, the discrepancy may be due to the correction terms needed in processing the neutron data or to an effect uniformly neglected in the computations. Comparisons are also made for self-diffusion coefficients obtained from molecular dynamics simulations. Overall, the SPC, SDT2, TIP2 and TIP4P models give reasonable structural and thermodynamic descriptions of liquid water and they should be useful in simulations of aqueous solutions. The simplicity of the SPC, TIP2, and TIP4P is also attractive from a computational standpoint.

Modern molecular modelling packages such as HyperChem usually include the option of a periodic solvent box, often the TIP3P box of 216 water molecules. The example shown in Figure 23.4 is from a MM optimization using phenylalanine in water.



Figure 23.4 Phenylalanine in periodic solvent box

The standard box is a cube of side 1870 pm, and it is conventional to choose the appropriate box with sides in multiples of this. I should not need to remind you by now of the significance of the word ‘periodic’.

References

- Debye, P. (1912) *Physik. Z.*, **13**, 97.
 Ermak, D.L. and Buckholtz, H. (1980) *J. Comput. Phys.*, **35**, 169.
 Jorgensen, W.L., Chandrasekhar, J. and Madura, J.D. (1983) *J. Phys. Chem.*, **79**, 926.
 Onsager, L. (1936) *J. Am. Chem. Soc.*, **58**, 1486.
 Tomasi, J. and Persico, M. (1994) *Chem. Rev.*, **94**, 2027.
 Tomasi, J., Mennucci, B. and Cammi, R. (2005) *Chem. Rev.*, **105**, 2999.

24

Hybrid Models; the QM/MM Approach

It all started with a paper (Warshel and Levitt 1976) in the *Journal of Molecular Biology*, which by and large is not read by chemists. Once again, I will let the two authors give you their ideas in their own words.

A general method for detailed study of enzymic reactions is presented. The method considers the complete enzyme–substrate complex together with the surrounding solvent and evaluates all the different quantum mechanical and classical energy factors that can affect the reaction pathway. These factors include the quantum mechanical energies associated with bond cleavage and charge redistribution of the substrate and the classical energies of steric and electrostatic interactions between the substrate and the enzyme. The electrostatic polarization of the enzyme atoms and the orientation of the dipoles of the surrounding water molecules is simulated by a microscopic dielectric model. The solvation energy resulting from this polarization is considerable and must be included in any realistic calculation of chemical reactions involving anything more than an isolated atom *in vacuo*. Without it, acidic groups can never become ionized and the charge distribution on the substrate will not be reasonable. The same dielectric model can also be used to study the reaction of the substrate in solution. In this way, the reaction in solution can be compared with the enzymic reaction. . .

What the authors did was to combine a molecular mechanics (MM) potential for the solvent with an early (MINDO/2) quantum mechanical model for the solute. By 1998 such hybrid methods had become sufficiently important to justify an American Chemical Society symposium (Gao and Thompson 1998).

The early hybrid models such as that of Warshel and Levitt are often referred to as ‘QM/MM’. Note that there is no mention of geometry optimization, and that attention focuses on the solute/solvent problem.

24.1 Link Atoms

In the case of solution-phase studies, it is clear where we should draw a boundary between the parts of a system that ought to be studied quantum mechanically, and those that can be treated by the techniques of MM.

Biochemists have different problems in mind; they want to divide parts of a protein chain into regions of interest that ought to be treated quantum mechanically, and the remainder of the chain that can be treated according to the methods of MM.

Consider for example the protein shown in Figure 24.1. The bottom left-hand amino acid is valine, which is linked to proline. Suppose for the sake of argument that we wanted to treat this valine quantum mechanically, and the rest of the protein chain according to the methods of MM. We would have to draw a QM/MM boundary somewhere between valine and the rest of the protein. The *link atoms* define the boundary between the QM and the MM regions. A great deal of care has to go into this choice of boundary. The boundary should not give two species whose chemical properties are quite different from those implied by the structural formulae on either side of this boundary.

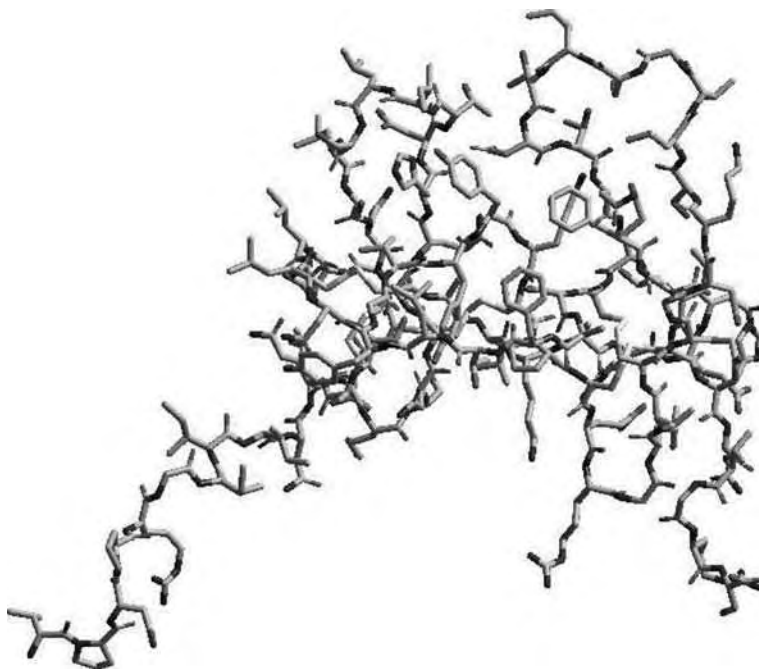


Figure 24.1 A typical protein (1PCN, from the Protein Databank)

In order to achieve this aim, it is usual to ‘cap’ the exposed valency at the boundary by adding an extra quantum mechanical atom. There are several reasons for doing this, not the least of which is that we do not want to be left with a radical for the quantum mechanical part. The extra quantum mechanical atom does not see any atoms within the MM region

and the final atom of the MM region is attached to this quantum mechanical atom by a harmonic MM potential.

24.2 IMOMM

When faced with the theoretical study of a system involving very large molecules, there are two possible approaches. One is to model the real system as a simpler one and proceed with (for example) an *ab initio* treatment of the simpler system. Thus for example, transition metal complexes often contain bulky ligands with large organic substituents; those far away from the metal atom are usually replaced by hydrogen atoms. This approach works well because the metal–ligand interaction remains well described. The only problem arises when the bulky ligands give rise to steric interactions.

Molecular mechanics methods have found favour in biochemical applications where very large molecules are involved, or areas such as MD where a large number of different calculations are required.

One possible way to overcome the shortcomings of both approaches is to mix quantum mechanics and MM in different parts of the same molecule. The *integrated ab initio + molecular mechanics (IMOMM)* scheme was introduced by Maseras and Morokuma (1995). As usual, I will give the abstract.

A new computational scheme integrating *ab initio* and molecular mechanics descriptions in different parts of the same molecule is presented. In contrast with previous approaches, this method is especially designed to allow the introduction of molecular mechanics corrections in full geometry optimizations concerning problems usually studied through *ab initio* calculations on model systems. The scheme proposed in this article intends to solve some of the systematic error associated with modelling through the use of molecular mechanics calculations. This method, which does not require any new parameter, evaluates explicitly the energy derivatives with respect to geometrical parameters and therefore has a straightforward application to geometry optimization. Examples of its performance on two simple cases are provided: the equilibrium geometry of cyclopropane and the energy barriers on S_N2 reactions of alkyl chloride systems. Results are in satisfactory agreement with those of full *ab initio* calculations in both cases.

Twenty years is a long time in molecular modelling, and it is interesting to note both the change of emphasis for hybrid models and the increased level of sophistication. Geometry optimization is *de rigueur*, and transition states are fair game for scrutiny.

To fix our ideas, I will discuss the (hypothetical) ‘real’ system $M(P(CH_3)_3)_2$ used by the authors quoted above. Here M is a transition metal such as platinum. Our ‘model’ system for the *ab initio* study is $M(PH_3)_2$ where we have replaced the methyl groups by hydrogen. In Figure 24.2 we divide the atoms in both the real (MM, lower molecule) and model (*ab initio*, upper molecule) systems into four different sets, each of them with corresponding coordinates.

- Set 1 atoms are present in both.
- Set 2 atoms are present in the *ab initio* system but not in the MM.
- Set 3 atoms are present in the MM calculation, but each atom is replaced by a simpler one in the *ab initio* calculation (these are of course the ‘link atoms’ discussed earlier).
- Set 4 atoms are present only in the MM calculation.

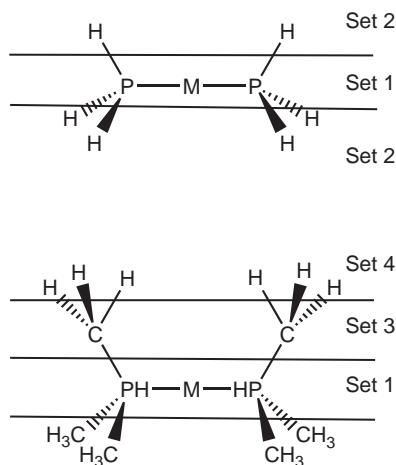


Figure 24.2 The (upper) model system is for *ab initio* study; the (lower) real system for MM study

The coordinates of atoms in set 3 are a function of those of atoms in sets 1 and 2, and the authors show how to write equations for the energies and gradients for the combined *ab initio* and MM descriptions. Care has to be taken to eliminate double counting of energy terms. The algorithm is as follows.

- Input coordinates for the set 1, set 2 and set 4 atoms.
- Use the *ab initio* procedure to calculate the energy and the gradients with respect to the coordinates of set 1, and of set 2.
- Carry out a MM optimization for the set 4 atoms keeping set 1 and set 3 frozen.
- Add the energies and gradients and test for self consistency.

24.3 IMOMO

The next step, integrated molecular orbital + MO, came in 1996 and once again I will let the originators (Humbel *et al.* 1996) spell out the good news.

A new theoretical method, called IMOMO (integrated MO (molecular orbital) + MO), for integration of two different levels of MO approximation is presented. Only the active or more difficult part of a molecule is treated at a higher level of approximation and the rest of the molecule at a lower level of approximation. The integrated total energy and energy derivatives are defined from three different calculations, and the structure of transition state as well as the equilibrium structure can be optimized using the integrated energy. Any combination of any molecular orbital approximations (*ab initio*, density functional to semi-empirical) can be used. Test calculations in the IMOMO method have been performed and compared with normal MO calculations for the conformation energy of ethane and *n*-butane and the S_N2 reaction of ethyl, propyl, isobutyl and neopentyl chloride with Cl^- . The results indicate that these methods have a tremendous potential for theoretical study of larger molecules, in particular for transition states.

24.4 ONIOM (Our Own *N*-layered Integrated Molecular Orbital and Molecular Mechanics)

We note several different schemes in which a molecule is divided into two parts, with one accurately described and the other treated at a lower level of theory. The total energy of the real system is a combination of both parts. For molecular systems normally treated with MM, the integrated IMOMM method is appropriate. For systems and reactions that require a more accurate description of the geometries and energies, the integrated IMOMO method has been shown to reproduce high-level *ab initio* results. The next logical step involves combining these approaches into an onion-like multilayered scheme, with MM used to describe steric and electrostatic effects in the exterior parts of a molecule, an intermediate MO method describing the electronic effects of functional groups or ligands close to the centre of action, and a highly accurate method to deal with the electron correlation on the most important parts of the molecule. The layers are conventionally known as the low, medium and high layers; these are treated with different model chemistries and the results automatically combined to give the final predicted results, and such a calculation would be denoted ONIOM3 (Svensson *et al.* 1996). Rather than quote the original abstract, I will give you a more up-to-date version due to Vreven and Morokuma (2000) that contains a wealth of useful detail.

Five years ago Morokuma and colleagues introduced the IMOMO method, which integrates two molecular orbital methods into one calculation. Since then, the method has been expanded in several ways; it has been generalized to consider up to three methods, and has been unified as the ONIOM method to include both MO and MM combinations. In this review we present the history of the method, a number of chemical problems that we have studied, how to assess IMOMO combinations and partitionings, and our latest efforts that take the method beyond the conventional investigation of ground state energy surfaces. In particular we emphasize the importance of the S-value test for validation of the ONIOM method/model combinations. The method combination depends much on the properties and accuracies required. Generally speaking, however, if the target level is CCSD(T) or G2, the best choice of low level is MP2. If MP2 or DFT is the target level, HF or eventually semiempirical MO methods are good choices of low level. These methods can be further combined with an outer-most layer of the MM level.

References

- Gao, J. and Thompson, M.A. (eds) (1998) *Combined Quantum Mechanical and Molecular Mechanical Methods*, ACS Symposium Series 712, American Chemical Society, Washington, DC.
- Humbel, S., Sieber, S. and Morokuma, K. (1996) *J. Chem. Phys.*, **105**, 1959.
- Maseras, F. and Morokuma, K. (1995) *J. Comput. Chem.*, **16**, 1170.
- Svensson, M., Humbel, S., Froese, R.D.J. *et al.* (1996) *J. Phys. Chem.*, **100**, 19357.
- Vreven, T. and Morokuma, K. (2000) *J. Comput. Chem.*, **21**, 1419.
- Warshel, A. and Levitt, M. (1976) *J. Mol. Biol.*, **103**, 227.

Appendix A: A Mathematical *Aide-Mémoire*

There are several mathematical topics such as vectors, vector calculus, determinants, matrices and linear operators that appear throughout this book. You should have them at your fingertips if you want to fully understand the subject; it is no use speaking glibly about the scalar and cross products of two vectors, gradients and Hessians, eigenvalues and eigenvectors, Hamiltonian operators and so on when you have not the slightest idea what the terms mean. I assume that you have come across most of these topics elsewhere, and so this appendix is essentially an *aide-mémoire*.

Angular momentum dominates the theory of atomic structure (and if you know about commuting linear operators, you will understand why), so I have included a section on this topic.

A.1 Scalars and Vectors

I should emphasize that physical quantities such as mass, temperature and time have two component parts: a measure (i.e. how many) and a unit (e.g. a kelvin or a second). It is usual these days to express all physical quantities in the system of units referred to as the *Système International* (SI). The International Union of Pure and Applied Physics and the International Union of Pure and Applied Chemistry both recommend SI units. These are based on the metre, kilogram, second and ampere as the fundamental units of length, mass, time and electric current. There are three additional units in SI: the kelvin, mole and candela that are units of thermodynamic temperature, amount of substance and luminous intensity, respectively.

A *scalar* s is a quantity such as mass, temperature, time that can be represented by a single value. The *modulus* (or *magnitude*, or *size*) of s , denoted $|s|$, is the value of s irrespective of its sign. So if T is a temperature and $T = -273$ K, the modulus of T is 273 K. If you

know about thermodynamics, you will realize that temperatures cannot be lower than 0 K, but that is another story.

A vector quantity is a quantity such as displacement, force, velocity that has both a magnitude and a direction in space. I am going to write vectors in bold script such as \mathbf{a} , \mathbf{b} , \mathbf{E} . Other authors follow different conventions, and you might find it easier to write vectors by underlining them \underline{a} , \underline{b} , \underline{E} rather than trying to write them in bold print with a biro. My continental European colleagues often write vectors with an arrow above the symbol \vec{a} , \vec{b} , \vec{E} .

The size or *modulus* of vector \mathbf{v} is denoted by v or $|\mathbf{v}|$. It is important to distinguish vector quantities from scalar ones; they are certainly not the same thing. Develop the habit of writing vectors and their moduli in a distinctive style. Once again, you should recognize that vectors representing physical quantities have a measure and an SI unit. A *unit vector* is one with unit modulus (apart from the SI unit), and a unit vector in the direction of vector \mathbf{v} is \mathbf{v}/v . Some authors write unit vectors with a special ‘hat’, typically $\hat{\mathbf{v}}$. I am not going to follow this convention, because I want to reserve the hat symbol for a *linear operator*, of which more shortly.

A.2 Vector Algebra

Vector notation makes many equations look simpler than they would otherwise be, and so I will make extensive use of vectors through the book. I need to remind you of certain elementary properties.

A.2.1 Vector Addition and Scalar Multiplication

It is common practice when drawing vectors to show their direction by including an arrow-head, as illustrated in Figure A.1. The vector $2\mathbf{v}$ is a vector in the same direction as \mathbf{v} but with twice the modulus. The vector $-\mathbf{v}$ has twice the modulus of \mathbf{v} but it points in the opposite direction. For many kinds of vector, it does not matter where on the page we draw them. Two vectors are equal if they have the same direction and magnitude.

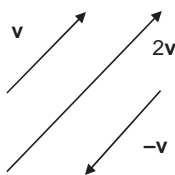


Figure A.1 *Scalar multiplication of a vector*

Two vectors \mathbf{u} and \mathbf{v} are added together according to the *parallelogram rule*; we draw \mathbf{u} and then add \mathbf{v} onto \mathbf{u} according to Figure A.2.

Vectors obey the law that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ and we could have equally drawn \mathbf{v} first and then added \mathbf{u} to get the sum of the two.

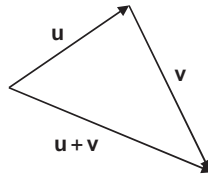


Figure A.2 Addition of two vectors

A.2.2 Cartesian Coordinates

The point P in Figure A.3 has Cartesian coordinates (x, y, z) . The Cartesian unit vectors \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z are directed parallel to the x , y and z axes respectively. They are sometimes written \mathbf{i} , \mathbf{j} and \mathbf{k} rather than \mathbf{e}_x , \mathbf{e}_y and \mathbf{e}_z . The *position vector* \mathbf{r} of point P can be expressed by drawing an arrow from the origin to P, and in view of the laws for vector addition and scalar multiplication we have

$$\mathbf{r} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z \quad (\text{A.1})$$

I have indicated the *components* x , y , z in Figure A.3.

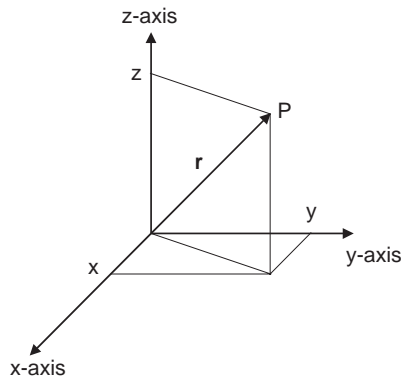


Figure A.3 Cartesian components of a vector

A.2.3 Cartesian Components of a Vector

Vectors \mathbf{u} and \mathbf{v} can be specified by their Cartesian components

$$\mathbf{u} = u_x\mathbf{e}_x + u_y\mathbf{e}_y + u_z\mathbf{e}_z$$

$$\mathbf{v} = v_x\mathbf{e}_x + v_y\mathbf{e}_y + v_z\mathbf{e}_z$$

so that the vector $k\mathbf{u}$ (where k is a scalar) has components

$$k\mathbf{u} = ku_x\mathbf{e}_x + ku_y\mathbf{e}_y + ku_z\mathbf{e}_z \quad (\text{A.2})$$

whilst the vector sum $\mathbf{u} + \mathbf{v}$ has components

$$\mathbf{u} + \mathbf{v} = (u_x + v_x)\mathbf{e}_x + (u_y + v_y)\mathbf{e}_y + (u_z + v_z)\mathbf{e}_z \quad (\text{A.3})$$

A.2.4 Vector Products

Vectors can be multiplied by scalars, but some care is needed when considering products of two vectors.

A.2.4.1 Scalar (or Dot) Product

The scalar (or dot) product of two vectors \mathbf{u} and \mathbf{v} is

$$\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \cos \theta \quad (\text{A.4})$$

where θ is the angle between \mathbf{u} and \mathbf{v} , and $|\mathbf{u}|$, $|\mathbf{v}|$ are the moduli of the vectors. If $\mathbf{u} \cdot \mathbf{v} = 0$ and neither of \mathbf{u} and \mathbf{v} are zero vectors, then we say that \mathbf{u} and \mathbf{v} are *orthogonal* (in other words, they are at right angles to each other).

Scalar products obey the rules

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= \mathbf{v} \cdot \mathbf{u} \\ \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) &= \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \end{aligned}$$

and the Cartesian unit vectors satisfy

$$\begin{aligned} \mathbf{e}_x \cdot \mathbf{e}_x &= \mathbf{e}_y \cdot \mathbf{e}_y = \mathbf{e}_z \cdot \mathbf{e}_z = 1 \\ \mathbf{e}_x \cdot \mathbf{e}_y &= \mathbf{e}_x \cdot \mathbf{e}_z = \mathbf{e}_y \cdot \mathbf{e}_z = 0 \end{aligned}$$

It follows that the scalar product of \mathbf{u} and \mathbf{v} can be written

$$\mathbf{u} \cdot \mathbf{v} = u_x v_x + u_y v_y + u_z v_z \quad (\text{A.5})$$

and the modulus of vector \mathbf{v} is

$$|\mathbf{v}| = (\mathbf{v} \cdot \mathbf{v})^{1/2} = (v_x^2 + v_y^2 + v_z^2)^{1/2}$$

It also follows that the angle θ between the vectors \mathbf{u} and \mathbf{v} is given by

$$\cos \theta = \frac{u_x v_x + u_y v_y + u_z v_z}{(u_x^2 + u_y^2 + u_z^2)^{1/2} (v_x^2 + v_y^2 + v_z^2)^{1/2}}$$

The scalar product of \mathbf{u} and \mathbf{v} gives the projection of \mathbf{u} in the direction of \mathbf{v} (Figure A.4), which by symmetry is the same as the projection of \mathbf{v} in the direction of \mathbf{u} .

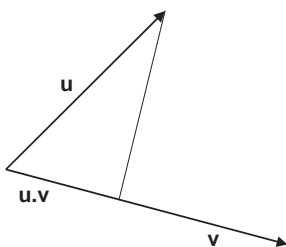


Figure A.4 Projection of \mathbf{u} in the direction of \mathbf{v}

A.2.4.2 Vector (or Cross) Product

In addition to the scalar product $\mathbf{u} \cdot \mathbf{v}$, there is a *vector* (or *cross*) product $\mathbf{u} \times \mathbf{v}$

$$\mathbf{u} \times \mathbf{v} = |\mathbf{u}| |\mathbf{v}| \sin \theta \mathbf{n} \quad (\text{A.6})$$

where θ is the angle between \mathbf{u} and \mathbf{v} and \mathbf{n} is a unit vector normal to the plane containing \mathbf{u} and \mathbf{v} . The direction of $\mathbf{u} \times \mathbf{v}$ is given by the direction a screw would advance if rotated from the direction of \mathbf{u} to the direction of \mathbf{v} . This is shown in Figure A.5.

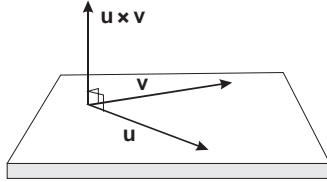


Figure A.5 Vector product of \mathbf{u} and \mathbf{v}

Vector products obey the rules

$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$$

$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$$

and the Cartesian unit vectors satisfy

$$\mathbf{e}_x \times \mathbf{e}_x = \mathbf{e}_y \times \mathbf{e}_y = \mathbf{e}_z \times \mathbf{e}_z = \mathbf{0}$$

$$\mathbf{e}_x \cdot \mathbf{e}_y = \mathbf{e}_z$$

$$\mathbf{e}_y \cdot \mathbf{e}_z = \mathbf{e}_x$$

$$\mathbf{e}_z \cdot \mathbf{e}_x = \mathbf{e}_y$$

It follows that the vector product of \mathbf{u} and \mathbf{v} can be written

$$\mathbf{u} \times \mathbf{v} = (u_y v_z - u_z v_y) \mathbf{e}_x + (u_z v_x - u_x v_z) \mathbf{e}_y + (u_x v_y - u_y v_x) \mathbf{e}_z \quad (\text{A.7})$$

A.3 Scalar and Vector Fields

Mathematically, a field is a function that describes a physical property at points in space. In a *scalar field*, this physical property is completely described by a single value for each point (e.g. temperature, electron density). A scalar field can be represented pictorially by contours, which are lines or surfaces that link together points with the same value of the field *in a given plane*. An example is shown in Figure A.6 (taken from a quantum mechanical study of the electron distribution in a diatomic molecule; the contour lines join points of equal electron density in a plane drawn so as to contain the two nuclei, which on close inspection can be identified as the circles joined by a ‘bond’).

For quantitative work we need to know the values of the field at each contour; this is sometimes achieved by labelling the contours or by a system of colour coding. For *vector fields* (e.g. gravitational force, electrostatic field intensity), both a direction and a magnitude are required at for each point.

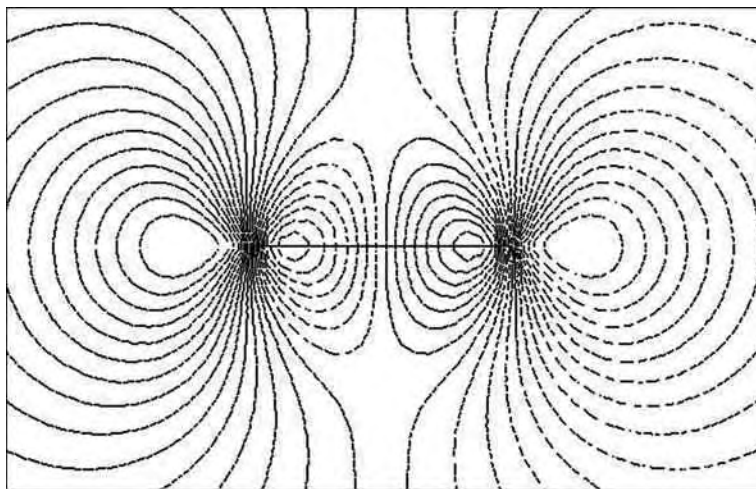


Figure A.6 Contour lines for a scalar field

A.4 Vector Calculus

In this section, I want to explain how fields can be differentiated. I will then discuss two aspects of scalar fields; their *gradient* (which describes their rate of change in space) and their *integrals* over surfaces and volumes. There are other advanced concepts that you might have come across, for example the divergence and curl of a vector field. I do not intend to draw on these ideas in this text.

A.4.1 Differentiation of Fields

Suppose that the vector field $\mathbf{u}(t)$ is a continuous function of the scalar variable t . As t varies, so does \mathbf{u} and if \mathbf{u} denotes the position vector of a point P, then P moves along a continuous curve in space.

For most of this book we will identify the variable t as time and we might be interested in studying the trajectory of a particle along the above curve in space. Suppose then that a particle moves from point P to point Q as in Figure A.7, in time δt .

Mathematicians tell us that we will profit from a study of what happens to the ratio $\delta\mathbf{u}/\delta t$ as δt gets smaller and smaller, just as we do in ordinary differential calculus. The first differential $d\mathbf{u}/dt$ is defined as the limit of this ratio as the interval δt becomes progressively smaller:

$$\begin{aligned} \frac{d\mathbf{u}}{dt} &= \text{Lim} \frac{\delta\mathbf{u}}{\delta t} \\ &= \text{Lim} \left(\frac{\delta u_x}{\delta t} \mathbf{e}_x + \frac{\delta u_y}{\delta t} \mathbf{e}_y + \frac{\delta u_z}{\delta t} \mathbf{e}_z \right) \\ &= \frac{du_x}{dt} \mathbf{e}_x + \frac{du_y}{dt} \mathbf{e}_y + \frac{du_z}{dt} \mathbf{e}_z \end{aligned} \quad (\text{A.8})$$

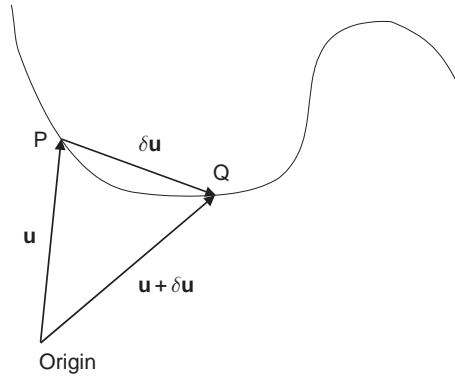


Figure A.7 A curve in space

It is a vector directed along the tangent at P (left to right in the figure). The derivative of a vector is the vector sum of the derivatives of its components. The usual rules for differentiation apply:

$$\begin{aligned}\frac{d}{dt}(\mathbf{u} + \mathbf{v}) &= \frac{d\mathbf{u}}{dt} + \frac{d\mathbf{v}}{dt} \\ \frac{d}{dt}(k\mathbf{u}) &= k \frac{d\mathbf{u}}{dt} \\ \frac{d}{dt}(f\mathbf{u}) &= \frac{df}{dt}\mathbf{u} + f \frac{d\mathbf{u}}{dt}\end{aligned}$$

where k is a scalar and f a scalar field.

A.4.2 The Gradient

One of the most important properties of a scalar field is the rate at which it changes in space. A vector called the *gradient vector* describes this. Suppose that f is a scalar field, and we wish to investigate how f changes between the points \mathbf{r} and $\mathbf{r} + d\mathbf{r}$. Here

$$d\mathbf{r} = \mathbf{e}_x dx + \mathbf{e}_y dy + \mathbf{e}_z dz$$

and the change in f is

$$df = f(\mathbf{r} + d\mathbf{r}) - f(\mathbf{r})$$

We know from the rules of partial differentiation that

$$df = \left(\frac{\partial f}{\partial x}\right) dx + \left(\frac{\partial f}{\partial y}\right) dy + \left(\frac{\partial f}{\partial z}\right) dz$$

and so we identify df as a certain scalar product

$$df = \left(\frac{\partial f}{\partial x}\mathbf{e}_x + \frac{\partial f}{\partial y}\mathbf{e}_y + \frac{\partial f}{\partial z}\mathbf{e}_z\right) \cdot (\mathbf{e}_x dx + \mathbf{e}_y dy + \mathbf{e}_z dz)$$

The first vector on the right-hand side is called the *gradient of f* , and it is written $\text{grad } f$ (in this text):

$$\text{grad } f = \frac{\partial f}{\partial x} \mathbf{e}_x + \frac{\partial f}{\partial y} \mathbf{e}_y + \frac{\partial f}{\partial z} \mathbf{e}_z \quad (\text{A.9})$$

An alternative notation, used in other texts, involves the use of the so-called gradient operator ∇ (pronounced *del*)

$$\nabla = \frac{\partial}{\partial x} \mathbf{e}_x + \frac{\partial}{\partial y} \mathbf{e}_y + \frac{\partial}{\partial z} \mathbf{e}_z$$

In this case the gradient of f is written ∇f .

The vector $\text{grad } f$ is a vector field whose direction at any point is the direction in which f is increasing most rapidly and whose magnitude is the rate of change of f in that direction. The spatial rate of change of the scalar field f in the direction of an arbitrary unit vector \mathbf{e} is given by the scalar product $\mathbf{e} \cdot \text{grad } f$.

A.4.3 Volume Integrals of Scalar Fields

In order to evaluate quantities such as the mass or electric charge contained within a region of space, it is necessary to evaluate certain *volume integrals*.

I can best illustrate the concept with the simple example shown in Figure A.8. Suppose that the electric charge density ρ inside a cubic box whose faces are the planes $x = 0$, $x = 1$; $y = 0$, $y = 1$; $z = 0$, $z = 1$ is given by $\rho = \rho_0 (x + y + z)$, where ρ_0 is a constant.

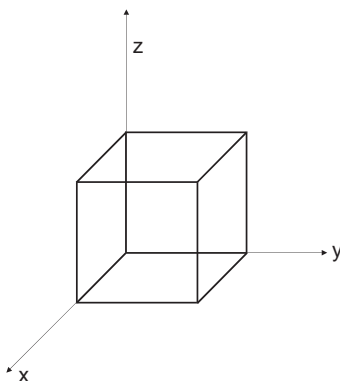


Figure A.8 A cubic volume

We divide each axis into differential elements, dx 's along the x axis, dy 's along the y axis and dz 's along the z axis, giving a number of infinitesimal differential volume elements each of volume $dx dy dz$. The charge enclosed by each differential volume element $dx dy dz$ is $\rho dx dy dz$ and so the total Q enclosed by the box is

$$Q = \int_0^1 \int_0^1 \int_0^1 \rho_0 (x + y + z) \, dx dy dz$$

We form three single integrals. First of all, we draw within the region a column having cross-section dz , and constant x and y .

In order to add these contributions to Q , we integrate between the limits of $z = 0$ and $z = 1$ to give

$$\int_0^1 \rho_0 (x + y + z) dz = \rho_0 \left(x + y + \frac{1}{2} \right)$$

Next we form the y integral by drawing a slice parallel to the yz plane and including the column. The limits of the y integration are $y = 0$ and $y = 1$:

$$\int_0^1 \rho_0 \left(x + y + \frac{1}{2} \right) dy = \rho_0 (x + 1)$$

Finally we complete the volume by adding together all the slabs, and the limits of the x integration are $x = 0$ to $x = 1$:

$$\int_0^1 \rho_0 (x + 1) dx = \frac{3}{2} \rho_0$$

I have used the symbol $d\tau$ to denote a volume element; other commonly used symbols are dV , $d\mathbf{r}$, $d\vec{r}$.

A.4.4 Line Integrals

Figure A.9 refers to a particle at point P moving along a curve in space under the influence of a force \mathbf{F} . I have taken the force to be centred at the origin, for example the particle at P might be a point charge moving under the influence of another point charge at the origin. The work done w in moving the short distance $\delta\mathbf{r}$ is the projection of \mathbf{F} along the displacement $\mathbf{F} \cdot \delta\mathbf{r}$. In order to calculate the total work done as the particle moves from initial point A to final point B, we divide the curve into small segments $\delta\mathbf{r}_1, \delta\mathbf{r}_2, \dots, \delta\mathbf{r}_N$. We then have

$$w \approx \sum_{i=1}^N \mathbf{F}(\mathbf{r}_i) \cdot \delta\mathbf{r}_i$$

We let the number of points become infinitely large, and the summation approaches a limit called the *line integral*:

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \text{Lim} \sum_{i=1}^N \mathbf{F}(\mathbf{r}_i) \cdot \delta\mathbf{r}_i \quad (\text{A.10})$$

If the points A and B coincide, then the line integral refers to a *closed curve* and we often write the line integral with a circle round the integral sign

$$\oint_C \mathbf{F} \cdot d\mathbf{r}$$

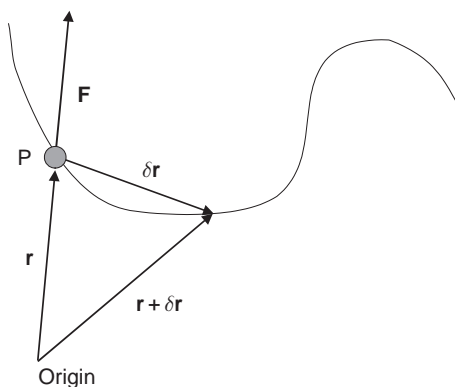


Figure A.9 Work done under the influence of a force

If the line integral of a certain vector field \mathbf{F} is zero around any arbitrary closed path, then the vector field is called a *conservative field*. It can be shown that every conservative field can be written as the gradient of a suitable scalar field

$$\mathbf{F} = -\text{grad}\phi \quad (\text{A.11})$$

The importance of scalar fields such as ϕ is that changes in ϕ depend only on the starting point and finishing point, and not on the path chosen to get between these points.

A.5 Determinants

The set of simultaneous linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned}$$

has solution

$$x_1 = \frac{b_1 a_{22} - b_2 a_{12}}{a_{11} a_{22} - a_{12} a_{21}} \quad x_2 = \frac{b_2 a_{11} - b_1 a_{21}}{a_{11} a_{22} - a_{12} a_{21}}$$

If we define a *determinant of order 2* by the symbol

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \equiv (ad - bc) \quad (\text{A.12})$$

then these solutions may be written as

$$x_1 = \frac{\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}} \quad x_2 = \frac{\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}}$$

We speak of the elements, rows, columns and diagonal of the determinant in an obvious manner. In a similar fashion, solution of the three simultaneous linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned}$$

involves a denominator

$$a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}$$

and so we define a determinant of order three by

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{32}a_{13} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13} \quad (\text{A.13})$$

By inspection, this can be written in terms of certain determinants of order 2

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

The three determinants of order 2 are called the *minors* of a_{11} , a_{21} and a_{31} respectively; they are the determinants produced by striking out the first column and successive rows of the determinant of order 3. A little analysis shows that this determinant may be expanded down any of its columns or along any of its rows by suitably combining products of elements and their minors.

The properties of determinants are as follows.

1. The value of a determinant is unchanged by interchanging the elements of any corresponding rows and columns.
2. The sign of a determinant is reversed by interchanging any two of its rows or columns.
3. The value of a determinant is zero if any two of its rows (or columns) are identical.

A.6 Matrices

A matrix is a set of $m \times n$ quantities arranged in a rectangular array of m rows and n columns, for example

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Throughout this text, I will denote matrices by bold letters, just like vectors. The matrix above is said to be of *order* m by n (denoted $m \times n$). The null matrix \mathbf{O} is one whose elements

are all zero. If $m = n$, the matrix is said to be square of order n , and a square matrix whose only nonzero elements are the diagonal elements is said to be a *diagonal matrix*. Thus

$$\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$$

is a diagonal matrix. The unit matrix (**1** or **I**) of order n is a diagonal matrix whose elements are all equal to 1.

A *row vector* is a matrix of order $1 \times n$ and a *column vector* is a matrix of order $n \times 1$. For example (a_1, a_2, \dots, a_n) is a row vector and

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

is a column vector.

A.6.1 Transpose of a Matrix

Interchanging rows and columns of a determinant leaves its value unchanged, but interchanging the rows and columns of a matrix **A** produces a new matrix called the *transpose* \mathbf{A}^T . Thus for example if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \text{ then } \mathbf{A}^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{pmatrix}$$

In the case that $\mathbf{A} = \mathbf{A}^T$ then we say that the matrix **A** is *symmetric*. A symmetric matrix is square and has $a_{ij} = a_{ji}$.

A.6.2 Trace of a Square Matrix

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

The elements a_{11} , a_{22} and a_{33} are called the *diagonal elements*. The sum of the diagonal elements is called the *trace* of the matrix:

$$\text{Tr}\mathbf{A} = a_{11} + a_{22} + a_{33} \quad (\text{A.14})$$

and the following rules are obeyed:

$$\text{Tr}(k\mathbf{A}) = k\text{Tr}\mathbf{A}$$

$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA}) \quad (\text{A.15})$$

$$\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}\mathbf{A} + \text{Tr}\mathbf{B}$$

A.6.3 Algebra of Matrices

We need to be aware of a few simple matrix properties as follows.

1. If \mathbf{A} and \mathbf{B} are two matrices of the same order with elements a_{ij} and b_{ij} , then their sum $\mathbf{S} = \mathbf{A} + \mathbf{B}$ is defined as the matrix whose elements $c_{ij} = a_{ij} + b_{ij}$.
2. Two matrices \mathbf{A} and \mathbf{B} with elements a_{ij} and b_{ij} are equal only if they are of the same order, and all their corresponding elements are equal $a_{ij} = b_{ij}$.
3. The result of multiplying a matrix \mathbf{A} whose elements are a_{ij} by a scalar k is a matrix whose elements are ka_{ij} .
4. The definition of *matrix multiplication* is such that two matrices \mathbf{A} and \mathbf{B} can only be multiplied together to form their product \mathbf{AB} when the number of columns of \mathbf{A} is equal to the number of rows of \mathbf{B} . Suppose \mathbf{A} is a matrix of order $(m \times p)$, and \mathbf{B} is a matrix of order $(p \times n)$. Their product $\mathbf{C} = \mathbf{AB}$ is a matrix of order $(m \times n)$ with elements

$$c_{ij} = \sum_{k=1}^p a_{ik}b_{kj} \quad (\text{A.16})$$

Thus for example if

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 6 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

then

$$\mathbf{AB} = \begin{pmatrix} 13 & 17 & 21 \\ 18 & 24 & 30 \\ 27 & 36 & 45 \end{pmatrix} \quad \text{and} \quad \mathbf{BA} = \begin{pmatrix} 14 & 29 \\ 32 & 68 \end{pmatrix}$$

so we see from this simple example that \mathbf{AB} is not necessarily equal to \mathbf{BA} . If $\mathbf{AB} = \mathbf{BA}$ then we say that the matrices *commute*.

A.6.4 Inverse Matrix

Let \mathbf{A} be the square matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

If a suitable matrix \mathbf{X} can be found such that $\mathbf{AX} = \mathbf{1}$ then we refer to \mathbf{X} as the *inverse* of \mathbf{A} , and write it \mathbf{A}^{-1} . We say that \mathbf{A} is *invertible* or *nonsingular*. Not all matrices are invertible; the rule is that square matrices with nonzero determinant such as

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

are invertible, whilst those with zero determinant such as

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

are not. Matrices with zero determinants are often called *singular*.

A.6.5 Matrix Eigenvalues and Eigenvectors

Consider once again an $(n \times n)$ matrix \mathbf{A} . If we form the product of \mathbf{A} with a suitable but arbitrary column vector \mathbf{u} then sometimes it will happen that the product $\mathbf{A}\mathbf{u}$ is a linear multiple λ of \mathbf{u} , $\mathbf{A}\mathbf{u} = \lambda\mathbf{u}$. In this case we say that \mathbf{u} is an *eigenvector* (or *eigenfunction*) of \mathbf{A} with *eigenvalue* λ . There are exactly n eigenvalues and eigenvectors. Thus for example if

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

then we find

$$\lambda_1 = \sqrt{2} \quad \text{and} \quad \mathbf{u}_1^T = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$$

$$\lambda_2 = 0 \quad \text{and} \quad \mathbf{u}_2^T = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

$$\lambda_3 = -\sqrt{2} \quad \text{and} \quad \mathbf{u}_3^T = \begin{pmatrix} \frac{1}{2} & -\frac{\sqrt{2}}{2} & \frac{1}{2} \end{pmatrix}$$

If two or more eigenvalues are the same, we say that they are *degenerate*.

A.7 Angular Momentum

Figure A.7 shows a particle moving along a curve in space. If the particle has mass m then its *linear momentum* \mathbf{p} is defined as

$$\mathbf{p} = m \frac{d\mathbf{r}}{dt} \quad (\text{A.17})$$

Newton's second law of motion relates force to the rate of change of linear momentum

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \quad (\text{A.18})$$

In the absence of an applied force, the linear momentum is constant. When the mass is constant, Equation (A.18) can be written $\mathbf{F} = m d^2\mathbf{r}/dt^2$.

The *angular momentum* \mathbf{l} is defined as a vector cross product

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \quad (\text{A.19})$$

Care has to be exercised when dealing with angular momentum, because \mathbf{l} depends on the point chosen to be the coordinate origin. If we were to choose a new coordinate origin such that the position vector of the particle were

$$\mathbf{r}' = \mathbf{r} + \mathbf{R}$$

where \mathbf{R} is a constant vector, we have

$$\begin{aligned}\mathbf{l}' &= \mathbf{r}' \times \mathbf{p} \\ &= \mathbf{l} + \mathbf{R} \times \mathbf{p}\end{aligned}$$

On differentiation of both sides of Equation (A.19) with respect to time we find

$$\begin{aligned}\frac{d\mathbf{l}}{dt} &= \frac{d}{dt} \mathbf{r} \times \mathbf{p} \\ \frac{d\mathbf{l}}{dt} &= \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt} \\ \frac{d\mathbf{l}}{dt} &= \mathbf{r} \times \mathbf{F}\end{aligned}$$

The vector cross product of \mathbf{r} and the applied force \mathbf{F} , $\mathbf{r} \times \mathbf{F}$, is known as the *torque* and so I have proved that in the absence of an applied torque, the angular momentum of a particle remains constant in time.

Consider now a particle of mass m executing circular motion around the z axis as shown in Figure A.10. We can write

$$\begin{aligned}\mathbf{r} &= x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z \\ &= a \sin \theta \mathbf{e}_x + a \cos \theta \mathbf{e}_y + d\mathbf{e}_z\end{aligned}$$

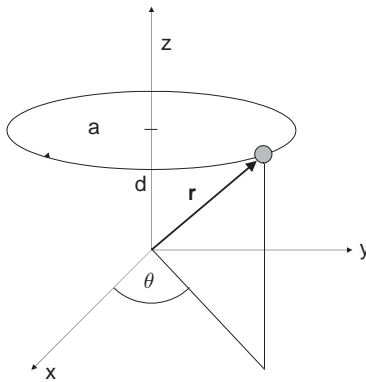


Figure A.10 Circular motion around the z axis

Differentiating we have

$$\begin{aligned}\frac{d\mathbf{r}}{dt} &= -a \frac{d\theta}{dt} \cos \theta \mathbf{e}_x + a \frac{d\theta}{dt} \sin \theta \mathbf{e}_y \\ &= \frac{d\theta}{dt} (-y\mathbf{e}_x + x\mathbf{e}_y) \\ &= \frac{d\theta}{dt} \mathbf{e}_z \times \mathbf{r}\end{aligned}$$

It also follows that

$$\begin{aligned}\mathbf{I} &= \mathbf{r} \times \mathbf{p} \\ &= m\mathbf{r} \times \frac{d\mathbf{r}}{dt} \\ &= m \frac{d\theta}{dt} \mathbf{r} \times (-y\mathbf{e}_x + x\mathbf{e}_y) \\ &= m \frac{d\theta}{dt} (-zx\mathbf{e}_x + zy\mathbf{e}_y + (x^2 + y^2) \mathbf{e}_z)\end{aligned}$$

The vector \mathbf{I} is not directed along the z axis because of the choice of ‘origin’ along the z axis but for rotation with $z = 0$ we have

$$\begin{aligned}\mathbf{I} &= m \frac{d\theta}{dt} (x^2 + y^2) \mathbf{e}_z \\ &= m \frac{d\theta}{dt} a^2 \mathbf{e}_z\end{aligned}\tag{A.20}$$

The quantity ma^2 turns out to have a special significance for circular motion, and we call it the *moment of inertia* about the z axis.

A.8 Linear Operators

I now need to remind you about linear operators, mathematical entities that form a key concept in the language of quantum mechanics. An operator \hat{A} can be thought of as a mathematical entity that turns one function into another. It is conventional to write them with the hat sign in order to distinguish them from other mathematical constructs. To keep the discussion simple, I will just deal in this section with operators that work on functions of a real single variable that I will write consistently as x .

Thus for example we might have

$$\hat{A}f(x) = (x^2 + 1)f(x)\tag{A.21}$$

which means that the action of \hat{A} is to replace $f(x)$ by $(x^2 + 1)f(x)$. So if $f(x) = \exp(x)$ we would have $f(0.1) = 1.105$ and $\hat{A}f(0.1) = (0.1^2 + 1) \times 1.105 = 1.116$. I should emphasize that this is not an equation that must somehow be solved for a value of x , it is a mapping between two sets of functions often represented as a Venn diagram (Figure A.11).

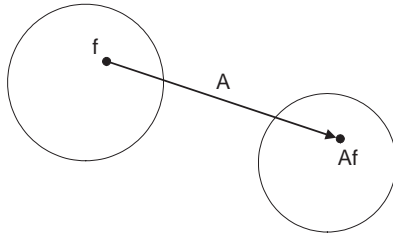


Figure A.11 Venn diagram

Now consider the operator \hat{B} which operates on a differentiable function $f(x)$ to give the first derivative $df(x)/dx$. We can write this as

$$\hat{B}f(x) = \frac{d}{dx}f(x) \quad (\text{A.22})$$

so \hat{B} acting on $\exp(x)$ gives $\exp(x)$. It is clear that action of these two operators on the same function generally gives a different result.

An operator is *linear* if it satisfied the following two properties:

$$\begin{aligned} \hat{A}(f(x) + g(x)) &= \hat{A}f(x) + \hat{A}g(x) \\ \hat{A}(cf(x)) &= c\hat{A}f(x) \text{ where } c \text{ is a scalar} \end{aligned}$$

and both of the two operators discussed above are linear.

Consider now the product of two operators \hat{A} and \hat{B} which we write as $\hat{A}\hat{B}$. We define this product as

$$(\hat{A}\hat{B})f(x) = \hat{A}(\hat{B}f(x))$$

that is to say, we calculate the result $\hat{B}f(x)$ first and then operate with \hat{A} on the result. Using the two operators discussed so far we have

$$\hat{A}\hat{B}f(x) = (x^2 + 1) \frac{df(x)}{dx}$$

whilst

$$\hat{B}\hat{A}f(x) = 2xf(x) + (x^2 + 1) \frac{df(x)}{dx}$$

They certainly do not give the same result and it therefore matters which way round we apply the operators. If we now define a third operator \hat{C} whose effect on a function is to multiply by x such that

$$\hat{C}f(x) = xf(x)$$

we see that

$$\hat{B}\hat{A}f(x) = 2\hat{C}f(x) + \hat{A}\hat{B}f(x)$$

This is true for all differentiable functions $f(x)$ and so it forms an *operator identity*. We would write it as

$$\hat{B}\hat{A} = 2\hat{C} + \hat{A}\hat{B} \quad (\text{A.23})$$

In the special case where two operators satisfy the equality

$$\hat{B}\hat{A} = \hat{A}\hat{B} \quad (\text{A.24})$$

we say that they *commute* (you will recognize much of the terminology that we used for matrices; once again, that is a different story). Commuting operators play a major role in quantum mechanics, as we will shortly see.

To every linear operator there belongs a special set of functions $u_i(x)$ and scalars a_i such that

$$\hat{A}u_i(x) = a_i u_i(x) \quad (\text{A.25})$$

The scalars can be real or complex. We say that $u_i(x)$ are the *eigenvectors* (or *eigenfunctions*) of the operator and a_i are the *eigenvalues*. Depending on the physical circumstances and the operator, there may be a finite set of eigenvalues, a countably infinite set or a continuum.

To finish this section, let me give you an important result from the theory of linear operators (together with a simplified proof).

Theorem

If two linear operators commute, then it is possible to find a set of simultaneous eigenfunctions.

Proof (simplified version)

Suppose that two operators have each a single eigenvector and eigenvalue, given by

$$\hat{A}u(x) = au(x)$$

$$\hat{B}v(x) = bv(x)$$

If we operate on the first equation by the second operator we find

$$\hat{B}\hat{A}u(x) = \hat{A}\hat{B}u(x)$$

but we know that the operators commute and so

$$\hat{A}\hat{B}u(x) = \hat{A}bv(x)$$

or

$$\hat{A}[\hat{B}u(x)] = a[\hat{B}u(x)]$$

The function in square brackets is therefore a linear multiple of the eigenfunction of the operator \hat{A}

$$\hat{B}u(x) = cu(x)$$

which shows that the eigenfunctions of the two operators are the same, apart from some arbitrary multiplicative constant.

Linear operators play an important role in quantum mechanics because:

- every observable such as position, momentum, energy and angular momentum can be represented by a linear operator;
- the results of measurements of these observables are given by the eigenvalues of the operators;
- observables can be measured simultaneously to arbitrary precision only when their operators commute.

We usually give some concrete form to the linear operators, especially those representing position and linear momentum. In the Schrödinger representation we substitute as follows.

- For a position operator such as \hat{x} we write x .
- For a momentum operator such as \hat{p}_x we substitute

$$-j\frac{h}{2\pi}\frac{\partial}{\partial x}$$

A little operator analysis shows that

$$xp_x - p_x x = j\frac{h}{2\pi} \quad (\text{A.26})$$

and this is consistent with the famous Heisenberg uncertainty principle.

A.9 Angular Momentum Operators

The classical angular momentum vector of a particle with position vector \mathbf{r} and linear momentum \mathbf{p} is

$$\mathbf{l} = \mathbf{r} \times \mathbf{p} \quad (\text{A.27})$$

In classical mechanics, we can measure the three components and the square of the length, l_x, l_y, l_z and $l^2 = (l_x^2 + l_y^2 + l_z^2)$, simultaneously with arbitrary precision. Things are far more interesting for quantum mechanical systems.

Expanding the vector product Equation (A.27) we see that the x component of \mathbf{l} is

$$l_x = yp_z - zp_y$$

The corresponding operator is

$$\hat{l}_x = \hat{y}\hat{p}_z - \hat{z}\hat{p}_y$$

which becomes, in Schrödinger representation

$$\hat{l}_x = -j\frac{h}{2\pi}\left(\hat{y}\frac{\partial}{\partial z} - \hat{z}\frac{\partial}{\partial y}\right)$$

After some operator algebra it can be shown that the various components of the angular momentum operator do not commute with each other and so cannot be measured simultaneously to arbitrary precision. They are subject to Heisenberg's uncertainty principle. We find in addition that the square of the magnitude and any of the three components can be measured simultaneously to arbitrary precision. We conclude that at best we can simultaneously measure the square of the magnitude and one component, which is by convention called the z component.

The eigenvalues and eigenfunctions of angular momentum can be found most easily in spherical polar coordinates

$$\begin{aligned} \hat{l}_z &= -j\frac{h}{2\pi}\frac{\partial}{\partial\phi} \\ \hat{l}^2 &= -\frac{h^2}{4\pi^2}\left(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) \end{aligned} \quad (\text{A.28})$$

This problem is tackled in detail in all the traditional quantum chemistry texts, such as Eyring *et al.* (1944), and we can write

$$\begin{aligned}\hat{l}^2 Y_{l,m_l}(\theta, \phi) &= l(l+1) \frac{h^2}{4\pi^2} Y_{l,m_l}(\theta, \phi) \\ \hat{l}_z Y_{l,m_l}(\theta, \phi) &= m_l \frac{h}{2\pi} Y_{l,m_l}(\theta, \phi)\end{aligned}\tag{A.29}$$

The eigenfunctions are well known from mathematical physics, and they are called *spherical harmonics*. The quantum numbers l and m_l are restricted to certain discrete values, $l = 0, 1, 2, 3, \dots$ and for each value of l , m_l can take integral values ranging from $-l$ to $+l$.

We visualize angular momentum as follows. The vector precesses round the z axis in such a way that its z component is constant in time, but that the x and y components vary. There are $2l + 1$ possible orientations of the vector. I have illustrated this for the case $l = 1$ in Figure A.12, where just one of the three possible vector alignments is shown.

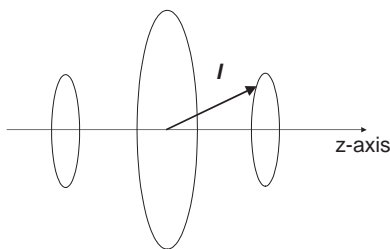


Figure A.12 Angular momentum with $l = 1$

For each orientation the vector precesses around the axis keeping the z component fixed at

$$m_l \frac{h}{2\pi}$$

The length of the vector is

$$\sqrt{l(l+1)} \frac{h}{2\pi}$$

A.10 Problem Set

A set of problems, together with solutions, is available on the associated website.

Reference

Eyring, H., Walter, J. and Kimball, G.E. (1944) *Quantum Chemistry*, John Wiley & Sons, Inc., New York.

Appendix B: Glossary

AM1	Austin model 1
Amber	Assisted model building and energy refinement
B3LYP	Becke 3-parameter/Lee, Yang and Parr exchange–correlation functional
CASSCF	Complete active space self-consistent field
CID	Configuration interaction with double excitations
CIS(D)	Configuration interaction with single excitations (doubles added as perturbation)
CNDO	Complete neglect of differential overlap
CNDO(S)	Complete neglect of differential overlap (spectroscopic version)
EHT	Extended Hückel theory
HF-LCAO	Hartree–Fock linear combination of atomic orbitals
G_n	Gaussian n (for accurate thermodynamic calculations)
GTO	Gaussian-type orbital
GVB	Generalized valence bond
HOMO	Highest occupied molecular orbital
IMOMM	Integrated molecular orbital + molecular mechanics
IMOMO	Integrated molecular orbital + molecular orbital
INDO	Intermediate neglect of differential overlap
IRC	Intrinsic reaction coordinate
KS-LCAO	Kohn–Sham linear combination of atomic orbitals
L-J	Lennard-Jones potential
LUMO	Lowest unoccupied molecular orbital
MINDO	Modified intermediate neglect of differential overlap
MM	Molecular mechanics
MNDO	Modified neglect of differential overlap
ONIOM	Our own N -layered integrated molecular orbital and molecular mechanics
OPLS	Optimized potentials for liquid simulations
PM3	Parameterized model 3

PPP	Pariser–Parr–Pople
QCISD	Quadratic configuration interaction with singles and doubles
QSAR	Quantitative structure and activity relationships
SAM1	Semi <i>ab initio</i> model 1
SCF	Self-consistent field
SCRF	Self-consistent reaction field
STO	Slater-type orbital
UFF	Unified force field
ZDO	Zero differential overlap
ZINDO/1	Zerner’s intermediate neglect of differential overlap 1
ZINDO(S)	ZINDO for spectroscopy

Appendix C: List of Symbols

$\langle \dots \rangle$	Mean value/time average
A	Thermodynamic Helmholtz energy
\hat{A}, \hat{B}, \dots	Generic operators
a_0	Atomic unit of length (the bohr)
\mathbf{B}	Wilson B-matrix
C_6, C_{12}	Lennard-Jones parameters
C_v, C_p	Heat capacities at constant volume and pressure
$D(\varepsilon)$	Density of states
D_0	Spectroscopic dissociation energy
D_e	Thermodynamic dissociation energy
d	Contraction coefficient (e.g. in STO- n G expansion)
$d\tau$	Volume element
E	Electron affinity
$\mathbf{E}(\mathbf{r})$	Electric field (\mathbf{r} = field point)
E_h	Atomic unit of energy (the hartree)
\mathbf{F}	Force
G	Thermodynamic Gibbs energy
\mathbf{g}	Gradient vector
$\hat{g}(\mathbf{r}_1, \mathbf{r}_2)$	Two-electron operator
\mathbf{H}	Hessian matrix
H	Thermodynamic enthalpy; classical Hamiltonian
\hat{H}	Hamiltonian operator
$H_\nu(\xi)$	Hermite polynomial of degree ν
\mathbf{h}_1	Matrix of one-electron integrals in LCAO models
$\hat{h}^{(1)}(\mathbf{r}_1)$	One-electron operator
\hat{I}, \hat{L}	Ionization energy
$\mathbf{J}, \mathbf{K}, \mathbf{G}$	Coulomb, exchange, G matrices from LCAO models

$\hat{J}(\mathbf{r})$	Coulomb operator
j	Square root of -1
$\hat{\mathbf{j}}, \hat{\mathbf{J}}$	Total angular momentum vector operators
$\hat{K}(\mathbf{r})$	Exchange operator
k_s	Force constant
L-J	Lennard-Jones (potential)
\mathbf{l}, \mathbf{L}	Angular momentum vectors
$\hat{\mathbf{l}}, \hat{\mathbf{L}}$	Orbital angular momentum vector operators
n	Amount of substance; principal quantum number
n^*	Effective quantum number
$\mathbf{P}(\mathbf{r})$	Dielectric polarization (\mathbf{r} = field point); electron density
p	Pressure
\mathbf{p}_e	Electric dipole moment
Q	Partition function
Q_A	Point charge
q	Normal coordinate; atomic charge; molecular partition function; Mulliken atom charge
\mathbf{q}	Quaternion
\mathbf{q}_e	Electric second moment
R	Gas constant
\mathbf{R}	Rotation matrix
\mathbf{R}_A	Position vector
R_e	Equilibrium bond length
R_H	Rydberg constant for hydrogen
R_∞	Rydberg constant for one-electron atom with infinite nuclear mass
\mathbf{r}, \mathbf{R}	Field point
S	Thermodynamic entropy
s	Shielding constant
$\hat{\mathbf{s}}, \hat{\mathbf{S}}$	Spin angular momentum vector operators
U	Mutual potential energy
U, U_{th}	Thermodynamic internal energy
V	Volume
Z	Atomic number
α	Electric polarizability
α	GTO orbital exponent; exchange parameter in X α DFT
α_A	Hückel π -electron Coulomb integral for atom A
α_e	Vibration-rotation coupling constant
β_{AB}	Hückel π -electron resonance integral for bonded pairs A, B
β_{AB}^0	Bonding parameter in semiempirical theories (e.g. CNDO)
ε	Energy; orbital energy
ε_0	Permittivity of free space
ε_r	Relative permittivity
ζ	STO orbital exponent
Θ_e	Electric quadrupole moment
μ	Reduced mass

$\rho(\mathbf{r})$	Electrostatic charge distribution (\mathbf{r} = field point)
$\rho_1(\mathbf{x}_1)$	One-electron density function
$\rho_2(\mathbf{x}_1, \mathbf{x}_2)$	Two-electron density function
Φ	Total mutual potential energy
$\phi(\mathbf{r})$	Electrostatic potential (\mathbf{r} = field point)
χ	Electronegativity; basis function in LCAO models
$\Psi(\mathbf{R}, t)$	Time-dependent wavefunction
$\Psi(\mathbf{R}_1, \mathbf{R}_2, \dots)$	Many-particle wavefunction
$\psi(\mathbf{r})$	Orbital (i.e. single-particle wavefunction)
ω	Angular vibration frequency
$\omega_e x_e$	Anharmonicity constant

Index

References to figures are given in italic type; references to tables are given in bold type

- Alkanes, quantitative structure-activity relationships (QSAR) **89**
- AMBER 82
- Amino acids 84
- Angular momentum 209–10, 397–8
and spin 209–10
summary 392–4
- Aromatic hydrocarbons 299
- Atomic beams 208
- Atomic charges 86–7
- Atomic models
Bohr 192–3
Hartree 229–30, 229–30
Hartree-Fock 230–1, 230–1
hydrogen 191–2
infinite nucleus approximation 194
- Atomic orbitals 197, 200–5
d-orbitals 205
- Atomic shielding constants 231–4
- Atomic spectra 192
- Atomic units 195
- Austin Model 1, 323
- Ball-and-spring models
basic principles 38–40
complex 49–51
more than 2 balls 51–2
quantum mechanical 165–8
- Balmer formula 192
- Basis sets
Hartree-Fock model 264–5
for semiempirical models 317–18
- Black body radiation 175–7
- Bohr, Niels 156–7, 192–3
- Bohr (unit of measurement) 195
- Boltzmann counting 188–9
- Boltzmann distribution 95–6, 100–4
conditions for applicability 104–6
- Boltzmann gas law 106–7
- Boltzmann, Ludwig 95–6
- Bonds
breaking 54–5
hydrogen 37–8
stretching 54
- Born, Max 152
- Bose factor 176
- Bosons 187
- Bound states 163
- Brillouin theorem 300–1
- Cambridge Crystallographic Data Centre (CCDC) 85
- Cambridge Structural Database (CSD) 85
- Canonical ensemble 114–15
Monte Carlo simulations 126
- Cartesian coordinates 381
- Central fields 197
- Chance 97–9
- Charge-dipole interactions 23–5
- Clausius-Mossotti equation 367
- Clausius' virial energy 120–1
- Clementi-Raimondi basis sets, HF model 264–5
- Cluster operator 330
- CNDO model 318–20

- Combination rules 31–2
- Conductors 177–8
- Configuration interaction (CI) 245, 296–7, 329–30
- Conformational studies, molecular dynamics 148–50
- Conjugate gradient method, minima determination 71–2
- Constant-temperature molecular dynamics 148
- Continuum approximation 171
- Correlation energy 231
- Coulomb interactions 3–4
 - in molecular mechanics 57
 - QM operator in HF-LCAO 256
- Coulomb's law 3–4
- Coupled cluster method 330–1
- Cubic lattice 138

- d-orbitals 205
- Databases 85
 - amino acid residues 84
- Density differences 243
- Density functional theory 337–46
 - numerical integration 343
 - worked example 345–6
- Density of states 171
- Determinants 388–9
- DFT *see* Density functional theory
- Dice 97–9, **100**
- Dielectric constant 3, 16–17
- Dielectric materials 16–17
- Dielectric polarization 16, 366
- Dielectric slab 16
- Differentiation, fields 384–5
- Dihydrogen, *see* Hydrogen; molecular
- Dipole-dipole interaction, repulsion 29–31
- Dipoles, electric 10–12
 - charge-dipole interactions 23–5
 - dipole-dipole interactions 25–6
 - repulsion 29–31
 - induction energy 27–8
 - polarization 17–18
- Dipoles, magnetic 107–8
- Dirac equation 211–12
- Dispersion energy, electrical 28–9
- Dreiding force field 60

- Effective core potential (ECP) 324
- Eigenvector following 356–7

- Electric charges
 - Coulomb interactions, *see* Coulomb interactions
 - Dipole, *see* Dipoles, electric
 - many-body problems 19–20
 - point charges, *see* Point charges
- Electric fields 5
 - charge distributions 8
 - gradients, and HF-LCAO 301–3
- Electric quadrupole moment 13
- Electric second moment 13–15
- Electron
 - charge 2
 - configuration interaction (CI) 245, 296–7, 329–30
 - configurations 102
 - Dirac treatment 210–212
 - excited states 296–300
 - orbitals, *see* Orbitals
 - spin 185–7
- Electron correlation
 - coupled cluster method 330–1
 - density functions 327–8
 - Fermi correlation 329
- Electron density function 327–9
- Electronegativity 86–7
- Electronic equation 197
- Electrostatic fields 5
- Electrostatic potential 15–16
 - calculation using HF-LCAO 283–5
- Electrostatics 2–3
- Energy levels, transition states 355–7
- Energy minima
 - characterisation 68
 - conformation searches 82–4
 - determination 68–9
 - choice of method 76
 - derivative methods 70
 - first order 70–2
 - second order 72–6
 - direct use of Cartesian coordinates 78–80
 - Hellmann Feynman theorem 290
 - Newton-Raphson method 72–5
 - Z-matrix 77–8
 - in HF-LCAO 285–7
 - local and global minima 65
 - multiple 65–6
 - multivariate grid search 69–70
 - saddle points 66–8

- Ensembles 114–15
- Entropy 117
- Enumeration 99, **100**
- Ethene 350
 - geometry optimization **334**
- Euler angles 132–3
- Exchange energy 231
- Exchange operator 256
- Exchange potential 337–8
 - Pauli and Thomas-Fermi models 338–40
- Excited states 296–300
- Exclusion principle 187–8, 226–8
- Exp-6 potential 29–30
- Experimentally determined parameters
 - electrical interactions 32–3
 - pair potentials 33

- Fermi energy 178, 179
- Fermions 187
- Field lines 5
- Fletcher-Powell algorithm 75–6
- Force fields, molecular mechanics 58–63
- Four-index transformation 297
- Functional 341

- G1 theory 349–51
- G2 theory 351–2
- Games of chance 97–9
- Gases
 - diatomic 293–4
 - imperfections 32
 - monatomic 292–3
 - quantum, overview 169–70
- Gaussian 03 software 275
- Gaussian-type orbital (GTO) 267–70
- GaussView 280–1
 - visualization 281
- Geometry optimization
 - HF-LCAO 285–7
 - molecular mechanics (MM) 81–2
- Gibbs, Josiah Willard 95–6
- Global minimum, definition 65
- Gradient vector 385–6
- Group orbitals 313

- Hamiltonians 118, 152
- Hard spheres, in molecular dynamics 140
- Hartree-Fock limit 237
- Hartree model 229–30
- Hartree (unit of measurement) 195
- Hartree units 195
- Heat bath 108
- Heitler-London dihydrogen model 244–5
- Helium atom
 - electron states **218**
 - Hartree model 229–30
- Hellmann Feynman theorem 290
- Hermite polynomials 166–8
- Hessian matrix 52, 356
- HF (Hartree-Fock) model
 - atomic shielding constants 231–4
 - basis sets 264–6
 - exchange potential 338
 - Hamiltonian matrix 258
 - Kohn-Sham equations 347
 - polarization functions 266
- HF-LCAO
 - basic equations 256–8
 - basis sets 264–6
 - computation 275–6
 - determination of molecular properties 282–5
 - and electric field gradients 301–3
 - examples, L-phenylalanine 275–279
 - Gaussian orbitals 267–70
 - geometry optimization 285–7
 - open shell systems 260–2
 - output 277–80
 - and particle indistinguishability 291–2
 - perturbation methods 331–5
 - thermodynamic properties 291–2
 - vibrational analysis 287–90
- Hoffmann, Roald 314
- Hohenberg-Kohn theorem 340–1
- Hooke's law 39
- Hückel π -electron theory 308–11
 - extension 311–14
 - organic molecules 314
- Hybrid models, quantum mechanics 374
- Hydrogen
 - bonding 37–8
 - molecular 238–50
 - bond population analysis 248
 - Heitler-London treatment 244–5
 - James-Coolidge treatment 247
 - molecular orbital treatment 246

- Hydrogen atom 191–2
 Bohr model 192–3
 p-orbitals 204
 Schrödinger equation 195–7
 s-orbitals 200–2
 spectra 192
- HyperChem software 82–4
 Hyperfine interactions 303–5
 Hyperpolarizability 18
- IMOMOMM model 375–6
 IMOMO model 376
 Importance sampling 128–9
 Incore calculations 276
 Indistinguishability 181–4
 and HF-LCAO 291–2
 and particle counting 188–9
- INDO model 320–1
 Induction energy, electrical 27–8
 Input-output bound calculations 276–7
 Integrated ab initio + molecular mechanics
 model (IMOMOMM) 375–6
 Integrated ab initio + molecular orbital model
 (IMOMO) 376
- Intermolecular forces
 Lennard-Jones potentials 30, **30**, 31–2, 34,
 142
 overview 21
 pair potential 21–3
 pair potential *see* Pair potential
- Intrinsic reaction coordinate (IRC) 359
- Kinetic energy, distribution, quantum gases
 173–5
- Kohn-Sham equations 342
 Kohn, Walter 337
 Koopmans' theorem 260
- Lagrange method of undetermined multipliers
 103–4
- Langevin dynamics 364–6
- LCAO model
 many-electron systems 249–50
 molecules 240–2
 Roothaan treatment 253–5
see also HF-LCAO
- Leapfrog algorithm 142–3
 Lennard-Jonesium 142
 Lennard-Jones potential 30, **30**, 34, 142
 combination rules 31–2
- Lindemann's theory of melting 109–10
- Linear combination of atomic orbitals (LCAO)
 model, *see* LCAO model
- Linear synchronous transit algorithm 356
- Linear variation method 223–6
- Link atoms 374–5
- Linnett potential 46–7
- Liquids, molecular dynamics 137–9
- Local minimum, definition 65
- Magnetic resonance effect 303–4
- Matrices 389–94
- Maxwell-Boltzmann distribution 173–5
- Mean 98–9
- Melting 109–11
- Metals
 Drude model 178–9
 Pauli treatment 180–1
 quantum mechanical modelling 177–81
- Metropolis Monte Carlo model (MMC) 128–9
- MINDO model 321–2
- Minima, *see* Energy minima
- MMC (Metropolis Monte Carlo) model 128–9
- MNDO model 322–3
- Molecular beams 32–3
- Molecular dynamics
 conformational studies 148–50
 molten salts 144
 pair correlation function 139
 rescaling 148
 time-dependent evolution 142–4
 Verlet algorithm 143
- Molecular mechanics (MM) 53–4
 bond stretching 54–5
 Coulomb interactions 57
 dihedral motion 55
 force fields 58–9
 AMBER 61–2
 Dreiding 60
 Johnson 62–3
 MM1 60–1
 MM2 61
 OPLS 62
 UFF 63
 geometry optimization 81–2
 hybrid models 375–6
 nonbonded interactions 56–7
 out-of-plane angle potential 56
 and quantum mechanics 373–6
 time-saving 57–9

- Molecular orbital model
 dihydrogen 246
 Roothaan treatment 254–5
- Molecular partition function 291–2
- Molecules
 challenges in modelling 237
 diatomic, vibration 41–3
 extension of LCAO model to many-electron systems 249–50
 hydrogen 238–40
 LCAO model 240–2
 quantum mechanics, general considerations 237
- Møller-Plesset perturbation theory 331–5
- Moment, electric 10
- Monte Carlo simulations
 cut-offs 131–2
 development 127–8
 flexible molecules 133
 overview 123–4, 125–6
 periodic box 129–31
 rigid molecules 132–3
- Morse potential 45–6
- Multipole expansion 23
- Multipoles, electric 10–15
- Mutual potential energy
 in electric field 8–9
 and force 9–10
 and force 41
- NDDO model 321
- Net atom and bond population 248
- Newton-Raphson method 355
 minima determination 72–5
- Nonbonded interactions, cutoffs 58–9
- One-electron density function 328
- ONIOM 377
- Onsager model 368–9
- Orbitals 197
 atomic
 d-orbitals 205
 p-orbitals 204
 s-orbitals 200–2
 Gaussian-type 267–70
 group 313
 molecular hydrogen 240–2
 multiple-electron atoms 217–18
 Slater-type 270–2
- p-orbitals 202–204, **204**
- Pair correlation function 139
- Pair potential 21–2
 improved modelling 33
 liquid water 147
 multipole expansion 23
 numerical form 33–4
 water 145
- Pair wise additivity 4
- Pariser, Rudolf 314–15
- Parr, Robert 314–15
- Partition function, Boltzmann gas law 106–7
- Pauli exclusion principle 187–8, 226–8
- PDB file format 85–6, 275–7
- Permittivity 3
- Perturbation methods 219–21
 electron correlation 331–5
- Phase space, thermodynamic 117–18
- L-phenylalanine 275–7, 282, 283–5, 295–6
- Planck, Max 175–6
- PM3 model 323
- Point charges 1–3
 additivity 4
 Coulomb's law 3–4
 di- and multipoles 10–12, 11
 multiple 19–20
- Polarizability 87–90
- Polarization
 dielectric 16, 366
 dipole 17–18
 electrical 16–18
 in HF model 266
- Pople, John 337–8
- Population analysis 246–9
- Potential energy
 in electric fields 8–9
 electrostatic 15–16, 283–5
 Lennard-Jones potential 30, 34, 142
 pair potentials, *see* Pair potential
 between point charge and electric field 9
- Potential well
 finite 162–4
 infinite
 one-dimensional 154–6
 three-dimensional 159–60
 two-dimensional 157–8
 one-dimensional 181–4
- Pressure 113–14
- Primary molecular properties 282–3

- Probability 97–104
 - Boltzmann probability 100–4
- Protein Data Bank files 275–7
- Proteins 374–5
- Proton, charge 2
- Pyridine 310–311

- Quadratic synchronous transit algorithm 356
- Quadrature, density functional theory 343
- Quadrupole coupling constant 302
- Quantitative structure-activity relationship (QSAR) 86–92
- Quantum mechanics
 - correspondence principle 156–7
 - hybrid models 374
 - indistinguishability 181–4
 - and many-body problems 216–17
 - measurement 212–13
 - methods of approximation 219–26
 - Pauli exclusion principle 187–8
 - potential wells 154–6
 - quantization
 - angular momentum 205–8
 - of energy levels, and vibration 43
 - Schrödinger equation 151–152
 - see also* Schrödinger equation
 - spin 185–7
- Quaternions 133

- Radial distribution function, molecular dynamics 136–9
- Radial function 197–200
- Randomness 96–7
- Rayleigh counting 171, 172–3, 172
- Rigid molecules, Monte Carlo simulations 132–3
- Rigid sphere potential 127, 127
- Rotational temperature 294
- Russell-Saunders scheme 218
- Rydberg constant 192

- s-orbitals 200–2
- Sackur-Tetrode equation 293
- SAM1 model 323
- Sample size
 - molecular dynamics 142
 - and probabilistic fluctuation 104–6
- Scalar field 383
- Scalar product 382
- Scalar quantities 379–80

- SCF, multiconfiguration 230, 335
- Schrödinger equation 151–2
 - approximate solutions
 - linear variation method 223–6
 - perturbation method 219–21
 - variation method 219–21
 - hydrogen atom 195–7
 - infinite potential wells
 - one-dimensional 154–5
 - two-dimensional 158
 - three-dimensional 159
 - many-electron atom 217–18
 - molecular hydrogen 239
 - time-independent 153–4
 - two noninteracting particles in
 - one-dimensional well 161–2
 - unbound particles 164–5
- Schrödinger's cat paradox 213
- Self-consistent field (SCF) 230, 335
- Semiempirical models
 - CNDO 318–19
 - CNDO/2 319–20
 - CNDO/S 320
 - INDO 320–1
 - MINDO 321–2
 - NDDO 321
 - overview 307–15
- Silberstein, L. 90
- Simple harmonic motion 39, 40, 41
- Sire-site potential 34
- Slater-Condon-Shortley rules 228–9, 255
- Slater determinants 226–8
 - Slater-Condon-Shortley rules 228–9
- Slater-type orbitals, fitting to Gaussian-type orbitals 270–2
- Solvents 57, 363–4
 - continuum models 366–8
 - Onsager model 368–9
 - Langevin dynamics 364–6
 - polarizable continuum models 370
- Spin 107–8, 208–9, 208–10
 - and electric field gradients 301–3
 - Slater-Condon-Shortley rules 228–9
 - Slater determinants 226–8
 - Stern-Gerlach experiment 205–8
- Square well potential 140–1
- Standard deviation 96
- Steepest descent method, determination of energy minima 70–1

- Stern-Gerlach experiment 205–8
 and measurement problem 212
- Surface area 90–1
- Surroundings, definition 6
- System, definition 6
- Temperature, and dipole-dipole interaction 26–7
- Thermal motion, and dipole-dipole interaction 26–7
- Thermodynamics
 configurational integral 119–20
 direct calculation 125
 ensembles 114–15
 entropy 117
 G1 theory 349–51
 G2 theory 351–2
 G3 theory 352–3
 Helmholtz energy 116, 117
 HF-LCAO computation 291–5
 internal energy 116
 overview 113–14
- Thomas-Fermi relation 339
- Time-independent Schrödinger equation 153–4
- Transition states 355–6
 example 357–8
 LST algorithm 356
 reaction pathways 359–61
- Two-electron operator, definition 216–17
- UHF model 262–4
- Unbound states 163, 164
- Unrestricted Hartree-Fock model 262–4
- Valence theory 37
- Variation method 221–3
- Velocity Verlet algorithm 143
- Verlet algorithm 143, 365
- Vibrational energy, diatomic gases 294–5
- Vibrational motion
 Linnett potential 46–7
 Morse potential 45–6
 normal modes of vibration 50–1
 overview 38–40
 quantization 43–5
 variable choice 53
 see also Ball-and-spring models
- Vibrational quantum number 43
- Vibration, HF-LCAO analysis 287–90
- Virial energy 120–1
- Visualization, HF-LCAO 281
- Volume 90–1
- Water
 molecular dynamics 144–8
 as solvent 370–1
- Wave function
 molecular nitrogen 167
 two-dimensional potential well 158
- Wilson B-matrix 52
- Wolfsberg, Max 312–14
- Work, and electric fields 7–8
- $X\alpha$ model 340
- Zener wavefunctions 232–4
- Zero differential overlap (ZDO) 315–17
- Zero point energy 166–8
- ZINDO models 323