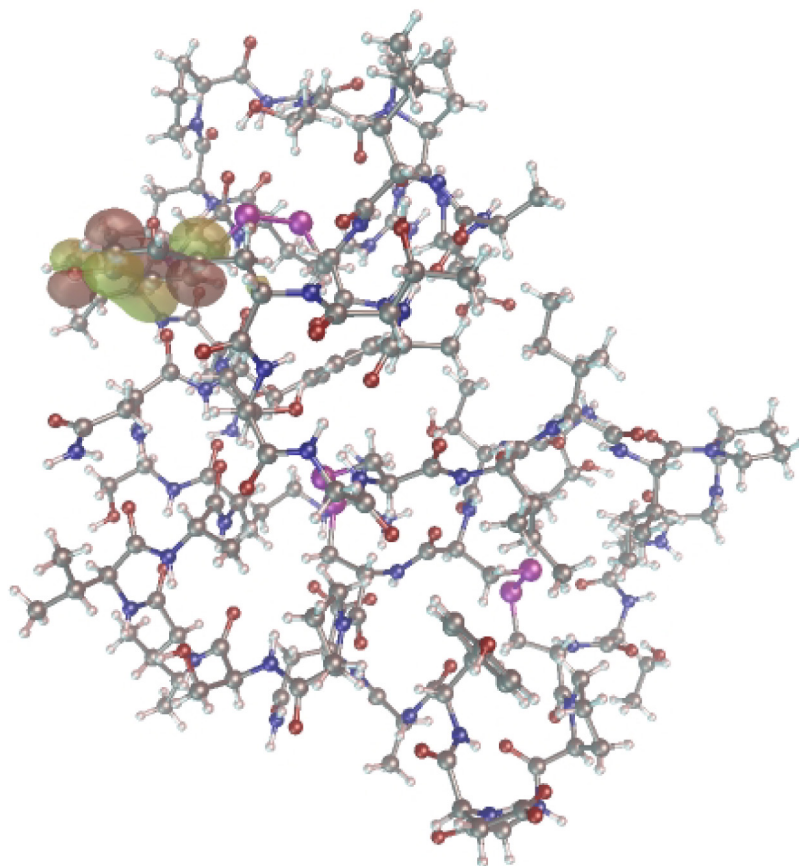


MOPAC 2000 V1.3 User's Manual

James J. P. Stewart



FUJITSU

Vertrieb durch:
ADDITIVE GmbH • Max-Planck-Straße 22b • 61381 Friedrichsdorf
<http://www.additive-net.de/software> • eShop: <http://eshop.additive-net.de>
Verkauf: +49-6172-5905-30 chemie@additive-net.de
Support: +49-6172-5905-20 support@additive-net.de

ADDITIVE
Soft- und Hardware für Technik und Wissenschaft

The contents of this manual may be revised without prior notice
All Rights Reserved, Copyright © FUJITSU LIMITED 2000

Preface

Purpose of this Manual

This manual describes how to use MOPAC 2000 (hereinafter called MOPAC), and includes background material on the theoretical techniques used in MOPAC.

Intended Readers

This manual is written for users of MOPAC.

The reader should have a knowledge of the basic concepts of quantum mechanics or molecular modeling.

Structure of this Manual

The manual has the following format:

Chapter 1: Introduction

An overview of MOPAC is given. The relationships of programs which can be used with MOPAC are described.

Chapter 2: Acknowledgments

In this chapter, contributors to MOPAC are recognized.

Chapter 3: Data files for MOPAC

This chapter describes how to construct data sets for MOPAC, and gives definitions for the keywords used.

Chapter 4: Examples

Examples of some types of calculations are described.

Chapter 5: Theory

In this chapter, the theoretical background to several parts of MOPAC is described.

Chapter 6: Accuracy of Methods in MOPAC

This chapter gives a quantitative description of the accuracy of MNDO, AM1, and PM3. The protocols used in minimizing bias are described.

Chapter 7: Program

This chapter describes the structure of the program MOPAC.

Chapter 8: Error Messages produced by MOPAC

In this chapter, a brief description is given of some of the error messages generated by MOPAC.

Chapter 9: Criteria

In this chapter, the degrees of precision used in various parts of MOPAC are described.

Chapter 10: Debugging

This chapter contains notes on how to locate faults in MOPAC.

Appendix A: Reference Heats of Formation

This appendix gives reference heats of formation for compounds of every element in each method in MOPAC. This is useful for validating parameter sets.

Appendix B: Bibliography

This appendix contains a short essay giving references to much of the material used in MOPAC.

How to Read this Manual

- Software developers should read Chapters 5 and 7.
- Users of MOPAC should read Chapters 3, 5, and 8.
- People interested in the accuracy of the methods in MOPAC should read Chapter 6.
- People interested in the theory used in MOPAC should read Chapter 5.

Precautions Concerning this Manual

MOPAC is a research program, and both it and the manual represent the current state of the science of semiempirical computational chemistry. This is a rapidly evolving field, and while care has been taken to ensure a high quality product, the fast pace of development means that many errors may exist in both the software and in the manual.

Users of MOPAC are also cautioned that the manual concentrates on how to use the program. The manual does not describe in detail either what MOPAC can be used for, or the difficulties which can occur in specific types of jobs.

Contents

1	Introduction	1
1.1	What is MOPAC?	1
1.1.1	New Functionalities of MOPAC	1
1.1.2	Technical and Other Changes	2
1.1.3	Cooperative Development of MOPAC	3
1.2	Graphical User Interfaces	4
1.3	Use of MOPAC as an Educational tool	4
1.4	A General Description of MOPAC	5
1.4.1	Validation	5
1.4.2	Geometry Optimization	6
1.4.3	Methods	7
1.4.4	SCF Procedures	7
1.4.5	Derived Properties	8
1.4.6	Graphics	9
1.4.7	Types of Species	9
1.4.8	Giant Molecules	10
1.4.9	Symmetry theory	12
1.4.10	Electric Fields	12
1.4.11	Electrostatic Potential	12
1.4.12	Solvent effects	13
1.4.13	Electronic Excited states	13
1.4.14	Intersystem Crossing	13
1.4.15	Vibrations	13
1.4.16	Thermodynamics	14
1.4.17	Molecular Dynamics	14
1.4.18	Reaction Paths	15
1.4.19	Saddle-Point Location	15
1.4.20	Data checking	16
1.4.21	Restarts	16
1.4.22	Portability	16
1.4.23	Program Structure	16
1.4.24	Dynamic Memory Allocation	16
1.4.25	Citation for MOPAC 2000	16
2	Acknowledgements	17
2.1	Contributors to MOPAC	17

3	Data files	19
3.1	Introduction	19
3.2	The Data-Set	19
3.2.1	Example of Data for Ethylene	19
3.2.2	Overview of Data-Set	21
3.3	Keywords	21
3.3.1	Specification of Keywords	21
3.3.2	Full List of Keywords	21
	Limitations	57
	Recommended use of MOZYME	58
3.3.3	Keywords that go together	85
3.4	Geometry Specification	85
3.4.1	Definition of Elements and Isotopes	86
3.4.2	Optimization flags	89
3.4.3	Internal Coordinate Definition	89
	Constraints	90
	Conversion to Cartesian Coordinates	90
3.4.4	Translation Vector	90
3.4.5	Gaussian Z-matrices	91
	Specification of Gaussian Z-matrices	91
	Exceptions to the full Gaussian standard	91
3.4.6	Cartesian Coordinate Definition	92
3.4.7	Protein Data Bank Format	92
	Preparing the data set	93
3.4.8	Conversion Between Various Formats	93
3.4.9	Examples of Coordinate Definitions	94
	Formic acid	94
3.5	Memory Considerations	94
4	Examples	97
4.1	To verify MOPAC	97
4.2	Excited states	97
4.2.1	Discussion of output from <code>meci.dat</code>	98
	The methyl radical	98
	Oxygen	98
	Methane	98
	Methane Cation	98
4.3	Simple geometry optimization	99
4.4	Highly excited methane dication	99
4.5	Polytetrahydrofuran	100
4.6	Diamond	100
5	Theory	105
5.1	Introduction	105
5.2	Semiempirical Theory	105
5.2.1	Approximations used in MNDO, MINDO/3, AM1, PM3, and MNDO- <i>d</i>	106

Basic Roothaan-Hall Equations	107
Neglect of diatomic overlap integral	107
Neglect of three and four center integrals	108
One-center two-electron integrals	109
NDDO two-electron two-center integrals	109
Final Assembly of Two-Electron Two-Center Integrals	112
MINDO/3 Two electron two center integrals	112
The one-center one-electron integral $H_{\mu\mu}$	113
The two-center one-electron integral $H_{\mu\nu}$	113
Core-core repulsion integrals	113
MINDO/3 modification to the core-core term	114
MNDO modification to the core-core term	114
AM1 and PM3 modifications to the core-core term	114
5.2.2 Self-Consistent Field Calculation	114
5.2.3 Calculation of ΔH_f	115
5.3 Miscellaneous Topics in Semiempirical Theory	115
5.3.1 Koopmans' Theorem.	115
5.3.2 Dipole moments.	116
5.3.3 Bond Orders	116
Atomic bond index	117
Anisotropy	117
Bond order	117
5.3.4 Mulliken populations	118
5.3.5 Localized orbitals	120
Eigenvectors	120
Localized M.O.s	120
Localization Theory	120
5.3.6 Outer Valence Green's Function	121
5.3.7 Using <i>ab initio</i> derivatives	122
5.3.8 Correction to the Peptide Linkage	122
Cautions	124
5.3.9 Convergence in SCF Calculation	124
Causes of failure to achieve an SCF	125
5.3.10 Use of C.I. in Reaction Path Calculations	125
5.3.11 Sparkles	127
5.3.12 Capped Bonds	128
5.3.13 Overlap Integrals	129
Rotation of Atomic Orbitals	131
5.3.14 Energies of Isolated Atoms	137
5.4 Multi-Electron Configuration Interaction	138
5.4.1 Starting electronic configuration	140
5.4.2 Microstates	141
Permutations	142
Energy of microstates	143
Zero of energy used in MECI	144
5.4.3 Construction of secular determinant	144

5.4.4	Atom Transition Moments	146
	Evaluation of Transition Dipole	147
5.4.5	States arising from various calculations	149
5.4.6	Spin angular momentum	150
	Calculation of spin-states	151
5.4.7	Choice of State to be Optimized	152
	Quantum Numbers	153
	Polarizability	153
	Franck-Condon considerations	153
5.4.8	Definition of some C.I. Keywords	154
5.4.9	Degenerate States	155
5.4.10	Calculation of Spin Density	155
5.5	An SCF Calculation	155
5.5.1	Starting density matrix	157
5.5.2	Assembly of the starting Fock matrix	157
5.5.3	Diagonalization of the Fock matrix	158
5.5.4	Exercises involving eigenvectors	158
5.5.5	Iterating density matrix	159
5.5.6	Iterating Fock matrix	159
5.5.7	Calculation of heat of formation	160
5.6	Localized Molecular Orbital Theory	160
5.6.1	Why Use Localized Molecular Orbitals?	160
	What is a Localized Molecular Orbital?	161
	Energy Considerations.	162
	Limitations on Systems.	162
5.6.2	Memory Management	163
	The Nature of the Problem.	163
	Distance Cutoffs.	163
5.6.3	Lewis Structures.	166
	Lewis Structure—Main Sequence.	166
	Detailed Description of Lewis Structure.	166
5.6.4	Construction of Starting Localized Molecular Orbitals	170
	Construction of Hybrids	171
	Construction of Starting Localized Molecular Orbitals	174
	Construction of Sigma Framework	174
	Identification of Lone Pairs	175
	Construction of Non-Cyclic π Bonds	175
	Construction of Cyclic π Bonds	176
	Identification of Ions	176
	Properties of Starting LMOs	177
5.6.5	Other Considerations	177
	Re-Orthogonalization of LMOs	177
	Localized Molecular Orbitals: Storage Considerations	178
5.6.6	Diagonalization	180
	Conventional Matrix Annihilation	180
	Localized Molecular Orbital Matrix Element Evaluation	181

Localized Molecular Orbital Matrix Element Annihilation	183
5.6.7 Reducing the Size of Localized Molecular Orbitals	184
Size of LMOs	184
5.6.8 Density Matrix Construction	184
5.6.9 Energy Effects of CUTOF2	185
5.7 Geometry optimization	187
5.7.1 EigenFollowing	187
5.7.2 The BFGS function optimizer	189
5.7.3 Optimization of one unknown	190
5.7.4 Considerations in Geometry Optimization	191
Overriding the default options	191
Locating Transition States	191
5.8 "Size" of a molecule	193
5.9 Solid state capability	193
5.9.1 Constructing Data Sets	193
5.9.2 The Cluster	194
5.9.3 Derivatives	194
5.9.4 Geometry Specification for Band Structure Calculations	195
5.9.5 Electronic Band Structure	195
5.9.6 Electronic Density of States	195
5.9.7 Brillouin Zone: Generation of Band Structures	196
Space-group operations	197
5.9.8 Examples of Solid State Data Set Construction	198
Diamond	198
5.10 Point Group Theory	201
5.10.1 Representation of Point Groups	201
5.10.2 Identification of Point-Groups	202
5.10.3 Orientation of the Abelian groups C_{2v} and D_{2h}	204
5.10.4 Molecular Orbitals	204
5.10.5 Normal Coordinates	205
5.10.6 States	205
5.11 Level of Precision within MOPAC	207
5.11.1 Fundamental Physical Constants	208
5.11.2 Various precision levels	208
5.11.3 Reasons for low precision	210
5.11.4 How large can a gradient be and still be acceptable?	210
5.11.5 Convergence tests in subroutine ITER	211
Self-consistency test	211
5.12 Torsion or Dihedral Angle Coherency	211
5.13 Gradients	211
5.13.1 Frozen density matrix finite difference derivatives	212
5.13.2 Hessian matrix in FORCE calculations	213
5.14 Normal Coordinate Calculation	214
5.14.1 Calculation of Vibrational Frequencies	214
5.14.2 Mechanism of the frame in FORCE calculation	215
5.14.3 Vibrational Analysis	215

5.14.4	Reduced masses	215
5.14.5	Effective masses	217
5.14.6	Travel	217
5.15	A Note on Thermochemistry	218
5.15.1	Basic Physical Constants	218
5.15.2	Thermochemistry from <i>ab initio</i> MO methods	218
	Vibrational terms	219
	Rotational terms	219
	Translational terms	220
5.15.3	Thermochemistry in MOPAC	221
	Vibration	222
	Rotation	222
	Translation	222
5.16	Force Constants	223
5.17	Reaction paths	224
5.18	Grid Calculation	226
5.19	Dynamic and Intrinsic Reaction Coordinates	226
	Period of vibration	233
	Initial dynamics of N ₂ with N–N distance = 1.094 Å	234
	Conservation of normal coordinate	235
	Rate of decay of starting mode	235
	Half-Life for decay of initial mode	235
	DRC print options	235
5.20	Use of SADDLE Calculation	236
5.20.1	How to escape from a hilltop	238
5.21	Polarizability and Hyperpolarizability Calculation	239
5.21.1	Time-Dependent Hartree-Fock	239
5.22	COSMO (Conductor-like Screening Model)	241
5.22.1	A Walk Through COSMO	241
5.22.2	COSMO Keywords	244
5.22.3	Solvent Accessible Surface	244
	Some hints on the use of COSMO	245
5.23	Parametric Molecular Electrostatic Potential (PMEP)	245
5.23.1	2-D Electrostatic Potential Plots	246
	Choice of Plane to be Calculated	247
5.23.2	Atomic Charges	247
5.24	Miertus-Scrocco-Tomasi Solvation Model	247
5.24.1	Outline of the MST Method	247
5.24.2	Data Requirements for MST Model	248
6	Accuracy of Methods in MOPAC	251
6.1	Protocols used in Determining Accuracy	251
6.2	Structure of the Tables	252
6.3	Limitations of MNDO	252
6.4	Accuracy of AM1	253
6.5	Accuracy of PM3	253

6.6	Faults and Errors in PM3	254
6.7	Comparison of Calc'd and Observed ΔH_f for MNDO, AM1, and PM3	279
6.8	Comp. of Calc'd and Exp'l Geometries for MNDO, AM1, and PM3	309
7	Program	341
7.1	Main geometric sequence	341
7.2	Main electronic flow	341
7.3	Control within MOPAC	341
7.4	Arrays.	344
7.4.1	Array Specification.	344
7.5	Names and Storage of Variables in MOPAC	344
8	Error messages produced by MOPAC	353
9	Criteria	377
9.1	SCF criterion	377
9.2	Geometric optimization criteria	377
10	Debugging	381
10.1	Keywords for use in debugging	381
	Bugs discovered in the electronics subroutines	383
	Bugs discovered in the geometric subroutines	384
A	Reference Heats of formation	385
B	Bibliography	387

List of Tables

3.1	Full list of Keywords used in MOPAC	21
3.2	Number of Configurations used in MECI	35
3.3	Use of C.I.=(n,m)	35
3.4	Use of CHARGE= n	36
3.5	Extra Data Required by MEP=1	54
3.6	Extra Data Required by MEP=2	55
3.7	States Arising from Various Microstates	56
3.8	Use of OPEN(n,m)	62
3.9	Data Required by PMEPR	65
3.10	Abbreviations for the 20 Amino Acids	70
3.11	Internal Coordinate Symmetry Functions	77
3.12	Cartesian Coordinate Symmetry Functions	78
3.13	Elements available within MNDO	86
3.14	Elements available within AM1	86
3.15	Elements available within PM3	87
3.16	Elements available within MNDO- d	87
3.17	Diatomics Parameterized within MINDO/3	88
5.1	Parameters used in Semiempirical Methods	106
5.2	Two-Electron Two-Center Integrals (Local Frame)	110
5.3	Types of Electron Density Distribution	110
5.4	Density Distributions Arising from Pairs of Atomic Orbitals	111
5.5	Additive Terms for Two-Electron Integrals	111
5.6	Comparison of Observed and Calculated Pyramidalization of Nitrogen	124
5.7	Values of C_{lmj}	131
5.8	Angular Dependence of the p -orbitals	132
5.9	Angular Dependence of the d -orbitals	134
5.10	Angular Dependence of the f -orbitals	135
5.11	Angular Dependence of the f -orbitals (cont.)	136
5.12	Two Electron Energy Contributions to EISOL for Atoms in their Ground States	139
5.13	Two Electron Energy Contributions to EISOL for Atoms in their Ground States	140
5.14	Examples of SCF configurations used in MECI	140
5.15	Microstates for 5 electrons in 5 M.O.s	142
5.16	Sets of Microstates for Various MECI Calculations	143
5.17	Number of Microstates for CIS, CISD, and CISDT	144
5.18	" x " Transition Integrals	147
5.19	" y " Transition Integrals	148
5.20	" z " Transition Integrals	149

5.21 States arising from C.I. = n	149
5.22 Examples of Use of C.I. Keywords	155
5.23 Values of $E_{(isol-atom)}$	161
5.24 Times for One SCF Calculation	162
5.25 MOPAC Memory Requirements for Proteins	163
5.26 Bond-Orders in GLY-GLY-GLY-GLY	164
5.27 Multipolar Representation of the Two Electron Two Center Integrals	164
5.28 Distance Cutoffs for Various Types of Interaction	165
5.29 One-Electron Integral Matrix for Ethane	172
5.30 Interaction Matrix used in Constructing Hybrid Orbitals for C ₁ in Ethane	172
5.31 Eigenvectors of Lowest Energy from Interaction Matrix	172
5.32 Localized Diatomic Molecular Orbitals for C ₁ in Ethane	173
5.33 Hybrid Atomic Orbitals for C ₁ in Ethane	173
5.34 Atomic Orbital and Electron Contributions in Benzaldehyde	174
5.35 Unused Electrons and Orbitals after Construction of σ Framework	175
5.36 Unused Electrons and Orbitals after Construction of Lone Pairs	175
5.37 Unused Electrons and Orbitals after Construction of Non-Cyclic π Bonds	176
5.38 Charged Sites in Proteins	176
5.39 Names of Arrays Used to Represent LMOs	179
5.40 Space-group operations for O_h^7 (diamond)	197
5.41 Point Groups available within Symmetry Code	202
5.42 Subset of Group D_{2h}	202
5.43 Fundamental Physical Constants	208
5.44 Table of Symmetry Numbers	219
5.45 Stretching Curve for Nitrogen Molecule	232
5.46 Velocities in DRC for N ₂ Molecule	234
5.47 Modified Velocities in DRC for N ₂ Molecule	234
5.48 Quantities Calculable using POLAR	240
5.49 Calculated and Observed Hydration Energies	242
5.50 Van der Waals radii (Å) used in COSMO	244
6.1 Worst ΔH_f Errors in MNDO	255
6.2 Worst ΔH_f Errors in AM1	256
6.3 Worst ΔH_f Errors in PM3	257
6.4 Average Signed and Unsigned Errors in ΔH_f by Compound Type	258
6.5 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	259
6.6 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	260
6.7 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	261
6.8 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	262
6.9 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	263
6.10 Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)	264
6.11 Average Errors in ΔH_f by Element	265
6.12 Average Errors in Bond-Length by Bond-Type	266
6.13 Average Errors in Bond-Length by Bond-Type (contd.)	267
6.14 Average Errors in Bond-Length by Bond-Type (contd.)	268
6.15 Average Errors in Bond-Length by Bond-Type (contd.)	269

6.16 Average Errors in Bond-Lengths for each Element	270
6.17 Worst Ionization Potential Errors in MNDO	271
6.18 Worst Ionization Potential Errors in AM1	272
6.19 Worst Ionization Potential Errors in PM3	273
6.20 Average Errors (eV) in Ionization Potential by Element	274
6.21 Worst Dipole Moment Errors in MNDO	275
6.22 Worst Dipole Moment Errors in AM1	276
6.23 Worst Dipole Moment Errors in PM3	277
6.24 Average Errors (Debye) in Dipole Moment by Element	278
6.25 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3	280
6.26 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	281
6.27 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	282
6.28 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	283
6.29 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	284
6.30 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	285
6.31 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	286
6.32 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	287
6.33 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	288
6.34 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	289
6.35 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	290
6.36 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	291
6.37 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	292
6.38 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	293
6.39 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	294
6.40 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	295
6.41 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	296
6.42 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	297
6.43 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	298
6.44 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	299
6.45 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	300
6.46 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	301
6.47 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	302
6.48 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	303
6.49 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	304
6.50 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	305
6.51 Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)	306
6.52 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3	311
6.53 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	312
6.54 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	313
6.55 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	314
6.56 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	315
6.57 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	316
6.58 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	317
6.59 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	318
6.60 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	319
6.61 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	320

6.62 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	321
6.63 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	322
6.64 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	323
6.65 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	324
6.66 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	325
6.67 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	326
6.68 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	327
6.69 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	328
6.70 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	329
6.71 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	330
6.72 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	331
6.73 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	332
6.74 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	333
6.75 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	334
6.76 Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)	335
7.1 Constants used in MOPAC	345
7.2 Channels Used in MOPAC	351
8.1 Names of Parameters for use with EXTERNAL=<file>	363
A.1 Reference Heats of Formation	386

List of Figures

3.1	Ethylene	20
3.2	Data set for Ethylene	20
3.3	Configurations used in BIRADICAL Calculation	33
3.4	Use of EXTERNAL= <i>name</i>	43
3.5	List of Covalent Bonds in Isoleucine	46
3.6	Standard MOPAC-type Data Set for Methane	49
3.7	Example of Output Generated by HYBRID	49
3.8	One Electron Matrix for Methane	49
3.9	Calculation of ESP Cross-Section in Formaldehyde using the Orozco-Luque Model	54
3.10	Molecular Electrostatic Potential around Formaldehyde	55
3.11	Calculation of ESP Connolly Surface in Formaldehyde using the Orozco-Luque Model	55
3.12	Example of Output Generated by SIZES	74
3.13	Ethane, showing use of SYMMETRY	79
3.14	Example of Cartesian SYMMETRY functions	80
3.15	Example of Parameter Set for PM3(tm)	87
3.16	Example of Gaussian Z-matrix geometry specification	92
3.17	Data set for Formic acid	94
4.1	Example of a Simple Geometry Optimization	99
4.2	Example of a Simple Geometry Optimization (2)	99
4.3	Example of an Electronic Excited State	100
4.4	Example of a One-Dimensional Polymer	101
4.5	Example of the use of MERS	102
4.6	Example of a Three-Dimensional Data-set	103
5.1	Example of the use of AIDER	123
5.2	Effect of C.I. on ΔH_f in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)	126
5.3	Effect of C.I. on the charge in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)	126
5.4	Mathematical Tools for use with Spherical Harmonics	132
5.5	Energies of Excited States	152
5.6	Examples of Initial Choice of Double Bond	167
5.7	Generation of Yne Bond	168
5.8	Kekule Structure for a Facet of Fullerene C_{1500}	169
5.9	Ethane, Showing Orientation and Numbering System	171
5.10	Electronic Structure of Benzaldehyde	174
5.11	Secular Equation used in Constructing Starting LMOs	175
5.12	Storage of Starting LMOs in Methane	179
5.13	Storage of Starting LMOs in Methane, Showing Space Available for Expansion	179

5.14	Storage of Iterating LMOs in Methane	180
5.15	Representation of Symmetry Operations	202
5.16	Example of Spurious Stationary Point	209
5.17	Example of an S_{N^2} reaction path calculation	224
5.18	Example of a rotation barrier calculation	224
5.19	Data set to stretch a polymer	225
5.20	Example of a GRID Calculation	226
5.21	Example of DRC calculation	230
5.22	Example of IRC calculation	230
5.23	Example of DRC calculation, monitoring a geometric variable	236
5.24	Example of data for SADDLE calculation	237
5.25	Data Set for PMEP Calculation of Formaldehyde	246
5.26	Parametric Molecular Electrostatic Potential around Formaldehyde	246
5.27	Data set for Calculation of Water using the Miertus-Scrocco-Tomasi Model	248
7.1	Diagram of Main Geometric Sequence in MOPAC	342
7.2	Diagram of Main Conventional Electronic Sequence in MOPAC	343

Chapter 1

Introduction

1.1 What is MOPAC?

MOPAC is a general-purpose semiempirical molecular orbital package for the study of solid state and molecular structures and reactions.

The semiempirical Hamiltonians MNDO [1], MINDO/3 [2], AM1 [3], PM3 [4], and MNDO-*d* [5, 6] are used in the electronic part of the calculation to obtain molecular orbitals, the heat of formation and its derivative with respect to molecular geometry. Using these results MOPAC calculates the vibrational spectra, thermodynamic quantities, isotopic substitution effects and force constants for molecules, radicals, ions, and polymers. For studying chemical reactions, a transition state location routine [7] and two transition state optimizing routines [8, 9, 10] are available. For users to get the most out of the program, they must understand how the program works, how to enter data, how to interpret the results, and what to do when things go wrong.

While MOPAC calls upon many concepts in quantum theory and thermodynamics and uses some fairly advanced mathematics, the user need not be familiar with these specialized topics. MOPAC is written with the non-theoretician in mind. The input data are kept as simple as possible, so users can give their attention to the chemistry involved and not concern themselves with quantum and thermodynamic exotica.

The simplest description of how MOPAC works is that the user creates a data-file which describes a molecular system and specifies what kind of calculations and output are desired. The user then commands MOPAC to carry out the calculation using that data-file. Finally the user extracts the desired output on the system from the output files created by MOPAC.

The name MOPAC should be understood to mean “Molecular Orbital PACKage”. The origin of the name is somewhat unusual, and might be of general interest: The original program was written in Austin, Texas. One of the roads in Austin is unusual in that the Missouri-Pacific railway runs down the middle of the road. Since this railway was called the MO-PAC, when names for the program were being considered, MO-PAC was an obvious contender.

1.1.1 New Functionalities of MOPAC

Addition of *d*-orbitals

In 1992, W. Thiel and A. A. Voityuk published a modified MNDO method [5] in which *d* orbitals were added to the conventional *s* – *p* basis set. The result of this was a new method, MNDO-*d*, which was more accurate than any other NDDO method. For example, using the *s* and *p* basis set only, none of the popular semiempirical methods correctly predict the point group for ClF_3 , whereas when *d*-orbitals are added, the correct point group is obtained. At present, parameters for only a few main-group elements are available, namely:

Al, Si, P, S, Cl, Br, and I.

Calculations using MNDO-*d* can be run by use of the new keyword MNDOD. In accordance with the description of the method, for elements for which MNDO-*d* parameters are not yet available, MNDO parameters are used by default. MNDO-*d* cannot be used with the solvation methods at present.

Parameters are available for the following transition metals: Fe, Cu, Ag, Mo, and Pt, within AM1. When other transition metal parameter sets become available they can be used in a MOPAC calculation using the keyword EXTERNAL.

The PM3(tm) method is not available by default in MOPAC, but can easily be added by use of the EXTERNAL keyword. See page 86 for details.

Excited States in Solution

The COSMO method has been extended [11] to allow calculation of electronic excited states in solution. Because the refractive index is needed in the calculation, it must be supplied in the data set. Accordingly, the keyword `N=n .nn` has been defined. Examples of the use of the COSMO method are provided in the data sets distributed with MOPAC.

Extension of the Tomasi Model of Solvation

In earlier versions of MOPAC containing the Tomasi model, a large amount of data defining the solvent had to be supplied. Three new keywords have now been defined, which automatically provide default values for common solvents. The keyword for water is `H2O`. Recently, the Tomasi method has been extended to organic solvents. As a result, two more keywords are provided, `CHCL3` (for chloroform) [12] and `CCL4` (for carbon tetrachloride) [13].

Dynamic Memory Allocation

Most of the arrays that depend on the size of the system are now allocated dynamically. Previous versions of MOPAC used static memory allocation, and in order to efficiently run calculations on systems of different sizes, two or three different versions of MOPAC had to be available. With dynamic memory allocation, a single executable can be used for small, medium, and large systems.

Not all arrays can be made dynamic. When MOPAC starts, it does not know how much memory will be needed for the geometry, so a few static arrays are needed. However, these are relatively small. Because of this, the default number of atoms can be set quite large.

Linear Scaling

A new method, called MOZYME, allows systems of several thousand atoms to be run in a practical time. The method uses localized molecular orbitals, and gives results that are the same as those of conventional methods. MOZYME should be regarded as an alternative way of solving the SCF equations, that is, it gives the same results, but by a different method. Commercial use of the MOZYME function is protected by a US patent, number: 5,604,686. Details of the patent can be obtained from <http://patent.womplex.ibm.com/patquery>. Currently, a patent on this procedure is being applied for in Japan.

1.1.2 Technical and Other Changes

As a result of the steady accumulation of donations, MOPAC had become quite difficult to modify. This had tended to discourage users from trying to add new features. To correct this, a major re-write of the program has been done. This involved removing most of the COMMON blocks, and passing information between subroutines using two large arrays.

Although the new program has a more complicated information passing method, once it is learned, the job of adding or modifying functions is greatly simplified.

Details of the theory and implementation are given in the "Theory" section of the Manual.

1.1.3 Cooperative Development of MOPAC

MOPAC has been developed, and hopefully will continue to be developed, by the addition of contributed code. As a policy, any supplied code which is incorporated into MOPAC will be described in the next release of the Manual, and the author or supplier acknowledged. In the subsequent release only journal references will be retained. The objective is to produce a good program. This is obviously not a one-person undertaking; if it were, then the product would be poor indeed. Instead, as we are in a time of rapid change in computational chemistry, a time characterized by a very free exchange of ideas and code, MOPAC has been evolving by accretion. The unstinting and generous donation of intellectual effort speaks highly of the donors.

The purpose of having this forum is to facilitate the development of semiempirical methods. Because MOPAC has been a public domain program, a large number of researchers have used it. Many of these have found defects in it, and many others have added functionalities which were missing in the original program. In an unknown number of instances, different research groups have added the same functionality. In other words, effort has been duplicated. Authors of such developments might be reluctant to distribute their improvements, particularly as frequently the improvements represent a considerable investment in time and resources. To help avoid this phenomenon of "re-inventing the wheel", developers of code are invited to submit their contribution for inclusion in the next version of MOPAC.

Contributions should satisfy one or more of the following conditions:

- For existing functionalities:
 - A significant increase (minimum of 15%) in efficiency.
 - The output format should be much easier to read.
 - The accuracy should be increased significantly.
- For new functionalities:
 - A description which would allow users to appreciate the usefulness of the new functionality must be provided, along with one or more data sets and results files, and clear instructions on how to use it.
 - For new functionalities which do not involve new data or new results, a description which would allow users to appreciate the usefulness of the new functionality must be provided.
- For bug reports:
 - A bug or fault in MOPAC must be identified, and the correction given. If possible, the effect of the bug should be described.

In addition, the new code should be provided as part of a complete functioning MOPAC program. Exceptions can be made where only a small number of lines of the original program need to be changed, e.g., a simple bug-fix.

How to acknowledge contributors and report new functionalities presents a problem. Ideally, new functionalities should first be reported in a more conventional scientific journal, and only summarized in the QCPE Bulletin. In addition, where possible, a description of the new functionality will be added to the Manual, so that interested users will have a working description in addition to the regular journal article.

If any errors are found, or if MOPAC does not perform as described, please contact:
Dr. James J. P. Stewart, Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs, CO 80921-2512, Tel: (719) 488-9416, FAX: (719) 488-9758, E-mail: jstewart@fujitsu.com.

1.2 Graphical User Interfaces

MOPAC does *not* have a specific graphical user interface. The program can be used on its own, or with an in-house G.U.I, or with a commercial G.U.I. The reason for the lack of a specific G.U.I. is simple: my objective is to produce a semiempirical computational chemistry 'engine'. This engine can then be joined, possibly with modifications, to different graphics packages. In order to facilitate this, the input and output of MOPAC has been kept as constant as possible.

Readily available and supported G.U.I.s include:

WinMOPAC

WinMOPAC is a GUI specifically designed for use with MOPAC 2000. It can be obtained from "WinMOPAC Distributor", Fujitsu Limited, 1-9-3 Nakase, Mihama-ku, Chiba 261-8588, Japan. See <http://www.winmopac.com/>. For more information, contact winmopac@ssd.se.fujitsu.co.jp, or call +81(43)299-3236 or fax +81(43)299-3019.

CHEM3D

The current version of CHEM3D uses MOPAC 97, Revision 2, but the next version will support MOPAC 2000. Distributor: CambridgeSoft Corporation, 875 Massachusetts Avenue, Cambridge MA 02139; <http://www.camsoft.com/>; E-mail: info@camsoft.com; Tel: (617) 491-2200 and (800) 315-7300; FAX: (617) 491-0555.

HyperChem

HyperChem has an interface to MOPAC for use on a PC. This interface simplifies the construction of molecules and MOPAC input data preparation. HyperChem visualization tools can also be used for the analysis of MOPAC results (orbitals, vibrations etc).

Distribution and/or contact information: FQS Poland Sp. z o.o., ul. Starowislna 13-15, 31-038 Krakow, Poland. <http://www.fqspl.com.pl/>

1.3 Use of MOPAC as an Educational tool

Although MOPAC is designed first and foremost as a research tool, a secondary, and still important, objective is that it should be suitable for educational purposes. In order to satisfy the demands of an educational tool, both the program and the Manual includes material which I hope will be of use to students.

The following list, while not exhaustive, illustrates the types of topic which MOPAC can be used for:

Semiempirical Theory

Hamiltonian matrices. One and two electron integrals. Semiempirical methods. Secular determinants. Diagonalization. Eigenvectors. Eigenvalues. Orthonormalization. Density matrices. Electronic energy. ΔH_f . See Section 5.2.

Symmetry theory

Point-groups. Irreducible representations. Normal modes. Electronic states. Symmetry products. Space groups. Little groups. Complex characters.

Electronic States

Space and spin quantization. Brillouin's theorem. Configuration interaction. Shift operators. Spin operators. Franck-Condon principle. Fluorescence. Red-shift. Pauli principle.

Normal Coordinates

Infra-red active modes. Reduced masses.

Time-dependent phenomena

1.4 A General Description of MOPAC

1.4.1 Validation

Before potential users start working with MOPAC, they must be reasonably sure that the program works as described. The testing of a new MOPAC is quite lengthy. The basic tests verify that all the functions work. This involves running many hundreds of systems, many of which are highly exotic. Once all tests of this type are successful, pairs of functions are tested. For example, two functions in MOPAC are (1) the ability to apply an external electric field, and (2) to optimize the geometry of a molecule. By combining these functions, the geometry of a molecular system can be optimized in the presence of an external electric field. Systems that would otherwise not interact can be made to weakly bond together under the influence of an external field. With the wide range of possible functions, testing all pairs of functions is not practical; however, most reasonable pairs of functions have been tested to ensure that they work correctly.

The number of possible combinations of three or more functions (an example of a three function calculation would be to calculate the path of an implanting ion as it approaches a solid surface while it is accelerating under the influence of a potential gradient). Several of these have been tested. There is no reason to assume that all such combinations of functions will work, but at the end of extensive testing, all calculations attempted did work.

Algorithmic validation

A powerful device to validate a program is by examination of the computer code involved. Reading the program is not practical: MOPAC 2000 is over 80,000 lines long. Instead various software validation tools can be used. The more important of these are:

- FTNCHEK: Checks conformance with the ANSI FORTRAN 77 Standard. MOPAC has known exceptions to the standard, but all of these have been independently checked. Exceptions include the use of machine-specific code for determining the current date and the amount of CPU time used.
- IEEE: Some machines support testing for conformance with the Institute of Electrical and Electronics Engineering standards (IEEE standard 754), for validity of arithmetic operations. Again, with a few known and tested exceptions, MOPAC conforms with this standard.
- Some machines allow the initial state of otherwise undefined numbers to be defined. By setting all undefined numbers to NaNs (Not a Number), any assumption of initial values is easily detected.
- The validation tests were carried out with array bounds checking. Of course, any errors detected during this test were corrected.
- Only one top copy of MOPAC exists. This copy is designed to run on many different platforms. Because of this, any errors detected on one platform can be corrected, and the correction is then automatically propagated to all platforms.

The Beta Test

MOPAC can be run on a wide range of platforms, from Windows 95, 98, and NT, through workstations such as the RS6000 and SGI UNIX machines, to parallel supercomputers, such as the Fujitsu VPP and AP3000. After porting, the program is sent to several different research groups around the world. These groups have volunteered to test MOPAC for use in their own on-house research programs. These tests are real in the sense that the testers are interested in performing serious research, and any limitations found would be the cause of severe irritation. In practice, very few bugs are uncovered by this test, but all such bugs are corrected before the final release.

Usage

Over 50,000 copies of MOPAC have been requested. This large installed base means that bugs and limitations are rapidly found and reported. Where practical, these bugs and limitations are corrected, and consequently the program has become quite robust. Of course, with any major new release, such as MOPAC 2000, new bugs are likely to be introduced. Most of these, however, will be detected and corrected rapidly.

Support

Although MOPAC is routinely maintained in the status of no known bugs, problems do occur. If a user suspects that there is a bug in the program, assistance can be requested by telephone, E-mail, or FAX. Suspected bugs are given a high priority, and most are resolved within one working day. In the case of genuine bugs, a bug-fix or patch is provided to the user.

1.4.2 Geometry Optimization

General

One of the commonest operations in computational chemistry is optimizing the geometry. This involves modifying the geometry until the energy is a minimum. At that point, the net forces acting on every atom vanish. Over the years, various methods for optimizing geometries have been developed. One of these, Baker's EigenFollowing procedure, has proven to be very robust, and this is now used as the default. If, for any reason, the EF method is not wanted, other methods are available.

Constrained by definition

When only part of a system is to be optimized, options are provided to constrain other parts of the system. For example, a bond length, angle, or dihedral for one or more atoms can be defined as fixed at some initial value, or the x or y or z coordinate can be defined as unchanging.

Constrained by symmetry

Some systems have symmetry that can be used to accelerate geometric operations. Thus in Fullerene, C_{60} , all atoms are in the same environment, and there are exactly two unique bond-lengths. The geometry optimization calculation can be reduced from order 174 ($3 \times 60 - 6$) to order two by use of symmetry. Symmetry can also be used in unsymmetric systems. For example, in determining the transition state for the S_N2 reaction $Br^- + CH_4 \rightarrow CH_3Br + H^-$, the C-Br and C-H distances can be set equal, and the geometry optimized. After releasing the symmetry constraint, the transition state geometry is determined in one step by use of Baker's EigenFollowing method.

Energy Minimization

The degree to which geometries can be optimized can be varied according to the user's need: from the default, which is the best compromise between precision and speed, to high precision for publication work, to low precision for rapid screening. Unconditionally, the degree of optimization can be displayed, although the default is for this information to be suppressed unless problems are encountered.

Unless special action is taken, only local minima are located. Thus optimization of the geometry of dimethyl ether would not yield ethanol, nor would optimization of *gauche* butane yield *trans* butane.

Transition State location

Several methods are provided for refining transition states. One, Baker's EigenFollowing, is usually sufficient for most systems; the others are provided for the rare instances when the default TS does not work.

1.4.3 Methods

MNDO-*d*

The full MNDO-*d* method is supported. Currently, parameters are available for Al, Si, P, S, Cl, Br, and I only, although new parameters will be added as they become available.

NDDO Methods

Three NDDO methods are supported: MNDO, AM1, and PM3. Parameters for PM3 are available for the following elements: H, Li, Be, B, C, N, O, F, Mg, Al, Si, P, S, Cl, Zn, Ga, Ge, As, Se, Br, Cd, In, Sn, Sb, Te, I, Hg, Tl, Pb, Bi. For all methods, 100% ionic charges are provided; these mimic Group I and II ions. A unique type of atom that has a single valence, and which takes on the electronegativity of whatever atom it is attached to is provided. This atom, called Cb for Capped Bond, is designed for use where dangling bonds might be created, for example, when a fragment of a system is being studied.

MINDO/3

The MINDO/3 method is available for most functions. However, it is likely that MINDO/3 will not continue to be supported.

New Parameters

As new parameter sets become available, a mechanism must exist to allow these parameters to be used by MOPAC. The simplest way is to convert the parameters into a small data set, and then to make this data set known to MOPAC, by use of a keyword. Only the unique parameters need be defined in the data set, any derived parameters are computed internally.

1.4.4 SCF Procedures

Restricted Hartree Fock

The default self-consistent method is restricted Hartree-Fock. This allows both closed shell and open shell systems. For open shell systems, errors due to the half electron approximation are automatically corrected.

Unrestricted Hartree Fock

At user's choice, unrestricted Hartree-Fock SCF calculations can be run. In these, the α and β spin molecular orbitals have different spatial forms. When UHF is used, the expectation values of the S and S^2 operators are printed. In addition to printing the α and β spin molecular orbitals, the associated α and β spin density matrices can be printed, as can the $\alpha - \beta$ spin density matrix.

SCF-CI

The C.I. in MOPAC is sufficiently versatile to allow almost any closed or open shell system, in any state, to be calculated. Unpaired spin densities can be generated. Geometry optimization can be performed on both ground and excited states, including degenerate states (both with and without Jahn Teller effects). This can be contrasted to the MCSCF method, which is not supported, in which the SCF is performed using several configurations.

1.4.5 Derived Properties

Bond orders

Several types of bond orders can be displayed. The simplest is the Wyberg indices, which mirror the simple ideas of single, double, and triple bonds. More complicated representations include Mulliken populations, delocalizability, superdelocalizability, free valence, and spin density.

Charges

By default, the net charge on each atom is printed. These are the Coulson charges, although if requested, Mulliken charges can also be printed.

Dipole moment

By default, the dipole vector (calculated from atomic charges and the lone pairs) is calculated and printed, along with the net dipole moment, in Debye.

Static (zero frequency) Polarizability

The polarizability of a system is a measure of the response of the electron density distribution to a static electric field. This can be calculated two ways, by the application of electric fields (the static method) and by direct analysis of the wavefunction (the sum over states method). For large systems, the application of electric fields method is faster, although the sum over states method is more precise for all systems. Because of the limited precision of the static method, a way has to be provided to allow the user to determine the precision. Static polarizability (and first and second non-linear optical responses) can be determined from the changing dipole or from the changing heat of formation. By printing the results of the calculation based on both dipole and ΔH_f , side by side, the user can get an estimate of the precision.

Frequency dependent Non-Linear Optics

In addition to the polarizability, the following non-linear optical properties can be calculated: first and second order properties (β and γ), Electric Field Second Harmonic Generation (EFISH), optical rectification, electrooptic Pockel's Effect, second and third harmonic generation. Where appropriate, averages are printed.

1.4.6 Graphics

Although MOPAC does not have any graphics, it has the capability of generating data for use by graphics packages. This data includes information on the molecular orbitals, the overlap matrix, and the orbital exponents, etc. Graphics packages can then use the data to generate contour maps of M.O.s, density, density differences, etc.

1.4.7 Types of Species

Atoms

Simple calculations on atoms yield conventional heats of formation. These are the heats of atomization of the elements. Most atoms have open shells in the ground state, and these can be allowed for by use of appropriate keywords. By default, the electronic state of the system will be printed. Excited states can be readily calculated, and the transition dipole for photoexcitation can be printed. If, as is common, the originating state (usually the ground state) or the terminating state (usually an excited state) is degenerate, the degeneracy is taken into account in the calculation of transition dipole. The conventional selection rules can be derived by an analysis of the states and transition dipoles.

Molecules

The commonest type of system calculated consists of neutral, polyatomic molecules. These can be simple, closed shell species, such as benzene, to radicals, e.g. nitric oxide, to zwitterions, to multiple open shell radicals, such as NH_4^+ , to excited electronic states, e.g. $n\pi^*$ pyridine.

Ions

Calculations can be performed on ionized species, both isolated and with counterions, in both gas phase and solvated. All degrees of ionization are allowed.

Polymers

Regular polymer systems for which periodic boundary conditions can be imposed can be calculated. The time for such calculations is about 30% greater than for a discrete molecule of the same size as the unit cell used. Geometries, including unit cell length, can be optimized.

Effect of Stretching

In polymers, a reaction can be set up, in which the translation vector distance can be steadily increased. The effect of this is to steadily increase the distance between the repeat units of the polymer. Initially, the energy would not change significantly, as any conformational flexibility in the polymer is used up, but once that is done, the energy of the system (a measure of the stress) would rise parabolically with increased strain. From an analysis of this parabola, and given the density of the polymer, the Young's modulus can readily be calculated. If the strain is increased without limit, a point would be reached at which the weakest bond in the polymer would break, and the stress would immediately drop to zero. From this, and the experimental density, the tensile strength can be calculated.

Vibrations (Phonon Spectrum)

As with molecules, the vibrational spectrum of polymers can be calculated. However, unlike molecular normal modes, which have quantized vibrational frequencies, polymers give rise to bands of frequencies, defined by the associated wave-vector k in the Brillouin zone. At $k=0$, four of these bands have zero frequency, corresponding to the four trivial vibrations of a polymer.

Defect polymers (solitons)

Excitations can occur in some polymers, to give rise to ion pairs, isolated electronic excited states, and other electronic defects. These can be modeled using oligomers. A simple excited state would be in polyacetylene, for example, in which one carbon atom was singly bonded or doubly bonded to two other carbon atoms. The effect of an applied electric field to induce hopping of solitons can be modeled.

Layer Systems

By using the Born-von Kármán periodic boundary conditions, layer systems can be modeled. Geometry optimization, including optimizing the unit cell dimensions, can be carried out.

Solids

By using the Born-von Kármán periodic boundary conditions, regular solid systems can be modeled. Geometry optimization, including optimizing the unit cell dimensions, can be carried out.

Compressivity

For cubic systems, the compressivity of the crystal can be deduced by calculating the ΔH_f in a reaction in which the unit cell dimension is systematically varied.

Electronic Structure (Brillouin Zone)

A utility program for analyzing the electronic structure of polymers, layer systems, and solids is provided. This uses output from MOPAC, the space-group symmetry operations, and interactive user input to generate the Little groups for points in k -space, band structures, and cross-sections through k -space for selected bands. The first step in this analysis is to symmetrize the energy matrix for the solid. After this is done, the resulting structures in k -space are fully symmetry adapted.

1.4.8 Giant Molecules

Linear Scaling

A new method for solving the SCF equations is implemented. In this method, the time required for calculating the SCF increases linearly with the size of the system. The memory demand is also considerably reduced (for the larger systems run, by 97 to 98 percent of what would be needed conventionally). This allows systems of many thousands of atoms to be calculated rapidly.

Lewis Structures

The Lewis structures for most systems can be printed. A limitation is that systems involving atoms with more than four bonds cannot be handled, although in practice, this can partly be circumvented by breaking bonds. Systems as complicated as fullerene, TCNQ, SF₆, cystein zwitterion, and benzene, can be described.

Determining Net Charge

Determining the net charge on a large system, such as a protein, can be tedious. The simplest way would involve examination of the structure, and identifying each ionized site. This task can be performed rapidly by running a simple calculation. If a MOZYME calculation is run, then any ionized sites, and the net charge on the system, will be printed. It is not necessary to perform an SCF calculation, so this operation is very rapid.

Frequently, the geometry supplied contains errors. If any of these errors would prevent a meaningful calculation being run, then a brief description of the errors detected is printed. This speeds up the preparation of data sets.

Residue Sequence

If requested, the residue sequence, along with information on the formal charges on each residue, can be printed. Both the one letter and three letter abbreviations are given. Non-standard residues can be accommodated by use of keywords.

PDB input/output

Molecular geometries can be read in using Brookhaven Protein Data Bank format. Regardless of the input format, geometries can be output in PDB format, under user control. There are several commonly used minor variations on the PDB format definition, most of which are accommodated in the input, and standard PDB format is used in the output.

Charge on Residues

The net charge on residues is of interest in protein chemistry. This quantity is printed for each residue, and is broken down into the contribution from the backbone atoms and the side chain atoms of each residue.

Description of Electronic Structure

The standard description of the electronic structure of giant molecules is given, although, because the default format would produce a very voluminous output, the format is changed. Normal descriptors include: Atomic charge, valence, bond orders $\sigma - \pi$ decomposition, and dipole.

Partial Geometry Optimization

When localized molecular orbitals are used, the M.O.s associated with each atom can readily be identified. This means that when a partial geometry optimization is run, only the LMOs involved in the moving atoms need be used. A consequence of this is that such calculations run much faster. For large systems, in which only a few atoms move, the largest increase in speed over conventional methods was a factor of 13,000 times.

1.4.9 Symmetry theory

Symmetry is very important in chemistry, and, in recognition of this, MOPAC makes extensive use of symmetry theory. Most of the symmetry theory is done automatically, without any user action being necessary. The simplest operation is the recognition of the point-group of a molecule. All non-magnetic single point groups up to order 7 and most of the groups up to order 8 are recognized. This set includes the three infinite groups (R_3 , the order of the sphere, $C_{\infty v}$, and $D_{\infty h}$) for all chemically realizable representations, and the seven cubic point-groups. There is an ambiguity in the definition of some irreducible representations of some groups, such as b_1 and b_2 in C_{2v} . This ambiguity is resolved using the conventional definition for orienting molecules. To allow for the fact that molecular geometries might not be completely precise, a certain tolerance is built in to the test for point-groups. Sometimes, a user might wish to use a sub-group of the full point-group. This can be done by using `NOREOR`, or, if symmetry theory is not wanted at all `NOSYM` can be specified.

Symmetry Labels

Symmetry labels are automatically assigned to molecular orbitals, vibrations, and electronic states. These labels are of the form nR , where R is the irreducible representation and n is the n th occurrence of that representation. For each eigenfunction, the symmetry label is unique, and for this and other reasons, the symmetry labels are true quantum numbers. This is particularly evident in electronic states, where the symmetry label includes information on the spin state.

Symmetrizing vibrations

To facilitate vibrational analysis of high-symmetry systems, symmetry theory is used to symmetrize the force matrix. This has the side effect of reducing any errors introduced by finite precision mathematics. The resulting normal modes are completely symmetry adapted, and subsequent analysis of these modes is made much easier.

Accelerating Calculation of Vibrations

For molecules that have symmetry, symmetry theory is automatically used to accelerate the construction of the force matrix. This reduces the time needed for the calculation of normal modes. For C_{60} , the increase in speed is a factor of about 40.

1.4.10 Electric Fields

The effect of applied external electric fields can be modeled. The fields are uniform, and the orientation and intensity of the field is under user control. The applied fields cannot be used in the translation directions of infinite systems.

1.4.11 Electrostatic Potential

Four electrostatic potential methods are available, of which two can generate data for use by graphics programs. These are the Wang-Ford Parametric Electrostatic Potential (PMEP) and the Merz-Bessler ESP methods. The PMEP method gives results similar to those from *ab initio* 6-31G calculations, but is limited in the range of atoms allowed. In contrast, the ESP method is quite general.

1.4.12 Solvent effects

Two solvent models are provided: the COSMO technique and the MST or Tomasi method. In the COSMO method, excited state systems in solution can be modeled, and the geometries of solvated systems optimized. The Tomasi method includes cavitation and surface tension effects. Both methods allow the user to define the properties of the solvent. In the Tomasi method, keywords can be used to define specific solvents, i.e., water, chloroform, and carbon tetrachloride.

1.4.13 Electronic Excited states

MOPAC contains an extensive configuration interaction package. This allows a wide range of open and closed shell ground and excited state phenomena to be modeled. Examples of the types of systems that can be modeled include: methane, stabilized by mixing in excited states; oxygen, with an open shell ground state; methane cation, with and without Jahn Teller effects, and high spin systems (up to nonet). Keywords are provided for the commonest types of C.I. (single electron excitation, single plus double, single plus paired double, single, double, and triple excitations). All other types of C.I. can be defined explicitly, under keyword control.

Internal checks are automatically carried out to ensure that the calculations do not violate any theoretical rules, although these constraints can be relaxed, at user discretion. State spin and symmetries are automatically assigned.

The effect of C.I. on electron density distributions can be modeled, both in ground and excited states, and for ground state and vibrational states.

1.4.14 Intersystem Crossing

In some reactions, the electronic state of the system can change. The geometry at which this change occurs can be modeled. This geometry can be defined as the minimum energy geometry for two degenerate states. This technique is likely to be of use in photochemical research, for example in the photographic industry.

1.4.15 Vibrations

The normal modes of vibration of a stationary system can be calculated. For ground states, this consists of the $3N - 6$ or $3N - 5$ non-zero modes, while for simple transition states, the $3N - 7$ real modes and 1 imaginary mode are reported. If desired, the force constants for the system can be printed.

Description of Vibrations

Because of the complexity of normal modes, an analysis of these modes is printed. This analysis allows the nature of the normal mode (i.e., X-Y stretch, A-B-C bend, etc) to be rapidly described. The normal coordinates are printed. However these are of limited use because they are velocity vectors and do not indicate the energy carried by each atom. An additional display shows the effect of mass-weighting the normal modes; this gives an alternative view of the molecular vibrations.

Effective Mass and Travel

To the degree to which normal modes can be described as a simple harmonic oscillator, the effective mass involved can be calculated. For a homonuclear diatomic, this is half the atomic mass; if one of the atoms is extremely massive, the mass is approximately that of the other atom.

From quantum theory, the energy of vibration is quantized. Therefore, given a knowledge of the force constant for the vibration, and the effective mass, the excursion distance (in mass weighted space) can readily be calculated. By default, when normal modes are calculated, these two quantities are printed.

Transition Dipole

The relative intensity of an infra-red active band is a simple function of the transition dipole. This quantity is printed whenever the vibrational analysis is printed.

Internal Coordinate Force Constants

Although not an observable, the internal coordinate force constants are often of interest. These quantities are printed at the end of a normal coordinate calculation.

Isotopic Substitution

Although calculating the force matrix in a normal coordinate analysis is often lengthy, the final stage, mass weighting and generating normal modes, is very rapid. To allow different isotopic masses to be used in a single normal coordinate analysis, the option exists to save the force matrix. The isotopic masses can then be changed, and the old force matrix used again. Such calculations are very rapid.

Trivial Modes, projecting out

A minor increase in precision is achieved by projecting out the six trivial modes. These can subsequently be printed out. The order of the modes is then defined as x , then y , then z translation, followed by the three rotations, about the moments of inertia.

1.4.16 Thermodynamics

Various thermodynamic quantities (Partition function, enthalpy, heat capacity, and entropy) can be calculated for any temperature, or range of temperatures. These quantities can be decomposed into vibrational, rotational, internal, and translational contributions. The effect of changing temperature on the ΔH_f can be monitored.

1.4.17 Molecular Dynamics

Conservation of Energy

The time evolution of a system can be investigated. The starting point can be either a stationary point (optimized geometry or a transition state) or a non-stationary point, and the initial velocity vector can be zero, or determined by one of the normal modes, or an arbitrary (user supplied) vector. At each step, the position of each atom is modified by (a) the forces acting on it and its isotopic mass, (b) the velocity vector of the atom, and (c) information on the acceleration, and rate of acceleration, of the atom. During the course of the molecular dynamics, the total energy (kinetic plus potential) is constant as the system moves down a potential energy surface, the velocity increases, and vice versa.

Simulated Annealing

Although the default is to conserve energy, the option exists to allow kinetic energy to be reduced, with the rate of reduction being a function of time, by specifying a half-life in femtoseconds. The effect of this option is to simulate cooling of the system, analogous to the molecular mechanics method of simulated annealing.

Simulated heating

No constraint is placed on the sign of the half-life, consequently, a negative half life can be used. This simulates heating of the system. Invariably, unlimited heating results in atomization of any compound.

By an appropriate choice of keywords, any combination of heating and cooling, including conserving energy, can be modeled in a single run.

Sampling

Sampling of the state of the system can be done several ways: each point calculated can be printed, or constant steps in time, energy, or position can be chosen. At each point printed, the amount of information printed is determined by keywords, from a single line giving simple energetics, to the energetics plus geometry plus velocity vector.

Simulated vibrations

The easiest way to calculate vibrational frequencies is to calculate the force matrix and perform a normal mode analysis. While this is acceptable most of the time, in some systems the normal modes are sufficiently non-simple-harmonic that these vibrational frequencies are unacceptable.

An alternative is to set the system in motion, using the normal modes calculated conventionally, and to determine directly the period of vibration. For well behaved systems, this is almost exactly the same as that given from the force matrix, verifying the internal consistency of the various methods.

For very simple systems, e.g. diatomics, a second alternative is to plot the energy coordinate graph, and determine the vibrational period from either the energies or the gradients.

Intrinsic Reaction Coordinate

The time-independent behavior of a system can be modeled using the IRC option. In this, the atoms in a system move in response to the forces acting on them, moderated by their isotopic masses, however, at each step all kinetic energy is annihilated. The effect is to produce a time-independent trajectory, the steepest descent from the starting geometry to a stationary point.

1.4.18 Reaction Paths

Where a definable reaction coordinate can be identified, this can be used to drive a chemical reaction. For example, in a bond-breaking bond-making reaction, the bond being made can be used as the reaction coordinate. Several options are provided for specifying reaction paths, the three most common of which are: (a) to supply the various values of the reaction coordinate as extra data (at each point on the reaction path, the gradient of the path is calculated); (b) to define a step size and number of point to be calculated (at the end of the calculation, the ΔH_f are printed in a form suitable for plotting); and (c) to give two step sizes and numbers of points in two directions. This option is useful in mapping out a potential energy surfaces.

1.4.19 Saddle-Point Location

The geometry of the transition state is not always easy to define. The option exists to allow the reactants and the products to be defined, and to allow the saddle or transition state between these two systems to be calculated. Recent changes in this function have made it more robust, so that now the failure rate in locating transition states is very low. (In practice, transition states have been located for all reactions investigated.)

1.4.20 Data checking

In practice, almost all the errors that occur are a result of user error. To allow for this, extensive data checking is done. Where possible, corrective action is automatically taken; if this is not possible, the run is stopped before much time is used, and an error message printed. Most error messages are self-explanatory.

1.4.21 Restarts

For various reasons, a job might stop before the calculation is complete. To allow for this, intermediate results are output in a form that can be used to restart the calculation. Restarts are a useful way to change the course of a calculation, thus a MOZYME job could be stopped and restarted as a MOPAC job. Likewise, an AM1 job could be restarted as a MNDO-*d* calculation.

1.4.22 Portability

The program is written in almost pure FORTRAN-77. Minor exceptions from the ANSI standard are necessary in order to allow CPU time and current date to be used. MOPAC 2000 currently runs on 16 different platforms. Extension of this set to still more platforms is expected to be straightforward.

1.4.23 Program Structure

The source code of MOPAC is designed for ease of reading and modification. Although there are many structures in MOPAC, once these are learned, navigating within the program is relatively easy. Every occurrence of the common variables has the same name, and the distinction between geometric operations and the electronics is complete and absolute. One common block holds information on all the variables and arrays, and control within the program is kept as simple as possible.

MOPAC is available as source code, so that researchers can both know what is in the program and make changes to add new functionalities.

1.4.24 Dynamic Memory Allocation

Almost all the arrays in MOPAC are fully dynamic. At the start of a run, the number of atoms in the system is not known, therefore some static arrays are needed. However, once the calculation begins, all the data in the static arrays is transferred to dynamic arrays.

Only those arrays needed by MOPAC are created, and those that are created are dimensioned to the smallest size possible. As a result, there is little wasted memory, and many jobs that could otherwise not be run will now work.

1.4.25 Citation for MOPAC 2000

To cite the program, please use: "MOPAC 2000", J. J. P. Stewart, Fujitsu Limited, Tokyo, Japan (1999).

Chapter 2

Acknowledgements

The following list details all the contributions to the current MOPAC made by collaborators, co-workers, and donors of code or ideas. Users of MOPAC are encouraged to cite donors in their publications.

2.1 Contributors to MOPAC

- **Walter Thiel**, Organisch-chemisches Institut Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and **Dr Alexander Voityuk**, Lehrstuhl F. Theoretische Chemie, Technische Universität München, Lichtenbergstrasse 4, D-85747 Garching, Germany.

The integral evaluation routines for MNDO-*d* were provided. These formed the basis for the MNDO-*d* method as implemented in MOPAC.

- **Andreas Klamt**, Bayer AG, Q18, D-5090 Leverkusen-Bayerwerk, Germany.

The COSMO method has been extended to allow electronic excited states of solvated systems to be calculated.

- **F. Javier Luque, et al**, Dept. Bioquímica, Fac. Química, Universitat Barcelona, Martí i Franquès 1, E-08028 Barcelona, Catalunya, Spain.

The Tomasi model was extended to CHCl_3 and CCl_4 . The data input for modeling aqueous and organic solvents was considerably simplified.

- **Josep M. Anglada**, C.I.D.-C.S.I.C., Jordi Girona Salgado 18-26, E-08034 Barcelona, Catalunya, Spain, and **Josep M. Bofill**, Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Catalunya, Spain.

A method has been added for geometry optimization of a system in which two electronic states are constrained to be degenerate. This is often referred to as the minimum energy geometry for intersystem crossing.

- For her unflagging patience in editing the manual for clarity of expression, and ease of use, I thank my wife, Anna.

Chapter 3

Data files

3.1 Introduction

Data defining a calculation are supplied to MOPAC in a file called the data-file. This file is normally named `<filename>.dat`, where `<filename>` is a name given by the user.

The data-set is the only file which *must* be present. Other files may need to be present if specified by the data-set. For example, if the data-set contains the keyword "RESTART", then the restart file must exist.

In a normal simple job, only two files are produced: the results file and the archive file.

As the data-set is the most important file in MOPAC, we will deal with it first.

3.2 The Data-Set

Before going into the precise definition of the data, let us examine a typical small data-set. Once the general structure is understood, learning about the more sophisticated data-sets will be easier.

3.2.1 Example of Data for Ethylene

Ethylene is a simple organic molecule, of formula C_2H_4 . The structure is as shown in Figure 3.1. For this calculation, the bond-lengths and angles can be assumed. Unless instructed *not* to do so, MOPAC will automatically optimize the geometry. For this system, the data set is as shown in Figure 3.2.

As can be seen, the first three lines are textual. The first line consists of keywords (here four keywords are shown). These control the calculation. The next two lines are comments or titles. The user might want to put the name of the molecule and why it is being run on these two lines.

Lines one through three are obligatory. If no name or comment is wanted, leave blank lines. If no keywords are specified, leave a blank line. A common error is to have a blank line before the keyword line: this error is quite tricky to find, so be careful not to have four lines before the start of the geometric data (lines 4-9 in the example). Whatever is decided, the three lines, blank or otherwise, are obligatory.

The next set of lines defines the geometry. In the example, the numbers are all neatly lined up; this is not necessary, but does make it easier when looking for errors in the data. The geometry is defined in lines 4 to 9; line 10 terminates both the geometry and the data-file. Any additional data, such as symmetry data, would follow line 10.

At any point in the data set, one or more comment lines can be introduced. These are identified by the presence of an asterisk (*) as the first character on the line.

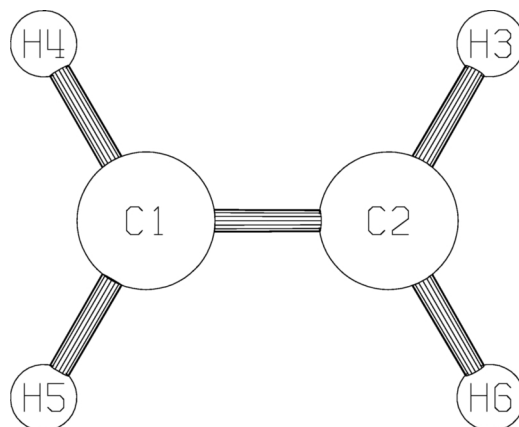


Figure 3.1: Ethylene

```

Line 1: MINDO3 VECTORS LOCAL PI
Line 2: EXAMPLE OF MOPAC DATA-SET FOR ETHYLENE
Line 3: The only geometric constraint is that the system must be planar
Line 4: C
Line 5: C    1.3 1
Line 6: H    1.1 1    120 1    0 0    2 1
Line 7: H    1.1 1    120 1    0 0    1 2 3
Line 8: H    1.1 1    120 1 180 0    1 2 3
Line 9: H    1.1 1    120 1    0 0    2 1 5
Line 10:

```

Figure 3.2: Data set for Ethylene

3.2.2 Overview of Data-Set

Data-sets consist of at least three parts. Later on in this Chapter we will elaborate on these parts, but for now it is sufficient to identify them.

In order, the three essential parts of a MOPAC data-set are:

Line 1: Keywords (one line). These control the calculation. The number of keywords used can range from none to about 20, space permitting. When more keywords are needed than will fit on one line, then more lines can be used (see Section 3.3.2 below).

Line 2: Description of the calculation, e.g. the name of the molecule or ion.

Line 3: Any other information describing the calculation.

Line 4: Internal or Cartesian coordinates (many lines). See Section 3.4 below.

Line 5: Blank line to terminate the geometry definition.

3.3 Keywords

3.3.1 Specification of Keywords

All control data are entered in the form of keywords, which form the first line of a data-file. A description of what each keyword does is given in Section 3.3.2. The order in which keywords appear is not important, although they must be separated by a space. Some keywords can be abbreviated; allowed abbreviations are noted in Section 3.3.2 (for example, 1ELECTRON can be entered as 1ELECT). However, the full keyword is preferred in order to more clearly document the calculation and to prevent the possibility that an abbreviated keyword might not be recognized. If there is insufficient space in the first line for all the keywords needed, then consider abbreviating the longer words. One type of keyword, those with an equal sign, such as BAR=0.05, may not be abbreviated, and the full word needs to be supplied.

Keywords which involve an equal sign, such as SCFCRT=1.D-12, cannot be written with spaces before or after the equal sign. Thus SCFCRT= 1.D-12, SCFCRT = 1.D-12, etc., are not allowed.

If two keywords which are incompatible, like UHF and C.I., are supplied, or a keyword which is incompatible with the species supplied, for instance TRIPLET and a methyl radical, then error trapping will normally occur, and an error message will be printed. This usually takes an insignificant amount of time, so data are quickly checked for obvious errors.

Keywords are normally supplied as the first line of a data-file. Options exist, however, which allow users to:

- Supply more than one line of keywords.
- To have frequently-used sets of keywords stored in an auxiliary file.

3.3.2 Full List of Keywords

& Turn next line into keywords

+ Add another line of keywords

OSCF Read in data, then stop

1ELECTRON Print final one-electron matrix

1SCF Do one scf and then stop

AIDER Read in ab-initio derivatives

AIGIN Geometry must be in gaussian format

AIGOUT In arc file, include ab-initio geometry

ALLBONDS Print final bond-order matrix, including bonds to hydrogen

ALLVEC Print all vectors (keywords vectors also needed)

ALT_a=A In pdb files with alternative atoms, select atoms A

ALT_r=A In pdb files with alternative residues, select residues A

ANALYT Use analytical derivatives of energy wrt geometry

AUTOSYM Symmetry to be imposed automatically

AM1 Use the AM1 hamiltonian

BAR=*n.n* reduce bar length by a maximum of *n.n*

BCC Solid is body-centered cubic (used by BZ)

BFGS Optimize geometries using bfgs procedure

BIGSCF If restart, do a full scf to re-form starting density matrix

BIGCYCLES Do a maximum of *n* big steps

BIRADICAL System has two unpaired electrons

BONDS Print final bond-order matrix

C.I. A multi-electron configuration interaction specified

CHARGE=*n* Charge on system = *n* (e.g. NH₄ ⇒ charge=1)

CIS C.I. uses 1 electron excitations only

CISD C.I. uses 1 and electron excitations

CISDT C.I. uses 1, 2 and 3 electron excitations

COMPFG Print heat of formation calculated in COMPFG

CONNOLLY Use Connolly surface

CROSS Determine intersystem crossing geometry

CUTOF1=*n.nn* Cutoff for polarization functions is *n.nn* ångstroms

CUTOF2=*n.nn* Cutoff for nddo approximation is *n.nn* ångstroms

CUTOFP=*n.nn* Madelung distance cutoff is *n.nn* ångstroms

CUTOFS=*n.nn* Overlap cutoff distance is *n.nn* ångstroms

CVB User defined chemical bonds (used in making Lewis structures)

CYCLES=*n* Do a maximum of *n* steps

DAMP=*n.nn* Use damping factor *n.nn* in SCF calculation

DCART Print part of working in DCART

DDMAX See EF code

DDMIN Minimum trust radius in a EF/TS calculation

DEBUG Debug option turned on

DENOUT Density matrix output (channel 10)

DENOUT=*n* Density matrix output every *n* seconds

DENSITY Print final density matrix

DEP Generate Fortran code for parameters for new elements

DERIV Print part of working in DERIV

DFORCE Force calculation specified, also print force matrix.

DFP Use Davidson-Fletcher-Powell method to optimize geometries

DIAGG2 Print part of working in DIAGG2

DIPOLE Fit the ESP to the calculated dipole

DIPX X component of dipole to be fitted

DIPY Y component of dipole to be fitted

DIPZ Z component of dipole to be fitted

DISEX Distance for interactions in fine grid in COSMO

DMAX Maximum stepsize in eigenvector following

DMIN See EF code

DOUBLET Doublet state required

DRC Dynamic reaction coordinate calculation

DUMP=*n* Write restart files every *n* seconds

ECHO Data are echoed back before calculation starts

EF Use ef routine for minimum search

EIGEN Generate canonical eigenvectors instead of LMOs

EIGS Print all eigenvalues in ITER

ENPART Partition energy into components

EPS Dielectric constant in COSMO calculation

ESP Electrostatic potential calculation

ESPRST Restart of electrostatic potential

ESR Calculate RHF spin density

EXCITED Optimize first excited singlet state

EXTERNAL Read parameters off disk

FIELD An external electric field is to be used

FILL=*n* In RHF open and closed shell, force M.O. *n* to be filled

FLEPO Print details of geometry optimization

FMAT Print details of working in FMAT

FOCK Print last Fock matrix

FORCE Force calculation specified

GEO-OK override interatomic distance check

GEOCHK Print part of working in GEOCHK

GNMIN Minimize gradients in EF

GNORM=*n.n* Exit when gradient norm drops below *n.n*

GRADIENTS Print all gradients

GRAPH Generate file for graphics

GREENF Correct I.P.s using Green's function technique

H2O Use water as the solvent in MST calculations

HCORE Print details of working in HCORE

HESS=*n* options for calculating hessian matrices in EF

H-PRIO Heat of formation takes priority in DRC

HYBRID Print details of working in HYBRID

HYPERFINE Hyperfine coupling constants to be calculated

INT Make all coordinates internal coordinates

IRC Intrinsic reaction coordinate calculation

ISOTOPE Force matrix written to disk (channel 9)

ITER Print details of working in ITER

ITRY=*n* Set limit of number of SCF iterations to *n*

IUPD Mode of Hessian update in eigenvector following

K=(*n*,*n*) Brillouin zone structure to be calculated

KINETIC Excess kinetic energy added to DRC calculation

LEWIS Print details of working in LEWIS

LINMIN Print details of line minimization

LARGE Print expanded output

LET Override certain safety checks

LOCALIZE Print localized orbitals

MAX In GRID, use maximum grid size (23*23)

MECI Print details of MECI calculation

MEMORY=*n* Try to keep memory usage below *n* MB.

MEP=1 Print Orozco-Luque electrostatic potentials for cross-section

MEP=2 Print Orozco-Luque electrostatic potentials for Connolly surface

MERS=(*n*) Keyword for BZ

MICROS Use specific microstates in the C.I.

MINDO/3 use the MINDO/3 hamiltonian

MINMEP Minimize MEP minima in the plane defined

MMOK Use molecular mechanics correction to CONH bonds

MNDO Use the MNDO hamiltonian

MNDO-d Use the MNDO-d hamiltonian

MODE=*n* In EF, follow Hessian mode no. *n*

MOLDAT Print details of working in MOLDAT

MS=*n* In MECI, magnetic component of spin

MULLIK Print the Mulliken population analysis

NEWGEO Make backbone atoms Cartesian and side-chain internal

NLLSQ Minimize gradients using NLLSQ

NLMO=*nnn* define average size of LMOs as *nnn*

NOANCI Do not use analytical C.I. derivatives

NOINT Do not print internal coordinates

NOLOG Suppress log file trail, where possible

NOMM Do not use molecular mechanics correction to CONH bonds

NONET NONET state required

NONR Do not use Newton-Raphson method in EF

NOREOR In symmetry work, use supplied orientation

NOSYM Point-group symmetry set to C1

NOTHIEL Do not use Thiel's FSTMIN technique

NSPA Sets number of geometric segments in COSMO

NSURF= n Number of surfaces in an ESP calculation

NOXYZ Do not print Cartesian coordinates

NSURF Number of layers used in electrostatic potential

OCTET Octet state required

OLDENS Read initial density matrix off disk

OLDFPC Use the old fundamental physical constants

OLDGEO Previous geometry to be used

OMIN In TS, minimum allowed overlap of eigenvectors

OPEN Open-shell RHF calculation requested

OPTBACK Set optimization flags for backbone atoms in a protein

OPTRES Set optimization flags for side-chain atoms in a protein

ORT Use Reynolds-Richards molecular electrostatic pot.

P= $n.nn$ An external pressure of $n.nn$ NM⁻² or N is applied to a solid or polymer.

PARASOK In AM1 calculations some MNDO parameters are to be used

PDB Geometry is in protein data bank format

PDB(*text*) User defined chemical symbols in protein data base

PDBOUT Output geometry in pdb format

PECI C.I. involves paired excitations only

PI Resolve density matrix into σ and π bonds

PINOUT Print localized M.O.s whenever PINOUT is called

PL Monitor convergence of density matrix in ITER

PM3 Use the MNDO-PM3 hamiltonian

PMEP Complete semiempirical MEP calculation

PMEPR Complete semiempirical MEP in a plane to be defined

POINT= n Number of points in reaction path

POINT1= n Number of points in first direction in grid calculation

POINT2= n Number of points in second direction in grid calculation

POLAR Calculate first, second and third order polarizabilities

POPS Print the self consistent field atomic orbital populations

POTWRT In ESP, write out electrostatic potential to unit 21

POWSQ Print details of working in POWSQ

PRECISE Criteria to be increased by 100 times

PRTMEP MEP contour data output to <filename>.mep

PULAY Use Pulay's converger to obtain a SCF

QPMEP Charges derived from Wang-Ford type AM1 MEP

QUARTET Quartet state required

QUINTET Quintet state required

RAPID In SCF, use atoms being optimized only

RECALC= n In EF, recalculate Hessian every n steps

RE-LOCAL Re-localize localized molecular orbitals

RELSCF= n Default SCF criterion multiplied by n

RELTHR= n Default threshold for LMOs multiplied by n

REORTH Reorthogonalize LMOs

RESEQ Resequence the atoms in a protein

RESIDUES Label all atoms in a protein

RESTART Calculation restarted

RMAX In TS, maximum allowed ratio for energy change

RMIN In TS, minimum allowed ratio for energy change

ROOT= n Root n to be optimized in a C.I. calculation

RSCAL In EF, scale p-RFO to trust radius

RSOLV Effective radius of solvent in COSMO

SADDLE Optimize transition state

SAFE Allow memory intensive scf convergers

SCALE Scaling factor for van der Waals distance in ESP

SCFCRT=*n* Default SCF criterion replaced by the value supplied

SCINCR Increment between layers in ESP

SEPTET Septet state required

SETUP Extra keywords to be read of setup file

SEXTET Sextet state required

SHIFT=*n* a damping factor of *n* defined to start SCF

SIGMA Minimize gradients using SIGMA

SINGLET Singlet state required

SIZES Print the sizes of all arrays created or destroyed

SLOG=*n.nn* In BFGS optimization, use fixed step of length *n.nn*

SLOPE Multiplier used to scale MNDO charges

SNAP Increase precision of symmetry angles

SPIN Print final UHF spin matrix

STEP Step size in path

STEP1=*n* Step size *n* for first coordinate in grid calculation

STEP2=*n* Step size *n* for second coordinate in grid calculation

STO-3G Deorthogonalize orbitals in STO-3G basis

SUPER Print superdelocalizabilities

SYMAVG Average symmetry equivalent ESP charges

SYMMETRY Impose symmetry conditions

T=*n* A time of *n* seconds requested

THERMO Perform a thermodynamics calculation

THRESH=*n.nn* Set the cutoff for LMO coefficients to *n.nn*

TIDY Print details of working in TIDY

TIMES Print times of various stages

TOM Use the Miertus-Scrocco-Tomasi solvation model

T-PRIO Time takes priority in DRC

TRANS The system is a transition state (used in thermodynamics calculation)

TRIPLET Triplet state required

TS Using EF routine for TS search

UHF Unrestricted Hartree-Fock calculation

UNSAFE Use reduced memory (may affect SCF convergence)

VDW Van der Waals radius for atoms in COSMO defined by user

VECTORS Print final eigenvectors

VELOCITY Supply the initial velocity vector in a DRC calculation

WILLIAMS Use Williams surface

WINMOPAC WINMOPAC data output to <filename>.wmp and <filename>.wob

XENO Specify fragments which are attached to residues

X-PRIO Geometry changes take priority in DRC

XYZ Do all geometric operations in Cartesian coordinates

The definitions below are given with some technical expressions which are not further defined. In many cases, more detail can be found in Chapter 5.

There are three classes of keywords:

1. those which CONTROL substantial aspects of the calculation, i.e., those which affect the final heat of formation,
2. those which determine which OUTPUT will be calculated and printed, and
3. those which dictate the WORKING of the calculation, but which do not affect the heat of formation. The assignment to one of these classes is designated by a (C), (O) or (W), respectively, following each keyword in the list below.

& (C)

An '&' means 'turn the next line into keywords'. An '&' on line 1 would mean that a second line of keywords should be read in. If that second line contained an '&', then a third line of keywords would be read in. If the first line has an '&' then the first description line is omitted; if the second line has an '&', then both description lines are omitted.

Examples: Use of one '&'

```
VECTORS DENSITY RESTART & NLLSQ T=1H SCFCRT=1.D-8 DUMP=30M
PM3 FOCK OPEN(2,2) ROOT=3 SINGLET SHIFT=30
Test on a totally weird system
```

Use of two '&'s

```
LARGE=-10 \& DRC=4.0 T=1H SCFCRT=1.D-8 DUMP=30M ITRY=300 SHIFT=30
PM3 OPEN(2,2) ROOT=3 SINGLET NOANCI ANALYT T-PRIORITY=0.5
LET GEO-OK VELOCITY KINETIC=5.0
```

+ (C)

A '+' sign means 'read another line of keywords'. A '+' on line 1 would mean that a second line of keywords should be read in. If that second line contains a '+', then a third line of keywords will be read in. Regardless of whether a second or a third line of keywords is read in, the next two lines would be description lines.

Example of '+' option

```
RESTART T=4D OPEN(2,2) SHIFT=20 PM3 +
SCFCRT=1.D-8 DEBUG + ISOTOPE FMAT ECHO singlet ROOT=3
THERMO(300,400,1)
```

Example of data set with three lines of keywords.

Note: This and the previous line are the two lines of description.

OSCF (O)

The data can be read in and then printed, but no actual calculation is performed when this keyword is used. This is useful as a check on the input data. All obvious errors are trapped, and warning messages printed. A second use is to convert from one format to another. The input geometry is printed in various formats in a OSCF calculation. Normally, only the Cartesian coordinate and MOPAC Z-matrix geometries are printed, however other geometries can also be printed if other keywords are used, e.g. AIGOUT or PDBOUT.

1ELECTRON (O)

The final one-electron matrix is printed out. This matrix is composed of atomic orbitals; the array element between orbitals φ_λ and φ_σ on different atoms is given by:

$$H_{\lambda\sigma} = 0.5 \times (\beta_\lambda + \beta_\sigma) \times S_{\lambda\sigma}$$

The matrix elements between orbitals φ_λ and φ_σ on the same atom are calculated from the electron-nuclear attraction energy, and also from the $U_{\lambda\lambda}$ value if $\lambda = \sigma$.

The one-electron matrix is unaffected by (a) the charge and (b) the electron density. It is only a function of the geometry. Abbreviation: 1ELEC.

1SCF (C)

When users want to examine the results of a single SCF calculation of a geometry, 1SCF should be used. 1SCF can be used in conjunction with RESTART, in which case a single SCF calculation will be done, and the results printed.

When 1SCF is used on its own (that is, RESTART is not also used), then derivatives will only be calculated if GRAD is also specified.

1SCF is helpful in a learning situation. MOPAC normally performs many SCF calculations, and in order to minimize output when following the working of the SCF calculation, 1SCF is very useful. See Section 5.5 for a worked example of an SCF.

When calculating the energies required to form electronic excited states, to avoid geometry reorganization, 1SCF must be used (see Section 5.4.7). This allows the vertical transition energies to be calculated (the Franck-Condon transitions).

AIDER (C)

AIDER allows MOPAC to optimize an *ab initio* geometry. To use it, calculate the *ab initio* gradients using, e.g., Gaussian [14]. Supply MOPAC with these gradients, after converting them into kcal/mol. The geometry resulting from a MOPAC run will be nearer to the optimized *ab initio* geometry than if a single step were taken using the geometry optimizer in Gaussian.

AIGIN (C)

If the geometry (Z-matrix) is specified using the Gaussian format [14], then this will normally be read in without difficulty. In the event that it is mistaken for a normal MOPAC-type Z-matrix, the keyword AIGIN is provided. AIGIN will force the data-set to be read in assuming Gaussian format. This is necessary if more than one system is being studied in one run.

AIGOUT (O)

The ARCHIVE file contains a data-set suitable for submission to MOPAC. If, in addition to this data-set, the Z-matrix for Gaussian input is wanted, then AIGOUT (*ab initio* geometry output), should be used.

The Z-matrix is in full Gaussian form. Symmetry, where present, will be correctly defined. Names of symbolics will be those used if the original geometry was in Gaussian format, otherwise 'logical' names will be used. Logical names are of form $\langle t \rangle \langle a \rangle \langle b \rangle [\langle c \rangle][\langle d \rangle]$ where $\langle t \rangle$ is

'r' for bond length, 'a' for angle, or 'd' for dihedral, $\langle a \rangle$ is the atom number, $\langle b \rangle$ is the atom with which $\langle a \rangle$ is related, $\langle c \rangle$, if present, is the atom number to which $\langle a \rangle$ makes an angle, and $\langle d \rangle$, if present, is the atom number to which $\langle a \rangle$ makes a dihedral.

ALLBONDS (MOZYME) (O)

This prints the rotationally invariant bond order [15] between atoms that are near to each other. In this context a bond is defined as the sum of the squares of the density matrix elements connecting any two atoms. For ethane, ethylene and acetylene the carbon-carbon bond orders are roughly 1.00, 2.00 and 3.00, respectively. See also (p. 34).

Valencies are calculated from the atomic terms only and are defined as the sum of the bonds the atom makes with other atoms. Only bond orders greater than 0.1 are printed.

ALLVEC (O)

When VECTORS is used then only those eigenvectors which are within about 8 M.O.s of the HOMO-LUMO gap will normally be printed. If you want all the M.O.s to be printed, add ALLVEC to the keyword line.

ALT_A=A (C)

Some PDB geometries have two or more possible positions for some atoms (positional disorder). These are indicated by the letters 'A', 'B', etc., before the three letter abbreviation for the residue. Only one position for an atom can be selected. To select all atoms of type 'A,' use ALT_A=A; to select all those of type 'B,' use ALT_A=B, etc.

ALT_R=A (C)

Some PDB geometries have two or more possible residues at a given site (structural disorder). These are indicated by the letters 'A', 'B', etc., after the three letter abbreviation for the residue. Only one residue can be selected. To select all residues of type 'A,' use ALT_R=A; to select all those of type 'B,' use ALT_R=B, etc. Sometimes a space is used instead of 'A', in which case the keyword would be "ALT_R= " (note the space after the equals sign).

AM1 (C)

The AM1 [3] method is to be used. By default MNDO is run.

ANALYT (W)

By default, finite difference derivatives of energy with respect to geometry are used. If ANALYT is specified, then analytical [16] derivatives are used instead. Since the analytical derivatives are over Gaussian functions—a STO-6G basis set [17] is used—the overlaps are also over Gaussian functions. This will result in a very small (less than 0.1 kcal/mol) change in heat of formation. Use analytical derivatives (a) when the mantissa used is less than about 51–53 bits, or (b) when comparison with finite difference is desired. Finite difference derivatives are still used when non-variationally optimized wavefunctions are present. See also NOANCI and Section 5.13.

AUTOSYM (W)

AUTOSYM is an alternative way of specifying symmetry. When AUTOSYM is present, symmetry data is generated. This data is similar to that supplied by the user when SYMMETRY is used. The 'rules' AUTOSYM uses are these:

- All bond-lengths which are within 0.0001Å are set equal.
- All bond-angles which are within 0.0057° are set equal.
- All dihedral angles which are within 0.0057° are set equal.
- All dihedral angles which are within 0.0057° of the negative of an existing dihedral are set equal to the negative of the existing dihedral.

These 'rules' are a re-wording of internal SYMMETRY functions 1, 2, 3, and 14. The .arc file will include the symmetry data. In order to change the .arc file symmetry data into normal symmetry data, change AUTOSYM in the .arc file to SYMMETRY.

BAR=n.nn (W)

In the SADDLE calculation [18], the distance between the two geometries is steadily reduced until the transition state is located. Sometimes, however, the user may want to alter the maximum rate at which the distance between the two geometries reduces. BAR is a ratio, normally 0.15, or 15 percent. This represents a maximum rate of reduction of the bar of 15 percent per step. Alternative values that might be considered are BAR=0.05 or BAR=0.10, although other values may be used. See also SADDLE.

If CPU time is not a major consideration, use BAR=0.03.

	Microstate 1	Microstate 2	Microstate 3	Microstate 4
LUMO	— ↓	— ↑	—	— ↑↓
HOMO	↑ ↑↓	↓ ↑↓	↑↓ ↑↓	— ↑↓

Figure 3.3: Configurations used in BIRADICAL Calculation

BFGS (W)

The default geometry optimizer, EF, is unsuitable for large (over about 500-1,000 atoms) systems. For large systems, the suggested optimizer is the Broyden-Fletcher-Goldfarb-Shanno [19, 20, 21, 22] procedure. By specifying BFGS, this procedure will be used instead of EF.

BIGCYCLES=*n* (C)

When **BIGCYCLES=*n*** is specified, a maximum of *n* complete geometry optimizations are allowed. After *n* optimizations, the calculation is stopped in the same manner as when the allowed time is exceeded.

BIGCYCLES is useful in comparing calculations using different computers. Different machines run at different speeds, so calculations based on time should not be used. Instead, by using **BIGCYCLES**, a specific number of cycles can be run, regardless of the machine speeds.

See also **CYCLES**.

BIGSCF (MOZYME) (W)

When an old density matrix is read in (a restart option), then the density matrix is assumed to be self consistent for the starting geometry. If this assumption is not valid, then **BIGSCF** can be specified. This will cause a SCF calculation to be performed on the starting geometry, starting with the LMOs read in by **OLDENS**.

BIRADICAL (C)

Note: **BIRADICAL** is a redundant keyword, and represents a particular configuration interaction calculation. Experienced users of MECI (q.v.) can duplicate the effect of the keyword **BIRADICAL** by using the MECI keywords **OPEN(2,2)** and **SINGLET**.

For molecules which are believed to have biradicaloid character, the option exists to optimize the lowest singlet energy state which results from the mixing of four microstates. These microstates are, in order: the microstate arising from a one electron excitation from the HOMO to the LUMO (Microstate 1 in Figure 3.3); the microstate resulting from the time-reversal operator acting on the parent microstate (Microstate 2); the microstate resulting from de-excitation from the formal LUMO to the HOMO (Microstate 3); and the microstate resulting from the single electron in the formal HOMO being excited into the LUMO (Microstate 4). Microstates 1 and 2 mix to form a singlet state, $1/\sqrt{2}(\Psi_1 - \Psi_2)$, and a triplet state $1/\sqrt{2}(\Psi_1 + \Psi_2)$. Microstates 3 and 4 are true singlet states. The singlet state arising from microstates 1 and 2 mixes with the (micro)states 3 and 4 to form three singlet states.

A configuration interaction calculation is involved here. A biradical calculation done without C.I. at the RHF level would be meaningless. Either rotational invariance would be lost, as in the D_{2d} form of ethylene, or very

artificial barriers to rotations would be found, such as in a methane molecule “orbiting” a D_{2d} ethylene. In both cases the inclusion of limited configuration interaction corrects the error. BIRADICAL should not be used if either the HOMO or LUMO is degenerate; in this case, the full manifold of HOMO \times LUMO should be included in the C.I., using MECI options. The user should be aware of this situation. When the biradical calculation is performed correctly, the result is normally a net stabilization. However, if the first singlet excited state is much higher in energy than the closed-shell ground state, BIRADICAL can lead to a destabilization. Abbreviation: BIRAD. See also MECI, C.I., OPEN, SINGLET.

BONDS (O)

The rotationally invariant bond order [15] between all pairs of atoms is printed. In this context a bond is defined as the sum of the squares of the density matrix elements connecting any two atoms. For ethane, ethylene, and acetylene, the carbon-carbon bond orders are roughly 1.00, 2.00 and 3.00 respectively. The diagonal terms are the valencies calculated from the atomic terms only and are defined as the sum of the bonds the atom makes with other atoms. In UHF and non-variationally optimized wavefunctions the calculated valency will be incorrect, the degree of error being proportional to the non-idempotency of the density matrix. For an RHF wavefunction the square of the density matrix is equal to twice the density matrix (see Section 5.3.3).

The bonding contributions of all M.O.s in the system are printed immediately before the bonds matrix. The idea of molecular orbital valency was developed by Gopinathan, Siddarth, and Ravimohan [23]. Just as an atomic orbital has a ‘valency’, so does a molecular orbital. This leads to the following relations: The sum of the bonding contributions of all occupied M.O.s is the same as the sum of all valencies which, in turn is equal to two times the sum of all bonds. The sum of the bonding contributions of all M.O.s is zero.

If LARGE is present, then the Medrano-Bochicchio-Reale population analysis is printed. For each atom, the following quantities are generated:

- The non-shared charge (sometimes called the self or inactive charge).
- The charge used to form bonds with other atoms (the active charge).
- The total charge (the sum of the first two terms).
- The valence (from the bonds matrix).
- The free valence (the difference of the last two terms).
- The statistical promotion (total charge minus core charge).
- The “Mulliken promotion”

Note that the last two terms are expressed in units of the electron, not the proton charge.

For more information, see [24, 25, 26, 27], and also see J. A. Medrano, R. C. Bochicchio, S. G. Das, “The ROHF Extension of the Statistical Population Analysis of Electron and Spin Densities”, J. Phys. B.

C.I.= n (C)

When C.I.= n is specified, the n M.O.s which ‘bracket’ the occupied-virtual energy levels will be used in a configuration interaction calculation. Thus, C.I.=2 will include both the HOMO and the LUMO, while C.I.=1 (implied for odd-electron systems) will only include the HOMO (This will do nothing for a closed-shell system, and leads to Dewar’s half-electron correction for odd-electron systems). Users should be aware of the rapid increase in the size of the C.I. with increasing numbers of M.O.s being used.

Table 3.2: Number of Configurations used in MECI

	No. of Electrons		Configs	No. of Electrons		Configs
	Alpha	Beta		Alpha	Beta	
C.I.=1	1	1	1	1	0	1
C.I.=2	1	1	4	1	0	2
C.I.=3	2	2	9	2	1	9
C.I.=4	2	2	36	2	1	24
C.I.=5	3	3	100	3	2	100
C.I.=6	3	3	400	3	2	300
C.I.=7	4	4	1225	4	3	1225
C.I.=8*	4	4	4900	4	3	3920

*: Do not use unless other keywords are also used. See MICROS, and the discussion in Section 5.4.7 (p. 152) .

Table 3.3: Use of C.I.=(n,m)

Keywords	No. of M.O.s	No. of Electrons*
C.I.=2	2	2 (1)
C.I.=(2,1)	2	2 (3)
C.I.=(3,1)	3	2 (3)
C.I.=(3,2)	3	4 (5)

* Odd electron systems given in parentheses.

Normally, configuration interaction (see Section 5.4) is invoked if the keywords which imply a C.I. calculation are used (BIRADICAL or EXCITED). Note that ROOT=, TRIPLET, QUARTET, etc. do not imply a C.I. calculation. Only when a C.I. calculation is performed (because of the presence of other keywords) are these keywords used. The numbers of microstates implied by the use of the keyword C.I.= n on its own are given in Table 3.2.

If a change of spin is defined, then larger numbers of M.O.s can be used up to a maximum of 13. If more than 13 M.O.s are needed, then `meci.h` would need to be changed. In `meci.h`, increase the value of NMECI as necessary. Normally, a full C.I. is carried out, in which case the spin-states are exact eigenstates of the spin operators. For systems with more than the normal number of configurations (Table 3.2), the configurations of lowest energy will be used. See also MICROS and the keywords defining spin-states.

Note that for any system, use of C.I.=5 or higher normally implies the diagonalization of quite large matrices. As a geometry optimization using a C.I. requires the derivatives to be calculated using derivatives of the C.I. matrix, geometry optimization with large C.I.s will require more time than smaller C.I.s.

Associated keywords: OPEN(n_1, n_2), C.I.=(n_1, n_2), CIS, CISD, PECI, CISDT, MECI, ROOT, MICROS, SINGLET, DOUBLET, etc.

C.I.=(n,m) (C)

In addition to specifying the number of M.O.s in the active space, the number of electrons can also be defined. In C.I.=(n,m), n is the number of M.O.s in the active space, and m is the number of doubly filled (that is, not empty or partially filled) levels to be used (see Table 3.3). If OPEN(n,m) is present, then the number of electrons may be increased.

Table 3.4: Use of CHARGE= n

Ion	Keyword	Ion	Keyword
NH_4^+	CHARGE=1	CH_3COO^-	CHARGE=-1
C_2H_5^+	CHARGE=1	$(\text{COO})^-$	CHARGE=-2
$\text{SO}_4^{=}$	CHARGE=-2	PO_4^{-3}	CHARGE=-3
HSO_4^-	CHARGE=-1	H_2PO_4^-	CHARGE=-1

CAMP (W)

The Camp-King converger (q.v.) is to be used. This is a very powerful, but CPU intensive, SCF converger.

CHARGE= n (C)

When the system being studied is an ion, the charge, n , on the ion must be supplied by CHARGE= n . For cations n can be 1, 2, 3, etc.; for anions -1 or -2 or -3 , etc., see Table 3.4.

Determining the charge on large systems is not at all obvious. A simple and rapid way to determine the charge is to run the system with keywords MOZYME and OSCF. If the system is charged, the charge will be printed, along with the atoms that carry a charge. This can then be used in a normal calculation.

CIS (C)

In configuration interaction calculations, only microstates resulting from single electron excitations are used if CIS is specified. (Read CIS as Configuration Interaction Singles.) Note that the ground state is *not* included in the set of configurations to be used.

CISD (C)

In configuration interaction calculations, the ground state and all microstates resulting from single and double electron excitations are used if CISD is specified. (Read CISD as Configuration Interaction Singles and Doubles.)

CISDT (C)

In configuration interaction calculations, the ground state and all microstates resulting from single, double, and triple electron excitations are used if CISDT is specified. (Read CISDT as Configuration Interaction Singles, Doubles, and Triples.)

COMPFG (O)

When COMPFG is specified, the ΔH_f is printed whenever COMPFG is called. In a conventional calculation (i.e., MOZ is not used), if DEBUG is also specified then the first few coordinates are printed, and if LARGE and DEBUG are both specified, then the whole geometry is printed.

CROSS (C)

When CROSS is specified, the minimum energy geometry subject to the constraint that two electronic states are degenerate is calculated using the method of Anglada and Bofill [28]. In other words, the lowest energy intersystem crossing geometry is generated.

For CROSS to work, an electronic excited state must be defined using $ROOT=n$, in which case the states n and $n - 1$ will be used in the intersystem crossing. These states can be further defined by use of keywords such as SINGLET or TRIPLET, etc. For example, for a singlet-triplet intersystem crossing, $ROOT=2$ could be used. For a singlet-singlet intersystem crossing $ROOT=2$ SINGLET might be suitable, and for a triplet-triplet crossing, $ROOT=2$ TRIPLET could be used.

CROSS works best when the states start off degenerate, or almost degenerate. If they are far apart in energy, the calculation will likely fail.

In addition to CROSS and $ROOT=n$, other keywords, defining the C.I., will be needed.

CUTOF1= $n.nn$ (C)

In MOZYME calculations, the cutoff distance for polarization functions is set by $CUTOF1=n.nn$. Beyond that distance, electrostatic interactions are considered only as point charges. At distances less than that given by $CUTOF1$, electrostatic interactions are represented by a point-charge and three polarization functions. Default: For systems of less than 30 atoms, $CUTOF1$ is 1000 Ångstroms; for systems of 30 or more atoms, $CUTOF1$ is 30 Ångstroms. See also p. 165.

CUTOF2= $n.nn$ (C)

In MOZYME calculations, the cutoff distance for two-electron two center and one-electron two center integrals is set by $CUTOF2=n.nn$. Below that distance, the interaction between two atoms is represented by the exact NDDO approximation. Above that distance, one-electron integrals that depend on the overlap are ignored, and two-electron integrals are simplified. Instead of using all 100 two-electron integrals between two heavy atoms, only seven are used. These represent the terms $\langle ss|ss \rangle$, $\langle ss|sx \rangle$, $\langle ss|sy \rangle$, $\langle ss|sz \rangle$, $\langle sx|ss \rangle$, $\langle sy|ss \rangle$, and $\langle sz|ss \rangle$. At still greater distances, beyond $CUTOF1$, only the $\langle ss|ss \rangle$ term is used. Default: For systems of less than 30 atoms, $CUTOF2$ is 1000 Ångstroms; for systems of 30 or more atoms, $CUTOF2$ is 12 Ångstroms. See also p. 165.

CUTOFP= $n.nn$ (C)

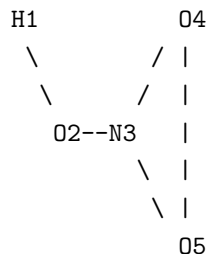
In polymers and solids, a cutoff distance is needed in order to ensure that every atom has a correct electric environment. If no cutoff is set, then equivalent atoms would experience different electric fields. The $CUTOFP$ distance is set, by default, to the largest distance that would ensure that equivalent atoms experience the same potential. In calculating the potential due to distance atoms, atoms separated by more than $CUTOFP$ are treated as if they were at distance $CUTOFP$.

CUTOFS= $n.nn$ (C)

The distance within which atomic orbital overlaps are calculated is set by $CUTOFS$. By default, $CUTOFS$ is 7.0 Ångstroms. There is no routine reason to change this.

CVB(text) (W)

Some molecules cannot be represented by simple Lewis structures. An example is HNO_3 . If artificial bonds are added, then a Lewis structure can be drawn, even if it is not very realistic. Thus, for nitric acid, the structure:



satisfies valency requirements. An unrealistic bond, such as the $\text{O}_4\text{-O}_5$ bond here, would be represented by `CVB(4:5)`. If several unrealistic bonds are needed, separate them in the keyword by commas. Thus if unrealistic bonds were needed between atoms 4 and 5, and between atoms 665 and 670, then `CVB(4:5,665:670)` would be used.

Note that `CVB` is used *only* in the generation of a starting Lewis structure. The presence of unrealistic bonds in this structure will not normally give rise to an incorrect SCF.

CYCLES=*n* (W)

When `CYCLES=n` is specified, a maximum of *n* distinct geometry cycles are allowed. After *n* cycles, the calculation is stopped in the same manner as when the allowed time is exceeded. Geometry cycles are (1) cycles of optimization in EF, SIGMA, NLLSQ, and BFGS, (2) steps in a FORCE calculation, and (3) points calculated in an IRC or DRC calculation.

`CYCLES` is intended to be an alternative to the `T=n.nn` keyword, when platforms of different speeds are being used, particularly in porting. See also `BIGCYCLES`.

DCART (O)

The Cartesian derivatives which are calculated in DCART for variationally optimized systems are printed if the keyword `DCART` is present. The derivatives are in units of kcal/Ångstrom, and the coordinates are displacements in *x*, *y*, and *z*.

DDMAX=*n.nn* (C)

The maximum value of the trust radius in EF and TS is set to *n.nn*. The largest geometry change on any cycle is `DDMAX`. Use this keyword to limit the rate of change of the geometry. If `DDMAX` is *not* set, a default of 0.3 is used in TS, and 0.5 is used in EF.

DDMIN=*n.nn* (C)

The minimum size of the trust radius in EF and TS is set to *n.nn*. If `DDMIN` is *not* set, the default of 0.001 is used. Set `DDMIN` to a small number, e.g. 0.00001, if the starting geometry is very near to the optimized geometry.

DEBUG (O)

Certain keywords have specific output control meanings, such as `FOCK`, `VECTORS` and `DENSITY`. If they are used, only the final arrays of the relevant type are printed. If `DEBUG` is supplied, then all arrays are printed. This is useful in debugging the subroutine `ITER`. `DEBUG` can also increase the amount of output produced when certain output keywords are used, e.g. `COMPFG`.

DENOUT (O)

In a MOPAC calculation, the density matrix at the end of the calculation is to be output in a form suitable for input in another job. If an automatic dump due to the time being exceeded occurs during the current run, `DENOUT` is invoked automatically. (see `RESTART`)

In a MOZYME calculation, the LMOs at the end of the calculation are to be output in a form suitable for input in another job. If an automatic dump due to the time being exceeded occurs during the current run, `DENOUT` is invoked automatically. (see `RESTART`)

A density matrix output by `DENOUT` can be read in using `OLDENS`.

DENSITY (O)

At the end of a MOPAC job, when results are being printed, the density matrix is also printed. For RHF the normal density matrix is printed. For UHF the sum of the alpha and beta density matrices is printed.

If density is not requested, then the diagonal of the density matrix, i.e., the electron density on the atomic orbitals, will be printed.

In a MOZYME job, the density matrix will be printed, *if* the size of the matrix is less than 201. If the array is larger than 200, only the submatrix of size 200 will be printed.

DEP (O)

For use only with `EXTERNAL=<filename>`. When new parameters are published, they can be entered at run-time by using `EXTERNAL=`, but as this is somewhat clumsy, a permanent change can be made by use of `DEP`.

If `DEP` is invoked, a complete block of FORTRAN code will be generated, and this can be inserted directly into the `BLOCK DATA` file.

Note that the output is designed for use with PM3. By modifying the names, the output can be used with the other methods.

DFP (W)

If `BFGS` is used, then by default, the Broyden [19]–Fletcher [20]–Goldfarb [21]–Shanno [22] method will be used to optimize geometries. The older Davidon–Fletcher–Powell method can be invoked by also specifying `DFP`. This is intended to be used for comparison of the two methods.

DIAGG2 (O)

Print some of the working in `DIAGG2`. Useful for debugging only.

DIPOLE (C)

Used in the ESP calculation, DIPOLE will constrain the calculated charges to reproduce the Cartesian dipole moment components calculated from the density matrix and nuclear charges.

DIPX=*n.nn* (C)

Similar to DIPOLE. The *x*-component of the dipole will be forced to *n.nn*. In order to use this keyword, it is necessary to also specify DIPOLE.

DIPY=*n.nn* (C)

Similar to DIPOLE. The *y*-component of the dipole will be forced to *n.nn*. In order to use this keyword, it is necessary to also specify DIPOLE.

DIPZ=*n.nn* (C)

Similar to DIPOLE. The *z*-component of the dipole will be forced to *n.nn*. In order to use this keyword, it is necessary to also specify DIPOLE.

DISEX=*n.nn* (C)

Used for the COSMO method (see EPS). In units of mean segment diameter, *n.nn* is the distance up to which the interactions of two segments is calculated as the sum of fine grid interactions. The default value is 2.0.

DISEX should not be altered unless there are problems with precision (note: precision applies to the mathematics, and not to the accuracy.) To increase the precision given by DISEX, set DISEX to a smaller number. If, after running a job with DISEX set smaller, e.g. DISEX=1.8, the results have changed significantly, then reduce DISEX again, and re-run the job.

DMAX=*n.nn* (W)

In the EF routine [7], the starting value for the trust radius is set to *n.nn*. After the first geometry optimization cycle, DMAX is changed by EF, and can increase by a factor of 2.0 on every cycle. If DMAX is *not* set, the default of 0.05 is used. (Try to avoid using this keyword.)

DOUBLET (C)

When a configuration interaction calculation is done, all spin states are calculated simultaneously, either for component of spin=0 or 1/2, unless other keywords which define the spin are present. When only doublet states are of interest, then DOUBLET can be specified, and all other spin states, while calculated, are ignored in the choice of root to be used.

Note that while almost every odd-electron system will have a doublet ground state, DOUBLET should still be specified if the desired state *must* be a doublet.

See also SINGLET, TRIPLET, QUARTET, QUINTET, SEXTET, SEPTET, OCTET, and NONET.

DOUBLET has no meaning in a UHF calculation.

DRC (C)

A Dynamic Reaction Coordinate calculation [29] is to be run. By default, total energy is conserved, so that as the 'reaction' proceeds in time, energy is transferred between kinetic and potential forms. See Section 5.19 for more details.

DRC=*n.nnn* (C)

In a DRC calculation, the 'half-life' for loss of kinetic energy is defined as *n.nnn* femtoseconds. This is equivalent to the system cooling down. If the 'half-life' is negative, energy will be added to the system. This is equivalent to heating it up. If *n.nnn* is set to zero, infinite damping simulating a very condensed phase is obtained.

DUMP=*nn.nn* (W)

Restart files are written automatically at one hour CPU time intervals to allow a long job to be restarted if the job is terminated catastrophically. To change the frequency of dump, set DUMP=*nn* to request a dump every *nn* seconds. Alternative forms are DUMP=*nn*M, DUMP=*nn*H, DUMP=*nn*D for a dump every *nn* minutes, hours, or days, respectively. DUMP only works with geometry optimization, gradient minimization, path, IRC, DRC, and FORCE calculations. It does not (yet) work with a SADDLE calculation.

ECHO (O)

Data are echoed back if ECHO is specified. Only useful if data are suspected to be corrupt.

EF (C)

The Eigenvector Following routine [7] is the default geometry optimization method. Keyword EF does nothing, but is provided to allow explicit definition of the optimizer. The alternative geometry optimizer is the BFGS [19, 20, 21, 22] method. See also DDMAX, DDMIN, DMAX, GNMIN, HESS, IUPD, LET, MODE=, NONR, NOUPD, OMIN, PRNT=, RECALC, RMAX, RMIN, RSCAL.

The current EF, while very reliable, is sometimes slow. To over-ride some safety checks, specify LET. This will sometimes make the job run faster.

EIGEN (O)

MOZYME works with localized molecular orbitals. To allow the canonical molecular orbitals, or eigenvectors, to be displayed, EIGEN should be used. It does not cause anything to be output, but it does modify the output generated by other keywords.

Keywords affected by EIGEN are:

- VECTORS: on its own, VECTORS will print out the localized molecular orbitals; if EIGEN is also present, then VECTORS will print out the eigenvectors (and eigenvalues).
- GRAPH: on its own, GRAPH will write out information for electron density graphics programs such as DENSITY and INSIGHT II. The molecular orbitals used in this case are the localized set.

If EIGEN is also present, then GRAPH will write out the eigenvectors instead of the LMOs. The graphical output would then be identical to that from MOPAC.

ENPART (O)

This is a very useful tool for analyzing the energy terms within a system. The total energy, in eV, obtained by the addition of the electronic and nuclear terms, is partitioned into mono- and bi-centric contributions, and these contributions in turn are divided into nuclear and one- and two-electron terms.

EPS=*n.nn* (C)

Sets the dielectric constant for the solvent to *n.nn*. Presence of this keyword will cause the COSMO (Conductor-like Screening Model) method [30] to be used to approximate the effect of a solvent model surrounding the molecule. Solvents with low dielectric constant are not likely to work well in this model. See also RSOLV, DISEX, NSPA, and VDW.

ESP (C)

This is the ElectroStatic Potential calculation of K. M. Merz and B. H. Besler [31]. ESP calculates the expectation values of the electrostatic potential of a molecule on a uniform distribution of points. The resultant ESP surface is then fitted to atom centered charges that best reproduce the distribution, in a least squares sense. To print out the ESP map, use POTWRT.

ESPRST (W)

ESPRST restarts a stopped ESP calculation. Do not use with RESTART.

ESR (O)

The unpaired spin density arising from an odd-electron system can be calculated both RHF and UHF. In a UHF calculation the alpha and beta M.O.s have different spatial forms, so spin density can naturally be present on in-plane hydrogen atoms such as in the phenoxy radical.

In the RHF formalism a MECI calculation is performed. In general, ESR can be used for *any* system for which $M_S \neq 0$. If the C.I. calculation results in many states, then the spin density for the state requested and the next few states will be printed. For example, if benzene cation, D_{6h} , is calculated, using ESR OPEN(3,2), then the spin density for the two degenerate components of E_{1g} will be printed. For this system, the total spin density on any given atomic orbital or any atom is given by the *average* of the spin densities for the two components. For benzene⁺, this would be 1/6 electrons.

Thus, for example, for ethylene, ESR TRIPLET C.I.=2 would give meaningful results, as would ESR MS=1 C.I.=2. However, ESR ROOT=2 C.I.=2 would not, as this would be used to calculate the spin density arising from the $M_S = 0$ component of a triplet state, which will have a zero spin density.

If the keywords OPEN and C.I.= are both absent, then only a single state is calculated. The spin density is then calculated from the state function. In order to have spin density on the hydrogens in, for example, the phenoxy radical, several states should be mixed.

EXCITED (C)

The state to be calculated is the first excited open-shell singlet state. If the ground state is a singlet, then the state calculated will be S(1); if the ground state is a triplet, then S(2). This state would normally be the state resulting from a one-electron excitation from the HOMO to the LUMO. Exceptions would be if the lowest singlet state were a biradical, in which case the EXCITED state could be a closed shell.

Line 1:	USS	Si	-34.08201495
Line 2:	UPP	Si	-28.03211675
Line 3:	BETAS	Si	-5.01104521
Line 4:	BETAP	Si	-2.23153969
Line 5:	ZS	Si	1.28184511
Line 6:	ZP	Si	1.84073175
Line 7:	ALP	Si	2.18688712
Line 8:	GSS	Si	9.82
Line 9:	GPP	Si	7.31
Line 10:	GSP	Si	8.36
Line 11:	GP2	Si	6.54
Line 12:	HSP	Si	1.32

Figure 3.4: Use of EXTERNAL=*name*

The EXCITED state will be calculated from a BIRADICAL calculation in which the second root of the C.I. matrix is selected. Abbreviation: EXCI.

Note: EXCITED is a redundant keyword, and represents a particular configuration interaction calculation. Experienced users of MECI can duplicate the effect of the keyword EXCITED by using the MECI keywords OPEN(2,2), SINGLET, and ROOT=2.

EXTERNAL=*name* (C)

Normally, PM3, AM1, MNDO, and MNDO-*d* parameters are taken from the BLOCK DATA files within MOPAC. When the supplied parameters are not suitable, as in an element recently parameterized, and the parameters have not yet installed in the user's copy of MOPAC, then the new parameters can be inserted at run time by use of EXTERNAL=<filename>, where <filename> is the name of the file which contains the new parameters.

<filename> consists of a series of parameter definitions in the format:

<Parameter><Element><Value of parameter>

where the possible parameters are USS, UPP, UDD, ZS, ZP, ZD, BETAS, BETAP, BETAD, GSS, GSP, GPP, GP2, HSP, ALP, FN_{nm}, n=1,2, or 3, and m=1 to 10, and the elements are defined by their chemical symbols, such as Si or Sl.

When new parameters for elements are published, they can be typed in as shown. This file is ended by a blank line, the word END or nothing i.e., no end-of-file delimiter). An example of a parameter data file is given in Figure 3.4 (put at least 2 spaces before and after the parameter name):

Derived parameters do not need to be entered; they will be calculated from the optimized parameters. All "constants" such as the experimental heat of atomization are already inserted for all elements.

NOTE: EXTERNAL can only be used to input parameters for MNDO, AM1, PM3, or MNDO-*d*. It is unlikely, however, that any more MINDO/3 parameters will be published.

See also DEP, to make a permanent change.

FIELD=(*n.nn*, *m.mm*, *l.ll*) (C)

An external electric field of intensity *n.nn* volts/Ångstrom in the *x*-direction, *m.mm* volts/Ångstrom in the *y*-direction, and *l.ll* volts/Ångstrom in the *z*-direction is to be applied. An ion will be strongly affected by such a field. For example, Cl⁻¹ at coordinates 1.00, 0.00, 0.00, would be stabilized by 1.0eV if a field were to be applied by specifying FIELD=(1.0,0.0,0.0).

FILL=*n* (C)

The *n*'th M.O. in an RHF calculation is constrained to be filled. It has no effect on a UHF calculation. After the first iteration (NOTE: not after the first SCF calculation, but after the first iteration within the first SCF calculation) the *n*'th M.O. is stored, and, if occupied, no further action is taken at that time. If unoccupied, then the HOMO and the *n*'th M.O.s are swapped around, so that the *n*'th M.O. is now filled. In all subsequent iterations the M.O. nearest in character to the stored M.O. is forced to be filled, and the stored M.O. replaced by that M.O. This is necessitated by the fact that in a reaction a particular M.O. may change its character considerably. A useful procedure is to run 1SCF and DENOUT first, in order to identify the M.O.s. The complete job is then run with OLDENS and FILL=*nn*, so that the eigenvectors at the first iteration are fully known. As FILL is known to give difficulty at times, consider using C.I.=*n* and ROOT=*m* instead.

FLEPO (O)

In a BFGS calculation, the predicted and actual changes in the geometry, the derivatives, and search direction for each geometry optimization cycle are printed. This is useful if there is any question regarding the efficiency of the BFGS geometry optimizer.

FMAT (O)

Details of the construction of the Hessian matrix for the force calculation are to be printed.

FORCE (C)

A force-calculation is to be run. The Hessian, that is the matrix in millidynes per Ångstrom) of second derivatives of the energy with respect to displacements of all pairs of atoms in *x*, *y*, and *z* directions, is calculated. On diagonalization this gives the force constants for the molecule. The force matrix, weighted for isotopic masses, is then used for calculating the vibrational frequencies. The system can be characterized as a ground state or a transition state by the presence of five (for a linear system) or six eigenvalues which are very small (less than about 30 reciprocal centimeters). A transition state is further characterized by one, and exactly one, negative force constant.

Before a FORCE calculation is run, the gradients are calculated to see if the geometry is at a stationary point. If it is not, then the calculation will be stopped, to allow the user to take corrective action.

Sometimes, the gradient norm at the start of a FORCE calculation will be larger than at the end of the geometry optimization which was used to generate the geometry for the force calculation. This is due to the FORCE calculation using a different method, double-sided derivatives, to calculate the gradients. In order to have the same GNORM at the end of a geometry optimization as at the start of a FORCE calculation, use PRECISE in the geometry optimization. Gradients calculated with PRECISE and with FORCE both use double-sided derivatives.

A FORCE calculation is a prerequisite for a THERMO calculation.

At the end of a FORCE calculation, the force constants for the coordinates supplied will be printed. If other force constants are needed, then use ISOTOPE to save the Hessian. The connectivity can then be changed, and the job restarted using RESTART. Of course care must be taken to ensure that the atoms are in exactly the same positions in both calculations.

Before a FORCE calculation is started, a check is made to ensure that a stationary point is being used. This check involves calculating the gradient norm (GNORM) and if it is significant, the calculation will be stopped. See also THERMO, LET, TRANS, ISOTOPE.

In a FORCE calculation, PRECISE will eliminate quartic contamination part of the anharmonicity). This is normally not important, therefore PRECISE should not routinely be used. In a FORCE calculation, the SCF criterion is automatically made more stringent; this is the main cause of the SCF failing in a FORCE calculation.

GEO-OK (W)

Normally the program will stop with a warning message if two atoms are within 0.8 Ångstroms of each other, or, more rarely, if the BFGS routine has difficulty optimizing the geometry.

Sometimes in a C.I., accidental degeneracy in the SCF energy levels will cause the C.I. to fail. In that case, GEO-OK should be used. See Chapter 8.

In order for C.I. calculations to be valid, the SCF energy levels used in the C.I. calculation must be different from those which are not used in the C.I. In other words, a degenerate manifold must not be partly included in a C.I. If, for any reason, a part of a degenerate manifold is to be used, then GEO-OK should be specified.

GEO-OK will override the job termination sequence and allow the calculation to proceed. In practice, most jobs that terminate due to these checks contain errors in data, so caution should be exercised if GEO-OK is used. An important exception to this warning is when the system contains, or may give rise to, a hydrogen molecule. GEO-OK will override other geometric safety checks such as the unstable gradient in a geometry optimization preventing reliable optimization.

See also the message ‘GRADIENTS OF OLD GEOMETRY, GNORM= *nn.nnnn*’.

GEOCHK (O)

GEOCHK prints some of the working in subroutine GEOCHK. The output is quite terse: it is designed for debugging. Possibly the only routine use for GEOCHK is to generate the list of covalent bonds, which is the first set of data printed by GEOCHK. In it, each atom is listed in the order in which it occurs in the data-set. The atom numbers for every atom covalently bonded to each atom are printed after the atom label. Thus for a protein in which isoleucine is the first residue, and in which the atoms are in the order determined by RESEQ, the list of covalent bonds in the first residue would be that shown in Figure 3.5.

GNMIN (C)

If GNMIN is present in a EF or TS calculation, then a step will be rejected unless the gradient norm drops. (Try to avoid using this keyword).

GNORM=*n.nn* (W)

The geometry optimization termination criteria (see Chapter 9) in both gradient minimization and energy minimization can be over-ridden by specifying a gradient norm requirement. For example, GNORM=20 would allow the geometry optimization to exit as soon as the gradient norm dropped below 20.0, the default being 1.0.

For high-precision work, GNORM=0.0 is recommended. Unless LET is also used, the GNORM will be set to the larger of 0.01 and the specified GNORM. Results from GNORM=0.01 are easily good enough for all high-precision work.

N.b.: Do not confuse GNORM, the keyword, with GNORM, the value of the scalar of the calculated gradient.

GRADIENTS (O)

In a 1SCF calculation gradients are not calculated by default: in non-variationally optimized systems this could take a lot of time. GRADIENTS allows the gradients to be calculated. Normally, gradients will not be printed if the

1	N(1	ILE*	1)	2	19	20		
2	C(2	ILE*	1)	1	3	5	9	
3	C(3	ILE*	1)	2	4	21		
4	O(4	ILE*	1)	3				
5	C(5	ILE	1)	2	6	7	10	
6	C(6	ILE	1)	5	11	12	13	
7	C(7	ILE	1)	5	8	14	15	
8	C(8	ILE	1)	7	16	17	18	
9	H(9	ILE*	1)	2				
10	H(10	ILE	1)	5		H12		H17
11	H(11	ILE	1)	6				
12	H(12	ILE	1)	6		H11--C6--H13	H16--C8--H18	
13	H(13	ILE	1)	6				
14	H(14	ILE	1)	7				
15	H(15	ILE	1)	7		H10--C5-----C7--H15		
16	H(16	ILE	1)	8				
17	H(17	ILE	1)	8	H20		O4	H14
18	H(18	ILE	1)	8	\		/	
19	H(19	ILE*	1)	1	N1---C2---C3			
20	H(20	ILE*	1)	1	/		\	
21	N(21	GLN*	2)	3	H19	H9	N21--(GLN)	

Figure 3.5: List of Covalent Bonds in Isoleucine

gradient norm is less than 2.0. However, if GRADIENTS is present, then the gradient norm and the gradients will unconditionally be printed. Abbreviation: GRAD.

GRAPH (O)

Information needed to generate electron density contour maps can be written to a file by using the keyword GRAPH. GRAPH first calls MULLIK in order to generate the inverse-square-root of the overlap matrix, which is required for the re-normalization of the eigenvectors. All data essential for the graphics package DENSITY are then output.

In MOZYME, by default, GRAPH will use LMOs. If EIGEN is also present, then the molecular orbitals used will be the eigenvectors.

GREENF (C)

The ionization potentials are to be corrected using the outer valence Green's Function (OVGF) technique. See Section 5.3.6 for more detail. The technique has been shown to be effective in improving the I.P.s of nitrogen heterocycles. See [32] for triazines and tetrazines, and [33] for pyridines. The technique uses several pre-set variables. These can be changed by the user by enclosing the names and values of the variables in brackets, thus: GREENF(VAR1=*n.nn*, VAR2=*m.mm*, ...). Users are *not* recommended to change these values until they know what they are doing. The OVGF technique and variables and their definitions were designed and written by Dr David Danovich.

Variables which can be modified by the user are:

NOCC The total number of occupied orbitals to be included in the Green's Function calculation. By default, all occupied M.O.s are included, up to a maximum of 20.

NVIR The total number of virtual or unoccupied orbitals to be included in the Green's Function calculation. By default, all virtual M.O.s are included, up to a maximum of 20.

PRINT Print flag. Normally set to 0.

IGGWW Total number of occupied M.O.s involved in the OVGf calculation. Default: NOCC.

IGGV Total number of unoccupied M.O.s involved in the OVGf calculation. Default: NVIR

NIZ The number of the first occupied molecular orbital for which OVGf correction must be started. (The numbering refers to the SCF calculation). Default: HOMO-7.

IVER The number of the last occupied M.O. to which OVGf correction is to be applied. (The numbering refers to the SCF calculation). Default: HOMO.

EPS The tolerance of the iteration procedure for solution of the Dyson Equation. Default: 0.0001.

ITER2 ITER2=0 if second-order contribution in the self-energy function is to be calculated, this is the default.
ITER2=1 if second-order contribution in the self-energy function is *not* to be calculated. See [34, 35, 36, 33, 37, 32].

ITER3 ITER3=0 if third-order contribution in the self-energy function is to be calculated, this is the default.
ITER3=1 if third-order contribution in the self-energy function is *not* to be calculated. See [34, 35, 36, 33, 37, 32].

IFUL The formula used for calculating the screening factor, which takes into account the higher than third order contributions in the self-energy function, as detailed in [35]. Default: IFUL=0.

- IFUL=0 Formula No. C25 from [35] to be used.
- IFUL=1 Formulas C27, C28 and C29 from [35] to be used.
- IFUL=2 Formula C30 from [35] to be used.

IPRINT IPRINT=0 if I.P.s used in the OVGf calculation are *not* to be printed. IPRINT=1 if I.P.s used in the OVGf calculation are to be printed. Default: IPRINT=0.

IT23 IT23=0 if the third-order contribution and complete expansion of the self-energy is to be calculated. IT23=1 if the third-order contribution and complete expansion of the self-energy is *not* to be calculated. Default: IT23=1

H-PRIORITY (O)

In a DRC calculation, results will be printed whenever the calculated heat of formation changes by 0.1 kcal/mol. Abbreviation: H-PRI0. See Section 5.19 for more details.

H-PRIORITY=*n.nn* (O)

In a DRC calculation, results will be printed whenever the calculated heat of formation changes by *n.nn* kcal/mol.

H20 (C)

Because the extra data for the MST calculations is so complicated, the keyword H20 has been written. When H20 is present, the extra data required by TOM is supplied by MOPAC. While this makes MST calculations easier to run, it does reduce flexibility. When H20 is present, the following conditions are used:

- The solvent is assumed to be water.

- The dielectric is set to 78.5 (the value for water).
- The temperature is set to 298.15K.
- The volume of the solvent molecule is set to 18.07\AA^3 .
- The diameter of the solvent molecule is set to 2.77\AA .
- The thermal expansion coefficient is set to 0.00026K^{-1} .
- The surface tension is set to $71.69\text{ dyne-cm}^{-2}$.
- The surface tension derivative is set to 0.657.
- The cavity microscopic coefficient is set to 1.227.
- The ORT option is used.
- The Van der Waals radii as defined in Orozco, Bachs, and Luque, J. Comp. Chem., 16 563-575 (1995) are used.

These conditions allow the results reported in J. Comp. Chem, Vol 16, p. 586, Table II, option (O) to be reproduced. From Orozco, et. al.'s article, the preferred method is AM1.

HESS=*n* (**W**)

In Baker's Eigenvector Following routine, options exist for deciding how to construct the initial Hessian matrix. The default is HESS=0. Options available are:

HESS=0

This is the default for geometry optimization (i.e., when EF is used). The initial Hessian is set equal to a diagonal matrix, with the diagonal terms set to $1000\text{ kcal/mol/\AA}^2$ for bond-lengths, $500\text{ kcal/mol/degrees}^2$ for angles, and $200\text{ kcal/mol/degrees}^2$ for dihedrals. If Cartesian coordinates are used, all diagonal elements are set to $200\text{ kcal/mol/\AA}^2$.

Do *not* specify HESS=0 unless there is a good reason to do so.

HESS=1

This is the default for transition-state location (i.e., when TS is used). The full Hessian matrix is constructed using single-sided derivatives, see Section 5.13.2, using the same density matrix throughout the entire construction of the Hessian.

HESS=2

A rarely used option. The Hessian matrix from an earlier run can be used to start the current job. In order for this to work, there must be a one-to-one correspondence of parameters to be optimized. For example, if a geometry optimization using EF and AM1 were to be followed by a similar geometry optimization using EF and PM3, then the Hessian from the earlier calculation could be used to start the PM3 job. (A simpler way of achieving this result is to use RESTART, but note that this will also use the old geometry.)

```

hybrid PM3 MOZ      1SCF      MOPAC
      METHANE HEAT OF FORMATION (PM3) = -13.026

H
C   1.08701794  1
H   1.08701784  0  109.4712210  0
H   1.08701784  0  109.4712210  0 -120.00  0   2  1  3
H   1.08701784  0  109.4712210  0  120.00  0   2  1  3

```

Figure 3.6: Standard MOPAC-type Data Set for Methane

```

ATOM NUMBER  2
0.0000
0.0000  0.0000
0.0000  0.0000  0.0000
0.0000  0.0000  0.0000  0.0000      (Secular Determinant)
-5.1852  3.1518  0.0300  0.0400  0.0000
-5.1852 -1.0239 -2.9227  0.0400  0.0000  0.0000
-5.1852 -1.0239  1.5063  2.5971  0.0000  0.0000  0.0000
-5.1852 -1.0239  1.5063 -2.5171  0.0000  0.0000  0.0000  0.0000
-10.3711 -3.6163 -3.6163 -3.6160  3.6160  3.6163  3.6163  10.3711
-0.7071  0.0031  0.0047  0.0062 -0.3545 -0.3519 -0.3554 -0.3524
0.0000  0.0000 -0.5657  0.4243  0.0000 -0.4619 -0.0691  0.5309
0.0000 -0.6565  0.1576  0.2102  0.5685 -0.0608 -0.4025 -0.1052
-0.0084 -0.2627 -0.3939 -0.5251  0.2261 -0.3989  0.4549 -0.2879

```

Hybrid orbitals

```

-0.3566 -0.2031 -0.5758  0.0020  0.0000 -0.7071  0.0000  0.0000
0.3557  0.2031 -0.2902  0.4980  0.0000  0.0000  0.0000  0.7071
0.3518 -0.6134 -0.0015 -0.0020  0.7071  0.0000  0.0000  0.0000
0.3500  0.2031 -0.2902 -0.5020  0.0000  0.0000  0.7071  0.0000

      -0.50431  -0.28724  -0.81434  0.00287
      0.50310   0.28724  -0.41040  0.70424
      0.49751  -0.86745  -0.00214  -0.00285
      0.49502   0.28725  -0.41038  -0.70995

```

Figure 3.7: Example of Output Generated by HYBRID

```

      S H 1   S C 2   Px C 2   Py C 2   Pz C 2   S H 3   S H 4   S H 5
-----
S H 1 -71.1558
S C 2 -5.1952 -84.0140
Px C 2  3.1318  0.0000 -72.1164
Py C 2  0.0000  0.0000  0.0000 -72.1164
Pz C 2  0.0000  0.0000  0.0000  0.0000 -72.1164
S H 3 -1.6988 -5.1952 -1.0439 -2.9527  0.0000 -71.1558
S H 4 -1.6988 -5.1952 -1.0439  1.4763  2.5571 -1.6988 -71.1558
S H 5 -1.6988 -5.1952 -1.0439  1.4763 -2.5571 -1.6988 -1.6988 -71.1558

```

Figure 3.8: One Electron Matrix for Methane

HYBRID (O)

The keyword `HYBRID` will print some of the working of the subroutine `HYBRID`, which constructs the hybrid orbitals used in making the starting LMOs. Thus, if `HYBRID` is used with methane (run using `PM3` with standard coordinates (Figure 3.6 , p. 49)) then `HYBRID` would generate the output shown in Figure 3.7 , p. 49 .

To understand the output, it is useful to be able to refer to the one-electron matrix, Figure 3.8 , p. 49 .

The eight lines below `ATOM NUMBER 2` show the elements of the secular determinant. This determinant can be described as follows:

- The rows and columns of the determinant represents the atomic orbitals of the atom being hybridized (here, carbon) and the *s* atomic orbitals of the atoms bonded to it, in order.
- A block of 10 elements, all zero. This is common to all determinants used in `HYBRID`. These terms represent pairs of atomic orbitals on carbon.
- The first 4 elements on line 5 are the interaction of the atomic orbitals on carbon with the 1*s* atomic orbital on hydrogen 1.
- The first 4 elements on line 6 are the interaction of the atomic orbitals on carbon with the 1*s* atomic orbital on hydrogen 3.
- The first 4 elements on line 7 are the interaction of the atomic orbitals on carbon with the 1*s* atomic orbital on hydrogen 4.
- The first 4 elements on line 8 are the interaction of the atomic orbitals on carbon with the 1*s* atomic orbital on hydrogen 5.
- All monatomic terms and all terms between the *s* orbitals of the ligands are set to zero.
- A small perturbation is applied to each interaction term. This amounts to 0.0001 times the address of the array element. This perturbation is necessary in order to destroy any accidental symmetry of the kind that can form in simple unsaturated systems such as benzene.

Diagonalization of this determinant gives eight eigenvalues: -10.3711 -3.6163, etc., and eight eigenvectors, the first four of which can be used in constructing the hybrids. A unitary transform of the first four eigenvectors generates hybrid orbitals: these are printed under the heading `Hybrid orbitals`. Only the four coefficients on the carbon are of interest, so these are extracted off and re-normalized. This gives the final block of output, which is the hybrid orbitals themselves.

The unitary transform to convert from eigenvectors to hybrids is the localization operation. See p. 171 for more detail.

HYPERFINE (O)

The quantity:

$$\frac{1}{3}(P_s^\alpha - P_s^\beta + 2(\psi_s^\alpha)^2)$$

is printed for all atoms in a radical when `HYPERFINE` is specified for a UHF system. Only the *s* orbital coefficients are used, as these are the only orbitals that have a non-zero coefficient at the nucleus. The units are in electrons. The ψ_s^α refers to the highest occupied alpha-spin molecular orbital. This quantity is of use in predicting hyperfine coupling coefficients.

INT (W)

Regardless of what type of coordinates are used in the data set, INT will force all the coordinates to be internal coordinates. Atom 1 is, by definition, always defined in Cartesian coordinates.

IRC (C)

An Intrinsic Reaction Coordinate calculation is to be run. All kinetic energy is shed at every point in the calculation. See Section 5.19 for more details.

IRC=*n* (C)

An Intrinsic Reaction Coordinate calculation to be run; an initial perturbation in the direction of normal coordinate *n* to be applied. If *n* is negative, then perturbation is reversed, i.e., initial motion is in the opposite direction to the normal coordinate.

ISOTOPE (O)

Generation of the FORCE matrix is very time-consuming, and in isotopic substitution studies several vibrational calculations may be needed. To allow the frequencies to be calculated from the (constant) force matrix, ISOTOPE is used. When a FORCE calculation is completed, ISOTOPE will cause the force matrix to be stored, regardless of whether or not any intervening restarts have been made. To re-calculate the frequencies, etc., starting at the end of the force matrix calculation, specify RESTART.

The two keywords RESTART and ISOTOPE can be used together. For example, if a normal FORCE calculation runs for a long time, the user may want to divide it up into stages and save the final force matrix. Once ISOTOPE has been used, it does not need to be used on subsequent RESTART runs.

ISOTOPE can also be used with FORCE to set up a RESTART file for an IRC=*n* calculation.

ITRY=*nn* (W)

The default maximum number of SCF iterations is 200. When this limit presents difficulty, ITRY=*nn* can be used to re-define it. For example, if ITRY=400 is used, the maximum number of iterations will be set to 400. ITRY should normally not be changed until all other means of obtaining a SCF have been exhausted, e.g. PULAY, CAMP-KING, etc.

IUPD=*n* (W)

IUPD is used only in the EF routine. IUPD should very rarely be touched. IUPD controls how the Hessian is updated. Values for IUPD are 0: skip the update; 1: Use Powell's method; 2: use the BFGS update. For more information, see Section 5.7.1.

K=(*n.nn*,*n*) (C)

Used in band-structure calculations, K=(*n.nn*,*n*) specifies the step-size in the Brillouin zone, and the number of atoms in the monomeric unit. Two band-structure calculations are supported: electronic and phonon. Both require a polymer to be used. If FORCE is used, a phonon spectrum is assumed, otherwise an electronic band structure is assumed. For both calculations, a density of states is also done. The band structure calculation is very fast,

so a small step-size will not use much time. The output is designed to be fed into a graphics package, and is not 'elegant'. For polyethylene, a suitable keyword would be $K=(0.01, 6)$.

The order in which the atoms in a polymer are defined is important when band phenomena are calculated.

See p. 195 for more information.

KINETIC=*n.nnn* (C)

In a DRC calculation *n.nnn* kcal/mol of excess kinetic energy is added to the system as soon as the kinetic energy builds up to 0.2 kcal/mol. The excess energy is added to the velocity vector, without change of direction. See Section 5.19 for more details.

KING (W)

The Camp-King converger (q.v.) is to be used. This is a very powerful, but CPU intensive, SCF converger.

LARGE (O)

Most of the time the output invoked by keywords is sufficient. LARGE will cause less-commonly wanted, but still useful, output to be printed. LARGE increases the amount of output generated by the keywords GEOCHK, DCART, LEWIS, TIDY, and MECI.

When COMPG is specified, LARGE will cause all coordinates to be printed; the default is the first 5 atoms.

When DCART is specified, LARGE will cause all derivatives in a solid to be printed; the default is the central unit cell only.

When DERNVO is specified, LARGE will cause details of the non-variationally optimized derivatives to be printed.

When FMAT is specified, LARGE will cause details of the construction of the Hessian to be printed.

In a FORCE calculation, LARGE will cause the force-constants to be printed. The default is to print the normal coordinates only.

When MECI is specified, LARGE will cause details of the multi-electron configuration interaction calculation to be printed. This includes the secular determinant and State vectors.

To save space, DRC and IRC outputs will, by default, only print the line with the percent sign. Other output can be obtained by use of the keyword LARGE, according to the following rules:

LARGE Print all internal and Cartesian coordinates and Cartesian velocities.

LARGE=1 Print all internal coordinates.

LARGE=-1 Print all internal and Cartesian coordinates and Cartesian velocities.

LARGE=*n* Print every *n*'th set of internal coordinates.

LARGE=-*n* Print every *n*'th set of internal and Cartesian coordinates and Cartesian velocities.

To reduce output, do not use LARGE.

LET (W)

As MOPAC evolves, the meaning of LET is changing.

Now LET means essentially "I know what I'm doing, override safety checks".

Currently, LET has the following meanings:

1. In a FORCE calculation, it means that the supplied geometry is to be used, even if the gradients are large.
2. In a geometry optimization, the specified GNORM is to be used, even if it is less than 0.01.
3. In a POLAR calculation, the molecule is to be orientated along its principal moments of inertia before the calculation starts. LET will prevent this step being done.
4. In a EF calculation, allow the ΔH_f to rise.

LEWIS (O)

In MOZYME, prints out the connectivity table generated by subroutine LEWIS, and the Lewis structure generated by MAKVEC.

Because LEWIS is intended to be used as a diagnostic tool only, the calculation is stopped after the Lewis structure is printed. An SCF will *not* be run.

LINMIN (O)

There are two line-minimization routines in MOPAC, an energy minimization (if BFGS is used) and a gradient norm minimization (if SIGMA is used). LINMIN will output details of the line minimization used in a given job.

LOCALIZE (O)

The occupied eigenvectors are transformed into a localized [38] set of M.O.s by a series of 2 by 2 rotations which maximize $\langle \psi^4 \rangle$. The value of $1/\langle \psi^4 \rangle$ is a direct measure of the number of centers involved in the MO: Thus, the value of $1/\langle \psi^4 \rangle$ is 2.0 for H₂, 3.0 for a three-center bond and 1.0 for a lone pair. Higher degeneracies than allowed by point group theory are readily obtained. For example, benzene would give rise to a 6-fold degenerate C–H bond, a 6-fold degenerate C–C sigma bond and a three-fold degenerate C–C pi bond. In principle, there is no single step method to unambiguously obtain the most localized set of M.O.s in systems where several canonical structures are possible, just as no simple method exists for finding the most stable conformer of some large compound. However, the localized bonds generated will normally be quite acceptable for routine applications.

In MOZYME, when LOCAL is specified, a description of the localized molecular orbitals is printed. Abbreviation: LOCAL.

MAX (W)

In a grid calculation, MAX requires that the maximum number of points (23) in each direction is to be used. The default is 11. If less than the maximum number of points is to be used, then the number of points in each direction can be set with POINT1 and POINT2.

MECI (O)

If MECI is specified, then details of the Multi Electron Configuration Interaction calculation [39] are printed at the end of the calculation. The state vectors can be printed by specifying MECI and LARGE. The MECI calculation is either invoked automatically, or explicitly invoked by the use of the C.I.=*n* keyword. See Section 5.4 for more detail.

```

Line
*      1scf AM1 mep=1
* Formaldehyde (Cross-section in plane of molecule)
* Generate a 2-D grid of MEP potentials for 'meplot' to use
* O  0.00000000 0  0.0000000 0  0.0000000 0  0 0 0
* C  1.22732374 1  0.0000000 0  0.0000000 0  1 0 0
* H  1.11047287 1 122.2253516 1  0.0000000 0  2 1 0
* H  1.11048351 1 122.2158646 1 179.9998136 1  2 1 3
*
1 -2.3 2.7 0 (Note: For 'meplot' to work, the first coordinate of line 2
2 -2.3 -2.7 0 here must be the same as the first coordinate of line 1)
3  4.1 2.7 0
4  0.2

```

Figure 3.9: Calculation of ESP Cross-Section in Formaldehyde using the Orozco-Luque Model

Table 3.5: Extra Data Required by MEP=1

Extra Line	Description
1	Bottom left-hand corner of cross-section (three numbers) (All coordinates are Cartesian and are in Ångstroms)
2	Top left-hand corner of cross-section (three numbers)
3	Bottom right-hand corner of cross-section (three numbers)
4	Step-size in Ångstroms for points from (1) to (2)

MEMORY=*n* (W)

If MEMORY is specified, then MOPAC will try to keep memory usage below *n* mega-bytes. Default value is 128 mega-bytes.

MEP=1 (O)

The Orozco-Luque molecular electrostatic potential map [40] for a cross-section through a system is calculated and printed. The output is written to <filename>.mep and <filename>.tab

In order to use MEP=1, extra data are needed at the end of the normal data-set. The extra data consist of four lines of numbers, as defined in Figure 3.5.

An example of a MEP=1 data set is given in Figure 3.9; the corresponding plot is shown in Figure 3.10. This plot can be compared with the on page 246.

MEP=2 (O)

The Orozco-Luque molecular electrostatic potential map for the Connolly surface is generated and printed. As with MEP=1, the output is written to <filename>.mep and <filename>.tab.

In order to use MEP=2, extra data are needed at the end of the normal data-set. The extra data consist of one line of numbers, as defined in Figure 3.6. An example of a MEP=2 data set is given in Figure 3.11.

MICROS=*n* (C)

The microstates used by MECI are normally generated by use of a permutation operator. When individually defined microstates are desired, then MICROS=*n* can be used, where *n* defines the number of microstates to be read in.

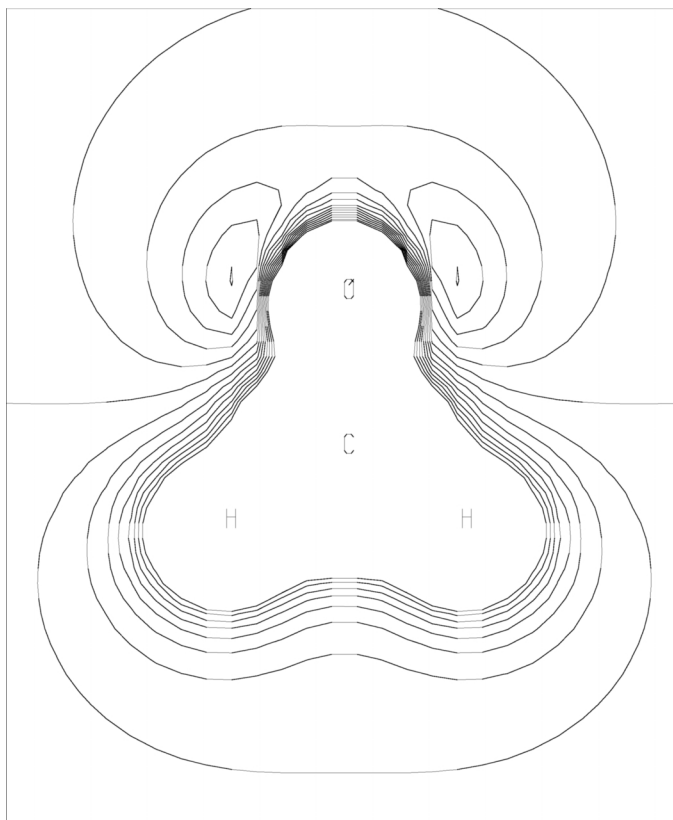


Figure 3.10: Molecular Electrostatic Potential around Formaldehyde

```

Line
*      1scf AM1 mep=1
* Formaldehyde Electrostatic Potential on Connolly Surface
*
* O      0.00000000  0    0.0000000  0    0.0000000  0    0    0    0
* C      1.22732374  1    0.0000000  0    0.0000000  0    1    0    0
* H      1.11047287  1   122.2253516  1    0.0000000  0    2    1    0
* H      1.11048351  1   122.2158646  1   179.9998136  1    2    1    3
*
1    1.4 0.2 0.5 1

```

Figure 3.11: Calculation of ESP Connolly Surface in Formaldehyde using the Orozco-Luque Model

Table 3.6: Extra Data Required by MEP=2

Layout of data: A B C D			
Datum	Value	Name	Description
A	1.4	SCALE	Scale factor which determines the size of the surface. The radii of the spheres (centered on each atom) are determined by applying this factor to the van der Waals' radii.
B	0.2	SCINCR	Separation (in Angstroms) between surfaces
C	0.5	DEN	Density of points on each surface
D	1	NSURF	Number of surfaces

Table 3.7: States Arising from Various Microstates

Microstate	No. of alpha	beta electrons	M_S	State
1100	2	0	1	Triplet
1010	1	1	0	Singlet
1001	1	1	0	Mixed
0110	1	1	0	Mixed
0101	1	1	0	Singlet
0011	0	2	-1	Triplet

Format for Microstates

After the geometry data and any symmetry data are read in, data defining each microstate is read in, using format 2011, at one microstate per line. The microstate data is preceded by the word "MICROS" on a line by itself. Examples of microstates for two electrons in two M.O.s are given in Table 3.7.

For a system with n M.O.s in the C.I. (use OPEN=(m, n) or C.I.= n to do this), the populations of the n alpha M.O.s are defined, followed by the n beta M.O.s. Allowed occupancies are zero and one. For $n=6$ the closed-shell ground state would be defined as 111000111000, meaning one electron in each of the first three alpha M.O.s, and one electron in each of the first three beta M.O.s.

Users are warned that they are responsible for completing any spin manifolds. Thus, while the state 111100110000 is a triplet state with component of spin = 1, the state 111000110100, while having a component of spin $M_S = 0$, is neither a singlet nor a triplet. In order to complete the spin manifold the microstate 110100111000 must also be included.

If a manifold of spin states is not complete, then the eigenstates of the spin operator will not be quantized.

There are two other limitations on possible microstates. First, the number of electrons in every microstate should be the same. If they differ, a warning message will be printed, and the calculation continued, but the results will almost certainly be nonsense. Second, the component of spin for every microstate must be the same, except for teaching purposes. Two microstates of different components of spin will have a zero matrix element connecting them. No warning will be given as this is a reasonable operation in a teaching situation. For example, if all states arising from two electrons in two levels are to be calculated, say for teaching Russell-Saunders coupling, then the microstates given in Table 3.7 would be used.

Constraints on the space manifold are just as rigorous, but much easier to satisfy. If the energy levels are degenerate, then all components of a manifold of degenerate M.O.s should be either included or excluded. If only some, but not all, components are used, the required degeneracy of the states will be missing.

As an example, for the tetrahedral methane cation, if the user supplies the microstates corresponding to a component of spin $M_S = 3/2$, neglecting Jahn-Teller distortion, the minimum number of states that can be supplied is $90 = (6!/(1!5!))(6!/(4!2!))$. This corresponds to the configuration $(t_2)^3(\bar{t}_2)^1(t_2^*)^1(\bar{t}_2^*)^0$.

The format is defined as 2011 so that spaces can be used for empty M.O.s.

MINDO/3 (C)

The default Hamiltonian within MOPAC is MNDO, with the alternatives of MNDO- d , PM3, AM1 and MINDO/3. To use the MINDO/3 Hamiltonian [2] the keyword MINDO/3 should be used. Acceptable alternatives to the keyword MINDO/3 are MINDO and MINDO3.

MINMEP (O)

Print minima in PMEP plot. Units used are kcal/mol.

MMOK (C)

If the system contains a peptide linkage, then `MMOK` will allow a molecular mechanics correction to be applied so that the barrier to rotation is increased (to 14.00 kcal/mol in N-methyl acetamide).

MNDO (C)

Use the MNDO Hamiltonian. This is the default, and therefore this keyword is not necessary. It is supplied so that the method can be explicitly defined.

MNDOD (C)

The MNDO-*d* method [5, 6] is to be used.

MODE=*n* (C)

`MODE` is used in the EF routine. Normally the default `MODE=1` is used to locate a transition state, but if this is incorrect, explicitly define the vector to be followed by using `MODE=n`. (`MODE` is not a recommended keyword). If you use the `FORCE` option when deciding which mode to follow, set all isotopic masses to 1.0. The normal modes from `FORCE` are normally mass-weighted: this can mislead. Alternatively, use `LARGE` with `FORCE`: this gives the force constants and vectors in addition to the mass-weighted normal modes.

MOLDAT (O)

Print some of the working in MOLDAT. Useful as a diagnostic only

MOPAC (W)

The keyword `MOPAC` is provided in order to allow compatibility with earlier versions of MOPAC. The main difference between `MOZYME` and `MOPAC` formats is that `MOPAC` does not require the connectivity of the second or third atoms to be explicitly defined. In general, `MOPAC` should be used for data sets constructed using an editor; data sets generated by GUIs normally give the connectivities of atoms 2 and 3.

MOZYME (W)

The keyword `MOZYME` alters the way in which the SCF is generated. When `MOZYME` is used, calculations on large systems require only a small fraction (a few percent) of the memory needed for an equivalent conventional calculation, and run very much faster (for 4,000 atoms, the jobs run about 200 times faster.)

Limitations

- Only closed shell RHF calculations are allowed. This means that `MOZYME` calculations are limited to species in their ground state. Radicals and electronic excited states cannot be run.

- For large systems, the recommended geometry optimizer is LBFSGS. This is a modified BFGS optimizer designed to minimize memory usage.

If LBFSGS is not wanted for any reason, then use BFGS, although it uses a lot more memory.

The default optimizer, EF, uses a large amount of memory, and should therefore not be used in optimizing the geometry of large systems. In addition, because it uses a matrix inversion, it becomes very time consuming for large systems.

- The results are not so precise, so for runs that need high precision (such as FORCE calculations), MOZYME should not be used.
- When the geometry of a *large* system is optimized, the run should be broken into a series of short calculations, of 50–200 cycles. This is necessary because the geometry might change considerably during the optimization, and atoms that were not interacting at the start of the calculation might interact strongly at the end of the run. By stopping and restarting the optimization, any new interactions are taken into account. Failure to break up an optimization in this way can lead to a ruined calculation. (This is really a bug—it will be corrected in a future release)

Recommended use of MOZYME

MOZYME can be used for simple geometric calculations, such as geometry optimization, transition state location, and dynamic reaction coordinates, and for the calculation of polarizability. For these calculations, MOZYME can be run as a “stand alone” calculation. If a partial geometry optimization is run, then the use of RAPID should be considered.

For calculation of vibrational frequencies, frequency-dependent polarizability, IRCs, and electronic excited states, MOZYME should be run first, to optimize the geometry, then a conventional MOPAC calculation run.

Another effective strategy is to run a MOZYME job, followed by a MOPAC job, using the OLDGEO feature. When geometry optimizations are being run, a MOZYME job can be run for a time, then a MOPAC job run, using RESTART. That is, the RESTART function will work when a geometry optimization or transition state location calculation is run, regardless of the method used in generating the SCF.

MS=*n* (**W**)

Most often used for checking the MECI calculation and for teaching. MS=*n* overrides the normal choice of magnetic component of spin. Normally, if a triplet is requested, an M_S of 1 will be used; this excludes all singlets. If MS=0 is also given, then singlets will also be calculated. For an odd-electron system, MS=-0.5 will give the same results as MS=0.5. The use of MS should not affect the values of the results at all.

MULLIK (**O**)

A full Mulliken Population analysis [41, 42] is to be done on the final RHF wavefunction. This involves the following steps:

1. The eigenvector matrix is divided by the square root of the overlap matrix, S .
2. The Coulson-type density matrix, P , is formed.
3. The overlap population is formed from $P_{\lambda\sigma}S_{\lambda\sigma}$.
4. Half the off-diagonals are added onto the diagonals.

NEWGEO (W)

NEWGEO sets up a new coordinate system for optimizing proteins. In this, all backbone atoms are defined using Cartesian coordinates, and all other atoms are defined in internal coordinates.

NLLSQ (W)

The gradient norm is to be minimized by Bartel's method [8]. This is a Non-Linear Least Squares gradient minimization routine. Gradient minimization will locate one of three possible points:

(a) A minimum in the energy surface. The gradient norm will go to zero, and the lowest five or six eigenvalues resulting from a FORCE calculation will be approximately zero.

(b) A transition state. The gradient norm will vanish, as in (a), but in this case the system is characterized by one, and only one, negative force constant.

(c) A local minimum in the gradient norm space. In this (normally unwanted) case the gradient norm is minimized, but does not go to zero. A FORCE calculation will not give the five or six zero eigenvalues characteristic of a stationary point. While normally undesirable, this is sometimes the only way to obtain a geometry. For instance, if a system is formed which cannot be characterized as an intermediate, and at the same time is not a transition state, but nonetheless has some chemical significance, then that state can be refined using NLLSQ.

NLMO=*nnn* (W)

The average number of atoms in a localized molecular orbital is defined as *nnn*. The default is usually adequate, if it is not then an advisory message will be printed, and the number can be raised using NLMO=*nnn*.

The amount of memory needed can be lowered by reducing *nnn*. However, lowering it too much will result in the job failing. If that happens, follow the instructions in the output in order to correct the fault.

NOANCI (W)

RHF open-shell derivatives are normally calculated using Liotard's analytical C.I. method [43]. In general, do *not* use NOANCI (NO ANalytical Configuration Interaction derivatives). NOANCI should *only* be used if there is cause to believe that the derivatives are faulty. Liotard's method is very fast, so using NOANCI will cause the job to take much longer than if NOANCI were *not* used.

In the unusual situation when analytical C.I. derivatives are *not* to be used, specify NOANCI. See also ANALYT and Section 5.13.

NOINTER (O)

The interatomic distances are printed by default. If you do not want them to be printed, specify NOINTER. By default, interatomic distances are not printed for systems of more than 200 atoms (unless LARGE is present), or for MOZYME calculations.

NOLOG (O)

Normally a copy of the archive file will be directed to the LOG file, along with a synopsis of the job. If this is not wanted, it can be suppressed completely by NOLOG.

NOMM (C)

All four semiempirical methods underestimate the barrier to rotation of a peptide bond. A Molecular Mechanics correction has been added which increases the barrier in N-methyl acetamide to 14 kcal/mol. If you do not want this correction, specify `NOMM` (NO Molecular Mechanics).

NONET (C)

RHF interpretation: The desired spin-state is a nonet: the state with component of spin $M_S = 4$ and spin $S = 4$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' nonet calculation, use `C.I.=8`.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to 4 are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to, or greater than 4, can be constructed. From this set, only a nonet state can be selected, all other states will be ignored. If `ROOT=n` is present, then the n 'th nonet state will be selected, otherwise the first nonet state will be chosen.

The nonet states are the highest spin states normally calculable using MOPAC in its unmodified form. If several nonets are to be calculated, say the second or third, then `C.I. (n1, n2)` should be used.

See also `SINGLET`, `DOUBLET`, `TRIPLET`, `QUARTET`, `QUINTET`, `SEXTET`, `SEPTET`, and `OCTET`.

UHF interpretation: The system will have eight more alpha electrons than beta electrons.

NONR (W)

By default, in EF the Newton-Raphson method is used. This can be suppressed if `FORMD` in EF fails while the maximum gradient is less than a preset limit (`RCUT`). To suppress it under these conditions, specify `NONR`.

NOREOR (O)

When the symmetry of a molecule is being worked out, the molecule is orientated by default. If `NOREOR` is specified, the molecule will *not* be reorientated. The main reason to not reorientate the molecule is to allow a lower point-group to be used, and to allow the x and y axes in Abelian groups to be defined by the user.

NOSYM (W)

Sometimes a system almost has the symmetry of point group higher than C_1 . Minor deviations from exact symmetry are allowed, but these can cause difficulties. If that happens, then the job will fail. To prevent this happening, the symmetry can be re-defined as C_1 . This effectively disables all the symmetry features, so some jobs, particularly `FORCE` calculations, will take longer to run.

Use `NOSYM` if there is cause to suspect that the symmetry features are causing problems.

NOTHIEL (W)

In a normal geometry optimization using the BFGS routine, Prof. Walter Thiel's FSTMIN technique [44] is used. If normal line-searches are wanted, specify `NOTHIEL`.

NOXYZ (O)

The Cartesian coordinates are printed by default. If you do not want them to be printed, specify NOXYZ. For big jobs this reduces the output file considerably.

NSPA= n (C)

Used for the COSMO method (see EPS) to set the number of geometrical segments per molecule to n . This keyword should seldom be used, as the default value (42) should be adequate. However, for high-precision work a larger value from the set $(10n^2+2)$ or $(30n^2+2)$ should be used (e.g., 92,122, or 162).

NSURF (C)

In an ESP calculation, NSURF= n specifies the number of surface layers for the Connolly surface.

OCTET (C)

RHF interpretation: The desired spin-state is an octet: the state with component of spin $M_S = 7/2$ and spin $S = 7/2$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' octet calculation, use C.I.=7.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to $7/2$ are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to, or greater than $7/2$, can be constructed. From this set, only an octet state can be selected, all other states will be ignored. If ROOT= n is present, then the n 'th octet state will be selected: otherwise the first octet state will be chosen.

If several octets are to be calculated, say the second or third, then C.I. (n_1, n_2) should be used.

See also SINGLET, DOUBLET, TRIPLET, QUARTET, QUINTET, SEXTET, SEPTET, and NONET.

UHF interpretation: The system will have seven more alpha electrons than beta electrons.

OLDENS (W)

A density matrix produced by an earlier run of MOPAC (by using DENOUT) is to be used to start the current calculation. This can be used in attempts to obtain an SCF when a previous calculation ended successfully but a subsequent run failed to achieve an SCF.

OLDFPC (C)

By default, MOPAC uses the CODATA fundamental physical constants. This means that the results of MOPAC 6 and earlier MOPAC calculations cannot be duplicated. When OLDFPC is specified, the fundamental physical constants used in earlier MOPACs is used. This allows these earlier calculations to be duplicated using the current MOPAC.

Users are advised to not use OLDFPC except when comparison or continuity with earlier MOPACs is necessary, as all future MOPACs will use the new CODATA constants.

OLDGEO (C)

If multiple geometries are to be run, and the final geometry from one calculation is to be used to start the next calculation, OLDGEO should be specified. Example: If MNDO, AM1, and PM3 calculation, were to be done on one

Table 3.8: Use of OPEN(n, m)

Keywords	Example	Number of M.O.s	No. of Electrons
OPEN(2,2)	Twisted Ethylene	2	2
OPEN(1,2)	O ₂ ⁺	2	1
OPEN(5,3)	CH ₄ ⁺	3	5

system, for which only a rough geometry was available, then after the MNDO calculation, the AM1 calculation could be done using the optimized MNDO geometry as the starting geometry, by specifying OLDGEO. After the AM1 calculation was done, then a PM3 calculation could also be done, starting with the old AM1 geometry, by again specifying OLDGEO.

OMIN= $n.nn$ (C)

In a TS calculation, the minimum allowable overlap of the transition state eigenvectors is set by $n.nn$. If OMIN is *not* set, the default of 0.8 is used.

OPEN(n_1, n_2) (C)

Do *not* use OPEN(n_1, n_2) for ground-state systems *except* for high symmetry systems with open shells, such as twisted (D_{2d}) ethylene or molecular oxygen O₂.

For most systems, the M.O. occupancy in the SCF calculation can be either two or zero. In order for some systems to have the correct symmetry, however, some M.O.s may need a fractional occupancy in the SCF. For example, ground-state O₂ has two half occupied π M.O.s. In order to allow such systems to be defined, OPEN(n_1, n_2) should be used. Examples of OPEN(n_1, n_2) are given in Table 3.8.

The M.O. occupancy during the SCF calculation can be defined in terms of doubly occupied, empty, and fractionally occupied M.O.s. The fractionally occupied M.O.s are defined by OPEN(n_1, n_2), where n_1 = number of electrons in the open-shell manifold, and n_2 = number of open-shell M.O.s; n_1/n_2 must be in the range 0 to 2.

Notes:

- OPEN(n_1, n_2) cannot be used instead of C.I.= n_2 . OPEN(n_1, n_2) modifies the SCF calculation, but C.I.= n_2 does not modify the SCF calculation. Both keywords cause a C.I. calculation to be done.
- Be careful to ensure that n_1 is odd if the system has an odd number of electrons, and even if the system has an even number of electrons.

OPEN(1,1) will be assumed for odd-electron systems unless an OPEN keyword is used. Errors introduced by use of fractional occupancy are automatically corrected in a MECI calculation when OPEN(n_1, n_2) is used. See also C.I.= n .

OPTBACK (C)

Only to be used with MOZYME. In a protein, the optimization flags for all backbone atoms are set to ' 1 '. The optimization flags for all atoms attached to the backbone (that is, the side-chains and the hydrogens and oxygens of the -NH-CHR-CO-) are set to ' 0 '.

OPTRES (C)

Only to be used with MOZYME. In a protein, the optimization flags for all atoms except the backbone atoms are set to ' 1 '. The optimization flags for all carbon and nitrogen atoms of the backbone are set to ' 0 '.

ORT (C)

Used with TOM only, ORT causes the Reynolds-Richards molecular electrostatic potential to be used. See page 247 for more details.

P=n.nn (C)

The effect of an applied pressure on a solid can be simulated by $P=n.nn$

For polymers, the applied strain is in units of Newtons per mole. That is, for one mole of polymer chains, stacked side by side, the applied force is $n.nn$ Newtons. Suitable values for $P=n.nn$ are in the order of 10^{13} Newtons. In general, polymers are subjected to tension only, not compression, therefore the sign of $n.nn$ should normally be negative, e.g. $P=-2.D13$.

For solids, the applied pressure is in units of Newtons per square meter. Typical bulk moduli are in the order of 10 - 700 GPa (gigapascals), and typical pressures should be $n.nn = 1.D9$ to $n.nn=1.D10 \text{ NM}^{-2}$. In general, solids are subjected to compression only, not tension, therefore the sign of $n.nn$ should normally be positive, e.g. $P=2.D9$, however, for mechanically strong solids, negative pressures, even quite large pressures, can be used.

Bulk moduli, B , in Pascals, can be calculated from the change in unit cell volume at zero pressure ($\text{Vol}(0)$) to the unit cell volume under a pressure of $P=n.nn$ ($\text{Vol}(n.nn)$) from:

$$B = n.nn * \text{Vol}(0) / (\text{Vol}(0) - \text{Vol}(n.nn))$$

Useful conversion factors are:

$$1 \text{ GPa} = 10^9 \text{ N.M}^{-2} = 10^{10} \text{ dynes.cm}^{-2} = 10^{16} \text{ erg.m}^{-3}$$

$$1 \text{ GPa} = 6.0221367 / (4.184 * 10) \text{ kcal.mol}^{-1} . \text{Angstrom}^{-3} = 0.14393 \text{ kcal.mol}^{-1} . \text{Angstrom}^{-3}.$$

PARASOK (W)

Use this keyword with extreme caution! The AM1 method has been parameterized for fewer elements than those available to MNDO or PM3. If any elements which are not parameterized at the AM1 level are specified, the MNDO parameters, if available, will be used. The resulting mixture of methods, AM1 with MNDO, has not been studied to see how good the results are, and users are strictly on their own as far as accuracy and compatibility with other methods is concerned. In particular, while all parameter sets are referenced in the output, other programs may not cite the parameter sets used and thus compatibility with other MNDO programs is not guaranteed.

PDB (W)

Normally, geometries supplied in PDB format are automatically recognized. However, if there are two or more PDB geometries in the one run, then only the first PDB geometry will be recognized by default. If PDB is present, then the format of the associated geometry is defined as being PDB. PDB is only necessary when more than one PDB geometry is present, although no harm will result from having PDB present for the first PDB geometry.

PDB(*text*) (W)

Although MOZYME will normally recognize all the elements in a PDB file, the possibility exists that an element or entity will be defined in a PDB file in an unusual way, and as a result it will not be recognized. To allow for this, the keyword PDB(*text*) is provided. The *text* is composed of a series of entries of form ‘Chemical-symbol:Atomic number’, separated by commas. Thus if a PDB had some hydrogen atoms represented by the symbol *J*, and a bromine atom represented by *QW*, then these symbols would be correctly recognized by MOZYME if PDB(J:1,QW:35) were specified on the keyword line.

Entities which do not correspond to elements can be excluded by assigning them the atomic number zero, e.g. PDB(LP:0) would exclude any explicitly defined lone-pairs.

PDBOUT (O)

To be used only with proteins, PDBOUT will cause the geometry to be printed in PDB format. Although the format printed is designed to match that used in the Brookhaven Protein Data Bank, the agreement is not exact. The main differences are:

- Only the geometry is output. No preamble is generated.
- The terminus of the protein is not identified.
- Every atom in a residue is assumed to belong to the residue. If a group is attached to a residue, it is considered to be part of the residue.
- The numbering system(s) used in the files deposited in the PDB are not consistent. Therefore the convention defined by the PDB itself is used.

PECI (C)

In a C.I., the microstates used are: (a) the ground state; (b) all one-electron excitations; and (c) all closed shell (paired) two-electron excitations. PECI = Paired Electron Configuration Interaction. Because of the definition of PECI, only Singlet and Triplet states can be generated.

PI (O)

The normal density matrix is composed of atomic orbitals, that is s , p_x , p_y and p_z . PI allows the user to see how each atom-atom interaction is split into σ , π , and δ bonds. The resulting “density matrix” is composed of the following basis-functions: $s - \sigma$, $p - \sigma$, $p - \pi$, $d - \sigma$, $d - \pi$, $d - \delta$. The on-diagonal terms give the hybridization state, so that an sp^2 hybridized system would be represented as $s - \sigma: 1.0, p - \sigma: 2.0, p - \pi: 1.0$.

PINOUT (O)

Not a very useful keyword, PINOUT will cause the LMOs to be printed whenever they are written to <filename>.den.

PL (O)

The largest change in any density matrix element on two successive iterations in the SCF calculation is printed when PL is specified. This is a useful tool in monitoring an SCF, particularly when it fails to converge.

Table 3.9: Data Required by PMEPR

Name of Datum	Allowed Values
ICASE	1, 2, 3
N1	1-NUMAT
N2	1-NUMAT, not N1
N3	1-NUMAT, not N1 or N2
X0	Any real number

PM3 (C)

The PM3 method [4] is to be used.

PMEP (C)

The Parametric Molecular Electrostatic Potential of Wang and Ford [45, 46] is generated. This method is very accurate, but has only been parameterized for H, C, N, O, F, P, S, Cl, and Br, and only for the AM1 method. By default, the plane used is the X-Y plane at Z=0. Use PMEPR for other planes. Other keywords, e.g., MINMEP (to print the minima) or PRTMEP (to write data for `esp1ot` to use) *must* be present.

PMEPR (C)

The Parametric Molecular Electrostatic Potential of Wang and Ford [45, 46] is generated. When PMEPR is used, extra data must follow the Z-matrix and symmetry data. This extra data defines the plane in which the electrostatic potential is plotted, and is summarized in Table 3.9. Other keywords, e.g., MINMEP (to print the minima) or PRTMEP (to write data for `esp1ot` to use) *must* be present.

The various planes which can be drawn through the system are identified by the datum ICASE:

ICASE=1 If X0 is zero, then the plane generated passes through atoms N1, N2 and N3. Obviously, these atoms should not lie on a straight line. If X0 is non-zero, then the plane is parallel to the plane of atoms N1, N2 and N3, but X0 Ångstroms above it.

ICASE=2 If X0 is zero, then the plane generated is perpendicular to the plane of N1, N2 and N3, and includes the line from atom N1 to N2. If X0 is non-zero, then the plane is perpendicular to the plane of N1, N2 and N3, and is X0 Ångstroms above, and parallel to, the line from atom N1 to N2.

ICASE=3 The plane used is that which bisects the angle N2-N1-N3. X0 is meaningless in this context, and should be set to zero.

POINT1=*n* (C)

In a grid calculation, the number of points to be calculated in the first direction is given by POINT1=*n*. '*n*' should be less than 24; default: 11.

POINT2=*n* (C)

In a grid calculation, the number of points to be calculated in the second direction is given by POINT2=*n*. '*n*' should be less than 24, default: 11;

POINT=*n* (C)

The number of points to be calculated on a reaction path is specified by POINT=*n*. Used only with STEP in a path calculation.

POLAR (C)

This calculates the polarizability and first and second hyperpolarizabilities [47] This routine has been completely re-written by Prof Henry Kurtz and Prakashan Korambath, of Memphis State University.

The POLAR calculation now gives the frequency-dependent NLO properties (α , β , and γ) at user-defined frequencies.

Specification of Parameters in POLAR.

By default, POLAR will calculate the hyperpolarizabilities at 0.0, 0.25, and 0.50 eV. Under user control, up to 10 energies can be specified. All user-specified data for POLAR is supplied after POLAR and inside brackets, thus POLAR(user-supplied-data). The form in which the energies of the radiation to be used is specified as POLAR(E=(n_1, n_2, n_3, \dots)), where n_1 up to n_{10} are the energies in eV. For example, if the radiation energies were 0, 1, and 2.1 eV, then POLAR(E=(0, 1, 2.1)) would be used. Other quantities that can be calculated by POLAR are second and third harmonic generation, the Electrooptic Pockel's Effect, and Electric Field Second Harmonic Generation.

There are several other parameters in POLAR which can be adjusted, see Section 5.21. To prevent re-orientation, use LET.

At present this calculation only works for RHF closed shell systems, and it does not work for polymers, or other extended solids.

The POLAR calculation cannot be run with the MOZYME function; instead the static polarizability calculation is done. This produces the polarizability (α) only. In other words, when the MOZYME function is used, POLAR and STATIC have the same meaning.

In the polarizability calculation the system is orientated along the principal moments of inertia, by default. This means that a constant system of orientation is used. Sometimes this produces problems, for example, a large change in orientation might make the SCF take longer or even fail (it starts with the old density matrix, which is not rotated). To prevent this, and to save some time, LET can be used.

Regardless of the coordinate system used in a POLAR calculation, the geometry that exists after the calculation will be in internal coordinates.

The various methods in MOPAC all underestimate the polarizability (ALPHA) by about 50%. However, one method, MNDO, has a monatomic correction for the elements H, C, N, O, F, Cl, Br, and I. This correction reduces the average error to only a few percent. Therefore, when polarizability is wanted, and the compound contains only the elements listed here, then the MNDO method should be used.

POPS (O)

When POPS is specified, the self-consistent-field atomic orbital electron density populations are printed.

POTWRT (W)

In an ESP calculation, write out surface points and electrostatic potential values to <filename>.esp.

POWSQ (C)

Details of the working of POWSQ are printed out. This is useful only in debugging.

PRECISE (W)

The criteria for terminating all optimizations, electronic and geometric, are to be increased by a factor, normally 100. This can be used where more precise results are wanted. If the results are going to be used in a FORCE calculation, where the geometry needs to be known quite precisely, then PRECISE is recommended; for small systems the extra cost in CPU time is minimal. PRECISE is not recommended for experienced users; instead, GNORM=*n.nn* and SCFCRT=*n.nn* are suggested. PRECISE should only rarely be necessary in a FORCE calculation: all it does is remove quartic contamination, which only affects the trivial modes significantly, and is very expensive in CPU time.

PRTMEP (O)

Print grid of electrostatic points calculated by PMP.

PULAY (W)

The default converger in the SCF calculation is to be replaced by Pulay's procedure [48] as soon as the density matrix is sufficiently stable. A considerable improvement in speed can frequently be achieved by the use of PULAY, particularly for excited states. If a large number of SCF calculations are envisioned, a sample calculation using 1SCF and PULAY should be compared with using 1SCF on its own, and if a saving in time results, then PULAY should be used in the full calculation. PULAY should be used with care in that its use will prevent the combined package of convergers (SHIFT, PULAY and the CAMP-KING convergers) from being used automatically in the event that the system fails to go SCF in (ITRY-10) iterations.

The combined set of convergers very seldom fails.

QPMEP (C)

Calculate and print the atomic point charges using the Ford-Wang Parametric Electrostatic Potential Calculation. By default, the Connolly surface is used. If the Williams surface is wanted, add WILLIAMS.

QUARTET (C)

RHF interpretation: the desired spin-state is a quartet, i.e., the state with component of spin $M_S = 3/2$ and spin $S = 3/2$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' quartet calculation, use C.I.=3. If the quartet state consists of three half-filled degenerate M.O.s, then OPEN(3,3) should be used.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to $3/2$ are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to or greater than $3/2$ can be constructed. From this set, only a quartet state can be selected; all other states will be ignored. If ROOT=*n* is present, then the *n*'th quartet state will be selected; otherwise, the first quartet state will be chosen.

See also SINGLET, DOUBLET, TRIPLET, QUINTET, SEXTET, SEPTET, OCTET, and NONET.

UHF interpretation: The system will have three more alpha electrons than beta electrons.

QUINTET (C)

RHF interpretation: The desired spin-state is a quintet: that is, the state with component of spin $M_S = 2$ and spin $S = 2$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' quintet calculation, use `C.I.=4`.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to 2 are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to or greater than 2 can be constructed. From this set, only a quintet state can be selected; all other states will be ignored. If `ROOT=n` is present, then the n 'th quintet state will be selected; otherwise, the first quintet state will be chosen.

See also `SINGLET`, `DOUBLET`, `TRIPLET`, `QUARTET`, `SEXTET`, `SEPTET`, `OCTET`, and `NONET`.

UHF interpretation: The system will have four more alpha electrons than beta electrons.

RAPID (W)

When only part of a geometry is being modified, the speed of the calculation can be increased by the use of `RAPID`. Thus, if a side-chain on a protein is being optimized, all atoms on the side-chain would be flagged with " 1"s, and all other atoms would be flagged with " 0"s. An atom is considered as being modified if any one of the three flags is " 1".

In order for `RAPID` to work, all atoms that move must be flagged for optimization. Thus, if a single backbone atom, near the middle of a protein chain, is flagged for optimization, and no other atoms are flagged for optimization, `RAPID` should not be used, because many atoms that are not marked for optimization will move as a result of the flagged atom moving.

One way to ensure that all atoms that move are flagged is to put all the flagged atoms at the end of the data-set. This is not essential, but it does prevent unflagged atoms from moving. To do this in Cartesian coordinates is simple. In internal coordinates, flag the desired atoms, convert to Cartesian coordinates, move the flagged atoms to the end of the data set, then convert back to internal coordinates.

The effort to set up a system with `RAPID` is well worth while, particularly if only a few (up to 10–20%) of the atoms are flagged.

RE-LOCAL (W)

The LMOs used by `MOZYME` in the SCF are not fully localized. At the start they are, but as the SCF proceeds, they become increasingly delocalized. At the end of a calculation, the LMOs may be 5 or 10 percent less than fully localized.

To generate localized M.O.s of the type that `MOPAC` makes, `RE-LOCAL` should be used. Note, however, that `RE-LOCAL` involves a relatively lengthy process, and the effect of `RE-LOCAL` on the results is usually quite small, therefore avoid using `RE-LOCAL` unless high-quality results are needed.

RECALC=n (W)

`RECALC=n` calculates the Hessian every n steps in the EF optimization. For small n this is costly but is also very effective in terms of convergence. `RECALC=10` and `DMAX=0.10` can be useful for difficult cases. In extreme cases `RECALC=1` and `DMAX=0.05` will always find a stationary point, if it exists.

RELSCF (W)

When `RELSCF=n` is present, the default SCF criterion is multiplied by *n*. This is useful if the value of the default SCF criterion is not readily available. Examples: `RELSCF=10` will make the SCF test easier to pass—the criterion will be made 10 times easier. Similarly, if the results are not precise enough, then `RELSCF=0.1` would increase the precision 10 times. However, if the precision is increased too much, the SCF test might never be passed. See also `SCFCRT`.

RELTHR (W)

When `RELTHR=n` is present in a `MOZYME` calculation, the default value for deciding whether an atom should be present in a LMO is multiplied by *n*. This is useful if the value of the default value is not readily available. Examples: `RELTHR=10` will tend to reduce the number of atoms in a LMO. This will reduce the storage requirement and make the iterations run faster, but might make the calculation too imprecise, preventing the SCF test from being passed, and causing the run to fail.

Similarly, if the results are not precise enough, then `RELTHR=0.1` would tend to increase the number of atoms in a LMO. This will make the calculation become more like a `MOPAC` calculation, but the time required will increase. For rough work—for example, the start of a geometry optimization—then `RELTHR=10` could be used. For high-precision work, either increase the precision by using `RELSCF=0.1` or, preferably, do not use `MOZYME`. See also `THRESH`.

REORTH (W)

During a `MOZYME` run, the steady accumulation of small errors (for example, roundoff, and more important, the elimination of atoms from LMOs) will result in the LMOs becoming non-orthogonal. These errors can be eliminated almost completely by re-orthogonalizing the LMOs. If `REORTH` is present, then the LMOs will be re-orthogonalized at the end of every 10th SCF calculation, starting with the first SCF.

A useful strategy would be to do geometry optimizations without using `REORTH`, and allow the LMOs to slowly become more and more non-orthogonal. Then, either at or near the end of the optimization, do a `1SCF` with `REORTH` to eliminate the accumulated errors.

The effect of `REORTH` is likely to be very small. If the value of `THRESH` is increased, however, then the LMOs will become non-orthogonal faster, and `REORTH` would then likely have a significant effect.

RESIDUES (O)

`RESIDUES` will put a label on each atom in a protein. The label is a short description of the atom, and is useful in identifying the atoms in a protein. The general form of the label for an atom is '*nnnnnn TYPfrrr*' where *nnnnnn* is the atom number (the 10th atom would have the atom number 10), *TYP* is a three letter symbol for each amino acid (the allowed symbols are shown in Table 3.10). If a residue is not recognized, then the symbol 'UNK' (unknown) will be used. Backbone atoms are indicated by an asterisk (*) after the residue symbol, i.e., 'f=*'; for all other atoms, 'f' is a space; *rrr* is the residue number. A modified residue can still be recognized if `ZENO=text` is used. The residue nearest to the NH_2 end of the protein is No. 1, the next is 2, and so on.

Small molecules are often found associated with proteins. To allow for this, some special groups are also recognized. A phosphate group is indicated by the symbol 'PO4'. In place of the residue number, the number of the phosphate will be given, the first phosphate being 1. More special groups will be added as needed.

Hydrogens added to these small molecules are given the symbol of the chemical element to which they are attached.

Table 3.10: Abbreviations for the 20 Amino Acids

Amino Acid	Formula of Residue [†]	Three-Letter Abbreviation	One-Letter Abbreviation
Glycine	C ₂ NOH ₃	GLY	G
Alanine	C ₃ NOH ₅	ALA	A
Valine	C ₅ NOH ₉	VAL	V
Leucine	C ₆ NOH ₁₁	LEU	L
Isoleucine	C ₆ NOH ₁₁	ILE	I
Serine	C ₃ NO ₂ H ₅	SER	S
Threonine	C ₄ NO ₂ H ₇	THR	T
Aspartic acid	C ₄ NO ₃ H ₅ (4)	ASP	D
Asparagine	C ₄ N ₂ O ₂ H ₆	ASN	N
Lysine	C ₆ N ₂ OH ₁₂ (13)	LYS	K
Glutamic acid	C ₅ NO ₃ H ₇ (6)	GLU	E
Glutamine	C ₅ N ₂ O ₂ H ₈	GLN	Q
Arginine	C ₆ N ₄ OH ₁₂ (13)	ARG	R
Histidine	C ₆ N ₃ OH ₇ (8)	HIS	H
Phenylalanine	C ₉ NOH ₉	PHE	F
Cysteine	C ₃ NOSH ₅ (4)	CYS	C
Tryptophan	C ₁₁ N ₂ OH ₁₀	TRP	W
Tyrosine	C ₉ NO ₂ H ₉ (8)	TYR	Y
Methionine	C ₅ NOSH ₉	MET	M
Proline	C ₅ NOH ₇	PRO	P

†: The number of hydrogen atoms in the ionized residue is given in parenthesis after the formula. Cysteine may exist in the neutral, ionized or reduced form.

If RESEQ is used, then RESIDUES is automatically run.

RESEQ (O)

According to the PDB format, atoms in a protein are specified starting with the nitrogen of the -NH₂ end and ending with the atoms at the -COOH end. RESEQ will re-arrange the atoms into the standard PDB sequence. RESEQ is useful particularly after adding hydrogen atoms. Because RESEQ changes the order of the atoms, further work is not possible, and the calculation is stopped, after the new geometry is printed.

RESTART (W)

When a job has been stopped, for whatever reason, and intermediate results have been stored, then the calculation can be restarted at the point where it stopped by specifying RESTART. The most common cause of a job stopping before completion is its exceeding the time allocated. A saddle-point calculation has no restart, but the output file contains information which can easily be used to start the calculation from a point near to where it stopped.

It is not necessary to change the geometric data to reflect the new geometry, although this can be done, if desired. A convenient way to monitor a long run is to specify 1SCF and RESTART; this will give a normal output file at very little cost.

Several functions in MOPAC are not allowed under MOZYME, examples being the FORCE calculation and the frequency-dependent hyperpolarizability. RESTART can be used to allow a calculation to start in MOPAC or MOZYME, and then change to the other if desired.

Note: Two restarts exist in the IRC calculation. If an IRC calculation stops while in the FORCE calculation, then a normal restart can be done. If the job stops while doing the IRC calculation itself then the keyword `IRC=n` should be changed to `IRC`, or it can be omitted if `DRC` is also specified. The absence of the string "IRC=" is used to indicate that the FORCE calculation was completed before the restart files were written.

RMAX=*n.nn* (C)

In a TS calculation, the calculated/predicted energy change must be less than *n.nn*. If RMAX is not set, the default of 4.0 is used.

RMIN=*n.nn* (C)

In a TS calculation, the calculated/predicted energy change must be more than *n.nn*. If RMIN is not set, the default of 0.0 is used.

ROOT=*n* (C)

The *n*'th root of a C.I. calculation is to be used in the calculation. If a keyword specifying the spin-state is also present, e.g. `SINGLET` or `TRIPLET`, then the *n*'th root of that state will be selected. Thus `ROOT=3` and `SINGLET` will select the third singlet root. If `ROOT=3` is used on its own, then the third root will be used, which may be a triplet, the third singlet, or the second singlet (the second root might be a triplet).

See also `C.I.=n`.

RSCAL (C)

In EF, the P-RFO step will be set to the trust radius if RSCAL is present. Do *not* use this keyword unless you have to.

RSOLV=*n.nn* (C)

Used by the COSMO method (see `EPS`) to set the effective radius of the solvent molecule to *n.nn*. The default value (1.0) is appropriate for water.

This quantity should be altered for each solvent. One way to get RSOLV is to try several values for RSOLV, and see which one gives the best solvation energies when compared with experimental results.

SADDLE (C)

The transition state in a simple chemical reaction is to be optimized. Extra data are required. After the first geometry, specifying the reactants, the second geometry, specifying the products, is defined, using the same format as that of the first geometry.

A SADDLE calculation works entirely in Cartesian coordinates, regardless of the coordinate system used in the data set. The output, however, is in internal coordinates. All coordinates are optimized, so `SYMMETRY` should not be used. See also `BAR=n.nn`.

SAFE (W)

If MOPAC has been compiled so that the memory demand is minimized, then some SCF convergers will not be available. For well-behaved systems, this is not a problem. However, if there is difficulty in generating an SCF, then these special SCF convergers may be needed. To allow them to be used, specify SAFE.

SCALE (C)

SCALE=*n.n* specifies the scaling factor for van der Waals' radii for the initial layer of the Connolly surface in the ESP calculation.

SCFCRT=*n.nnn* (W)

The default SCF criterion is to be replaced by that defined by SCFCRT=*n.nnn*.

The SCF criterion is the change in energy in kcal/mol on two successive iterations. Other minor criteria may make the requirements for an SCF slightly more stringent. The SCF criterion can be varied from about 1.0 to 1.D-25, although numbers in the range 0.1 to 1.D-9 will suffice for most applications.

An overly tight criterion can lead to failure to achieve an SCF, and the consequent failure of the run.

SCINCR=*n.nn* (W)

In an ESP calculation, SCINCR=*n.nn* specifies the increment between layers of the surface in the Connolly surface. (default: 0.20)

SEPTET (C)

RHF interpretation: The desired spin-state is a septet: the state with component of spin $M_S = 3$ and spin $S = 3$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' septet calculation, use C.I.=6.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to 3 are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to or greater than 3 can be constructed. From this set, only a septet state can be selected; all other states will be ignored. If ROOT=*n* is present, then the *n*'th septet state will be selected; otherwise, the first septet state will be chosen.

The septet states are the highest spin states normally calculable using MOPAC in its unmodified form. If several septets are to be calculated, say the second or third, then C.I. (*n₁, n₂*) should be used.

See also SINGLET, DOUBLET, TRIPLET, QUARTET, QUINTET, SEXTET, OCTET, and NONET.

UHF interpretation: The system will have six more alpha electrons than beta electrons.

SETUP (C)

If, on the keyword line, the word 'SETUP' is specified, then one or two lines of keywords will be read from a file with the logical name SETUP. The logical file SETUP must exist, and must contain at least one line. If the second line is defined by the first line as a keyword line, and the second line contains the word SETUP, then one line of keywords will be read from a file with the logical name SETUP.

SETUP=name (C)

Same as SETUP, only the logical or actual name of the SETUP file is 'name'.

SEXTET (C)

RHF interpretation: The desired spin-state is a sextet: the state with component of spin $M_S = 5/2$ and spin $S = 5/2$. In order to define this type of calculation other keywords *must* also be used. For a 'simple' sextet calculation, use C.I.=5.

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to 5/2 are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to or greater than 5/2 can be constructed. From this set, only a sextet state can be selected; all other states will be ignored. If ROOT= n is present, then the n 'th sextet state will be selected; otherwise, the first sextet state will be chosen.

If several sextets are to be calculated, say the second or third, then C.I. (n_1, n_2) should be used.

See also SINGLET, DOUBLET, TRIPLET, QUARTET, QUINTET, SEPTET, OCTET, and NONET.

UHF interpretation: The system will have five more alpha electrons than beta electrons.

SHIFT= $n.nn$ (W)

In an attempt to obtain an SCF by damping oscillations which slow down the convergence or prevent an SCF being achieved, the virtual M.O. energy levels are shifted up or down in energy by a shift technique [49]. The principle is that, if the virtual M.O.s are changed in energy relative to the occupied set, then the polarizability of the occupied M.O.s will change *pro rata*. Normally, oscillations are due to autoregenerative charge fluctuations.

The SHIFT method has been re-written so that the value of SHIFT changes automatically to give a critically-damped system. This can result in a positive or negative shift of the virtual M.O. energy levels. If a non-zero SHIFT is specified, it will be used to start the SHIFT technique, rather than the default 15eV. If SHIFT=0 is specified, the SHIFT technique will not be used unless normal convergence techniques fail and the automatic "ALL CONVERGERS . . ." message is produced.

SIGMA (C)

In cases where TS is *not* wanted, the McIver-Komornicki gradient norm minimization [9, 10] routines, POWSQ and SEARCH, can to be used by including SIGMA. These are very rapid routines, but do not work for all species. If the gradient norm is low, i.e., less than about 5 units, then SIGMA will probably work; in most cases, NLLSQ is recommended. SIGMA first calculates a Hessian matrix, a slow step, then works out the direction of fastest descent, and searches along that direction until the gradient norm is minimized. The Hessian is then partially updated in light of the new gradients, and a fresh search direction found. Clearly, if the Hessian changes markedly as a result of the line-search, the update done will be inaccurate, and the new search direction will be faulty.

SIGMA should be avoided if at all possible when non-variationally optimized calculations are being done.

If the Hessian is suspected to be corrupt within SIGMA, it will be automatically recalculated. This frequently speeds up the rate at which the transition state is located. If you do not want the Hessian to be reinitialized—it is costly in CPU time—specify LET on the keyword line.

SINGLET (C)

When a configuration interaction calculation is done, all spin states are calculated simultaneously, either for component of spin = 0 (for even electron systems) or 1/2 (for odd electron systems). When only singlet states are of

```

REAL: 1923, Left: 7132352 Used: 1924 3 CSETUP: GEO
REAL: 1282, Left: 7131069 Used: 3207 34 CSETUP: TXTATM
REAL: 1923, Left: 7129145 Used: 5131 7 CSETUP: COORD
REAL: 1923, Left: 7127221 Used: 7055 5 CSETUP: XPARAM
REAL: 1622, Left: 7125598 Used: 8678 11 CSETUP: USPD
REAL: 1622, Left: 7123975 Used: 10301 12 CSETUP: PSPD
REAL: 641, Left: 7123333 Used: 10943 51 CSETUP: ATMASS
REAL: 1923, Left: 7121409 Used: 12867 52 CSETUP: REACT
REAL: 1923, Left: 7119485 Used: 14791 65 CSETUP: PROFIL
REAL: 641, Left: 7118843 Used: 15433 73 CSETUP: CH
.
.
.
INTEGER: 641, Left: 176111 Used: 778680 97 RMOPAC: IOPT
INTEGER: 641, Left: 176753 Used: 778038 0 RMOPAC: IOPT*

```

Figure 3.12: Example of Output Generated by SIZES

interest, then SINGLET can be specified, and all other spin states, while calculated, are ignored in the choice of root to be used.

Note that while almost every even-electron system will have a singlet ground state, SINGLET should still be specified if the desired state *must* be a singlet.

See also DOUBLET, TRIPLET, QUARTET, QUINTET, SEXTET, SEPTET, OCTET, and NONET.

SINGLET has no meaning in an UHF calculation, but see also TRIPLET.

SIZES (O)

When SIZES is specified, details of memory managements, including the amount of dynamic memory reserved, the sizes of all arrays created, both permanent and temporary, are printed whenever a new array is created.

At the start of a calculation, the amount of dynamic memory necessary for the arrays is printed. If an error message indicating that there is insufficient dynamic memory to create a new array is printed, then SIZES should be run in order to identify which arrays have been allowed for.

If there is insufficient memory available on the computer, then SIZES can be run in an attempt to reduce the memory needed. For example, if the only purpose of a run is to convert from internal to Cartesian coordinates (using XYZ), then a run using SIZES might indicate that storage is being reserved for the arrays in the SCF calculation. These arrays can be avoided by use of OSCF.

After the dynamic memory is reserved, most or all of the permanent arrays are created. Each of these is characterized by a unique number in the range 3 to 96. Only those arrays necessary for the current run will be created. Each time an array is created, the amount of available memory is decreased by the size of the array plus 1 (the extra array element is needed for an array bounds check).

Finally, temporary arrays are created and destroyed as needed.

A few examples of the output generated by SIZES are given in Figure 3.12 , p 74 .

The numbers on each line representing an array being created have the following meaning:

1. The size of the array (for GEO, this is 1923).
2. The storage left after the array has been created (7132352).
3. The storage used up to and including the current array (1924).
4. The address in NCORE of the start of the array. If the array is temporary, then the address 97 is used, otherwise the address is unique.

The numbers on each line representing a temporary array being destroyed have the following meaning:

1. The size of the array (for IOPT, this is 641).
2. The storage left after the array has been destroyed (176753).
3. The storage used up to the present (778038).
4. The number 0. This indicates that an array is being destroyed.

The name of an array that is being destroyed is followed by an asterisk, thus: IOPT*.

Every array created uses one element more than the size of the array, so that, for example the array COORD requires 1923 elements, therefore the storage used increases by 1924 elements, from 3207 to 5131.

SLOG=*n.nn* (W)

The line-search mechanism in the BFGS geometry optimization is replaced with a constant step size. This prevents large steps that could waste time or even cause a failure to achieve an SCF. If SLOG is used, the step is 0.25 of the default step, otherwise if SLOG=*n.nn* is used, the step is *n.nn*.

SLOPE (C)

In an ESP calculation, SLOPE=*n.nn* specifies the scale factor for MNDO charges. (default=1.422)

SNAP (W)

Many symmetry-defined angles, such as the tetrahedral angle 109.471221 ... are difficult to type. If SNAP is used, then any angle near to a symmetry-defined angle will be adjusted to that angle. Thus 109.471 would become arcos(-1/3) accurate to 15 or 16 figures. The tolerance of SNAP is quite strict, so 109.47 would NOT be recognized.

SPIN (O)

The spin matrix, defined as the difference between the alpha and beta density matrices, is to be printed. If the system has a closed-shell ground state, e.g. methane run UHF, the spin matrix will be null.

If SPIN is not requested in a UHF calculation, then the diagonal of the spin matrix, that is the spin density on the atomic orbitals, will be printed.

STATIC (C)

The static polarizability is calculated. An electric field gradient is applied to the system, and the response is calculated. The dipole and polarizability are calculated two different ways, from the change in ΔH_f and from the change in dipole. A measure of the imprecision of the calculation can be obtained by comparing the two quantities. Greater precision can be obtained by increasing the precision of the SCF (RELSCF=0.1 or RELSCF=0.01). For large systems, the MOZYME function should be used, but the polarizability calculated using the MOZYME function is not as precise as that calculated using conventional methods. If the results are not precise enough, increase the SCF precision (preferably by using RELSCF=*n.nn*).

In STATIC, 37 or 38 SCF calculations are involved. The polarizability involves 36 calculations, one for each electric field. The fields are +X, -X, +2X, -2X, +Y, -Y, +2Y, -2Y, +X+Y, -X+Y, -X-Y, +X-Y, +2X+2Y, -2X+2Y, -2X-2Y, +2X-2Y, +Z, -Z, +2Z, -2Z, +X+Z, -X+Z, -X-Z, +X-Z, +2X+2Z, -2X+2Z, -2X-2Z, +2X-2Z, +Y+Z, -Y+Z, -Y-Z, +Y-Z,

+2Y+2Z, -2Y+2Z, -2Y-2Z, and +2Y-2Z. These are used in the construction of a three by three secular matrix, which is then diagonalized, giving the orthogonal polarizabilities.

STEP (C)

In a reaction path, if the path step is constant, STEP can be used instead of explicitly specifying each point. The number of steps is given by POINT.

STEP1=*n.nnn* (C)

In a grid calculation the step size in degrees or Ångstroms for the first of the two parameters is given by *n.nnn*. By default, an 11 by 11 grid is generated. See POINT1 and POINT2 on how to adjust this number. The first point calculated is the supplied geometry, and is in the upper left hand corner. See also STEP2=*n.nnn*.

STEP2=*n.nnn* (C)

In a grid calculation the step size in degrees or Ångstroms for the second of the two parameters is given by *n.nnn*.

ST03G (W)

In an ESP calculation ST03G means "Use the STO-3G basis set to de-orthogonalize the semiempirical orbitals".

SUPER (O)

The nucleophilic and electrophilic delocalizabilities, the charge densities and bond orders are calculated and printed. These include the Mulliken electronegativity, the Parr and Pople absolute hardness, and the Schüürmann Molecular Orbital shift alpha quantities.

The superdelocalizabilities are calculated according to the method described in Schüürmann, G. *Env. Tox. Chem.* (9), 417 (1990), and Schüürmann, G. *Quant. Struct.-Act. Relat.* (9), 326 (1990).

SYMAVG (W)

Used by the PMEP, SYMAVG will average charges which should have the same value by symmetry. Thus, to generate symmetry-equivalent charges, use QPMEP and SYMAVG.

SYMMETRY (C)

Symmetry data defining related bond lengths, angles and dihedrals, or *x*, *y*, and *z* coordinates, can be included by supplying additional data after the geometry has been entered. If there are any other data, such as values for the reaction coordinates, or a second geometry, as required by SADDLE, then it would follow the symmetry data. Symmetry data are terminated by one blank line. For non-variationally optimized systems symmetry constraints can save a lot of time because many derivatives do not need to be calculated. At the same time, there is a risk that the geometry may be wrongly specified, e.g. if methane radical cation is defined as being tetrahedral, no indication that this is faulty will be given until a FORCE calculation is run. (This system undergoes spontaneous Jahn-Teller distortion.)

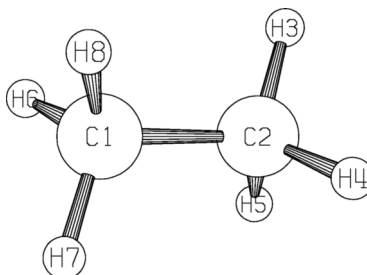
Usually a lower heat of formation can be obtained when SYMMETRY is specified. To see why, consider the geometry of benzene. If no assumptions are made regarding the geometry, then all the C–C bond lengths will be very slightly different, and the angles will be almost, but not quite 120 degrees. Fixing all angles at 120

Table 3.11: Internal Coordinate Symmetry Functions

- 1 Bond length is set equal to the reference bond length
- 2 Bond angle is set equal to the reference bond angle
- 3 Dihedral angle is set equal to the reference dihedral angle
- 4 Dihedral angle varies as $90^\circ - \text{reference dihedral}$
- 5 Dihedral angle varies as $90^\circ + \text{reference dihedral}$
- 6 Dihedral angle varies as $120^\circ - \text{reference dihedral}$
- 7 Dihedral angle varies as $120^\circ + \text{reference dihedral}$
- 8 Dihedral angle varies as $180^\circ - \text{reference dihedral}$
- 9 Dihedral angle varies as $180^\circ + \text{reference dihedral}$
- 10 Dihedral angle varies as $240^\circ - \text{reference dihedral}$
- 11 Dihedral angle varies as $240^\circ + \text{reference dihedral}$
- 12 Dihedral angle varies as $270^\circ - \text{reference dihedral}$
- 13 Dihedral angle varies as $270^\circ + \text{reference dihedral}$
- 14 Dihedral angle varies as the negative of the reference angle
- 15 Bond length varies as half the reference bond length
- 16 Bond angle varies as half the reference bond angle
- 17 Bond angle varies as 180 degrees – reference bond angle
- 18 (not used)
- 19 Bond length is a multiple of reference bond-length

Table 3.12: Cartesian Coordinate Symmetry Functions

- 1 X coordinate is set equal to the reference X coordinate
- 2 Y coordinate is set equal to the reference Y coordinate
- 3 Z coordinate is set equal to the reference Z coordinate
- 4 X coordinate is set equal to $-$ the reference X coordinate
- 5 Y coordinate is set equal to $-$ the reference Y coordinate
- 6 Z coordinate is set equal to $-$ the reference Z coordinate
- 7 X coordinate is set equal to the reference Y coordinate
- 8 Y coordinate is set equal to the reference Z coordinate
- 9 Z coordinate is set equal to the reference X coordinate
- 10 X coordinate is set equal to $-$ the reference Y coordinate
- 11 Y coordinate is set equal to $-$ the reference Z coordinate
- 12 Z coordinate is set equal to $-$ the reference X coordinate
- 13 X coordinate is set equal to the reference Z coordinate
- 14 Y coordinate is set equal to the reference X coordinate
- 15 Z coordinate is set equal to the reference Y coordinate
- 16 X coordinate is set equal to $-$ the reference Z coordinate
- 17 Y coordinate is set equal to $-$ the reference X coordinate
- 18 Z coordinate is set equal to $-$ the reference Y coordinate



```

SYMMETRY
ETHANE, D3D                NA NB NC

C
C   1.53 1                    1
H   1.10 1  110 1            2  1
H   1.10 0  110 0  120 0    2  1  3
H   1.10 0  110 0  240 0    2  1  3
H   1.10 0  110 0   60 0    1  2  3
H   1.10 0  110 0  180 0    1  2  3
H   1.10 0  110 0  300 0    1  2  3
O   0.00 0   0 0   0 0   0 0 0
3,   1,   4,   5,   6,   7,   8,
3,   2,   4,   5,   6,   7,   8,

```

Figure 3.13: Ethane, showing use of SYMMETRY

degrees, dihedrals at 180 or 0 degrees, and only optimizing one C–C and one C–H bond-length will result in a 2–D optimization, and exact D_{6h} symmetry. Any deformation from this symmetry must involve error, so by imposing symmetry some error is removed.

The layout of the symmetry data is:

```
<defining atom><symmetry relation><defined atom><defined atom>, ...
```

where the numerical codes for <symmetry relation> are given in the tables of symmetry functions, (Tables 3.11 and 3.12).

For function 19 in internal coordinates, the format of the symmetry data is

```
<defining atom> 19 <multiplying factor><defined atom><defined atom>, ...
```

For example, ethane, with three independent internal coordinate variables, can be defined as shown in Figure 3.13. Here atom 3, a hydrogen, is used to define the bond lengths (symmetry relation 1) of atoms 4,5,6,7 and 8 with the atoms they are specified to bond with in the NA column of the data file; similarly, its angle (symmetry relation 2) is used to define the bond-angle of atoms 4,5,6,7 and 8 with the two atoms specified in the NA and NB columns of the data file. The other angles are point-group symmetry defined as a multiple of 60 degrees.

Spaces, tabs or commas can be used to separate data. Note that only three parameters are marked to be optimized. The symmetry data can be the last line of the data file unless more data follows, in which case a blank line must be inserted after the symmetry data.

Internal coordinate symmetry function 19 (see Table 3.11) is intended for use in polymers, in which the translation vector may be a multiple of some bond-length. 1,2,3 and 14 are most commonly used. Abbreviation: SYM.

Methane, specified using Cartesian coordinates, can be described with one unknown, the C–H bond length, as shown in Figure 3.14.

```

XYZ SYMMETRY
Methane, Td

C  0.0000000 0   0.0000000 0   0.0000000 0
H  0.6375302 1   0.6375302 0   0.6375302 0
H  0.6375302 0  -0.6375302 0  -0.6375302 0
H -0.6375302 0   0.6375302 0  -0.6375302 0
H -0.6375302 0  -0.6375302 0   0.6375302 0

  2  1  3
  2  4  4  5
  2  9  2  5
  2 12  3  4
  2 14  2  4
  2 17  3  5

```

Figure 3.14: Example of Cartesian SYMMETRY functions

T-PRIORITY (O)

In a DRC calculation, results will be printed whenever the calculated time changes by 0.1 femtoseconds. Abbreviation, T-PRIO. See Section 5.19 for more details.

T-PRIORITY=*n.nn* (O)

In a DRC calculation, results will be printed whenever the calculated time changes by *n.nn* femtoseconds.

T=*n* [M,H,D] (W)

This is a facility to allow the program to shut down in an orderly manner on computers with execution time cpu limits.

The total cpu time allowed for the current job is limited to *nn.nn* seconds; by default this is one hour, i.e., 3600 seconds. If the next cycle of the calculation cannot be completed without running a risk of exceeding the assigned time the calculation will write a restart file and then stop. The safety margin is 100 percent; that is, to do another cycle, enough time to do at least two full cycles must remain.

If several systems are run in one job, then the time allowed for *all* the jobs is one hour, unless T=*n.nn* is used. If T=*n.nn* is used, then the allowed time *for the remainder of the job* is *n.nn* seconds. This means that if the allowed time for each system is to be an hour, then the keyword T=1h must be specified in each system.

Alternative specifications of the time are T=*nn.nn*M, which defines the time in minutes, T=*nn.nn*H, in hours, and T=*nn.nn*D, in days, for very long jobs.

THERMO (O)

The thermodynamic quantities, internal energy, heat capacity, partition function, and entropy can be calculated [50] for translation, rotation and vibrational degrees of freedom for a single temperature, or a range of temperatures. Special situations such as linear systems and transition states are accommodated. The approximations used in the THERMO calculation are invalid below 100K, and checking of the lower bound of the temperature range is done to prevent temperatures of less than 100K being used. See Section 5.15 for more detail.

Another limitation, for which no checking is done, is that there should be no internal rotations. If any exist, they will not be recognized as such, and the calculated quantities will be too low as a result.

If THERMO is specified on its own, then the default values of the temperature range are assumed. This starts at 200K and increases in steps of 10 degrees to 400K. Three options exist for overriding the default temperature range. These are:

THERMO(*nnn*) or THERMO=(*nnn*) **(O)**

The thermodynamic quantities for a 200 degree range of temperatures, starting at *nnn*K (integer) and with an interval of 10 degrees are to be calculated.

THERMO(*nnn*,*mmm*) or THERMO=(*nnn*,*mmm*) **(O)**

The thermodynamic quantities for the temperature range limited by a lower bound of *nnn* Kelvin and an upper bound of *mmm* Kelvin (both integer numbers), the step size being calculated in order to give approximately 20 points, and a reasonable value for the step. The size of the step in Kelvin degrees will be 1, 2, or 5, or a power of 10 times these numbers.

THERMO(*nnn*,*mmm*,*lll*) or THERMO=(*nnn*,*mmm*,*lll*) **(O)**

Same as for THERMO(*nnn*,*mmm*), (three integer numbers) only now the user can explicitly define the step size. The step size, *lll*, cannot be less than 1K.

THRESH=*n.nn* **(W)**

In TIDY the LMOs are tidied up: this involves deleting atoms that have very low intensity in an LMO, and removing the unused space that is generated as a result. The intensity, ρ , of an atom in a LMO is given by:

$$\rho_A = \sum_{\lambda \in A} \psi_{\lambda}^2 \quad (3.1)$$

The criterion for 'very low intensity' has a default of 10^{-10} , and can be reset by THRESH=*n.nn*. See also RELTHR.

TIDY **(O)**

Some of the working in subroutine TIDY is printed when TIDY is specified.

TIMES **(O)**

The times for various stages of the calculation are printed when TIMES is specified. This is very useful in finding out what is using a lot of time. A second use is to find out precisely where in a run a failure occurred.

TOM **(O)**

The MIERTUS-SCROCCO-TOMASI self-consistent reaction field model for solvation is used when TOM is specified.

For more detail on this technique, see page 247.

When TOM is specified, an additional keyword (H2O, CHCL3, or CCL4) is necessary or extra data must be supplied at the end of the input file.

A description of this data is given on page 248.

TRANS (C)

The imaginary frequency due to the reaction vector in a transition state calculation must not be included in the thermochemical calculation. The number of genuine vibrations considered can be: $3N - 5$ for a linear ground state system, $3N - 6$ for a non-linear ground state system, or $3N - 6$ for a linear transition-state complex, $3N - 7$ for a non-linear transition-state complex.

This keyword must be used in conjunction with THERMO if a transition state is being calculated.

TRANS=*n* (C)

The facility exists to allow the THERMO calculation to handle systems with internal rotations. TRANS=*n* will remove the *n* lowest vibrations. Note that TRANS=1 is equivalent to TRANS on its own. For xylene, for example, TRANS=2 would be suitable.

TRIPLET (C)

The triplet state is defined. In order to define this type of calculation other keywords *must* also be used. For a 'simple' triplet calculation, use C.I.=2. Results from such calculations can be compared with ground state calculations. If the triplet state consists of two half-filled degenerate M.O.s, such as molecular oxygen, then OPEN(2,2) should be used.

From experience, OPEN(2,2) and TRIPLET run faster than C.I.=2 and TRIPLET.

If the system has an odd number of electrons, an error message will be printed.

See also SINGLET, DOUBLET, QUARTET, QUINTET, SEXTET, SEPTET, OCTET, and NONET.

UHF interpretation:

The number of alpha electrons exceeds that of the beta electrons by 2. If TRIPLET is not specified, then the numbers of alpha and beta electrons are set equal. This does not necessarily correspond to a singlet.

RHF interpretation:

TRIPLET cannot be used unless other keywords are present. If C.I.=2 is used, then a single state corresponding to:

$$\frac{1}{\sqrt{2}}(\psi_{\text{homo}}^{\alpha} \cdot \psi_{\text{lumo}}^{\beta} + \psi_{\text{homo}}^{\beta} \cdot \psi_{\text{lumo}}^{\alpha})$$

is calculated. See keywords C.I.=*n* and OPEN(*n*₁,*n*₂).

When the configuration interaction calculation is done, all microstates having a component of spin, M_S , equal to 1 are selected. These microstates are then used in the construction of states. Because of the way in which the microstates were chosen, only states of spin equal to or greater than 1 can be constructed. From this set, only a triplet state can be selected, all other states will be ignored. If ROOT=*n* is present, then the *n*'th triplet state will be selected, otherwise the first triplet state will be chosen.

TS (C)

Within the Eigenvector Following routine [7], the option exists to optimize a transition state. To do this, use TS. Preliminary indications are that the TS method is much faster and more reliable than either SIGMA or NLLSQ.

TS appears to work well with Cartesian coordinates.

In the event that TS does not converge on a stationary point, try adding RECALC=5 to the keyword line.

UHF (C)

The unrestricted Hartree-Fock Hamiltonian is to be used.

UNSAFE (W)

If MOPAC has been compiled so that the special SCF convergers can be used, then the memory demand will be higher than if they are not allowed. These convergers are not needed for well-behaved systems, and therefore more memory is needed than is used. To reduce the memory demand for a run, add UNSAFE to the keyword line. A side effect of this command is that some systems will fail to generate an SCF, and the run may be wasted as a result. Do not use UNSAFE unless there is a need to conserve memory resources.

VDW(*text*) (W)

The Van der Waals radii of atoms can be set or changed by use of VDW. The format of the command is:

```
VDW(:<chemical symbol1>=n.nn:<chemical symbol2>=n.nn:<chemical symbol3>=n.nn: ... )
```

For example, VDW(:Cl=2.33:Br=2.50) would override the default values of the van der Waals radii of chlorine and bromine (1.65 and 1.80, respectively). Note that all chemical symbols (including the first one) must be preceded by a colon (:).

This keyword is used in setting the VDW radii in the COSMO solvation method.

The VDW radii are also used in determining connectivity in the generation of the Lewis structure. Thus, to make sure that a magnesium atom is fully ionic, the keyword VDW(:Mg=--1.0) could be used. Note that a large negative VDW radius ensures that no atoms are within the VDW radius.

VECTORS (O)

MOPAC:

The eigenvectors are to be printed. In UHF calculations both alpha and beta eigenvectors are printed; if ALLVEC is specified, in all cases the full set, occupied and virtual, are output. By default, the nine highest occupied and the seven lowest unoccupied levels are printed. The eigenvectors are normalized to unity: that is, the sum of the squares of the coefficients is exactly one. If DEBUG is specified, then the eigenvectors on every iteration of every SCF calculation will be printed. This is useful in a learning context, but would normally be undesirable.

MOZYME:

The molecular orbitals are printed if VECTORS is present. The type of M.O.s printed depends on the presence of another keyword, EIGEN.

If EIGEN is *not* present, then the final set of occupied localized molecular orbitals is to be printed.

If EIGEN is present, then the conventional M.O.s or eigenvectors are printed. By default, only the M.O.s near the HOMO-LUMO gap are printed. In order to print all the M.O.s, use the keywords VECTORS, EIGEN, and ALLVECS.

VELOCITY (C)

The user can supply the initial velocity vector to start a DRC calculation. For obvious reasons, the input geometry should be in Cartesian coordinates. If internal coordinates must be used, add GEO=OK.

Put the velocity vector after the geometry as three data per line, representing the *x*, *y*, and *z* components of velocity for each atom. The units of velocity are centimeters per second.

If KINETIC=*n.n* is also specified, the velocity vector will be scaled to equal the velocity corresponding to *n.n* kcal/mol. This allows the user to define the direction of the velocity vector; the magnitude is given by KINETIC=*n.n*.

WILLIAMS (C)

Within the ESP calculation, the Connolly surface is used as the default. If the surface generation procedure of Donald Williams is wanted, the keyword WILLIAMS should be used.

WINMOPAC (O)

WINMOPAC will write out the geometry and MO information to <filename>.wmp and <filename>.wob using WINMOPAC format.

X-PRIORITY (O)

In a DRC calculation, results will be printed whenever the calculated geometry changes by 0.05 Å. The geometry change is defined as the linear sum of the translation vectors of motion for all atoms in the system. Abbreviation, X-PRI0. See Section 5.19 for more details.

X-PRIORITY=*n.nn* (O)

In a DRC calculation, results will be printed whenever the calculated geometry changes by *n.nn* Å.

XENO (O)

By default, MOZYME only recognizes the standard twenty amino-acid residues, but some proteins contain residues which have extra molecular fragments attached. In order to allow these unusual species to be recognized, the keyword XENO is provided.

XENO, from the Greek *ξένος*, *xenos*, for 'stranger', defines the unusual species in terms of the extra atoms which are added to the normal residue. The format of XENO is:

XENO=(*n_C*,*n_N*,*n_O*,*n_S*,*name* [;*n_C*,*n_N*,*n_O*,*n_S*,*name*] [;*n_C*,*n_N*,*n_O*,*n_S*,*name*] ...)

Up to four fragments can be defined. To specify a fragment, the number of extra carbon, nitrogen, oxygen, and sulfur atoms in the fragment is used. A *name* should be selected which describes the fragment.

As an example, consider the extra fragment in bacteriorhodopsin, on Lys216:

–COCHRNH–; R = CH₂CH₂CH₂CH₂N=CHCH=CM_eCH=CHCH=CM_eCH=CHC₉H₁₅

Not counting hydrogens, the empirical formula for a lysine fragment is C₆N₂O (Table 3.10, p. 70). For residue 216, the empirical formula is C₂₆N₂O, so the extra fragment, the Schiff base, accounts for C₂₀. Therefore, the number of extra atoms is, in the order C, N, O, S, '20,0,0,0'. To specify the retinal fragment, the keyword would be XENO=(20,0,0,0,RETINAL).

Only the number of extra atoms for the elements C, N, O, and S need be specified, because these are the elements used in identifying the residues. Thus, a phosphorylated serine, of formula: –COCHRNH–; R = CH₂–(H₂PO₄) would be specified by the keyword XENO(0,0,3,0,P04).

In the output, all the atoms of the residue are labeled with the three letter abbreviation of the amino-acid. Ideally, the atoms of the extra fragment would be labeled differently, but it is not easy to algorithmically 'recognize' the fragment. Instead, the unusual residue is indicated in the residue sequence by an asterisk (*), and a one-line description given immediately before the sequence is printed.

If XENO is *not* used, the calculation will still work, but the label for the modified fragment will be UNK, instead of the more descriptive label which would result from using XENO.

XYZ (W)

Regardless of the coordinate system used in defining the geometry, if XYZ is present, then all atoms will be converted to Cartesian coordinates, and the calculation will be run entirely in Cartesian coordinates.

When a large ring system is being optimized, the closure is sometimes difficult, in which case XYZ will normally work.

3.3.3 Keywords that go together

Normally only a subset of keywords are used in any given piece of research. Keywords which are related to each other in this way are:

1. In getting an SCF: SHIFT, PULAY, ITRY, CAMP, SCFCRT, 1SCF, PL.
2. In MECI work: SINGLET, DOUBLET, etc., OPEN(n,m), C.I.=(n,m), CIS, CISD, PECI, CISDT, LARGE, MECI, MS= n , VECTORS, ESR, ROOT= n , MICROS. See Section 5.4 for more details.
3. In excited states: UHF with (TRIPLET, QUARTET, etc.), C.I.= n , C.I.=(n,m). See Section 5.4 for more details.
4. In geometry optimization:
 - (a) Using BFGS: GNORM= $n.n$, XYZ, PRECISE.
 - (b) Using EF: GNORM= $n.n$, XYZ, PRECISE, NOUPD, RSCAL, NONR, GNMIN, CYCLES, RECALC, IUPD, MODE, DDMIN, DDMAX, DMAX, HESS, RMIN, RMAX, OMIN.
 - (c) Using NLLSQ: GNORM= $n.n$, XYZ, PRECISE
 - (d) Using SIGMA: GNORM= $n.n$, XYZ, PRECISE
5. In COSMO calculations: RSOLV, DISEX, NSPA, VDW, EPS.
6. In Gaussian work: AIGIN, AIGOUT, AIDER.
7. In SADDLE: BAR= $n.n$
8. In DRC or IRC: see Section 5.19 for more details.

See also Sections 5.4.8, 5.19 and Chapter 9 for other keywords that are usually associated together.

3.4 Geometry Specification

The molecular or unit cell geometry is supplied in the data-set as one atom per line. Sometimes it is useful to have points within a geometry defined. These points need not represent atoms. To allow for this, users can add entities called “dummy atoms”.

The format in which the data is supplied is atoms essentially the “Free-Format” style of FORTRAN-77. In fact, a character input is used in order to accommodate the chemical symbols, but the numeric data can be regarded as “free-format”. This means that integers and real numbers can be interspersed, and numbers can be separated by one or more spaces, a tab and/or by one comma. If a number is not specified, its value is set to zero.

The geometry can be defined in terms of either internal or Cartesian coordinates, or a mixture of the two, or it can be in PDB or Gaussian format.

Table 3.13: Elements available within MNDO

Elements	Dummy atom, sparkles and Translation Vector												
H													
Li	Be	B	C	N	O	F							
Na [†]		Al	Si	P	S	Cl							
K [†]	Zn	*	Ge	*	*	Br	XX [‡]	Cb	++	+	--	-	Tv ⁺
Rb [†]	...	*	Sn	*	*	I	99	102	103	104	105	106	107
*	...	Hg	*	Pb	*								

[†]These symbols refer to elements which lack a basis set.

[‡]This is the dummy atom for assisting with geometry specification.

* Element not parameterized.

+ This is the translation vector for use with polymers.

Table 3.14: Elements available within AM1

Elements	Dummy atom, sparkles and Translation Vector															
H																
*	*			B	C	N	O	F								
Na [†]	*			Al	Si	P	S	Cl								
K [†]	*	Fe	Cu	Zn	*	Ge	As	Se	Br	XX [‡]	Cb	++	+	--	-	Tv ⁺
Rb [†]	*		Ag	*	*	*	Sb	Te	I	99	102	103	104	105	106	107
*	*	Mo	Pt	Hg	*	*	*									

[†]These symbols refer to elements which lack a basis set.

[‡]This is the dummy atom for assisting with geometry specification.

* Element not parameterized.

+ This is the translation vector for use with polymers.

3.4.1 Definition of Elements and Isotopes

Elements are defined in terms of their atomic numbers or their chemical symbols, case insensitive. Thus, chlorine could be specified as 17, or Cl.

Acceptable symbols for MNDO are given in Table 3.13.

Old parameters for some elements are available. These are provided to allow compatibility with earlier copies of MOPAC. To use these older parameters, use a keyword composed of the chemical symbol followed by the year of publication of the parameters. Keywords currently available: Si1978, S1978.

Table 3.14 lists all elements available in AM1.

If users need to use other elements, such as beryllium or lead, they can be specified, in which case MNDO-type atoms will be used. As the behavior of such systems is not well investigated, users are cautioned to exercise unusual care. To alert users to this situation, the keyword PARASOK is given.

For PM3, acceptable symbols are shown in Table 3.15. Transition metals are not yet available by default. However, parameters for the method PM3(tm) are readily available, and can be added to MOPAC by means of the "EXTERNAL" command. For example, by specifying EXTERNAL=pm3_tm, and having a file called pm3_tm in the same subdirectory, parameters for PM3(tm) can be added to MOPAC.

The form of the file pm3_tm is shown in Figure 3.15.

Acceptable symbols for MNDO-*d* are given in Table 3.16.

MINDO/3 uses diatomic pairs. In order for a system to be calculable, all pairs of types of atoms must be available within MINDO/3. Pairs available are shown in Table 3.17.

Note: MINDO/3 should now be regarded as being of historical interest only. MOPAC contains the original parameters. These do not reproduce the original reported results in the case of P, Si, or S. The original work was faulty [51]. See also G. Frenking, H. Goetz, and F. Marschner, *J. Am. Chem. Soc.*, 100:5295 (1978). Re-optimized

Table 3.15: Elements available within PM3
Dummy atom, sparkles and Translation Vector

Elements										Dummy atom, sparkles and Translation Vector							
H																	
Li	Be			*	C	N	O	F									
Na [†]	Mg			Al	Si	P	S	Cl									
K [†]	*	...	Zn	Ga	Ge	As	Se	Br	XX [‡]	Cb	++	+	--	-		Tv ⁺	
Rb [†]	*	...	Cd	In	Sn	Sb	Te	I	99	102	103	104	105	106		107	
*	*	...	Hg	Tl	Pb	Bi											

[†]These symbols refer to elements which lack a basis set.

[‡]This is the dummy atom for assisting with geometry specification.

* Element not parameterized.

+ This is the translation vector for use with polymers.

Figure 3.15: Example of Parameter Set for PM3(tm)

```

atnum      22
n           4.0
uss        -26.45829779
upp        -21.17197024
udd        -36.34653108
betas      -23.30450933
.
.
.
a2         -0.01221641
b2         4.07610129
c2         2.79140274
hform      112.30000000
zval       4.00000000

```

If this data set is called pm3_tm, it can be made available to MOPAC by using EXTERNAL=pm3_tm. If more than one transition metal parameter set is needed, the additional parameter sets are attached to the end of the file pm3_tm.

Table 3.16: Elements available within MNDO-d
Dummy atom, sparkles and Translation Vector

Elements										Dummy atom, sparkles and Translation Vector							
<i>H</i>																	
<i>Li</i>	<i>Be</i>			<i>B</i>	<i>C</i>	<i>N</i>	<i>O</i>	<i>F</i>									
Na [§]	Mg [§]			Al	Si	P	S	Cl									
<i>K</i> [†]		...	Zn [§]	*	<i>Ge</i>	*	*	Br	XX [‡]	Cb	++	+	--	-		Tv ⁺	
<i>Rb</i> [†]		...	Cd [§]	*	<i>Sn</i>	*	*	I	99	102	103	104	105	106		107	
*		...	Hg [§]		<i>Pb</i>	*											

Elements in *italics* use the old MNDO parameters.

Elements in **bold face** have $s - p - d$ basis sets.

[†]These symbols refer to elements which lack a basis set.

[‡]This is the dummy atom for assisting with geometry specification.

* Element not parameterized.

[§]These atoms have $s - p$ basis sets only.

+ This is the translation vector for use with polymers.

Table 3.17: Diatomics Parameterized within MINDO/3

	H	B	C	N	O	F	Si	P	S	Cl
H	*	*	*	*	*	*	*	*	*	*
B	*	*	*	*	*	*				
C	*	*	*	*	*	*	*	*	*	*
N	*	*	*	*	*	*			*	*
O	*	*	*	*	*	*		*		
F	*	*	*	*	*	*				
Si	*		*							
P	*		*		*	*		*		*
S	*		*	*	*				*	*
Cl	*		*	*				*	*	*

A star (*) indicates that the atom-pair is parameterized within MINDO/3

parameters for P–C and P–Cl were derived later which gave better results. These are:

- Alpha(P–C): 0.8700 G. Frenking, H. Goetz, F. Marschner,
- Beta(P–C): 0.5000 *J. Am. Chem. Soc.*, 100:5295-5296 (1978).
- Alpha(P–Cl): 1.5400 G. Frenking, F. Marschner, H. Goetz,
- Beta(P–Cl): 0.2800 *Phosphorus and Sulfur*, 8:337-342 (1980).

Although better than the original parameters, these have not been adopted within MOPAC because to do so at this time would prevent earlier calculations from being duplicated. Parameters for P–O and P–F have been added: these were abstracted from Frenking's 1980 paper. No inconsistency is involved as MINDO/3 historically did not have P–O or P–F parameters.

Extra entities available to MNDO, MINDO/3, AM1 and PM3:

- + A 100% ionic alkali metal.
- ++ A 100% ionic alkaline earth metal.
- A 100% ionic halogen-like atom
- A 100% ionic group VI-like atom.
- Cb A special type of monovalent atom

Elements 103, 104, 105, and 106 are the sparkles; elements 11 and 19 are sparkles tailored to look like the alkaline metal ions; Tv is the translation vector for polymer calculations. See "Sparkles" in Section 5.3.11

Element 102, symbol Cb, is designed to satisfy valency requirements of atoms for which some bonds are not completed. Thus in "solid" diamond the usual way to complete the normal valency if the cluster model is *not* used is to use hydrogen atoms. This approach has the defect that the electronegativity of hydrogen is different from that of carbon. The "capped bond" atom, Cb, is designed to satisfy these valency requirements without acquiring a net charge.

Cb behaves like a monovalent atom, with the exception that it can alter its electronegativity to achieve an exactly zero charge in whatever environment it finds itself. It is thus all things to all atoms. On bonding to hydrogen it behaves similarly to a hydrogen atom. On bonding to fluorine it behaves like a very electronegative atom. If several capped bond atoms are used, each will behave independently. Thus if the two hydrogen atoms in formic acid were replaced by Cbs, then each Cb would independently become electroneutral.

Capped bonds internal coordinates should not be optimized. A fixed bond-length of 1.7 Å is recommended; if two Cbs are on one atom, a contained angle of 109.471221 degrees is suggested, and if three Cbs are on one atom, a contained dihedral of –120 degrees (note sign) should be used.

Element 99, X, or XX, is known as a dummy atom and is used in the definition of the geometry; it is deleted automatically from any Cartesian coordinate geometry files. Dummy atoms are pure mathematic points, and are useful in defining geometries; for example, in ammonia the definition of C_{3v} symmetry is facilitated by using one dummy atom and symmetry relating the three hydrogens to it.

Isotopes are used in conjunction with chemical symbols. If no isotope is specified, the average isotopic mass is used, thus chlorine is 35.453. This is different from early versions of MOPAC (before 1993), in which the most abundant isotope was used by default. This change was justified by the removal of any ambiguity in the choice of isotope. Also, the experimental vibrational spectra involve a mixture of isotopes. If a user wishes to specify any specific isotope it should immediately follow the chemical symbol (no space), e.g., H2, H2.0140, C(meta)13, or C13.00335.

The sparkles ++, +, --, and - have no mass; if they are to be used in a force calculation, then appropriate masses should be used.

3.4.2 Optimization flags

Each internal coordinate is followed by an flag, to indicate the action to be taken.

Flag	Action
1	Optimize the internal coordinate.
0	Do not optimize the internal coordinate.
-1	Reaction coordinate, or grid index.
T	Monitor turning points in DRC

Remarks: Only one reaction coordinate is allowed, but this can be made more versatile by the use of SYMMETRY. Two methods of specifying the points to be used on the reaction coordinate are allowed. If the user wants to explicitly specify each point, then the values of the reaction coordinate should follow immediately after the geometry and any symmetry data. No terminator is required, and free-format-type input is acceptable.

If the points to be used form a regular sequence, then the preferred method of specifying them is by use of the two keywords POINT=*n* and STEP=*n.nn*. For example, to rotate a torsion angle through 360° in steps of 10°, the keywords POINT=36 STEP=10 would be used.

If two “reaction coordinates” are used, then MOPAC assumes that the two-dimensional space in the region of the supplied geometry is to be mapped. The two dimensions to be mapped are in the plane defined by the “-1” labels. Step sizes in the two directions must be supplied using STEP1 and STEP2 on the keyword line.

Using internal coordinates, the first atom has three unoptimizable coordinates, the second atom two, (the bond-length can be optimized) and the third atom has one unoptimizable coordinate. None of these six unoptimizable coordinates at the start of the geometry should be marked for optimization. If any are so marked, a warning is given, but the calculation will continue.

In Cartesian coordinates all parameters can be optimized.

In IRC/DRC calculations, the flag “T” specifies that turning points for the coordinate are to be printed, see Section 5.19 for more detail.

3.4.3 Internal Coordinate Definition

For any one atom, *i*, this consists of an interatomic distance in Ångstroms from an already-defined atom, *j*, an interatomic angle in degrees between atoms *i* and *j* and an already defined *k*, (*k* and *j* must be different atoms), and, finally, a torsional angle in degrees between atoms *i*, *j*, *k*, and an already defined atom *l* (*l* cannot be the same as *k* or *j*). See also Section 5.12. Exceptions:

1. Atom 1 has no internal coordinates at all. The coordinates of atom 1 are, by definition, Cartesian. Normally, the coordinates of atom 1 are (0,0,0), but can be set to any value desired.
2. Atom 2 must be connected to atom 1 by an interatomic distance only. If atom 1 is not at the origin, then the care must be taken in defining atom 2: if internal coordinates are used, then the connectivity must be given. If the connectivity is not specified, then the coordinate of atom 2 is, by definition, Cartesian.
3. Atom 3 can be connected to atom 1 or 2, and must make an angle with atom 2 or 1 (thus 3–2–1 or 3–1–2); no dihedral is possible for atom 3. Again, if the connectivity is not given, the coordinate is defined as Cartesian.

Constraints

1. Interatomic distances must be greater than zero. Zero Ångstroms is acceptable only if the parameter is symmetry-related to another atom, and is the dependent function.
2. Angles must be in the range 0.0 to 180.0, inclusive. This constraint is for the benefit of the user only; negative angles are the result of errors in the construction of the geometry, and angles greater than 180 degrees are fruitful sources of errors in the dihedrals.
3. Dihedral angles must be definable. If atom *i* makes a dihedral with atoms *j*, *k*, and *l*, and the three atoms *j*, *k*, and *l* are in a straight line, then the dihedral has no definable angle. During the calculation this constraint is checked continuously, and if atoms *j*, *k*, and *l* lie within 0.02 Ångstroms of a straight line, an attempt will be made to re-number the connectivity. If this fails, the calculation will output an error message and then stop. The exceptions to this constraint are:
 - (a) if the angle is zero or 180 degrees, in which case the dihedral is not used.
 - (b) if atoms *j*, *k*, and *l* lie in an exactly straight line (usually the result of a symmetry constraint), as in acetylene, acetonitrile, but-2-yne, etc.

If the exceptions are used, care must be taken to ensure that the program does not violate these constraints during any optimizations or during any calculations of derivatives—see also FORCE.

Conversion to Cartesian Coordinates

Normally, atom 1 is at the origin of Cartesian coordinate space. However, if atom 1 has non-zero coordinates, then it will be at the position defined by its coordinates. Be careful, however, if atom 1 is a dummy atom. Atom 2 is defined as being displaced in the X-direction from atom 1 by a distance equal to its bond length. Atom 3 is in the X-Y plane of atoms 1 and 2 unless the angle 3–2–1 is exactly 0 or 180 degrees. Atom 4, 5, 6, etc. can lie anywhere in 3-D space.

3.4.4 Translation Vector

The translation vector is the distance through which an atom must be moved (translated) in order to be in the next unit cell. The symbol for a translation vector is T_v . The T_v should be specified at the end of the geometry. No real or dummy atoms should come after a T_v , although symmetry data, etc, are allowed. The order of the data is thus: (real and dummy atoms) (translation vector(s)) (blank line) (symmetry data or MECI data etc, if needed).

Each translation vector is specified by one T_v . The 'bond length' of the T_v is the translation vector distance, and the direction of the vector is given by the connectivity. An example of a single T_v is given on page 101, where the translation vector distance is 12.3Ångstroms, and moves atom 1 in the direction of dummy atom 11.

Dummy atoms are useful in defining the direction of the T_V . There must not be real atoms in two unit cells related by a T_V , but at the same time we want a way to specify where an atom will be after it has been translated. Therefore it is convenient to put a dummy atom where an original atom will be moved to, and then define the T_V in terms of that dummy atom and the original atom.

Note that in the polytetrahydrofuran example, the angle and dihedral of the T_V are not flagged for optimization. This is because the direction of the T_V can be defined as being from a real atom (here atom 1) to a dummy atom (atom 11), and the coordinates of the dummy atom are marked for optimization. Unless absolutely necessary, do *not* optimize the angle and dihedral of a T_V .

Two and three dimensional systems have two and three T_V respectively.

3.4.5 Gaussian Z-matrices

With certain limitations, geometries can be specified within MOPAC using the Gaussian [14] Z-matrix format.

Specification of Gaussian Z-matrices

The information contained in the Gaussian Z-matrix is identical to that in a MOPAC Z-matrix, but the order of presentation is different. Atom N, (real or dummy) is specified in the format:

Element N1 Length N2 Alpha N3 Beta

where Element is the same as for the MOPAC Z-matrix. N1, N2, and N3 are the connectivity, the same as the MOPAC Z-matrix NA, NB, and NC: bond lengths are between N and N1, angles are between N, N1 and N2, and dihedrals are between N, N1, N2, and N3. The same rules apply to N1, N2, and N3 as to NA, NB, and NC.

Length, Alpha, and Beta are the bond lengths, the angle, and dihedral. They can be 'real', e.g. 1.45, 109.4, 180.0, or 'symbolic'. A symbolic is an alphanumeric string of up to 8 characters, e.g. R51, A512, D5213, CH, CHO, CHOC, etc. Two or more symbolics can be the same. Dihedral symbolics can optionally be preceded by a minus sign (Figure 3.16), in which case the value of the dihedral is the negative of the value of the symbolic. This is the equivalent of the normal MOPAC SYMMETRY operations 1, 2, 3, and 14.

If an internal coordinate is real, it will not be optimized. This is the equivalent of the MOPAC optimization flag "0". If an internal coordinate is symbolic, it can be optimized.

The Z-matrix is terminated by a blank line, after which comes the starting values of the symbolics, one per line. If there is a blank line in this set, then all symbolics after the blank line are considered fixed; that is, they will not be optimized. The set before the blank line will be optimized.

Exceptions to the full Gaussian standard

1. The option of defining an atom's position by one distance and two angles is not allowed. In other words, the N4 variable described in the Gaussian manual must either be zero or not specified. MOPAC requires the geometry of atoms to be defined in terms of, at most, one distance, one angle, and one dihedral.
2. Gaussian Cartesian coordinates are not supported.
3. Chemical symbols must not be followed by an integer identifying the atom. Numbers after a symbol are used by MOPAC to indicate isotopic mass. If labels are desired, they should be enclosed in parentheses, thus $C1(ON\ C5)34.96885$.
4. The connectivity (N1, N2, N3) must be integers. Labels are not allowed.

```

Line 1  AM1
Line 2  Ethane
Line 3
Line 4  C
Line 5  C      1      r21
Line 6  H      2      r32      1      a321
Line 7  H      2      r32      1      a321      3  d4213
Line 8  H      2      r32      1      a321      3 -d4213
Line 9  H      1      r32      2      a321      3   60.
Line 10 H      1      r32      2      a321      3  180.
Line 11 H      1      r32      2      a321      3  d300
Line 12
Line 13      r21      1.5
Line 14      r32      1.1
Line 15      a321     109.5
Line 16      d4313    120.0
Line 17
Line 18      d300     300.0
Line 19

```

Figure 3.16: Example of Gaussian Z-matrix geometry specification

3.4.6 Cartesian Coordinate Definition

A definition of geometry in Cartesian coordinates consists of the chemical symbol or atomic number, followed by the Cartesian coordinates and optimization flags but no connectivity.

MOPAC uses the lack of connectivity to indicate that Cartesian coordinates are to be used.

3.4.7 Protein Data Bank Format

Protein geometries are usually defined using the Brookhaven Protein Data Bank format. A full definition of this format is given in the web-page:

http://www.pdb.bnl.gov/pdb-docs/Format.doc/Contents_Guide_21.html, and in

<http://pdb.pdb.bnl.gov/pdb-docs/atoms.htm>

The home-page of the PDB is <http://www.pdb.bnl.gov>. This is an excellent resource for obtaining biomolecular macromolecules.

There are several ways of converting a PDB file into a form suitable for use by MOPAC. For most systems in the PDB the hydrogen atoms are missing. If they are present, that is, the data set is complete, then the file can simply be edited as follows:

1. Delete the header of the PDB file, down to the first line that starts with the word ATOM.
2. Delete the tail of the file back to the last occurrence of a line starting with the word HETATM.
3. Add the keyword line(s) and comment line(s).

An alternative way of generating a MOPAC data set is to read in the geometry using Cambridge Soft's CHEM3D program. This is particularly useful if the hydrogens are *not* present in the PDB file.

After reading in the file, hydrogens can be added using the RECTIFY function in CHEM3D. Once this is done, the data set can be exported for use by MOPAC using the SAVE AS option.

Preparing the data set

Even when the hydrogens are present in the PDB file, for example in crambin, file=P1CBN, there are usually problems that prevent the file being run. Therefore, before attempting to calculate the electronic structure, any faults in the data set should be corrected. The following procedure has been found to be useful:

- First, the atoms should be rearranged into the standard PDB sequence. This is a quick operation, and only one keyword, RESEQ, should be used. Examination of the output will indicate whether any severe problems exist. Potential problems are:
 - The system consists of many short sections of protein. This usually occurs when only a fragment of the system is present, e.g., the active site. Working with such a system is difficult, because the geometry does not correspond to anything that can exist naturally.
 - One or more UNK residues exist. Before continuing, identify the unknown residues. Ensure that they are not artifacts of a faulty geometry.
 - One or more unknown elements exist. The author of the PDB file may have used non-standard labels for atoms. This fault can be corrected by use of PDB(...).
 - Structural or positional disorder may exist. Use ALT_R=*n* and/or ALT_A=*n* to correct this.
- Edit the output file to create a new data set.
- If any very small interatomic distances (less than 0.8 Ångstroms) were reported, edit the data set to correct these faults. A quick way to find faulty distances is to search the data set for “) 0.6”, then “) 0.7”, etc.
- Run the corrected data set using keywords 1SCF, RESIDUES, and MOZ. Almost certainly, the run will fail. Usually the charge on the system, as determined by the MOZYME function, will be incorrect. If it is in the range +2 to -2, that is, in a chemically meaningful range, then simply add CHARGE=*n* and re-submit.
- Usually the charge is quite unreasonable (outside the range +2 to -2). Examination of the charged atoms will usually suggest how to proceed. If the charge is large and negative, add hydrogens to the oxygen atoms of ionized acid residues, to neutralize them. If it is large and positive, delete hydrogen atoms from nitrogen atoms of ionized amine groups. Be careful to neutralize an even number of ions—if an odd number is neutralized, then the system will have an odd number of electrons, and the job will stop as soon as the number of electrons is determined.
- After correcting the errors, run a single SCF, followed by more complicated calculations, as needed.

3.4.8 Conversion Between Various Formats

MOPAC can accept any of the following formats: Cartesian, MOPAC internal coordinates, a mixture of Cartesian and MOPAC internal coordinates, Brookhaven Protein Data Bank, and Gaussian internal coordinates. Both MOPAC and Gaussian Z-matrices can also contain dummy atoms. If the 0SCF option is requested, the geometry defined on input will be printed in MOPAC Z-matrix format, along with other optional formats.

The type(s) of geometry printed at the end of a 0SCF calculation depend only on the keywords XYZ, INT, AIGOUT, PDBOUT, and NOXYZ. The geometry printed is independent of the type of input geometry, and therefore makes a convenient conversion mechanism.

If XYZ is present, all dummy atoms are removed and the internal coordinate definition remade. All symmetry relations are lost if XYZ is used.

If NOXYZ is present, Cartesian coordinates will not be printed.

```

MINDO/3
Formic acid
Example of normal geometry definition
O
C   1.20 1
O   1.32 1 116.8 1   0.0 0   2 1
H   0.98 1 123.9 1   0.0 0   3 2 1
H   1.11 1 127.3 1 180.0 0   2 1 3
O   0.00 0   0.0 0   0.0 0   0 0 0

```

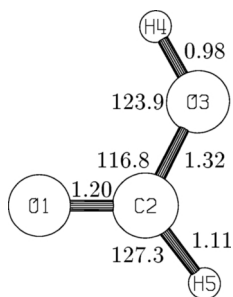


Figure 3.17: Data set for Formic acid

If AIGOUT is present, a data set using Gaussian Z-matrix format is printed.

Note: Only if the keyword SYMMETRY is present in a MOPAC internal coordinate geometry, or two or more internal coordinates in a Gaussian Z-matrix have the same symbolic will symmetry be present in the MOPAC or Gaussian geometries output.

3.4.9 Examples of Coordinate Definitions

Formic acid

The data set for formic acid is given in Figure 3.17.

The geometry in this data-set can be understood as follows: Atom 1, an oxygen, is at the origin of internal coordinate space, and has coordinates (0,0,0). Atom 2, a carbon, is positioned at coordinate (1.20,0,0), that is, it is related to the oxygen by a bond-length of 1.20 Ångstroms, and to atom 3, an oxygen, by a bond-length of 1.32 Ångstroms. The O-C-O angle is 116.8 degrees. The first hydrogen is bonded to the hydroxyl oxygen and the second hydrogen is bonded to the carbon atom. The H-C-O-O dihedral angle is 180 degrees.

3.5 Memory Considerations

Of their nature, large systems use a large amount of memory. This will sometimes prevent a calculation from being run. There are, however, several ways to reduce the amount of memory needed by a specific calculation. These can be summarized as follows:

- Run using the MOZYME function.

For large systems, the MOZYME function uses much less memory than the default MOPAC. The MOZYME function only works for closed shell RHF ground state systems. Also, it cannot be used when high precision is needed, such as in FORCE, hyperpolarizability, or IRC calculations.

- Run the job with `SIZES`.

Run the job that fails using `SIZES`. This will print out the sizes of all arrays created dynamically. By inspection of these arrays, ways can usually be found to reduce the memory demand. For example, the default geometry optimizer using the `MOZYME` function, `EF`, uses a large amount of memory. Switching to the `BFGS` optimizer (`BFGS`) will save a lot of memory.

- Run in `UNSAFE` mode.

Normally, by default, MOPAC runs in `SAFE` mode. This means that when things start to go wrong in the SCF, special procedures can be used in an attempt to generate a SCF. These procedures use a large amount of memory, and are normally not used. Most likely, if memory is limited, the job can be run in `UNSAFE` mode. This will save a lot of memory, but at the same time, if things start to go wrong, the job has an increased chance of failing. If the job fails, the results will usually indicate the failure in an obvious way.

- Reduce the cutoffs.

In a `MOZYME` calculation, useful results can often be obtained even when the values of `CUTOF1` and `CUTOF2` are reduced. When these are reduced, the amount of memory needed drops quite rapidly.

If the objective of the calculation is to optimize the geometry, then reducing the cutoffs will both increase the speed and reduce the memory demand without significant loss of accuracy.

Chapter 4

Examples

A large number of examples of the type of calculation that can be run using MOPAC are supplied with the program. These examples can be found in the directory `... /mopac_examples/`. Each subdirectory contains data sets that illustrate one topic within MOPAC. In order to run any tests, the contents of the directory should be copied to a new, empty, directory, and the tests run from this new directory. Once the tests are complete, the contents of the new directory can be deleted. It is important that *all* the files in the `... /mopac_examples/` subdirectory be copied.

4.1 To verify MOPAC

After MOPAC has been modified, either to make an in-house improvement, or to port MOPAC to a new platform, you might want to verify that the program has not been damaged. Data-set `port.dat` is intended to be used for this purpose. This data set makes use of a large fraction (about 85%) of the code in MOPAC, and is therefore very sensitive to changes made to the program. Because it does test most of MOPAC, `port.dat` is very exotic, and does not represent a realistic calculation.

If the run fails for any reason, the data-set can be edited to remove all the tests which have been successfully passed, and the new data-set given a new name. The new data-set can then be run without having to re-run all the tests which have been successful.

A test should be considered successful if the criteria on the comment lines are satisfied.

The verification data-set also illustrates several of the options within MOPAC, and may be useful in learning how data-sets are constructed.

Other data sets are included in the examples directory—these test various options, and are a useful set of templates for designing user specific data sets.

4.2 Excited states

In order to understand the calculations described in this section, users should have a good working knowledge of the theory of electronic states. The example data-set is called `meci.dat`. This job should be run. In this discussion, users will be assumed to have access to the output from `meci.dat`.

4.2.1 Discussion of output from `meci.dat`

The methyl radical

Although configuration interaction is primarily used for calculating excited states, the first calculation is in fact not an excited state, but a radical, the methyl radical, CH_3 . This system is used to illustrate a very simple configuration interaction calculation. The methyl radical has 7 electrons, 4 from carbon and 1 from each of the three hydrogen atoms. The lowest three M.O.s ($1a'_1$, at -28.6eV, and $1e'$ at -14.3eV) each have two electrons, followed by a half-filled M.O., ($1a''_2$ at -4.1eV) with only one electron, then there are three empty levels ($2e'$ at 5.0eV and $2a'_1$ at 5.3eV).

Only one M.O., the $1a''_2$, is involved in the C.I. In this calculation, the energy due to the half-electron approximation is removed, and the energy of the pure microstate, in which the odd electron is given α spin, is calculated. Since only one configuration is involved, this new energy is the energy of the pure state *relative to the energy of the system calculated using half an α and half a β electron*. This amounts to -2.77eV. The energy of a methyl radical is then equal to the SCF energy, 89.7 kcal/mol, plus the C.I. energy, -2.77eV (= -63.9 kcal/mol) = 25.8 kcal/mol.

The symmetry of the state is $^2A''_2$.

Oxygen

Molecular oxygen contains 12 valence electrons, with two unpaired electrons in a π_g M.O. In order to preserve symmetry, $D_{\infty h}$, the SCF must be carried out with one electron in each of the two degenerate π_g M.O.s. To allow for this, `OPEN(2,2)` is used. The energy of the ground-state is -6.15eV relative to the SCF energy of 126.4 kcal/mol. This means that the predicted energy of $^3\Sigma_g^-$ molecular oxygen is -15.4 kcal/mol.

Because `TRIPLET` was specified, the microstate chosen has an $M_S=1$; in other words, both electrons in the active space were of α spin. As a result, the spin density can be calculated using `ESR`, and this gives exactly 1.0 unpaired electrons in the $p - \pi$ atomic orbitals.

When specifying `TRIPLET`, information regarding the other states is not calculated. These other states can be calculated by simply eliminating `TRIPLET`. When this is done, we see that the triplet state is, indeed, the ground state, but now two other states, excited states, are also generated. These are the $^1\Delta_g$ and $^1\Sigma_g$ states at 0.79eV and 1.59eV above the ground state, respectively.

The choice of M_S can be over-ridden if `M_S` is used. By specifying $M_S=0$ and `TRIPLET`, all the states are generated, although only the triplet state would be used in determining which state was to be selected.

Methane

In this calculation, methane is specified, and an `OPEN(6,3)` is used. Normally, `OPEN(n1,n2)` implies a C.I. calculation, but since the result is obvious (A C.I. on a simple closed-shell system involving only one microstate does exactly nothing.), C.I. is not used.

Methane Cation

Now, with 5 electrons in a three-fold degenerate t_2 M.O., three microstates are involved in the C.I. These microstates form a degenerate 2T_2 state.

In earlier MOPACs, systems with degenerate states could not be optimized. Such systems would undergo Jahn-Teller distortion, and the degeneracy would be lost. Starting with MOPAC 93, such states could be optimized (see Section 5.4.9). Several other open shell calculations are given in `meci.dat`; these illustrate various types of configuration interaction calculations.

```

Line 1: MOPAC SYMMETRY
Line 2: Formaldehyde, for Demonstration Purposes
Line 3:
Line 4: 0
Line 5: C 1.2 1
Line 6: H 1.1 1 120 1
Line 7: H 1.1 0 120 0 180 0 2 1 3
Line 8:
Line 9: 3 1 4
Line 10: 3 2 4
Line 11:

```

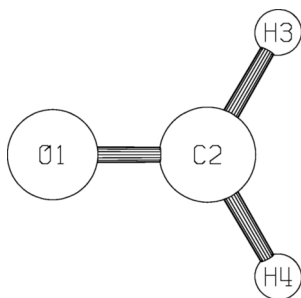


Figure 4.1: Example of a Simple Geometry Optimization

4.3 Simple geometry optimization

The data-set for a simple system is shown in Figure 4.1. The old MOPAC format is used here, so keyword `MOPAC` must be used, otherwise the first hydrogen atom would be positioned at Cartesian coordinate (1.1,120.0,0).

These data could be more neatly written as shown in Figure 4.2. Now the new coordinate convention is used—the position of the first hydrogen is defined in terms of atoms 1 and 2.

These two data-files will produce identical results files.

4.4 Highly excited methane dication

In this calculation we see that the vibrational frequencies of a highly excited state can be easily calculated.

```

Line 1: SYMMETRY
Line 2: Formaldehyde, for Demonstration Purposes
Line 3:
Line 4: 0
Line 5: C 1.20 1 0.00 0 0.00 0 1
Line 6: H 1.10 1 120.00 1 0.00 0 2 1
Line 7: H 1.10 0 120.00 0 180.00 0 2 1 3
Line 8:
Line 9: 3, 1, 4,
Line 10: 3, 2, 4,
Line 11:

```

Figure 4.2: Example of a Simple Geometry Optimization (2)

```

MOPAC CHARGE=2 OPEN(4,3) SINGLET ROOT=6 FORCE
Methane RHF dication, 6th singlet state (totally symmetric)
Check that vibrational frequencies are 1338(T2) 2482(A1) 2551(E) 4616(T2) +/-1
H
C 1.2298156
H 1.2298156 0 109.471221
H 1.2298156 0 109.471221 0 -120.000000 0 2 1 3
H 1.2298156 0 109.471221 0 120.000000 0 2 1 3

```

Figure 4.3: Example of an Electronic Excited State

The system being studied is CH_4^{++} , in the A_1 electronic state. The t_2 level, which normally holds six electrons, has lost two, and therefore has an occupancy of 4/3 electrons per M.O. These 4 electrons in 3 levels give rise to 9 microstates. The 6th singlet state is non-degenerate, and is the one chosen for the FORCE calculation. This completes the description of the calculation. All that remains is to generate the data-set; this is presented in Figure 4.3.

4.5 Polytetrahydrofuran

The data set shown in Figure 4.4 illustrates the data file for a polytetrahydrofuran calculation. As you can see the layout of the data is almost the same as that for a molecule, the main difference being the presence of the translation vector atom "Tv". This data set would allow calculation of all polymer properties *except* band structure and density of states. The reason why these properties could not be calculated is that the polymer units are not correctly specified. In order for the band-structure to be calculated, the order in which each atom in every mer occurs must be exactly the same. An example of such a numbering system is given for polyethylene, Figure 4.5.

Polytetrahydrofuran has a repeat unit of $(\text{C}_4\text{H}_8\text{O})_2$: i.e., twice the monomer unit. This is necessary in order to allow the lattice to repeat after a translation through 12.3 Å. See Section 5.9 on Solid State Capability for further details.

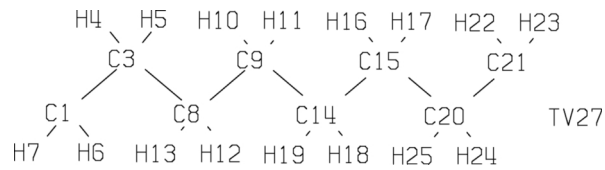
Note the two dummy atoms on lines 4k and 4l. These are useful, but not essential, for defining the geometry. The atoms on lines 4y to 4B use these dummy atoms, as does the translation vector on line 4C. The translation vector has only the length marked for optimization.

4.6 Diamond

Data sets for solids are very complicated. The full data set for diamond is quite large. Part of this data set is shown in Figure 4.6.

```
Line 1 : T=4H F
Line 2 : POLY-TETRAHYDROFURAN (C4 H8 O)2
Line 3 :
Line 4a: C 0.000000 0 0.000000 0 0.000000 0 0 0 0
Line 4b: C 1.551261 1 0.000000 0 0.000000 0 1 0 0
Line 4c: O 1.401861 1 108.919034 1 0.000000 0 2 1 0
Line 4d: C 1.401958 1 119.302489 1 179.392581 1 3 2 1
Line 4e: C 1.551074 1 108.956238 1 179.014664 1 4 3 2
Line 4f: C 1.541928 1 113.074843 1 179.724877 1 5 4 3
Line 4g: C 1.551502 1 113.039652 1 179.525806 1 6 5 4
Line 4h: O 1.402677 1 108.663575 1 179.855864 1 7 6 5
Line 4i: C 1.402671 1 119.250433 1 -179.637345 1 8 7 6
Line 4j: C 1.552020 1 108.665746 1 -179.161900 1 9 8 7
Line 4k: XX 1.552507 1 112.659354 1 -178.914985 1 10 9 8
Line 4l: XX 1.547723 1 113.375266 1 -179.924995 1 1110 9
Line 4m: H 1.114250 1 89.824605 1 126.911018 1 1 3 2
Line 4n: H 1.114708 1 89.909148 1 -126.650667 1 1 3 2
Line 4o: H 1.123297 1 93.602831 1 127.182594 1 2 4 3
Line 4p: H 1.123640 1 93.853406 1 -126.320187 1 2 4 3
Line 4q: H 1.123549 1 90.682924 1 126.763659 1 4 6 5
Line 4r: H 1.123417 1 90.679889 1 -127.033695 1 4 6 5
Line 4s: H 1.114352 1 90.239157 1 126.447043 1 5 7 6
Line 4t: H 1.114462 1 89.842852 1 -127.140168 1 5 7 6
Line 4u: H 1.114340 1 89.831790 1 126.653999 1 6 8 7
Line 4v: H 1.114433 1 89.753913 1 -126.926618 1 6 8 7
Line 4w: H 1.123126 1 93.644744 1 127.030541 1 7 9 8
Line 4x: H 1.123225 1 93.880969 1 -126.380511 1 7 9 8
Line 4y: H 1.123328 1 90.261019 1 127.815464 1 91110
Line 4z: H 1.123227 1 91.051403 1 -125.914234 1 91110
Line 4A: H 1.113970 1 90.374545 1 126.799259 1 101211
Line 4B: H 1.114347 1 90.255788 1 -126.709810 1 101211
Line 4C: Tv 12.299490 1 0.000000 0 0.000000 0 11110
Line 5 : O 0.000000 0 0.000000 0 0.000000 0 0 0 0
```

Figure 4.4: Example of a One-Dimensional Polymer



SYMMETRY MERS=4
Polyethylene

C	0.00	0	0.0	0	0.0	0	0	0
XX	7.73	1	0.0	0	0.0	0	1	0
C	1.50	1	0.0	0	0.0	0	1	2
H	1.10	1	109.0	1	0.0	1	3	1
H	1.10	0	109.0	0	120.0	1	3	1
H	1.10	0	109.0	0	180.0	1	1	2
H	1.10	0	109.0	0	60.0	1	1	2
C	1.54	1	110.2	1	60.0	0	3	1
C	1.50	0	110.2	0	180.0	0	8	3
H	1.10	0	109.0	0	-60.0	0	9	8
H	1.10	0	109.0	0	60.0	0	9	8
H	1.10	0	109.3	1	-60.1	1	8	3
H	1.10	0	109.3	0	60.1	0	8	3
C	1.54	0	110.2	0	179.9	0	9	8
C	1.50	0	110.2	0	180.0	0	14	9
H	1.10	0	109.0	0	-60.0	0	15	14
H	1.10	0	109.0	0	60.0	0	15	14
H	1.10	0	109.3	0	-60.1	0	14	9
H	1.10	0	109.3	0	60.1	0	14	9
C	1.54	0	110.2	0	180.0	0	15	14
C	1.50	0	110.2	0	180.0	0	20	15
H	1.10	0	109.0	0	-60.0	0	21	20
H	1.10	0	109.0	0	60.0	0	21	20
H	1.10	0	109.3	0	-60.1	0	20	15
H	1.10	0	109.3	0	60.1	0	20	15
XX	1.54	0	110.2	0	179.9	0	21	20
Tv	10.00	1	0.0	0	0.0	0	1	26

```

3 1 9 15 21
4 1 5 6 7 10 11 12 13 16 17 18 19 22 23 24 25
4 2 5 6 7 10 11 16 17 22 23
6 3 9 15 20 21
6 14 14 26
7 3 8 11 17 23
7 14 10 16 22
8 1 14 20 26
8 2 9 14 15 20 21 26
12 2 13 18 19 24 25
12 3 18 24
12 14 13 19 25

```

Figure 4.5: Example of the use of MERS


```

MERS=(4,4,4) BCC PM3 1SCF GRADIENTS PL PRECISE
Diamond, 64 atoms

C 0.0000000 0 0.0000000 0 0.0000000 0 0 0 0
C 1.54478440 1 0.0000000 0 0.0000000 0 1 0 0
C 2.95803659 1 100.0249853 1 0.0000000 0 2 1 0
C 1.54478440 0 100.0249853 0 -179.9999981 1 3 2 1
C 1.54478419 0 109.4712159 1 -59.9999996 1 2 1 3
C 1.54478440 0 109.4712162 0 -179.9999988 0 5 2 1
C 1.54478417 0 109.4712193 0 120.0000002 1 4 3 2
C 1.54478440 0 109.4712186 0 180.0000000 0 7 4 3
C 2.52262210 1 90.0000000 1 -90.0000000 1 5 2 1
C 1.54478436 0 90.0000010 0 -109.4712164 1 9 5 2

```

(Many lines missing)

```

C 1.54478398 0 109.4712072 0 -179.9999979 0 38 37 18
C 1.54478440 0 109.4711954 0 179.9999991 0 57 38 37
C 1.54478338 0 109.4711915 0 179.9999966 0 40 39 20
C 1.54478440 0 109.4711797 0 179.9999988 0 59 40 39
C 1.54478408 0 109.4712137 0 180.0000000 0 42 41 22
C 1.54478463 0 109.4712020 0 179.9999991 0 61 42 41
C 1.54478352 0 109.4711940 0 179.9999991 0 44 43 24
C 1.54478440 0 109.4711782 0 179.9999985 0 63 44 43
XX 2.52262206 0 90.0000020 0 89.9999974 0 7 4 3
XX 2.52262208 0 90.0000018 0 90.0000001 0 29 10 9
XX 2.52262210 0 89.9999999 0 -90.0000008 0 53 34 33
Tv 7.13505280 1 0.0000000 0 0.0000000 0 1 65 2
Tv 7.13505280 0 0.0000000 0 0.0000000 0 1 66 2
Tv 7.13505280 0 0.0000000 0 0.0000000 0 1 67 2

```

This data set was generated by MAKPOL acting on make_diamond.dat. Contents of make_diamond.dat:

```

MERS=(4,4,4) BCC PM3 1SCF GRADIENTS PL PRECISE
Diamond, 64 atoms

c
xx 0.7723922 1
c 0.7723922 0 180 0
xx 1.0000000 0 54.73561 0 0 0 1 2 3
xx 1.0000000 0 54.73561 0 120 0 1 2 4
xx 1.0000000 0 54.73561 0 240 0 1 2 4
TV 1.7837632 1 0.0 0 0 0 1 4 2
TV 1.7837632 1 0.0 0 0 0 1 5 2
TV 1.7837632 1 0.0 0 0 0 1 6 2

```

Figure 4.6: Example of a Three-Dimensional Data-set

Chapter 5

Theory

5.1 Introduction

Most of the theory used in MOPAC is in the literature, so that in principle one could read and understand the algorithm. However, it is convenient to have the theory gathered together in one document. The theory given here is intended for users who (a) want to modify MOPAC, or (b) want to understand how MOPAC does what it does. All derivations can be followed with a little patience, and, as far as possible, “It can be shown that” jumps in logic have been avoided.

5.2 Semiempirical Theory

There are five distinct methods available within MOPAC: MINDO/3, MNDO, AM1, PM3, and MNDO-*d*. All are semiempirical, and have roughly the same structure. A complete knowledge of these methods is not necessary in order to use MOPAC; however, a superficial understanding of these methods and their relationship to *ab initio* methods is important for using MOPAC and particularly for interpreting the results.

The five methods within MOPAC have many features in common. They are all self-consistent field (SCF) methods, they take into account electrostatic repulsion and exchange stabilization, and, in them, all calculated integrals are evaluated by approximate means. Further, they all use a restricted basis set of one *s* orbital and three *p* orbitals (p_x , p_y , and p_z) per atom (except MNDO-*d*, which has five *d* orbitals in addition to the *s* – *p* basis set) and ignore overlap integrals in the secular equation. Thus, instead of solving

$$|H - ES| = 0,$$

the expression

$$|H - E| = 0,$$

in which *H* is the secular determinant, *S* is the overlap matrix, and *E* is the set of eigenvalues, is solved. These approximations considerably simplify quantum mechanical calculations on systems of chemical interest. As a result, larger systems can be studied. Computational methods are only models, and there is no advantage in rigorously solving Schrödinger’s equation for a large system if that system has had to be abbreviated in order to make the calculations tractable. Semiempirical methods are thus seen to be well balanced: they are accurate enough to have useful predictive powers, yet fast enough to allow large systems to be studied.

All five semiempirical methods contain sets of parameters. For MINDO/3 atomic and diatomic parameters exist, while MNDO, AM1, PM3, and MNDO-*d* use only single-atom parameters. Not all parameters are optimized for all methods; for example, in MINDO/3, MNDO and AM1 the two electron one center integrals are normally

Table 5.1: Parameters used in Semiempirical Methods

Parameter	Description	MINDO/3	MNDO	AM1	PM3
U_{ss} and U_{pp}	s and p atomic orbital one-electron one-center integrals	+	*	*	*
β_s and β_p	s and p atomic orbital one-electron two-center resonance integral terms		*	*	*
I_s	s atomic orbital ionization potential for two center resonance integral term	+			
I_p	p atomic orbital ionization potential for two center resonance integral term	+			
$\beta_{A,B}$	Diatomic two center one-electron resonance integral multiplier	*			
ξ_s	s -type Slater atomic orbital exponent	*	*	*	*
ξ_p	p -type Slater atomic orbital exponent	*	*	*	*
α_A	Atom A core-core repulsion term		*	*	*
α_{AB}	Atoms A and B core-core repulsion term	*			
G_{ss}	$s-s$ atomic orbital one center two electron repulsion integral	+	+	+	*
G_{sp}	$s-p$ atomic orbital one center two electron repulsion integral	+	+	+	*
G_{pp}	$p-p$ atomic orbital one center two electron repulsion integral	+	+	+	*
G_{p2}	$p-p'$ atomic orbital one center two electron repulsion integral	+	+	+	*
H_{sp}	$s-p$ atomic orbital one-center two-electron exchange integral	+	+	+	*
K_{n_A} or a_{n_A}	A Gaussian multiplier for n th Gaussian of atom A			*	*
L_{n_A} or b_{n_A}	A Gaussian exponent multiplier for n th Gaussian of atom A			*	*
M_{n_A} or c_{n_A}	A Radius of center of n th Gaussian of atom A .			*	*

taken from atomic spectra. In the list given in Table 5.1, parameters optimized for a given method are indicated by '*'. A '+' indicates that the value of the parameter was obtained from experiment (not optimized). Where neither symbol is given, the associated parameter is not used in that method.

All five semiempirical methods also use two experimentally determined constants per atom: the atomic mass of the most abundant isotope and the heat of atomization.

5.2.1 Approximations used in MNDO, MINDO/3, AM1, PM3, and MNDO- d

All five methods are similar enough to describe simultaneously. In the following discussion, applications specific to each method will be indicated in the text. This will allow easy comparison among the methods, a comparison which is not obvious from their names. MINDO/3 stands for Modified Intermediate Neglect of Differential Overlap, version 3. MNDO stands for Modified Neglect of Diatomic Overlap, and AM1 is Austin Model 1: these are the first two of the MNDO-type methods. PM3 is the Modified Neglect of Diatomic Overlap, Parametric Method Number 3, and MNDO- d , is really just MNDO with d orbitals.

Earlier methods, such as CNDO (Complete Neglect of Differential Overlap) ignored all terms involving two different atomic orbitals on one atom. Because of this, CNDO was an example of a ZDO (Zero Differential Over-

lap) method. (For a good introduction to CNDO, see "Approximate Molecular Orbital Theory", J. A. Pople, D. L. Beveridge, McGraw-Hill, New York, 1970. CNDO was first described in: J. A. Pople, D. P. Santry, and G. A. Segal, "Approximate Self-Consistent Molecular Orbital Theory I. Invariant Procedures", J. Chem. Phys., 1965, 43, S129-S135 (1965); J. A. Pople and G. A. Segal, "Approximate Self-Consistent Molecular Orbital Theory II. Calculations with Complete Neglect of Differential Overlap", J. Chem. Phys., 1965, 43, S136-S149 (1965); and J. A. Pople and G. A. Segal, "Approximate Self-Consistent Molecular Orbital Theory. III CNDO Results for AB₂ and AB₃ Systems", J. Chem. Phys., 1966, 44, 3289-3296 (1966).),

MNDO, AM1, PM3, and MNDO-d belong to the family of NDDO (Neglect of Diatomic Differential Overlap) methods. In these methods all terms arising from the overlap of two atomic orbitals which are on different centers or atoms are set to zero. As this is not the forum for developing the ideas of Hartree-Fock theory, the derivation of the Roothaan-Hall equations will be assumed, and our description of the methods will start with the final Roothaan [52]-Hall [53] equations.

Basic Roothaan-Hall Equations

Secular Equation

$$c_i|F - \epsilon_i S|c_i = 0$$

c_i : eigenvector; F : Fock matrix; ϵ_i : eigenvalue; S : overlap matrix. The total electronic energy of the system is given by:

$$E = 1/2P(H + F),$$

in which P : density matrix; H : one-electron matrix. The general Fock Matrix element is

$$F_{\mu\nu}^{\alpha} = H_{\mu\nu} + \sum_{\lambda} \sum_{\sigma} [P_{\lambda\sigma}^{\alpha+\beta} \langle \mu\nu | \lambda\sigma \rangle - P_{\lambda\sigma}^{\alpha} \langle \mu\lambda | \nu\sigma \rangle].$$

Or, spin-free

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda} \sum_{\sigma} [P_{\lambda\sigma} \langle \mu\nu | \lambda\sigma \rangle - \frac{1}{2} P_{\lambda\sigma} \langle \mu\lambda | \nu\sigma \rangle].$$

The methods all use a minimum basis set consisting of a maximum of one atomic orbital for each angular quantum number. The normal basis set for any atom consists of one s and three p orbitals (p_x , p_y and p_z).

Neglect of diatomic overlap integral

All overlap integrals arising from the overlap of two different atomic orbitals are neglected. This reduces the overlap matrix to a unit matrix. The secular equation thus reduces to:

$$c_i|F - \epsilon_i|c_i = 0.$$

In semiempirical theory the Coulson density matrix is used, e.g.:

$$P_{\lambda\sigma}^{\alpha} = \sum_i^{occ} c_{\lambda i}^{\alpha} c_{\sigma i}^{\alpha},$$

where the sum is over all occupied spin molecular orbitals. In RHF calculations, only the total density matrix is calculated:

$$P_{\lambda\sigma} = 2 \sum_i^{occ} c_{\lambda i} c_{\sigma i},$$

where the sum is over all occupied molecular orbitals.

When a system has more than half the available M.O.s, N , filled, it is computationally faster to calculate the positron electron equivalent:

$$P_{\mu\nu}^{\alpha} = 1 - \sum_{i=occ+1}^N c_{\lambda i}^{\alpha} c_{\sigma i}^{\alpha}$$

and

$$P_{\mu\nu} = 2 - 2 \sum_{i=occ+1}^N c_{\lambda i} c_{\sigma i}.$$

An important exception to this rule is the calculation of the one-electron two-center integral $H_{\mu\nu}$, which is approximated by:

$$H_{\mu\nu} = S_{\mu\nu} \frac{1}{2} (U_{\mu\mu} + U_{\nu\nu}),$$

where $S_{\mu\nu}$ is the overlap integral between atomic orbital φ_{μ} on an atom, and φ_{ν} on another atom, and the U values are atomic orbital constants, supplied as data.

Neglect of three and four center integrals

Continuing with the neglect of differential overlap, all two-electron integrals involving charge clouds arising from the overlap of two atomic orbitals on different centers are ignored. Since no rotation can convert a two center two-electron integral into a set of integrals involving three and four center terms, rotational invariance is not compromised by this approximation. Rotational invariance is present if the calculated observables [ΔH_f , Dipole, I.P., etc] are not dependent on the orientation of the system. The effects of this approximation on the Roothaan equations are as follows:

In the Fock matrix, if φ_{μ} and φ_{ν} are on different centers the NDDO matrix element $F_{\mu\nu}^{\alpha}$ reduces to

$$F_{\mu\nu}^{\alpha} = H_{\mu\nu} - \sum_{\lambda}^A \sum_{\sigma}^B P_{\lambda\sigma}^{\alpha} \langle \mu\lambda | \nu\sigma \rangle,$$

while the MINDO/3 matrix element becomes:

$$F_{\mu\nu}^{\alpha} = H_{\mu\nu} - P_{\mu\nu}^{\alpha} \langle \mu\nu | \mu\nu \rangle.$$

Equivalent expressions exist for $F_{\mu\nu}^{\beta}$ and $P_{\mu\nu}^{\beta}$. Thus no Coulombic terms are present in the two-center Fock matrix elements.

If φ_{μ} and φ_{ν} are different but on the same center, then, since a minimal basis set is being used, all integrals of the type $\langle \mu\nu | \lambda\sigma \rangle$ are zero by the orthogonality of the atomic orbitals unless $\mu = \nu$ and $\lambda = \sigma$, or $\mu = \lambda$ and $\nu = \sigma$. The off-diagonal one-center NDDO Fock matrix elements become:

$$F_{\mu\nu}^{\alpha} = H_{\mu\nu} + 2P_{\mu\nu}^{\alpha+\beta} \langle \mu\nu | \mu\nu \rangle - P_{\mu\nu}^{\alpha} (\langle \mu\nu | \mu\nu \rangle + \langle \mu\mu | \nu\nu \rangle),$$

while the MINDO/3 element becomes:

$$F_{\mu\nu}^{\alpha} - P_{\mu\nu}^{\alpha} \langle \mu\nu | \mu\nu \rangle.$$

If φ_{μ} is the same as φ_{ν} , then, because of the symmetry of the two-electron integrals, the diagonal NDDO Fock matrix elements reduce to:

$$F_{\mu\mu}^{\alpha} = H_{\mu\mu} + \sum_{\nu}^A (P_{\nu\nu}^{\alpha+\beta} \langle \mu\mu | \nu\nu \rangle - P_{\nu\nu}^{\alpha} \langle \mu\nu | \mu\nu \rangle) + \sum_B \sum_{\lambda}^B \sum_{\sigma}^B P_{\lambda\sigma}^{\alpha+\beta} \langle \mu\mu | \lambda\sigma \rangle,$$

and the corresponding MINDO/3 element becomes:

$$F_{\mu\mu}^{\alpha} = H_{\mu\mu} + \sum_{\nu}^A (P_{\nu\nu}^{\alpha+\beta} \langle \mu\mu | \nu\nu \rangle - P_{\nu\nu}^{\alpha} \langle \mu\nu | \mu\nu \rangle) + \sum_B \langle AA | BB \rangle + \sum_{\lambda}^B P_{\lambda\lambda}^{\alpha+\beta}.$$

One-center two-electron integrals

The MNDO, AM1, and MNDO-*d* one-center two-electron integrals are derived from experimental data on isolated atoms. Most were taken from Oleari's [54] work, but a few were obtained by optimization to fit molecular properties. The values of PM3 one-center two-electron integrals were optimized to reproduce experimental molecular properties.

For each atom there are a maximum of five one-center two-electron integrals. These are $\langle ss|ss \rangle$, $\langle ss|pp \rangle$, $\langle sp|sp \rangle$, $\langle pp|pp \rangle$, and $\langle pp|p'p' \rangle$, where p and p' are two different p -type atomic orbitals. In the original formulation [55] there was a sixth integral, $\langle pp'|pp' \rangle$, but it can be shown that this integral is related to two of the other integrals by:

$$\langle pp'|pp' \rangle = \frac{1}{2}(\langle pp|pp \rangle - \langle pp|p'p' \rangle).$$

Proof: If the molecular frame is rotated by 45° about the z axis the atomic bases mix thus:

$$\begin{aligned} R(45^\circ)p_x &= \frac{1}{\sqrt{2}}(p_x + p_y) \\ R(45^\circ)p_y &= \frac{1}{\sqrt{2}}(p_y - p_x) \end{aligned} \quad (5.1)$$

$$\begin{aligned} R(45^\circ) \langle p_x p_y | p_x p_y \rangle &= \frac{1}{4} \langle (p_x + p_y)(p_y - p_x) | (p_x + p_y)(p_y - p_x) \rangle \\ &= \frac{1}{4} (\langle p_x p_x | p_x p_x \rangle + \langle p_y p_y | p_y p_y \rangle - \langle p_x p_x | p_y p_y \rangle - \langle p_y p_y | p_x p_x \rangle) \end{aligned}$$

or

$$R(45^\circ) \langle p_x p_y | p_x p_y \rangle = \frac{1}{2} (\langle p_x p_x | p_x p_x \rangle - \langle p_x p_x | p_y p_y \rangle)$$

For convenience these five integrals are given the following names:

$$\begin{aligned} \langle ss|ss \rangle &= G_{ss} \\ \langle pp|pp \rangle &= G_{pp} \\ \langle sp|sp \rangle &= H_{sp} \\ \langle pp|pp \rangle &= G_{pp} \\ \langle pp|p'p' \rangle &= G_{p2} \end{aligned}$$

Using these definitions, the two-electron one-center contributions to the Fock matrix become:

$$\begin{aligned} F_{ss}^\alpha &: P_{ss}^\beta G_{ss} + (P_{p_x}^{\alpha+\beta} + P_{p_y}^{\alpha+\beta} + P_{p_z}^{\alpha+\beta}) G_{sp} - (P_{p_x}^\alpha + P_{p_y}^\alpha + P_{p_z}^\alpha) H_{sp} \\ F_{sp}^\alpha &: 2P_{sp}^{\alpha+\beta} H_{sp} - P_{sp}^\alpha (H_{sp} + G_{sp}) \\ F_{pp}^\alpha &: P_{ss}^{\alpha+\beta} G_{sp} - P_{ss}^\alpha H_{sp} + P_{pp}^\beta G_{pp} + (P_{p'p'}^{\alpha+\beta} + P_{p''p''}^{\alpha+\beta}) G_{p2} - \frac{1}{2} (P_{p'p'}^\alpha + P_{p''p''}^\alpha) (G_{pp} - G_{p2}) \\ F_{pp'}^\alpha &: P_{pp'}^{\alpha+\beta} (G_{pp} - G_{p2}) - \frac{1}{2} P_{pp'}^\alpha (G_{pp} + G_{p2}) \end{aligned}$$

These expressions are common to all methods.

NDDO two-electron two-center integrals

In a local diatomic frame there are 22 unique two-electron two-center integrals for each pair of heavy (non-hydrogen) atoms. These are shown in Table 5.2.

Each integral represents the energy of an electron density distribution (electron 1) arising from the product of the first two atomic orbitals interacting with the electron density distribution (electron 2), which in turn arises

Table 5.2: Two-Electron Two-Center Integrals (Local Frame)

1	$\langle ss ss \rangle$	12	$\langle sp_\sigma p_\pi p_\pi \rangle$
2	$\langle ss p_\pi p_\pi \rangle$	13	$\langle sp_\sigma p_\sigma p_\sigma \rangle$
3	$\langle ss p_\sigma p_\sigma \rangle$	14	$\langle ss sp_\sigma \rangle$
4	$\langle p_\pi p_\pi ss \rangle$	15	$\langle p_\pi p_\pi sp_\sigma \rangle$
5	$\langle p_\sigma p_\sigma ss \rangle$	16	$\langle p_\sigma p_\sigma sp_\sigma \rangle$
6	$\langle p_\pi p_\pi p_\pi p_\pi \rangle$	17	$\langle sp_\pi sp_\pi \rangle$
7	$\langle p_\pi p_\pi p'_\pi p'_\pi \rangle$	18	$\langle sp_\sigma sp_\sigma \rangle$
8	$\langle p_\pi p_\pi p_\sigma p_\sigma \rangle$	19	$\langle sp_\pi p_\pi p_\sigma \rangle$
9	$\langle p_\sigma p_\sigma p_\pi p_\pi \rangle$	20	$\langle p_\pi p_\sigma sp_\pi \rangle$
10	$\langle p_\sigma p_\sigma p_\sigma p_\sigma \rangle$	21	$\langle p_\pi p_\sigma p_\pi p_\sigma \rangle$
11	$\langle sp_\sigma ss \rangle$	22	$\langle p_\pi p'_\pi p_\pi p'_\pi \rangle$

Table 5.3: Types of Electron Density Distribution

Monopole (1 charge)	Unit negative charge centered on the nucleus
Dipole (2 charges)	+1/2 charge located at position (x,y,z), -1/2 charge located at position (-x,-y,-z)
Linear Quadrupole (3 charges)	+1/2 charge located at the nucleus, -1/4 charge at positions (x,y,z) and at (-x,-y,-z)
Square Quadrupole (4 charges)	Four charges of magnitude +1/4, -1/4, +1/4 and -1/4 forming a square centered on the nucleus.

from the product of the second two atomic orbitals. Only if the first two atomic orbitals are the same and the second two are the same will the interaction energy have to be positive, in which case the integral represents an electron-electron repulsion term. In all other cases the sign of the integral value may be positive or negative.

With the exception of integral 22, all the integrals can be calculated using different techniques without loss of rotational invariance. That is, no integral depends on the value of another integral, except for number 22. As with the H_{pp} monocentric integral, it is easy to show that:

$$\langle p_\pi p'_\pi|p_\pi p'_\pi \rangle = \frac{1}{2}(\langle p_\pi p_\pi|p_\pi p_\pi \rangle - \langle p_\pi p_\pi|p'_\pi p'_\pi \rangle).$$

The electron density distributions are approximated by a series of point charges. There are four possible types of distribution. These are given in Table 5.3.

These are used to represent the four types of atomic orbital products (Table 5.4).

Each two electron interaction integral is the sum of all the interactions arising from the charge distribution representing one pair of atomic orbitals with the charge distribution representing the second pair of atomic orbitals. Thus, in the simplest case, the $\langle ss|ss \rangle$ interaction is represented by the repulsion of two monopoles, while a $\langle p_\pi p_\pi|p'_\pi p'_\pi \rangle$, a much more complicated interaction, is represented by 16 separate terms, arising from the four charges representing the monopole and linear quadrupole on one center interacting with the equivalent set on the second center.

While the repulsion of two like charges is proportional to the inverse distance separating the charges, boundary conditions preclude using this simple expression to represent the interelectronic interactions. Instead, the

Table 5.4: Density Distributions Arising from Pairs of Atomic Orbitals

Atomic Orbitals	Multipole Distribution	Number of Charges
$\langle ss $	Monopole	1
$\langle sp $	Dipole	2
$\langle pp $	Monopole plus Linear Quadrupole	4
$\langle pp' $	Square Quadrupole	4

Table 5.5: Additive Terms for Two-Electron Integrals

Multipole	Monocentric Equivalent	Name
Monopole	G_{ss}	AM
Dipole	H_{sp}	AD
Quadrupole	$H_{pp} = 1/2(G_{pp} - G_{p2})$	AQ

interaction energy is approximated by:

$$\langle ss | ss \rangle = \frac{27.21}{\sqrt{(R + c_A + c_B)^2 + \frac{1}{4}(\frac{1}{G_A} + \frac{1}{G_B})^2}}$$

All that remains is to specify functional forms for the terms c and A . c , the distance of a multipole charge from its nucleus, is a simple function of the atomic orbitals; in the case of a $s - p$ product, this is a vector of length D_1 Bohr pointing along the p axis, where

$$D_1 = \frac{(2n + 1)(4\xi_s\xi_p)^{(n+1/2)}}{\sqrt{3}(\xi_s + \xi_p)^{(2n+2)}}$$

The principal quantum number is always the same in these methods for s and p orbitals on any given atom.

The corresponding distances of the charges from the nucleus for the linear and square quadrupoles are $2D_2$ and $\sqrt{2}D_2$ Bohr, respectively, where

$$D_2 = \left(\frac{4n^2 + 6n + 2}{20} \right)^{1/2} \frac{1}{\xi_p}$$

Now that the distances of the charges from the nucleus have been defined, the upper boundary condition can be set. In the limit, when $R = 0.0$, the value of the two-electron integral should equal that of the corresponding monocentric integral. Three cases can be identified:

1. A monopole-monopole interaction, in which case the integral must converge on G_{ss} .
2. A dipole-dipole interaction, where the integral must converge on H_{sp} .
3. The quadrupole-quadrupole interaction where the integral must converge on H_{pp} .

For convenience, the G_A terms are given special names. These are given in Table 5.5. In practice, $\frac{1}{2}(G_{pp} - G_{p2})$ is used instead of H_{pp} . This eliminates any possibility of loss of rotational invariance due to an incorrect value of H_{pp} .

While AM is given simply by $G_{ss}/27.21$, AD and AQ are complicated functions of one-center terms and the orbital exponents—recall that, in the limit, the associated charges are not all coincident. AD and AQ are solved iteratively. Given an initial estimate of AD of

$$AD = \left[\frac{H_{sp}}{27.21D_1^2} \right]^{1/3},$$

then, by iterating, an exact value of AD can be found. On iteration n the value of AD is given by

$$AD_n = AD_{n-2} + (AD_{n-1} - AD_{n-2}) \frac{\left(\frac{H_{sp}}{27.21} - a_{n-2}\right)}{a_{n-1} - a_{n-2}},$$

where

$$a_n = 2AD_n - 2(4D_1^2 + AD_n^{-2})^{-1/2}.$$

About 5 iterations are needed in order to get AD specified with acceptable accuracy.

Similarly, for AQ an initial estimate of $\left[\frac{H_{pp}}{27.21(3D_2^2)}\right]^{1/5}$ is made and, again, by iterating using

$$AQ_n = AQ_{n-2} + (AQ_{n-1} - AQ_{n-2}) \frac{\left(\frac{H_{pp}}{27.21} - a_{n-2}\right)}{a_{n-1} - a_{n-2}},$$

where, now,

$$a_n = 4AQ_n - 8(4D_2^2 + AQ_n^{-2})^{-1/2} + 4(8D_2^2 + AQ_n^{-2})^{-1/2},$$

an exact value of AQ can be obtained. About 5 iterations are necessary.

Note that these equations are intrinsically unstable on finite-precision computers. The denominator ($a_{n-1} - a_{n-2}$) rapidly becomes vanishingly small; this is, however, necessary in order to accurately define AD and AQ .

Final Assembly of Two-Electron Two-Center Integrals

With all the component parts defined, the two-electron two-center integrals are assembled from the sum of all the interactions of the charges on one center with those on the other center. The distance between the two charges must be determined—this is the vector addition of R , the interatomic distance in Bohr, and the two c terms defining the location of the charges from the nucleus—as well as the appropriate additive terms, AM , AD or AQ selected. Two examples will illustrate this assembly:

$\langle ss|ss \rangle$: This is represented by a single term. For monopoles, $c = 0$ and $G_A = AM_A$, $G_B = AM_B$ giving:

$$\langle ss|ss \rangle = \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AM_B}]^2 0.529177^2)^{1/2}}.$$

$\langle ss|p_\pi p_\pi \rangle$: $p_\pi p_\pi$ is expressed as the sum of a monopole and a linear quadrupole. This gives rise to a total of four charges, hence four terms. However, since the interaction of the monopole with each of the two negative charges of the dipole are the same, only three terms need to be evaluated. In general, symmetry considerations lower the total number of terms that need to be evaluated, so the maximum number in any integral is 8. The full integral is then represented as:

$$\begin{aligned} \langle ss|p_\pi p_\pi \rangle &= \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AM_B}]^2 0.529177^2)^{1/2}} \\ &+ \frac{\frac{1}{2} 14.399}{(R_{AB}^2 + (2D_2^B)^2 1/2[\frac{1}{AM_A} + \frac{1}{AQ_B}]^2 0.529177^2)^{1/2}} \\ &- \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AQ_B}]^2 0.529177^2)^{1/2}}. \end{aligned}$$

MINDO/3 Two electron two center integrals

In marked contrast to the other methods, MINDO/3 Coulomb and exchange integrals are given by the simple Dewar-Sabelli [56, 57] - Klopman [58] approximation. The integral is a function of the atom types and the interatomic distance only, and is of form

$$\langle AA|BB \rangle = \frac{14.399}{(R_{AB}^2 + \frac{1}{2}[\frac{14.399}{G_A} + \frac{14.399}{G_B}]^2)^{1/2}},$$

where G_A and G_B , sometimes called the Gamma-values, are the averages of the appropriate one-center two-electron integrals. All finite integrals over atomic orbitals on two centers are set equal. Thus, $\langle s_A s_A | s_B s_B \rangle = \langle s_A s_A | p_B p_B \rangle = \langle p_A p_A | p_B p_B \rangle = \langle AA | BB \rangle$.

The one-center one-electron integral $H_{\mu\mu}$

This represents the energy an electron in atomic orbital φ_μ would have if all electrons were removed from the system. This is approximated by adding on to the one-electron energy of the atomic orbital in the fully ionized atom the potential due to all the other nuclei in the system. The one-electron energy is obtained parametrically, and is given the symbol $U_{\mu\mu}$. $H_{\mu\mu}$ is derived from the fundamental equation

$$H_{\mu\mu} = U_{\mu\mu} - \sum_{B \neq A} Z_B \langle \mu\mu | BB \rangle$$

by equating the core-electron integral $\langle \mu\mu | BB \rangle$ to the corresponding two-electron integral, thus: (a) NDDO:

$$\langle \mu\mu | BB \rangle = Z_B \langle \mu\mu | ss \rangle$$

(b) MINDO/3:

$$\langle \mu\mu | BB \rangle = Z_B \langle AA | ss \rangle$$

The two-center one-electron integral $H_{\mu\nu}$

Sometimes called the resonance integral, $H_{\mu\nu}$ is approximated using the overlap integral, $S_{\mu\nu}$. Note that this violates the NDO approximation, but since resonance integrals are large, this integral is retained. This is the origin of Modified in the name of the method. With Slater atomic orbitals of type

$$\varphi_\mu = N r^{n-1} e^{-\xi r},$$

the overlap integral is given by:

$$S_{\mu\nu} = \langle \varphi_\mu \varphi_\nu \rangle.$$

Within NDDO, $H_{\mu\nu}$ is approximated by:

$$H_{\mu\nu} = S_{\mu\nu} \frac{1}{2} (\beta_\mu + \beta_\nu),$$

while MINDO/3 has a very different form:

$$H_{\mu\nu} = S_{\mu\nu} \beta_{AB} (I_\mu + I_\nu).$$

This use of a diatomic parameter is the most distinctive difference between the NDDO philosophy and that of MINDO/3. Because of the difficulty in parameterizing an element at the MINDO/3 level—the number of β -parameters rises as the square of the number of elements parameterized—it is unlikely that any further development of MINDO/3 will be made.

Core-core repulsion integrals

From simple electrostatics the core-core repulsion integral in eV is:

$$E_N(A, B) = \frac{14.399 Z_A Z_B}{R_{AB}}.$$

However, the electron-electron and electron-core integrals do not collapse to the form c/R_{AB} (R in Å) for distances beyond the van der Waal's radii. If the simple term given above is used, there would be a net repulsion between two neutral atoms or molecules. To correct for this the core-core repulsion is approximated by:

$$E_N(A, B) = Z_A Z_B \langle AA | BB \rangle.$$

MINDO/3 modification to the core-core term

In addition to this term, account must be taken of the decreasing screening of the nucleus by the electrons as the interatomic distance becomes very small. At very small distances the core-core term should approach the classical form. To account for this, an additional term is added to the basic core-core repulsion to give:

$$E_N(A, B) = Z_A Z_B (\langle AA|BB \rangle + (\frac{14.399}{R_{AB}} - \langle AA|BB \rangle) e^{-\alpha_{AB} R_{AB}}).$$

Here, α_{AB} is a diatomic parameter. In the case of N-H and O-H interactions only, this expression is replaced by:

$$E_N(A, B) = Z_A Z_B (\langle AA|BB \rangle + (\frac{14.399}{R_{AB}} - \langle AA|BB \rangle) \alpha_{AB} e^{-R_{AB}}).$$

MNDO modification to the core-core term

The MNDO approximation to the screening effect is similar to that of MINDO/3 in practice, but has a different functional form:

$$E_N(A, B) = Z_A Z_B (\langle s_A s_A | s_B s_B \rangle (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}})).$$

Again, O-H and N-H interactions are treated differently. For these interactions, use

$$E_N(A, B) = Z_A Z_B (\langle s_A s_A | s_B s_B \rangle (1 + e^{-\alpha_B R_{AB}} R_{AB} + e^{-\alpha_H R_{AB}})).$$

AM1 and PM3 modifications to the core-core term

These modifications are the same as that for MNDO with the addition of an extra term to reduce the excessive core-core repulsions just outside bonding distances. The additional term may be considered as a van der Waal's attraction term. The AM1 and PM3 core-core terms are:

$$E_N(A, B) = E_N^{MNDO}(A, B) + \frac{Z_A Z_B}{R_{AB}} (\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2}).$$

The extra terms define spherical Gaussian functions, the a , b , and c are adjustable parameters. PM3 has two Gaussians per atom, while AM1 has between two and four.

When I originally developed the theory and software for AM1, my purpose for having the Gaussian modification was to correct for the excessive repulsions in the region outside the normal bonding distance and less than the van der Waals distance. My reasoning was as follows: The excessive repulsion depends on distance only. This must be due to a fault in the approximations. It is not necessary to identify either the faulty approximation, or to modify it so as to correct the fault. All that is necessary is to add a new term, using the well-known method of operator equivalents. In this, I decided to treat the symptom of the fault, and not the cause. As subsequent events have shown, this has proved to be an effective approach. However, it is still intellectually unsatisfactory, in that the final justification is that "the end justifies the means." Hopefully, a better approximation than the *ad hoc* one I chose will eventually be developed.

5.2.2 Self-Consistent Field Calculation

Once all the integrals needed are calculated, the SCF calculation can be started. A trial density matrix is constructed; this is a diagonal matrix with the diagonal terms chosen so that every atom starts off electrically neutral (except for ions, when each atom is given an equal charge).

Using this trial density matrix, the one-electron matrix, and the two-electron integrals, a trial Fock matrix is constructed.

Diagonalization produces a set of eigenvectors, from which a better density matrix can be made.

This sequence (constructing the density matrix—constructing the Fock matrix—diagonalizing the Fock to get new eigenvectors) is repeated until the density matrix has become self-consistent to within a pre-set limit (SELCON).

5.2.3 Calculation of ΔH_f

The SCF calculation produces a density, P , and Fock matrix, F . These, together with the one-electron matrix, H , allow the total electronic energy to be calculated via

$$E_{elect} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}).$$

The total core-core repulsion energy is given by:

$$E_{nuc} = \sum_A \sum_{B < A} E_N(A, B).$$

The addition of these two terms represents the energy released when the ionized atoms and valence electrons combine to form a molecule.

A more useful quantity is the heat of formation of the compound from its elements in their standard state. This is obtained when the energy required to ionize the valence electrons of the atoms involved (calculated using semiempirical parameters), $E_{isol}(A)$, and heat of atomization, $E_{atom}(A)$, are added to the electronic plus nuclear energy. This yields:

$$\Delta H_f = E_{elect} + E_{nuc} + \sum_A E_{isol}(A) + \sum_A E_{atom}(A).$$

This is the quantity which MOPAC calls the “Heat of Formation”. An alternative but equivalent definition of ΔH_f , more suited for comparison with experimental ΔH_f s, is:

“ ΔH_f is the calculated gas-phase heat of formation at 298K of one mole of a compound from its elements in their standard state.”

Things to note about this definition: unlike *ab initio* methods, which yield the energy at 0K, semiempirical methods give ΔH_f at 298K. This follows from the way in which semiempirical methods are parameterized: the reference ΔH_f are conventionally given at 298K. This means that semiempirical methods will reproduce ΔH_f for 298K. Secondly, note that ΔH_f are for gas-phase systems. To calculate ΔH_f in the liquid or solid phases, additional terms are necessary.

5.3 Miscellaneous Topics in Semiempirical Theory

5.3.1 Koopmans' Theorem.

Koopmans' theorem [59] can be understood as follows: for closed-shell systems, the negative of the HOMO energy is the ionization potential. That is, the energy required to form the cation *provided that the ionization process is adequately represented by the removal of an electron from an orbital without change in the wave-functions of the other electrons*.

MOZYME does not use eigenvectors (LMOs are not eigenvectors), therefore, Koopmans' theorem cannot be used unless eigenvectors are generated. To use Koopmans' theorem, add VECTORS and EIGEN to the keyword line. This will cause the eigenvectors to be generated, and from the eigenvalues the ionization potential can be calculated.

The only alternative way to calculate the I.P. is to calculate the ΔH_f of the parent species, then, without allowing the geometry to relax, calculate the ΔH_f of the ionized system. The difference in ΔH_f , in kcal.mol⁻¹, divided by 23.06, is the predicted I.P., in eV.

5.3.2 Dipole moments.

For neutral systems, the dipole moment is calculated from the atomic charges and the lone-pairs as

$$\mu_x = cC \sum_A Q_A x_A + cC a_o 2 \sum_A P(s - p_x)_A D_1(A) \quad (5.2)$$

$$\mu_y = cC \sum_A Q_A y_A + cC a_o 2 \sum_A P(s - p_y)_A D_1(A) \quad (5.3)$$

$$\mu_z = cC \sum_A Q_A z_A + cC a_o 2 \sum_A P(s - p_z)_A D_1(A) \quad (5.4)$$

$$\mu = \mu_x + \mu_y + \mu_z \quad (5.5)$$

Where c = speed of light, C = charge on the electron, and a_o = Bohr radius, or $cC = 2.99792458 \times 1.60217733 = 4.8032066$, and $cC a_o 2 = 2.99792458 \times 1.60217733 \times 0.529177249 \times 2.0 = 5.0834948$. $D_1(A)$ is defined in Table 5.5 (see p. 111).

Formally, the dipole moment for an ion is undefined; however, it is convenient to set up a 'working definition.' Consider a heteronuclear diatomic ion in a uniform electric field. The ion will accelerate. To compensate for this, it is convenient to consider the ion in an accelerating frame of reference. The ion will experience a torque which acts about the center of mass, in a manner similar to that of a polar molecule. This allows us to define the dipole of an ion as the dipole the system would exhibit while accelerating in a uniform electric field. To formalize this definition:

$$\mu_x = cC \sum_A Q_A (x_A - x_{cog}) + cC a_o 2 \sum_A P(s - p_x)_A D(A) \quad (5.6)$$

$$\mu_y = cC \sum_A Q_A (y_A - y_{cog}) + cC a_o 2 \sum_A P(s - p_y)_A D(A) \quad (5.7)$$

$$\mu_z = cC \sum_A Q_A (z_A - z_{cog}) + cC a_o 2 \sum_A P(s - p_z)_A D(A) \quad (5.8)$$

$$\mu = \mu_x + \mu_y + \mu_z, \quad (5.9)$$

where x_{cog} is the x -coordinate of the center of gravity of the system

$$x_{cog} = \sum_A M_A x_A, \quad (5.10)$$

and y_{cog} and z_{cog} have similar definitions. This general expression will work for all discrete species, charged and uncharged, and is rotation and position invariant.

5.3.3 Bond Orders

Three quantities can be derived [15] from the density matrix for use in discussing bonding. These are: atomic bond index, anisotropy, and bond order.

The density matrix, P , can be decomposed into sub-matrices representing atoms or interactions between atoms. The three quantities just mentioned can then be defined in terms of these sub-matrices.

Atomic bond index

A measure of the valency of an atom.

$$V_A = \sum_{\lambda \in A} 2P_{\lambda\lambda} - \sum_{\lambda \in A} \sum_{\sigma \in A} P_{\lambda\sigma}^2. \quad (5.11)$$

Typical valencies are: 1.0 for hydrogen, 2.0–2.4 for oxygen to 3.8–4.0 for carbon. The maximum valency of an atom is equal to the number of atomic orbitals, e.g. 1 or 4 (for a sp^3 system), or 9 (in MNDO- d). This maximum is only achieved when the orbital population is 1.00, and all off-diagonal terms on the atom are zero.

Anisotropy

A measure of the number of lone-pairs on an atom.

$$L_A = \sum_{\lambda \in A} \sum_{\sigma \in A} P_{\lambda\sigma}^2 - \sum_{l=0}^k \frac{1}{2l+1} \left(\sum_{\lambda=l^2+1}^{(l+1)^2} P_{\lambda\lambda} \right)^2. \quad (5.12)$$

Typical numbers of lone-pairs are: 0 in, for example, hydrogen and carbon, 1 for nitrogen in amines, and 2 in oxygen.

To see how this expression is derived, consider an atom having valence orbitals defined by angular quantum numbers $l = 0, 1, \dots, k$. For H, $k = 0$, for all other elements, $k = 1$. In order to have spherical symmetry, all orbitals in any shell must be equally occupied. In addition, since the product of any two different atomic orbitals is non-spherical, all off-diagonal density matrix terms on any one atom must be zero.

For a spherical atom having the atomic populations $s^p p^q d^r$, the valency would be:

$$V_A = p^2 + \frac{q^2}{3} + \frac{r^2}{5}, \quad (5.13)$$

or, in general:

$$V_A = \sum_{l=0}^k \frac{1}{2l+1} \left(\sum_{\lambda=l^2+1}^{(l+1)^2} P_{\lambda\lambda} \right)^2. \quad (5.14)$$

Since, in general, atoms are non-spherical, then:

$$V_A = \sum_{\lambda \in A} \sum_{\sigma \in A} P_{\lambda\sigma}^2. \quad (5.15)$$

The difference between equations 5.14 and 5.15 is a measure of how unspherical the atom is.

Bond order

A measure of the number of bonds between atoms in a compound.

$$B_{AB} = \sum_{\lambda \in A} \sum_{\sigma \in B} P_{\lambda\sigma}^2 \quad (5.16)$$

Typical bond-orders are: 1.0, e.g., C-C in ethane; 2.0, e.g., C=C in ethylene; 3.0, e.g., C≡C in acetylene. Bond orders of less than about 0.1–0.2 are indicative of “no bond”.

The ideas here are an extension of Wiberg's indices [60].

Given the normal semiempirical density matrix, it is easy to show that

$$P^2 = 2P, \quad (5.17)$$

from which it follows that

$$P_{\lambda\lambda} = 1/2 \sum_{\sigma} P_{\lambda\sigma}^2. \quad (5.18)$$

This is the starting point for the derivation of

$$V_A = \sum_{B \neq A} B_{AB}, \quad (5.19)$$

from which the above definitions follow.

5.3.4 Mulliken populations

By default, the density matrix printed is the Coulson matrix, which assumes that the atomic orbitals are orthogonalized.

If the assumption of orthogonality is not made, then the Mulliken density matrix can be constructed. To construct the Mulliken density matrix (also known as the Mulliken population analysis), the M.O.s must first be re-normalized, using the overlap matrix, S :

$$\psi'_i = \psi_i \times S^{-\frac{1}{2}}.$$

From these M.O.s, a Coulson population is carried out. The off diagonal terms are simply the Coulson terms multiplied by the overlap:

$$P'_{\lambda\sigma \neq \lambda} = S_{\lambda\sigma} 2 \sum_{i=1}^{occ} c_{\lambda i} c_{\sigma i},$$

while the on-diagonal terms are given by the Coulson terms, plus half the sum of the off-diagonal elements:

$$P'_{\lambda\lambda} = S_{\lambda\lambda} 2 \sum_{i=1}^{occ} c_{\lambda i} c_{\lambda i} + \frac{1}{2} \sum_{\sigma \neq \lambda} P'_{\lambda\sigma}.$$

A check of the correctness of the Mulliken populations is to add the diagonal terms: these should equal the number of electrons in the system.

Theory of Mulliken Populations

The NDDO methods (MNDO, AM1, PM3, and MNDO-*d*) all use Slater orbitals, but an implication of one of the approximations made, that $\sum (F_{\mu\nu} - E_i \delta_{\mu\nu}) C_{\nu i} = 0$, is that the conventional molecular orbitals are normalized to unity:

$$\psi_i = \sum_{\lambda} c_{\lambda i} \phi_{\lambda}$$

with

$$\langle \psi_i^2 \rangle = 1 = \sum_{\lambda} c_{\lambda i}^2$$

For example, for H_2 , the occupied M.O. is:

$$\psi_1 = \sqrt{\frac{1}{2}} (\phi_{H_1} + \phi_{H_2}),$$

and the unoccupied M.O. is:

$$\psi_2 = \sqrt{\frac{1}{2}} (\phi_{H_1} - \phi_{H_2}).$$

The diagonal of the density matrix is then constructed using the Coulson formula:

$$P_{1,1} = P_{2,2} = 2.0 \times \left(\sqrt{\frac{1}{2}} \right)^2 = 1.0.$$

The off-diagonal terms are constructed in the same way:

$$P_{1,2} = P_{2,1} = 2.0 \times \left(\sqrt{\frac{1}{2}} \right)^2 = 1.0.$$

If, instead of using $\sum (F_{\mu\nu} - E_i \delta_{\mu\nu}) C_{\nu i} = 0$, $\sum (F_{\mu\nu} - E_i) C_{\nu i} = 0$ is used, then the occupied and unoccupied M.O.s become:

$$\psi_1 = \sqrt{\frac{1}{2(1+S)}} (\phi_{H_1} + \phi_{H_2}),$$

and the unoccupied M.O. is:

$$\psi_2 = \sqrt{\frac{1}{2(1-S)}} (\phi_{H_1} - \phi_{H_2}).$$

where S is the overlap integral: $\int \phi_{H_1} \phi_{H_2} dv$.

In this case, the Coulson population would give

$$P = \begin{vmatrix} \frac{1}{1+S} & \frac{1}{1+S} \\ \frac{1}{1+S} & \frac{1}{1+S} \end{vmatrix}$$

From this we see that the Coulson representation is unsuitable for two reasons: first, the number of electrons in the system, represented by the diagonal terms, does not add to 2.0. Second, the off-diagonal terms, which should represent the number of electrons resulting from the overlap of the two atomic orbitals, becomes unity as the overlap *decreases*.

To correct for this, it is physically meaningful to multiply the matrix elements by the overlap. This gives:

$$P = \begin{vmatrix} \frac{1}{1+S} & \frac{S}{1+S} \\ \frac{S}{1+S} & \frac{1}{1+S} \end{vmatrix}$$

Now the off-diagonal terms accurately represent the number of electrons which are associated with the overlap electron density. The total number of electrons in the system is now correct: $\frac{1}{1+S}$ on atom 1, $\frac{1}{1+S}$ on atom 2, and $\frac{2S}{1+S}$ in the overlap region, giving a total of 2.0.

Although this representation is correct, it is potentially misleading, in that the diagonal terms do not add to the number of electrons. Mulliken reasoned that the electron density resulting from the overlaps should be divided into two equal parts and added to the diagonal terms. When that is done, we get:

$$P = \begin{vmatrix} \frac{1}{1+S} + \frac{S}{1+S} & \frac{S}{1+S} \\ \frac{S}{1+S} & \frac{1}{1+S} + \frac{S}{1+S} \end{vmatrix}$$

or

$$P = \begin{vmatrix} 1.0 & \frac{S}{1+S} \\ \frac{S}{1+S} & 1.0 \end{vmatrix}$$

This simple example can be extended to systems involving heteroatoms and to polyatomics, and is fully general.

The Mulliken analysis can be applied to semiempirical methods. To do this, it is necessary to first convert the M.O.s from solutions of $\sum(F_{\mu\nu} - E_i\delta_{\mu\nu})C_{\nu i} = 0$ to solutions of $\sum(F_{\mu\nu} - E_i)C_{\nu i} = 0$. The simplest way to do this is to take the conventional M.O.s and multiply them by $S^{-\frac{1}{2}}$. In the case of H_2 , the resulting M.O.s are exactly correct; in general, a small error is introduced. This error arises from the incomplete annihilation of the secular matrix elements, and is quite unimportant.

5.3.5 Localized orbitals

The molecular orbitals generated by diagonalization are normally delocalized over the system. By using a unitary transform of the occupied M.O.s, it is possible to generate a set of molecular orbitals which are localized on from one up to three centers.

These localized M.O.s are not eigenvectors of the Hamiltonian, nor are their energies eigenvalues. However, localized orbitals can be equated with the single, double, triple, and delocalized π -bonds of classical organic chemistry.

Eigenvectors

Conventional M.O.s are one-electron solutions to Schrödinger's equation, and can be studied by photoelectron and other methods. Such methods work on a timescale of about 10^{-17} s, roughly the time it takes light to cross a molecule. Normal chemical methods (study of reactions, etc.) yield little information about conventional M.O.s. Some information is obtained by inference, such as aromatic substitution directing considerations.

Localized M.O.s

In addition to being equivalent to the classical bonds, localized M.O.s are useful in understanding chemical reactions, and other phenomena which take place slowly relative to the speed of light: i.e., phenomena which take place in 10^{-12} s or slower.

Consider a S_{N^2} reaction. All the bonds forming and breaking can be seen in the localized M.O.s. Consider the reactive sites in a molecule (double bonds, lone pairs, etc.). These have exact equivalents in the localized M.O.s. The energies of localized M.O.s are indicative of the reactivity of the associated electron pair. Consider an excited, insulating polymer, such as excited polyethylene. The excited state is usually written with an asterisk (*), and is generated simply by localizing the M.O.s of an excited polymer cluster.

Localization Theory

Various methods of localizing M.O.s have been proposed [61, 62, 63]. The method described here is a modification of Von Niessen's technique, and is ideally suited for semiempirical methods.

For a set of LMOs,

$$\sum_i \langle \psi_i^4 \rangle$$

is a maximum. Since

$$\sum_i \sum_j \langle \psi_i^2 \rangle \langle \psi_j^2 \rangle$$

is a constant,

$$\sum_i \sum_{j < i} \langle \psi_i^2 \rangle \langle \psi_j^2 \rangle$$

must be a minimum.

The operation to localize M.O. consists of a series of binary unitary transforms of the type:

$$|\psi_i \rangle = a|\psi_k \rangle + b|\psi_l \rangle$$

$$|\psi_j \rangle = -b|\psi_k \rangle + a|\psi_l \rangle$$

where $|\psi_k \rangle$ and $|\psi_l \rangle$ are normal M.O.s, and $|\psi_i \rangle$ and $|\psi_j \rangle$ are the LMOs.

The ratio a/b is given by

$$a/b = \frac{1}{4} \arctan \left(\frac{4(\langle \psi_k \psi_l^3 \rangle - \langle \psi_k^3 \psi_l \rangle)}{\langle \psi_k^4 \rangle + \langle \psi_l^4 \rangle - 6 \langle \psi_k^2 \psi_l^2 \rangle} \right)$$

Note that in normal semiempirical work: $\langle \phi_\lambda | \phi_\sigma \rangle = \delta(\lambda, \sigma)$.

From this it follows that, given $\psi_k = \sum_\lambda C_{\lambda k} \phi_\lambda$,

$$\langle \psi_k \psi_l^3 \rangle = \sum_\lambda C_{\lambda k} C_{\lambda l}^3$$

In order to preserve rotational invariance, all contributions on each atom must be added together. This gives:

$$\begin{aligned} \langle \psi_k^4 \rangle &= \sum_A \left(\sum_{\lambda \in A} C_{\lambda k}^2 \right)^2 \\ \langle \psi_k^3 \psi_l \rangle &= \sum_A \left(\sum_{\lambda \in A} C_{\lambda k}^2 \right) \sum_{\lambda \in A} C_{\lambda k} C_{\lambda l} \\ \langle \psi_k^2 \psi_l^2 \rangle &= \sum_A \left(\sum_{\lambda \in A} C_{\lambda k}^2 \right) \left(\sum_{\lambda \in A} C_{\lambda l}^2 \right) \end{aligned}$$

5.3.6 Outer Valence Green's Function

This section is based on materials supplied by

Dr David Danovich
The Fritz Haber Research Center for Molecular Dynamics
The Hebrew University of Jerusalem
91904 Jerusalem
Israel

The OVGf technique was used with the self-energy part extended to include third order perturbation corrections, [34]. The higher order contributions were estimated by the renormalization procedure. The actual expression used to calculate the self-energy part, $\sum_{pp}(w)$, chosen in the diagonal form, is given in equation (5.20), where $\sum_{pp}^{(2)}(w)$ and $\sum_{pp}^{(3)}(w)$ are the second- and third-order corrections, and A is the screening factor accounting for all the contributions of higher orders.

$$\sum_{pp}(w) = \sum_{pp}^{(2)}(w) + (1 - A)^{-1} \sum_{pp}^{(3)}(w). \quad (5.20)$$

The particular expression which was used for the second-order corrections is given in equation (5.21).

$$\sum_{pp}^{(2)}(w) = \sum_a \sum_{i,j} \frac{(2V_{paij} - V_{paji})V_{paij}}{w + e_a - e_i - e_j} + \sum_{a,b} \sum_i \frac{(2V_{piab} - V_{piba})V_{piab}}{w + e_i - e_a - e_b}, \quad (5.21)$$

where

$$V_{pqrs} = \int \int \psi_p^*(1)\psi_q^*(2)(1/r_{12})\psi_r^*(1)\psi_s^*(2)d\tau_1d\tau_2.$$

In equation (5.21), i and j denote occupied orbitals, a and b denote virtual orbitals, p denotes orbitals of unspecified occupancy, and e denotes an orbital energy. The equations are solved by an iterative procedure, represented in equation (5.22).

$$w_p^{i+1} = e_p + \sum_{pp} (w^i). \quad (5.22)$$

The SCF energies and the corresponding integrals, which were calculated by one of the semiempirical methods (MNDO, AM1, or PM3), were taken as the zero'th approximation and all M.O.s may be included in the active space for the OVGf calculations.

The expressions used for $\sum_{pp}^{(3)}$ and A are given in [35].

The OVGf method itself, is described in detail in [34].

5.3.7 Using *ab initio* derivatives

AIDER will allow gradients to be defined for a system. MOPAC will calculate gradients, as usual, and will then use the supplied gradients to form an error function. This error function is (supplied gradients – initial calculated gradients), which is then added to the computed gradients, so that for the initial SCF, the apparent gradients will equal the supplied gradients.

A typical data-set using AIDER is shown in Figure 5.1.

Each supplied gradient goes with the corresponding internal coordinate. In the example given, the gradients came from a 3-21G calculation on the geometry shown. Symmetry will be taken into account automatically. Gaussian prints out gradients in atomic units; these need to be converted into kcal/mol/Ångstrom or kcal/mol/radian for MOPAC to use. The resulting geometry from the MOPAC run will be nearer to the optimized 3-21G geometry than if the normal geometry optimizers in Gaussian had been used.

5.3.8 Correction to the Peptide Linkage

The residues in peptides are joined together by peptide linkages, –HNCO–. These linkages are almost flat, and normally adopt a trans configuration, the hydrogen and oxygen atoms being on opposite sides of the C–N bond. Experimentally, the barrier to interconversion in N-methyl acetamide is about 14 kcal/mol, but all four methods within MOPAC predict a significantly lower barrier, PM3 giving the lowest value.

The low barrier can be traced to the tendency of semiempirical methods to give pyramidal nitrogens. The degree to which pyramidalization of the nitrogen atom is preferred can be seen in the series of compounds given in Table 5.6.

To correct this, a molecular-mechanics correction has been applied. This consists of identifying the –R–HNCO– unit, and adding a torsion potential of form:

$$k \sin^2\theta$$

where θ is the X–N–C–O angle, X=R or H, and k varies from method to method. This has two effects: there is a force constraining the nitrogen to be planar, and the HNCO barrier in N–methyl acetamide is raised to 14.00 kcal/mol. When the MM correction is in place, the nitrogen atom for all methods for the last three compounds shown above is planar. The correction should be user-transparent.

```
PM3 AIDER AIGOUT GNORM=0.01
Cyclohexane

X
X  1  1.0
C  1  CX    2  CXX
C  1  CX    2  CXX    3  120.000000
C  1  CX    2  CXX    3  120.000000
X  1  1.0    2  90.0    3  0.000000
X  1  1.0    6  90.0    2  180.000000
C  1  CX    7  CXX    3  180.000000
C  1  CX    7  CXX    3  60.000000
C  1  CX    7  CXX    3 -60.000000
H  3  H1C   1  H1CX   2  0.000000
H  4  H1C   1  H1CX   2  0.000000
H  5  H1C   1  H1CX   2  0.000000
H  8  H1C   1  H1CX   2  180.000000
H  9  H1C   1  H1CX   2  180.000000
H  10 H1C   1  H1CX   2  180.000000
H  3  H2C   1  H2CX   2  180.000000
H  4  H2C   1  H2CX   2  180.000000
H  5  H2C   1  H2CX   2  180.000000
H  8  H2C   1  H2CX   2  0.000000
H  9  H2C   1  H2CX   2  0.000000
H  10 H2C   1  H2CX   2  0.000000

CX    1.46613
H1C   1.10826
H2C   1.10684
CXX   80.83255
H1CX  103.17316
H2CX  150.96100

AIDER
  0.0000
 13.7589 -1.7383
 13.7589 -1.7383  0.0000
 13.7589 -1.7383  0.0000
  0.0000  0.0000  0.0000
  0.0000  0.0000  0.0000
 13.7589 -1.7383  0.0000
 13.7589 -1.7383  0.0000
 13.7589 -1.7383  0.0000
-17.8599 -2.1083  0.0000
-17.8599 -2.1083  0.0000
-17.8599 -2.1083  0.0000
-17.8599 -2.1083  0.0000
-17.8599 -2.1083  0.0000
-17.8599 -2.1083  0.0000
-17.5612 -0.6001  0.0000
-17.5612 -0.6001  0.0000
-17.5612 -0.6001  0.0000
-17.5612 -0.6001  0.0000
-17.5612 -0.6001  0.0000
```

Figure 5.1: Example of the use of AIDER

Table 5.6: Comparison of Observed and Calculated Pyramidalization of Nitrogen

Compound	MINDO/3	MNDO	AM1	PM3	Exp
Ammonia	Py	Py	Py	Py	Py
Aniline	Py	Py	Py	Py	Py
Formamide	Py	Py	Flat	Py	Py
Acetamide	Flat	Py	Flat	Py	Flat
N-methyl formamide	Flat	Py	Flat	Py	Flat
N-methyl acetamide	Flat	Flat	Flat	Py	Flat

Cautions

1. This correction will lead to errors of 0.5–1.5 kcal/mol if the peptide linkage is made or broken in a reaction calculation.
2. If the correction is applied to formamide the nitrogen will be flat, contrary to experiment.
3. When calculating rotation barriers, take into account the rapid rehybridization which occurs. When the dihedral is 0 or 180 degrees the nitrogen will be planar (sp^2), but at 90 degrees the nitrogen should be pyramidal, as the partial double bond is broken. At that geometry the true transition state involves motion of the nitrogen substituent so that the nitrogen in the transition state is more nearly sp^2 . In other words, a simple rotation of the HNCO dihedral will not yield the activation barrier, however it will be within 2 kcal/mol of the correct answer. The 14 kcal barrier mentioned earlier refers to the true transition state.

5.3.9 Convergence in SCF Calculation

A brief description of the convergence techniques used in subroutine ITER follows.

ITER, the SCF calculation, employs six methods to achieve a self-consistent field. In order of usage, these are:

1. Intrinsic convergence by virtue of the way the calculation is carried out. Thus a trial Fock gives rise to a trial density matrix, which in turn is used to generate a better Fock matrix.

This is normally convergent, but many exceptions are known. The main situations when the intrinsic convergence does not work are:

- (a) A bad starting density matrix. This normally occurs when the default starting density matrix is used. This is a very crude approximation, and is only used to get the calculation started. A large charge is generated on an atom in the first iteration, the second iteration overcompensates, and an oscillation is generated.
 - (b) The equations are only very slowly convergent. This can be due to a long-lived oscillation or to a slow transfer of charge.
2. Oscillation damping. If, on any two consecutive iterations, a density matrix element changes by more than 0.05, then the density matrix element is set equal to the old element shifted by 0.05 in the direction of the calculated element. Thus, if on iterations 3 and 4 a certain density matrix element was 0.55 and 0.78, respectively, then the element would be set to 0.60 (= 0.55 + 0.05) on iteration 4. The density matrix from iteration 4 would then be used in the construction of the next Fock matrix. The arrays which hold the old

density matrices are not filled until after iteration 2. For this reason they are not used in the damping before iteration 3.

3. Three-point interpolation of the density matrix. Subroutine CNVG monitors the number of iterations, and if this is exactly divisible by three, and certain other conditions relating to the density matrices are satisfied, a three-point interpolation is performed. This is the default converger, and is very effective with normally convergent calculations. It fails in certain systems, usually those where significant charge build-up is present.
4. Energy-level shift technique (the SHIFT technique). The virtual M.O. energy levels are normally shifted to more positive energy. This has the effect of damping oscillations, and intrinsically divergent equations can often be changed to intrinsically convergent form. With slowly-convergent systems the virtual M.O. energy levels can be moved to a more negative value.

The precise value of the shift used depends on the behavior of the iteration energy. If it is dropping, then the HOMO-LUMO gap is reduced; if the iteration energy rises, the gap is increased rapidly.

5. Pulay's method. If requested, when the largest change in density matrix elements on two consecutive iterations has dropped below 0.1, then routine CNVG is abandoned in favor of a multi-Fock matrix interpolation. This relies on the fact that the eigenvectors of the density and Fock matrices are identical at self-consistency, so $[P.F] = (P.F - F.P) = 0$ at SCF. The extent to which this condition does not occur is a measure of the deviance from self-consistency. Pulay's Direct Inversion of the Iterative Subspace (DIIS) method uses this relationship to calculate that linear combination of Fock matrices which minimize $[P.F]$. This new Fock matrix is then used in the SCF calculation.

Under certain circumstances, Pulay's method can cause very slow convergence, but sometimes it is the only way to achieve a self-consistent field. At other times the procedure gives a ten-fold increase in speed, so care must be exercised in its use. (started by the keyword PULAY)

6. The Camp-King converger. If all else fails, the Camp-King converger is just about guaranteed to work every time. However, it is time-consuming, and therefore should only be started as a last resort.

It evaluates that linear combination of old and current eigenvectors which minimize the total energy. One of its strengths is that systems which otherwise oscillate due to charge surges, e.g. CHO-H, the C-H distance being very large, will converge using this very sophisticated converger.

Causes of failure to achieve an SCF

In a system where a biradical can form, such as ethane decomposing into two CH_3 units, the normal RHF procedure can fail to go self-consistent. If the system has marked biradicaloid character, then BIRADICAL or UHF and TRIPLET can often prove successful. These options rely on the assumption that two unpaired electrons can represent the open shell part of the wave-function.

5.3.10 Use of C.I. in Reaction Path Calculations

Although closed-shell methods are suitable for normal systems, when a reaction occurs such that a bond makes or breaks, then configuration interaction can help in the description of the system.

Consider CH_2O , with the interatomic distance between carbon and one of the hydrogen atoms being steadily increased. At first the covalent bond will be strong, and a self-consistent field is readily obtained. Gradually the bond will become more ionic, and if configuration interaction is not used, a highly strained system will result. This exotic system will still have a large C-H bond order, despite the fact that the C-H distance is very large.

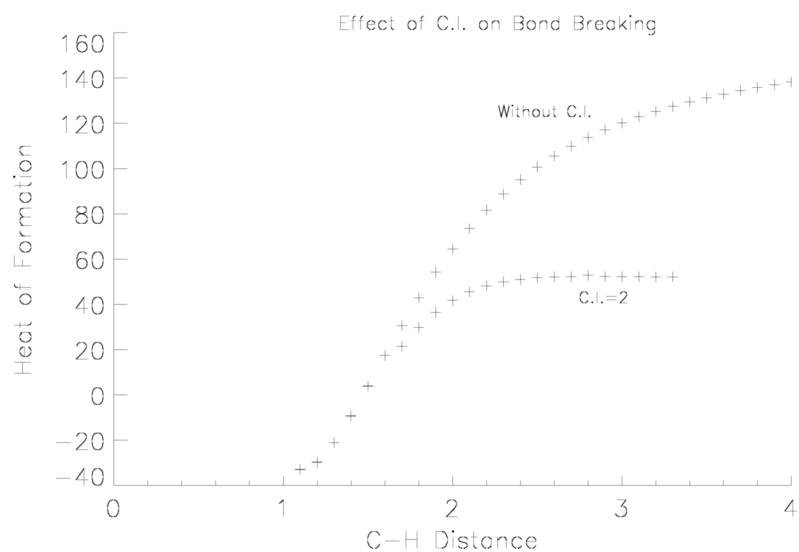


Figure 5.2: Effect of C.I. on ΔH_f in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)

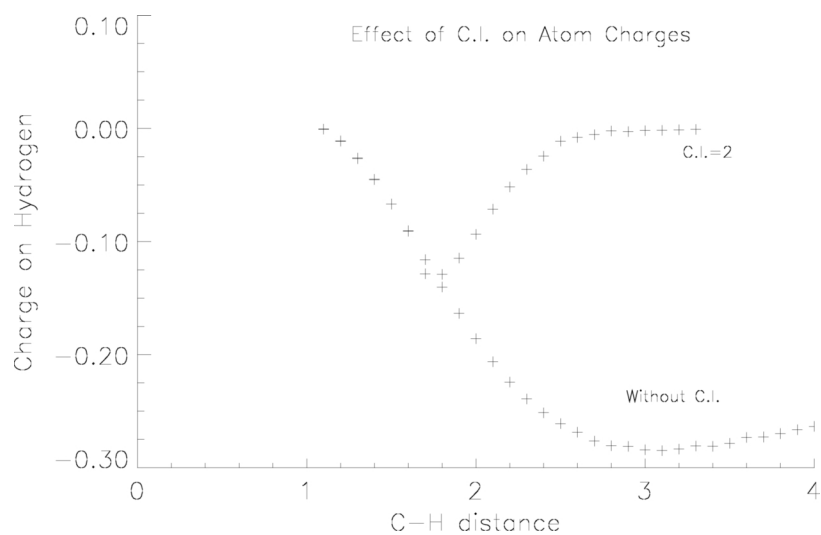


Figure 5.3: Effect of C.I. on the charge in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)

To a degree, configuration interaction can correct this picture. When C.I.=2 is used, Figure 5.2 and 5.3, a more realistic description of the dissociation is obtained. Now the leaving hydrogen atom becomes neutral as the distance increases, and the energy becomes constant at large distances. A C–H bond in formaldehyde is being stretched. The effect of C.I. is to make the dissociated state more realistic. Without C.I., the energy rises continuously, and the charge on the departing hydrogen atom becomes unrealistic.

5.3.11 Sparkles

Four extra “elements” have been put into MOPAC. These represent pure ionic charges, roughly equivalent to the following chemical entities:

Chemical Symbol	Equivalent to
+	Tetramethyl ammonium radical Potassium atom or Cesium atom.
++	Barium atom.
–	Borohydride radical, Halogen, or Nitrate radical
—	Sulfate, oxalate.

For the purposes of discussion these entities are called ‘sparkles’: the name arises from consideration of their behavior.

Behavior of sparkles in MOPAC

Sparkles have the following properties:

1. Their nuclear charge is integer, and is +1, +2, –1, or –2; there are an equivalent number of electrons to maintain electroneutrality, +1, +2, –1, and –2 respectively. For example, a ‘+’ sparkle consists of a unipositive nucleus and an electron. The electron is donated to the quantum mechanics calculation.
2. They all have an ionic radius of 0.7 Å. Any two sparkles of opposite sign will form an ion-pair with a interatomic separation of 1.4 Å.
3. They have a zero heat of atomization, no orbitals, and no ionization potential.
4. The associated one-center two-electron integral, G_{ss} is 27.21 for all sparkles. Because of this, the monopole-monopole interaction, AM , is set to 1.0. This is different to that in earlier (before MOPAC 6) MOPACs where the value of AM was set to 0.5D0.

They can be regarded as unpolarizable ions of diameter 1.4Å. They do not contribute to the orbital count, and cannot accept or donate electrons.

Since they appear as uncharged species which immediately ionize, attention should be given to the charge on the whole system. For example, if the alkaline metal salt of formic acid was run, the formula would be: HCOO+ where ‘+’ is the unipositive sparkle. The charge on the system would then be zero.

A water molecule polarized by a positive sparkle would have the formula H₂O⁺, and the charge on the system would be +1.

At first sight, a sparkle would appear to be too ionic to be a point charge and would combine with the first charge of opposite sign it encountered.

This representation is faulty, and a better description would be of an ion, of diameter 1.4Å, and the charge delocalized over its surface. Computationally, a sparkle is an integer charge at the center of a repulsion sphere having the exponential form $\exp(-\alpha r)$. The hardness of the sphere is such that other atoms or sparkles can approach within about 2Å quite easily, but only with great difficulty come closer than 1.4Å.

Uses of Sparkles

1. They can be used as counterions, e.g. for acid anions or for cations. Thus, if the ionic form of an acid is wanted, then the moieties $H\cdot X$, $H\cdot-$, and $+X$ could be examined.
2. Two sparkles of equal and opposite sign can form a dipole for mimicking solvation effects. Thus water could be surrounded by six dipoles to simulate the solvent cage. A dipole of value D can be made by using the two sparkles $+$ and $-$, or using $++$ and $--$. If $+$ and $-$ are used, the inter-sparkle separation would be $D/4.803\text{\AA}$. If $++$ and $--$ are used, the separation would be $D/9.606\text{\AA}$. If the inter-sparkle separation is less than 1.0\AA (a situation that cannot occur naturally) then the energy due to the dipole on its own is subtracted from the total energy.

5.3.12 Capped Bonds

Sometimes the system being studied is too large to be calculated using MOPAC. When only a part of the system is of interest, a mechanism exists to allow only that part to be calculated, and to ignore or not consider the rest of the system. Capped bonds are used to satisfy valency requirements.

The procedure for using capped bonds (Cb) is as follows:

1. Identify all atoms which are important to the calculation. For an enzyme, this would be the residues of the active site, for example.
2. Identify bonds which would be broken in order to isolate the atoms of interest. Make sure these are single bonds, and try to ensure that there is not more than one broken bond on any atom. Examples of "good" broken bonds are: CH_2-CH_2 , CH_2-NH , $NH-NH$.
3. Attach a Cb to each atom which has a broken bond. The Cb should be positioned in the direction of the atom which has deleted, and should have a bond-length of 1.7\AA , exactly. Do not mark the Cb bond length for optimization. If two broken bonds exist on an atom, use two Cb, but make the Cb-atom-Cb angle 109.471221° , exactly, and do not let it optimize.

A Cb behaves like a monovalent atom, but always has a zero charge. In other words, a Cb has a core charge of $+1$, and always has an electron population of 1.0 . Cbs have one orbital, and so can be regarded as being hydrogen-like.

Capped bonds are different from hydrogen atoms, however, in that they have a large β -value. The β value is used in the calculation of the one-electron two-center integral. Because the β value is so large, the difference in electronegativity of the Cb and the atom it is attached to, A , becomes negligible. Therefore the bonding M.O. consists of $1/\sqrt{2}(Cb + A)$. From this, it follows that the bonding M.O. contributes 1.0 electrons to the Cb.

In addition to the bonding M.O., there is an antibonding M.O. This M.O. is of form $1/\sqrt{2}(Cb - A)$, and is of very high energy.

To prevent capped bonds from forming bonds to all nearby atoms, overlaps from capped bonds to all atoms further away than 1.8\AA are set to zero. Because of this, the coupling between capped bonds and all atoms, other than A , is zero. Of course, A can interact with other atoms, once it has satisfied the demands of the attached capped bond.

Because of the huge β -value, the energy of a Cb-atom bond is very large. To prevent this interfering with the calculation, when the electronic energy is calculated, contributions from Cb are ignored. For this reason it is important that the bond-length for the Cb should not be optimized.

The M.O. energy levels due to Cb-type bonds are also enormous. Before the M.O. energy levels are printed in the normal output, energy levels arising from Cb-atom bonds are first set to zero.

The electronic behavior of capped bonds can easily be studied by use of `1SCF DENSITY VECTORS`.

5.3.13 Overlap Integrals

The particular technique used for the evaluation of the overlap integral depends on the atoms involved and whether analytic derivatives are used. All five semiempirical methods use Slater-type orbitals, STOs, although when analytic derivatives are involved [16], a Gaussian expansion [16] of STOs is normally used.

Specific expressions for various of the overlap integrals have appeared in the literature. These are normally used for those overlaps which involve only small principal quantum numbers, PQN, n , and a low angular quantum number, l . For the general case, however, in which any PQN may be encountered, the general overlap integral is used. As the final expression is rather ungainly, a simple derivation of the overlap integral will be given.

Slater atomic orbitals are of form

$$\varphi = \frac{(2\xi)^{n+1/2}}{(2n)!^{1/2}} r^{n-1} e^{-\xi r} Y_l^m(\theta, \phi),$$

where the $Y_l^m(\theta, \phi)$ are the normalized complex spherical harmonics. Complex spherical harmonics are chosen for convenience; $Y_l^m(\theta, \phi)$ real orbitals have a similar behavior, but require more manipulation. The θ dependence of spherical harmonics are the Laguerre polynomials, of form

$$Y_l^m(\theta, \phi) = \frac{e^{im\phi}}{(2\pi)^{1/2}} \left[\frac{(2l+1)(l-m)!}{2(l+m)!} \right]^{1/2} \frac{\sin^m \theta d^{l+m}(\cos^2 \theta - 1)^l}{2^l l! (d \cos \theta)^{l+m}}.$$

For convenience the phase factor is set to +1; this varies according to which source is used and the purpose for which the Laguerre polynomials are used.

Solving the differential gives

$$\frac{d^{l+m}}{(d \cos \theta)^{l+m}} (\cos^2 \theta - 1)^l = \sum_j \frac{l! (2j)! (-1)^{l+j}}{(l-j)! j! (2j-l-m)!} (\cos \theta)^{2j-l-m},$$

which, on rearranging to have the summation start at zero, becomes

$$\frac{d^{l+m}}{(d \cos \theta)^{l+m}} (\cos^2 \theta - 1)^l = \sum_{j=0}^{1/2(l-m)} \frac{l! (2(l-j))! (-1)^j}{j! (l-j)! (l-m-2j)!} (\cos \theta)^{l-m-2j}.$$

Substituting this into the STO yields

$$\begin{aligned} \varphi &= \frac{(2\xi)^{n+1/2}}{(2n)!^{1/2}} \left[\frac{(2l+1)(l-m)!}{2(l+m)!} \right]^{1/2} \left[\frac{\sin^m \theta}{2^l} r^{n-1} e^{-\xi r} \frac{e^{im\phi}}{(2\pi)^{1/2}} \right] \\ &\quad \sum_{j=0}^{1/2(l-m)} \frac{(2(l-j))! (-1)^j}{j! (l-j)! (l-m-2j)!} (\cos \theta)^{l-m-2j}. \end{aligned}$$

At this point it is convenient to collect some of the constants together; thus,

$$C_{lmj} = \left[\frac{(l-m)!}{(l+m)!} \right]^{1/2} \frac{(2(l-j))! (-1)^j}{2^l j! (l-j)! (l-m-2j)!},$$

which allows us to represent the STO in a considerably simplified form:

$$\varphi = \frac{(2\xi)^{n+1/2}}{(2n)!^{1/2}} \frac{(2l+1)^{1/2}}{2^{1/2}} \sin^m \theta r^{n-1} e^{-\xi r} \frac{e^{im\phi}}{(2\pi)^{1/2}} \sum_{j=0}^{1/2(l-m)} C_{lmj} (\cos \theta)^{l-m-2j}.$$

The overlap integral of two STOs can then be represented as

$$\langle \varphi_a \varphi_b \rangle = \frac{(2\xi_a)^{n_a+1/2} (2\xi_b)^{n_b+1/2} [(2l_a+1)(2l_b+1)]^{1/2}}{((2n_a)! (2n_b)!)^{1/2}} \frac{2}{2}$$

$$\int_0^\infty \sin^m \theta_a \sin^m \theta_b r_a^{n_a-1} r_b^{n_b-1} e^{-r_a \xi_a} e^{-r_b \xi_b} \frac{e^{im\phi} e^{im\phi^*}}{2\pi} \\ \sum_{j_a=0}^{1/2(l_a-m)} C_{i_a m j_a} (\cos \theta_a)^{l_a-m-2j_a} \sum_{j_b=0}^{1/2(l_b-m)} C_{i_b m j_b} (\cos \theta_b)^{l_b-m-2j_b} d\nu$$

It is impractical to solve this integral using polar coordinates. Instead, a prolate spheroidal coordinate system is used. Using the identities:

$$r_a = \frac{R(\mu+\nu)}{2}; \quad \cos \theta_a = \frac{(1+\mu\nu)}{(\mu+\nu)}; \quad \sin \theta_a = \frac{((\mu^2-1)(1-\nu^2))^{1/2}}{(\mu+\nu)} \\ r_b = \frac{R(\mu-\nu)}{2}; \quad \cos \theta_b = \frac{(1-\mu\nu)}{(\mu-\nu)}; \quad \sin \theta_b = \frac{((\mu^2-1)(1-\nu^2))^{1/2}}{(\mu-\nu)},$$

this gives $d\nu = \frac{R^3}{8} (\mu+\nu)(\mu-\nu) d\mu d\nu d\phi$.

Substituting these identities into the previous expression we get:

$$\langle \varphi_a \varphi_b \rangle = \int_0^{2\pi} \int_{-1}^1 \int_1^\infty \xi_a^{n_a+1/2} \xi_b^{n_b+1/2} \left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!} \right]^{1/2} \\ \frac{((\mu^2-1)(1-\nu^2))^m}{(\mu+\nu)^m (\mu-\nu)^m} \frac{R^{n_a-1}}{2^{n_a-1}} (\mu+\nu)^{n_a-1} \frac{R^{n_b-1}}{2^{n_b-1}} (\mu-\nu)^{n_b-1} \\ \frac{e^{-R\xi_a(\mu+\nu)/2} e^{-R\xi_b(\mu-\nu)/2}}{2\pi} \\ \sum_{j_a=0}^{(l_a-m)/2} \sum_{j_b=0}^{(l_b-m)/2} C_{l_a m j_a} C_{l_b m j_b} \frac{(1+\mu\nu)^{l_a-m-2j_a} (1-\mu\nu)^{l_b-m-2j_b}}{(\mu+\nu)^{l_a-m-2j_a} (\mu-\nu)^{l_b-m-2j_b}} \\ \frac{R^3}{8} (\mu+\nu)(\mu-\nu) d\mu d\nu d\phi,$$

which, on integrating over and rearranging, gives:

$$\langle \varphi_a \varphi_b \rangle = \int_{-1}^1 \int_1^\infty \frac{\xi_a^{n_a+1/2} \xi_b^{n_b+1/2}}{2} \left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!} \right]^{1/2} R^{n_a+n_b+1} \\ \sum_{j_a=0}^{(l_a-m)/2} \sum_{j_b=0}^{(l_b-m)/2} C_{l_a m j_a} C_{l_b m j_b} (\mu^2-1)^m (1-\nu^2)^m (\mu+\nu)^{n_a-l_a+2j_a} \\ (\mu-\nu)^{n_b-l_b+2j_b} (1+\mu\nu)^{l_a-m-2j_a} (1-\mu\nu)^{l_b-m-2j_b} e^{-R\xi_a(\mu+\nu)/2} e^{-R\xi_b(\mu-\nu)/2} d\mu d\nu.$$

This is a product of six simple expressions of type $(a+b)^n$. Expanding each term as a binomial generates six summations:

$$\langle \varphi_a \varphi_b \rangle = \int_{-1}^1 \int_1^\infty \frac{\xi_a^{n_a+1/2} \xi_b^{n_b+1/2}}{2} \left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!} \right]^{1/2} R^{n_a+n_b+1} \\ \sum_{j_a=0}^{(l_a-m)/2} \sum_{j_b=0}^{(l_b-m)/2} C_{l_a m j_a} C_{l_b m j_b} \sum_{k_a=0}^m \sum_{k_b=0}^m \sum_{P_a}^{n_a-l_a+2j_a} \sum_{P_b}^{n_b-l_b+2j_b} \sum_{q_a}^{l_a-m-2j_a} \sum_{q_b}^{l_b-m-2j_b} \\ \frac{(l_b-m-2j_b)!}{(l_b-m-2j_b-q_b)! q_b!} \frac{(l_a-m-2j_a)!}{(l_a-m-2j_a-q_a)! q_a!} \frac{(n_b-l_b+2j_b)!}{(n_b-l_b+2j_b-P_b)! P_b!} \\ \frac{(n_a-l_a+2j_a)!}{(n_a-l_a+2j_a-P_a)! P_a!} \frac{m!^2}{(m-k_b)! k_b! (m-k_a)! k_a!} \\ (-1)^{k_a+k_b+m+P_b+q_b} \nu^{2k_a+P_a+P_b+q_a+q_b} \mu^{2k_b+n_a-l_a+2j_a+n_b-l_b+2j_b-P_a-P_b+q_a+q_b} d\mu d\nu.$$

Using integration by parts, and making use of the following integrals (The "A" and "B" overlap integrals):

$$\int_1^\infty x^n e^{-ax} dx = e^{-a} \sum_{\mu=1}^{n+1} \frac{n!}{a^\mu (n-\mu+1)} = A_n(a)$$

Table 5.7: Values of C_{lmj}

l	m	j	C_{lmj}	l	m	j	C_{lmj}
0	0	0	1.0	3	0	0	5/2
1	0	0	1.0	3	1	0	$(225/48)^{1/2}$
1	1	0	$(1/2)^{1/2}$	3	2	0	$(15/8)^{1/2}$
2	0	0	3/2	3	3	0	$(5/16)^{1/2}$
2	1	0	$(3/2)^{1/2}$	3	0	1	-3/2
2	2	0	$(3/8)^{1/2}$	3	1	1	$-(3/16)^{1/2}$
2	0	1	-1/2				

Note: In subroutine SS the array $AFF(l,m,2j)$ corresponds to C_{lmj} here.

$$\int_{-1}^1 x^n e^{-ax} dx = -e^{-a} \sum_{\mu=1}^{n+1} \frac{n!}{a^\mu (n-\mu+1)} - e^a \sum_{\mu=1}^{n+1} \frac{n(-1)^{n-\mu}}{a^\mu (n-\mu+1)!} = B_n(a),$$

the overlap integral becomes

$$\begin{aligned} \langle \varphi_a \varphi_b \rangle &= \frac{\xi^{n_a+1/2} \xi^{n_b+1/2}}{2} \left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!} \right]^{1/2} R^{n_a+n_b+1} \\ &\sum_{j_a=0}^{(l_a-m)/2} C_{l_a m j_a} \sum_{j_b=0}^{(l_b-m)/2} C_{l_b m j_b} \sum_{k_a=0}^m \sum_{k_b=0}^m \sum_{P_a=0}^{n_a-l_a+2j_a} \sum_{P_b=0}^{n_b-l_b+2j_b} \sum_{q_a=0}^{l_a-m-2j_a} \sum_{q_b=0}^{l_b-m-2j_b} \\ &\frac{(l_b-m-2j_b)!}{(l_b-m-2j_b-q_b)!q_b!} \frac{(l_a-m-2j_a)!}{(l_a-m-2j_a-q_a)!q_a!} \frac{(n_b-l_b+2j_b)!}{(n_b-l_b+2j_b-P_b)!P_b!} \\ &\frac{(n_a-l_a+2j_a)!}{(n_a-l_a+2j_a-P_a)!P_a!} \frac{m!^2}{(m-k_b)!k_b!(m-k_a)!k_a!} (-1)^{k_a+k_b+m+P_b+q_b} \\ &B_{2k_a+P_a+P_b+q_a+q_b} \left(\frac{R(\xi_a - \xi_b)}{2} \right) A_{2k_b+n_a-l_a+2j_a+n_b-l_b+2j_b-P_a-P_b+q_a+q_b} \\ &\left(\frac{R(\xi_a + \xi_b)}{2} \right) d\mu d\nu, \end{aligned}$$

in which the coefficients C_{lmj} have the numerical values given in Table 5.7. which is the most convenient form for algorithmic use.

In this form, the overlap integral can be found in function SS.

Rotation of Atomic Orbitals

As calculated, the overlap integrals represent the overlap of atomic orbitals that are aligned along the z -axis. In general, this will not be the case, and the diatomic overlap integral matrix must be rotated in order to represent the actual orientation used.

The rotation matrices are well known, but the method by which they are constructed is by no means simple. Consider the general case of one atom at the origin, and a second atom at some point $p(x, y, z)$. Let the angle from the second atom to the first atom to z axis be b , and the angle of the projection of the second atom onto the xy plane to the y axis be a . Then, the p -orbital transform is as shown in Figure 5.8. The rotation matrices for the higher harmonics are fairly difficult to construct. In this work, several mathematical tools will be used. These are presented in Figure 5.4.

Consider the rotation of $d(x^2 - y^2)$. The rotation components involved are:

$$R(x) = (\cos a \cdot \cos b)x + (\sin a \cdot \cos b)y + (-\sin b)z,$$

Table 5.8: Angular Dependence of the p -orbitals

	$p(x)$	$p(y)$	$p(z)$
$\pi+$	$\cos a \cdot \cos b$	$-\sin a$	$\cos a \cdot \sin b$
$\pi-$	$\sin a \cdot \cos b$	$\cos a$	$\sin a \cdot \sin b$
σ	$-\sin b$	0	$\cos b$

1. All odd integrals, i.e., integrals of the type $\langle x^n \rangle$, where n is an odd integer, are zero.
2. $\langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle \neq 0$
3. $\langle x^2 y^2 \rangle = \langle y^2 x^2 \rangle = \langle y^2 z^2 \rangle = \langle z^2 x^2 \rangle \neq 0$
4. $\langle x^4 \rangle = \langle y^4 \rangle = \langle z^4 \rangle \neq 0$
5. All integrals of the type $\langle a^n b^m c^l \rangle$ where a, b , and c are different members of the set (x, y, z) are equal.
6. The normalization condition for the spherical harmonics is:

$$\int_0^{2\pi} \int_0^\pi Y_l^m(\theta, \phi)^2 \sin \theta d\theta d\phi = 1$$

7. $\cos^2 \theta + \sin^2 \theta = 1$
8. $\cos 2\theta = 2 \cos^2 \theta - 1$
9. $\sin 2\theta = 2 \sin \theta \cdot \cos \theta$

10. All atomic orbitals are assumed to be normalized:

$$\langle \psi^2 \rangle = 1$$

11. $x = r \sin \theta \cdot \cos \phi$, $y = r \sin \theta \cdot \sin \phi$, $z = r \cos \theta$.

When working with the angular components, it is normally easier to use the Cartesian symbols x , y , and z instead of the trigonometric forms. This avoids the potentially confusing expressions that involve the angles a , b , θ , and ϕ .

Figure 5.4: Mathematical Tools for use with Spherical Harmonics

and

$$R(y) = (-\sin a)x + (\cos a)y.$$

The rotation matrix element $\langle d(x^2 - y^2)|R|d(x^2 - y^2) \rangle$ can be derived using the relationship:

$$\begin{aligned} R|d(x^2 - y^2) \rangle &= (\cos^2 a \cdot \cos^2 b)x^2 + (2 \cos a \cdot \cos b \cdot \sin a \cdot \cos b)xy + (-2 \cos a \cdot \cos b \cdot \sin b)xz + \\ &(\sin^2 a \cdot \cos^2 b)y^2 + (-2 \sin a \cdot \cos b \cdot \sin b)yz + (\sin^2 b)z^2 + \\ &(-\sin^2 a)x^2 + (2 \sin a \cdot \cos a)xy + (-\cos^2 a)y^2. \end{aligned}$$

In this expression, the integral over all odd terms vanish, therefore:

$$\begin{aligned} \langle d(x^2 - y^2)|R|d(x^2 - y^2) \rangle &= \langle x^2(\cos^2 a \cdot \cos^2 b - \sin^2 a)x^2 \rangle + \\ &\langle x^2(\sin^2 a \cdot \cos^2 b - \cos^2 a)y^2 \rangle + \\ &\langle x^2(\sin^2 b)z^2 \rangle + \\ &\langle y^2(-\cos^2 a \cdot \cos^2 b + \sin^2 a)x^2 \rangle + \\ &\langle y^2(-\sin^2 a \cdot \cos^2 b + \cos^2 a)y^2 \rangle + \\ &\langle y^2(-\sin^2 b)z^2 \rangle. \end{aligned}$$

The integral $\langle x^4 \rangle$ has the same value as $\langle y^4 \rangle$, and $\langle x^2y^2 \rangle$ has the same value as $\langle x^2z^2 \rangle$, etc., therefore this expression can be simplified. Collecting together terms of the type $\langle x^4 \rangle$ gives

$$\cos^2 a \cdot \cos^2 b - \sin^2 a - \sin^2 a \cdot \cos^2 b + \cos^2 a,$$

or

$$(\cos^2 b + 1) \cdot (2 \cos^2 a - 1).$$

Collecting terms of the type $\langle y^2x^2 \rangle$ gives

$$\sin^2 a \cdot \cos^2 b - \cos^2 a + \sin^2 b - \cos^2 a \cdot \cos^2 b + \sin^2 a - \sin^2 b,$$

or

$$-(\cos^2 b + 1) \cdot (2 \cos^2 a - 1).$$

The rotation matrix element is thus:

$$\langle d(x^2 - y^2)|R|d(x^2 - y^2) \rangle = \langle x^4 - 2x^2y^2 + y^4 \rangle \frac{1}{2}(\cos^2 b + 1) \cdot (2 \cos^2 a - 1)$$

As a result of the fact that the angular components of the atomic orbitals are normalized, i.e., $\langle x^4 - 2x^2y^2 + y^4 \rangle = 1$:

$$\langle d(x^2 - y^2)|R|d(x^2 - y^2) \rangle = (2 \cos^2 a - 1) \cos^2 b + \frac{1}{2}(2 \cos^2 a - 1) \sin^2 b,$$

which is the form used in the calculation of the d -rotation matrix. The full set of d -orbital rotation matrix elements is presented in Table 5.9.

The f -orbitals are even more complicated. Unlike the d -orbitals, there are two complete and equivalent sets of f -orbitals. The set selected is that set which transforms as σ , π , δ , and ϕ when viewed along the z axis. The f -orbital transform is presented in Table 5.10 and 5.11.

Table 5.9: Angular Dependence of the d -orbitals

	$\delta+ \equiv d(x^2 - y^2)$
$d(x^2 - y^2)$	$(2 \cos^2 a - 1) \cos^2 b + \frac{1}{2}(2 \cos^2 a - 1) \sin^2 b$
$d(xz)$	$(2 \cos^2 a - 1) \sin b \cos b$
$d(2z^2 - x^2 - y^2)$	$\sqrt{\frac{3}{4}}(2 \cos^2 a - 1) \sin^2 b$
$d(yz)$	$-2 \sin a \cos a \sin b$
$d(xy)$	$-2 \sin a \cos a \cos b$
	$\pi+ \equiv d(xz)$
$d(x^2 - y^2)$	$-\cos a \sin b \cos b$
$d(xz)$	$\cos a (2 \cos^2 b - 1)$
$d(2z^2 - x^2 - y^2)$	$\sqrt{3} \cos a \sin b \cos b$
$d(yz)$	$-\sin a \cos b$
$d(xy)$	$\sin a \sin b$
	$\sigma \equiv d(2z^2 - x^2 - y^2)$
$d(x^2 - y^2)$	$\sqrt{\frac{3}{4}} \sin^2 b$
$d(xz)$	$-\sqrt{3} \sin b \cos b$
$d(2z^2 - x^2 - y^2)$	$\cos^2 b - \frac{1}{2} \sin^2 b$
$d(yz)$	0
$d(xy)$	0
	$\pi- \equiv d(yz)$
$d(x^2 - y^2)$	$-\sin a \sin b \cos b$
$d(xz)$	$\sin a (2 \cos^2 b - 1)$
$d(2z^2 - x^2 - y^2)$	$\sqrt{3} \sin a \sin b \cos b$
$d(yz)$	$\cos a \cos b$
$d(xy)$	$-\cos a \sin b$
	$\delta- \equiv d(xy)$
$d(x^2 - y^2)$	$2 \sin a \cos a \cos^2 b + \sin a \cos a \sin^2 b$
$d(xz)$	$2 \sin a \cos a \sin b \cos b$
$d(2z^2 - x^2 - y^2)$	$\sqrt{3} \sin a \cos a \sin^2 b$
$d(yz)$	$(2 \cos^2 a - 1) \sin b$
$d(xy)$	$(2 \cos^2 a - 1) \cos b$

Table 5.10: Angular Dependence of the f -orbitals

	$\sigma \equiv f(z^3) \equiv z(5z^2 - 3r^2)$
$f(z^3)$	$\frac{1}{2} \cos b \cdot (5 \cos^2 b - 3)$
$f(xz^2)$	$-\frac{\sqrt{6}}{4} \sin b \cdot (5 \cos^2 b - 1)$
$f(yz^2)$	0
$f(xyz)$	0
$f(z(x^2 - y^2))$	$\frac{\sqrt{15}}{2} \cos b \cdot \sin^2 b$
$f(x(x^2 - 3y^2))$	$-\frac{\sqrt{10}}{4} \sin^3 b$
$f(y(3x^2 - y^2))$	0
	$\pi+ \equiv f(xz^2) \equiv x(5z^2 - r^2)$
$f(z^3)$	$\frac{\sqrt{6}}{4} \cos a \cdot \sin b \cdot (5 \cos^2 b - 1)$
$f(xz^2)$	$\frac{1}{4} \cos a \cdot \cos b \cdot (15 \cos^2 b - 11)$
$f(yz^2)$	$-\frac{1}{4} \sin a \cdot (5 \cos^2 b - 1)$
$f(xyz)$	$\frac{\sqrt{10}}{2} \sin a \cdot \cos b \cdot \sin b$
$f(z(x^2 - y^2))$	$\frac{\sqrt{10}}{4} \cos a \cdot \sin b \cdot (1 - 3 \cos^2 b)$
$f(x(x^2 - 3y^2))$	$\frac{\sqrt{15}}{4} \cos a \cdot \cos b \cdot \sin^2 b$
$f(y(3x^2 - y^2))$	$-\frac{\sqrt{15}}{4} \sin a \cdot \sin^2 b$
	$\pi- \equiv f(yz^2) \equiv y(5z^2 - r^2)$
$f(z^3)$	$\frac{\sqrt{6}}{4} \sin a \cdot \sin b \cdot (5 \cos^2 b - 1)$
$f(xz^2)$	$\frac{1}{4} \sin a \cdot \cos b \cdot (15 \cos^2 b - 11)$
$f(yz^2)$	$\frac{1}{4} \cos a \cdot (5 \cos^2 b - 1)$
$f(xyz)$	$-\frac{\sqrt{10}}{2} \cos a \cdot \cos b \cdot \sin b$
$f(z(x^2 - y^2))$	$\frac{\sqrt{10}}{4} \sin a \cdot \sin b \cdot (1 - 3 \cos^2 b)$
$f(x(x^2 - 3y^2))$	$\frac{\sqrt{15}}{4} \sin a \cdot \cos b \cdot \sin^2 b$
$f(y(3x^2 - y^2))$	$\frac{\sqrt{15}}{4} \cos a \cdot \sin^2 b$

Table 5.11: Angular Dependence of the f -orbitals (cont.)

	$\delta+ \equiv f(xyz) \equiv xyz$
$f(z^3)$	$\sqrt{15} \sin a. \cos a. \cos b. \sin^2 b$
$f(xz^2)$	$\frac{\sqrt{10}}{2} \sin a. \cos a. \sin b. (3 \cos^2 b - 1)$
$f(yz^2)$	$\frac{\sqrt{10}}{2} (\cos^2 a - \sin^2 a) \sin b. \cos b$
$f(xyz)$	$(\cos^2 a - \sin^2 a)(\cos^2 b - \sin^2 b)$
$f(z(x^2 - y^2))$	$\sin a. \cos a. \cos b. (3 \cos^2 b - 1)$
$f(x(x^2 - 3y^2))$	$-\frac{\sqrt{6}}{2} \sin a. \cos a. \sin b. (1 + \cos^2 b)$
$f(y(3x^2 - y^2))$	$-\frac{\sqrt{6}}{2} (\cos^2 a - \sin^2 a) \sin b. \cos b$
	$\delta- \equiv f(z(x^2 - y^2)) \equiv z(x^2 - y^2)$
$f(z^3)$	$\frac{\sqrt{15}}{2} (\cos^2 a - \sin^2 a) \cos b. \sin^2 b$
$f(xz^2)$	$\frac{\sqrt{10}}{4} (\cos^2 a - \sin^2 a) \sin b. (3 \cos^2 b - 1)$
$f(yz^2)$	$-\sqrt{10} \sin a. \cos a. \sin b. \cos b$
$f(xyz)$	$-2 \sin a. \cos a. (\cos^2 b - \sin^2 b)$
$f(z(x^2 - y^2))$	$\frac{1}{2} (\cos^2 a - \sin^2 a) \cos b. (3 \cos^2 b - 1)$
$f(x(x^2 - 3y^2))$	$-\frac{\sqrt{6}}{4} (\cos^2 a - \sin^2 a) \sin b. (1 + \cos^2 b)$
$f(y(3x^2 - y^2))$	$\sqrt{6} \sin a. \cos a. \sin b. \cos b$
	$\phi+ \equiv f(x(x^2 - 3y^2)) \equiv x(x^2 - 3y^2)$
$f(z^3)$	$\frac{\sqrt{10}}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) \sin^3 b$
$f(xz^2)$	$\frac{\sqrt{15}}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) \cos b. \sin^2 b$
$f(yz^2)$	$-\frac{\sqrt{15}}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \sin^2 b$
$f(xyz)$	$-\frac{\sqrt{6}}{2} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \sin b. \cos b$
$f(z(x^2 - y^2))$	$\frac{\sqrt{6}}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) \sin b. (1 + \cos^2 b)$
$f(x(x^2 - 3y^2))$	$\frac{1}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) \cos b. (3 + \cos^2 b)$
$f(y(3x^2 - y^2))$	$-\frac{1}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) (1 + 3 \cos^2 b)$
	$\phi- \equiv f(y(3x^2 - y^2)) \equiv y(3x^2 - y^2)$
$f(z^3)$	$\frac{\sqrt{10}}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \sin^3 b$
$f(xz^2)$	$\frac{\sqrt{15}}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \cos b. \sin^2 b$
$f(yz^2)$	$\frac{\sqrt{15}}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) \sin^2 b$
$f(xyz)$	$\frac{\sqrt{6}}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) 2 \sin b. \cos b$
$f(z(x^2 - y^2))$	$\frac{\sqrt{6}}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \sin b. (1 + \cos^2 b)$
$f(x(x^2 - 3y^2))$	$\frac{1}{4} ((\cos^2 a - \sin^2 a) \sin a + 2 \sin a. \cos a. \cos a) \cos b. (3 + \cos^2 b)$
$f(y(3x^2 - y^2))$	$\frac{1}{4} ((\cos^2 a - \sin^2 a) \cos a - 2 \sin a. \cos a. \sin a) (1 + 3 \cos^2 b)$

5.3.14 Energies of Isolated Atoms

The ΔH_f calculated by semiempirical methods is defined as the energy in kcal.mol⁻¹ required to form one mole of the system in the gas phase at 298K from its elements in their standard state:

$$\Delta H_f = E_{elect} + E_{nuc} + \sum_A E_{isol}(A) + \sum_A E_{atom}(A)$$

In order to calculate ΔH_f , the quantity E_{isol} must be determined; this is the energy required to form the isolated atom from its valence electrons:

$$E_{isol}(A) = E_{neutral\ atom}(A) - E_{nucleus}(A) - E_{valence\ electrons}(A)$$

In the calculation of E_{elect} , the energy of valence electrons is defined as zero, likewise in calculating E_{nuc} , the energy of the isolated nucleus is defined as zero, therefore the calculation of E_{isol} simplifies to the calculation of $E_{neutral\ atom}$.

The energy of E_{isol} is the energy released when the valence electrons are added to the nucleus. For example, for the hydrogen atom, this would be U_{ss} . For poly-electronic atoms, the electron-electron interactions must be included, in addition to the one-electron contributions. Most elements have open shell ground states, and for these systems, the nature of the state is important.

For all main group elements, that is, elements with valence shell configurations of the form $ns^a np^b$, other than the alkali metals, the value of E_{isol} is given by:

$$E_{isol} = aU_{ss} + bU_{pp} + (a-1)G_{ss} + a.bG_{sp} + (b(b-1))/2G_{p2} - bH_{sp} - cH_{pp}$$

in which $c = \min(b(b-1)/2, (6-b).(5-b)/2)$. Except for the H_{pp} term, all the contributions to E_{isol} are obvious. Non-zero H_{pp} terms occur when there are two or more unpaired electrons in the ground state, in which case there is an exchange stabilization that is otherwise absent.

Because H_{pp} is usually written as $1/2(G_{pp} - G_{p2})$, the expression for systems with 2 to 4 p electrons is recast as:

$$E_{isol} = aU_{ss} + bU_{pp} + (a-1)G_{ss} + a.bG_{sp} + ((b(b-1))/2 + c/2)G_{p2} - (a-1)bH_{sp} - c/2G_{pp},$$

or

$$E_{isol} = 2U_{ss} + bU_{pp} + G_{ss} + 2.bG_{sp} + ((b(b-1))/2 + c/2)G_{p2} - (a-1)bH_{sp} - c/2G_{pp}.$$

For the alkali metals, the equation for E_{isol} is the same as that for hydrogen.

For the transition metals, the coefficients for the $d-d$ interactions are more complicated.

The general form for E_{isol} for a transition metal of configuration $s^m d^n s$, in which there are m_α α s -electrons and m_β β s -electrons, and n_α α d -electrons and n_β β d -electrons, and the total angular quantum number is L , is:

$$\begin{aligned} E_{isol} = & mU_{ss} + nU_{dd} + (m(m-1))/2G_{ss} + m.nG_{sd} - (m_\alpha n_\alpha + m_\beta n_\beta)H_{sd} \\ & + (n(n-1))/2 \frac{G_{dd}^0}{5} + (-4(n_\alpha^2 + n_\beta^2) + 13n - 3/2(L(L+1))) \frac{G_{dd}^2}{49} \\ & + (-(n_\alpha^2 + n_\beta^2)/2 - 9/2n + 5/6(L(L+1))) \frac{G_{dd}^4}{49}. \end{aligned}$$

As might be imagined, derivation of this expression is by no means obvious, particularly the terms for G_{dd}^2 and G_{dd}^4 . Interested readers are referred to Racah's paper in *Phys Rev*, **61**, 186 (1942). In this, Racah derived an expression for the d orbital energy of the ground state in terms of three quantities, A , B , and L , the total angular momentum:

$$\langle^{n+1}L|H|^{n+1}L\rangle = \frac{1}{2}n(n-1)(A-8B) + \frac{3}{2}[6n-L(L+1)]B.$$

The quantities A and B , and a third quantity, C , not used here, are related to the G_k as follows:

$$\begin{aligned} A &= G_{dd}^0 - 49G_{dd}^2 \\ B &= G_{dd}^2 - 5G_{dd}^4 \\ C &= 35G_{dd}^4 \end{aligned}$$

Using Racah's equation, derivation of E_{isol} is straightforward. In texts on transition metal ion theory, the quantities G_{dd}^0 , G_{dd}^2 , and G_{dd}^4 are usually represented by the symbols F_0 , F_2 , and F_4 , respectively. However, care should be exercised when reading these texts: sometimes other quantities, F^0 , F^2 , and F^4 are used. The relationship between these three sets of symbols is as follows:

$$\begin{aligned} G_{dd}^0 = F_0 &= F^0 \\ G_{dd}^2 = F_2 &= \frac{1}{49}F^2 \\ G_{dd}^4 = F_4 &= \frac{1}{441}F^4 \end{aligned}$$

Because the coefficients for the two electron terms are so complicated, values for all elements likely to be parameterized for semiempirical methods are presented in Table 5.12. From this table, the values of some coefficients are readily derived. Thus for the s - d coulomb integral, G_{sd} , the coefficient is simply the number of s electrons times the number of d electrons. One s - d exchange integral, H_{sd} , exists for each electron in the s shell for which there is an electron of the same spin in the d shell. For elements with two s electrons, this is simply the number of d electrons, for elements with one s electron, the Aufbau principle indicates that the d shell with higher occupancy has the same spin as that of the s electron. Finally, the coefficients for the simple d - d repulsion integral, G_{dd}^0 , are given by the number of possible d - d interactions.

Note also that there are no elements with both p and d valence electrons, therefore terms of the type G_{pd} are not necessary.

5.4 Multi-Electron Configuration Interaction

For some systems a single determinant is insufficient to describe the electronic wave function. For example, square cyclobutadiene and twisted ethylene require at least two configurations to describe their ground states. More than one configuration is also needed if an excited state is required—the RHF SCF converges on a ground state or, if half-electron methods are used, on a mixture of states, while the excited state involves a different configuration. Radicals also present a difficulty at the RHF level in that the SCF wavefunction corresponds to an equal mixture of the two doublets, with a corresponding error in the total energy. In order to correct for this error, MOPAC contains a very large Multi-Electron Configuration Interaction calculation, MECI [39] (pronounced “me-sigh”) which, in addition to automatically correcting “half-electron” energies, allows almost any configuration interaction calculation to be performed. Because of its complexity, two distinct levels of input are supported; the default values will be of use to the novice while an expert has available an exhaustive set of keywords from which a specific C.I. can be tailored.

MECI is a completely general C.I. The resulting states are space and spin-quantized, there is no restriction on total spin, the starting wavefunction can be closed or open shell, and both even and odd electron systems are allowed, although for simplicity in describing the method, the starting configuration is assumed to be closed-shell.

Table 5.12: Two Electron Energy Contributions to EISOL for Atoms in their Ground States

Element	Orbital Config.	State	Mult.:	G_{ss} 1	G_{sp} 1	H_{sp} -1	G_{pp} -1/2	G_{p2} 1/2	G_{sd} 1	H_{sd} -1/5	G_{dd}^0 1	G_{dd}^2 -1/49	G_{dd}^4 -1/49
1	H	$1s^1$	2S										
2	He	$1s^2$	1S	1									
3	Li	$2s^1$	2S										
4	Be	$2s^2$	1S	1									
5	B	$2s^2 2p^1$	2P	1	2	1							
6	C	$2s^2 2p^2$	3P	1	4	2	1	3					
7	N	$2s^2 2p^3$	4S	1	6	3	3	9					
8	O	$2s^2 2p^4$	3P	1	8	4	1	13					
9	F	$2s^2 2p^5$	2P	1	10	5		20					
10	Ne	$2s^2 2p^6$	1S	1	12	6		30					
11	Na	$3s^1$	2S										
12	Mg	$3s^2$	1S	1									
13	Al	$3s^2 3p^1$	2P	1	2	1							
14	Si	$3s^2 3p^2$	3P	1	4	2	1	3					
15	P	$3s^2 3p^3$	4S	1	6	3	3	9					
16	S	$3s^2 3p^4$	3P	1	8	4	1	13					
17	Cl	$3s^2 3p^5$	2P	1	10	5		20					
18	Ar	$3s^2 3p^6$	1S	1	12	6		30					
19	K	$4s^1$	2S										
20	Ca	$4s^2$	1S	1									
21	Sc	$4s^2 3d^1$	2D	1					2	1			
22	Ti	$4s^2 3d^2$	3F	1					4	2	1	8	1
23	V	$4s^2 3d^3$	4F	1					6	3	3	15	8
24	Cr	$4s^1 3d^5$	7S						5	5	10	35	35
25	Mn	$4s^2 3d^5$	6S	1					10	5	10	35	35
26	Fe	$4s^2 3d^6$	5D	1					12	6	15	35	35
27	Co	$4s^2 3d^7$	4F	1					14	7	21	43	36
28	Ni	$4s^2 3d^8$	3F	1					16	8	28	50	43
29	Cu	$4s^1 3d^{10}$	2S						10	5	45	70	70
30	Zn	$4s^2$	1S	1									
31	Ga	$4s^2 4p^1$	2P	1	2	1							
32	Ge	$4s^2 4p^2$	3P	1	4	2	1	3					
33	As	$4s^2 4p^3$	4S	1	6	3	3	9					
34	Se	$4s^2 4p^4$	3P	1	8	4	1	13					
35	Br	$4s^2 4p^5$	2P	1	10	5		20					
36	Kr	$4s^2 4p^6$	1S	1	12	6		30					
37	Rb	$5s^1$	2S										
38	Sr	$5s^2$	1S	1									
39	Y	$5s^2 4d^1$	2D	1					2	1			
40	Zr	$5s^2 4d^2$	3F	1					4	2	1	8	1
41	Nb	$5s^1 4d^4$	6D						4	4	6	21	21
42	Mo	$5s^1 4d^5$	7S						5	5	10	35	35
43	Tc	$5s^2 4d^5$	6S	1					10	5	10	35	35
44	Ru	$5s^1 4d^7$	5F						7	5	21	43	36
45	Rh	$5s^1 4d^8$	4F						8	5	28	50	43
46	Pd	$5s^0 4d^{10}$	1S								45	70	70
47	Ag	$5s^1 4d^{10}$	2S						10	5	45	70	70

Table 5.13: Two Electron Energy Contributions to EISOL for Atoms in their Ground States

Element	Orbital Config.	State	Mult.:	G_{ss} 1	G_{sp} 1	H_{sp} -1	G_{pp} -1/2	G_{p^2} 1/2	G_{sd} 1	H_{sd} -1/5	G_{dd}^0 1	G_{dd}^2 -1/49	G_{dd}^4 -1/49
48	Cd	$5s^2$	$1S$	1									
49	In	$5s^2 5p^1$	$2P$	1	2	1							
50	Sn	$5s^2 5p^2$	$3P$	1	4	2	1	3					
51	Sb	$5s^2 5p^3$	$4S$	1	6	3	3	9					
52	Te	$5s^2 5p^4$	$3P$	1	8	4	1	13					
53	I	$5s^2 5p^5$	$2P$	1	10	5		20					
54	Xe	$5s^2 5p^6$	$1S$	1	12	6		30					
55	Cs	$6s^1$	$2S$										
56	Ba	$6s^2$	$1S$	1									
72	Hf	$6s^2 5d^2$	$3F$	1					4	2	1	8	1
73	Ta	$6s^2 5d^3$	$4F$	1					6	3	3	15	8
74	W	$6s^2 5d^4$	$5D$	1					8	4	6	21	21
75	Re	$6s^2 5d^5$	$6S$	1					10	5	10	35	35
76	Os	$6s^2 5d^6$	$5D$	1					12	6	15	35	35
77	Ir	$6s^2 5d^7$	$4F$	1					14	7	21	43	36
78	Pt	$6s^1 5d^9$	$3D$						9	5	36	56	56
79	Au	$6s^1 5d^{10}$	$2S$						10	5	45	70	70
80	Hg	$6s^2$	$1S$	1									
81	Tl	$6s^2 6p^1$	$2P$	1	2	1							
82	Pb	$6s^2 6p^2$	$3P$	1	4	2	1	3					
83	Bi	$6s^2 6p^3$	$4S$	1	6	3	3	9					
84	Po	$6s^2 6p^4$	$3P$	1	8	4	1	13					
85	At	$6s^2 5p^5$	$2P$	1	10	5		20					

Table 5.14: Examples of SCF configurations used in MECI

System	KeyWords used	Starting Configuration
Methane	none	2.00 2.00 2.00 2.00
Methyl Radical	none	2.00 2.00 2.00 1.00
Twisted Ethylene	OPEN(2,2)	2.00 2.00 1.00 1.00
Twisted Ethylene Cation	OPEN(1,2)	2.00 2.00 0.50 0.50
Methane Cation	CHARGE=1 OPEN(5,3)	2.00 1.67 1.67 1.67

5.4.1 Starting electronic configuration

As MECI requires the space parts of the α and β molecular orbitals to be identical, only RHF wavefunctions are used. However, this is not a severe restriction in that any starting configuration will be supported. Examples of starting configurations are shown in Table 5.14.

Choice of starting configuration is important. For example, if twisted ethylene, a ground-state triplet, is not defined using OPEN(2,2), then the closed-shell ground-state structure will be calculated. Obviously, this configuration is a legitimate microstate, but from the symmetry of the system a better choice would be to define one electron in each of the two formally degenerate π -type M.O.s.

Each configuration which can be generated in a molecule may be represented by a single Slater determinant; this is called a microstate. The final states will be linear combinations of these microstates. In general, microstates will not be eigenfunctions of the total spin operator, but will be mixtures of different spin states.

The initial configuration used to generate the SCF is arbitrary; for half-electron systems it will not even correspond to a microstate, each M.O. having a fractional electron occupancy. Even if the starting wavefunction is a closed shell it would still correspond to only one of a large number of microstates to be used in the MECI. As a result, before the MECI is started all electronic terms arising from the electrons in the initial configuration, which

will be used by MECI, are removed. The starting wavefunction will thus consist of a low-lying doubly occupied set of M.O.s and a high-lying empty set of M.O.s, neither of which will be involved in the MECI, and in between a small set of M.O.s from which the electrons have been removed. This set of M.O.s will be involved in the MECI.

5.4.2 Microstates

Microstates, which are normally represented by a Slater determinant, are normally written as an antisymmetrized product of p α - and q β -electrons:

$$\Psi_g = [(p+q)!]^{-\frac{1}{2}} \sum_P (-1)^P P[\psi_1(1)\alpha(1) \psi_2(2)\alpha(2) \dots \psi_p(p)\alpha(p) \psi_1(p+1)\beta(p+1) \dots \psi_q(p+q)\beta(p+q)],$$

where $[(p+q)!]^{-\frac{1}{2}}$ is the normalization constant, P is an operator which permutes the electron coordinates, and $(-1)^P$ assumes the values -1 or $+1$ for odd and even permutations respectively. A more compact and useful notation for representing a general microstate is:

$$\Psi_j = \frac{1}{\sqrt{N!}} \sum_{P=1}^{N!} (-1)^P P\left(\prod_{k=1}^N \psi_k^j\right)$$

where Ψ_j is any microstate consisting of N electrons. Given the full set of M.O.s, a subset of these is used in the microstate. This subset is defined by the M.O.s ψ_k^j , $k=1, N$. Each microstate will consist of a different set of M.O.s from the full set.

Rather than having all the α electrons appearing first in a microstate, it is more convenient to order the one electron wavefunctions in the order in which their indices occur in the full set of M.O.s. If both α and β M.O.s of the same index occur, then α precedes β , thus:

$$\Psi_g = [(p+q)!]^{-\frac{1}{2}} \sum_P (-1)^P P[\psi_1(1)\alpha(1) \psi_1(2)\beta(2) \psi_2(3)\alpha(3) \psi_2(4)\beta(4) \dots \psi_j(p+q)\beta(p+q)]$$

This numbering scheme follows the Aufbau principle, in that the order of filling is in order of energy. This point is critically important in deciding the sign of matrix elements. For a 5 M.O. system, then, the order of filling is:

$$(1)(\bar{1})(2)(\bar{2})(3)(\bar{3})(4)(\bar{4})(5)(\bar{5})$$

A triplet state arising from two microstates, each with a component of spin = 0, will thus be the positive combination:

$$(\bar{1})(2) + (1)(\bar{2}).$$

This standard sign convention was chosen in order to allow the signs of the microstate coefficients to conform to those resulting from the spin step-down operator.

Only those M.O.s involved in the MECI are of interest, thus from the full set of M.O.s, filled and empty

$$\begin{bmatrix} \psi_1(1)\alpha(1) & \psi_2(3)\alpha(3) & \dots \\ \psi_1(2)\beta(2) & \psi_2(4)\beta(4) & \dots \end{bmatrix}$$

the ground-state configuration (assumed to be closed shell for simplicity) can be represented by

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

where a 1 represents a spin molecular orbital occupied by one electron and 0 represents an empty M.O.

The M.O.s involved in the C.I. are called the "active space". For convenience, the index of the M.O. at the lower bound of the active space will be called "B", and the index of the M.O. at the upper bound of the active space will

Table 5.15: Microstates for 5 electrons in 5 M.O.s

M.O.	Electron Configuration					M_S	M.O.	Electron Configuration					M_S										
	Alpha		Beta					Alpha		Beta													
	1	2	3	4	5	1	2	3	4	5		1	2	3	4	5							
1	1	1	1	0	0	1	1	0	0	0	1/2	4	1	1	1	1	1	0	0	0	0	0	5/2
2	1	1	0	0	0	1	1	1	0	0	-1/2	5	1	1	0	1	0	1	1	0	0	0	1/2
3	1	1	1	0	0	0	0	0	1	1	1/2	6	1	1	0	1	0	1	0	1	0	0	1/2

be called “A”. All M.O.s below the active space can be considered as part of the core while those above it are empty and can likewise be ignored. We can thus focus our attention on the M.O.s in the active space. Most of the time, MECI calculations will involve between 1 and 5 M.O.s, so a system such as pyridine, with 15 filled levels and 29 M.O.s, would use M.O.s 13–17 in a large C.I.

For convenience, microstates will be expressed as a sum of molecular orbital occupancies, so that:

$$\Psi_p = \sum_{i=B}^A (O_i^{\alpha p} + O_i^{\beta p}).$$

For example, if the ground state configuration Ψ_g is closed shell, then the occupancy of the M.O.s would be

$$O^{\alpha g} = O^{\beta g} = |1, \dots, 1, 0, \dots, 0|$$

Microstates are particular electron configurations. Examples of microstates involving 5 electrons in 5 levels are given in Table 5.15.

Permutations

For 5 electrons in 5 M.O.s there are 252 microstates ($10!/(5!5!)$), but as states of different spin do not mix, we can use a smaller number. If doublet states are needed, then 100 states ($5!/(2!3!)(5!/(3!2!))$) are needed. If only quartet states are of interest, then 25 states ($5!/(1!4!)(5!/(4!1!))$) are needed and if the sextet state is required, then only one state is calculated.

In the microstates listed, state 1 is the ground-state configuration. This can be written as (2,2,1,0,0), meaning that M.O.s 1 and 2 are doubly occupied, M.O. 3 is singly occupied by an alpha electron, and M.O.s 4 and 5 are empty. Microstate 1 has a component of spin of 1/2, and is a pure doublet. By Kramer's degeneracy—sometimes called time-inversion symmetry—microstate 2 is also a doublet, and has a spin of 1/2 and a component of spin of $-1/2$.

Microstate 3, while it has a component of spin of 1/2, is not a doublet, but is in fact a component of a doublet, a quartet and a sextet. The coefficients of these states can be calculated from Wigner's symbol, also called the Clebsch-Gordon 3-J symbol¹. Thus, the coefficient in the doublet is $\sqrt{1/2}$ ($j_1 = 3/2$, $m_1 = 3/2$, $j_2 = 1$, $m_2 = -1$, $j = 1/2$), in the quartet is $\sqrt{4/10}$ ($j_1 = 3/2$, $m_1 = 3/2$, $j_2 = 1$, $m_2 = -1$, $j = 3/2$), and in the sextet, $\sqrt{1/10}$ ($j_1 = 3/2$, $m_1 = 3/2$, $j_2 = 1$, $m_2 = -1$, $j = 5/2$).

¹The symbol is of form

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle = \frac{\alpha}{(j_1 + m_1)!(j_2 + m_2)!(j_1 - j_2 + j)!(j_2 - j_1 + j)!(j_1 + j_2 + j + 1)!} \frac{(j + m)!(j - m)!(j_1 - m_1)!(j_2 - m_2)!(j_1 + j_2 - j)!(2j + 1)^{\frac{O\epsilon_1}{2}}}{\delta(m, m_1 + m_2) \sum_r (-1)^{j_1 + r - m_1} \frac{(j_1 + m_1 + r)!(j_2 + j - r - m_1)!}{r!(j - m - r)!(j_1 - m_1 - r)!(j_2 - j + m_1 + r)!}}$$

where the summation is over all values of r such that all factorials occurring are of non-negative integers ($0! = 1$). See [64]. To use the symbol, the coefficient of momentum (j, m) due to two momenta (j_1, m_1) and (j_2, m_2) is $\langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle$

Table 5.16: Sets of Microstates for Various MECI Calculations

Odd Electron Systems				Even Electron Systems			
Alpha Beta			No. of Configs.	Alpha Beta			No. of Configs.
C.I.=1	(1,1) × (0,1)	=	1	(1,1) × (1,1)	=	1	1
2	(1,2) × (0,2)	=	2	(1,2) × (1,2)	=	4	4
3	(2,3) × (1,3)	=	9	(2,3) × (2,3)	=	9	9
4	(2,4) × (1,4)	=	24	(2,4) × (2,4)	=	36	36
5	(3,5) × (2,5)	=	100	(3,5) × (3,5)	=	100	100

(n,m) means n electrons in m M.O.s.

Microstate 4 is a pure sextet. If all 100 microstates of component of spin = 1/2 were used in a C.I., one of the resulting states would have the same energy as the state resulting from microstate 4.

Microstate 5 is an excited doublet, and microstate 6 is an excited state of the system, but not a pure spin-state.

By default, if n M.O.s are included in the MECI, then all possible microstates which give rise to a component of spin = 0 for even electron systems, or 1/2 for odd electron systems, will be used.

MOPAC is configured to allow a maximum of MAXCI states, where MAXCI is defined in the file `meci.h`. If more states are needed (see Table 5.16), then MAXCI in `meci.h` should be modified. Of course, if MAXCI is changed, MOPAC should be recompiled.

If CIS, CISD, or CISDT are specified, then the number of microstates is defined by C.I.= k and the keyword. The number of microstates is a function of k . Let n and m be integers, such that:

$$n = \frac{k}{2}$$

$$m = \frac{k+1}{2}$$

If k is odd, then round down to the next lower integer. Then the number of microstates n_{CIS} , n_{CISD} , and n_{CISDT} , for even-electron systems is:

$$\begin{aligned} n_{CIS} &= 2nm \\ n_{CISD} &= 1 + 2nm + (nm)^2 + \frac{n!m!}{2(n-2)!(m-2)!} \\ n_{CISDT} &= 1 + 2nm + (nm)^2 + \frac{n!m!}{2(n-2)!(m-2)!} + \frac{n!m!}{18(n-3)(m-3)} + \frac{nm \times n!m!}{2(n-2)!(m-2)!} \end{aligned}$$

Note that when CIS is used, the ground state is *not* included in the list of microstates. Values for the more important k are given in Table 5.17.

Energy of microstates

The electronic energy, E_r , of any microstate Ψ_r is the sum on the one and two-electron energies:

$$E_r = \sum_i^p H_{ii} + \sum_i^q H_{ii} + \frac{1}{2} \sum_{ij}^p (J_{ij} - K_{ij}) + \frac{1}{2} \sum_{ij}^q (J_{ij} - K_{ij}) + \sum_i^p \sum_j^q J_{ij}$$

where H_{ii} is the one-electron energy of M.O. ψ_i , J_{ij} is the two-electron Coulomb integral

$\langle \psi_i \psi_i | \psi_j \psi_j \rangle$, and K_{ij} is the two-electron exchange integral $\langle \psi_i \psi_j | \psi_i \psi_j \rangle$.

In this section it is more convenient to express it in terms of molecular orbital occupancies:

$$E_r = \sum_{i=B}^A H_{ii} (O_i^{\alpha r} + O_i^{\beta r}) + \sum_{ij=B}^A \left(\frac{1}{2} (J_{ij} - K_{ij}) (O_i^{\alpha r} O_j^{\alpha r} + O_i^{\beta r} O_j^{\beta r}) + J_{ij} O_i^{\alpha r} O_j^{\beta r} \right)$$

Table 5.17: Number of Microstates for CIS, CISD, and CISDT

C.I.= <i>k</i>	CIS	CISD	CISDT
1	0	1	1
2	2	4	4
3	4	9	9
4	8	27	35
5	12	55	91
6	18	118	282
7	24	205	635
8	32	361	1545

(for even electron systems only)

Similarly, the orbital energies can be written

$$\epsilon_{ii}^{\alpha r} = H_{ii} + \sum_j^p (J_{ij} - K_{ij}) + \sum_j^q J_{ij}$$

or, in terms of orbital occupancies

$$\epsilon_{ii}^{\alpha r} = H_{ii} + \sum_{j=B}^A (J_{ij} - K_{ij}) O_j^{\alpha r} + \sum_{j=B}^A J_{ij} O_j^{\beta r}.$$

Zero of energy used in MECI

The energy of the system after all the electronic terms arising from the electrons of the M.O.s involved in the starting configuration are removed is a useful quantity. Removal of these terms lowers the orbital energies thus:

$$\epsilon_{ii}^+ = \epsilon_{ii} - \sum_{j=B}^A (J_{ij} - K_{ij}) O_j^g.$$

The arbitrary zero of energy in a MECI calculation is the starting ground state, without any correction for errors introduced by the use of fractional occupancies. In order to calculate the energy of the various configurations, the energy of the vacuum state (i.e., the state resulting from removal of the electrons used in the C.I.) needs to be evaluated. This energy is given by:

$$GSE = E_g^+ = - \sum_{i=B}^A 2\epsilon_{ii}^+ O_i^g + J_{ii} (O_i^g)^2 + \sum_{i=B}^A \sum_{j=B}^{i-1} 2(2J_{ij} - K_{ij}) O_i^g O_j^g$$

(Within the MECI routine, GSE refers to E_g^+ .)

By redefining the system so that those filled M.O.s which are not used in the MECI are considered part of an unpolarizable core, the new energy levels ϵ_i^+ can be identified with the one-electron energies H_{ii} and the total electronic energy E_r of any microstate is set equal to the sum of the energy of the electrons considered in the microstate plus E_g^+ .

5.4.3 Construction of secular determinant

Microstates can be generated by permuting available electrons among the available levels. Elements of the C.I. matrix are then defined by

$$\langle \Psi_a | H | \Psi_b \rangle$$

Evaluation of these matrix elements is difficult. Each microstate is a Slater determinant, and the Hamiltonian operator involves all electrons in the system. Fortunately, most matrix elements are zero because of the orthogonality of the M.O.s. Only the non-zero elements need be evaluated; three types of interaction are possible:

1. $\Psi_a = \Psi_b$. Since the two wavefunctions are the same, this corresponds to the energy of a microstate. As the electronic energy of the closed shell is common to all configurations considered in the C.I., it is sufficient to add on to E_g^+ the energy terms which are specific to the microstate, thus

$$\begin{aligned} \langle \Psi_a | H | \Psi_b \rangle &= E_g^+ + \sum_{i=B}^A \left(\epsilon_{ii} + \sum_{j=B}^A (J_{ij} - K_{ij}) O_j^{\alpha p} \right) O_i^{\alpha p} \\ &\quad + \sum_{i=B}^A \left(\epsilon_{ii} + \sum_{j=B}^A (J_{ij} - K_{ij}) O_j^{\beta p} \right) O_i^{\beta p} + \sum_{i=B}^A \sum_{j=B}^A J_{ij} O_i^{\alpha p} O_j^{\beta p}. \end{aligned}$$

2. Except for ψ_i in Ψ_a and ψ_j in Ψ_b ; $\Psi_a = \Psi_b$. Assuming ψ_i and ψ_j to be α -spin the interaction energy is

$$\langle \Psi_a | H | \Psi_b \rangle = (-1)^W (\epsilon_{ij}^+ + \sum_{k=B}^A (\langle ij|kk \rangle - \langle ik|jk \rangle) O_k^{\alpha a} + \langle ij|kk \rangle O_k^{\beta a}).$$

This presents a problem. Unlike ϵ_{ii}^+ , which has already been defined, there is no easy way to calculate ϵ_{ij}^+ . Rather than undertake this calculation, use can be made of the fact that, for the starting configuration:

$$\epsilon_{ij} = \langle \psi_i | H | \psi_j \rangle = H_{ij} + \sum_{k=B}^A (\langle ij|kk \rangle - \langle ik|jk \rangle) O_k^{\alpha g} + \langle ij|kk \rangle O_k^{\beta g}$$

or

$$\epsilon_{ij} = \epsilon_{ij}^+ + \sum_{k=B}^A (\langle ij|kk \rangle - \langle ik|jk \rangle) O_k^{\alpha g} + \langle ij|kk \rangle O_k^{\beta g}.$$

ϵ_{ij} corresponds to an off-diagonal term in the Fock matrix, which at self-consistency is, by definition, zero. Therefore:

$$\epsilon_{ij}^+ = - \sum_{k=B}^A (\langle ij|kk \rangle - \langle ik|jk \rangle) O_k^{\alpha g} + \langle ij|kk \rangle O_k^{\beta g},$$

which can be substituted directly into the expression for $\langle \Psi_a | H | \Psi_b \rangle$ to give

$$\langle \Psi_a | H | \Psi_b \rangle = (-1)^W \sum_{k=B}^A (\langle ij|kk \rangle - \langle ik|jk \rangle) (O_k^{\alpha a} - O_k^{\alpha g}) + \langle ij|kk \rangle (O_k^{\beta a} - O_k^{\beta g}).$$

All that remains is to determine the phase factor. One of the microstates is permuted until the two unmatched M.O.s occupy the same position. The number of permutations needed to do this when the two M.O.s are of α spin is

$$W = \sum_{k=i+1}^{j-1} (O_k^{\alpha p} - O_k^{\beta p}),$$

assuming $j > i$; otherwise:

$$W = O_j^{\alpha p} + \sum_{k=i+1}^{j-1} (O_k^{\alpha p} - O_k^{\beta p}).$$

3. Except for ψ_i and ψ_j in Ψ_a and ψ_k and ψ_l in Ψ_b ; $\Psi_a = \Psi_b$. Two situations exist: (a) when all four M.O.s are of the same spin; and (b) when two are of each spin. Thus,

(a) All four M.O.s are of the same spin. The interaction energy is

$$\langle \Psi_a | H | \Psi_b \rangle = (-1)^W [\langle ik | jl \rangle - \langle il | jk \rangle],$$

in which the phase factor is:

$$W = \sum_{m=i+1}^{j-1} (O_m^{\alpha a} - O_m^{\beta a}) + \sum_{m=k+1}^{l-1} (O_m^{\alpha a} - O_m^{\beta a}) + O_i^{\beta a} + O_k^{\beta a},$$

if the four M.O.s are of α spin; otherwise,

$$W = \sum_{m=i+1}^{j-1} (O_m^{\alpha a} - O_m^{\beta a}) + \sum_{m=k+1}^{l-1} (O_m^{\alpha a} - O_m^{\beta a}) + O_j^{\beta a} + O_l^{\beta a}.$$

(b) Two M.O.s are of each spin. In this case there is no exchange integral, therefore the interaction energy is

$$\langle \Psi_a | H | \Psi_b \rangle = (-1)^W \langle ik | jl \rangle$$

and the phase factor is:

$$W = \sum_{m=k}^i (O_m^{\alpha a} - O_m^{\beta a}) + \sum_{m=j}^l (O_m^{\alpha a} - O_m^{\beta a}).$$

If $i > k$, then $W = W + O_k^{\alpha a} + O_i^{\beta a}$, if $j > l$, then $W = W + O_k^{\alpha b} + O_i^{\beta b}$, finally, if $i > k$ and $j > l$ or $i < k$ and $j < l$, then $W = W + 1$.

All other matrix elements are zero. The completed secular determinant is then diagonalized. This yields the state vectors and state energies, relative to the starting configuration. In turn, the state vectors can be used to generate spin density (at the RHF level) for pure spin states. If the density matrix for the state is of interest, such as in the calculation of transition dipoles for vibrational modes of excited or open shell systems, or for other use, the perturbed density matrix is automatically reconstructed.

5.4.4 Atom Transition Moments

A system can go from the ground state to an excited state as the result of the absorption of a photon. The probability of this happening, κ , is given² in terms of the oscillator integral:

$$\langle \Psi_0 | \vec{r} | \Psi_* \rangle, \quad (5.23)$$

by

$$\kappa = \frac{8\pi^3}{3ch} \nu_{n'n''} (N_{n'} - N_{n''}) \langle \Psi_0 | \vec{r} | \Psi_* \rangle^2.$$

For electronic photoexcitations, Ψ_A are state functions:

$$\Psi_A = \sum_i c_i \Psi_i,$$

and the Ψ_i are microstates; see p. 141 for a definition of microstates.

²Wilson Decius and Cross, "Molecular Vibrations", p 163, McGraw-Hill (1955)

Table 5.18: "x" Transition Integrals

	<i>s</i>	<i>p_x</i>	<i>p_y</i>	<i>p_z</i>	<i>d_{x²-y²}</i>	<i>d_{xz}</i>	<i>d_{z²}</i>	<i>d_{yz}</i>	<i>d_{xy}</i>
<i>s</i>	X_A								
<i>p_x</i>	sp	X_A							
<i>p_y</i>	0	0	X_A						
<i>p_z</i>	0	0	0	X_A					
<i>d_{x²-y²}</i>	0	pd	0	0	X_A				
<i>d_{xz}</i>	0	0	0	pd	0	X_A			
<i>d_{z²}</i>	0	$-\frac{1}{\sqrt{3}}$ pd	0	0	0	0	X_A		
<i>d_{yz}</i>	0	0	0	0	0	0	0	X_A	
<i>d_{xy}</i>	0	0	pd	0	0	0	0	0	X_A

Note: $X_A = \langle \phi_\lambda | \vec{x} | \phi_\lambda \rangle$; sp = $\langle ns | \vec{r} | np \rangle$; pd = $\langle np | \vec{r} | nd \rangle$ (see below).

Some Mathematical tools

In order to evaluate 5.23, a property of integrals of the type:

$$\langle \psi_i | \vec{r} | \psi_j \rangle$$

will be used several times. This property is:

$$\langle \psi_i | \vec{r} | \psi_i \rangle = 0.$$

From this, it follows that, if

$$\langle \psi_j | \vec{r} | \psi_i \rangle \neq 0,$$

then

$$\langle \psi_i | \vec{r} | \psi_j \rangle = - \langle \psi_j | \vec{r} | \psi_i \rangle.$$

To prove this relationship, consider the integral

$$\langle (\psi_i + \psi_j) | \vec{r} | (\psi_i + \psi_j) \rangle.$$

Obviously, this integral has a value of zero, therefore

$$\langle \psi_i | \vec{r} | \psi_i \rangle + \langle \psi_j | \vec{r} | \psi_i \rangle + \langle \psi_i | \vec{r} | \psi_j \rangle + \langle \psi_j | \vec{r} | \psi_j \rangle = 0.$$

In this expression, the first and fourth terms are obviously zero, therefore

$$\langle \psi_j | \vec{r} | \psi_i \rangle = - \langle \psi_i | \vec{r} | \psi_j \rangle.$$

Evaluation of Transition Dipole

The probability, $B_{0 \rightarrow *}$, that a photon will be absorbed by a system that has a ground state Ψ_0 and an excited state Ψ_* separated by an energy ϵ when irradiated by an energy density ρ_ϵ is given by

$$B_{0 \rightarrow *} = \frac{2\pi}{3\hbar^2} |R_{0*}|^2 \rho_\epsilon,$$

in which

$$|R_{0*}|^2 = |X_{0*}|^2 + |Y_{0*}|^2 + |Z_{0*}|^2.$$

Table 5.19: "y" Transition Integrals

	<i>s</i>	<i>p_x</i>	<i>p_y</i>	<i>p_z</i>	<i>d_{x²-y²}</i>	<i>d_{xz}</i>	<i>d_{z²}</i>	<i>d_{yz}</i>	<i>d_{xy}</i>
<i>s</i>	Y_A								
<i>p_x</i>	0	Y_A							
<i>p_y</i>	sp	0	Y_A						
<i>p_z</i>	0	0	0	Y_A					
<i>d_{x²-y²}</i>	0	0	-pd	0	Y_A				
<i>d_{xz}</i>	0	0	0	0	0	Y_A			
<i>d_{z²}</i>	0	0	$-\frac{1}{\sqrt{3}}pd$	0	0	0	Y_A		
<i>d_{yz}</i>	0	0	0	pd	0	0	0	Y_A	
<i>d_{xy}</i>	0	pd	0	0	0	0	0	0	Y_A

X_{0*} is the matrix element for the *x* component of the dipole moment:

$$X_{0*} = \int \Psi_0 |e \sum_j \vec{x}_j | \Psi_* d\tau.$$

Evaluation of this integral requires evaluating the effect of the operators \vec{x}_j , \vec{y}_j , and \vec{z}_j acting on an atomic orbital. Tables 5.18, 5.19, and 5.20 show the integrals of the type $\langle \phi_\lambda | \vec{r}_j | \phi_\sigma \rangle$, where ϕ_λ and ϕ_σ are pairs of atomic orbitals.

The integral $\langle \phi_\lambda | \vec{r} | \phi_\lambda \rangle$ is simply the appropriate Cartesian coordinate, that is, the *x*, *y*, or *z* coordinate of the atom that ϕ_λ is on.

$\langle ns | \vec{r} | np \rangle$ and $\langle np | \vec{r} | nd \rangle$ can be evaluated using the following expressions:

$$\langle ns | \vec{r} | np \rangle = a_0 \frac{(2n+1) \cdot 2^{2n+1} \cdot (\xi_s \xi_p)^{n+1/2}}{\sqrt{3}(\xi_s + \xi_p)^{2n+2}}$$

$$\langle np | \vec{r} | nd \rangle = a_0 \frac{(n_p + n_d + 1)! \cdot 2^{n_p+n_d+1} \cdot \xi_p^{n_p+1/2} \cdot \xi_d^{n_d+1/2}}{\sqrt{5}(\xi_p + \xi_d)^{n_p+n_d+2} \cdot \sqrt{(2n_p)!(2n_d)!}},$$

where *ns*, *np*, and *nd* are *s*, *p*, and *d* quantum numbers, respectively. For the *sp* transition, $n = ns = np$. The Slater orbital exponents, ξ_s , ξ_p , and ξ_d , are usually given in atomic units, that is, in inverse Bohr, therefore they must be converted to Ångstroms before use, hence the presence of the $a_0 = 0.529$ in these expressions.

All integrals of the type used here are in Ångstroms, therefore the units of the integral of the dipole operator on a M.O. is also in Ångstroms:

$$\langle \psi_i | \vec{r} | \psi_j \rangle = \sum_\lambda \sum_\sigma c_{\lambda i} c_{\sigma j} \langle \phi_\lambda | \vec{r} | \phi_\sigma \rangle,$$

Once the value of the integral is known, the phase to be used must be determined. The simplest way to achieve this is to reverse the sign of the oscillator whenever the second M.O. has a higher index than the first.

Evaluation of the integrals over microstates is straightforward, in that all integrals are zero, unless the number of differences between the microstates is exactly two, in which case the integral is equal to that of the two M.O.s involved, times a phase factor. That is, for each pair of microstates that are identical, except for ψ_i in Ψ_a and ψ_j in Ψ_b , the integral is:

$$\langle \Psi_a | \vec{r} | \Psi_b \rangle = \langle \psi_i | \vec{r} | \psi_j \rangle * (-1)^n,$$

where *n* is the number of permutations necessary to move ψ_i in microstate Ψ_a to the position occupied by ψ_j in microstate Ψ_b . This is similar to the 'b' option on page 145. As with the molecular orbitals, the oscillators for

Table 5.20: "z" Transition Integrals

	<i>s</i>	<i>p_x</i>	<i>p_y</i>	<i>p_z</i>	<i>d_{x²-y²}</i>	<i>d_{xz}</i>	<i>d_{z²}</i>	<i>d_{yz}</i>	<i>d_{xy}</i>
<i>s</i>	Z _A								
<i>p_x</i>	0	Z _A							
<i>p_y</i>	0	0	Z _A						
<i>p_z</i>	sp	0	0	Z _A					
<i>d_{x²-y²}</i>	0	0	0	0	Z _A				
<i>d_{xz}</i>	0	pd	0	0	0	Z _A			
<i>d_{z²}</i>	0	0	0	$\frac{2}{\sqrt{3}}$ pd	0	0	Z _A		
<i>d_{yz}</i>	0	0	pd	0	0	0	0	Z _A	
<i>d_{xy}</i>	0	0	0	0	0	0	0	0	Z _A

Table 5.21: States arising from C.I.=*n*

No. of M.O.s in MECI	States Arising from Odd Electron Systems			States Arising from Even Electron Systems		
	Doublets	Quartets	Sextets	Singlets	Triplets	Quintets
1	1			1		
2	2			3	1	
3	8	1		6	3	
4	20	4		20	15	1
5	75	24	1	50	45	5

microstates change sign when the order of the microstates is reversed. The simplest way to achieve this is to use the same device that was used with the M.O.s; that is, to reverse the sign of the oscillator whenever the second microstate has a higher index than the first.

For completeness, the sign of R_{0*} should be reversed if $k > l$, but since only the modulus is used, this operation does not need to be done.

Finally, the state transition dipole can be calculated from:

$$\langle \Psi_A | \vec{r} | \Psi_B \rangle = \sum_a \sum_b c_{Aa} c_{Bb} \langle \Psi_a | \vec{r} | \Psi_b \rangle,$$

Although the transition dipole is normally regarded as involving the ground and an excited state, it is possible to calculate the transition between two excited states. The initial state is, by default, the ground state, however if ROOT=*n* $n \neq 1$, or any other keyword that specifies a state other than the ground state, then the initial state will be an excited state.

For degenerate states, the transition dipole is the sum over all states involved.

5.4.5 States arising from various calculations

Each MECI calculation invoked by use of the keyword C.I.=*n* normally gives rise to states of quantized spins. When C.I. is used without any other modifying keywords, the states shown in Table 5.21 will be obtained. These numbers of spin states will be obtained irrespective of the chemical nature of the system.

5.4.6 Spin angular momentum

State functions are eigenvalues of the S_z and S^2 operators. The derivation of the expectation value of the S^2 operator is given in this section.

The fundamental spin operators have the following effects:

$$\begin{aligned} S_x\alpha &= \frac{1}{2}\beta & S_x\beta &= \frac{1}{2}\alpha \\ S_y\alpha &= \frac{i}{2}\beta & S_y\beta &= -\frac{i}{2}\alpha \\ S_z\alpha &= \frac{1}{2}\alpha & S_z\beta &= -\frac{1}{2}\beta \end{aligned}$$

Using these expressions, various useful identities can be established:

$$\begin{aligned} S^2 &= S_x^2 + S_y^2 + S_z^2 \\ I^+ &= (S_x + iS_y); & I^+\beta &= \alpha \\ I^- &= (S_x - iS_y); & I^-\alpha &= \beta \end{aligned}$$

$$\begin{aligned} S_x^2 + S_y^2 &= (I^+I^-) + i(S_xS_y - S_yS_x) \\ &= (I^-I^+) + i(S_yS_x - S_xS_y) \\ &= \frac{1}{2}(I^+I^- + I^-I^+) \end{aligned}$$

and finally $i(S_yS_x - S_xS_y) = S_z$.

For any microstate Ψ , the expectation value of the S^2 operator is given by

$$\langle S^2 \rangle = \langle \Psi | S_z^2 + S_y^2 + S_x^2 | \Psi \rangle .$$

The first part of this expression is obvious, *vis*:

$$\langle \Psi | S_z^2 | \Psi \rangle = \frac{1}{4}(N^\alpha + N^\beta).$$

However, the effect of $S_y^2 + S_x^2$ is not so simple. By making use of the fact that the operators involve two electrons, a large number of integrals resulting from the expansion of the Slater determinants can be readily eliminated. The only integrals which are not zero due to the orthogonality of the eigenvectors, i.e., those which may be finite due to the spin operators, are

$$\langle \Psi | S_y^2 + S_x^2 | \Psi \rangle = 2 \sum_{i < j} [\langle \psi_i \psi_i | S_y^2 + S_x^2 | \psi_j \psi_j \rangle - \langle \psi_i \psi_j | S_y^2 + S_x^2 | \psi_i \psi_j \rangle].$$

Using the relationships already defined, this expression simplifies [65] as follows:

$$S_1 S_2 = S_{1z} S_{2z} + \frac{1}{2}(I_1^+ I_2^- + I_1^- I_2^+)$$

$$\langle \Psi | S^2 | \Psi \rangle = 2 \sum_{i < j} \left[\frac{1}{4}(2\delta(m_{s_i} m_{s_j}) - 1 - \frac{1}{2}(1 - \delta(m_{s_i} m_{s_j}))) \langle \psi_i \psi_j \rangle^2 \right]$$

or,

$$\langle \Psi | S^2 | \Psi \rangle = \frac{3(p+q)}{4} + \frac{p(p-1)}{2} + \frac{q(q-1)}{2} - \frac{(p+q)(p+q-1)}{4} - \sum_{ij}^{pq} \langle \psi_i \psi_j \rangle^2 .$$

Recall that p is the number of α electrons, and q , the number of β electrons. This expression simplifies to yield

$$\langle \Psi | S^2 | \Psi \rangle = \frac{1}{2}(p+q) + \frac{1}{4}(p-q)^2 - \sum_i^p \sum_j^q \langle \psi_i \psi_j \rangle^2 .$$

For the general case, in which the state function Φ , is a linear combination of microstates, the expectation value of S is more complicated:

$$\langle \Phi_k | S^2 | \Phi_k \rangle = \sum_i \sum_j C_{ik} C_{jk} \langle \Psi_i | S^2 | \Psi_j \rangle .$$

As with the construction of the C.I. matrix, the elements of this expression can be divided into a small number of different types:

1. $\Psi_a = \Psi_b$: Since the two wavefunctions are the same, this corresponds to the expectation value of a microstate, and has already been derived.
2. Except for ψ_i in Ψ_a and ψ_j in Ψ_b ; $\Psi_a = \Psi_b$: Assuming ψ_i and ψ_j to have alpha-spin the expectation value is

$$\langle \Psi_a | S_y^2 + S_x^2 | \Psi_b \rangle = (-1)^W \sum_{k=B}^A (\langle ij | kk \rangle - \langle ik | jk \rangle) O_k^{\alpha\alpha} + \langle ij | kk \rangle O_k^{\beta\alpha} .$$

The effect of the spin operator is to change the spin of the electrons but leave the space part unchanged. All integrals vanish identically due to one or more of the following identities:

$$\begin{aligned} \langle \psi_i \psi_j \rangle &= 0; & \langle m_i m_j \rangle &= \delta(i, j) \\ \langle \psi_i \psi_k \rangle &= \delta(i, k); & \langle \psi_j \psi_k \rangle &= \delta(j, k). \end{aligned}$$

Therefore, $\langle \Psi_a | S^2 | \Psi_b \rangle = 0$.

3. Except for ψ_i and ψ_j in Ψ_a and ψ_k and ψ_l in Ψ_b ; $\Psi_a = \Psi_b$. Two situations exist: (a) when all four M.O.s are of the same spin; and (b) when two are of each spin.

When all four M.O.s have the same spin, the effect of the spin operator is to reverse the spin of two M.O.s in the ket half of the integral. By spin orthogonality this results in an integral value of zero.

In the case where two M.O.s are of α spin and two are of β spin, the matrix elements, after elimination of those terms which are zero due to space orthogonality, are

$$\langle \Psi_a | S^2 | \Psi_b \rangle = (-1)^W (\langle \psi_i \psi_k | S^2 | \psi_j \psi_l \rangle - \langle \psi_i \psi_l | S^2 | \psi_j \psi_k \rangle)$$

The effect of S^2 on ψ_k and ψ_l is to reverse the spin of these functions; this gives

$$\langle \Psi_a | S^2 | \Psi_b \rangle = (-1)^W (\langle \psi_i \psi'_k \rangle \langle \psi_j \psi'_l \rangle - \langle \psi_i \psi'_l \rangle \langle \psi_j \psi'_k \rangle),$$

where ψ' has the opposite spin to that of ψ .

Thus, only if ψ_i and ψ_j are spatially identical with ψ_k and ψ_l will $\langle \Psi_a | S^2 | \Psi_b \rangle$ be non-zero. The phase-factor W is such that if $i = k$ and $j = l$ then $W = -1$, and if $i = l$ and $j = k$ then $W = 1$; for all other cases the matrix element is zero, so the phase of W is irrelevant. For these two cases, the matrix element is $\langle \Psi_a | S^2 | \Psi_b \rangle = 1$ if $(I^+ + I^-)(\psi_i + \psi_j) = (\psi_k + \psi_l)$, otherwise $\langle \Psi_a | S^2 | \Psi_b \rangle = 0$.

4. If more than two differences exist, $\langle \Psi_a | S^2 | \Psi_b \rangle = 0$.

Calculation of spin-states

In order to calculate the spin-state, the expectation value of S^2 is calculated.

$$\begin{aligned} \langle \Phi_k | S^2 | \Phi_k \rangle &= S(S+1) = S_z^2 + 2I^+ I^- \\ &= \frac{1}{2}(p+q) - \end{aligned}$$

Figure 5.5: Energies of Excited States

State	Energy (eV)		Q.N.	Spin	Symmetry
	Absolute	Relative			
1	-0.0049	0.0000	1	Singlet	A1
2	2.7109	2.7158	1	Triplet	A2
3	3.1029	3.1078	2	Singlet	A2
4	7.8630	7.8679	2	Singlet	A1

$$\sum_i \left\{ C_{ik} C_{ik} \left((1/4)(N_i^\alpha - N_i^\beta)^2 + \sum_l O_{lk}^\alpha O_{lk}^\beta \right) + \sum_j 2 [C_{ik} C_{jk} [\delta(\Psi_i, (I^+ I^-) \Psi_j)]] \right\}$$

where C_{ik} is the coefficient of microstate Ψ_i in State Φ_k , N_i^α is the number of alpha electrons in microstate Ψ_i , N_i^β is the number of beta electrons in microstate Ψ_i , O_{lk}^α is the occupancy of alpha M.O. l in microstate Ψ_k , O_{lk}^β is the occupancy of beta M.O. l in microstate Ψ_k , I^+ is the spin shift up or step up operator, and I^- is the spin shift down or step down operator.

The spin state is calculated from:

$$S = (1/2)[\sqrt{(1 + 4S^2)} - 1]$$

In practice, S is calculated to be exactly integer, or half integer. That is, there is insignificant error due to approximations used. This does not mean, however, that the method is accurate. The spin calculation is completely precise, in the group theoretic sense, but the accuracy of the calculation is limited by the Hamiltonian used, a space-dependent function.

5.4.7 Choice of State to be Optimized

MECI can calculate a large number of states of various total spin. Two schemes are provided to allow a given state to be selected. First, `ROOT=n` will, when used on its own, select the n 'th state, irrespective of its total spin. By default, $n=1$. If `ROOT=n` is used in conjunction with a keyword from the set `SINGLET`, `DOUBLET`, `TRIPLET`, `QUARTET`, `QUINTET`, `SEXTET`, `SEPTET`, `OCTET`, or `NONET`, then the n 'th root of that spin-state will be used. For example, `ROOT=4` and `SINGLET` will select the 4th singlet state. If there are two triplet states below the fourth singlet state then this will mean that the sixth state will be selected.

Sometimes the energy required to form an excited state is wanted. By this we mean the energy of the excited state relative to the energy of the ground state, and not the heat of formation of the excited state. To calculate this quantity, the keywords `PRECISE`, `GNORM=0.01`, `MECI` and `C.I.=2` should be used. For formaldehyde, these keywords would produce the output shown in Figure 5.5. This output can be read as follows: The first state (the one at -0.004891eV) is the new ground state. C.I. will lower the energy of the ground state, relative to the SCF ground state, and for formaldehyde this extra stabilization amounts to 0.0049 eV. The ground state is a singlet, and has A_1 symmetry. The second state is a triplet, with energy 2.7109eV above the SCF energy, or 2.7158eV above the ground state, and has A_2 symmetry. The third and fourth states are both singlets.

Using the two keywords given, the system would optimize on the ground singlet state, and the bond orders and density matrix would reflect this. If the first excited singlet state were wanted, then the extra keywords `ROOT=2` and `SINGLET` would also be used. Alternatively, the single extra keyword `ROOT=3` could be used. If the first triplet state were wanted, then `TRIPLET` or `ROOT=2` (but not both!) could be used.

Quantum Numbers

When MECI is used, the output contains information on the symmetry of each state. States of different symmetries are automatically orthogonal, but states of the same symmetry do not need to be orthogonal. Of course they are orthogonal, and, to emphasize this fact, an extra symmetry label is added. This label is, in fact, a quantum number, and is given under the heading "Q.N." in the output. The first occurrence of a given irreducible representation is given the Q.N. "1", the second, "2", etc. By using the Q.N. and the symmetry label, each state can be assigned a unique label.

Polarizability

The expectation value of the polarization operator is given under "POLARIZABILITY." This is an approximation to the transition moment for the absorption or emission of a photon. One of the two states involved is the state defined by the keywords. By default, this is the ground state, but might be an excited state, for example. For a description of this calculation, see p. 146.

Franck-Condon considerations

This section was written based on discussions with

Victor I. Danilov
Department of Quantum Biophysics
Academy of Sciences of the Ukraine
Kiev 143
Ukraine

The Frank-Condon principle states that electronic transitions take place in times that are very short compared to the time required for the nuclei to move significantly. Because of this, care must be taken to ensure that the calculations actually do reflect what is wanted.

Examples of various phenomena which can be studied are:

Photoexcitation If the purpose of a calculation is to predict the energy of photoexcitation, then the ground-state should first be optimized. Once this is done, then a C.I. calculation can be carried out using 1SCF. With the appropriate keywords (MECI C.I.=*n* etc.), the energy of photoexcitation to the various states can be predicted.

A more expensive, but more rigorous, calculation would be to optimize the geometry using all the C.I. keywords. This is unlikely to change the results significantly, however.

Fluorescence If the excited state has a sufficiently long lifetime, so that the geometry can relax, then if the system returns to the ground state by emission of a photon, the energy of the emitted photon will be less (it will be red-shifted) than that of the exciting photon. To do such a calculation, proceed as follows:

- Optimize the ground-state geometry using all the keywords for the later steps, but specify the ground state, e.g. C.I.=3 GNORM=0.01 MECI.
- Optimize the excited state, e.g. C.I.=3 ROOT=2 GNORM=0.01 MECI.
- Calculate the Franck-Condon excitation energy, using the results of the ground-state calculation only.
- Calculate the Franck-Condon emission energy, using the results of the excited state calculation only.
- If indirect emission energies are wanted, these can be obtained from the ΔH_f of the optimized excited and optimized ground-state calculations.

In order for fluorescence to occur, the photoemission probability must be quite large, so only transitions of the same spin are allowed. For example, if the ground state is S_0 , then the fluorescing state would be S_1 .

Phosphorescence If the photoemission probability is very low, then the lifetime of the excited state can be very long (sometimes minutes). Such states can become populated by $S_1 \rightarrow T_1$ intersystem crossing. Of course, the geometry of the system will relax before the photoemission occurs.

Indirect emission If the system relaxes from the excited electronic, ground vibrational state to the ground electronic, ground vibrational state, then a more complicated calculation is called for. The steps of such a calculation are:

- Optimize the geometry of the excited state.
- Using the same keywords, except that the ground state is specified, optimize the geometry of the ground state.
- Take the difference in ΔH_f of the optimized excited and optimized ground-state calculations.
- Convert this difference into the appropriate units.

Excimers An excimer is a pair of molecules, one of which is in an electronic excited state. Such systems are usually stabilized relative to the isolated systems. Optimization of the geometries of such systems is difficult. Suggestions on how to improve this type of calculation would be appreciated.

5.4.8 Definition of some C.I. Keywords

It has been my policy, ever since the first release of MOPAC in 1983, to resist changing the definition of keywords. This policy has allowed users to confidently use a new MOPAC in the belief that old keywords will have their old, familiar, meaning. However, an ambiguity was found in certain keywords, an ambiguity which has, at times, resulted in severe frustration.

Consider the word TRIPLET. This meant (but no longer means) "Do an SCF calculation in which the M.O. populations are [... ,2,2,2,1,1,0,0, ...], then do a C.I. on the two half-occupied M.O.s, and select the triplet state." This definition meant that twisted ethylene would have the correct symmetry, as would triplet oxygen. However, if a user wanted to examine triplet formaldehyde, and compare it with the ground state, problems arose. The keywords C.I.=2 ROOT=2 would generate the correct energy, but a user might expect that TRIPLET should achieve the same result. Because of the definition of TRIPLET, the SCF starting configuration was different, and as a result, the ΔH_f was also different. Under earlier MOPACs, there was no way to set up a calculation using the keyword TRIPLET and go SCF on a closed-shell configuration as the precursor to a C.I. calculation.

Because of the limitations of the earlier definitions of spin-states (TRIPLET, QUARTET, QUINTET, SEXTET, etc.), these words were all redefined in 1993, in MOPAC 93. In order to reproduce the earlier keywords, pairs of keywords, such as TRIPLET OPEN(2,2) or SEXTET OPEN(5,5) must now be used. Spin-states which result from SCF calculations on ground-state configurations can now be specified by the following pairs of keywords: TRIPLET C.I.=2; QUARTET C.I.=3; QUINTET C.I.=4; SEXTET C.I.=5.

Using these new definitions, spin-states of a system can now be more easily related. Consider the various states of formaldehyde (Table 5.22), in which all calculations use the ground-state geometry and 1SCF.

Now we see that C.I.=3 ROOT=2 and C.I.=3 TRIPLET do, in fact, give the same result. The "old" MOPAC (pre-1993) result of using TRIPLET can still be generated by OPEN(2,2) TRIPLET. Note that C.I.=1 generates the normal ΔH_f of CH_2O , and that increasing the C.I. lowers the energy steadily.

Table 5.22: Examples of Use of C.I. Keywords

Keywords Used	ΔH_f
(No keywords)	-32.9040
C.I.=1	-32.9040
C.I.=2	-33.0166
C.I.=3	-39.7234
C.I.=4	-39.9665
C.I.=5	-40.1743
C.I.=2 TRIPLET	29.6348
C.I.=3 ROOT=2	28.2840
C.I.=3 TRIPLET	28.2840
C.I.=3 TRIPLET MS=0	28.2840
OPEN(2,2) TRIPLET	27.9318

5.4.9 Degenerate States

By the Jahn-Teller theorem, systems with orbital degeneracy will distort so as to remove the degeneracy. However, many dynamic Jahn-Teller systems are known in which the time-average geometry is of the higher point-group. These systems are the kind that will be addressed here.

The analytical RHF configuration interaction first derivative calculation developed by Liotard [43] has been modified to allow systems with degenerate states to be run.

Each of the degenerate states is a linear combination of microstates. Each microstate can be described by a Slater determinant [66, 67], which represents a specific pattern of occupancy of molecular orbitals. Each M.O. is a linear combination of Slater atomic orbitals.

The whole state is best described by an equal mixture of the degenerate states of which it is composed. Note that this is NOT a combination of states, rather it is a mixture. An example of a combination of states is a state function, composed of a linear combination of microstates. In such a combination the phase-factor between microstates is significant, thus state(1) = $1/\sqrt{2}(\text{Microstate}(a) + \text{Microstate}(b))$ is different from state(2) $1/\sqrt{2}(\text{Microstate}(a) - \text{Microstate}(b))$. An example of a mixture of states is the ${}^2T_{2g}$ state of TiF_6^{3-} , a d^1 system, in which the best description of the state is an equal mixture of the three degenerate space components of T_{2g} , and an equal mixture of the two spin components of the Kramer's doublet. The overall state is thus $1/6(\alpha(T_{2g}(x)+T_{2g}(y)+T_{2g}(z))+\beta(T_{2g}(x)+T_{2g}(y)+T_{2g}(z)))$.

If equimixtures are not used, then the Jahn-Teller theorem applies, and the system would immediately distort so as to remove the degeneracy. In the case of TiF_6^{3-} , this would result in distortion from O_h to D_{4h} symmetry.

5.4.10 Calculation of Spin Density

Starting with the state functions as linear combinations of configurations, the spin density, corresponding to the alpha spin density minus the beta spin density, will be calculated for the first few states. This calculation is straightforward for diagonal terms, and only those terms are used.

5.5 An SCF Calculation

Semiempirical calculations can be run on computers using readily available programs such as MOPAC. It is possible to use MOPAC for research without having any knowledge of its workings. This does not imply any failing on

the part of the researcher; after all, it is possible to write extensive computer programs while having little knowledge of how a computer works. However, in order to more efficiently use MOPAC, a more than casual knowledge of the theory involved is desirable. In the following section the details of a very simple calculation will be described. This calculation can be carried out as a 'black box' calculation, but the following exercise may help to satisfy the intellectual curiosity of users of semiempirical methods regarding the mechanics of carrying out an SCF calculation.

The MNDO method will be used because it is the oldest of the "NDO" methods. The CNDO/2 method is very similar, and the example we will look at will emphasize the similarity. The system to be examined is a regular hexagon of hydrogen atoms in which the H-H distance is 0.98316 Ångstrom. Of course, a regular hexagon of hydrogen atoms is not a stable system; the only reason we are using it here is to demonstrate the working of an SCF calculation. The optimized geometry was obtained by defining all bond lengths to be equal, constraining all bond angles to be 120 degrees and defining the system as being planar. We will need various reference data in order to follow the calculation. MOPAC contains a large data-set, BLOCK.F, of atomic and diatomic parameters for all the elements which have been parameterized. By reference to this source file we find that, for hydrogen:

$$\begin{aligned} G_{ss} &= \langle \varphi_s \varphi_s | 1/r | \varphi_s \varphi_s \rangle = 12.848 \text{eV} \\ U_{ss} &= \langle \varphi_s | H | \varphi_s \rangle = -11.906276 \text{eV} \\ \xi_s &= 1.331967 \text{Bohr} \\ \beta_s &= -6.989064 \text{eV} \\ E_{atom} &= 52.102 \text{kcal/mol} \end{aligned}$$

This exercise is designed to allow the reader to reproduce each step. All that is needed in order to follow the working is a hand calculator.

Interatomic Distance Matrix (Å)						
Atom	1	2	3	4	5	6
1	0.0000					
2	0.9832	0.0000				
3	1.7029	0.9832	0.0000			
4	1.9663	1.7029	0.9832	0.0000		
5	1.7029	1.9663	1.7029	0.9832	0.0000	
6	0.9832	1.7029	1.9663	1.7029	0.9832	0.0000

The overlap integral of two Slater orbitals between two hydrogen atoms is particularly simple:

$$\langle \varphi | \varphi \rangle = (e^{-a}) \left(\frac{a^2}{3} + a + 1 \right)$$

where $a = \xi R/a_0$.

At the optimum H-H distance of 0.9831571 Å, this yields an overlap integral of 0.4643. The nearest-neighbor one-electron integral is thus

$$H(1,2) = S_{1,2}(\beta_s + \beta_s)/2 = -3.2457 \text{eV}.$$

In general, overlap integrals are more complicated and also involve angular components, but the principles involved are the same. You may want to check other off-diagonal terms in the one-electron matrix, or you may accept the results given here.

One-electron matrix (eV)						
Atom	1	2	3	4	5	6
1	-51.7124					
2	-3.2457	-51.7124				
3	-1.0970	-3.2457	-51.7124			
4	-0.6992	-1.0970	-3.2457	-51.7124		
5	-1.0970	-0.6992	-1.0970	-3.2457	-51.7124	
6	-3.2457	-1.0970	-0.6992	-1.0970	-3.2457	-51.7124

On-diagonal one-electron integrals are more complicated than the off-diagonal terms. The one-electron energy of an electron in an atomic orbital is the sum of its kinetic energy and stabilization due to the positive nucleus of its own atom, U_{ss} or U_{pp} , plus the stabilization due to all the other nuclei in the system. Each electron on a hydrogen atom experiences a stabilization due to the five other unipositive nuclei in the system. Within semiempirical theory the electron-nuclear interaction is related to the electron-electron interaction via

$$E_{e,n} = -Z_n \langle \varphi_s \varphi_s | \varphi_s \varphi_s \rangle .$$

Given the two-electron two-center integral matrix the calculation of the diagonal terms of the one-electron matrix is straightforward:

$$H_{n,n} = -11.9063 - 2(9.6585) - 2(7.0635) - 6.3622 = -51.7124.$$

For interactions between an atomic orbital and a non-hydrogen atom there will be ten terms; these arise from all permutations of the basis set, s , p_x , p_y , p_z with the atomic orbital under the neglect of differential overlap approximation. The ten integrals are $\langle ii|ss \rangle$, $\langle ii|sp_x \rangle$, $\langle ii|p_x p_x \rangle$, $\langle ii|sp_y \rangle$, $\langle ii|p_x p_y \rangle$, $\langle ii|p_y p_y \rangle$, $\langle ii|sp_z \rangle$, $\langle ii|p_x p_z \rangle$, $\langle ii|p_y p_z \rangle$, and $\langle ii|p_z p_z \rangle$.

Two-Electron Integrals (eV)						
Atom	1	2	3	4	5	6
1	12.8480					
2	9.6585	12.8480				
3	7.0635	9.6585	12.8480			
4	6.3622	7.0732	9.6585	12.8480		
5	7.0635	6.3622	7.0732	9.6585	12.8480	
6	9.6585	7.0635	6.3622	7.0732	9.6585	12.8480

5.5.1 Starting density matrix

The density matrix is necessary in order to calculate the Fock matrix, but, in turn, the Fock matrix is necessary in order to calculate the density matrix. To break this impasse, a guessed density matrix is used. The guess is very crude: all off-diagonal matrix elements are set to zero, and all on-diagonal terms on any atom are set equal to the core charge of that atom divided by the number of atomic orbitals. Our starting guess for H_6 consists of a unit matrix.

Each iteration of the SCF calculation consists of assembling a Fock matrix from the one-electron matrix, the two-electron integrals, and the density matrix, diagonalizing it to obtain the eigenvectors, and finally reassembling the density matrix. At some point the change in density matrix drops below a preset limit. When this happens we say that the field is self-consistent. We will now carry out these steps for the H_6 system.

5.5.2 Assembly of the starting Fock matrix

In the first iteration this is particularly simple, as there are no off-diagonal terms in the density matrix. Only the on-diagonal terms are affected. Each on-diagonal term in the Fock matrix F_{aa} is modified by the electrostatic field

of all the electrons in the system except the electron or fraction of an electron in the atomic orbital φ_a . Consider $F(1,1)$. The total initial population of φ_1 is 1.0, composed of equal amounts of α and β electron density. An electron in φ_1 would therefore experience the electrostatic repulsion of half an electron. An electron cannot repel itself; however, it will be repelled by its partner electron of opposite spin.

In addition, each electron will be affected, normally repelled, by the electrostatic field of all the electrons on all the other atoms. Each atom has one electron, so the total energy of an electron, i.e., the diagonal Fock matrix element, is given by:

$$F(1,1) = -51.7124 + \frac{1}{2}(12.848) + 2(9.6585 + 7.0635) + 6.3622.$$

The Fock matrix is obtained by adding the two-electron terms to the one electron matrix. The elements of the Fock matrix represent the sum of the one and two electron interactions. For the system of six hydrogen atoms, this has the following form:

Initial Fock Matrix (eV)						
Atom	1	2	3	4	5	6
1	-5.4823					
2	-3.2457	-5.4823				
3	-1.0970	-3.2457	-5.4823			
4	-0.6992	-1.0970	-3.2457	-5.4823		
5	-1.0970	-0.6992	-1.0970	-3.2457	-5.4823	
6	-3.2457	-1.0970	-0.6992	-1.0970	-3.2457	-5.4823

5.5.3 Diagonalization of the Fock matrix

The Fock matrix is then diagonalized to yield the following set of eigenvalues, or one-electron energies, and eigenvectors, or molecular orbitals:

Energy Level	Molecular Orbital Coefficients					
	1	2	3	4	5	6
6 -0.4857	0.4082	-0.4082	0.4082	-0.4082	0.4082	-0.4082
5 -1.8388	0.5774	-0.2887	-0.2887	0.5774	-0.2887	-0.2887
4 -1.8388	0.0000	0.5000	-0.5000	0.0000	0.5000	-0.5000
3 -6.9317	0.5774	0.2887	-0.2887	-0.5774	-0.2887	0.2887
2 -6.9317	0.0000	0.5000	0.5000	0.0000	-0.5000	-0.5000
1 -14.8670	0.4082	0.4082	0.4082	0.4082	0.4082	0.4082

These form a normalized, orthogonal set. Under the NDDO approximation, overlaps between different atomic orbitals are ignored, i.e., $\langle \varphi_i | \varphi_j \rangle = \delta(i, j)$, so instead of

$$\langle \psi_i | \psi_j \rangle = \sum_{\lambda} \sum_{\sigma} c_{\lambda i} c_{\sigma j} \langle \varphi_{\lambda} | \varphi_{\sigma} \rangle$$

we have

$$\langle \psi_i | \psi_j \rangle = \sum_{\lambda} c_{\lambda i} c_{\lambda j} = \delta(i, j).$$

5.5.4 Exercises involving eigenvectors

In the following exercises 'verify' means using a hand calculator. They are intended to confirm understanding of the theory involved. Work through one or more examples to confirm the validity of the statement that follows.

1. Verify that the eigenvectors are normalized.
2. Verify that the eigenvectors are orthogonal to each other.

3. Verify that the eigenvalues are correct.
4. Verify that the eigenvectors diagonalize the Fock matrix.
5. Verify that the diagonal sum rule is obeyed; i.e., that the sum of the eigenvalues is equal to the sum of the diagonal matrix elements (the trace) of the Fock matrix.

5.5.5 Iterating density matrix

The density matrix is then reformed using the occupied set of eigenvectors, i.e., the lowest three levels. This yields:

		Density Matrix (eV)					
Atom		1	2	3	4	5	6
1		1.0000					
2		0.6667	1.0000				
3		0.0000	0.6667	1.0000			
4		-0.3333	0.0000	0.6667	1.0000		
5		0.0000	-0.3333	0.0000	0.6667	1.0000	
6		0.6667	0.0000	-0.3333	0.0000	0.6667	1.0000

Verify that the density matrix is correct.

5.5.6 Iterating Fock matrix

The second Fock matrix can then be constructed using this density matrix. The on-diagonal terms are identical to those in the first Fock matrix, since the atomic orbital electron densities are unchanged, but the off-diagonal terms are now changed. The off-diagonal terms are modified to allow for exchange interactions. (Note that not all exchange terms are stabilizing.)

Let us evaluate the matrix element $F(1,2)$:

$$F(1,2) = -3.2457 - \frac{1}{2}(0.6667)(9.6583)\text{eV}.$$

The second Fock matrix is thus:

		Second Fock Matrix (eV)					
Atom		1	2	3	4	5	6
1		-5.4823					
2		-6.4652	-5.4823				
3		-1.0970	-6.4652	-5.4823			
4		+0.3611	-1.0970	-6.4652	-5.4823		
5		-1.0970	+0.3611	-1.0970	-6.4652	-5.4823	
6		-6.4652	-1.0970	+0.3611	-1.0970	-6.4652	-5.4823

Diagonalization of this matrix yields the same set of eigenvectors as we had initially. In general, several iterations are necessary in order to obtain an SCF; however, a few systems exist for which symmetry restrictions on the form of the eigenvectors allow them to achieve an SCF in one iteration. Hexagonal H_6 is one such system. Although the eigenvectors are the same, the eigenvalues obviously have to be different.

Exercise: Verify that the SCF energy levels of H_6 are -20.2457, -11.2116, -11.2116, 2.4411, 2.4411, and 4.8929 eV.

Once an SCF is achieved we need to calculate the heat of formation.

5.5.7 Calculation of heat of formation

The heat of formation is defined as:

$$\Delta H_f = E_{elect} + E_{nuc} - E_{isol} + E_{atom},$$

where E_{elect} is the electronic energy, E_{nuc} is the nuclear-nuclear repulsion energy, $-E_{isol}$ is the energy required to strip all the valence electrons off all the atoms in the system, and E_{atom} is the total heat of atomization of all the atoms in the system.

E_{elect} is calculated from $\frac{1}{2}\mathbf{P}(\mathbf{H} + \mathbf{F})$, or

$$E_{elect} = \frac{1}{2} \sum_{\lambda=1}^6 \sum_{\sigma=1}^6 P_{\lambda\sigma} (H_{\lambda\sigma} + F_{\lambda\sigma}).$$

Using the data we have already derived, we can calculate E_{elect} as

$$\begin{aligned} E_{elect} &= 3(+1.0000)(-51.7124 + -5.4823) \\ &\quad + 6(+0.6667)(-3.2457 + -6.4652) \\ &\quad + 3(-0.3333)(-0.6992 + -0.3611) \end{aligned}$$

or

$$E_{elect} = -210.0898\text{eV}.$$

E_{nuc} is a relatively straightforward calculation, and is equal to 130.2902eV. The total energy of the system is thus -79.7996 eV.

We are now ready to calculate the ΔH_f . As the total energy and E_{isol} are in eV, we must first convert them into kcal/mol:

$$\Delta H_f = 23.061(-79.7996 + 71.4377) + 6(52.1020) \text{ kcal/mol}$$

or

$$\Delta H_f = 119.780\text{kcal/mol}.$$

It is convenient to combine E_{isol} and E_{atom} together, to simplify this calculation. In order to convert any total energy ($E_{elect} + E_{nuc}$) into a ΔH_f , the following operation must be performed:

$$\Delta H_f = 23.061(E_{elect} + E_{nuc} - \sum_i E_{(isol-atom)})$$

in which the index i is over all atoms in the system.

Users of MOPAC may wish to verify this calculation for a system of their own choice. To facilitate this, the data in Table 5.23 may prove useful.

These numbers may be used in conjunction with the semiempirical electronic and nuclear energies to calculate the heat of formation.

5.6 Localized Molecular Orbital Theory

5.6.1 Why Use Localized Molecular Orbitals?

Using conventional SCF methods, the time required for a SCF calculation rises as the third or higher power of the size of the system (number of atoms, orbitals, or electrons). For semiempirical methods, this places a practical limit on the number of atoms in a molecule: about 500–1,000, using a supercomputer.

Table 5.23: Values of $E_{(isol-atom)}$

Element	$E_{(isol-atom)}$ (eV/atom)			
	MINDO/3	MNDO	AM1	PM3
Hydrogen	-14.764312	-14.165588	-13.655739	-15.332633
Lithium		-6.793583		
Beryllium		-27.541992		
Boron	-67.584394	-70.200344	-69.601659	
Carbon	-126.880346	-127.910952	-128.226140	-118.640263
Nitrogen	-192.410048	-207.466249	-207.307791	-162.513823
Oxygen	-309.652672	-320.451178	-318.682192	-291.924879
Fluorine	-475.817831	-477.502913	-483.109715	-438.336301
Aluminum		-47.931017		-50.311708
Silicon	-95.576505	-87.539565	-83.701885	-72.488357
Phosphorus	-154.270388	-156.236921	-124.436836	-121.236135
Sulfur	-231.996798	-228.891710		-186.333060
Chlorine	-347.185366	-354.374768	-373.455532	-316.452049
Zinc		-31.231065		
Germanium		-80.129955		
Bromine		-347.840783	-353.473742	-353.699430
Tin		-95.454929		
Iodine		-341.704860	-347.970786	-289.422586
Mercury		-29.456154		
Lead		-107.856099		

The origin of the N^3 dependence lies in the way conventional molecular orbitals are generated and used. They are generated from a Fock matrix either by a diagonalization, or, if approximate M.O.s exist, by a pseudo-diagonalization. Both processes are of type N^3 . Once generated, they are used in the construction of a density matrix. Because every M.O. extends over every atom, this process is also of order N^3 . Most of the time in a semiempirical calculation is spent in these two operations, therefore the overall time dependency of conventional semiempirical calculations rises as N^3 .

To understand how limiting this dependency is, consider how the time requirement rises with increasing numbers of atoms. For the sake of discussion, assume that the time required for an SCF calculation of an amino acid is one second. Table 5.24 shows the time required for larger systems.

Many enzymes containing thousands of residues are known. It is obvious that these systems cannot be studied using conventional SCF methods. By using localized molecular orbitals, the time dependency can be changed from N^3 to N^1 . LMO methods require much more arithmetic manipulation, in that every array which holds atomic orbital information has one or more associated integer arrays which specify the atoms involved and their location. For this reason, LMO methods are much slower than conventional methods for small systems.

What is a Localized Molecular Orbital?

Unlike a conventional molecular orbital, which extends over all atoms, a localized molecular orbital is localized in a small region of the molecule. Almost all of a localized M.O. can be found on one, two, or, at most, three atoms. The nature and behavior of the LMO is dictated by these few atoms. Small amounts of the LMO can be found on the surrounding atoms. Although the effect on these small contributions on the nature of the LMO is very small, their existence is responsible for almost all of the time spent on mathematical manipulation of the LMOs.

Table 5.24: Predicted Times Required for One SCF Calculation Using Conventional SCF Methods*

No. of Residues	Time	
2	8.0	seconds
5	2.1	minutes
10	16.7	minutes
20	2.2	hours
50	1.4	days
100	11.6	days
200	3.1	months
500	4.0	years
1,000	31.7	years
2,000	253.6	years

*: Using Conventional SCF Methods, and assuming that one amino acid runs in one second.

Energy Considerations.

The electronic energy of a system is given by

$$E_{elect} = \frac{1}{2} \sum_{\lambda} \sum_{\sigma} P_{\lambda\sigma} (H_{\lambda\sigma} + F_{\lambda\sigma}) \quad (5.24)$$

where $P_{\lambda\sigma}$ is the density matrix element connecting atomic orbitals ϕ_{λ} and ϕ_{σ} .

In practice, the density matrix elements rapidly become small as the interatomic distance increases. Only for atoms which are chemically bonded together will $P_{\lambda\sigma}$ be large. As an example, in propanolamine, the smallest bond order between two chemically bonded atoms is 0.98. Between any two neighboring but non-chemically bound atoms (that is, atoms separated by one atom), the largest bond order is 0.015, and for all other interactions, the bond-orders are very small.

Because the electronic energy is determined by the density matrix, the contribution to the electronic energy arising from density matrix elements between atoms which are not near to each other is small, and becomes very small with increasing distance. The contributions from atoms separated by more than about 10 Ångstroms is quite negligible, and can safely be ignored. From this it follows that the only density matrix terms which need to be considered are those arising from atoms separated by less than 10 Ångstroms. Since the density matrix is constructed from the M.O.s only, then, in LMO theory, the LMOs need only extend over a distance of about 10 Ångstroms from the center of the LMO.

Limitations on Systems.

The use of localized molecular orbitals to help solve the self-consistent field equations is limited to what we will call 'normal' compounds. In order for a system to be classified as 'normal', it must be possible to sketch the molecular structure using only the following drawing elements:

1. Chemical symbols to represent the atoms.
2. A maximum valency of 4 for any non-hydrogen atom.
3. Lines to represent σ bonds.
4. Pairs of parallel lines to represent double bonds, and three parallel lines to represent triple bonds.

Table 5.25: MOPAC Memory Requirements for Proteins

N	No. [†] of Residues	Memory Needed (Mb)
1000	65	7,414
2000	131	29,616
3000	196	66,609
4000	262	118,389
5000	327	184,961

†: Number of residues is approximate only; the number will vary from protein to protein

- Pairs of dots to represent lone-pairs.
- '+' and '-' signs, to represent charges.

This definition allows aromatic rings and delocalized π systems to be present, as these structures can be represented as localized π bonds.

What is *not* allowed are radicals, open-shell systems, electronic excited states, etc. Hypervalent systems are allowed, as they can always be written as non-hypervalent Lewis structures, for example, SF_6 can be submitted as a system of point group O_h , in which case the Lewis structure generated would be represented by SF_4^{++} plus 2F^- .

5.6.2 Memory Management

The Nature of the Problem.

Conventional semiempirical software is unsuitable for the calculation of large systems, due to the heavy memory demands made. For example, in MOPAC 93, the memory requirement (in Kb) rises as $2249 + 17.15N + 7.395N^2$, where N is the number of heavy (non-hydrogen) atoms in the molecule, assuming that the number of hydrogen and non-hydrogen atoms are equal. In proteins, the number of hydrogen and non-hydrogen atoms are roughly equal, so the expression for MOPAC 93 would be applicable to proteins. Table 5.25 illustrates how rapidly the memory requirement increases for proteins.

Clearly, calculation of systems of only a few hundred residues is impractical. By making approximations, by altering the way arrays are specified, and by use of localized molecular orbitals, the memory requirements can be reduced. As a result of these changes, the memory requirement for MOZYME is considerably less than that for MOPAC. For example, a MOPAC calculation of a system of 3,686 atoms (bacteriorhodopsin) would require over 100Gb; the same calculation using MOZYME would require only 408Mb. This represents less than 0.5% of the memory required by MOPAC. A more detailed description of array specification appears elsewhere.

Distance Cutoffs.

In conventional SCF calculations, all interactions regardless of distance are calculated. This is both impractical and unnecessary in LMO work. Impractical, because the storage required would rapidly become very large: the storage necessary for merely the two-electron two-center integrals of a system of M non-hydrogen and N hydrogen atoms would be $50M(M-1) + 10M \times N + \frac{1}{2}N(N-1)$ array elements. Unnecessary, because many of the integrals would never be used in the LMO calculation, anyway. Consider the electronic energy terms. These all depend on the density matrix elements. If a density matrix element, $P_{\lambda\sigma}$ is zero, then the associated energy

Table 5.26: Bond-Orders in GLY-GLY-GLY-GLY

Bond to	Order	Bond to	Order
C ₂	1.017090	C ₈	0.000009
C ₃	0.015249	C ₉	0.000001
N ₄	0.004989	N ₁₀	0.000000
C ₅	0.001219	C ₁₁	0.000000
C ₆	0.000086	C ₁₂	0.000000
N ₇	0.000031		

Bonds are between the terminal nitrogen and the backbone atoms.

GLY-GLY-GLY-GLY has the structure: H₂N-CH₂-CO-NH-CH₂-CO-NH-CH₂-CO-NH-CH₂-COOH.

Table 5.27: Multipolar Representation of the Two Electron Two Center Integrals

Multipole	Example	No.	Distance Dependence	
			Energy	Gradient
Monopole	$\langle ss ss\rangle$	16	r^{-1}	r^{-2}
Dipole	$\langle ss sx\rangle$	24	r^{-2}	r^{-3}
Quadrupole	$\langle ss xy\rangle$	33	r^{-3}	r^{-4}
Octapole	$\langle sx xy\rangle$	18	r^{-4}	r^{-5}
Hexadecapole	$\langle xy xy\rangle$	9	r^{-5}	r^{-6}

term is independent of $H_{\lambda\sigma}$ and $F_{\lambda\sigma}$. Similarly, the Fock element $F_{\lambda\sigma}$ depends on $P_{\lambda\sigma}$; therefore, if $P_{\lambda\sigma}$ is zero, then $F_{\lambda\sigma}$ is independent of the value of the exchange integral $\langle \lambda\sigma | \frac{e^2}{r} | \lambda\sigma \rangle$.

The bond-order is a measure of the electron density between two atoms. Table 5.26 shows the bond-orders between the terminal nitrogen and the backbone atoms in the tetrapeptide GLY-GLY-GLY-GLY. For atoms separated by more than 3 - 4 bonds, that is, by more than 5 - 7 Ångstroms, the bond-orders rapidly become very small. Also, at such distances the overlap of two atomic orbitals becomes extremely small. Since the one-electron interactions depend on the overlap

$$H_{\lambda\sigma} = S_{\lambda\sigma}(\beta_\lambda + \beta_\sigma), \quad (5.25)$$

where $S_{\lambda\sigma}$ is the atomic orbital overlap, and β_λ and β_σ are atomic orbital parameters, it follows that the one-electron interactions at distances greater than about 6–7 Ångstroms will also be vanishingly small.

By calculating and storing only the one-electron integrals representing interactions of less than a given distance, a considerable saving in storage requirements can be achieved.

Unlike the one-electron integrals, the value of some two-electron two-center integrals does not fall off rapidly with increasing distance. The 100 two-electron integrals involving two atoms can be divided into five groups, as shown in Table 5.27. Some of these integrals, such as the $\langle pp|pp\rangle$ integrals, can be expressed in terms of two multipoles, here a monopole and a quadrupole; however, for the purpose of this discussion the effect of the lower multipole—the monopole in this case—dominates.

At distances greater than about 5 or 6 Ångstroms, the behavior of these integrals becomes very simple. Of the 100 integrals, 60 represent quadrupoles or higher multipoles and have a negligible value. The 16 monopole terms – representing simple electrostatic repulsion – are all composed of the same monopole term and various quadrupolar components, and can be set equal. The 24 dipole terms can be expressed as a dipole interacting with either a simple monopole (an $\langle ss|sx\rangle$, for example), or a monopole plus quadrupole, e.g., $\langle xx|sx\rangle$. Since the quadrupolar terms can be ignored, the 24 dipolar terms can be represented by 6 simpler terms, of generic form $\langle sp|ss\rangle$ and $\langle ss|sp\rangle$. Thus, out of the 100 integrals needed at small distances, only 7 are needed at larger

Table 5.28: Distance Cutoffs for Various Types of Interaction

Type	Less than Cutoff	Greater than Cutoff	Cutoff Name	Default
Two electron	All multipoles	Dipole plus Monopole	CUTOF2	12Å
Two electron	Dipole plus Monopole	Monopole	CUTOF2	30Å
One electron	Overlap calculated	Overlap not calculated	CUTOF2	7Å

distances. At still larger distances, 30 Ångstroms or more, even the dipolar terms become negligible. Therefore, at such large distances only the single monopole term, representing simple electrostatic repulsion, need be used.

Two-electron two-center integrals must therefore be represented in a different way from the one-electron two-center integrals. As with the one-electron integrals, a cutoff distance can be specified. For convenience, the one-electron and two-electron cutoff distances are set equal. As a result, all one- and two-electron integrals which represent interactions between atoms that are separated by less than the cutoff distance would then be treated using standard NDDO approximations. The simplest electrostatic repulsion ($\langle ss|ss \rangle$) between electrons on atoms that are separated by more than the cutoff distance is approximated by the conventional NDDO term

$$\langle \phi_\lambda \phi_\lambda | e^2 / r_{AB} | \phi_\sigma \phi_\sigma \rangle = \frac{1}{\left(R_{AB}^2 + \frac{1}{2} \left(\frac{1}{\langle \phi_s \phi_s | e^2 / r_{AA} | \phi_s \phi_s \rangle} + \frac{1}{\langle \phi_s \phi_s | e^2 / r_{BB} | \phi_s \phi_s \rangle} \right) \right)^{\frac{1}{2}}}. \quad (5.26)$$

This was chosen in order to minimize the discontinuity at the cutoff distance. Similar NDDO approximations were used for the other dipolar terms.

In recognition of this fact, SCF calculations can be simplified considerably by including only those NDDO terms which apply to atoms separated by less than a preset limit, and including only the electrostatic term for interactions between atoms separated by more than that limit.

The various cutoffs used have specific names and defaults. These are summarized in Table 5.28.

By default, the cutoff distances are set sufficiently large that any ΔH_f calculated using these values will agree with that obtained by using conventional methods within 0.1 kcal/mol. CUTOF2 can be reduced, under user control, to save computational time and to reduce the memory demand. To avoid serious errors, CUTOF2 should not be set below about 5 Ångstroms.

Once CUTOF2 is set, all one-electron integrals involving atoms separated by less than CUTOF2 can be evaluated. To save space, only those integrals which are evaluated are stored in an array. For the one-electron integral, this array is called *H*. All integrals relating to any pair of atoms are stored contiguously. Given the starting address of an atom-pair, the sequence in which the integrals occur is determined only by the number of atomic orbitals on each atom. However, the order in which atom pairs are stored is not so simple. The only way to find the starting address of any atom pair is to have that address stored in a new array. This array is called IJBO, and has the following structure:

IJBO is a square array, of size *N* by *N*, where *N* is the number of atoms. The starting address of integral string for atoms *A* and *B* is stored in array element IJBO(A,B). If $A \neq B$, then the starting address is also stored in IJBO(B,A). All pairs of atoms separated by more than CUTOF2 do not have associated integral strings. To recognize this, the relevant array elements in IJBO are set to '-1', and if the interatomic distance is greater than CUTOF1 then the array element is set to '-2'. Zero cannot be used, because of (see Page 178).

To reiterate: All integrals involving atoms, *A* and *B*, separated by less than CUTOF2 are stored, and the starting address of the integrals is given by IJBO(A,B) and IJBO(B,A). With the exception of some electrostatic terms, no integrals involving atoms separated by more than CUTOF2 are calculated or stored. The relevant array element in IJBO is set to '-1' or '-2'.

To summarize:

There are three regions around each atom. The sizes of these regions are determined by $CUTOF2=m.mm$ (default: $m.mm=12$) and $CUTOF1=n.nn$ (default $n.nn=30$). In the first region, 0.0 to $m.mm$ Ångstroms, all NDDO interactions are used. Between $m.mm$ and $n.nn$ Ångstroms, only monopolar and dipolar electrostatic terms are used, and beyond $n.nn$ Ångstroms, the only term considered is the monopolar electrostatic interaction.

5.6.3 Lewis Structures.

One of the more difficult operations involved in using localized molecular orbitals is the generation of a Lewis structure. While this operation is almost trivially easy for a competent chemist, setting up the instructions so that a program can do the same operation has proved to be a daunting task.

In the following section, the steps involved in calculating the Lewis structure are described. This description is intended to be definitive, in the sense that it should allow the Lewis structure for *any* compound to be generated. At the same time, any deficiency in the description should be reflected in the inability of MOZYME to generate the Lewis structure for certain systems.

Because of its complexity, the main sequence involved will be given first, followed by a more detailed explanation of the individual steps.

Lewis Structure—Main Sequence.

1. The connectivity is calculated. This determines which atom is bonded to which atom.
2. All atoms that have explicit charges are identified, and the charges assigned.
3. The σ framework is determined.
4. Most of the lone pairs are identified.
5. Open-ended (non-aromatic) π -bonds are identified.
6. Aromatic π -bonds are identified.
7. All cations, anions, and any remaining lone pairs are identified.

Detailed Description of Lewis Structure.

Calculation of Connectivity

Hydrogen atoms are monovalent. Because of this, they can only bond to one other atom. Therefore, the first set of bonds formed are the X–H σ bonds. The criterion used is that each hydrogen atom is connected to the atom nearest to it, except that a hydrogen atom is not allowed to be bonded to another hydrogen atom.

The connectivity of all other atoms is determined. Atoms are considered as being connected (bonded together) if the interatomic distance is less than 110% of the sum of their Van der Waals radii.

Any bridging hydrogen bonds are identified. These usually indicate a faulty geometry. If any are present, then the SCF calculation will not be run, unless LET is present.

Any user-defined chemical bonds are identified. This is useful in cases where a Lewis structure could not otherwise be created. An example is the simple system HNO_3 .

Finally, the number of atoms bonded to each atom is checked. Conventional Lewis structures do not allow more than four bonds to each atom (assuming an sp^3 basis set), so if there are more than four atoms bonded to any atom, a conventional Lewis structure cannot be generated, and the SCF cannot be run. Before abandoning the calculation, an attempt is made to reduce the number of bonds to a hypervalent

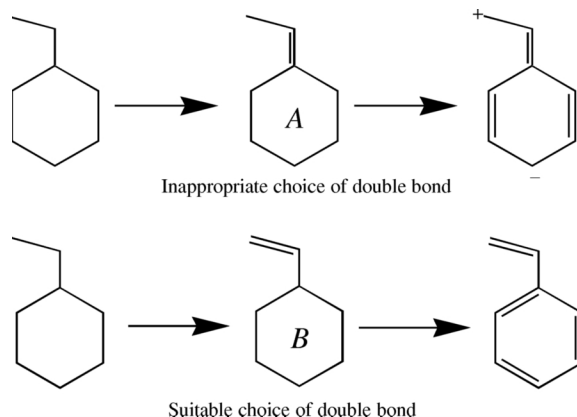


Figure 5.6: Examples of Initial Choice of Double Bond

atom. First, any bonds from a hypervalent atom to a halogen are broken. If this makes the hypervalent atom normal, then the calculation can proceed. If that does not work, then bonds to elements of group VI are broken. If that still does not make the atom normal-valent, then the calculation is stopped.

σ -framework

This is the simplest set of bonds to identify. A σ bond exists for every pair of atoms that are connected. The number of σ bonds is equal to the number of connections in the system. For water, this would be 3; for benzene, 12; and for ethylene, 5.

Each time a bond is formed, the number of available atomic orbitals on the atoms involved is decremented by 1 and the number of available electrons is decremented by 1. For example, before the σ framework of ammonia is formed, the number of available orbitals on nitrogen is 4 and the number of electrons is 5. After the σ bonds are formed, there is 1 orbital and 2 electrons left.

Lone pairs

The next set of Lewis elements formed are the lone pairs. The rule used here is that if there are more electrons than orbitals on an atom, the extra electrons are used in the construction of lone pairs. Each lone pair uses up two electrons and one orbital. Thus, one lone pair would be assigned to the nitrogen in ammonia, two lone pairs would be assigned to oxygen in water, and three lone pairs would be assigned to chlorine in HCl.

The lone pairs could only be assigned after the σ bonds were created. If they were assigned before the σ bonds, then some ionic systems, such as NH_4^+ , could not be represented.

Open-ended π -bonds

Once the lone pairs are assigned, the number of unused atomic orbitals on an atom will be equal to the number of unused electrons on that atom. These unused orbitals are then available for forming multiple bonds between atoms. If two atoms, that are σ bonded together, both have unused orbitals, then they can form a π bond.

The order in which π bonds are generated is important. If in styrene, for example, a π bond is assigned to the ring- C_1 as in Figure 5.6, *A*, then when the remaining double bonds are created, there are two unused atomic orbitals on atoms that are not bonded together. In order for a Lewis structure to be generated, two electrons are put into one of these unused orbitals, creating an anionic center, and no electrons are put into the other orbital, making it a cationic center.

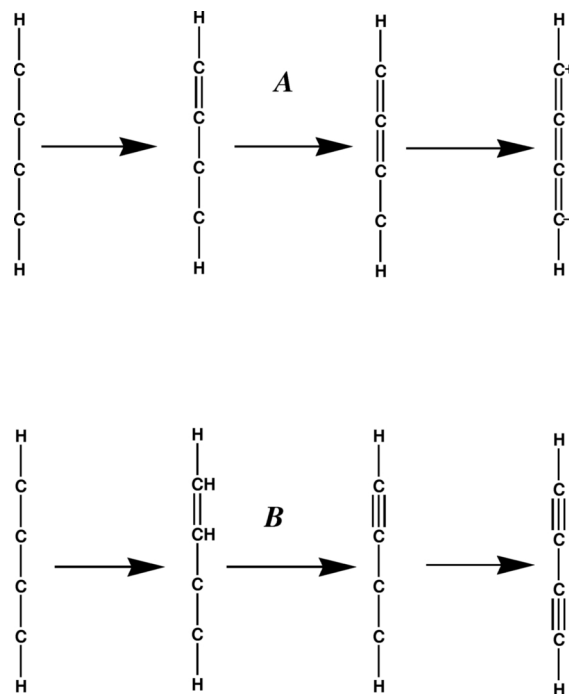


Figure 5.7: Generation of Yne Bond

A better choice is to identify open-ended π systems, and to assign these first, option *B* in Figure 5.6.

The order in which the π bonds are assigned in an open-ended π system is important. In the case of a simple conjugated polyene, the order is simple - the carbon atom that has only one atom π bonding to it is identified. A π bond is constructed between the two atoms. This is repeated until all π bonds in the polyene are identified.

Problems arise in more complicated systems, such as buta-1,3-diyne. If the simple rule just described is used, then a zwitterionic cumulated polyene results, Figure 5.7 *A*, instead of a diyne, Figure 5.7 *B*.

Also, if both ends of the olefinic group are connected to aromatic rings, as in stilbene, then identification of the olefin group is not obvious.

To allow for this, the following two rules are used:

1. Where there is the possibility of forming a triple bond, do so.
2. When a delocalized π system is opened, the π bond formed should involve atoms that are π bonded to exactly two other atoms.

In the case of stilbene ($\phi-\text{CH}=\text{CH}-\phi$), this prevents a π bond forming between the ring and a carbon atom of the olefin.

Whenever a delocalized π system is encountered, as soon as a π bond is formed and the delocalization destroyed, then the rest of the π system is treated as a simple conjugated polyene. This ensures that the maximum number of π bonds is formed, and prevents unconnected π bonds from being created. Thus benzene would have the three π bonds: C_1-C_2 , C_3-C_4 , and C_5-C_6 , and not the quinoidal C_1-C_2 and C_4-C_5 bonds.

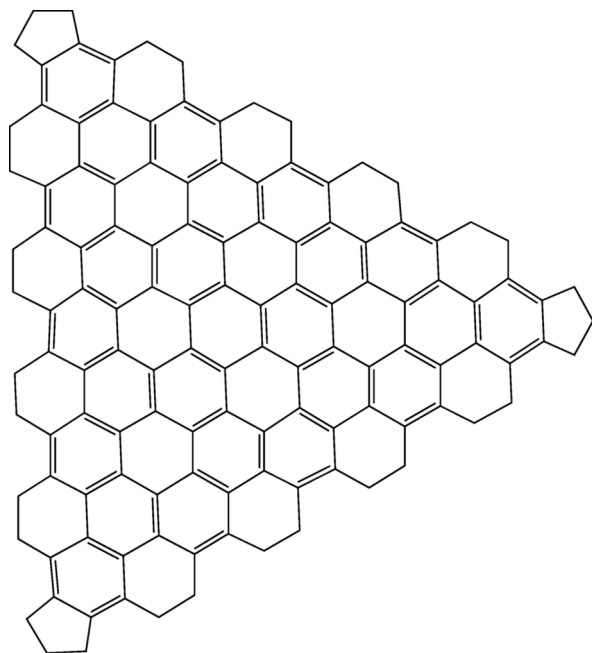


Figure 5.8: Kekule Structure for a Facet of Fullerene C₁₅₀₀

To allow compounds that contain several fused delocalized π -systems, such as the higher buckyballs (specifically, C₉₆₀) to be modeled, two extra rules are needed. These rules can be regarded as minor qualifications to the earlier rules:

1. If any atoms attached to an open-ended π system belong to a delocalized π system, then when delocalized π systems are opened, the opening is done using these atoms.
2. If a five-membered π ring is attached to another delocalized π system, then when the π system is opened, at least one atom must be in the other delocalized system.

The effect of the first rule is that as soon as a graphitic network is encountered, all the π bonds in the network are assigned in one pass. Without this rule, individual parts of the network could be assigned separately, and at the junctions of the various domains, the potential for isolated π orbitals exists. This would lead to charges that would cause severe problems with the SCF calculation.

For extended graphite-like systems, two more rules are needed. These are:

1. If a six-membered ring has two π bonds already, then add a third π bond to make it an aromatic ring.
2. If a six-membered ring has one π bond, then add another π bond to the same ring, so that the previous rule can be used.

The effect of these rules is that when a graphitic lattice is encountered, all the atoms in the lattice will be assigned in such a way as to maximize the number of aromatic rings. An example of such a system is provided by the large icosahedral fullerene C₁₅₀₀. A facet of this system is shown in Figure 5.8.

Remaining unused atomic orbitals

All that remains is to identify any unused orbitals and to assign them to either the occupied or virtual sets. The action taken depends mainly on the group in the periodic table to which the atom belongs, to a lesser degree on the nature of its environment, and sometimes on the charge on the system.

Each group has different properties. Thus:

Group I

The alkali metals are extremely electropositive. Therefore, without exception, the unused orbital is assigned to the virtual set, and the atom is charged unipositive.

Group II

The alkaline earth elements are very electropositive. If the atom does not form any bonds, then the charge is set to +2, and all orbitals are assigned to the virtual set.

An atom that has formed N bonds will have a charge of $2-N$. This can allow the atom to have a negative charge, in which case the data set should indicate that the atoms attached to it have a positive charge.

Group III

These elements are electropositive. The action taken depends on the number of unused valence electrons:

1 unused electron The atom is unipositive.

2 unused electrons The atom is neutral: the two unused electrons are used in the formation of a 'lone pair'.

3 unused electrons The atom is unipositive.

Group IV

This is the most complicated group, with the charge on the atom depending on many factors. If there are two unused valence electrons, then the atom will be neutral (a carbene, for example). Otherwise, if the first or second nearest neighboring atom is of Group 6, then the atom is assigned a negative charge; if the first or second nearest neighboring atom is of Group 5, then the atom is assigned a positive charge. If the charge is still not determined, then the assignment of charge is deferred until all other atoms have been assigned. At that point, the charge is assigned as either +1 or -1, depending on the calculated charge on the system and the charge supplied by the data set. The charges are assigned so that the calculated charge equals the supplied charge.

Group V

If there is one unused valence electron, then the charge assigned to Group V elements is +1. If there are 2, then the atom is neutral.

Group VI

If oxygen and there are two unused orbitals, then both are filled, and the charge is -2. If the atom has one unused orbital, the charge is -1; otherwise the atom is neutral.

Group VII

A very electronegative group, the charge is invariably -1.

5.6.4 Construction of Starting Localized Molecular Orbitals

In conventional semiempirical methods, molecular orbitals are created by the diagonalization of an atomic orbital secular determinant. Because of the size of the matrices involved in proteins, this operation is impractical; therefore the starting orthogonal localized molecular orbitals must be constructed in a different way.

Monatomic and diatomic LMOs are constructed using all the atomic orbitals in the system. Care must be taken to ensure that a realistic set of starting LMOs are generated. If the set is not realistic, then, although an SCF could be calculated, more work would be involved. The most obvious set of LMOs is that set corresponding to the classical line-drawing of a molecule: a sigma framework, lone pairs, π -bonds, and charges, localized on specific

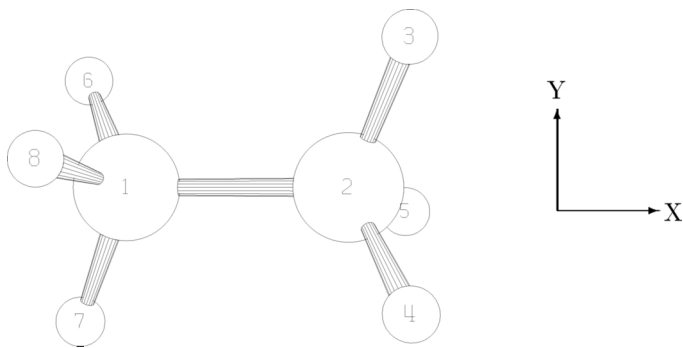


Figure 5.9: Ethane, Showing Orientation and Numbering System

atoms. Without loss of rigor, delocalized π systems can be represented as localized π bonds. The purpose in constructing the mono and di-atomic LMOs is to have a starting set of LMOs, which can then be used as the basis for a self-consistent set of LMOs.

The steps involved in constructing the LMOs are as follows:

1. Hybrid atomic orbitals are constructed for each heavy atom.
2. All sigma bonds are identified.
3. Diatomic σ LMOs are constructed.
4. All lone pairs are identified, and monatomic LMOs constructed.
5. All non-cyclic π systems are identified, and diatomic π LMOs bonds constructed.
6. All cyclic π systems are identified, and diatomic π LMOs constructed.
7. Any unused atomic orbitals are identified as cationic or anionic centers, and monatomic LMOs constructed.

Inspection of this sequence shows that the resulting set of LMOs will, in fact, correspond to the classical line-sketch of the molecular structure.

There are one or two less than obvious details involved in this sequence, which will now be described.

Construction of Hybrids

With the exception of hydrogen, all atoms have an $s-p$ basis set, which must be mixed to form four hybrid atomic orbitals. Ideally, the main lobes of the hybrid orbitals involved in σ bonding would point towards the other atoms involved in those σ bonds. To achieve this, information regarding the immediate environment of the atom must be used. As only the direction of the neighboring atoms is important, the energy terms relating to the one-electron integrals between the atom and the s orbital of the neighboring atoms can be used.

The process of using these energy terms to generate hybrid orbitals involves two steps. First, a set of M.O.s which involve the atomic orbitals of an atom and the s orbitals of its immediate neighbors is constructed by diagonalizing the interaction matrix. Consider a carbon atom in ethane. Using the numbering scheme and orientation shown in Figure 5.9, the complete MNDO one-electron matrix for ethane is shown in Table 5.29.

The one-electron matrix is used in the construction of the interaction matrix. This matrix is of size $4+n$, where n is the number of neighbors, i.e., $n = 1, 2, 3,$ or 4 . The first four functions are the s and p atomic orbitals of the atom; the remaining functions are the s orbitals of the neighbors. The only non-zero terms are those representing

Table 5.29: One-Electron Integral Matrix for Ethane

A.O.	$C(1)_s$	$C(1)_{p_x}$	$C(1)_{p_y}$	$C(1)_{p_z}$	$C(2)_s$	$C(2)_{p_x}$	$C(2)_{p_y}$
$C(1)_s$	-126.7891						
$C(1)_{p_x}$	-6.5701	-114.2822					
$C(1)_{p_y}$	0.0000	0.0000	-111.8502				
$C(1)_{p_z}$	0.0000	0.0000	0.0000	-111.8502			
$C(2)_s$	-5.3977	-4.2826	0.0000	0.0000	-126.7891		
$C(2)_{p_x}$	4.2826	2.4743	0.0000	0.0000	6.5701	-114.2822	
$C(2)_{p_y}$	0.0000	0.0000	-1.1990	0.0000	0.0000	0.0000	-111.8502
$C(2)_{p_z}$	0.0000	0.0000	0.0000	-1.1990	0.0000	0.0000	0.0000
H(1)	-1.0116	-0.6430	-0.3460	0.0000	-5.7834	-1.2013	-3.0984
H(1)	-1.0116	-0.6430	0.1730	-0.2996	-5.7834	-1.2013	1.5492
H(1)	-1.0116	-0.6430	0.1730	0.2996	-5.7834	-1.2013	1.5492
H(1)	-5.7834	1.2013	-1.5492	2.6833	-1.0116	0.6430	-0.1730
H(1)	-5.7834	1.2013	3.0984	0.0000	-1.0116	0.6430	0.3460
H(1)	-5.7834	1.2013	-1.5492	-2.6833	-1.0116	0.6430	-0.1730

A.O.	$C(2)_{p_z}$	H(1)	H(2)	H(3)	H(4)	H(5)	H(6)
$C(2)_{p_z}$	-111.8502						
H(1)	0.0000	-99.6727					
H(1)	2.6833	-0.9455	-99.6727				
H(1)	2.6833	-0.9455	-0.9455	-99.6727			
H(1)	0.2996	-0.2444	-0.0814	-0.2444	-99.6727		
H(1)	0.0000	-0.0814	-0.2444	-0.2444	-0.9455	-99.6727	
H(1)	0.2996	-0.2444	-0.2444	-0.0814	-0.9455	-0.9455	-99.6727

Table 5.30: Interaction Matrix used in Constructing Hybrid Orbitals for C_1 in Ethane

A.O.	$C(1)_s$	$C(1)_{p_x}$	$C(1)_{p_y}$	$C(1)_{p_z}$	$C(2)_s$	H(1)	H(2)	H^3
$C(1)_s$	0.0000							
C_{p_x}	0.0000	0.0000						
C_{p_y}	0.0000	0.0000	0.0000					
C_{p_z}	0.0000	0.0000	0.0000	0.0000				
$C(2)_s$	-5.3977	-4.2826	0.0000	0.0000	0.0000			
H(1)	-5.7834	1.2013	-1.5492	2.6833	0.0000	0.0000		
H(1)	-5.7834	1.2013	3.0984	0.0000	0.0000	0.0000	0.0000	
H(1)	-5.7834	1.2013	-1.5492	-2.6833	0.0000	0.0000	0.0000	0.0000

Table 5.31: Eigenvectors of Lowest Energy from Interaction Matrix

Ψ	$C(1)_s$	$C(1)_{p_x}$	$C(1)_{p_y}$	$C(1)_{p_z}$	$C(2)_s$	H(1)	H(2)	H(3)
1	0.7069	0.0150	0.0000	0.0000	0.3409	0.3577	0.3577	0.3577
2	-0.0150	0.7069	0.0000	0.0000	0.6195	-0.1968	-0.1968	-0.1968
3	0.0000	0.0000	0.6112	0.3556	0.0000	-0.0019	-0.4990	0.5009
4	0.0000	0.0000	-0.3556	0.6112	0.0000	-0.5773	0.2903	0.2870

Table 5.32: Localized Diatomic Molecular Orbitals for C₁ in Ethane

Ψ	C(1) _s	C(1) _{p_x}	C(1) _{p_y}	C(1) _{p_z}	C(2) _s	H(1)	H(2)	H(3)
1	0.3277	0.6266	0.0000	0.0000	0.7071	0.0000	0.0000	0.0000
2	-0.3617	0.1892	-0.2887	0.5000	0.0000	-0.7071	0.0000	0.0000
3	0.3618	-0.1892	0.2887	0.5000	0.0000	0.0000	0.0000	0.7071
4	0.3618	-0.1892	-0.5774	0.0000	0.0000	0.0000	0.7071	0.0000

Table 5.33: Hybrid Atomic Orbitals for C₁ in Ethane

Hybrid	C(1) _s	C(1) _{p_x}	C(1) _{p_y}	C(1) _{p_z}
1	0.46345	0.88612	0.00002	-0.00004
2	-0.51159	0.26761	-0.40825	0.70711
3	0.51161	-0.26756	0.40825	0.70711
4	0.51161	-0.26756	-0.81650	0.00000

the interaction of the atom with its neighbors. Therefore, for the first carbon in ethane, the interaction matrix would be that shown in Table 5.30.

The four eigenvectors of lowest energy (Table 5.31) are used to form the hybrid orbitals. A simple unitary transformation converts these eigenvectors into localized di-atomic bonds (Table 5.32). This transformation is the localization procedure, which is very rapid. Finally, before the hybrid functions can be used, they must be normalized to unity. For C₁ in ethane, the final hybrid atomic orbitals are presented in Table 5.33.

This simple procedure works not just for atoms of sp^3 hybridization. Consider an sp^2 system. Now, with only three ligands, the interaction matrix for the atom will be of order seven. The hybrid orbitals are constructed from the lowest four eigenvectors resulting from diagonalization of this matrix. As with the sp^3 system, the hybrid orbitals are first localized, then re-normalized. This gives four hybrid orbitals, three of which are the sp^2 hybrids, with the fourth being a pure p orbital.

This procedure is quite general, as can be illustrated by oxygen, first in a hydroxy and then in a double-bonded oxygen environment. In a hydroxy environment an oxygen will have two ligands: so the interaction matrix will be of order six. Two sp hybrids and two p orbitals will be created. Later on, we will see that the two p orbitals become lone-pairs.

In the case of the double-bonded oxygen, the interaction matrix is of order five. The lowest eigenvector of this matrix corresponds to the oxygen-ligand σ -bond, and consists of a mixture of oxygen s and p orbitals. To satisfy orthogonality requirements, two other $s-p$ hybrids are also made, so the overall hybridization is three sp^2 orbitals plus one pure p orbital. As with the hydroxy oxygen, two of these orbitals will eventually be used to create lone pairs.

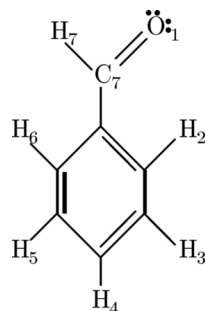
When there is only one or two atoms bonded to an atom, the non-bonding hybrid orbitals generated are ill-defined by a unitary transform. This ill-definition is unimportant in that at self-consistency the density matrix is independent of the starting LMOs. It is important, however, in that the starting LMOs are very sensitive to minute numerical differences, of the type encountered on porting the program to different platforms.

To prevent this, the ill-definition is removed by adding an extra constraint, subroutine MINLOC. If the atom has only one ligand, oxygen in carbonyl, for example, then the three non-bonding degenerate hybrid orbitals are mixed so that the p_x and p_y orbitals on one hybrid have zero coefficients, and the p_x orbital on another hybrid also has a zero coefficient.

If the atom has two ligands, then the two non-bonding degenerate hybrids are mixed so that p_x orbital on one

Table 5.34: Atomic Orbital and Electron Contributions in Benzaldehyde

Quantity	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	H ₂	H ₃	H ₄	H ₅	H ₆	C ₇	O ₁	H ₇
Orbitals	4	4	4	4	4	4	1	1	1	1	1	4	4	1
Electrons	4	4	4	4	4	4	1	1	1	1	1	4	6	1



Benzaldehyde - Electronic Structure

Number of Atomic Orbitals	38
Number of Electrons	40
Number of Diatomic Bonds	18
Number of Lone Pairs	2

Figure 5.10: Electronic Structure of Benzaldehyde

of the hybrids has a zero coefficient.

Hybrid atomic orbitals form an orthonormal set. This is a natural consequence of the fact that the hybrids are obtained by a unitary transform of the starting orthonormal atomic orbitals.

Construction of Starting Localized Molecular Orbitals

In order to illustrate the procedure for generating starting LMOs, we will use benzaldehyde as the test example. The line-sketch of benzaldehyde is shown in Figure 5.10. Before starting the procedure, let us spend a little time examining the sketch. From the empiric formula of benzaldehyde, C₇H₆O, we see that there are $(7 \times 4 + 6 \times 1 + 1 \times 6) = 40$ valence electrons, and $(7 \times 4 + 6 \times 1 + 1 \times 4) = 38$ valence orbitals. Since each diatomic bond involves two electrons and two orbitals, and each lone-pair involves two electrons and one orbital, we see that there must be 18 diatomic bonds and two lone pairs. The diatomic bonds can be divided into three sets: 14 σ bonds, one localized π bond and three π bonds in the aromatic ring. Every atom contributes both atomic orbitals and electrons to the system. These contributions are shown in Table 5.34.

Construction of Sigma Framework

Because it is the simplest set of bonds to identify, the σ framework is constructed first. σ bonds are defined as existing between every pair of atoms which are chemically bonded together. In our example, there are 14 σ bonds. Each di-atomic σ bond is constructed by identifying which pair of unused hybrid orbitals on the atoms involved has the largest interaction energy. When one of the atoms is a hydrogen, only the hybrid orbitals on the other atom need to be examined at this point.

Once the appropriate orbitals have been identified, bonding and antibonding diatomic LMOs are constructed. This is accomplished by diagonalizing the associated two by two secular matrix. The elements of this matrix are shown in Figure 5.11.

For every σ bond formed, two electrons are used, one from each of the atoms involved in forming the bond, and two atomic orbitals are used - again, one from each atom. After all 14 σ bonds are formed, the number of still unused atomic orbitals and electrons is much reduced. The remaining unused atomic orbitals and electrons are

Figure 5.11: Secular Equation used in Constructing Starting LMOs

$$C \begin{pmatrix} \langle \text{Hybrid}_1 | H | \text{Hybrid}_1 \rangle - E_{1,1} & \langle \text{Hybrid}_1 | H | \text{Hybrid}_2 \rangle \\ \langle \text{Hybrid}_2 | H | \text{Hybrid}_1 \rangle & \langle \text{Hybrid}_2 | H | \text{Hybrid}_2 \rangle - E_{2,2} \end{pmatrix} C^t = 0$$

Table 5.35: Unused Electrons and Orbitals after Construction of σ Framework

Quantity	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	H ₂	H ₃	H ₄	H ₅	H ₆	C ₇	O ₁	H ₇
Orbitals	1	1	1	1	1	1	0	0	0	0	0	1	3	0
Electrons	1	1	1	1	1	1	0	0	0	0	0	1	5	0

shown in Table 5.35. The oxygen forms only one σ bond, so there are still three unused hybrid orbitals and five valence electrons left.

Identification of Lone Pairs

After the σ framework has been constructed, all lone pairs are identified. They could not be identified earlier, because sometimes a lone pair may be used up in making the σ framework. An example of this would be the lone pair on an $\text{-}\ddot{\text{N}}\text{H}_2$ group, which is used up in forming a bond when a proton is added to form a -NH_3^+ ion.

The maximum number of lone pairs on each atom is given by the difference between the number of electrons and the number of orbitals on an atom. Since a lone pair consists of two electrons in one atomic orbital, the remaining atomic orbitals are used up, one at a time, until either all the atomic orbitals on the atom are accounted for or all the potential lone pairs have been formed.

In benzaldehyde, only the oxygen atom has the potential to form lone pairs: five electrons and three atomic orbitals imply a maximum of two lone pairs. The fact that there are three available atomic orbitals means that both lone pairs can in fact be constructed. This uses up two atomic orbitals and four electrons, to leave the remaining orbital and electron as shown in Table 5.36.

When a lone pair is constructed, an occupied LMO is formed, but, unlike the σ bonds, no unoccupied LMO is formed at the same time.

Construction of Non-Cyclic π Bonds

A potential complication exists when the π bonds are constructed, in that it is possible to start constructing π bonds in such a way that a di-radical is formed. For example, in butadiene, if a π bond is constructed between carbon atoms 2 and 3, then radical sites would be formed at atoms 1 and 4. To avoid this, the first set of π bonds constructed is that set in which the way the π bond is drawn is unambiguous. The rule for assigning π bonds is simple: a π bond is constructed if an atom can make a π bond with exactly one other atom.

In benzaldehyde, this situation exists only for the carbon and the oxygen of the -CHO group. The π bonds in the ring cannot be assigned using this rule, because every atom in the ring can potentially make a π bond with *two*

Table 5.36: Unused Electrons and Orbitals after Construction of Lone Pairs

Quantity	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	H ₂	H ₃	H ₄	H ₅	H ₆	C ₇	O ₁	H ₇
Orbitals	1	1	1	1	1	1	0	0	0	0	0	1	1	0
Electrons	1	1	1	1	1	1	0	0	0	0	0	1	1	0

Table 5.37: Unused Electrons and Orbitals after Construction of Non-Cyclic π Bonds

Quantity	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	H ₂	H ₃	H ₄	H ₅	H ₆	C ₇	O ₁	H ₆
Orbitals	1	1	1	1	1	1	0	0	0	0	0	0	0	0
Electrons	1	1	1	1	1	1	0	0	0	0	0	0	0	0

Table 5.38: Charged Sites in Proteins

Neutral	Ion	Charge
-COOH	-COO	-
-NH ₂	-NH ₃	+
=NH	=NH ₂	+
-SH	-S	-

other atoms. In more complicated systems, repeated application of the rule automatically results in the correct assignment of π bonds in all non-cycle conjugated systems.

As with the σ bonds, when a π bond is made, two electrons and two atomic orbitals are used up, and an occupied and an unoccupied LMO are formed. After the non-cyclic π bonds are formed, the number of available orbitals and electrons is as shown in Table 5.37.

Construction of Cyclic π Bonds

At this point, the only remaining bonds are those in cyclic π systems. The sequence in which these are assigned is similar to that for the non-cyclic π bonds. One of the remaining π bonds is assigned arbitrarily, then the rule for assigning non-cyclic π bonds is used until no more π bonds can be made. This two-step sequence is repeated until all cyclic π systems have been accounted for.

In the case of benzaldehyde, this completes the construction of the occupied and virtual starting LMOs. Two more structures are possible in enzymes, however. These will be considered next.

Identification of Ions

For un-ionized systems, all the atomic orbitals on all atoms are accounted for at this point. For ionized systems, however, one or more atomic orbitals will still be unused. At most, there would be one unused atomic orbital on any given atom. The only problem left is to decide whether to assign these unused orbitals as belonging to the occupied or unoccupied sets of LMOs.

In proteins, charged sites occur in only a limited number of environments. Table 5.38 lists the more common species. The following general rules can be derived from this Table:

1. If the immediate environment has an oxygen or a sulfur atom, the unused atomic orbital is assigned to the occupied set.
2. If the immediate environment has a nitrogen atom, the unused atomic orbital is assigned to the unoccupied set.

In other words, all ions associated with oxygen or sulfur are anions, and all ions associated with nitrogen are cations.

As with the lone pairs, LMOs representing ionized sites involve only one orbital, an occupied LMO in the case of an anion, and an unoccupied LMO for a cation.

If the calculated number of cationic and anionic sites is different, then the entire system will have a net charge. The numerical value of that charge is given by subtracting the number of anionic sites from the number of cationic sites.

Properties of Starting LMOs

The starting LMOs have two properties which are essential for the SCF calculation. First, since they are constructed from atomic or hybrid atomic orbitals, either directly or as a linear combination of two hybrids, they form an orthogonal set. Because of the way bonding and antibonding LMOs are constructed, all occupied LMOs are orthogonal to all unoccupied LMOs. The second essential property is that every LMO is normalized to unity.

Both properties are essential for the SCF calculation, in that their existence means that any unitary transform involving any two LMOs (usually one occupied and one unoccupied LMO) will preserve the orthonormality of the entire set of occupied plus virtual LMOs.

5.6.5 Other Considerations

The first time a large system is run, there is a high probability that errors will exist in the data-set which would make the full calculation useless. To save time, a check is run on the system to determine the number of each type of LMO. If an atom bonds to more than four neighbors, then the σ framework cannot be made, and the run is stopped with a warning message.

If the calculated net charge is different from the charge specified on the keyword line, then the assigned number of occupied LMOs will not match the calculated number of occupied LMOs, i.e., the number of σ and π LMOS plus the lone pairs plus the anions. If that happens, then the job is stopped, and a warning message printed out.

Re-Orthogonalization of LMOs

During the course of a long calculation, the LMOs will become increasingly non-orthogonal. Subroutine CHECK ensures that the LMOs remain normalized; this is a very rapid calculation. Ensuring that the LMOs are all orthogonal is not so simple. Re-orthogonalizing the LMOs is a lengthy calculation, and is not routinely performed. However, the option exists to re-orthogonalize the LMOs, and this can be done either during a 1SCF calculation (the preferred way), or periodically during a geometry optimization.

Re-orthogonalizing consists of taking pairs of LMOs, ψ_i and ψ_j , and forming linear combinations such that the overlap ($\langle \psi_i | \psi_j \rangle$) is zero. The LMOs involved form the full set, that is, both occupied plus virtual sets are used.

The two LMOs can be regarded as unit vectors that are almost at 90° to each other. Let the difference from 90° be α . If the vectors are each rotated by $-\frac{1}{2}\alpha$, then they will become perfectly orthogonal. This operation is most conveniently performed using perturbation theory. Let:

$$\langle \psi_i | \psi_i \rangle \simeq 1 \simeq \langle \psi_j | \psi_j \rangle$$

and

$$|\langle \psi_i | \psi_j \rangle| = S_{ij} \ll 1,$$

then

$$\psi'_i = \psi_i - \frac{1}{2} S_{ij} \psi_j$$

and

$$\psi'_j = \psi_j - \frac{1}{2} S_{ij} \psi_i.$$

That the new LMOs are orthogonal can readily be demonstrated:

$$\begin{aligned}
 \langle \psi'_i | \psi'_j \rangle &= \langle (\psi_i - \frac{1}{2} S_{ij} \psi_j) | (\psi_j - \frac{1}{2} S_{ij} \psi_i) \rangle \\
 &= \langle \psi_i | \psi_j \rangle - \frac{1}{2} S_{ij} \langle \psi_i | \psi_i \rangle - \frac{1}{2} S_{ij} \langle \psi_j | \psi_j \rangle + \frac{1}{4} S_{ij}^2 \langle \psi_j | \psi_i \rangle \\
 &= S_{ij} - \frac{1}{2} S_{ij} - \frac{1}{2} S_{ij} + \frac{1}{4} S_{ij}^3 \\
 &= 0
 \end{aligned}$$

The calculation of the overlaps, S_{ij} , is most conveniently done for one LMO, ϕ_i , with all other LMOs. Because of this, ϕ_i should not be modified while the re-orthogonalization is done. In order to avoid modifying ϕ_i , the rotation is changed so that ϕ_i remains stationary and all the rotation is incurred by ϕ_j , thus:

$$\psi'_i = \psi_i$$

and

$$\psi'_j = \psi_j - S_{ij} \psi_i.$$

Before the re-orthogonalization, the LMOs are almost orthogonal, and the use of perturbation theory here is fully justified.

Localized Molecular Orbitals: Storage Considerations

When they are first made, the LMOs are highly compact functions. Each LMO consists of one or two atom labels and a set of four to eighteen atomic orbital coefficients. However, as soon as the LMOs start to interact, mixing of the LMOs will occur, and the size of the LMOs will increase rapidly. This expansion must be allowed for. Later on, further mixing of the LMOs may result in the contribution to an atom to a certain LMO becoming negligible, in which case the atom can be deleted from the LMO. This results in the LMO becoming smaller. A mechanism must be constructed to allow the size of the LMOs to be changed. Because the occupied and unoccupied LMO sets are treated in a similar manner, only the storage of the occupied set will be described. Therefore, for the rest of this section, all reference to LMOs should be understood as applying to the occupied set only.

Before going further, we need to spend a little time to understand how LMOs are stored. The atomic orbital coefficients of all LMOs are stored in a single REAL array, COCC. A second array, NCOCC, holds the starting location in COCC of each LMO. For convenience, all starting locations are defined as the location immediately *before* the location of the first element of an array. The atom numbers (not the atomic numbers) used in all LMOs are stored in an INTEGER array, ICOCC, and a second INTEGER array, NNCF, holds the starting location in ICOCC of each LMO. Finally, the number of atoms in each LMO are stored in an array NCF. Figure 5.12 illustrates these arrays for methane.

During the SCF calculation, the number of atoms in each LMO will increase. To allow for this, some space must be left between the end of one LMO and the start of the next LMO. The amount of space is determined by dividing all the unused space by the number of LMOs. If the size of the vector COCC is 40, and the size of ICOCC is 25, then the storage of the LMO vectors, after adjusting to allow for expansion, is as shown in Figure 5.13.

After a few iterations, the number of atoms in each LMO will have increased. In general, this number will be less than the number of atoms in the system, but in the example we are using, every LMO will contain all the atoms. This is shown in Figure 5.14. For large systems, the number of atoms in any one LMO will vary. Very large LMOs may have 300 or more atoms, while small LMOs may have only 25 atoms. On average, LMOs will have about 150–200 atoms.

For every array used in describing the occupied LMOs, there exists an equivalent array for the virtual LMOs. Table 5.39 gives the names of the arrays for the virtual LMOs.

	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
COCC:	$H_1, C_s, C_{px}, C_{py}, C_{pz}$ $C_s, C_{px}, C_{py}, C_{pz}, H_2$ $C_s, C_{px}, C_{py}, C_{pz}, H_3$ $C_s, C_{px}, C_{py}, C_{pz}, H_4$
	1 2 3 4
NCOCC:	0 5 10 15
	1 2 3 4 5 6 7 8
ICOCC:	1 2 2 3 2 4 2 5
	1 2 3 4
NNCF:	0 2 4 6
	1 2 3 4
NCF:	2 2 2 2

Figure 5.12: Storage of Starting LMOs in Methane

	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
COCC:	$H_1, C_s, C_{px}, C_{py}, C_{pz}$ $C_s, C_{px}, C_{py}, C_{pz}, H_2$...
	1 2 3 4
NCOCC:	0 10 20 30
	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
ICOCC:	1 2 2 3 2 4 2 5 ...
	1 2 3 4
NNCF:	0 6 12 18
	1 2 3 4
NCF:	2 2 2 2

Figure 5.13: Storage of Starting LMOs in Methane, Showing Space Available for Expansion

Table 5.39: Names of Arrays Used to Represent LMOs

Occupied	Unoccupied	Contents
NCF	NCE	Number of Atoms in LMO
NNCF	NNCE	Starting Address of Atoms
ICOCC	ICVIR	Atom Numbers
NCOCC	NCVIR	Starting Address of LMO
COCC	CVIR	LMO Atomic Orbital Coefficients

COCC:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">5</td><td style="padding: 2px 5px;">6</td><td style="padding: 2px 5px;">7</td><td style="padding: 2px 5px;">8</td><td style="padding: 2px 5px;">9</td><td style="padding: 2px 5px;">10</td><td style="padding: 2px 5px;">11</td><td style="padding: 2px 5px;">12</td><td style="padding: 2px 5px;">13</td><td style="padding: 2px 5px;">14</td><td style="padding: 2px 5px;">15</td><td style="padding: 2px 5px;">16</td><td style="padding: 2px 5px;">17</td><td style="padding: 2px 5px;">18</td><td style="padding: 2px 5px;">19</td><td style="padding: 2px 5px;">20</td> </tr> <tr> <td colspan="10" style="padding: 2px 5px; text-align: center;">$H_1, C_s, C_{px}, C_{py}, C_{pz}, H_2, H_3, H_4$</td> <td colspan="10" style="padding: 2px 5px; text-align: center;">$C_s, C_{px}, C_{py}, C_{pz}, H_2, H_1, H_3, H_4$</td> </tr> </table>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	$H_1, C_s, C_{px}, C_{py}, C_{pz}, H_2, H_3, H_4$										$C_s, C_{px}, C_{py}, C_{pz}, H_2, H_1, H_3, H_4$...
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																							
$H_1, C_s, C_{px}, C_{py}, C_{pz}, H_2, H_3, H_4$										$C_s, C_{px}, C_{py}, C_{pz}, H_2, H_1, H_3, H_4$																																

NCOCC:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td> </tr> <tr> <td style="padding: 2px 5px; text-align: center;">0</td><td style="padding: 2px 5px; text-align: center;">10</td><td style="padding: 2px 5px; text-align: center;">20</td><td style="padding: 2px 5px; text-align: center;">30</td> </tr> </table>	1	2	3	4	0	10	20	30
1	2	3	4						
0	10	20	30						

ICOCC:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td><td style="padding: 2px 5px;">5</td><td style="padding: 2px 5px;">6</td><td style="padding: 2px 5px;">7</td><td style="padding: 2px 5px;">8</td><td style="padding: 2px 5px;">9</td><td style="padding: 2px 5px;">10</td><td style="padding: 2px 5px;">11</td><td style="padding: 2px 5px;">12</td><td style="padding: 2px 5px;">13</td><td style="padding: 2px 5px;">14</td><td style="padding: 2px 5px;">15</td><td style="padding: 2px 5px;">16</td><td style="padding: 2px 5px;">17</td><td style="padding: 2px 5px;">18</td><td style="padding: 2px 5px;">19</td><td style="padding: 2px 5px;">20</td><td style="padding: 2px 5px;">21</td><td style="padding: 2px 5px;">22</td> </tr> <tr> <td style="padding: 2px 5px; text-align: center;">1</td><td style="padding: 2px 5px; text-align: center;">2</td><td style="padding: 2px 5px; text-align: center;">3</td><td style="padding: 2px 5px; text-align: center;">4</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px; text-align: center;">2</td><td style="padding: 2px 5px; text-align: center;">3</td><td style="padding: 2px 5px; text-align: center;">1</td><td style="padding: 2px 5px; text-align: center;">4</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px; text-align: center;">2</td><td style="padding: 2px 5px; text-align: center;">4</td><td style="padding: 2px 5px; text-align: center;">1</td><td style="padding: 2px 5px; text-align: center;">3</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px;"></td><td style="padding: 2px 5px; text-align: center;">2</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px; text-align: center;">1</td><td style="padding: 2px 5px; text-align: center;">3</td> </tr> </table>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	1	2	3	4	5		2	3	1	4	5		2	4	1	3	5		2	5	1	3	...
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22																									
1	2	3	4	5		2	3	1	4	5		2	4	1	3	5		2	5	1	3																									

NNCF:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td> </tr> <tr> <td style="padding: 2px 5px; text-align: center;">0</td><td style="padding: 2px 5px; text-align: center;">6</td><td style="padding: 2px 5px; text-align: center;">12</td><td style="padding: 2px 5px; text-align: center;">18</td> </tr> </table>	1	2	3	4	0	6	12	18
1	2	3	4						
0	6	12	18						

NCF:	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="padding: 2px 5px;">1</td><td style="padding: 2px 5px;">2</td><td style="padding: 2px 5px;">3</td><td style="padding: 2px 5px;">4</td> </tr> <tr> <td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px; text-align: center;">5</td><td style="padding: 2px 5px; text-align: center;">5</td> </tr> </table>	1	2	3	4	5	5	5	5
1	2	3	4						
5	5	5	5						

Figure 5.14: Storage of Iterating LMOs in Methane

5.6.6 Diagonalization

At self-consistency, all matrix elements connecting every occupied LMO with every unoccupied LMO is zero. Annihilation of these matrix elements is performed by the subroutine DIAGG. The only matrix elements which need to be considered are those which involve an occupied LMO with an unoccupied LMO which shares one or more common atoms. With increasing size of system, the number of interactions of this type compared with the product of the number of occupied times unoccupied LMOs becomes increasingly small.

Although DIAGG looks quite formidable, the working of this subroutine can be easily understood in terms of the principles involved in matrix element annihilation. Therefore, before going into the details of the LMO matrix annihilation, the conventional procedure used in MOPAC will be reviewed.

Conventional Matrix Annihilation

The raw material DIAGG is supplied with consists of the occupied LMOs and the Fock matrix over atomic orbitals. The first step, therefore, is to calculate the occupied and virtual energy levels, and the potentially non-zero matrix elements connecting the occupied and virtual sets.

In a system of N atoms, let a molecular orbital, ψ_i , be represented by

$$\psi_i = \sum_{A=1}^N \sum_{\lambda \in A} c_{\lambda i} \quad (5.27)$$

Then the energy of a M.O. can be evaluated from

$$\varepsilon_{ii} = \sum_A^N \sum_B^N \sum_{\lambda \in A} \sum_{\sigma \in B} c_{\lambda i} F_{\lambda \sigma} c_{\sigma i} \quad (5.28)$$

and the matrix element representing the energy term between occupied M.O. ψ_i and virtual M.O. ψ_j would be

$$\varepsilon_{ij} = \sum_A^N \sum_B^N \sum_{\lambda \in A} \sum_{\sigma \in B} c_{\lambda i} F_{\lambda \sigma} c_{\sigma j} \quad (5.29)$$

Annihilation of a matrix element is achieved by performing a unitary transform on the M.O.s involved

$$\begin{aligned}\psi'_i &= \alpha\psi_i + \beta\psi_j \\ \psi'_j &= -\beta\psi_i + \alpha\psi_j\end{aligned}\quad (5.30)$$

where the rotation angles α and β are calculated from the molecular orbital energy matrix elements

$$\begin{aligned}\alpha &= \sqrt{\frac{1}{2}(1 + D/\sqrt{4\varepsilon_{ij}^2 + D^2})} \\ \beta &= \phi\sqrt{1 - \alpha^2}\end{aligned}\quad (5.31)$$

where $D = \varepsilon_{jj} - \varepsilon_{ii}$ and $\phi = 1$ if ε_{ij} is negative, $\phi = -1$, otherwise.

By evaluating partial sums, the calculation of ε_{ij} can be made more efficient. If ε_{ij} is re-written as

$$\varepsilon_{ij} = \sum_A^N \sum_B^N \sum_{\lambda \in A} c_{\lambda i} \sum_{\sigma \in B} F_{\lambda\sigma} c_{\sigma j}; \quad (5.32)$$

then the partial sum

$$F'_\lambda(j) = \sum_{\sigma \in B} F_{\lambda\sigma} c_{\sigma j} \quad (5.33)$$

can be used to simplify ε_{ij} :

$$\varepsilon_{ij} = \sum_A^N \sum_{\lambda \in A} c_{\lambda i} F'_\lambda(j). \quad (5.34)$$

Because the vector $F'_\lambda(j)$ is evaluated once, and then used for all i , the calculation of the ε_{ij} is changed from an n^3 process to a n^2 process. The use of the partial sum $F'_\lambda(j)$ also speeds up the evaluation of the virtual M.O. energies. Unfortunately, it does not speed up the calculation of the occupied M.O. energies.

Localized Molecular Orbital Matrix Element Evaluation

It is important in LMO work that only those matrix elements which are potentially non-zero be evaluated; therefore, modifications need to be made to the equations just described. For the purpose of the following discussion, the system being studied should be assumed to be very large, i.e., to contain thousands of atoms.

To assist in understanding how the modifications are made, it is convenient at this time to write the occupied LMO as:

$$\psi_i = \sum_{j=1}^{NCF(i)} \sum_{\lambda \in ICOCC(j+NNCF(i))} COCC(\lambda + NCOCC(i)) \quad (5.35)$$

and the virtual LMO as:

$$\psi_j = \sum_{l=1}^{NCE(j)} \sum_{\sigma \in ICVIR(l+NNCE(j))} CVIR(\sigma + NCVIR(j)). \quad (5.36)$$

These expressions also illustrate the relationships of the five arrays which are used to represent the LMOs.

Then the partial sum, $F'_\lambda(j)$, can be represented as:

$$F'_\lambda(j) = \sum_{l=1}^{NCE(j)} \sum_{\sigma \in ICVIR(l+NNCE(j))} CVIR(\sigma + NCVIR(j)) F_{\lambda\sigma}. \quad (5.37)$$

The number of terms in this sum is already much smaller than in conventional matrix annihilation in that the number of atoms represented in each LMO, $NCE(j)$, is much less than the total number of atoms.

Because all Fock matrix elements connecting atoms which are separated by more than $CUTOF2$ are automatically zero, only those terms which refer to atoms separated by less than $CUTOF2$ need be evaluated.

All LMOs consist of a central part consisting of one to three atoms, which accounts for almost all of the wavefunction. In the regions of an LMO far away from the center, the contribution of any atom to the wavefunction becomes very small. If the Fock matrix elements connecting this distant atom to any other atom is also very small, then quantities which depend of both of these terms becomes quite negligible. To test whether this condition exists, two scalar quantities need to be calculated. The first is the contribution of each atom to the LMO:

$$\rho_A(j) = \sum_{\lambda \in A} \psi_{\lambda j}^2 \quad (5.38)$$

and second is the magnitude of the Fock matrix vector connecting each pair of atoms:

$$F_{AB} = \sum_{\lambda \in A \sigma \in B} F_{\lambda \sigma}^2. \quad (5.39)$$

During the evaluation of $F'_\lambda(j)$, the quantity $\rho_A F_{AB}$ is computed. Only if it is above a preset limit are the terms involving atoms A and B used.

As soon as $F'_\lambda(j)$ is calculated, the magnitude of each atom's contribution to $F'_\lambda(j)$ is determined, and stored:

$$F_A(j) = \sum_A \sum_{\lambda \in A} F'_\lambda(j)^2. \quad (5.40)$$

Together, these three modifications result in a large reduction in the time necessary to compute the partial sum $F'_\lambda(j)$.

The virtual energy levels are calculated using $F'_\lambda(j)$:

$$\varepsilon_{jj} = \sum_{l=1}^{NCE(j)} \sum_{\sigma \in ICVIR(l+NNCE(j))} CVIR(\sigma + NCVIR(j)) F'_\sigma(j). \quad (5.41)$$

In the same manner as $\rho_A F_{AB}$ was used in deciding whether the terms involving atoms A and B should be evaluated in the calculation of $F'_\lambda(j)$, the quantity $F_A(j) \rho_A$ is evaluated and used to decide which terms in the current summation should be used.

Now the calculation of the occupied-virtual matrix elements can be performed. The quantity to be calculated is

$$\varepsilon_{ij} = \sum_{k=1}^{NCF(i)} \sum_{\lambda \in ICOCC(k) + NNCF(i)} COCC(\lambda + NCOCC(i)) F'_\lambda(j). \quad (5.42)$$

Because this is a single sum over atoms, not much time is saved by testing to see if any terms can be omitted. However, the test is a simple one, and it does result in a small increase in speed by evaluating the quantity $F_A(j) \rho_A(i)$, and comparing it to a preset limit. Since both $F_A(j)$ and $\rho_A(i)$ have already been calculated, this test is very rapid.

Calculation of the occupied-virtual matrix elements presents a new problem. There are a large number of these elements, many of which involve only atoms which are far from the LMO centers. To see why this is so, consider two LMOs which have exactly one atom in common. For this to happen, the atom in question must be very far from the centers of both LMOs. The energy terms arising from such an atom must, of necessity, be very small. In the interest of efficiency, all calculations involving such atoms should be ignored. Unfortunately, it is not possible to *a priori* determine which terms to include and which to leave out. Therefore, at the start of the SCF calculation, all terms must be evaluated. As soon as large changes in the LMOs have stopped, a list can be constructed of those occupied-virtual matrix elements which need to be considered for annihilation.

Localized Molecular Orbital Matrix Element Annihilation

Once the matrix elements have been calculated, annihilation is relatively straightforward. As with conventional SCF matrix annihilation, the operation to be performed is:

$$\begin{aligned}\psi'_i &= \alpha\psi_i + \beta\psi_j \\ \psi'_j &= -\beta\psi_i + \alpha\psi_j\end{aligned}\quad (5.43)$$

where α and β are calculated using the matrix elements just derived.

Because of the way LMOs are represented (the five arrays for each set, occupied and virtual), care must be taken to ensure that the two by two rotation is done correctly. For the rotation to be performed, each atomic orbital in ψ_i must be matched with the same atomic orbital in ψ_j . Three distinct situations occur: the same atom might be present in both ψ_i and ψ_j ; ψ_i might have an atom which ψ_j does not have; and ψ_j might have an atom which ψ_i does not have. The procedure for dealing with these situations is as follows:

- Both LMOs have the same atom.

Simple rotation of the coefficients is performed:

$$\begin{aligned}\sum_{\lambda \in A} \psi'_i(\lambda) &= \alpha\psi_i(\lambda) + \beta\psi_j(\lambda) \\ \sum_{\lambda \in A} \psi'_j(\lambda) &= -\beta\psi_i(\lambda) + \alpha\psi_j(\lambda)\end{aligned}\quad (5.44)$$

- Occupied LMO ψ_i has atom A , virtual LMO ψ_j does not have atom A .

If the product $\beta\rho_A(i)$ is above a threshold, then the rotation

$$\begin{aligned}\sum_{\lambda \in A} \psi'_i(\lambda) &= \alpha\psi_i(\lambda) \\ \sum_{\lambda \in A} \psi'_j(\lambda) &= -\beta\psi_i(\lambda)\end{aligned}\quad (5.45)$$

is performed. If it is less than the threshold, no rotation is done. The most important effect of this rotation is to place an atom in the virtual LMO ψ_j , an atom which was not in ψ_j before the two by two rotation was started. That is, the number of atoms in ψ_j is increased by one.

- Virtual LMO ψ_j has atom A , occupied LMO ψ_i does not have atom A .

Again, if the product $\beta\rho_A(i)$ is above a threshold, then the rotation

$$\begin{aligned}\sum_{\lambda \in A} \psi'_i(\lambda) &= \beta\psi_j(\lambda) \\ \sum_{\lambda \in A} \psi'_j(\lambda) &= \alpha\psi_j(\lambda)\end{aligned}\quad (5.46)$$

is performed. The number of atoms in the occupied LMO is increased by one.

At the start of the first SCF calculation, each LMO contains at most two atoms, so the most dramatic effect of matrix element annihilation is to cause the LMOs to expand so that they involve more atoms. In the first few iterations, the number of atoms in each LMO increases rapidly. After about 10-30 iterations, the rate of increase becomes very small, as the size of the LMOs becomes stable.

During the first SCF calculation, the number of atoms in certain LMOs may increase so rapidly that the gap between LMOs might vanish. Any further expansion would then cause vector overwriting to occur. To avoid this happening, the annihilation step is modified as follows: The situation where vector overwriting is about to occur is detected. The degree of mixing of the two LMOs is then halved, and the annihilation is attempted again. If this procedure does not work, then the degree of mixing is halved again. This is done repeatedly until the danger of vector overwriting has been removed. This technique is only used near the end of the SCF calculation; at other times the degree of mixing is simply set to zero.

5.6.7 Reducing the Size of Localized Molecular Orbitals

If, as a result of matrix element annihilation, the LMOs are allowed to expand so that they include more and more atoms without limit, then eventually the efficiency of the calculation would be severely impaired. To prevent this occurring, a procedure is needed which will allow the size of LMOs to be reduced.

Atoms can be added at an LMO as a result of matrix element annihilation. What is not so obvious is that annihilation can also make the contribution of an atom to the LMO so small that the atom can subsequently be deleted from that LMO without harm. Of course, if an atom is deleted from an LMO, changes should be made to the five arrays representing the LMOs. The operation of removing atoms from LMOs and adjusting the appropriate arrays is done in subroutine TIDY.

This compression is carried out by copying all the information on each atom in the LMOs. During the copy operation, any atom whose contribution to an LMO is insignificant is not copied. In addition, unused space at the end of each LMO is also not copied. The result of the copy is to produce a set of LMOs in which only atoms which contribute significantly are present, and with no unused space between the LMOs.

Before the LMOs can be used once more, they must be redistributed in their arrays so that each LMO has some unused space at the end to allow for expansion during the next matrix annihilation operation.

Size of LMOs

When a SCF is achieved, the LMOs extend over more atoms than might be expected. Each LMO is about 90–99% on one or two atoms, and if the surrounding few atoms are included, almost 100% of the LMO can be accounted for. From this, it would appear that the intensity of the LMO would continue to decrease rapidly with distance from the center. This is normally not the case. Instead, LMOs usually have intensity on a large number of atoms, sometimes several hundred atoms. As a result, the calculations take a much longer time than would otherwise be expected.

Some effort has been expended in trying to find ways of reducing the size of LMOs. These attempts have not been successful. The definitive failure was provided as follows:

Using MOPAC, the LMOs for a large system were generated using PRECISE. The resulting LMOs were as expected, over 90% on two atoms, with most of the rest of the wavefunction on the nearby atoms. However, the intensity did not drop rapidly to zero with increasing distance. Instead, it held more or less constant at about 10^{-5} to 10^{-8} for a large number of atoms before finally dropping to a negligible value.

This behavior did not change on increasing the precision of the localization.

Because of this result, it was obvious that further localization of the MOZYME LMOs would not be useful.

5.6.8 Density Matrix Construction

The general expression for constructing the density matrix from the occupied set of molecular orbitals is:

$$P_{\lambda\sigma} = 2 \sum_{i=1}^{occ} c_{\lambda i} c_{\sigma i}. \quad (5.47)$$

This same basic equation is used in the construction of the density matrix in LMO theory. The main difference is that not all density matrix elements are, or need to be, evaluated.

The only elements of the density matrix which need to be evaluated are those relating to atoms which are (see Page 163). To understand why, consider how the density matrix is used in the SCF calculation. The Fock matrix is constructed from products involving the density matrix, the one-electron matrix, and the two-electron integrals. All terms which involve density matrix elements connecting any two atoms only, involve one electron integrals and two electron exchange integrals for the same two atoms. All one-electron integrals and two-electron exchange

integrals for atoms separated by more than `CUTOF2` are zero. Therefore, the value of density matrix elements for atoms separated by more than `CUTOF2` is unimportant. To save time and to reduce array size, therefore, these density matrix elements are not calculated.

Having stated that the construction of the density matrix is the same in principle as that used in conventional M.O. theory, the way in which LMOs are stored does introduce a technical difference. The sequence in which the atoms are represented in LMOs changes from LMO to LMO. While it is possible for two LMOs to have the same sequence (a σ and π LMO involving the same two atoms is an example), in general the LMOs should be considered as consisting of a random set of atoms, and should be treated as such.

This unpredictable nature of the composition of the LMOs means that the order in which density matrix elements are calculated is determined by the LMOs, and not by the sequence of atoms in the molecule.

5.6.9 Energy Effects of `CUTOF2`

Energy terms arise from pairs of atoms which are separated by distances greater than the cutoff distance. These terms are purely electrostatic in origin: their magnitude is simply proportional to the net charge, Q_A , on the atoms involved. Net charge is defined as

$$Q_A = Z_A - Ne_A \quad (5.48)$$

where Ne_A is the total electron population on atom A

$$Ne_A = \sum_{\lambda \in A} P_{\lambda\lambda}. \quad (5.49)$$

From simple electrostatics, the overall effect of these net charges is to contribute a long-range energy term, E_{lre} to the energy of the system thusly

$$E_{lre} = \frac{1}{2} \sum_A \sum_B \phi_{AB} Q_A Q_B \gamma_{AB}, \quad (5.50)$$

where ϕ_{AB} is 1 if R_{AB} is greater than the cutoff distance but 0 otherwise, and γ_{AB} is the $\langle ss|1/R_{AB}|ss \rangle$ integral. This simple description must be modified because of the effect of net charges on the electron density. Each distant net positive charge will induce a small, but in general not negligible, stabilizing effect on the electron distribution, while a distant net negative charge will have the opposite effect. This can be expressed formally in terms of the Fock matrix

$$F_{\lambda \in A \lambda \in A} = F'_{\lambda \in A \lambda \in A} - \sum_B \phi_{AB} Q_B \gamma_{AB}. \quad (5.51)$$

The effect on the nuclear term is similar, but opposite in sign:

$$E_{A(nuc)} = E'_{A(nuc)} + \sum_B \phi_{AB} Q_B Z_A \gamma_{AB}. \quad (5.52)$$

The total energy is given by the sum of the electronic plus nuclear energies. Given that the electronic energy is

$$E_{ee} = \frac{1}{2} \sum_{\lambda} \sum_{\sigma} P_{\lambda\sigma} (H_{\lambda\sigma} + F_{\lambda\sigma}), \quad (5.53)$$

the electronic term due to distant nuclei is

$$E_{lre(ee)} = \frac{-1}{2} \sum_A \sum_{\lambda \in A} P_{\lambda\lambda} \sum_B \phi_{AB} Q_B \gamma_{AB} \quad (5.54)$$

or

$$E_{lre(ee)} = \frac{-1}{2} \sum_A N e_A \sum_B \phi_{AB} Q_B \gamma_{AB}. \quad (5.55)$$

Likewise, given that the nuclear energy is

$$E_{ne} = \sum_A \sum_{B < A} Z_A Z_B \gamma_{AB}, \quad (5.56)$$

the nuclear energy term due to distant nuclei is

$$E_{lre(ne)} = \sum_A \sum_{B < A} \phi_{AB} Q_A Z_B \gamma_{AB}. \quad (5.57)$$

This can be re-written in a more symmetric form as

$$E_{lre(ne)} = \frac{1}{4} \sum_A \sum_B \phi_{AB} (Q_A Z_B + Q_B Z_A) \gamma_{AB}. \quad (5.58)$$

Together, the total contribution due to the electronic and nuclear terms arising from distant atoms is

$$E_{lre(ee)} + E_{lre(ne)} = \frac{1}{4} \sum_A \sum_B \phi_{AB} (Q_A (Z_B - N e_B) + Q_B (Z_A - N e_A)) \gamma_{AB}. \quad (5.59)$$

Rearranging gives

$$E_{lre(ee)} + E_{lre(ne)} = \frac{1}{4} \sum_A \sum_B \phi_{AB} (Q_A Q_B + Q_B Q_A) \gamma_{AB} \quad (5.60)$$

or

$$E_{lre(ee)} + E_{lre(ne)} = \frac{1}{2} \sum_A \sum_B \phi_{AB} Q_A Q_B \gamma_{AB}, \quad (5.61)$$

which is identical to the simple equation this discussion began with.

In the 1996 version of MOZYME, this sum was calculated explicitly. However, evaluation of the sum can be avoided by adding into the one-electron matrix the electrostatic stabilization terms arising from atoms beyond CUTOFF2. This concisely and effectively takes into account the quantities just discussed.

Although the point-charge electrostatic effects are important, they are not the only long-range effect. The effect of a point-charge on a lone pair of electrons is also significant. Consider a lone pair on an atom pointing in the direction of a positive charge. Clearly this will lead to a stabilizing effect. A lone pair pointing in the opposite direction will be destabilized. Similarly, a lone pair pointing at 90° to a charge will be subjected to a torque. In order to include these effects in the Hamiltonian some extra terms are needed.

The size of a lone pair on an atom is represented in the density matrix by the value of the associated $s - p$ terms. For each atom, there are three such terms: P_{sp_x} , P_{sp_y} , and P_{sp_z} . The effect of distant charges on F_{sp_x} is given by

$$F'_{sp_x} = F_{sp_x} - \sum_B Q_B \langle ss | sx \rangle P_{sp_x} \quad (5.62)$$

Similar terms exist for F_{sp_y} and F_{sp_z} .

5.7 Geometry optimization

The default geometry optimizer in MOPAC uses Baker's EigenFollowing method. If this is *not* wanted, for example, if there is a need to reduce memory demands, then the Broyden Fletcher Goldfarb Shanno method can be used.

The most common use of MOPAC is for geometry optimization. This involves starting with an approximation to the desired geometry and, by calculating the forces acting on the system, changing the geometry so as to lower the total energy. The objective of geometry optimization is to achieve a structure in which all the atoms are at equilibrium, that is, one in which the forces acting on every atom are very small, and in which the second derivatives are everywhere positive. Such a geometry is called a ground state stationary point.

5.7.1 EigenFollowing

Description of the EF and TS function
by
Dr Frank Jensen
Department of Chemistry
Odense University
5230 Odense
Denmark

The current version of the EF optimization routine is a combination of the original EF algorithm of Simons et al. (J. Phys. Chem. 89, 52) as implemented by Baker (J. Comp. Chem. 7, 385) and the QA algorithm of Culot et al. (Theo. Chim. Acta 82, 189), with some added features for improving stability.

The geometry optimization is based on a second order Taylor expansion of the energy around the current point. At this point the energy, the gradient and some estimate of the Hessian are available. There are three fundamental steps in determining the next geometry based on this information:

- finding the "best" step within or on the hypersphere with the current trust radius.
- possibly reject this step based on various criteria.
- update the trust radius.

1. For a minimum search the correct Hessian has only positive eigenvalues. For a Transition State (TS) search the correct Hessian should have exactly one negative eigenvalue, and the corresponding eigenvector should be in the direction of the desired reaction coordinate. The geometry step is parameterized as $g/(s - H)$, where s is a shift factor which ensure that the step-length is within or on the hypersphere. If the Hessian has the correct structure, a pure Newton-Raphson step is attempted. This corresponds to setting the shift factor to zero. If this step is longer than the trust radius, a P-RFO step is attempted. If this is also too long, then the best step on the hypersphere is made via the QA formula. This three step procedure is the default. The pure NR step can be skipped by giving the keyword `NONR`. An alternative to the QA step is to simply scale the P-RFO step down to the trust radius by a multiplicative constant, this can be accomplished by specifying `RSCAL`.
2. Using the step determined from 1), the new energy and gradient are evaluated. If it is a TS search, two criteria are used in determining whether the step is "appropriate". The ratio between the actual and predicted energy change should ideally be 1. If it deviates substantially from this value, the second order Taylor expansion is no longer accurate. `RMIN` and `RMAX` (default values 0 and 4) determine the limits on how far

from 1 the ratio can be before the step is rejected. If the ratio is outside the R_{MIN} and R_{MAX} limits, the step is rejected, the trust radius reduced by a factor of two and a new step is determined. The second criteria is that the eigenvector along which the energy is being maximized should not change substantially between iterations. The minimum overlap of the TS eigenvector with that of the previous iteration should be larger than O_{MIN} , otherwise the step is rejected. Such a step rejection can be recognized in the output by the presence of (possibly more) lines with the same $CYCLE$ number. The default O_{MIN} value is 0.8, which allows fairly large changes to occur, and should be suitable for most uncomplicated systems. See below for a discussion of how to use R_{MIN} , R_{MAX} and O_{MIN} for difficult cases. The selection of which eigenvector to follow towards the TS is given by $MODE=n$, where n is the number of the Hessian eigenvector to follow. The default is $MODE=1$. These features can be turned off by giving suitable values as keywords, e.g. $R_{MIN}=-100$ $R_{MAX}=100$ effectively inhibits step rejection. Similarly setting $O_{MIN}=0$ disables step rejection based on large changes in the structure of the TS mode. The default is to use mode following even if the TS mode is the lowest eigenvector. This means that the TS mode may change to some higher mode during the optimization. To turn off mode following, and thus always follow the mode with lowest eigenvalue, set $MODE=0$. If it is a minimum search the new energy should be lower than the previous.

The acceptance criteria used is that the actual/predicted ratio should be larger than R_{MIN} , which for the default value of $R_{MIN}=0$ is equivalent to a lower energy. If the ratio is below R_{MIN} , the step is rejected, the trust radius reduced by a factor of two and a new step is predicted. The R_{MIN} , R_{MAX} and O_{MIN} features has been introduced in the current version of EF to improve the stability of TS optimizations. Setting R_{MIN} and R_{MAX} close to one will give a very stable, but also very slow, optimization. Wide limits on R_{MIN} and R_{MAX} may in some cases give a faster convergence, but there is always the risk that very poor steps are accepted, causing the optimization to diverge. The default values of 0 and 4 rarely rejects steps which would lead to faster convergence, but may occasionally accept poor steps. If TS searches are found to cause problems, the first try should be to lower the limits to 0.5 and 2. Tighter limits like 0.8 and 1.2, or even 0.9 and 1.1, will almost always slow the optimization down significantly but may be necessary in some cases.

In minimum searches it is usually desirable that the energy decreases in each iteration. In certain very rigid systems, however, the initial diagonal Hessian may be so poor that the algorithm cannot find an acceptable step larger than DD_{MIN} , and the optimization terminates after only a few cycles with the "TRUST RADIUS BELOW DD_{MIN} " warning long before the stationary point is reached. In such cases the user can add $DD_{MIN}=0.0$ and R_{MIN} set to some negative value, say -10, thereby allowing steps which allow the energy to increase. An alternative is to use `LET $DD_{MIN}=0.0$` .

The algorithm has the capability of following Hessian eigenvectors other than the one with the lowest eigenvalue toward a TS. Such higher mode following are always much more difficult to make converge. Ideally, as the optimization progresses, the TS mode should at some point become the lowest eigenvector. Care must be taken during the optimization, however, that the nature of the mode does not change all of a sudden, leading to optimization to a different TS than the one desired. O_{MIN} has been designed for ensuring that the nature of the TS mode only changes gradually, specifically the overlap between successive TS modes should be higher than O_{MIN} . While this concept at first appears very promising, it is not without problems when the Hessian is updated.

As the updated Hessian in each step is only approximately correct, there is a upper limit on how large the TS mode overlap between steps can be. To understand this, consider a series of steps made from the same geometry (e.g. at some point in the optimization), but with steadily smaller step-sizes. The update adds corrections to the Hessian to make it a better approximation to the exact Hessian. As the step-size become small, the updated Hessian converges toward the exact Hessian, at least in the direction of the step. The old Hessian is constant, thus the overlap between TS modes thus does not converge toward 1, but rather to

a constant value which indicate how well the old approximate Hessian resembles the exact Hessian. Test calculations suggest a typical upper limit around 0.9, although cases have been seen where the limit is more like 0.7. It appears that an updated Hessian in general is not of sufficient accuracy for reliably rejecting steps with TS overlaps much greater than 0.80. The default OMIN of 0.80 reflects the typical use of an updated Hessian. If the Hessian is recalculated in each step, however, the TS mode overlap does converge toward 1 as the step-size goes toward zero, and in this cases there is no problems following high lying modes.

Unfortunately setting RECALC=1 is very expensive in terms of computer time, but used in conjecture with OMIN=0.90 (or possibly higher), and maybe also tighter limits on RMIN and RMAX, it represents an option of locating transitions structures that otherwise might not be possible. If problems are encountered with many step rejections due to small TS mode overlaps, try reducing OMIN, maybe all the way down to 0. This most likely will work if the TS mode is the lowest Hessian eigenvector, but it is doubtful that it will produce any useful results if a high lying mode is followed. Finally, following modes other than the lowest toward a TS indicates that the starting geometry is not "close" to the desired TS. In most cases it is thus much better to further refined the starting geometry, than to try following high lying modes. There are cases, however, where it is very difficult to locate a starting geometry which has the correct Hessian, and mode following may be of some use here.

5.7.2 The BFGS function optimizer

The alternative heat of formation minimization routine in MOPAC is a modified Broyden [19]-Fletcher [20]-Goldfarb [21]-Shanno [22] or BFGS method. Minor changes were made necessary by the presence of phenomena peculiar to chemical systems.

Starting with a user-supplied geometry x_o , MOPAC computes an estimate to the inverse Hessian H_o . The geometry optimization proceeds by

$$x_{k+1} = x_k + \alpha d_k,$$

where

$$d_k = H g_k,$$

and each element of H is defined by

$$H_{k+1} = H_k - \frac{H y_k p_k^t + p_k y_k^t H}{S} + \frac{Q(p_k p_k^t)}{S},$$

where

$$Q = 1 + \frac{y_k^t H y_k}{p_k^t y_k},$$

and g_k is the gradient vector on step k .

Although this expression for the update of the Hessian matrix looks very complicated, the operation can be summarized as follows:

The initial Hessian matrix used in geometry optimization is chosen as a diagonal matrix, with the diagonal elements determined by a simple formula based on the gradients at two geometries. As the optimization proceeds, the gradients at each point are used to improve the Hessian. In particular, the off-diagonal elements are assigned based on the old elements and the current gradients.

Two different methods are used to calculate the displacement of x in the direction d . During the initial stages of geometry optimization, a line search is used. This proceeds as follows:

The geometry is displaced by $(\alpha/4)d$ and the energy evaluated via an SCF calculation. If this energy is lower than the original value, then a second step of the same size is made. If it is higher, then a step of $-(\alpha/4)d$ is made. The energy is then re-evaluated. Given the three energies, a prediction is made as to the value of α which will

yield the minimum value of the energy in the direction d . Of course, the size of the steps are constrained so that the system would not suddenly become unrealistic (e.g., break bonds, superimpose atoms, etc.). Similarly, the contingency in which the energy versus α function is inverse parabolic is considered, as are rarely-encountered curves, e.g., almost perfectly linear regressions. By default, Thiel's FSTMIN technique is used [44]. This uses gradient information from the starting point of the search, and the calculated ΔH_f , to decide when to end the line search. If NOTHIEL is specified, the older line-search is used, in which case the search is stopped when the drop in energy on any step becomes less than 5% of the total drop or 0.5 kcal/mol, whichever is smaller.

An important modification has been made to the BFGS routine. For the line-search, Thiel's FSTMIN technique is used. This modification make the algorithm run faster most of the time. However, one unfortunate result of these changes is that there is no guarantee that as the cycles increase, the energy will drop monotonically. If the calculation does not converge on a stationary point, then re-run the job with NOTHIEL.

As the geometry converges on a local minimum, the prediction of the search direction becomes less accurate. There are many reasons for this. For example, the finite precision of the SCF calculation may lead to errors in the density matrix, or finite step sizes in the derivative calculation (if analytical derivatives are not used) may result in errors in the derivatives. For whatever reason, the gradient norm and energy minimum may not coincide. The difference is typically less than 0.00001 kcal/mol and less than 0.05 units of gradient norm.

Normally, the initial guess to H , the inverse Hessian, is the unit matrix. However, in chemical systems where the second derivatives are very large, use of the unit matrix would result in large changes in the geometry. Thus a slightly elongated bond length could, in the first step, change from 1.6Å to -6.5Å. To prevent this catastrophe, the initial geometry is perturbed by a small amount, thus

$$x_1 = x_0 + 0.01 \times \text{sign}(g_0),$$

from which a trial inverse Hessian can be constructed:

$$H_1(i, i) = 0.01 \times \text{sign}(g_0(i))/y_1(i).$$

A negative value for $H_1(i, i)$ would lead to difficulties within the BFGS optimization. To avoid this, $H_1(i, i)$ is set to $0.06/\text{abs}(g(i))$ whenever $\text{sign}(g_0(i))/y_1(i)$ is negative.

As the optimization proceeds, the inverse Hessian matrix becomes more accurate. However, as the geometry steadily changes, the inverse Hessian will contain information which does not reflect the current point. This can lead to the predicted search direction vector making an angle of more than 90° with the gradient vector. In other words, the search direction vector may point uphill in energy. To guard against this, the inverse Hessian is re-initialized whenever the cosine of the angle between the search direction and the gradient vector drops below 0.05.

Originally the Davidson-Fletcher-Powell technique was used, but in rare instances it failed to work satisfactorily. The BFGS formula appears to work as well as or better than the DFP method most of the time. In the infrequent case when the DFP is more efficient, the increase in efficiency of the DFP can usually be traced to a fortuitous choice of a search direction. Small changes in starting conditions can destroy this accidental increased efficiency and make the BFGS method appear more efficient. A keyword, DFP, is provided to allow the DFP optimizer to be used.

5.7.3 Optimization of one unknown

If a system has exactly one coordinate to be optimized, then obviously one line-search will optimize the geometry. Because of this, the geometry optimization is done a little differently. Given the initial geometry, the ΔH_f is calculated, and the line-search started. Unlike the normal line-search, however, the search is not stopped when the minimum is almost reached, instead, the minimum is located with quite high precision. After the line-search

is complete, the gradients are not calculated (unless requested by GRADIENTS). Instead, it is assumed that the gradient is small, and the results are output. This saves some time. However, if GRADIENTS is *not* present, and the geometry is not at a stationary point (because other coordinates are not optimized), then the warning message that the geometry is not at a stationary point will not be printed.

5.7.4 Considerations in Geometry Optimization

The default settings in MOPAC are designed to allow most systems to be optimized in an efficient way. Quite often, however, problems arise. The following notes are intended as background material for use when things go wrong.

Overriding the default options

In the EigenFollowing geometry optimization method, the geometry is changed on each cycle; if the ΔH_f decreases, the cycle is completed. If it does not drop, the step-size is reduced, and the ΔH_f recalculated. Only when the ΔH_f decreases, compared to the previous cycle, is the current cycle considered to be successful. During the calculation, the confidence level or trust radius is continuously checked. If this becomes too small, the calculation will be stopped. This can readily happen if (a) the geometry was already almost optimized; (b) a reaction path or grid calculation is being performed; (c) if the geometry is in internal coordinates and “big rings” are involved; or (d) if the gradients are not correctly calculated (in a complicated C.I., for example).

For cases (a) and (b), add LET and DDMIN=0. In case (c) use either mixed coordinates or entirely Cartesian coordinates. Case (d) is difficult—if nothing else works, add NOANCI; this will always cause the derivatives to be correctly calculated, but will also use a lot of time.

Adding LET and DDMIN=0 is often very effective, particularly when reaction paths are being calculated. The first geometry optimization might take more cycles, but the resulting Hessian matrix is better tempered, and subsequent steps are generally more efficient.

Locating Transition States

Unlike optimizing ground states, locating transition states involves deciding on an efficient strategy. In general, there are three stages in locating transition states:

1. Generating a geometry in the region of the transition state.
2. Refining the transition state geometry.
3. Characterizing the transition state.

Of these three, the first is by far the most difficult. The following approaches are suggested as potential strategies for generating a geometry in the region of the transition state.

For narcissistic reactions (reactions in which the reactants and products are the same, e.g. the inversion of ammonia.

- Use geometry constrains, e.g. SYMMETRY, to lock the geometry in the symmetry of the potential transition state.
- Minimize the ΔH_f .
- Verify that the system is a transition state. If it has more than one negative force constant, use another method.

For a bond making-bond breaking reaction (e.g., an S_N2 reaction)

- Use SYMMETRY to set the two bonds equal. It does *not* matter that the bonds are of different type. For example, to locate the transition state for Br^- reacting with CH_4 to give CH_3Br , the C–Br and C–H bonds would be set equal.
- Optimize the geometry, to minimize the ΔH_f . Any geometry optimizer could be used, but of course the default optimizer should be tried first.
- Remove the symmetry constraint, and locate the transition state using TS. At this point, the main geometric change is to adjust the two bond lengths involved in the reaction.

For barriers to rotation, inversion, or other simple reaction that does not involve making or breaking bonds

- Optimize the starting geometry.
- Optimize the final geometry.
- Identify the coordinate that corresponds to the reaction. This is likely to be an angle or a dihedral.
- Starting with the higher energy geometry, use a (see p. 224) to drive the reaction in the direction of the other geometry. Use about 20 points, and go about half way to the other geometry—the transition state is likely to be between the higher energy geometry and the half-way point.
- From the output, locate the highest energy point—this will be near to the transition state.
- Starting with the geometry of the highest energy point, repeat the path calculation. Use smaller steps (0.1 times the previous step is usually OK), and again do 20 points.
- Inspect the reaction gradient. It should drop as the transition state is approached. If it does, then use TS to refine the transition state.

For bond making or bond breaking reactions

- Identify the reaction coordinate (the bond that makes or breaks)
- Use a (see p. 224) to drive the reaction.
- The geometry of the highest point on the reaction path should then be used to start a TS calculation.

For complicated reactions (e.g. Diels Alder) For these systems, the SADDLE calculation is a suitable method.

- Optimize the reactant geometry.
- Using the same atoms in the same sequence, optimize the product geometry.
- Run the SADDLE calculation.
- If the calculation ends because “both reactants and products are on the same side of the transition state,” use two of the geometries to set up a new SADDLE calculation. Use a smaller $\text{BAR}=n.nn$, e.g., $\text{BAR}=0.03$, and re-run the calculation. If CPU time is not important, run the original data set with $\text{BAR}=0.03$.
- Use the final geometry, or the highest energy geometry, if the SADDLE does not run to completion, as the starting point for a TS calculation.

5.8 "Size" of a molecule

A useful measure of a molecule is its size. There are several possible ways of defining the size of a molecule. The definition used in MOPAC is as follows:

The first dimension is the maximum distance between any pair of atoms.

For systems of 20 or fewer atoms, this distance, and the atoms involved, is worked out explicitly. For systems of 21 or more atoms, an atom is selected; the atom, K , most distant from it is then identified, then the atom, L , most distant from K is identified. In most systems, the distance $R(K - L)$ is the first dimension. To ensure that it is, the point half-way between K and L is selected, and atom K is then re-defined as the atom most distant from that point. A new L is determined. This sequence is repeated up to 10 times, or until atoms K and L no longer change. There is no guarantee that the first dimension is, in fact, the largest distance, but it is likely to be close to the largest possible value.

The second dimension is the maximum distance in the plane perpendicular to the first dimension between any pair of atoms.

The technique that was used in determining the first dimension for systems of over 21 atoms is used here.

The third dimension is the maximum distance between any two atoms on the line perpendicular to the plane of the first two atoms.

This quantity is explicitly calculated.

Note that the second and third dimensions do *not* define the smallest rectangular slot that a molecule would go through; it will normally be slightly larger than the minimum slot. Nevertheless, the "dimensions" of a molecule can be regarded as a good measure of the size of hole that the molecule could pass through. Of course, allowance must be made for the finite size of atoms.

Monatomic systems have no "dimension", linear systems have two zero "dimensions", and flat systems have one zero "dimension".

5.9 Solid state capability

5.9.1 Constructing Data Sets

Setting up the data set for a solid is much more complicated than setting up the equivalent data set for a molecule. For example, the data set for iodine, I_2 , might take say 20 seconds to type up. For crystalline I_2 , the data set might take 20 minutes to an hour to set up. The reason for this is the presence in solids of the translation vectors. In MOPAC, these are represented by the symbol "Tv", short for Translation Vector. Translation vectors must be the last entries in the Z-matrix, are defined in terms of the positions of atoms or dummy atoms.

The recommended procedure for constructing a data set for a solid is as follows:

Build a primitive unit cell. In this work, the action of simple translations on a primitive unit cell would make the solid. That is, other operations, such as rotation, non-primitive translation, or reflection should not be necessary.

Use MAKPOL to build a cluster. The number of primitive unit cells in each direction should be sufficiently large that the shortest distance between two opposite faces in the cluster is at least 10 Ångstroms.

Using an editor (vi, Notepad, or WORD, in order of preference), add symmetry to the system.

Run using MOPAC

Worked examples of this procedure, with explanation of each step are shown at the end of this section, see P. 198.

5.9.2 The Cluster

Unlike more conventional methods, MOPAC does not normally use a fundamental unit cell. Neither does it sample the Brillouin Zone in order to model the electronic structure. Instead, it uses a large unit cell, called a 'cluster', and applies the Born-von Kármán [68, 69] periodic boundary conditions. In this discussion, the term 'solid' is intended to include polymers, layer systems, and true solids, unless otherwise indicated by the context.

If a unit cell of a solid is large enough, then a single point in k -space, the Γ point, is sufficient to specify the entire Brillouin zone. The secular determinant for this point can be constructed by adding together the Fock matrix for the central unit cell plus those for the adjacent unit cells. The periodic boundary conditions are satisfied, and diagonalization yields the correct density matrix for the Γ point.

At this point in the calculation, conventionally, the density matrix for each unit cell is constructed. Instead, the Γ -point density and one-electron density matrices are combined with " Γ -point-like" Coulomb and exchange integral strings to produce a new Fock matrix. The calculation can be visualized as being done entirely in reciprocal space, at the Γ point.

Most solid-state calculations take a very long time. These calculations, called "Cluster" calculations after the original publication, require between 1.3 and 2 times the equivalent molecular calculation.

A minor 'fudge' is necessary to make this method work. The contribution to the Fock matrix element arising from the exchange integral between an atomic orbital and all atomic orbitals which are more than half a unit cell away must be ignored.

The unit cell must be large enough that an atomic orbital in the center of the unit cell has an insignificant overlap with the atomic orbitals at the ends of the unit cell. In practice, a translation vector of more than about 7 or 8 Å is sufficient. For one rare group of compounds a larger translation vector is needed. Solids with delocalized π -systems, and solids with very small band-gaps will require a larger translation vector, in order to accurately sample k -space. For these systems, a translation vector in the order of 15–20 Ångstroms is needed.

5.9.3 Derivatives

Solid-state derivatives with respect to geometry are handled differently from molecule derivatives. If the Cartesian coordinate derivatives are printed, using `DEBUG` and `DCART`, then for a molecule with an optimized geometry all the derivatives will be zero. This is not the case for an infinite system.

An infinite system is represented by cell supplied by the user, called the Central Unit Cell, or the CUC, and the cells surrounding this CUC. When `DCART`, `LARGE`, and `DEBUG` are used in an infinite system calculation for which the geometry has been optimized, the Cartesian derivatives for all unit cells are output. Many of these will be quite large, up to about 60 kcal/mol/Å. This is not an error, rather it is a peculiarity of the way solid-state derivatives are stored.

The Cartesian derivatives of the CUC represent the sum of all forces acting on the atoms of the CUC due to all the atoms in the CUC. Thus, if the atoms in the CUC are the set (a,b,c,d,e,f) , then the Cartesian derivatives for atom a represent the forces on a due to the set (b,c,d,e,f) . The Cartesian derivatives of atom a do *NOT* include terms from the surrounding unit cells. Because of this, those atoms in the CUC which are at the cell boundaries are likely to have large derivatives.

The Cartesian derivatives of the surrounding unit cells represent the forces acting on the atoms in those cells arising from the atoms of the CUC. Again, this is an unbalanced set of forces, and those atoms near to the cell boundaries are likely to have large resultant forces.

It is possible to evaluate the total, balanced, forces acting on the atoms of the CUC. This is done by simply adding the forces acting on the atoms of the three unit cells. When the keywords given above are used, the last part of the derivative output consists of the forces acting on the CUC itself.

Only by representing the forces in this unusual manner can the information necessary for calculating the derivative of the translation vector be generated.

5.9.4 Geometry Specification for Band Structure Calculations

Before electronic band structure calculations can be done, the sequence of atoms in the polymer must be supplied in a highly specific order. For a simple polymer, the coordinates of all the atoms in the first fundamental unit cell are given. These atoms can be in any order. The next set of atoms defined are those for the next unit cell. These atoms *must* be in the same order as the atoms in the first unit cell. For band structures at least two unit cells must be defined. If more than two unit cells are defined, the atoms in the other unit cells must be defined in the same order as those in the first unit cell.

For all polymer calculations *except* band structures, the order of atoms is not important. An example of such a data set is shown on p 101 for polytetrahydrofuran. When band structures are to be calculated, . For an example, see p. 102.

Because of the difficulty in generating data sets for band-structure work, program BZ was written. Given a suitable data-set, BZ will generate a MOPAC data set which can then be used for the calculation of band structures.

5.9.5 Electronic Band Structure

In a normal cluster calculation, the Fock matrix is diagonalized to yield eigenvalues corresponding to various points in the Brillouin zone. For m unit cells, the points generated are $0, 1/m, 2/m, \dots$ up to $1/2$. If m is odd, the upper bound becomes $(m - 1)/(2m)$. No other points in the Brillouin zone can be generated by diagonalization.

In order to represent a general point, k , in the Brillouin zone, a complex secular determinant, F_k , of size n must be constructed. The elements of this matrix are

$$F_k(\lambda, \sigma) = \sum_{r=-\infty}^{r=\infty} E(\lambda, \sigma + nr) e^{-ikr2\pi}.$$

Because interactions between atomic orbitals fall off rapidly with distance, the limits of r can be truncated to include all non-vanishing elements of E , for the sake of convenience. However, these elements are precisely those which were used in the construction of the Fock matrix. Using this, and the fact that periodic boundary conditions were employed in the construction of the Fock matrix, this summation can be simplified to

$$F_k(\lambda, \sigma) = \sum_{r=0}^{r=m-1} E(\lambda, \sigma + nr) \exp(-ikr'2\pi),$$

where r' , the index of the unit cell, equals r while r is less than $m/2$, otherwise $r' = m - r$. Band structures can then readily be constructed by varying the wave-vector k over the range 0–0.5. Units of k are $2\pi/a$, where a is the fundamental unit cell repeat distance. The band structure is then constructed by simply joining the points in the order in which they are generated. Within band structures, bands of different symmetry are allowed to cross. Simply joining the points does not allow for band crossing. However, when the resulting bands are represented graphically, visual inspection readily reveals which bands should, in fact, cross.

5.9.6 Electronic Density of States

The density of states, DoS, is the spectrum of the number of energy levels per eV versus energy. While the energy levels resulting from the calculation of the band structure could be used directly for the calculation of the DoS, the

resulting DoS would be very rough as a result of the relatively coarse mesh used. A better procedure is to assume continuity of the bands, and, by using an interpolation procedure, numerically integrate. A possible complication arises from the incorrect representation of bands which should cross. In practice, however, errors due to such causes are so small as to not show up in a normal graphical representation of the DoS.

At present, the DoS is calculated in MOPAC (not in BZ), and only for one-dimensional systems, i.e., polymers.

5.9.7 Brillouin Zone: Generation of Band Structures

Using a modified cluster technique, band structures of polymers can readily be calculated. When a sufficiently large repeat unit is used, errors introduced due to the methodology of the cluster procedure become vanishingly small. Even for delocalized π systems, such as polyacetylene, accurate band structures can be generated using repeat units of about 20Å. For less highly conjugated systems, a shorter cluster length should be sufficient. In contrast to earlier oligomer work, no allowance need be made for end-effects. In addition, the set of points in the BZ to be used is determined explicitly by the step-size.

The technique outlined here is very fast compared to earlier methods [70].

Geometry optimization of clusters of the size reported here (i.e., having translation vectors of about 25Å) require only a little more time than molecules of similar size, the extra time being used to calculate the inter-unit cell interactions. Band structure calculations are also very fast. The time required depends on the size of the fundamental unit cell. For polyacetylene, this amounted to 3% of the time for a single self consistent field calculation of the cluster.

Band structures calculated using the program BZ are accurate in the sense that any errors are due to the Hamiltonian used. A more accurate method, for example a large basis set *ab initio* calculation, should yield highly accurate band structures. In addition, limited use of symmetry in the construction of the cluster secular determinant and in the geometry optimization should increase the speed of such a calculation considerably. Electrical conductivity in semiconductors is caused by holes in the valence band and electrons in the conduction band. Conductivity also depends on the hole and electron effective masses, which are readily calculable from the second derivative of the energy of the band with respect to wave-vector. Band structures for linear polymers, calculated using semiempirical methods, should be suitable for calculation of effective masses, and consequently electrical conductivity. Unfortunately, NDDO type semiempirical methods have not proven very accurate at predicting conduction band levels. As a result, in order to rapidly calculate electrical phenomena, it is likely that a combination of *ab initio* methods and the cluster technique will be necessary.

As generated by MOPAC, the Fock matrix is unsuitable for band-structure work. First, the matrix represents the cluster, not the unit cell, and second, the Fock matrix will not exhibit the high symmetry of the associated space-group. The perturbation is small, but fortunately it can readily be eliminated.

The steps involved in converting the MOPAC Fock matrix into one suitable for band-structure work are as follows:

Generation of solid-state Fock matrix

BZ assumes that the unit cells used in constructing the MOPAC cluster were supplied in the order defined in MAKPOL. Based on this assumption, the first unit cell will have the index (0,0,0). If there are N atomic functions in a unit cell, then the first N rows of the MOPAC Fock matrix will correspond to the central unit cell (CUC). Of all the unit cells, this one is the only one for which the entire Fock matrix is not present; instead only the lower-half triangle is available. However, since the CUC is symmetric, the missing data are generated by forming the transpose, i.e., $H_{i,j} = H_{j,i}$.

The Fock matrix representing the interaction of the CUC with the next unit cell, (0,0,1), or (0,0,2) if BCC is specified, is then extracted, as are all the small Fock matrices. Each Fock matrix, representing the CUC

Table 5.40: Space-group operations for O_h^7 (diamond)

$\{E 000\}$	$\{C_2(1, 1, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{I \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{\sigma_d(0, 1, -1) 000\}$
$\{C_3(1, 1, 1) 000\}$	$\{C_2(1, 0, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4(0, 0, 1) 000\}$	$\{S_6(1, 1, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3(1, -1, -1) 000\}$	$\{C_2(0, 1, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4(0, 1, 0) 000\}$	$\{S_6(1, -1, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3(-1, 1, -1) 000\}$	$\{C_2(0, 1, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4(1, 0, 0) 000\}$	$\{S_6(-1, 1, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3(-1, -1, 1) 000\}$	$\{C_2(1, 0, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4^2(0, 0, 1) 000\}$	$\{S_6(-1, -1, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3^2(1, 1, 1) 000\}$	$\{C_2(1, -1, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4^2(0, 1, 0) 000\}$	$\{S_6^5(1, 1, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3^2(1, -1, -1) 000\}$	$\{C_4(0, 0, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4^2(1, 0, 0) 000\}$	$\{S_6^5(1, -1, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3^2(-1, 1, -1) 000\}$	$\{C_4(0, 1, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{S_4(1, 1, 0) 000\}$	$\{S_6^5(-1, 1, -1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_3^2(-1, -1, 1) 000\}$	$\{C_4(1, 0, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{\sigma_d(1, 0, 1) 000\}$	$\{S_6^5(-1, -1, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_2(0, 0, 1) 000\}$	$\{C_4^3(0, 0, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{\sigma_d(0, 1, 1) 000\}$	$\{\sigma_h(0, 0, 1) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_2(0, 1, 0) 000\}$	$\{C_4^3(0, 1, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{\sigma_d(1, -1, 0) 000\}$	$\{\sigma_h(0, 1, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$
$\{C_2(1, 0, 0) 000\}$	$\{C_4^3(1, 0, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$	$\{\sigma_d(1, 0, -1) 000\}$	$\{\sigma_h(1, 0, 0) \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\}$

interacting with each unit cell, is stored in a large array, of size N^2 times the number of unit cells used. As phrases of the type “The Fock matrix representing the interaction of the CUC with unit cell (i,j,k)” are cumbersome, from here on, the term “unit cell (i,j,k)” should be understood as having the same meaning.

The indices of each unit cell is also generated and stored. However, the cluster theory assumes that the interaction matrix relating two unit cells which are separated by more than half the distance of the translation vector does not represent that interaction. Rather, it represents the interaction of two unit cells which are separated by less than half the translation vector distance. In order to conform with this definition, all unit cell indices more than half of the number of unit cells specified by the MERS keyword are changed. For example, if MERS(4, 4, 4) is used, the unit cells (0,1,1) and (2,2,2) would be unchanged, but unit cells(0,1,3) and (0,0,4) would become (0,1,-2) and (0,0,-1), respectively.

As a result of this operation, most of the unit cells surrounding the CUC are generated. The next step is to symmetrize the Fock matrices so that they have the symmetry of the space group. Note that if symmetrization is not done, the band-structures generated would be almost, but not quite, identical to those which use symmetrized Fock matrices.

Symmetrization of Fock matrices

From group-theory we know that if a matrix is operated on by every operation of a group exactly once, the resulting matrix will have the symmetry of that group. In other words,

$$F_{sym} = \frac{1}{M} \sum_{i=1}^M \langle R_i | F_{unsym} | R_i^T \rangle .$$

The index i covers all operations of the group, including the identity.

Space-group operations

Space-group operations differ from point-group operations in that in addition to the point-group operation, a non-primitive translation may be involved. Thus far, we have been using as our example the diamond lattice which is suitable for illustrating space-group operations. For convenience, we will specify a space-group operation thus: $\{R|T\}$, where “R” is a point-group operation, e.g. $C_2(0,0,1)$ or $S_6(1,1,-1)$, and “T” is a non-primitive translation, e.g. $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, or $(0,0,0)$. The term in parentheses following the point group operation indicates the axis about which the operation is to be performed. Finally, to complete the specification of a space-group operation, the point about which the operation acts must be defined. As this is most conveniently done in fractional unit cell coordinates, or crystal coordinates, the Cartesian coordinates are converted at this time into fractional unit cell coordinates.

Diamond belongs to the $Fd3m$ or O_h^7 space-group, and has octahedral symmetry; its associated point-group is O_h . The space-group operations are given in Table 5.40.

5.9.8 Examples of Solid State Data Set Construction

Diamond

The unit cell of diamond consists of two carbon atoms. If symmetry operations were allowed, only one atom would be needed, but only translation operations are allowed, so two atoms must be used.

Each carbon atom forms four single bonds with other carbon atoms. This can be used in defining the translation vectors: the effect of a translation vector acting on atom 1 would be to move it to one of the atoms attached to atom 2. Using this fact, the data set can easily be made:

First attempt at a data set for MAKPOL.

```
MERS=(4,4,4) Diamond
C
C 1.545 1 0 0 0 0 1
Tv 2.523 1 35.26439 10 0 1 2
Tv 2.523 1 35.26439 1 120 1 1 2 3
Tv 2.523 1 35.26439 1 240 1 1 2 3
```

In this description, the number of unit cells to be used is too small: 4 by 4 by 4. Ideally, at least a 6 by 6 by 6 system should be used. The smaller number is used here purely for the purposes of illustration.

This data set is, however, not ideal, for the following reasons:

The angle between the translation vectors is 60 degrees. This means that the unit cell must be very large in order for the distance between opposite faces to be at least 10 Ångstroms. Diamond is a Body-Centered-Cubic lattice. This means that every odd cell (e.g. 001,111,012, etc.) is missing. By specifying BCC, the angle between the translation vectors can be increased to 90 degrees. This is the ideal angle to maximize the distance between faces. The system has a lot of symmetry. By adding "SYMMETRY" two objectives can be met: First, symmetry can be used in defining the Tv. This is not very important in this unit cell, but does make it easier to change the bond-length of atom 2, in that any change in this distance is automatically made in the Tv. Second, if SYMMETRY is present, MAKPOL will automatically add symmetry to the data set.

```
Data set for MAKPOL. File name: Make_diamond.dat
SYMMETRY MERS=(4,4,4) BCC
Diamond, 64 atoms
C
C 1.545 1 0 0 0 0 1
Tv 1.784 0 54.73561 0 0 0 1 2
Tv 1.784 0 54.73561 0 120 0 1 2 3
Tv 1.784 0 54.73561 0 240 0 1 2 3

2 19 1.1547005 3 4 5
```

When this data set is run using MAKPOL, the following data set is generated:

```
SYMMETRY MERS=(4,4,4) BCC
Diamond, 64 atoms
```



```

C 0.000000 0 0.000000 0 0.000000 0
C 1.545000 1 0.000000 0 0.000000 0 1
C 3.568025 1 54.735610 1 0.000000 0 1 2
C 1.545000 0 125.264390 1 0.000000 1 3 1 2
C 2.522974 1 35.264390 1 60.000000 1 1 2 3
C 1.545000 0 144.735610 1 0.000000 0 5 1 2
C 2.522974 0 135.000000 1 45.000000 1 3 1 2
C 1.545000 0 144.735610 0 -90.000000 1 7 3 1
C 2.522974 0 90.000000 1 90.000000 0 5 1 2
C 1.545000 0 90.000000 0 -144.735610 1 9 5 1

```

(many lines deleted)

```

C 1.545000 0 90.000000 0 160.528779 1 61 57 41
C 2.522974 0 60.000000 0 -35.264390 0 59 43 26
C 1.545000 0 90.000000 0 160.528779 0 63 59 43
XX 2.522974 0 90.000000 0 180.000000 0 7 3 1
XX 2.522974 0 90.000000 0 -125.264390 0 29 9 5
XX 2.522974 0 180.000000 0 0.000000 0 53 41 42
Tv 7.136049 1 0.000000 0 0.000000 0 1 65 2
Tv 7.136049 0 0.000000 0 0.000000 0 1 66 2
Tv 7.136049 0 0.000000 0 0.000000 0 1 67 2

```

```

2 1 4 6 8 10 12 14 16 18 20 22
2 1 24 26 28 30 32 34 36 38 40 42
2 1 44 46 48 50 52 54 56 58 60 62
2 1 64
4 3 6 11 12 15 16 18 33 34 35 36
4 3 39 40 43 44 45 51 52 53 54 55
4 3 56 59 60 67
5 1 7 9 11 13 15 17 19 21 23 25
5 1 27 29 31 33 35 37 39 41 43 45
5 1 47 49 51 53 55 57 59 61 63 65
5 1 66 67
5 2 17
5 14 17

```

(many lines deleted)

Before running this data set, more symmetry can be added "by hand." Every angle is symmetry defined, and does not need to be optimized, therefore every angle and dihedral optimization flag can be set to zero. Every angle and dihedral symmetry relation defined at the end of the data set can also be deleted. These are the lines that start with a number followed by a 2, a 3, or a 14. Every distance can be related to the bond-length of atom 2. The position of atoms 3 and 5, and the length of the translation vector Tv can be defined by distances that are exactly $\text{Sqrt}(16/3)$, $\text{Sqrt}(8/3)$ and $\text{Sqrt}(64/3)$ times the C_2-C_1 distance, respectively. These symmetry relations can be defined using MOPAC symmetry function 19. This has the form: Defining-atom 19 multiplier dependent atom(s)

When these changes are made to the data set, the final data set is produced. This is:

SYMMETRY MERS=(4,4,4) BCC

Diamond, 64 atoms

C	0.000000	0	0.000000	0	0.000000	0			
C	1.545000	1	0.000000	0	0.000000	0	1		
C	3.568025	0	54.735610	0	0.000000	0	1	2	
C	1.545000	0	125.264390	0	0.000000	0	3	1	2
C	2.522974	0	35.264390	0	60.000000	0	1	2	3
C	1.545000	0	144.735610	0	0.000000	0	5	1	2
C	2.522974	0	135.000000	0	45.000000	0	3	1	2

(many lines deleted)

C	1.545000	0	90.000000	0	160.528779	0	61	57	41
C	2.522974	0	60.000000	0	-35.264390	0	59	43	26
C	1.545000	0	90.000000	0	160.528779	0	63	59	43
XX	2.522974	0	90.000000	0	180.000000	0	7	3	1
XX	2.522974	0	90.000000	0	-125.264390	0	29	9	5
XX	2.522974	0	180.000000	0	0.000000	0	53	41	42
Tv	7.136049	0	0.000000	0	0.000000	0	1	65	2
Tv	7.136049	0	0.000000	0	0.000000	0	1	66	2
Tv	7.136049	0	0.000000	0	0.000000	0	1	67	2

2	19	2.3094	3								
2	19	1.6330	5								
2	19	4.6188	68								
2	1	4	6	8	10	12	14	16	18	20	22
2	1	24	26	28	30	32	34	36	38	40	42
2	1	44	46	48	50	52	54	56	58	60	62
2	1	64									
5	1	7	9	11	13	15	17	19	21	23	25
5	1	27	29	31	33	35	37	39	41	43	45
5	1	47	49	51	53	55	57	59	61	63	65
5	1	66	67								
68	1	69	70								

Is the use of symmetry worth all this effort? Most definitely! If symmetry is NOT used, then for this small system of 64 atoms, 195 parameters would need to be optimized. That is, $3 \times 64 - 6$ for the 64 atoms plus 9 parameters for the three translation vectors. Optimization of a system with 195 unknowns would take much longer than for a system with precisely one unknown. For a more realistic system, involving 6 by 6 by 6 primitive unit cells, symmetry would lower the complexity of the calculation from 651 unknowns to precisely 1.

5.10 Point Group Theory

This Section is based on the original work of Peter Bischof in the UMNDO program, and made available to me by Dr David Danovich.

Some point-group theory has been added to MOPAC. The main functionalities added are:

- ‘Normal’ symmetry relationships are now automatically recognized if AUTOSYM is specified.
- The symmetry of the system is printed both at the start of the run and at the end. If the point-group changes, the change will be shown in the different point-group symbols.
- Molecular orbitals will be characterized by Irreducible Representation (I.R.).
- Normal coordinates generated in the vibrational calculation will be characterized by I.R.
- State functions will be characterized by I.R.
- All rotation groups up to order 8, except D_{8d} , are available.
- The cubic groups T , T_h , T_d , O , O_h , I , and I_h are available.
- The infinite groups $C_{\infty v}$, $D_{\infty h}$ and R_3 are available.
- In FORCE or vibrational frequency calculations, symmetry will be used to accelerate the calculation, thus a calculation of benzene would involve two atoms, a C and a H atom, to be calculated, rather than the normal 12 atoms.
- In vibrational frequency calculations, the Hessian or force matrix will be symmetrized.

$$F_{ij} = \frac{1}{h} \sum_h R(h)^T F'_{ij} R(h)$$

This eliminates the normal small deviations from exact symmetry (see also p. 223 for a qualification).

Limitations

- Group D_{8d} is missing. This group is characterized by the presence of a 16-fold S_n axis. Only S_n operations up to S_{12} are checked for. As a result, D_{8d} would not be recognized. However, this is a rare point-group, and its loss should not be important.
- Some systems which are insufficiently near to a given point group will be assigned to the nearest sub-group. For example, if SF_6 is distorted so that two opposite F atoms are at a different distance to the other four, the system might be classified as O_h or D_{4h} , depending on the degree of distortion. This shows up mainly in methyl groups, e.g. neopentane, in which optimization normally stops before the angles of the hydrogens are fully optimized.

5.10.1 Representation of Point Groups

The 57 groups recognized in MOPAC are given in Table 5.41.

Each point group is represented by a subset of the associated point-group table. For example, the group D_{2h} is represented by the subset shown in Table 5.42. The operations selected for the subgroup are the identity, E, and that minimum set of operations which is sufficient to allow all the operations to be generated as products of these

Table 5.41: Point Groups available within Symmetry Code

C ₁	C _s	C _i					O
C ₂	C _{2v}	C _{2h}	D ₂	D _{2d}	D _{2h}		T
C ₃	C _{3v}	C _{3h}	D ₃	D _{3d}	D _{3h}		T _d
C ₄	C _{4v}	C _{4h}	D ₄	D _{4d}	D _{4h}	S ₄	T _h
C ₅	C _{5v}	C _{5h}	D ₅	D _{5d}	D _{5h}		O _h
C ₆	C _{6v}	C _{6h}	D ₆	D _{6d}	D _{6h}	S ₆	I I _h
C ₇	C _{7v}	C _{7h}	D ₇	D _{7d}	D _{7h}		C _{∞v} D _{∞h}
C ₈	C _{8v}	C _{8h}	D ₈		D _{8h}	S ₈	R ₃

Table 5.42: Subset of Group D_{2h}

Γ	E	C _{2y}	C _{2z}	I
A _g				
B _{1g}	1	1	-1	1
B _{2g}	1	-1	1	1
B _{3g}	1	-1	-1	1
A _u	1	1	1	-1
B _{1u}	1	1	-1	-1
B _{2u}	1	-1	1	-1
B _{3u}	1	-1	-1	-1

operations. Thus, for the highest finite point group, I_h, the generating operations are: E, I, C₃, and C₅. Although it is not obvious, all 120 operations of the group can be generated as products of these four operations.

Each point-group is assumed to contain the totally symmetric representation, here A_{1g}. Operations are represented as 3 × 3 Euler matrices, thus C_{2x}, C_{2y} and C_{2z} would be represented as in Figure 5.15 All operations not given can be generated as products of operations already known, thus C_{2x} = C_{2y} × C_{2z}.

In order to minimize storage, the characters in character tables are stored separately from the point groups. This allows, e.g., C_{2v}, C_{2h}, and D₂ to use the same character table.

5.10.2 Identification of Point-Groups

Infinite Groups

In order to identify the molecular point-group the system must be oriented in a specific way. Four families of point-groups are checked for: (1) the infinite groups, (2) the cubic groups, (3) groups with one high-symmetry axis, and (4) the Abelian groups. Each family is treated differently. First, the moments of inertia are calculated. If all are zero, the system is a single atom, and the associated group is R₃. If two moments are zero, the system is either C_{∞v} or D_{∞h}; the presence of a horizontal plane of symmetry distinguishes between them.

$$\begin{array}{ccc}
 C_{2x}: \begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{vmatrix} &
 C_{2y}: \begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{vmatrix} &
 C_{2z}: \begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{vmatrix}
 \end{array}$$

Figure 5.15: Representation of Symmetry Operations

Cubic Groups

Having eliminated the infinite groups, the three moments of inertia are checked to see if they are all the same. If they are, then the system is cubic. Cubic systems are oriented by identifying atoms of the set nearest to the center of symmetry. If there are 4, 6, 8, 12, or 20 of these, and the number of equidistant nearest neighbors is 3, 4, 3, 5, or 3, respectively, then the atoms are at the vertices of one of the Platonic solids (tetrahedron, octahedron, cube, icosahedron, dodecahedron), and therefore all atoms of the set lie on high-symmetry axes. The first atom is selected and used to define the z axis.

If the number of atoms in the set does not correspond to any of the Platonic solids, then the set is checked for the existence of an equilateral triangle, a square, or a regular pentagon. When one of these is found, the center of the polygon is used to define the z axis. An example of this type of system is C_{60} , Buckminsterfullerene, which has a five-fold axis going through the center of a pentagonal face.

Once the z axis is identified, the system is checked for C_n axes, $n=3$ to $n=8$. To complete the orientation, the system is rotated about the z axis so that two atoms, having equal z coordinates, have equal y coordinates. The existence of rotation axes which are not coincidental with the z axis and the presence or absence of a center of inversion are then used to identify which cubic group the system belongs to.

Other Degenerate Groups

If the system has still not been identified, then the two equal moments of inertia indicate a degenerate point group. As with the cubic groups, the y and z axes (and, by implication, the x axis) are identified. The system is oriented, and the C_n and S_n axes identified.

The degenerate groups, C_n , C_{nv} , C_{nh} , D_n , D_{nd} , D_{nh} , S_n , are distinguished by the existence or absence of C_2 axes perpendicular to the z axis, and by planes of symmetry.

Abelian Groups

All that remains are the Abelian groups, C_1 , C_2 , C_i , C_s , C_{2v} , C_{2h} , D_2 , and D_{2h} . After orienting the molecule, the axes are swapped around so that the normal convention for orienting Abelian systems is obeyed. For groups C_1 , C_2 , C_i , and C_s , there is no possibility for ambiguity. For C_{2v} and D_2 , however, the orientation of the system affects the labels of the irreducible representations. To prevent ambiguity, the convention for orienting Abelian molecules is:

- The axis with the largest number of atoms is the z axis.
- The plane with the largest number of atoms that includes the z axis is the yz plane.

Thus for ethylene, the π orbitals point along the x axis.

Tolerance

Normally, molecular geometries do not exactly correspond to the idealized point-group. Thus benzene might have slightly different bond-lengths and angles. Of course, symmetry could be used to prevent this, but in the discussion here we assume that the symmetry of the system is unknown. To allow for these slight distortions, a small tolerance is built in to the tests for symmetry elements. This starts off at 0.1\AA , but may be tightened automatically if ambiguities are detected. An example of such an ambiguity is found in tropylium, $C_7H_7^+$ ion, where the C-C distance is 1.4\AA . Rotating the ring by 45 degrees (a C_8 operation) would place the atoms at a distance of only 0.18\AA from equivalent positions. C_7 and C_8 would thus give almost identical results. To resolve such ambiguities, when they arise, the tolerance is reduced, and the test re-run.

Even with this feature, some systems still resist classification. A distorted geometry might have some, but not all, elements of a high point group. Perhaps a distorted benzene has a $C_2(z)$ and a $C_3(z)$, but not a $C_6(z)$, impossible in a real system. As such it would appear to be different from all real point groups. To accommodate such defects a descent in symmetry is carried out. This consists of checking each point-group in turn, in order of decreasing symmetry. Once all of the elements of a point group are satisfied, the system is assigned to that point group, even if the system contains more symmetry than the point group.

By these two devices, a variable tolerance and the descent in symmetry, most systems should be identified correctly, or at least as a sub-group of the full point group.

5.10.3 Orientation of the Abelian groups C_{2v} and D_{2h}

Unlike all other groups, two of the Abelian groups, C_{2v} and D_{2h} , present novel problems in assigning the irreducible representations. For most groups, the symmetry axis is obvious, or if there are several axes, the principal axis is obvious. For C_{2v} and D_{2h} an ambiguity exists. Consider, for example, ethylene, a system of point-group D_{2h} . Should the z axis be perpendicular to the plane of the molecule—that is the unique direction, or should it go through the two carbon atoms—that is also a unique direction, but for a different reason, or should it be the third orthogonal direction—which is also unique. The choice of z axis is important in order to correctly assign the B_{1g} and B_{1u} of point-group D_{2h} . For both C_{2v} and D_{2h} the x and y axes must also be unambiguously defined in order to distinguish between B_{2g} and B_{3g} and between B_{2u} and B_{3u} of D_{2h} , and between B_1 and B_2 of C_{2v} . Clearly a convention has to be decided upon, otherwise one person's B_{1g} might be a second person's B_{2g} and a third person's B_{3g} .

The convention used in MOPAC is the following:

If there are three C_2 axes, the one with the largest number of atoms unmoved by a C_2 operation is z . If there is only one C_2 axis, that is z .

Once z is defined, the y axis is defined as the axis of the remaining two axes which has the larger number of atoms unmoved by the σ symmetry operations.

The x axis is the remaining axis.

To see how this works, consider ethylene, with the C–C axis being along the x direction, and the plane of the system being xy . Under the eight operations of D_{2h} , E , C_{2z} , C_{2y} , C_{2x} , σ_{xy} , σ_{xz} , σ_{yz} , and i , the number of atoms unmoved are 6, 0, 0, 2, 6, 2, 0, and 0 respectively.

From this it follows that the old x axis is now re-defined as the z axis. The new y axis has to be chosen based on the number of atoms unmoved under the σ_{xy} and σ_{xz} operations (6 and 2). The new y axis is defined as being the old y axis. The remaining new axis x therefore is the old z axis.

The overall result is that the symmetry axes in ethylene are defined as: z - along the C–C bond; y - in the molecular plane, perpendicular to the C–C bond, and x - out-of-plane.

In MOPAC the orientation of the molecule is defined by the user, therefore the assignment of the symmetry axes might be confusing. If the irreducible representations of ethylene are assigned, and the atoms are defined using internal coordinates in the order C, C, H, H, H, H, then the p orbitals will reflect the orientation used in the previous discussion, but the representations will be correct according to the conventions just defined.

5.10.4 Molecular Orbitals

Each M.O. is subjected to the operation

$$\psi' = |R|\psi \rangle$$

from which the expectation value

$$\chi = \langle \psi' | \psi \rangle$$

can readily be calculated.

All χ 's within a given degenerate manifold are summed:

$$\chi_i = \sum_j \chi_j^{(i)}$$

where j runs over all components of the degenerate manifold i .

This results in a set of characters which can be compared to those stored in the character tables.

Since molecular orbitals involve single electrons, the irreducible representations are rendered into lower case before printing.

5.10.5 Normal Coordinates

Analysis of normal coordinates is a little simpler than that of molecular orbitals in that every atom contributes precisely three components to each normal coordinate, an x , y , and z component. These transform in the same way as the p_x , p_y , and p_z atomic orbitals.

Because normal coordinates are states, the first letter of the irreducible representation is capitalized.

5.10.6 States

Calculating the characters for electronic states is much more complicated than that for M.O.s or normal coordinates.

Consider the effect of an operation, R , on a state, Φ_a . The character of the operation is given by

$$\chi_{R,a} = \langle \Phi_a | R | \Phi_a \rangle .$$

A state function can be expressed as a linear combination of microstates:

$$\Phi_a = \sum_j C_{ja} \Psi_j,$$

so the character of the operation on the state function can be written in terms of microstates as

$$\chi_{R,a} = \sum_i \sum_j C_{ia} C_{ja} \langle \Psi_i | R | \Psi_j \rangle .$$

Each microstate, Ψ_j , can be represented by a Slater determinant of N molecular orbitals :

$$\Psi_j = \frac{1}{\sqrt{N!}} \sum_{P=1}^{N!} (-1)^P P \left(\prod_{k=1}^N \psi_k^j \right)$$

where the molecular orbitals in the microstate consist of a selection of the M.O.s in the active space. Before we continue, let us examine this idea:

Consider a full set of M.O.s:

$$\psi_1 \psi_2 \psi_3 \psi_4 \psi_5 \psi_6 \psi_7 \psi_8 \psi_9 \psi_{10} \psi_{11} \psi_{12} \psi_{13} \psi_{14} \psi_{15} \psi_{16} .$$

Let the active space be the M.O.s from 8 to 11. Then microstates containing two electrons would be:

$$\psi_8 \psi_9 \quad \psi_8 \psi_{10} \quad \psi_8 \psi_{11} \quad \psi_9 \psi_{10} \quad \psi_9 \psi_{11} \quad \psi_{10} \psi_{11} .$$

These microstates could be represented by M.O. orbital occupancies.

1100 1010 1001 0110 0101 0011.

Remember that the M.O.s here can be of either α or β spin.

To continue, we need to evaluate $\langle \Psi_i | R | \Psi_j \rangle$. This can be expressed in terms of M.O.s as:

$$\langle \Psi_i | R | \Psi_j \rangle = \frac{1}{N!} \sum_{P=1}^{N!} (-1)^P P \langle \prod_{k=1}^N \psi_k^i | R | \sum_{Q=1}^{N!} (-1)^Q Q \left(\prod_{l=1}^N \psi_l^j \right) \rangle.$$

For convenience, we will represent the integral $\langle \psi_k^i | R | \psi_l^j \rangle$ by χ_{kl}^{ij} . This integral can be described as "The integral over M.O. ψ_k in microstate Ψ_i with the result of operator R acting on M.O. ψ_l in microstate Ψ_j ."

Using this abbreviation, $\langle \Psi_i | R | \Psi_j \rangle$ can be written as:

$$\langle \Psi_i | R | \Psi_j \rangle = \frac{1}{N!} \sum_{P=1}^{N!} \sum_{Q=1}^{N!} (-1)^P (-1)^Q P \prod_{k=1}^{N!} Q \prod_{l=1}^{N!} \chi_{kl}^{ij}.$$

Although it is not immediately obvious, the right-hand term is a determinant, of order N :

$$\langle \Psi_i | R | \Psi_j \rangle = \begin{vmatrix} \chi_{11}^{ij} & \chi_{21}^{ij} & \chi_{31}^{ij} & \cdots \\ \chi_{12}^{ij} & \chi_{22}^{ij} & \chi_{32}^{ij} & \cdots \\ \chi_{13}^{ij} & \chi_{23}^{ij} & \chi_{33}^{ij} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix}.$$

For our purposes, solution of the determinant is best done explicitly. To see why, note that the number of M.O.s involved in the C.I. (the active space) is very small. Because of this, the number of electrons, N , in the Slater determinants is also small; N has a maximum value of 20. Next, use can be made of the fact that no point-group operation can mix α and β electrons. This allows the integral to be split into two parts, each of which has a maximum value of $N=10$. Finally, remember that N is the number of electrons, not M.O.s, used in the active space. A system of N electrons has the same symmetry as a system in which all the M.O.s which were occupied were replaced with all the M.O.s which were not occupied (the positron equivalent). (This assumes that if every M.O. were occupied, then the state of the system would be totally symmetric.) Using this fact, we can replace the N occupied M.O.s with N' unoccupied M.O.s, if $N' < N$.

When these three points are considered, we see that N has a maximum value of 5 (for a system of 10 M.O.s). Each case can be considered separately.

For $N = 1$:

$$\langle \Psi_i | R | \Psi_i \rangle = \frac{1}{1} \sum_{P=1}^1 \sum_{Q=1}^1 (-1)^P (-1)^Q P \prod_{k=1}^1 Q \prod_{l=1}^1 \chi_{kl}^{ii}$$

or

$$\langle \Psi_i | R | \Psi_i \rangle = \langle \psi_1^i | R | \psi_1^i \rangle = \chi_{11}^{ii}.$$

For $N=2$:

$$\langle \Psi_i | R | \Psi_j \rangle = \frac{1}{2!} \sum_{P=1}^{2!} \sum_{Q=1}^{2!} (-1)^P (-1)^Q P \prod_{k=1}^2 Q \prod_{l=1}^2 \chi_{kl}^{ij}$$

or

$$\langle \Psi_i | R | \Psi_j \rangle = \langle \psi_1^i | R | \psi_1^j \rangle - \langle \psi_2^i | R | \psi_2^j \rangle + \langle \psi_1^i | R | \psi_2^j \rangle - \langle \psi_2^i | R | \psi_1^j \rangle$$

or

$$\langle \Psi_a | R | \Psi_a \rangle = \chi_{11}^{ij} \chi_{22}^{ij} - \chi_{12}^{ij} \chi_{21}^{ij}.$$

For $N=3$:

$$\begin{aligned} \langle \Psi_i | R | \Psi_j \rangle &= \frac{1}{3!} \sum_{P=1}^{3!} \sum_{Q=1}^{3!} (-1)^P (-1)^Q \\ \langle \Psi_i | R | \Psi_j \rangle &= \langle \psi_1^i | R | \psi_1^j \rangle \langle \psi_2^i | R | \psi_2^j \rangle \langle \psi_3^i | R | \psi_3^j \rangle \\ &\quad - \langle \psi_1^i | R | \psi_1^j \rangle \langle \psi_2^i | R | \psi_3^j \rangle \langle \psi_3^i | R | \psi_2^j \rangle \\ &\quad - \langle \psi_1^i | R | \psi_2^j \rangle \langle \psi_2^i | R | \psi_1^j \rangle \langle \psi_3^i | R | \psi_3^j \rangle \\ &\quad + \langle \psi_1^i | R | \psi_2^j \rangle \langle \psi_2^i | R | \psi_3^j \rangle \langle \psi_3^i | R | \psi_1^j \rangle \\ &\quad + \langle \psi_1^i | R | \psi_3^j \rangle \langle \psi_2^i | R | \psi_1^j \rangle \langle \psi_3^i | R | \psi_2^j \rangle \\ &\quad - \langle \psi_1^i | R | \psi_3^j \rangle \langle \psi_2^i | R | \psi_2^j \rangle \langle \psi_3^i | R | \psi_1^j \rangle \end{aligned}$$

or

$$\begin{aligned} \langle \Psi_i | R | \Psi_j \rangle &= \chi_{11}^{ij} \chi_{22}^{ij} \chi_{33}^{ij} + \chi_{12}^{ij} \chi_{23}^{ij} \chi_{31}^{ij} + \chi_{13}^{ij} \chi_{21}^{ij} \chi_{32}^{ij} \\ &\quad - \chi_{11}^{ij} \chi_{23}^{ij} \chi_{32}^{ij} - \chi_{12}^{ij} \chi_{21}^{ij} \chi_{33}^{ij} - \chi_{13}^{ij} \chi_{22}^{ij} \chi_{31}^{ij} \end{aligned}$$

For higher numbers of electrons, the associated determinant is solved using standard methods.

The total character, $\langle \Psi_\alpha | R | \Psi_\alpha \rangle$, is obtained by multiplying the characters for the α and β parts together:

$$\langle \Psi_\alpha | R | \Psi_\alpha \rangle = \langle \Psi_\alpha^\alpha | R | \Psi_\alpha^\alpha \rangle \langle \Psi_\alpha^\beta | R | \Psi_\alpha^\beta \rangle .$$

If the positron equivalent is taken for only one set of electrons, e.g. either the α or the β set, but not both, then the character has to be multiplied by the determinant of the M.O. transform.

These expressions can then be used in

$$\chi_{R,a} = \sum_i \sum_j C_{ia} C_{ja} \langle \Psi_i | R | \Psi_j \rangle .$$

to give the expectation value for the state. Finally, if the state is degenerate, the character is given by summing the components of the state.

For the atom, the Russell-Saunders coupling scheme can be reproduced. States allowed are $S, P, D, F, G, H, I, K, L$, and M . This set is more than sufficient to allow all possible Russell-Saunders states spanned by a basis set of s, p , and d orbitals to be represented. The highest angular momentum achievable with such a basis set is 8, i.e. L . For simpler atoms (ones with only a $s-p$ basis set) the allowed states are $p^0, p^6: {}^1S_g, p^1, p^5: {}^2P_u, p^2, p^4: {}^1S_g + {}^3P_g + {}^1D_g, p^3: {}^4S_u + {}^2P_u + {}^2D_u$.

For the axial infinite groups, allowed states are: $\Sigma, \Pi, \Delta, \Phi$, and Γ . Even quite simple systems can achieve quite high angular momentum, thus acetylene, with a C.I. = 4 (the HOMO π and LUMO π^*) will contain a ${}^1\Gamma_g$ state, i.e., the angular momentum will be 4.

At present J-J coupling is not supported.

5.11 Level of Precision within MOPAC

Several users have criticized the tolerances within MOPAC. The point made is that significantly different results have been obtained when different starting conditions have been used, even when the same conformer should have resulted. Of course, different results must be expected—there will always be small differences—nonetheless any differences should be small, for example, heats of formation (ΔH_f) differences should be less than about 0.1 kcal/mol. MOPAC has the flexibility to allow users to specify a much higher precision than the default when circumstances warrant it.

Table 5.43: Fundamental Physical Constants

Physical Constant	Symbol	Value	Units
Speed of Light	c	299 792 458	m sec^{-1} (Definition)
Planck constant	h	$6.626\ 075\ 5(40) \times 10^{-34}$	J sec
		$6.626\ 075\ 5(40) \times 10^{-27}$	erg s
Avogadro constant	N	$6.022\ 136\ 7(36) \times 10^{23}$	mol^{-1}
Molar gas constant	R	1.987 215 6	cal/mol/degree
		8.314 510(70)	J/mol/K
Volume of 1 mol of gas	V_0	22.414 10(19)	l/mol (at 1 atm, 25 C)
Electron volt	eV	$1.602\ 177\ 33(49) \times 10^{-19}$	J
Electron charge	e	$1.602\ 177\ 33(49) \times 10^{-19}$	C
Hartree	E_h	27.211 396 1(81)	eV
Electrostatic energy	$E_h a_0$	14.399 651 782 565	eV
Bohr radius	a_0	$0.529\ 177\ 249(24) \times 10^{-10}$	m
Boltzmann constant	$k = R/N$	$1.380\ 658(12) \times 10^{-23}$	J/K
		$1.380\ 658(12) \times 10^{-16}$	erg/K
pi	π	3.141 592 653 589 79	
Joule	J/cal	4.184	J/cal (Definition)
cm^{-1}	hc/eV	$1.239\ 842\ 4 \times 10^{-4}$	eV
cm^{-1}	$hcN/(1000\text{J/cal})$	$2.859\ 144 \times 10^{-3}$	kcal/mol
cm^{-1}	$hcN/(\text{J/cal})$	2.859 144	cal/mol
cm^{-1}		$1.196\ 266 \times 10^8$	erg = dyne \AA^{-1}
Atomic unit (a.u.)		$8.657\ 10 \times 10^{-33}$	e.s.u.
a.u.		$2.541\ 747\ 8 \times 10^{-40}$	Debye
a.u.		51.422 082	V m^{-1}
kcal/mol		$6.947\ 700 \times 10^{-3}$	erg
1 J		$1. \times 10^7$	erg
1 eV		23.060 542 301 389	kcal/mol
		627.509 6	kcal/mol
1 atm		$1.013\ 25 \times 10^5$	Pa
		$1.013\ 25 \times 10^6$	dyn/cm^2 (Definition)

Note: The precision of derived constants should *not* be used as an indication of their accuracy. The uncertainty in the fundamental constants is given in parenthesis after the value.

5.11.1 Fundamental Physical Constants

The fundamental physical constants used in MOPAC were updated in 1993 to conform with the 1986 CODATA recommendations [71]. The constants used in MOPAC are given in Table 5.43. As a result of this update, all calculated quantities in MOPAC, except molecular weight, will change slightly when compared to earlier MOPACs (MOPAC 6 and earlier). Most of the time, changes in ΔH_f are less than 0.1 kcal/mol. It is not anticipated that the physical constants will change again. If they do, however, the effect on calculated properties should be very small.

The derived quantities, AM, AD, AQ, EISOL, DD, and QQ are functions of the fundamental constants. Rather than change all of these, starting with MOPAC 93, they are evaluated at the start of each calculation. This is a quick operation, taking only about 0.1s, and prevents any mistakes being introduced due to human error.

5.11.2 Various precision levels

In normal (non-publication quality) work the default precision of MOPAC is recommended. This will allow reasonably precise results to be obtained in a reasonable time. Unless this precision proves unsatisfactory, use this

```

Line 1: GRADIENTS 1SCF DEBUG DCART
Line 2: EXAMPLE OF INTERNAL COORDINATE DERIVATIVES ZERO
Line 3: AND CARTESIAN DERIVATIVES LARGE
Line 4: 0
Line 5: C 1.2075664 1
Line 6: H 1.2114325 1 103.347931 1 0 0 2 1
Line 7: H 1.0904182 1 180.000000 1 90 1 2 1 3

```

Figure 5.16: Example of Spurious Stationary Point

default for all routine work.

The best way of controlling the precision of the geometry optimization and gradient minimization is by specifying a gradient norm which must be satisfied. The gradient norm is the scalar of the vector of derivatives of the energy with respect to the geometric variables flagged for optimization. I.e.,

$$\text{GNORM} = \sqrt{\sum_i \left(\frac{d(\Delta H_f)}{dx_i} \right)^2}$$

where x_i refers to coordinates flagged for optimization. Note that the calculated GNORM may be very small and at the same time the geometry might not be at a stationary point. This can easily happen when (a) less than $3N-6$ coordinates are flagged for optimization, (b) SYMMETRY has been used incorrectly, or (c) (less common) only $3N-6$ coordinates are flagged for optimization and dummy atoms are used. If any one of these three conditions occurs, then the warning message, "WARNING – GEOMETRY IS NOT AT A STATIONARY POINT" will be printed.

A less common, but not unknown, situation arises when internal coordinates are used. In this strange situation, the internal derivatives might all be zero, but the Cartesian derivatives are large. An example of such a system is shown in Figure 5.16.

In this strange system, the bond-angle of the second hydrogen is 180° , and the dihedral is 0° . Obviously, the derivative of the dihedral will be zero. The derivative of the angle is not so obvious. If the angle changes, then the fourth atom will move out of the plane of the other three atoms. The energy will change in the same way regardless of whether the angle increases or decreases; therefore, the derivative of the angle must be zero.

Because of the unusual nature of this type of system, users may be unaware of the danger. If such a system is detected, a warning will be given and the job stopped. For all other cases, the "WARNING – GEOMETRY IS NOT AT A STATIONARY POINT" message will be printed on completion of the calculation. In the unlikely event that the calculation should not be stopped when the strange system is detected, calculation can be continued by specifying GEO-OK.

Modification of GNORM is done via the keyword GNORM=*n.nn*. Altering the GNORM automatically disables the other termination tests resulting in the gradient norm dominating the calculation. This works both ways: GNORM=20 will give a very crude optimization while GNORM=0.01 will give a very precise optimization. The default is GNORM=1.0.

When the highest precision is needed, such as in exacting geometry work, or when you want results whose precision cannot be improved, then use the combination keywords GNORM=0.0 and either RELSCF=0.01 or SCFCRT=1.D-NN; (NN should be in the range 5–15). By default, EigenFollowing is used in geometry optimization. One reason is that EigenFollowing is nearer to a gradient minimizer than it is to an energy minimizer. Because of this, if there is any difference between the gradient minimum and the energy minimum, it will give better reproducibility of the optimized geometry than the alternative BFGS method.

In practice, optimized geometries for "well behaved" systems can be obtained with GNORMs of less than 0.0001.

Increasing the SCF criterion (the default is SCFCRT=1.D-4) improves the precision of the gradients; however, it can lead to excessive run times, so take care. Also, there is an increased chance of not achieving an SCF when

the SCF criterion is excessively increased.

Superficially, requesting `GNORM=0` might seem excessively stringent, but as soon as the run starts, it will be cut back to 0.01. Even that might seem too stringent. The geometry optimization will continue to lower the energy, and hopefully the GNORM, but frequently it will not prove possible to lower the GNORM to 0.01. If, after 10 cycles, the energy does not drop then the job will be stopped. At this point you have the best geometry that MOPAC, in its current form, can give.

If a slightly less than highest precision is needed, such as for normal publication quality work, set the GNORM to the limit wanted. For example, for a flexible system, a GNORM of 0.1 to 0.5 will normally be good enough for all but the most demanding work.

If higher than the default, but still not very high precision is wanted, then use the keyword `PRECISE`. This will tighten up various criteria so that higher-than-routine precision will be given.

If high precision is used, so that the printed GNORM is 0.000, and the resulting geometry resubmitted for one SCF and gradients calculation, then normally a GNORM higher than 0.000 will result. This is *not* an error in MOPAC: the geometry printed is only precise to eight figures after the decimal point. Geometries may need to be specified to more than eight decimals in order to drive the GNORM to less than 0.000.

If you want to test MOPAC, or use it for teaching purposes, the GNORM lower limit of 0.01 can be overridden by specifying `LET`, in which case you can specify any limit for GNORM. However, if it is too low the job may finish due to an irreducible minimum in the heat of formation being encountered. If this happens, the "STATIONARY POINT" message will be printed.

Examples of highly optimized geometries can be found in the `port.dat` file. When this job is run, most gradients will be less than 0.001 kcal/mol/Å. A few will be larger. These exceptions fall into two classes: diatomics, for which a simple line-search is sufficient to locate the optimum geometry, in which case the GNORM criterion is *not* used; and non-variationally optimized systems, where the analytical C.I. derivatives are used. These derivatives are of lower precision than the variational derivatives, but are still much better than finite difference derivatives using full SCFs. Finally there is a full analytical derivative function [16] within MOPAC. These use STO-6G Gaussian wavefunctions because the derivatives of the overlap integral are easier to calculate in Gaussians than in STOs. Consequently, there will be a small difference in the calculated ΔH_f when analytical derivatives are used. If there is any doubt about the accuracy of the finite derivatives, try using the analytical derivatives. They are a bit slower than finite derivatives but are more precise (a rough estimate is 12 figures for finite difference, 14 for analytical).

Some calculations, mainly open shell RHF or closed shell RHF with C.I., have untracked errors which prevent very high precision. For these systems GNORM should be in the range 1.0 to 0.1.

5.11.3 Reasons for low precision

There are several reasons for obtaining low quality results, the most obvious of which is that for general work the default criteria will result in a difference in ΔH_f of less than 0.1 kcal/mol. This is only true for fairly rigid systems, e.g. formaldehyde and benzene. For systems with low barriers to rotation or flat potential surfaces, such as aniline or water dimer, quite large ΔH_f errors can result.

5.11.4 How large can a gradient be and still be acceptable?

A common source of confusion is the limit to which the GNORM should be reduced in order to obtain acceptable results. There is no easy answer. However, a few guidelines can be given.

First of all, setting the GNORM to an arbitrarily small number is not sensible. If `GNORM=0.000001` and `LET` are used, a geometry can be obtained which is precise to about 0.000001 Å. If `ANALYT` is also used, the results

obtained will be slightly different. Chemically, this change is meaningless, and no significance should be attached to such numbers. In addition, any minor change to the algorithm, such as porting it to a new machine, will give rise to small changes in the optimized geometry. Even the small changes involved in going from one version of MOPAC to another causes small changes in the optimized geometry of test molecules.

As a guide, a GNORM of 0.1 is sufficient for all heat-of-formation work, and a GNORM of 0.01 for most geometry work. If the system is large, you may need to settle for a GNORM of 1.0–0.5.

This whole topic was raised by Dr. Donald B. Boyd while he was at Lilly Research Laboratories, who provided unequivocal evidence for a failure of MOPAC and convinced me of the importance of increasing precision in certain circumstances.

5.11.5 Convergence tests in subroutine ITER

Self-consistency test

The SCF iterations are stopped when two tests are satisfied. These are (1) when the difference in electronic energy, in eV, between any two consecutive iterations drops below the adjustable parameter, SELCON, and the difference between any three consecutive iterations drops below ten times SELCON, and (2) the difference in density matrix elements on two successive iterations falls below a preset limit, which is a multiple of SELCON.

SELCON is set initially to 0.0001 kcal/mol; this can be made 100 times smaller by specifying PRECISE or FORCE. It can be over-ridden by explicitly defining the SCF criterion *via* SCFCRT=1.D-12, or by use of RELSCF=0.1.

SELCON is further modified by the value of the gradient norm, if known. If GNORM is large, then a more lax SCF criterion is acceptable, and SCFCRT can be relaxed up to 50 times its default value (using RELSCF=50). As the gradient norm drops, the SCF criterion returns to its default value.

The SCF test is performed using the energy calculated from the Fock matrix which arises from a density matrix, and not from the density matrix which arises from a Fock. In the limit, the two energies would be identical, but the first converges faster than the second, without loss of precision.

5.12 Torsion or Dihedral Angle Coherency

MOPAC calculations do not distinguish between enantiomers, consequently the sign of the dihedrals can be multiplied by -1 and the calculations will be unaffected. However, if chirality is important, a user should be aware of the sign convention used.

The dihedral angle convention used in MOPAC is that defined by Klyne and Prelog [72]. In this convention, four atoms, *AXYB*, with a dihedral angle of 90 degrees, will have atom B rotated by 90 degrees clockwise relative to A when X and Y are lined up in the direction of sight, X being nearer to the eye. In their words, "To distinguish between enantiomeric types the angle τ is considered as positive when it is measured clockwise from the front substituent A to the rear substituent B, and negative when it is measured anticlockwise." The alternative convention was used in programs which preceded MOPAC.

5.13 Gradients

By "gradients" we generally mean "the derivative of the energy with respect to coordinates". The two most commonly used gradients are with respect to Cartesian coordinates, in which case the units are kcal/mol/Ångstrom, or with respect to internal coordinates, in which case the units are either kcal/mol/Ångstrom or kcal/mol/radian, depending on whether the coordinate is a distance (in which case it would be kcal/mol/Ångstrom) or an angle or

dihedral (in which case it would be kcal/mol/radian). The particular gradient actually being used at any given point should be clear from the context. In all cases, the gradient can be regarded as the following derivative

$$g_i = \frac{d(\Delta H_f)}{dx_i}$$

In discussion “gradient” will be reserved for the derivative with respect to coordinates flagged for optimization (internal or Cartesian), and “derivative” will be used for both gradients and terms which are used to calculate gradients, such as Cartesian derivatives which are used to calculate internal coordinate gradients.

There are four very different ways to calculate gradients, although all four result in the same type of derivative. The four ways are:

Frozen density matrix finite difference derivatives

In these procedures, once an SCF has been achieved, the derivatives can be calculated using the density matrix from the SCF calculation. These methods can only be used with variationally optimized wavefunctions.

By default, the derivatives are worked out by calculating the energy of each pair of atoms, then re-calculating the energy after a small displacement has been made, and then calculating the derivative from the differences in the energies and the step. This is the default, and is the fastest. If this method is *not* wanted, specify ANALYT.

Analytical derivatives, using frozen density matrix approximation

Not as fast as the first method, but more accurate. Useful when finite difference derivatives are suspected to be of insufficient accuracy. When analytical derivatives are wanted, specify ANALYT. Analytical derivatives cannot be used with non-variational finite difference derivatives.

Non-variational finite difference derivatives

For non-variational wavefunctions (systems for which the electronic energy is modified after the SCF calculation is done, e.g. C.I. calculations), a sophisticated derivative routine in DERNVO calculates the effect on the derivative of the post-SCF energy terms. This method is used automatically in RHF C.I. calculations. If this method is *not* wanted, specify NOANCI.

Brute force gradients

These should be avoided whenever possible. To calculate the gradient, a small change is made in the desired coordinate, then a full SCF is done, and the gradient calculated from

$$g_i = \frac{\Delta H_f - \Delta H'_f}{x - x'}$$

These gradients are very slow, and are of poor accuracy, but sometimes they are the only way to obtain gradients. These derivatives cannot be used with variationally optimized wavefunctions, but can be used with non-variational wavefunctions by specifying NOANCI.

Note that ANALYT and NOANCI apply to two very different things: ANALYT applies to the derivatives using a frozen density matrix approximation, and uses true analytical methods. NOANCI prevents Liotard's C.I. derivative method being used. Of course NOANCI has no meaning for variationally optimized wavefunctions.

5.13.1 Frozen density matrix finite difference derivatives

The first step in calculating the gradients is to calculate the derivatives with respect to Cartesian coordinates. This is done in subroutine DCART.

DCART calculates the energy of each pair of atoms, then moves one atom a small distance (10^{-4}\AA) in each of the three Cartesian directions. The density matrix for the atom-pair is not changed during this calculation, but is set equal to the SCF density matrix. The derivative for each atom is then calculated from:

$$\left(\frac{dE}{dx}\right)_A = \sum_{B \neq A} \frac{E_{AB} - E'_{AB}}{\delta_x},$$

$$\left(\frac{dE}{dy}\right)_A = \sum_{B \neq A} \frac{E_{AB} - E'_{AB}}{\delta_y},$$

$$\left(\frac{dE}{dz}\right)_A = \sum_{B \neq A} \frac{E_{AB} - E'_{AB}}{\delta_z}.$$

where E'_{AB} is the energy of the pair of atoms after displacement in the appropriate direction. For a stationary point, these derivatives are zero.

To convert from Cartesian coordinate (c.c.) derivatives into gradients (i.c.), the sum

$$g_i = \sum_j \frac{dE}{d(\text{c.c.}j)} \frac{d(\text{c.c.}i)}{d(\text{i.c.}j)}$$

must be evaluated. Evaluation of $\frac{d(\text{c.c.}i)}{d(\text{i.c.}j)}$ is quite simple, and is done in routine JCARIN.

5.13.2 Hessian matrix in FORCE calculations

The Hessian matrix is the matrix of second derivatives of the energy with respect to geometry. The most important Hessian is that used in the FORCE calculation. Normal modes are expressed as Cartesian displacements, consequently the Hessian is based on Cartesian rather than internal coordinates.

Although first derivatives are relatively easy to calculate, second derivatives are not. The simplest, although not an elegant, way to calculate [73] second derivatives is to calculate first derivatives for a given geometry, then perturb the geometry, do an SCF calculation on the new geometry, and re-calculate the derivatives. The second derivatives can then be calculated from the difference of the two first derivatives divided by the step size. This method, which is used in the EigenFollowing routine, is called 'single-sided' derivatives.

The Hessian is quite sensitive to geometry, and should only be evaluated at stationary points. Because of this sensitivity, "double-sided" derivatives are used:

$$H_{i,j} = \frac{g_i^{+\delta_j} - g_i^{-\delta_j}}{2\delta}.$$

Note the asymmetry in the treatment of the Cartesian coordinates i and j . It can be shown that

$$\frac{g_j^{+\delta_i} - g_j^{-\delta_i}}{2\delta} = \frac{g_i^{+\delta_j} - g_i^{-\delta_j}}{2\delta}.$$

To help improve precision, the Hessian is calculated from

$$H_{i,j} = \frac{1}{2} \left(\frac{g_j^{+\delta_i} - g_j^{-\delta_i}}{2\delta} + \frac{g_i^{+\delta_j} - g_i^{-\delta_j}}{2\delta} \right).$$

5.14 Normal Coordinate Calculation

5.14.1 Calculation of Vibrational Frequencies

For a simple harmonic oscillator the period r is given by:

$$r = 2\pi\sqrt{\frac{\mu}{k}}$$

where k is the force constant. A molecule can absorb a photon that vibrates at the same frequency as one of its normal vibrational modes. That is, if a molecule, initially in its ground vibrational state, could be excited so that it vibrated at a given frequency, then that molecule could absorb a photon that vibrates at the same frequency. Although vibrational frequencies are usually expressed as kilohertz or megahertz, in chemistry vibrational frequencies are normally expressed in terms of the number of vibrations that would occur in the time that light travels one centimeter, i.e., $\bar{\nu} = 1/cr$. Using this equation for simple harmonic motion, the vibrational frequency can be written as:

$$\bar{\nu} = \frac{1}{2\pi c}\sqrt{\frac{k}{\mu}}$$

In order for $\bar{\nu}$ to be in cm^{-1} , c , the speed of light must be in $\text{cm}\cdot\text{sec}^{-1}$, k , the force constant in erg/cm^2 , and μ the reduced mass in grams.

For a molecule, the force constants are obtained by diagonalization of the mass-weighted Hessian matrix. Most of the work in calculating vibrational frequencies is spent in constructing the Hessian.

Calculation of the Hessian matrix

The elements of the Hessian are defined as:

$$H_{i,j} = \frac{\delta^2 E}{\delta x_i \delta x_j}$$

and are generated by use of finite displacements, that is, for each atomic coordinate x_i , the coordinate is first incremented by a small amount, the gradients calculated, then the coordinate is decremented and the gradients re-calculated. The second derivative is then obtained from the difference of the two derivatives and the step size:

$$H_{i,j} = \frac{(\frac{\delta E}{\delta x_i})_{+0.5\Delta x_j} - (\frac{\delta E}{\delta x_i})_{-0.5\Delta x_j}}{\Delta x_j}$$

This is done for all $3N$ Cartesian coordinates. Because the Hessian is symmetric, that is $H_{i,j} = H_{j,i}$, the random errors that occur in the gradient calculation can be reduced (by a factor of $\sqrt{2}$) by re-defining the Hessian as:

$$H_{i,j} = \frac{1}{2} \left(\frac{(\frac{\delta E}{\delta x_i})_{+0.5\Delta x_j} - (\frac{\delta E}{\delta x_i})_{-0.5\Delta x_j}}{\Delta x_j} + \frac{(\frac{\delta E}{\delta x_j})_{+0.5\Delta x_i} - (\frac{\delta E}{\delta x_j})_{-0.5\Delta x_i}}{\Delta x_i} \right)$$

A call to the energy - gradient function COMPFG will generate the gradients in $\text{kcal}/\text{mol}/\text{\AA}$ at a given geometry. These can then be converted into millidynes/ \AA (or 10^5 dynes/cm) as follows:

$$H_{i,j}(\text{millidynes}/\text{\AA}) = 10^5 \frac{(\text{Kcal to ergs})}{(\text{\AA to cm})^2(\text{Mole to molecule})} H_{i,j}(\text{kcal}/\text{mol}/\text{\AA}^2)$$

or

$$H_{i,j}(\text{millidynes}/\text{\AA}) = 10^5 \frac{4.184 * 10^3 * 10^7}{(10^{-8*2})(6.022 * 10^{23})} H_{i,j}(\text{kcal}/\text{mol}/\text{\AA}^2)$$

Diagonalization of this matrix yields the force constants of the system.

In order to calculate the vibrational frequencies, the Hessian matrix is first mass-weighted:

$$H_{i,j}^m = \frac{H_{i,j}}{\sqrt{M_i * M_j}}$$

Then the Hessian is converted from millidynes per Ångstrom to dynes per centimeter by multiplying by 10^5 .

Diagonalization of this matrix yields eigenvalues, ϵ , which represent the quantities $\sqrt{k/\mu}$, from which the vibrational frequencies can be calculated:

$$\bar{\nu}_i = \frac{1}{2\pi c} \sqrt{\epsilon_i}$$

5.14.2 Mechanism of the frame in FORCE calculation

The FORCE calculation uses Cartesian coordinates, and all 3N modes are calculated, where N is the number of atoms in the system. Clearly, there will be 5 or 6 “trivial” vibrations, which represent the three translations and two or three rotations. If the molecule is exactly at a stationary point, then these “vibrations” will have a force constant and frequency of precisely zero. If the force calculation was done correctly, and the molecule was not exactly at a stationary point, then the three translations should be exactly zero, but the rotations would be non-zero. The extent to which the rotations are non-zero is a measure of the error in the geometry.

If the distortions are non-zero, the trivial vibrations can interact with the low-lying genuine vibrations or rotations, and with the transition vibration if present.

To prevent this the analytic form of the rotations and vibrations is calculated, and arbitrary eigenvalues assigned; these are 500, 600, 700, 800, 900, and 1000 millidynes/Ångstrom for Tx, Ty, Tz, Rx, Ry and Rz (if present), respectively. The rotations are about the principal axes of inertia for the system, taking into account isotopic masses. The “force matrix” for these trivial vibrations is determined, and added on to the calculated force matrix. After diagonalization the arbitrary eigenvalues are subtracted off the trivial vibrations, and the resulting numbers are the “true” values. Interference with genuine vibrations is thus avoided.

5.14.3 Vibrational Analysis

Analyzing normal coordinates is very tedious. Users are normally familiar with the internal coordinates of the system they are studying, but not familiar with the Cartesian coordinates. To help characterize the normal coordinates, a very simple analysis is done automatically, and users are strongly encouraged to use this analysis first, and then to look at the normal coordinate eigenvectors.

In the analysis, each pair of bonded atoms is examined to see if there is a large relative motion between them. By bonded is meant within the van der Waals’ distance. If there is such a motion, the indices of the atoms, the relative distance in Ångstroms, and the percentage radial motion are printed. Radial plus tangential motion adds to 100%, but as there are two orthogonal tangential motions and only one radial, the radial component is printed.

Vibrations in the range $+50$ to -50 cm^{-1} cannot be described accurately in the vibrational analysis, due to numerical difficulties. However, the normal coordinates and frequencies are printed in the section above “DESCRIPTION OF VIBRATIONS”.

5.14.4 Reduced masses

A molecular vibration normally involves all the atoms moving simultaneously. This is clearly very different from the simple harmonic motion of a mass attached to a spring that is attached to an immovable object. Nevertheless, it is convenient to visualize a molecular vibration as consisting of a single mass, M , on the end of a spring of force

constant k . For such a system, the period of vibration, T , is given by:

$$T = 2\pi\sqrt{\frac{M}{k}}.$$

How, then, do we relate the complicated motion of a molecular vibration to the mass and spring model?

During a molecular vibration, each atom follows a simple harmonic motion. So the problem is, to what extent does each atom contribute to the mass, and to what extent does each atom contribute to the spring?

In order to answer this, first consider some simple systems. In the system H–X, where X has a very large mass, compared to that of the H, the effective mass is obviously that of H. In H₂, the effective mass is half that of a single H. Why is this so? In H–X, particle X is stationary, and particle H contributes 100% of the energy to the vibration. In H₂, each particle obviously contributes 50%, but now the center of mass is half way between the two particles. If the force constants are the same in H–X and in H–H, then the period of vibration of H–X will be $\sqrt{2}$ times that of H–H. This is the same period as for a system of two particles, each of which having a mass twice that of a H particle. For a system of two particles, A and B , having masses M_A and M_B , the vibrational wavefunction, ψ_v , is:

$$\psi_v = \sqrt{\frac{M_B}{M_A + M_B}}\psi_A - \sqrt{\frac{M_A}{M_A + M_B}}\psi_B.$$

This can be interpreted as particle A moves $(\sqrt{\frac{M_B}{M_A + M_B}})^2$ in the time particle B moves $(\sqrt{\frac{M_A}{M_A + M_B}})^2$. The center of mass, ρ , stays constant:

$$\rho = \sum_i M_i \delta x_i = M_A \frac{M_A}{M_A + M_B} - M_B \frac{M_A}{M_A + M_B} = 0.$$

The square of the coefficients of the wavefunction represent the contribution to the motion. The effective mass, μ , for this system is:

$$\mu = \frac{M_A \times M_B}{M_A + M_B}.$$

What fraction is due to A and what fraction is due to B ? From the wavefunction, the intensity of A is $\frac{M_B}{M_A + M_B}$, and the relative rate of motion is also $\frac{M_B}{M_A + M_B}$, so the contribution to the effective mass due to A is:

$$\left(\frac{M_B}{M_A + M_B}\right)M_A;$$

likewise, for B :

$$\left(\frac{M_A}{M_A + M_B}\right)M_B.$$

Consider two particles, A and B , of mass 1 and 4, respectively. The wavefunction for the vibration is:

$$\psi_v = \sqrt{\frac{4}{5}}\psi_A - \sqrt{\frac{1}{5}}\psi_B,$$

where A contributes

$$\frac{16}{25} \times 1 = 0.64$$

and B contributes

$$\frac{1}{25} \times 4 = 0.16$$

to the effective mass of $\frac{4}{5}$.

In other words, the contribution to the effective mass is equal to the intensity of the wavefunction on each atom, times the mass of the atom, times the intensity of the wavefunction. This is intuitively correct: the total vibration

is composed of contributions from each particle, and the amount each particle contributes is proportional to its intensity in the wavefunction. The mass of each particle is also proportional to its intensity in the wavefunction.

Extension to polyatomic molecules is now trivial. The effective mass is given by:

$$\mu = \sum_A \langle \psi_A | M_A \psi_A \rangle \times \langle \psi_A | \psi_A \rangle.$$

When written in this way, the quantum nature of the expression is obvious. However, for computational convenience, the effective mass is rewritten as:

$$\mu = \sum_A (q_A^2)^2 \times M_A$$

or

$$\mu = \sum_A \left(\sum_{i=1}^3 c_{A_i}^2 \right)^2 M_A.$$

This expression is suitable for most systems. However, it is not a well-defined quantity. Under certain circumstances involving degenerate vibrations, the quantity μ can become ill-defined. This phenomenon can be attributed to the fact that the reduced mass is not an observable.

5.14.5 Effective masses

Another way of regarding the effective mass of a mode can be derived from consideration of the simple harmonic oscillator:

$$E = \sqrt{\frac{k}{\mu}}.$$

Diagonalization of the mass-weighted Hessian yields the energies, and from the normal coordinates the force-constants can readily be derived. From these two quantities, the effective mass can readily be calculated:

$$\mu = \frac{k}{E^2}.$$

For a homonuclear diatomic, the effective mass calculated this way is equal to the mass of one atom.

5.14.6 Travel

To continue the idea of representing a normal mode as a simple harmonic oscillator, the distance the atoms move through can be represented as the distance the idealized mass moves through. This can be calculated knowing the energy of the mode and the force constant:

$$E = \frac{1}{2} k x^2.$$

Here k is the force-constant for the mode, and is given by

$$k = \langle \psi | \text{Hessian} | \psi \rangle;$$

E is the energy of the mode.

From this, the distance, x , which the system moves through, can be calculated from

$$x = \sqrt{\frac{2 \times 1.196266 \times 10^8 \times 1000 \times 10^8 \nu}{Nk}},$$

where 1.196266×10^8 is the conversion factor from cm^{-1} to ergs, 1000 converts from millidynes to dynes, 10^8 converts from cm to Å, and N converts from moles to molecules.

Note that x , which in the output is called TRAVEL, is in mass weighted space, not simple space. This quantity can also be calculated using the DRC, by depositing one quantum of energy into a vibrational mode. For a system at a stationary point, the relevant keywords would be IRC=1 DRC t=1m. For larger systems, the time may need to be increased. At least one coordinate must have an optimization flag set to 1. This is required in order to instruct the DRC to print the turning points.

5.15 A Note on Thermochemistry

by
Tsuneo Hirano
Department of Chemistry
Faculty of Science
Ochanomizu University
2-1-1 Otsuka
Bunkyo-ku
Tokyo 112
Japan

5.15.1 Basic Physical Constants

All basic physical constants should be taken from: "Quantities, Units and Symbols in Physical Chemistry," Blackwell Scientific Publications Ltd, Oxford OX2 0EL, UK, 1987 (IUPAC, based on CODATA of ICSU, 1986 [71]). pp. 81–82. A summary of these constants can be found in Table 5.43 (p. 208) .

Some derived quantities, which will be used in this section only are:

Moment of inertia: I 1 amu Ångstrom² = 1.660540×10^{-40} g cm².

Rotational constants: A , B , and C (e.g. $A = h/(8\pi^2I)$)

With I in amu Ångstroms² then: A (in MHz) = $5.053791 \times 10^5/I$

A (in cm⁻¹) = $5.053791 \times 10^5/cI = 16.85763/I$

5.15.2 Thermochemistry from *ab initio* MO methods

Ab initio MO methods provide total energies, E_{eq} , as the sum of electronic and nuclear-nuclear repulsion energies for molecules, isolated in vacuum, without vibration at 0 K.

$$E_{\text{eq}} = E_{\text{el}} + E_{\text{nuclear-nuclear}}$$

From the 0 K potential surface and using the harmonic oscillator approximation, we can calculate the vibrational frequencies, ν_i , of the normal modes of vibration. Using these, we can calculate vibrational, rotational and translational contributions to the thermodynamic quantities such as the partition function and heat capacity which arise from heating the system from 0 to T K.

Q : partition function, E : energy, S : entropy, and C : Heat capacity at constant pressure = C_p . In *ab initio* calculations, the heat capacity calculated is C_v . The relationship between C_p and C_v (in cal.degree⁻¹.mol⁻¹) is:

$$C_p = C_v + R = C_v + 1.987.$$

Table 5.44: Table of Symmetry Numbers

C ₁	C _I	C _S :	1	D ₂	D _{2d}	D _{2h} :	4	C _{∞v} :	1
C ₂	C _{2v}	C _{2h} :	2	D ₃	D _{3d}	D _{3h} :	6	D _{∞h} :	2
C ₃	C _{3v}	C _{3h} :	3	D ₄	D _{4d}	D _{4h} :	8	T, T _h T _d :	12
C ₄	C _{4v}	C _{4h} :	4	D ₅	D _{5d}	D _{5h} :	10	O, O _h :	24
C ₅	C _{5v}	C _{5h} :	5	D ₆	D _{6d}	D _{6h} :	12	I, I _h :	24
C ₆	C _{6v}	C _{6h} :	6	D ₇	D _{7d}	D _{7h} :	14	S ₄ :	2
C ₇	C _{7v}	C _{7h} :	7	D ₈	D _{8d}	D _{8h} :	16	S ₆ :	3
C ₈	C _{8v}	C _{8h} :	8					S ₈ :	4

Vibrational terms

$$Q_{\text{vib}} = \prod_i \frac{1}{[1 - \exp(-h\nu_i/kT)]} \quad (5.63)$$

E_{vib} , for a molecule at the temperature T as:

$$E_{\text{vib}} = \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\}$$

where h is Planck's constant, ν_i the i -th normal vibration frequency, and k the Boltzmann constant. For 1 mole of molecules, E_{vib} should be multiplied by the Avogadro number $N_a = R/k$. Thus:

$$E_{\text{vib}} = N_a \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\} \quad (5.64)$$

Note that the first term in equation (5.64) is the zero-point vibration energy. Hence, the second term in eq. (5.64) is the additional vibrational contribution due to the temperature increase from 0 K to T K. Namely,

$$\begin{aligned} E_{\text{vib}} &= E_{\text{zero}} + E_{\text{vib}}(T) \\ E_{\text{zero}} &= N_a \sum_i \frac{h\nu_i}{2} \\ E_{\text{vib}}(T) &= N_a \sum_i \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \end{aligned} \quad (5.65)$$

The value of E_{vib} from GAUSSIAN 82 and 86 includes E_{zero} as defined by Eqs. (5.64,5.65).

$$\begin{aligned} S_{\text{vib}} &= R \sum_i \left\{ \frac{(h\nu_i/kT) \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} - \ln[1 - \exp(-h\nu_i/kT)] \right\} \\ C_{\text{vib}} &= R \sum_i \left\{ \frac{(h\nu_i/kT)^2 \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]^2} \right\} \end{aligned}$$

At temperature $T > 0$ K, a molecule rotates about the x, y, and z-axes and translates in x, y, and z-directions. By assuming the equipartition of energy, energies for rotation and translation, E_{rot} and E_{tr} , are calculated.

Rotational terms

σ is symmetry number (Examples of symmetry numbers are shown in Table 5.44). I is moment of inertia. I_A , I_B , and I_C are moments of inertia about A, B, and C axes.

Linear molecule

$$\begin{aligned}
 Q_{\text{rot}} &= \frac{8\pi^2 I k T}{\sigma h^2} \\
 E_{\text{rot}} &= (2/2)RT \\
 S_{\text{rot}} &= R \ln \left[\frac{8\pi^2 I k T}{\sigma h^2} \right] + R \\
 &= R \ln I + R \ln T - R \ln \sigma - 4.349203
 \end{aligned}$$

where $-4.349203 = R \ln[8 \times 10^{-16} \pi^2 k / (N_a h^2)] + R$.

$$C_{\text{rot}} = (2/2)R$$

Non-linear molecule

$$\begin{aligned}
 Q_{\text{rot}} &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\frac{8\pi^2 k T}{h^2} \right]^{3/2} \sqrt{I_A I_B I_C} \\
 &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \right]^{1/2} \left(\frac{k T}{hc} \right)^{3/2} \\
 E_{\text{rot}} &= (3/2)RT \\
 S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \left(\frac{\pi}{\sqrt{\sigma}} \right) \left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \left(\frac{k T}{hc} \right)^3 \right\} + (3/2)R \\
 &= (R/2) \ln (I_A I_B I_C) + (3/2)R \ln T - R \ln \sigma - 5.3863921
 \end{aligned}$$

Here, -5.3863921 is calculated as:

$$R \ln \left\{ \frac{1}{h^3} \left(\frac{10^{-16}}{N_a} \right)^{3/2} \sqrt{(3 \times 2^9 \times \pi^7 \times k)} \right\} + (3/2)R.$$

$$C_{\text{rot}} = (3/2)R$$

Translational terms

M is Molecular weight.

$$\begin{aligned}
 Q_{\text{tra}} &= \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 \\
 E_{\text{tra}} &= (3/2)RT \\
 S_{\text{tra}} &= R \left\{ \frac{5}{2} + \frac{3}{2} \ln \left(\frac{2\pi k}{h^2} \right) + \ln k + \frac{3}{2} \ln \left(\frac{M}{N_a} \right) + \frac{5}{2} \ln T - \ln p \right\} \\
 &= (5/2)R \ln T + (3/2)R \ln M - R \ln p - 2.31482 \\
 C_{\text{tra}} &= (5/2)R
 \end{aligned}$$

or $H_{\text{tra}} = (5/2)RT$ due to the pV term (cf. $H = U + pV$). The internal energy U at T is:

$$U = E_{\text{eq}} + [E_{\text{vib}} + E_{\text{rot}} + E_{\text{tra}}]$$

or

$$U = E_{\text{eq}} + [(E_{\text{zero}} + E_{\text{vib}}(T)) + E_{\text{rot}} + E_{\text{tra}}] \quad (5.66)$$

Enthalpy H for one mole of gas is defined as

$$H = U + pV \quad (5.67)$$

Assumption of an ideal gas (i.e., $pV = RT$) leads to

$$H = U + pV = U + RT \quad (5.68)$$

Thus, Gibbs free energy G can be calculated as:

$$G = H - TS(T) \quad (5.69)$$

5.15.3 Thermochemistry in MOPAC

It should be noted that M.O. parameters for MINDO/3, MNDO, AM1, PM3, and MNDO- d are optimized so as to reproduce the experimental heat of formation (i.e., standard enthalpy of formation or the enthalpy change to form a mole of compound at 25°C from its elements in their standard state) as well as observed geometries (mostly at 25°C), and not to reproduce the E_{eq} and equilibrium geometry at 0 K.

In this sense, E_{SCF} (defined as Heat of formation, ΔH_f), force constants, normal vibration frequencies, etc. are all related to the values at 25°C, not to 0 K. Therefore, the E_{zero} calculated in FORCE is not the true E_{zero} . Its use as E_{zero} should be made at your own risk, bearing in mind the situation discussed above.

Since E_{SCF} is standard enthalpy of formation (at 25°C):

$$E_{SCF} = E_{\text{eq}} + E_{\text{zero}} + E_{\text{vib}}(298.15) + E_{\text{rot}} + E_{\text{tra}} + pV + \sum [-E_{\text{elec}}(\text{atom}) + \Delta H_f(\text{atom})]. \quad (5.70)$$

To avoid the complication arising from the definition of E_{SCF} , within the thermodynamics calculation the Standard Enthalpy of Formation, ΔH , is calculated by

$$\Delta H = E_{SCF} + (H_T - H_{298}). \quad (5.71)$$

Here, E_{SCF} is the heat of formation (at 25°C) given in the output list, and H_T and H_{298} are the enthalpy contributions for the increase of the temperature from 0 K to T and 298.15, respectively. In other words, the enthalpy of formation is corrected for the difference in temperature from 298.15 K to T .

There is a problem in that H_T is the heat of formation at T relative to the heat of formation of the elements in their standard state at 298K. This involves mixing standard and not standard terms. There is no easy way to get the correct value for H_T , but for rough work H_T is useful. For more correct work, calculate ΔH for the elements in their standard state at T , and use these ΔH 's to get the ΔH for the compound you're working with (or use tables from the literature).

The method of calculation for T and H_{298} will be given below.

In MOPAC, the variables defined below are used:

$$C_1 = \frac{hc}{kT}.$$

The wavenumber, ω_i , in cm^{-1} :

$$\nu_i = \omega_i c$$

$$E_{\omega_i} = \exp(-h\nu_i/kT) = \exp(-\omega_i hc/kT) = \exp(-\omega_i C_1)$$

The rotational constants A , B , and C in cm^{-1} :

$$A = \frac{h}{(8\pi^2 c I_A)}$$

Energy and Enthalpy in cal/mol, and Entropy in cal/mol/K. Thus, eqs. (5.63–5.69) can be written as follows:

Vibration

$$\begin{aligned}
Q_{\text{vib}} &= \prod_i \frac{1}{(1 - E_{\omega_i})} \\
E_0 &= \frac{0.5 N_a h c}{4.184 \times 10^7} \sum_i \omega_i \\
&= 1.429572 \sum_i \omega_i \\
E_{\text{vib}}(T) &= N_a h c \sum_i \frac{\omega_i E_{\omega_i}}{1 - E_{\omega_i}} = (R/k) h c \sum_i \frac{W_i E_{\omega_i}}{1 - E_{\omega_i}} \\
S_{\text{vib}} &= R(hc/kT) \sum_i \left\{ \frac{\omega_i E_{\omega_i}}{(1 - E_{\omega_i})} \right\} - R \sum_i \ln(1 - E_{\omega_i}) \\
&= RC_1 \sum_i \left\{ \frac{\omega_i E_{\omega_i}}{(1 - E_{\omega_i})} \right\} - R \sum_i \ln(1 - E_{\omega_i}) \\
C_{\text{vib}} &= R(hc/kT)^2 \sum_i \left\{ \frac{\omega_i^2 E_{\omega_i}}{(1 - E_{\omega_i})^2} \right\} \\
&= RC_1^2 \sum_i \left\{ \frac{\omega_i^2 E_{\omega_i}}{(1 - E_{\omega_i})^2} \right\}
\end{aligned}$$

Rotation**Linear molecule**

$$\begin{aligned}
Q_{\text{rot}} &= (1/\sigma)(1/\text{\AA})(kT/hc) = \frac{1}{\sigma AC_1} \\
E_{\text{rot}} &= (2/2)RT \\
S_{\text{rot}} &= R \ln \left(\frac{kT}{\sigma hc \text{\AA}} \right) + R = R \ln \left(\frac{1}{\sigma \text{\AA} C_1} \right) + R = R \ln \left(\frac{kT}{\sigma hc \text{\AA}} \right) + R \\
C_{\text{rot}} &= (2/2)R
\end{aligned}$$

Non-linear molecule

$$\begin{aligned}
Q_{\text{rot}} &= \frac{1}{\sigma} \left[\frac{\pi}{(ABC C_1^3)} \right]^{1/2} \\
E_{\text{rot}} &= (3/2)RT \\
S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \frac{\pi}{\sigma^2 ABC} \left(\frac{kT}{hc} \right)^3 \right\} + (3/2)R \\
&= 0.5R3 \ln(kT/hc) - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3 \\
&= 0.5R - 3 \ln C_1 - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3 \\
C_{\text{rot}} &= (3/2)R
\end{aligned}$$

Translation

$$Q_{\text{tra}} = \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 = \left(\frac{\sqrt{1.660540 \times 10^{-24} \times 2\pi M k T}}{h} \right)^3$$

$$\begin{aligned}
 E_{\text{tra}} &= (3/2)RT \\
 H_{\text{tra}} &= (3/2)RT + pV = (5/2)RT \text{ cf. } pV = RT \\
 S_{\text{tra}} &= (R/2)[5 \ln T + 3 \ln M] - 2.31482 \text{ cf. } p = 1\text{atm} \\
 &= 0.993608[5 \ln T + 3 \ln M] - 2.31482
 \end{aligned}$$

In MOPAC:

$$H_{\text{vib}} = E_{\text{vib}}(T)$$

(Note: E_{zero} is *not* included in H_{vib} ; ω_i is not derived from force-constants at 0 K) and for T :

$$H_T = [H_{\text{vib}} + H_{\text{rot}} + H_{\text{tra}}]$$

while for $T = 298.15$ K:

$$H_{298} = [H_{\text{vib}} + H_{\text{rot}} + H_{\text{tra}}]$$

Note that H_T (and H_{298}) is equivalent to:

$$(E_{\text{vib}} - E_{\text{zero}}) + E_{\text{rot}} + (E_{\text{tra}} + pV)$$

except that the normal frequencies are those obtained from force constants at 25°C, or at least not at 0 K.

Thus, Standard Enthalpy of Formation, ΔH , can be calculated according to Eqs. (5.66,5.67) and (5.70), as shown in Eq. (5.71);

$$\Delta H = E_{SCF} + (H_T - H_{298})$$

Note that E_{zero} is already counted in E_{SCF} , see Eq. (5.70).

By using Eq. (5.68), Standard Internal Energy of Formation, ΔU , can be calculated as:

$$\Delta U = \Delta H - R(T - 298.15)$$

Standard Gibbs Free-Energy of Formation, ΔG , can be calculated by taking the difference from that for the isomer or that at different temperature:

$$\Delta G = [\Delta H - TS] \text{ (for the state under consideration)} - [\Delta H - TS] \text{ (for reference state)}$$

Taking the difference is necessary to cancel the unknown values of standard entropy of formation for the constituent elements.

5.16 Force Constants

Internal coordinate force constants, f_{ic} , can be derived from the Cartesian coordinates and the Cartesian force constant matrix by use of:

$$f_{ic}(l) = \sum_j \sum_k \frac{d(\text{c.c.}j)}{d(\text{i.c.}l)} \frac{d^2 E}{d(\text{c.c.}j)d(\text{c.c.}k)} \frac{d(\text{c.c.}k)}{d(\text{i.c.}l)}$$

As with the gradients, the calculation of $\frac{d(\text{c.c.}j)}{d(\text{i.c.}l)}$ is quite simple, and is done in routine JCARIN.

During the testing of this function, a minor fault in the conventional force calculation was revealed. To reduce any error introduced by finite arithmetic, the Hessian matrix is symmetrized before the vibrational frequencies and normal coordinates are calculated. This is done by as described on p. 201.

```

CHARGE=-1
SN2 reaction, Cl(-) + CH3F = CH3Cl + F(-)
C
F  1.4  1
H  1.1  1  109.5  1    0    0    1  2
H  1.1  1  109.5  1  120.0  0    1  2  3
H  1.1  1  109.5  1  120.0  0    1  2  3
Cl 20.0 -1  127.3  1  180.0  0    1  2  3
0  0.00  0    0.0  0    0.0  0    0  0  0
10.0 5.0 4.0 3.0 2.9 2.8 2.7 2.6 2.5
2.4 2.3 2.2 2.1 2.0 1.9 1.8 1.7 1.6

```

Figure 5.17: Example of an S_N2 reaction path calculation

```

step=5 points=13 SYMMETRY
Ethane, Barrier to Rotation
C
C  1.5  1    0  0    0  0    1  0  0
H  1.0  1  111  1    0  0    2  1  0
H  1.0  0  111  0  120  0    2  1  3
H  1.0  0  111  0 -120  0    2  1  3
H  1.0  0  111  0   60 -1    1  2  3
H  1.0  0  111  0  180  0    1  2  3
H  1.0  0  111  0  -60  0    1  2  3
0  0.0  0    0  0    0  0    0  0  0
3  1  4  5  6  7  8
3  2  4  5  6  7  8
6  7  7
6  11  8

```

Figure 5.18: Example of a rotation barrier calculation

In addition to the requirement that the symmetry of the Hessian should be the same as that of the nuclear coordinates, a second requirement is that the diagonal elements of the Hessian should be equal to the negative of the sum of the off-diagonal elements, that is, that:

$$F_{ii} = - \sum_{j \neq i} F_{ij}.$$

During the testing of the internal force constants, very small variations in the force constants were found where no variation was expected. This was traced back to a failure of the above expression. To correct this, the diagonal terms of the force constant matrix were modified. This resulted in a perfect equivalence of equivalent force constants. An incidental benefit would be that the associated error in the calculated frequencies would be eliminated.

5.17 Reaction paths

MOPAC has the capability to model the effects of changing an internal coordinate. In the data-set, the relevant internal coordinate is flagged with a '-1' rather than a '1' or '0'. Two options then exist to allow the values of the changing coordinate to be defined.

```
step=0.05 points=20
Trans-polyparaphenylene benzobisthiazole
Stretching the polymer
C 0.0 0 0 0 0 0 0 0 0
N 1.3 1 0 0 0 0 1 0 0
S 1.7 1 115 1 0 0 1 2 0
C 1.6 1 92 1 0 1 3 1 2
C 1.4 1 109 1 -0 1 2 1 3
C 1.4 1 124 1 -180 1 5 2 1
C 1.4 1 116 1 180 1 6 5 2
C 1.4 1 121 1 0 1 7 6 5
C 1.4 1 129 1 180 1 4 3 1
S 1.6 1 129 1 180 1 7 6 5
C 1.7 1 92 1 180 1 10 7 6
N 1.4 1 113 1 -180 1 8 7 6
C 1.4 1 121 1 -180 1 11 10 7
C 1.4 1 120 1 -90 1 13 11 10
C 1.4 1 120 1 180 1 14 13 11
C 1.4 1 120 1 0 1 15 14 13
C 1.4 1 118 1 -0 1 16 15 14
C 1.4 1 120 1 0 1 17 16 15
H 1.0 1 121 1 -0 1 6 5 2
H 1.0 1 121 1 0 1 9 4 3
H 1.0 1 120 1 -0 1 14 13 11
H 1.0 1 119 1 -180 1 15 14 13
H 1.0 1 120 1 -180 1 17 16 15
H 1.0 1 119 1 180 1 18 17 16
xx 1.4 1 120 1 180 1 16 15 14
Tv 12.6 -1 0 0 0 0 1 25 24
```

Figure 5.19: Data set to stretch a polymer

```

SYMMETRY STEP1=0.01 STEP2=1
Water, potential energy surface for

H
O  0.92 -1    0  0    0  0    1
H  0.92  0   104 -1    0  0    2  1

2 1 3

```

Figure 5.20: Example of a GRID Calculation

First, the various values of the coordinate can be supplied after the geometry and any symmetry data have been entered. An example for the S_{N2} reaction $Cl^- + CH_3F \rightarrow CH_3Cl + F^-$ is given in Figure 5.17.

Second, if the step-size is a constant, then the step-size and number of steps can be defined on the keyword line. An example of such a “reaction” would be the rotation of a methyl group in, e.g., ethane, Figure 5.18. Here, symmetry is used to maintain D_3 symmetry as the rotation takes place. Note that SYMMETRY can be used to relate coordinates to the reaction coordinate. The path calculations work by optimizing the geometry while the reaction coordinate is fixed at the starting value. Once the geometry is optimized, the reaction coordinate is changed, and the geometry re-optimized. This is done for all points on the reaction path.

Reaction paths can be used for calculating mechanical properties. For example, to calculate Hook’s force constant for stretching polyethylene, the translation vector could be steadily increased, Figure 5.19.

5.18 Grid Calculation

The GRID calculation is the two-dimensional analog of the PATH calculation. In a PATH calculation, one coordinate is flagged with a ‘-1’. In a GRID calculation, two coordinates are flagged by ‘-1’s. An example of a GRID calculation is shown in Figure 5.20. Note that the keywords $STEP1=n.nn$ and $STEP2=m.mm$ are essential.

In this example, the potential energy surface for water is generated. For one axis of the 2-D plot, the O–H bond length is varied from 0.92Å to 1.02Å, in 11 steps of 0.01Å, and in the other axis, the H–O–H angle is varied from 104 to 114° in 11 steps of 1.0 degree. Because of the use of symmetry, there are no variables to be optimized. If symmetry were not used, then the second O–H bond length could either be optimized, by setting its flag to 1, or held constant, by setting its flag to 0.

Other keywords which can be used with the GRID option are: MAX: to use 23 points in each direction, rather than the default 11; POINT1=*n* and POINT2=*m*: to use *n* and *m* points in directions 1 and 2; and keywords to specify how the geometry should be optimized.

5.19 Dynamic and Intrinsic Reaction Coordinates

The Intrinsic Reaction Coordinate method pioneered and developed by Mark Gordon, North Dakota State University, has been incorporated into MOPAC in a modified form. As this facility is quite complicated all the keywords associated with the IRC have been grouped together in this section (these can be seen later on in this section).

The Dynamic Reaction Coordinate is the path followed by all the atoms in a system assuming conservation of energy; i.e., as the potential energy changes the kinetic energy of the system changes in exactly the opposite way so that the total energy (kinetic plus potential) is a constant. It is equivalent to the molecular mechanics molecular dynamics calculation, except that bond-breaking and bond-making are supported, as are all the electronic phenomena of the semiempirical methods.

If started at a ground state geometry, no significant motion should be seen. Similarly, starting at a transition state geometry should not produce any motion—after all it is a stationary point and during the lifetime of a calculation it is unlikely to accumulate enough momentum to travel far from the starting position.

In order to calculate the DRC path from a transition state, either an initial deflection is necessary or some initial momentum must be supplied.

Because of the time-dependent nature of the DRC the time elapsed since the start of the reaction is meaningful, and is printed.

Description

The course of a molecular vibration can be followed by calculating the potential and kinetic energy at various times. Two extreme conditions can be identified: (a) gas phase, in which the total energy is a constant through time, there being no damping of the kinetic energy allowed, and (b) liquid phase, in which kinetic energy is always set to zero, the motion of the atoms being infinitely damped.

All possible degrees of damping are allowed. In addition, the facility exists to dump energy into the system, appearing as kinetic energy. As kinetic energy is a function of velocity, a vector quantity, the energy appears as energy of motion in the direction in which the molecule would naturally move. If the system is a transition state, then the excess kinetic energy is added after the intrinsic kinetic energy has built up to at least 0.2 kcal/mol.

For ground-state systems, the excess energy sometimes may not be added; if the intrinsic kinetic energy never rises above 0.2 kcal/mol then the excess energy will not be added.

Equations used

Force acting on any atom:

$$g(i) + g'(i)t + g''(i)t^2 = \frac{dE}{dx(i)} + \frac{d^2E}{dx(i)^2} + \frac{d^3E}{dx(i)^3}$$

Acceleration due to force acting on each atom:

$$a(i) = \frac{1}{M(i)}(g(i) + g'(i)t + g''(i)t^2)$$

New velocity:

$$V(o) + \frac{1}{M(i)} (\Delta t g(i) + (1/2)\Delta t^2 g'(i) + (1/3)\Delta t^3 g''(i))$$

or:

$$V(i) = V(o) + V'(i)t + V''(i)t^2 + V'''(i)t^3$$

That is, the change in velocity is equal to the integral over the time interval of the acceleration.

New position of atoms:

$$X(i) = X(o) + V(o)t + (1/2)V't^2 + (1/3)V''t^3 + (1/4)V'''t^4$$

That is, the change in position is equal to the integral over the time interval of the velocity.

The velocity vector is accurate to the extent that it takes into account the previous velocity, the current acceleration, the predicted acceleration, and the change in predicted acceleration over the time interval. Very little error is introduced due to higher order contributions to the velocity; those that do occur are absorbed in a re-normalization of the magnitude of the velocity vector after each time interval.

The magnitude of Δt , the time interval, is determined mainly by the factor needed to re-normalize the velocity vector. If it is significantly different from unity, Δt will be reduced; if it is very close to unity, Δt will be increased.

Even with all this, errors creep in and a system, started at the transition state, is unlikely to return precisely to the transition state unless an excess kinetic energy is supplied, for example 0.2 kcal/mol.

The calculation is carried out in Cartesian coordinates, and converted into internal coordinates for display. All Cartesian coordinates must be allowed to vary, in order to conserve angular and translational momentum.

IRC

The Intrinsic Reaction Coordinate is the path followed by all the atoms in a system and assumes that all kinetic energy is completely lost at every point; i.e., as the potential energy changes the kinetic energy generated is annihilated so that the total energy (kinetic plus potential) is always equal to the potential energy only.

The IRC is intended for use in calculations in which the starting geometry is that of the transition state. A normal coordinate is chosen, usually the reaction coordinate, and the system is displaced in either the positive or negative direction along this coordinate. The internal modes are obtained by calculating the mass-weighted Hessian matrix in a force calculation and translating the resulting Cartesian normal mode eigenvectors to conserve momentum. That is, the initial Cartesian coordinates are displaced by a small amount proportional to the eigenvector coefficients plus a translational constant; the constant is required to ensure that the total translational momentum of the system is conserved as zero. At the present time there may be small residual rotational components which are not annihilated; these are considered unimportant, and will not materially affect the calculation.

General description of the DRC and IRC

As the IRC usually requires a normal coordinate, a force constant calculation normally has to be done first. If IRC is specified on its own, a normal coordinate is not used and the IRC calculation is performed on the supplied geometry.

A recommended sequence of operations to start an IRC calculation is as follows:

1. Calculate the transition state geometry. If the transition state is not first optimized, then the IRC calculation may give very misleading results. For example, if NH_3 inversion is defined as the planar system but without the N–H bond length being optimized, the first normal coordinate might be for N–H stretch rather than inversion. In that case the IRC will relax the geometry to the optimized planar structure.
2. Do a normal FORCE calculation, specifying ISOTOPE in order to save the FORCE matrices. (Note: Do not attempt to run the IRC at this point directly unless you have confidence that the FORCE calculation will work as expected. If the IRC calculation is run directly, specify ISOTOPE anyway: that will save the FORCE matrix and if the calculation has to be re-done then RESTART will work correctly.)
3. Using $\text{IRC}=n$ and RESTART, run the IRC calculation. If RESTART is specified with $\text{IRC}=n$ then the restart is assumed to be from the FORCE calculation. If, in an IRC calculation, RESTART is specified, and $\text{IRC}=n$ is *not* present, then the restart is assumed to be from an earlier IRC calculation that was shut down before going to completion.

A DRC calculation is simpler, in that a force calculation is not a prerequisite; however, most calculations of interest normally involve use of an internal coordinate. For this reason $\text{IRC}=n$ can be combined with DRC to give a calculation in which the initial motion (0.3kcal worth of kinetic energy) is supplied by the IRC, and all subsequent motion obeys conservation of energy. The DRC motion can be modified in three ways:

1. It is possible to calculate the reaction path followed by a system in which the generated kinetic energy decays with a finite half-life. This can be defined by $\text{DRC}=n.nnn$, where $n.nnn$ is the half-life in femtoseconds. If $n.nnn$ is 0.0 this corresponds to infinite damping simulating the IRC. A limitation of the program is that time only has meaning when DRC is specified without a half-life.

2. Excess kinetic energy can be added to the calculation by use of `KINETIC=n.nn`. After the kinetic energy has built up to 0.2 kcal/mol or if `IRC=n` is used then *n.nn* kcal/mol of kinetic energy is added to the system. The excess kinetic energy appears as a velocity vector in the same direction as the initial motion.
3. The RESTART file `<filename>.res` can be edited to allow the user to modify the velocity vector or starting geometry. This file is formatted.

Frequently, the DRC leads to a periodic, repeating orbit. One special type—the orbit in which the direction of motion is reversed so that the system retraces its own path—is sensed for and if detected the calculation is stopped after exactly one cycle. If the calculation is to be continued, the keyword `GEO-OK` will allow this check to be by-passed.

Sometimes the system will enter a stable state in which the geometry is always changing, but nothing new is occurring. One example would be a system which decomposed into fragments, and the fragments were moving apart. If all forces acting on the atoms become small, then the calculation will be stopped. If the calculation should be continued, then specify `GNORM=0 LET`.

Due to the potentially very large output files that the DRC can generate, extra keywords are provided to allow selected points to be printed. Two types of control are provided: one controls which points to print, the other controls what is printed.

By default, every point calculated is printed. Often, this is not desirable, and three keywords are provided to allow printing to be done whenever the system changes by a preset amount. These keywords are:

KeyWord	Default	User Specification
X-PRIO	0.05 Ångstroms	X-PRIORITY= <i>n.nn</i>
T-PRIO	0.10 Femtoseconds	T-PRIORITY= <i>n.nn</i>
H-PRIO	0.10 kcal/mol	H-PRIORITY= <i>n.nn</i>

By default, only the energies involved are printed (one line per point). To allow the geometry to be printed, , see p. 52 for more detail. Using `LARGE` a wide range of control is provided over what is printed.

Option to allow only extrema to be output

In the geometry specification, if an internal coordinate is marked for optimization then when that internal coordinate passes through an extremum a message will be printed and the geometry output.

Difficulties can arise from the way internal coordinates are processed. The internal coordinates are generated from the Cartesian coordinates, so an internal coordinate supplied may have an entirely different meaning on output. In particular the connectivity may have changed. For obvious reasons dummy atoms should not be used in the supplied geometry specification. If there is any doubt about the internal coordinates or if the starting geometry contains dummy atoms then run a 1SCF calculation specifying `INT`. This will produce an ARC file with the “ideal” numbering—the internal numbering system used by MOPAC. Use this ARC file to construct a data file suitable for the DRC or IRC.

Notes:

1. Any coordinates marked for optimization will result in only extrema being printed.
2. If extrema are being printed then kinetic energy extrema will also be printed.

Keywords for use with the IRC and DRC

1. Setting up the transition state: `NLLSQ`, `SIGMA`, or `TS`.

```

KINETIC=6 RESTART  IRC=-1 DRC T=600
WATER

H  0.000000 0  0.000000 0  0.000000 0  0 0 0
O  0.911574 0  0.000000 0  0.000000 0  1 0 0
H  0.911574 0 180.000000 0  0.000000 0  2 1 0
O  0.000000 0  0.000000 0  0.000000 0  0 0 0

```

Figure 5.21: Example of DRC calculation

```

RESTART  IRC  T=600
WATER

H  0.000000 0  0.000000 0  0.000000 0  0 0 0
O  0.911574 0  0.000000 0  0.000000 0  1 0 0
H  0.911574 0 180.000000 0  0.000000 0  2 1 0
O  0.000000 0  0.000000 0  0.000000 0  0 0 0

```

Figure 5.22: Example of IRC calculation

2. Constructing the FORCE matrix: FORCE or IRC=*n*, ISOTOPE, LET.
3. Starting an IRC: RESTART and IRC=*n*, X-PRIO, H-PRIO.
4. Starting a DRC: DRC or DRC=*n.nn*, KINETIC=*n.nn*, T-PRIO, etc..
5. Starting a DRC from a transition state: (DRC or DRC=*n*) and IRC=*n*, KINETIC=*n*.
6. Restarting an IRC: RESTART and IRC.
7. Restarting a DRC: RESTART and (DRC or DRC=*n.nn*).
8. Restarting a DRC starting from a transition state: RESTART and (DRC or DRC=*n.nn*).

Other keywords, such as T=*nnn* or GEO-OK can be used any time.

Examples of DRC/IRC data

Use of the IRC/DRC facility is quite complicated. In the following examples various ‘reasonable’ options are illustrated for a calculation on water. It is assumed that an optimized transition-state geometry is available.

Example 1: Figure 5.21 illustrates a Dynamic Reaction Coordinate calculation, starting at the transition state for water inverting, the initial motion being opposite to the transition normal mode, with 6kcal of excess kinetic energy added in. Every point calculated is to be printed (Note all coordinates are marked with a zero, and T-PRIO, H-PRIO and X-PRIO are all absent). The results of an earlier calculation using the same keywords is assumed to exist. The earlier calculation would have constructed the force matrix. While the total cpu time is specified, it is in fact redundant in that the calculation will run to completion in less than 600 seconds.

Example 2: Figure 5.22 shows an Intrinsic Reaction Coordinate calculation. Here the restart is from a previous IRC calculation which was stopped before the minimum was reached. Recall that RESTART with IRC=*n* implies a restart from the FORCE calculation. Since this is a restart from within an IRC calculation the keyword IRC=*n* has been replaced by IRC. IRC on its own (without the “=*n*”) implies an IRC calculation from the starting position—here the RESTART position—without initial displacement.

Output format for IRC and DRC

The IRC and DRC can produce several different forms of output. Because of the large size of these outputs, users are recommended to use search functions to extract information. To facilitate this, specific lines have specific characters. Thus, a search for the “%” symbol will summarize the energy profile while a search for “AA” will yield the coordinates of atom 1, whenever it is printed. The main flags to use in searches are:

% Energies for all points calculated, excluding extrema

%M Energies for all turning points

%MAX Energies for all maxima

%MIN Energies for all minima

% Energies for all points calculated

AA* Internal coordinates for atom 1 for every point

AE* Internal coordinates for atom 5 for every point

123AB* Internal coordinates for atom 2 for point 123

As the keywords for the IRC/DRC are interdependent, the following list of keywords illustrates various options.

DRC The Dynamic Reaction Coordinate is calculated. Energy is conserved, and no initial impetus.

DRC=0.5 In the DRC kinetic energy is lost with a half-life of 0.5 femtoseconds.

DRC=1.0 Energy is put into a DRC with an half-life of -1.0 femtoseconds, i.e., the system gains energy.

IRC The Intrinsic Reaction Coordinate is calculated. No initial impetus is given. Energy not conserved.

IRC=4 The IRC is run starting with an impetus in the negative of the 4th normal mode direction. The impetus is one quantum of vibrational energy.

IRC1 KINETIC=1 The first normal mode is used in an IRC, with the initial impetus being 1.0 kcal/mol.

DRC KINETIC=5 In a DRC, after the velocity is defined, 5 kcal of kinetic energy is added in the direction of the initial velocity.

IRC=1 DRC KINETIC=4 After starting with a 4 kcal impetus in the direction of the first normal mode, energy is conserved.

DRC VELOCITY KINETIC=10 Follow a DRC trajectory which starts with an initial velocity read in, normalized to a kinetic energy of 10 kcal/mol.

Instead of every point being printed, the option exists to print specific points determined by the keywords T-PRIORITY, X-PRIORITY and H-PRIORITY. If any one of these words is specified, then the calculated points are used to define quadratics in time for all variables normally printed. In addition, if the flag for the first atom is set to “T” then all kinetic energy turning points are printed. If the flag for any other internal coordinate is set to “T” then, when that coordinate passes through an extremum, that point will be printed. As with the PRIORITY’s, the point will be calculated via a quadratic to minimize non-linear errors.

N.B.: Quadratics are unstable in the regions of inflection points; in these circumstances linear interpolation will be used. A result of this is that points printed in the region of an inflection may not correspond exactly to those requested. This is not an error and should not affect the quality of the results.

Table 5.45: Stretching Curve for Nitrogen Molecule

N-N DIST (Ångstroms)	ΔH_f (kcal/mol)	GRADIENT (kcal/mol/Ångstrom)
1.11800	8.69441	60.84599
1.11700	8.63563	56.70706
1.11600	8.58100	52.54555
1.11500	8.53054	48.36138
1.11400	8.48428	44.15447
1.11300	8.44224	39.92475
1.11200	8.40444	35.67214
1.11100	8.37091	31.39656
1.11000	8.34166	27.09794
1.10900	8.31672	22.77620
1.10800	8.29611	18.43125
1.10700	8.27986	14.06303
1.10600	8.26799	9.67146
1.10500	8.26053	5.25645
1.10400	8.25749	0.81794
1.10300	8.25890	-3.64427
1.10200	8.26479	-8.12993
1.10100	8.27517	-12.63945
1.10000	8.29007	-17.17278
1.09900	8.30952	-21.73002
1.09800	8.33354	-26.31123
1.09700	8.36215	-30.91650
1.09600	8.39538	-35.54591
1.09500	8.43325	-40.19953
1.09400	8.47579	-44.87745
1.09300	8.52301	-49.57974
1.09200	8.57496	-54.30648
1.09100	8.63164	-59.05775
1.09000	8.69308	-63.83363

Test of DRC—verification of trajectory path

Introduction: Unlike a single-geometry calculation or even a geometry optimization, verification of a DRC trajectory is not a simple task. In this section a rigorous proof of the DRC trajectory is presented; it can be used both as a test of the DRC algorithm and as a teaching exercise. Users of the DRC are asked to follow through this proof in order to convince themselves that the DRC works as it should.

The nitrogen molecule

For the nitrogen molecule (using MNDO) the equilibrium distance is 1.103816 Å, the heat of formation is 8.25741 kcal/mol and the vibrational frequency is 2738.8 cm⁻¹. For small displacements, the energy curve versus distance is parabolic and the gradient curve is approximately linear, as is shown in Table 5.45. A nitrogen molecule is thus a good approximation to a harmonic oscillator.

Period of vibration

The period of vibration (time taken for the oscillator to undertake one complete vibration, returning to its original position and velocity) can be calculated in three ways. Most direct is the calculation from the energy curve; using the gradient constitutes a faster, albeit less direct, method, while calculating it from the vibrational frequency is very fast but assumes that the vibrational spectrum has already been calculated.

1. From the energy curve. For a simple harmonic oscillator the period r is given by:

$$r = 2\pi\sqrt{\frac{\mu}{k}}$$

where k is the force constant. The reduced mass, μ , (in amu) of a nitrogen molecule is $14.0067/2 = 7.00335$, and the force-constant, k , can be calculated from:

$$E - c = (1/2)k(R - R_o)^2.$$

Given $R_o = 1.1038$, $R = 1.092$, $c = 8.25741$ and $E = 8.57496$ kcal/mol then:

$$k = 2 * 0.31755 / (0.0118)^2 \text{ (per mole)}$$

$$k = 4561.2 \text{ kcal/mol/\AA}^2 \text{ (per mole)}$$

$$k = 1.9084 * 10^{30} \text{ ergs/cm}^2 \text{ (per mole)}$$

$$k = 31.69 * 10^5 \text{ dynes/cm (per molecule)}$$

(Experimentally, for N_2 , $k = 23 * 10^5$ dynes/cm)

Therefore:

$$r = 2 \times 3.14159 \times \sqrt{\frac{7.0035}{1.9084 \times 10^{30}}} \text{ seconds} = 12.037 \times 10^{-15} \text{ s} = 12.037 \text{ fs.}$$

If the frequency is calculated using the other half of the curve ($R = 1.118$, $E = 8.69441$), then $k = 12.333$ fs, or k , average, = 12.185 fs.

2. From the gradient curve. The force constant is the derivative of the gradient wrt distance:

$$k = \frac{dG}{dx}.$$

Since we are using discrete points, the force constant is best obtained from finite differences:

$$k = \frac{(G_2 - G_1)}{(x_2 - x_1)}.$$

For $x_2 = 1.1100$, $G_2 = 27.098$ and for $x_1 = 1.0980$, $G_1 = -26.311$, giving rise to $k = 4450.75$ kcal/mol/ \AA^2 and a period of 12.185 fs.

3. From the vibrational frequency. Given a "frequency" (wavenumber) of vibration of N_2 of $\bar{\nu} = 2738.8 \text{ cm}^{-1}$, the period of oscillation, in seconds, is given directly by:

$$r = \frac{1}{c\bar{\nu}} = \frac{1}{2738.8 \times 2.998 \times 10^{10}},$$

or as 12.179 fs.

Summarizing, by three different methods the period of oscillation of N_2 is calculated to be 12.1851, 12.185 and 12.179 fs, average 12.183 fs.

Table 5.46: Velocities in DRC for N₂ Molecule

Time	Calculated Velocity	Linear Velocity	Diff. in Velocity
0.000	0.0	0.0	0.0
0.100	1340.6	1340.5	-0.1
0.200	2678.0	2681.0	-3.0
0.300	4007.0	4021.5	-14.5
0.400	5325.3	5362.0	-36.7
0.500	6628.4	6702.5	-74.1
0.600	7912.7	8043.0	-130.3

Table 5.47: Modified Velocities in DRC for N₂ Molecule

Time	Calculated Velocity	Simple Harmonic 25325.Sin(0.5296t)	Diff.
0.000	0.0	0.0	0.0
0.100	1340.6	1340.6	0.0
0.200	2678.0	2677.4	+0.6
0.300	4007.0	4006.7	+0.3
0.400	5325.3	5324.8	+0.5
0.500	6628.4	6628.0	+0.4
0.600	7912.7	7912.5	0.0

Initial dynamics of N₂ with N–N distance = 1.094 Å

A useful check on the dynamics of N₂ is to calculate the initial acceleration of the two nitrogen atoms after releasing them from a starting interatomic separation of 1.094 Å.

At R(N-N) = 1.094 Å, $G = -44.877$ kcal/mol/Å or -18.777×10^{19} erg/cm. Therefore acceleration, $f = -18.777 \times 10^{19} / 14.0067$ cm/sec/sec, or -13.405×10^{18} cm/s², which is $-13.405 \times 10^{15} \times$ Earth surface gravity.

Distance from equilibrium = 0.00980 Å. After 0.1 fs, velocity is $0.1 \times 10^{-15}(-13.405 \times 10^{18})$ cm/sec or 1340.5 cm/s.

In the DRC the time-interval between points calculated is a complicated function of the curvature of the local surface. By default, the first time-interval is 0.105fs, so the calculated velocity at this time should be $0.105/0.100 \times 1340.5 = 1407.6$ cm/s, in the DRC calculation the predicted velocity is 1407.6 cm/s.

The option is provided to allow sampling of the system at constant time-intervals, the default being 0.1 fs. For the first few points the calculated velocities are given in Table 5.46.

As the calculated velocity is a fourth-order polynomial of the acceleration, and the acceleration, its first, second and third derivatives, are all changing, the predicted velocity rapidly becomes a poor guide to future velocities.

For simple harmonic motion the velocity at any time is given by:

$$v = v_0 \sin(2\pi t/r).$$

By fitting the computed velocities to simple harmonic motion, a much better fit is obtained (Table 5.47).

The repeat-time required for this motion is 11.86 fs, in good agreement with the three values calculated using static models. The repeat time should not be calculated from the time required to go from a minimum to a maximum and then back to a minimum—only half a cycle. For all real systems the potential energy is a skewed

parabola, so that the potential energy slopes are different for both sides; a compression (as in this case) normally leads to a higher force-constant, and shorter apparent repeat time (as in this case). Only the addition of the two half-cycles is meaningful.

Conservation of normal coordinate

So far this analysis has only considered a homonuclear diatomic. A detailed analysis of a large polyatomic is impractical, and for simplicity a molecule of formaldehyde will be studied.

In polyatomics, energy can transfer between modes. This is a result of the non-parabolic nature of the potential surface. For small displacements the surface can be considered as parabolic. This means that for small displacements interconversion between modes should occur only very slowly. Of the six normal modes, mode 1, at 1209.5 cm^{-1} , the in-plane C–H asymmetric bend, is the most unsymmetric vibration, and is chosen to demonstrate conservation of vibrational purity.

Mode 1 has a frequency corresponding to 3.46 kcal/mol and a predicted vibrational time of 27.58 fs. By direct calculation, using the DRC, the cycle time is 27.59 fs. The rate of decay of this mode has an estimated half-life of a few thousands femtoseconds.

Rate of decay of starting mode

For trajectories initiated by an IRC= n calculation, whenever the potential energy is a minimum the current velocity is compared with the supplied velocity. The square of the cosine of the angle between the two velocity vectors is a measure of the intensity of the original mode in the current vibration.

Half-Life for decay of initial mode

Vibrational purity is assumed to decay according to zero'th order kinetics. The half-life is thus $-0.6931472t / \log(\langle \psi^2 \rangle^2)$ fs, where $\langle \psi^2 \rangle^2$ is the square of the overlap integral of the wavefunction for the original vibration with that of the current vibration. Due to the very slow rate of decay of the starting mode, several half-life calculations should be examined. Only when successive half-lives are similar should any confidence be placed in their value.

DRC print options

The amount of output in the DRC is controlled by three sets of options. These sets are:

- Equivalent Keywords H-PRIORITY, T-PRIORITY, and X-PRIORITY.
- Potential Energy Turning Point option.
- Geometry Maxima Turning Point options.

If T-PRIORITY is used then turning points cannot be monitored.

To monitor geometry turning points, put a "T" in place of the geometry optimization flag for the relevant geometric variable. In the example shown in Figure 5.23, the geometry of formaldehyde would first be optimized, then a FORCE calculation run, then a DRC calculation started, using the first normal mode for the starting velocity. Whenever the C=O bond length becomes a maximum or a minimum, a message is printed.

To monitor the potential energy turning points, put a "T" for the flag for atom 1 bond length (Do not forget to put in a bond-length (zero will do)!).

To monitor the geometry, use LARGE= n . This will cause the geometry to be printed once every n steps.

The effect of using these flags together is as follows.

```

IRC=1 DRC T=20
Formaldehyde
Monitoring the C=O Bond-length turning points
O   0.0 0   0.0 0   0.000000 0   0 0 0
C   1.2 T   0.0 0   0.000000 0   1 0 0
H   1.0 1 120.0 1   0.000000 0   2 1 0
H   1.0 1 120.0 1 180.000000 0   2 1 3
O   0.0 0   0.0 0   0.000000 0   0 0 0

```

Figure 5.23: Example of DRC calculation, monitoring a geometric variable

1. No options: All calculated points will be printed. No turning points will be calculated.
2. Atom 1 bond length flagged with a "T": If T-PRI0, etc. are NOT specified, then potential energy turning points will be printed.
3. Internal coordinate flags set to "T": If T-PRI0, etc. are NOT specified, then geometry extrema will be printed. If only one coordinate is flagged, then the turning point will be displayed in chronologic order; if several are flagged then all turning points occurring in a given time-interval will be printed as they are detected. In other words, some may be out of chronologic order. Note that each coordinate flagged will give rise to a different geometry: minimize flagged coordinates to minimize output.
4. Potential and geometric flags set: The effect is equivalent to the sum of the first two options.
5. T-PRI0 set: No turning points will be printed, but constant time-slices (by default 0.1 fs) will be used to control the print.

5.20 Use of SADDLE Calculation

The SADDLE technique is used for locating a transition state, given two geometries, one on each side of the transition state. In order for the SADDLE technique to work, the Z-matrix must be specified as follows:

- The first geometry, defining one geometry is defined as usual. If symmetry data is supplied, it should follow the first geometry. After the geometry, or geometry and symmetry data, there should be a blank line to indicate the end of the data
- The second geometry should then be specified. There must be a one-to-one correspondence of the atoms in the second geometry to those of the first geometry.

From this specification, it follows that if two molecules react to form one molecule, then the first geometry must contain all the atoms of the two molecules. The easiest way of defining such a geometry is to define one molecule, then have an unusually long bond-length from one atom in the first molecule to the first atom in the second molecule. The two molecules together form the first geometry. Likewise, if a molecule decomposes, e.g. $\text{C}_2\text{H}_5\text{OH} \rightarrow \text{C}_2\text{H}_4 + \text{H}_2\text{O}$, every atom in the product must be defined in the same order as the atoms in the reactant.

An example of a data-set for a SADDLE calculation, modeling the ethyl radical hydrogen migration from one methyl group to the other is given in Figure 5.24.

Details of the mathematics of SADDLE appeared in print in 1984 (M. J. S. Dewar, E. F. Healy, J. J. P. Stewart, *J. Chem. Soc. Faraday Trans. II*, 3, 227, (1984)), so only a superficial description will be given here.

The main steps in the saddle calculation are as follows:

```

Line 1:  UHF  SADDLE
Line 2:  ETHYL RADICAL HYDROGEN MIGRATION
Line 3:
Line 4:  C   0.000000  0    0.000000  0    0.000000  0    0  0  0
Line 5:  C   1.479146  1    0.000000  0    0.000000  0    1  0  0
Line 6:  H   1.109475  1   111.328433  1    0.000000  0    2  1  0
Line 7:  H   1.109470  1   111.753160  1   120.288410  1    2  1  3
Line 8:  H   1.109843  1   110.103163  1   240.205278  1    2  1  3
Line 9:  H   1.082055  1   121.214083  1    38.110989  1    1  2  3
Line 10: H   1.081797  1   121.521232  1   217.450268  1    1  2  3
Line 11: O   0.000000  0    0.000000  0    0.000000  0    0  0  0
Line 12: C   0.000000  0    0.000000  0    0.000000  0    0  0  0
Line 13: C   1.479146  1    0.000000  0    0.000000  0    1  0  0
Line 14: H   1.109475  1   111.328433  1    0.000000  0    2  1  0
Line 15: H   1.109470  1   111.753160  1   120.288410  1    2  1  3
Line 16: H   2.109843  1    30.103163  1   240.205278  1    2  1  3
Line 17: H   1.082055  1   121.214083  1    38.110989  1    1  2  3
Line 18: H   1.081797  1   121.521232  1   217.450268  1    1  2  3
Line 19: O   0.000000  0    0.000000  0    0.000000  0    0  0  0
Line 20:

```

Figure 5.24: Example of data for SADDLE calculation

1. The heats of formation of both systems are calculated.
2. A vector R of length $3N$ defining the difference between the two geometries in Cartesian coordinates is calculated. The scalar of the vector is called the BAR, and represents the distance between the two geometries.
3. The BAR is reduced by some fraction, normally about 5 to 15 percent. Normally, the default step is used, but this can be changed by use of $\text{BAR}=n.nn$, where $n.nn$ is the fraction. $\text{BAR}=0.15$ is the default.
4. The geometry of lower energy is identified; call this G .
5. Geometry G is optimized, subject to the constraint that it maintains a constant distance P from the other geometry.
6. If the newly-optimized geometry is higher in energy than the other geometry, and the last two steps involved the same geometry moving, make the other geometry G without modifying P , and go to 5.
7. Otherwise go back to 2.

The mechanism of 5 involves the coordinates of the moving geometry being perturbed by an amount equal to the product of the discrepancy between the calculated and required P and the vector R .

SADDLE works with Cartesian coordinates, so before the calculation starts, the two geometries are superimposed as much as possible. This is done as follows:

1. Both geometries are converted into Cartesian coordinates.
2. Both geometries are centered about the origin of Cartesian space.
3. One geometry is rotated until the difference vector is a minimum — this minimum is within 1 degree of the absolute bottom.
4. The SADDLE calculation then proceeds as described above.

The two geometries must be related by a continuous deformation of the coordinates. For this, internal coordinates are unsuitable in that while bond lengths and bond angles are unambiguously defined (being both positive), the dihedral angles can be either positive or negative. Clearly 300 degrees could equally well be specified as -60 degrees. A wrong choice of dihedral would mean that instead of the desired reaction vector being used, a completely incorrect vector was used, with disastrous results.

To prevent this, a SADDLE calculation will always convert coordinates into Cartesian before starting the run. If symmetry is to be used, then the geometry must be supplied in Cartesian coordinates, because internal symmetry relations are not meaningful here.

5.20.1 How to escape from a hilltop

A particularly irritating phenomenon sometimes occurs when a transition state is being refined. A rough estimate of the geometry of the transition state has been obtained by either a SADDLE or reaction path or by good guesswork. This geometry is then refined by TS, SIGMA or by NLLSQ, and the system characterized by a force calculation. Remember that NLLSQ is preferred over SIGMA when the GNORM is large, so NLLSQ is probably the method of choice, if for any reason TS does not work. It is at this point that things often go wrong. Instead of only one negative force constant, two or more are found. In the past, the recommendation has been to abandon the work and to go on to something less masochistic. It is possible, however, to systematically progress from a multiple maximum to the desired transition state. The technique used will now be described.

If a multiple maximum is identified, most likely one negative force constant corresponds to the reaction coordinate, in which case the objective is to render the other force constants positive. The associated normal mode eigenvalues are complex, but in the output are printed as negative frequencies, and for the sake of simplicity will be described as negative vibrations. Use a graphical user interface program to display the negative vibrations, and identify which mode corresponds to the reaction coordinate. This is the one we need to retain.

Hitherto, simple motion in the direction of the other modes has proved difficult. However the DRC provides a convenient mechanism for automatically following a normal coordinate. Pick the largest of the negative modes to be annihilated, and run the DRC along that mode until a minimum is reached. At that point, refine the geometry once more using TS and repeat the procedure until only one negative mode exists.

To be on the safe side, after each DRC+TS sequence do the DRC+TS operation again, but use the negative of the initial normal coordinate to start the trajectory. After both stationary points are reached, choose the lower point as the starting point for the next elimination. The lower point is chosen because the transition state wanted is the highest point on the lowest energy path connecting reactants to products. Sometimes the two points will have equal energy: this is normally a consequence of both trajectories leading to the same point or symmetry equivalent points.

After all spurious negative modes have been eliminated, the remaining normal mode corresponds to the reaction coordinate, and the transition state has been located.

This technique is relatively rapid, and relies on starting from a stationary point to begin each trajectory. If any other point is used, the trajectory will not be even roughly simple harmonic. If, by mistake, the reaction coordinate is selected, then the potential energy will drop to that of either the reactants or products, which, incidentally, forms a handy criterion for selecting the spurious modes: if the potential energy only drops by a small amount, and the time evolution is roughly simple harmonic, then the mode is one of the spurious modes. If there is any doubt as to whether a minimum is in the vicinity of a stationary point, allow the trajectory to continue until one complete cycle is executed. At that point the geometry should be near to the initial geometry.

Superficially, a line-search might appear more attractive than the relatively expensive DRC. However, a line-search in Cartesian space will normally not locate the minimum in a mode. An obvious example is the mode corresponding to a methyl rotation.

Keyword Sequences to be Used

1. To locate the starting stationary point given an approximate transition state: TS
2. To define the normal modes: FORCE ISOTOPE
At this point, copy all the files to a second filename, for use later.
3. Given vibrational frequencies of -654 , -123 , 234 , and 456 , identify *via* a GUI the normal coordinate mode, say the -654 mode. Eliminate the second mode by:

```
IRC=2 DRC T=30M RESTART LARGE
```

Use is made of the FORCE restart file.
4. Identify the minimum in the potential energy surface by inspection or using the “grep” command, of form:

```
grep '%' <Filename>.out
```
5. Edit out of the output file the data file corresponding to the lowest point, and refine the geometry using: TS
6. Repeat the last three steps but for the negative of the normal mode, using the copied files. The keywords for the first of the two jobs are:

```
IRC=-2 DRC T=30M RESTART LARGE
```
7. Repeat the last four steps as often as there are spurious modes.
8. Finally, carry out a DRC to confirm that the transition state does, in fact, connect the reactants and products. The drop in potential energy should be monotonic. If you are unsure whether this last operation will work successfully, do it at any time you have a stationary point. If it fails at the very start, then we are back where we were before—give up and go home!!

5.21 Polarizability and Hyperpolarizability Calculation

This section was written based on material provided by:

Henry Kurtz and Prakashan Korambath
Department of Chemistry
Memphis State University
Memphis TN 38152.

5.21.1 Time-Dependent Hartree-Fock

This procedure is based on the detailed description given by M. Dupuis and S. Karna (J. Comp. Chem. 12, 487 (1991)). The program is capable of calculating the quantities shown in Table 5.48.

Keywords for the POLAR calculation are given inside the POLAR keyword. Quantities under user-control are:

IWFLB= n The type of β calculation to be performed. This variable is only important if iterative beta calculations are chosen.

IWFLB=0 static (This is the default)

IWFLB=1 second harmonic generation

IWFLB=2 electrooptic Pockels effect

Table 5.48: Quantities Calculable using POLAR

Type of Phenomenon	Symbol
Frequency Dependent Polarizability	$\alpha(-\omega; \omega)$
Second Harmonic Generation	$\beta(-2\omega; \omega, \omega)$
Electrooptic Pockels Effect	$\beta(-\omega; 0, \omega)$
Optical Rectification	$\beta(0; -\omega, \omega)$
Third Harmonic Generation	$\gamma(-3\omega; \omega, \omega, \omega)$
DC-EFISH	$\gamma(-2\omega; 0, \omega, \omega)$
Optical Kerr Effect	$\gamma(-\omega; 0, 0, \omega)$
Intensity Dependent Index of Refraction	$\gamma(-\omega; \omega, -\omega, \omega)$

IWFLB=3 optical rectification

E=(n_1, n_2, n_3, \dots) The energies, in eV, of the radiation to be used. Up to 10 energies can be specified. If this option is not used, the default energies of 0.0, 0.25, and 0.50 eV will be used.

BETA= n Type of beta calculation.

BETA=0 $\beta(0;0)$ static (This is the default)

BETA=1 iterative calculation with type of β chosen by IWFLB

BETA=-1 Noniterative calculation of second harmonic generation

BETA=-2 Noniterative calculation of electrooptic Pockels effect

BETA=-3 Noniterative calculation of optical rectification

GAMMA= n Type of gamma calculation:

GAMMA=0 No gamma calculation

GAMMA=1 third harmonic generation (This is the default)

GAMMA=2 DC-EFISH

GAMMA=3 intensity dependent index of refraction

GAMMA=4 optical Kerr effect

TOL= $n.nn$ Cutoff tolerance for α calculations, default=0.001.

MAXITU= nnn Maximum number of interactions for beta, default: 500.

MAXITA= nnn Maximum number of iterations for α calculations, default: 150.

BTOL= $n.nn$ Cutoff tolerance for β calculations The default is 0.001.

Examples of POLAR keyword

To calculate the NLO quantities α , β , and γ at 1.0eV:

POLAR(E=(1.0))

This same calculation can be set up by setting all the variables to their default value:

POLAR(IWFLB=0, E=(1.), BETA=0, GAMMA=1, TOL=1.D-3, MAXITU=500, MAXITA=150, BTOL=1.D-3)

This takes up the entire keyword line. If more than one line is needed to hold the keyword, use the + option, as in:

```
+ symmetry 1scf uhf POLAR(IWFLB=0,E=(1.),BETA=0,GAMMA=1,TOL=1.D-3,MAXITU=501,
MAXITA=151,BTOL=1.D-3)
```

Note: This is not a recommended way of writing a keyword. In order for a keyword to be recognized, the 'join' of the two lines must be perfect. In other words, the last character of the first line must be in column 80, unless character 1 was not blank, in which case the last character must be in column 79. Anyhow, it is unlikely that such long keywords would be used very often.

5.22 COSMO (Conductor-like Screening Model)

Based on materials provided by

Dr Andreas Klamt
Burscheider Strasse 524
D-51381 Leverkusen
Germany

Unlike the Self-Consistent Reaction Field model [74], the **Conductor-like Screening Model** (COSMO) is a continuum approach [30] which, while more complicated, is computationally quite efficient. The expression for the total screening energy is simple enough to allow the first derivatives of the energy with respect to atomic coordinates to be easily evaluated.

The COSMO procedure generates a conducting polygonal surface around the system (ion or molecule), at the van der Waals' distance. By introducing a ε -dependent correction factor,

$$f(\varepsilon) = \frac{(\varepsilon - 1)}{(\varepsilon + \frac{1}{2})},$$

into the expressions for the screening energy and its gradient, the theory can be extended to finite dielectric constants with only a small error.

The accuracy of the method can be judged by how well it reproduces known quantities, such as the heat of solution in water (water has a dielectric constant of 78.4 at 25°C), Table 5.49. Here, the keywords used were:

```
NSPA=60 GRADIENTS 1SCF EPS=78.4 AM1 CHARGE=1
```

From the Table we see that the glycine zwitterion becomes the stable form in water, while the neutral species is the stable gas-phase form.

(After the COSMO paper was published, improvements in the method made the results shown in Table 5.49 invalid. However, the general conclusion that the method is of useful accuracy is still true.)

The COSMO method is easy to use, and the derivative calculation is of sufficient precision to allow gradients of 0.1 to be readily achieved.

5.22.1 A Walk Through COSMO

To explain the COSMO program, it is best to do a walk through the program first.

INITSV (INITialize SolVation)

This subroutine is called by the main program, if the `EPS=nn.n` keyword is set. Here, all initializations for the COSMO calculation are done.

Table 5.49: Calculated and Observed Hydration Energies

Compound	Method	ΔH_f (kcal/mol)		Hydration	
		gas phase	solution phase	ΔH (calc.)	Enthalpy(exp.) †
NH_4^+	AM1	150.6	59.5	91.1	88.0
$\text{N}(\text{Me})_4^+$	AM1	157.1	101.1	56.0	59.9
$\text{N}(\text{Et})_4^+$	AM1	132.1	84.2	47.9	57.0
Glycine					
neutral	AM1	-101.6	-117.3	15.7	--
zwitterion	AM1	-59.2	-125.6	66.4	--

†: Y. Nagano, M. Sakiyama, T. Fujiwara, Y. Kondo, J. Phys. Chem., **92**, 5823 (1988).

- In the DATA statement, the van der Waals (VDW) radii are set.
- The dielectric constant ϵ (EPSI) is read in from KEYWRD, and transformed to the dielectric factor FEPSI = $(\epsilon-1)/(\epsilon+0.5)$.
- The number of interatomic density matrix elements NDEN is calculated from NORBS and NUMAT:

$$\text{NDEN} = 3 \times \text{NORBS} - 2 \times \text{NUMAT}$$
- The solvent radius, RSOLV, for the construction of the SAS is read in off RSOLV=*n.nnn* (default: 1.0 Ångstroms).
- DISEX is set (default = 2). DISEX controls the distance within which the interaction of two segments is calculated accurately using the basic grid.
- The solvation radius SRAD is set for each atom. Be careful: the distance of the SAS will be SRAD-RDS.
- NSPA (Number of Segments Per Atom) is set (default = 42). NSPA is the number of segments created on a full VDW sphere.
- To guarantee the most homogeneous initial segment distribution, it is best to create one of the “magic numbers” (see DVFILL) of segments. Therefore the next smallest “magic number” compared to NSPA is evaluated and the corresponding set of direction vectors is stored in DIRSM.

For hydrogens the number of segments can safely be reduced by a factor of three or four, since the potential is rather homogeneous on hydrogen spheres. This is done and the corresponding direction vectors are stored in DIRSM.

- By the equation:

$$\text{NSPA} \times \pi \times \text{RSEG}^2 = 4\pi R^2$$

the mean segment radius can easily be found to be

$$\text{RSEG} = 2R/\sqrt{\text{NSPA}}.$$

With a mean VDW-radius of 1.5 Ångstroms, the mean radius is $R = 1.5 + \text{RSOLV} - \text{RDS}$. Thus DISEX2 is set to the squared mean distance of two segments multiplied by DISEX.

DVFILL (Direction Vector FILLing)

This routine constructs a homogeneous set of points on a unit sphere (direction vectors). It starts with 12 face-centers on a regular dodecahedron, i.e. the 12 corners of a regular icosahedron. These form 20 regular triangles. If the centers of these 20 triangles are added to the 12 initial directions, a new set of triangles is created, which, by a Wigner-Seitz construction on the unit sphere corresponds to 32 faces, 12 pentagons, and 20 hexagons. Thus we have constructed a soccerball. By iterative addition of the triangle centers, any number N of homogeneously distributed points can be generated, which can be written in the form $N = 3^i \times 10 + 2$, with i being an integer ($N = 12, 32, 92, 272, \dots$).

But there is a second way of generating a finer mesh of regular triangles from a cruder one. Instead of additional centers, the midpoints of the edges can be added. This roughly corresponds to an increase in the number of directions by a factor of four, instead of three. By combining these two procedures, a number $N = 3^i \times 4^j \times 10 + 2$ of directions can be created. These are the allowed values of NPPA in the DVFILL routine (“magic numbers”). The smallest are 12, 32, 42, 92, 122, 162, The default for the construction of the fine grid is $1082 = 3^3 \times 4 \times 10 + 2$.

COSCAV/COSCAN (COSmo CAVity)

This routine constructs the SAS and calculates and inverts the surface charge interaction matrix A .

Then the new transformation matrix is built from the local coordinate system defined by the nearest neighbors. The transformation to a local coordinate system guarantees a higher stability of the segmentation during geometry optimization. Then the basic grid of direction vectors is transformed according to the new transformation matrix.

MKBMAT

This routine calculates the B -matrix, i.e., the density-surface charge interaction matrix.

- First the array COSURF is filled with the explicit segment center coordinates instead of directions only.
- Then in a loop over all atoms, all segments for all density matrix elements of each atom the Coulomb interaction of the density with the segment charge is calculated. With r being the atom-segment distance vector, the various elements are:

$$\begin{array}{ll}
 ss & r^{-1} \\
 sp_k & \frac{DD(i)r_k}{r^3} \\
 p_k p_k & r^{-1} + QQ(i)^2 \left(\frac{3r_k^2}{r^5} - r^{-3} \right) \\
 p_k p_l & 6QQ(i)^2 \frac{r_k r_l}{r^5}
 \end{array}$$

Here, $DD(i)$ and $QQ(i)$ are the atomic dipole and quadrupole lengths, respectively, as coded in MOPAC.

- The B -matrix is stored in BMAT.

DIEGRD (DIElectric GRaDient)

The dielectric part of the gradient is calculated and added to DXYZ. There are two dielectric gradient contributions:

$$d E_{die}/d R_k^\alpha = Q \left(\frac{d}{d R_k^\alpha} B \right) G - \frac{1}{2} G \left(\frac{d}{d R_k^\alpha} A \right) G$$

with $G = A^{-1} B Q$ being the vector of charges on the segments.

Table 5.50: Van der Waals radii (Å) used in COSMO

I	R	II	R	III	R	IV	R	V	R	VI	R	VII	R
H	1.08												
Li	1.80					C	1.53	N	1.48	O	1.36	F	1.30
Na	2.30			Al	2.05	Si	2.10	P	1.75	S	1.70	Cl	1.65
K	2.80	Ca	2.75									Br	1.80
												I	2.05

- First the COSURF transformation is done.
- Then Q and QS are calculated. (If you are interested in a visualization of the screening charges, you may write out QS and COSURF at this point.)
- Then the 'second part' of the gradient is calculated.
- In the next part the first part is calculated. Here the gradient of the density-segment-charge interactions have to be evaluated.
- Finally, COSURF is re-transformed.

5.22.2 COSMO Keywords

EPS=*n.n* Defines the dielectric constant of the solvent. This keyword triggers the whole COSMO.

NSPA=*nn* Controls the number of segments, default = 42.

DISEX=*n.n* Controls the radius, up to which the segment-segment interactions are evaluated on the basis of the basic grid points. Default = 2.0. For accurate calculations or very high dielectric energies³, (e.g. ions) larger values may be preferable. The calculation time may increase as DISEX² (until all interactions are calculated accurately).

RSOLV=*n.n* Effective VDW radius of the solvent molecule. Default = 1.0Å.

5.22.3 Solvent Accessible Surface

The solvent accessible surface is a continuous surface of the molecule which can be reached by the center of charge of a solvent molecule. The calculation of the SAS is carried out as follows:

- Each atom is assigned a van der Waals' radius. VdW radii used in COSMO are given in Table 5.50.
- To each radius is added a distance equal to the radius of the solvent. By default, this is 1.0Å, but may be changed by the user using RSOLV=*n.nn*. This gives the distance from the nucleus to the center of a solvent molecule.
- A set of points is generated on this surface. These points produce a basic grid.
- All points which are inside the surface of any other atom are excluded.

³The dielectric energy is the energy of stabilization arising from the interaction of the charges in the solute with the induced charges on the solvent accessible surface plus the electrostatic energy due to the charges on the SAS interacting with each other.

- The remaining points are moved towards the center of the atom. The distance moved is equal to the distance of the center of charge of the solvent molecule from the center of the solvent molecule. By default, this distance is set to `RSOLV`, but may be set explicitly by keyword `RSOLV=n.nn`.
- Each of the remaining points represents a small area of the solvent accessible surface. The total SAS is calculated from the number of points.

From this definition of the SAS we see that the SAS of each atom is a surface of radius equal to the van der Waals' radius plus the radius of the solvent molecule minus the distance of the center of charge of the solvent molecule to the center of the solvent molecule. In other words, the radius is the VdW radius plus the distance from the surface of the solvent molecule to the center of charge of the solvent molecule. By default, this extra distance is zero. Only that part of the atom surface which can be touched by the solvent molecule is used. This means that only those atoms on the surface of the molecule can contribute to the SAS. Of those atoms that are on the surface of the molecule there will be parts of the surface which cannot be reached by the solvent because the solvent molecule is too bulky.

Some hints on the use of COSMO

- 1SCF calculations run in general without problems. On gas-phase geometries they give useful solvation energies for neutral rigid molecules.
- For geometry optimization Eigenvector Following has proved to be most efficient in combination with COSMO. Gradient norms of about 1% of the dielectric energy should be reachable, in many cases even less. Nevertheless, don't use a too small `GNORM` criterion, since the calculation may have convergence problems.
- Keep in mind that energy differences of about 1% of the dielectric energy may arise due to small differences in the segmentation.
- Dr Klamt does not recommend the use of COSMO in `FORCE` calculations at the present time.
- UHF calculations should run without additional problems.
- C.I. calculations can now be done, and C.I. gradients are now valid; this has been the result of recent work by Dr Klamt.

5.23 Parametric Molecular Electrostatic Potential (PMEP)

The PMEP procedure [45, 46] is a technique for rapidly calculating the electrostatic properties of a molecule. Written by Prof. George Ford and Dr. Bingze Wang⁴ at Southern Methodist University, Dallas, Texas, the procedure is ideally suited for large systems.

The PMEP procedure has two main functionalities: first, to generate a 2-D grid of points giving the Electrostatic Potential (ESP) in a cross-section through a system, and second, to generate atomic charges based on the calculated ESP. At present, the method is limited to AM1 systems containing H, C, N, O, F, Cl, only.

⁴Current address: IRBM, Via Pontina Km.30.600, 00040 Pomezia (Roma), Italy

```

1scf AM1 PMEPE MINMEP PRTMEP
Formaldehyde (Cross-section in plane of molecule)
Generate a 2-D grid of PMEPE potentials for 'esplot' to use
O 0.00000000 0 0.0000000 0 0.0000000 0 0 0 0 -0.2759
C 1.22732374 1 0.0000000 0 0.0000000 0 1 0 0 0.1384
H 1.11047287 1 122.2253516 1 0.0000000 0 2 1 0 0.0688
H 1.11048351 1 122.2158646 1 179.9998136 1 2 1 3 0.0687

```

Figure 5.25: Data Set for PMEPE Calculation of Formaldehyde

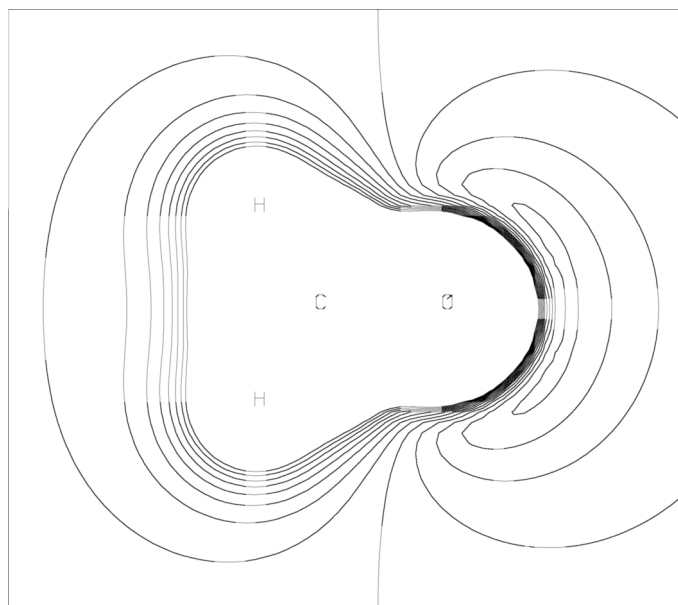


Figure 5.26: Parametric Molecular Electrostatic Potential around Formaldehyde

5.23.1 2-D Electrostatic Potential Plots

ESP plots are generated in two steps. First, a MOPAC calculation generates a 2-D grid of points. This grid is then converted into a picture by the utility program ESPLLOT. ESPLLOT is very simple to use: the command is `esplot <filename>`, where `<filename>` is the name of the data-set. ESPLLOT generates an on-line picture of the PMEPE, and a HPGL file suitable for use in generating hard-copy. Because ESPLLOT is so simple, it will not be discussed further. Instead, the rest of this discussion applies to the MOPAC calculation.

The grid generated by MOPAC consists of a 2-D array of points representing a cross-section through the system. The distance between points is a constant 0.1 Ångstroms. The size represented by the grid is roughly 4 Ångstroms plus twice the size of the system. For example, N_2 has a N–N distance of 1.1 Å, and the default associated grid represents a rectangular area of 5.8 by 4.6 Å. Each grid point represents the potential in kcal/mol which a unit positive charge would experience due to the electrostatic field of the system.

ESP grids are generated by specifying (see p. 65) and PRTMEPE. An example of a data-set for the PMEPE procedure is shown in Figure 5.25. The PMEPE plot for this data set is shown in Figure 5.26. This plot can be compared with the on page 55.

Choice of Plane to be Calculated

By default, the grid is centered on the center of the molecule, and the X-Y plane at Z=0 is selected. Other grids can be chosen using PMEPR. PMEPR uses three atoms and an optional offset to define the plane to be used. It has enough options to allow any plane to be easily specified.

5.23.2 Atomic Charges

By use of QPMEP, a set of atomic charges can be calculated. This set of charges is the best least squares fit to the charges which reproduce the ESP of the Connolly or Williams surfaces.

5.24 Miertus-Scrocco-Tomasi Solvation Model

Based on materials provided by

Modesto Orozco, Dep. Bioquímica, Facultat de Química, Universitat de Barcelona, C/ Martí i Franqués 1, Barcelona 08028. SPAIN. E-mail: modesto@far1.far.ub.es.

and

F. J. Luque, Dep. Farmacia, Unitat Físicoquímica, Facultat de Farmacia, Universitat de Barcelona, Avgda Diagonal sn, Barcelona 08028, SPAIN. E-mail: javier@far1.far.ub.es.

5.24.1 Outline of the MST Method

Solvation is computed using a modified version of Miertus, Scrocco and Tomasi [75, 76] self-consistent reaction field method (MST/SCRF), which has been modified [77, 78, 79, 80, 81] to allow semiempirical Hamiltonians to be used. Following this strategy the solute is placed in a molecular-shaped cavity, which is built using GEPOL routines [82] which use optimized van der Waals' radii [83, 84, 85]. The solvent is represented as a continuum which reacts against the solute charge distribution *via* a perturbation operator. The perturbation operator is obtained by solving the Laplace equation at the solute/solvent interface (the solute cavity). The solute electrostatic potential, which is necessary to solve the Laplace equation, is rigorously computed as the expectation value of the r^{-1} operator [75, 76, 77, 78, 79, 80, 81].

The free energy of hydration is computed as the addition of three contributions:

1. The electrostatic term, which is computed from the linear free energy response theory [75, 76, 77, 78, 79, 80, 81].
2. The cavitation contribution, which is computed from Pierotti's scaled particle theory [83, 84].
3. The van der Waals terms, which is computed using a linear relation with the solute accessible surface, and optimized "hardness" parameters [85].

In addition to the free energy of hydration a "solvent-adapted" wavefunction is obtained. Such a wavefunction can be used to determine changes in solute properties due to the solvent [86, 87, 88, 89].

The program allows the use of two different strategies to compute the semiempirical Molecular Electrostatic Potential (MEP). The first one [90, 91, 40] computes the MEP via the deorthogonalization of the semiempirical wavefunction (no keyword), while the second one [92, 93] maintains the orthogonality of the wavefunction in the calculation of the MEP (keyword ORT).

The strategy has been tested in different systems. Optimized versions for MNDO, AM1 and PM3 are available. The comparison of experimental and theoretical free energies of hydration for 23 small neutral molecules yields to

```

Line
*  PM3 TOM
*
*  O  .0000000 0      .000000 0      .000000 0      0 0 0
*  H  .9512971 0      .000000 0      .000000 0      1 0 0
*  H  .9512971 0 107.457468 0      .000000 0      1 2 0
*
1  78.5  1.0  1  1  0  3
2  298.15  18.07  2.77  0.00026  71.690  0.657  1.277
3  40.0  1.5  0.2  0.7  0.02  5  3  1
4  1  1.75000  8
5  2  1.00000  99
6  3  1.00000  99
7  1.00

```

Figure 5.27: Data set for Calculation of Water using the Miertus-Scrocco-Tomasi Model

RMS deviations in the range of 1 kcal/mol. The best results (RMS 1.0 kcal/mol, $r=0.94$) are obtained with the AM1 Hamiltonian with orthogonal MEP's. The solvent-induced dipoles matches Monte Carlo/QM and *ab initio* SCRF values.

Semiempirical/SCRF methods have been successfully applied to several problems. Examples: i) tautomerization of nitroimidazole, and pyridones, ii) solvent effects in the isomerization of amides, and iii) H-bond dimerization.

Warnings

- The solute geometry cannot be properly optimized in solution. Optimization flags at the different coordinates should be set to 0
- The program has been carefully parametrized to work with neutral molecules. A reduction of the cavity for charged solutes may be necessary [94].
- We cannot recommend the use of the program for solvents other than water, chloroform, or carbon tetrachloride at the present moment. A simple change in the dielectric constant cannot be enough to reproduce the characteristics of other solvents.

5.24.2 Data Requirements for MST Model

For the three solvent for which parameters have been published, special keywords exist. These are H2O, CHCL3, and CCL4. If other solvents are to be modeled, extra data must be provided at the end of the data set. This extra data is described by example, using water as the solvent. (This extra data could be avoided by using H2O.)

For water, the data needed is shown in Figure 5.27.

The extra data are indicated by the lines numbered 1 to 7 in the Figure. These data are described as follows:

Line 1. Parameters indicating the level of computation

(see Miertus, Scrocco, Tomasi, Chem. Phys., 55 (1981) 117 and Miertus, Tomasi, Chem. Phys., 65 (1982) 239 for further details)

Layout of data: A B C D E F

Datum	Value	Name	Description
A	78.5	EPS	Solvent dielectric constant
B	1.0	DMP	Self-polarization acceleration factor
C	1	MC	Self-polarization cycles
D	1	ICOMP	Charge compensation
E	0	IFIELD	Reaction field calculation on nuclei
F	3	ICAV	Cavitation energy computation

Line 2. Parameters related to the cavitation energy

(This line is needed only if ICAV \neq 0)

Layout of data: A B C D E F G

Datum	Value	Name	Description
A	298.15	TABS	Absolute temperature (in kelvin)
B	18.07	VMOL	Solvent molar volume (\AA^3)
C	2.77	DMOL	Solvent molar diameter (\AA)
D	0.00026	TCE	Thermal expansion coefficient (kelvin^{-1})
E	71.69	STEN	Surface tension ($\text{dyne} \times \text{cm}^{-1}$)
F	0.657	DSTEN	Surface tension derivative
G	1.227	CMF	Cavity microscopic coefficient

If ICAV=1 parameters A-D are needed, and Pierotti's cavitation energy is computed.

If ICAV=2 parameters E-G are needed, and Sinanoglu's cavitation energy is computed.

If ICAV=3 parameters A-G are needed, and Pierotti's and Sinanoglu's cavitation energies are computed.

Line 3. Parameters related to the cavity surface

(see Pacual-Ahuir, Silla, Tomasi, Bonaccorsi, J. Comput. Chem. 77 (1982) 3654 for further details)

Layout of data: A B C D E F G H

Datum	Value	Name	Description
A	40.0	OMEGA	Sphere overlapping parameter
B	1.5	RD	Solvent molecular radius
C	0.2	RET	Minimal radius to define new spheres
D	0.7	FRO	Numerical factor to define new spheres
E	0.02	DR	Step to compute electric field on cavity surface
F	5	NDIV	Parameter of cavity surface partition
G	3	NESF	Number of spheres
H	1	ICENT	Definition of spheres

Lines 4–6. Coordinates and radii of spheres used to build up cavity surface

ICLASS specifies the type of atom. This information is used in the calculation of the van der Waals' energy, determines as the sum of products between the surface of each atom and a parameter which is characteristic for each atom.

ICLASS is, in fact, the atomic number of each atom, with the only exception of hydrogen, for which we distinguish between hydrogen atoms bound to a heteroatom (ICLASS=99) and those bound to a carbon (ICLASS=1).

If ICENT=0, then five parameters are needed for each sphere.

Layout of data: A B C D E

Datum	Value	Name	Description
A	n/a	XE	X-coordinate
B	n/a	YE	X-coordinate
C	n/a	ZE	X-coordinate
D	n/a	RE	Radius of sphere
E	n/a	ICLASS	Type of atom

If $ICENT \neq 0$, then three parameters are needed for each sphere. The spheres are defined as being centered on each atom.

Layout of data: A B C

Datum	Values			Name	Description
A	1	2	3	NC1	Atom Number
B	1.75	1.0	1.0	RE	Radius of sphere
C	8	99	99	ICLASS	Type of atom

Line 7. Factor used to scale the MEP in the deorthogonal procedure

Layout of data: A

Datum	Value	Name	Description
A	1.0	FACTOR	Factor used to scale the MEP

Chapter 6

Accuracy of Methods in MOPAC

The following tables illustrate the relative accuracy of three semiempirical methods: MNDO, AM1, and PM3. PM3 is a re-parameterization of the MNDO method, in which the AM1 form of the core-core interaction is used. In cases where two sets of MNDO parameters have been published, the more recent set will be used.

6.1 Protocols used in Determining Accuracy

The selection of species to be used to demonstrate the accuracy of the various methods presented some difficulty. As the designer of the AM1 method, I have an understandable bias to show that AM1 is better than MNDO. As the sole author of PM3, I also have an understandable bias to show that PM3 is better than AM1. However, in order to be credible from a reader's point of view, protocols need to be set up which will convince the reader that these biases have not affected the discussions or conclusions in this Chapter.

After some thought, I decided to adopt the following protocols:

1. All species for which reliable experimental data were available would be used in generating the raw data for the statistical analyses.
2. Experimental data of low or unknown accuracy was used as reference data if, and only if, a calculation using that data was done.
3. All results of all data would be reported in un-corrected and un-modified form.
4. Only data for which references were available would be used.
5. Experimental data included some high-level *ab initio* results. Where *ab initio* results were used, the *ab initio* method (basis set plus post-Hartree Fock corrections, if necessary) was of much higher reliability than the semiempirical method. In other words, any differences between *ab initio* and semiempirical results would be due to errors in the semiempirical method.

The second of these protocols needs a word of explanation. Consider the case of a compound whose ΔH_f is of highly questionable accuracy, say 10 ± 30 kcal/mol. Should this system be used in discussions of accuracy? If the calculated PM3 ΔH_f was 11 kcal/mol and the calculated AM1 ΔH_f was 40 kcal/mol, then PM3 would obviously look better than AM1. In that case, my bias would be to include the compound. Conversely, if the results were reversed, then my bias would suggest that clearly the reference datum was very inaccurate, and therefore should not be included. This is obviously faulty reasoning. Therefore, in order to prevent bias affecting the selection of compounds to be used in these discussions, the decision of whether to include any specific species was made before the calculation was done.

By adopting these protocols, and by extensive use of automatic table generation, and by an absolute minimum of human intervention, I hope the reader will be convinced that bias has been kept to an absolute minimum.

In the Tables, the accuracy of the reference data is not given. This is a known fault. However, addition of the accuracy data would be a lot of work, and has not yet been done.

6.2 Structure of the Tables

Because of the size of these Tables, finding any specific system can be quite time-consuming. To help locate species, the order of occurrence of species within a Table follows very specific rules. These are:

- The order of occurrence is dictated by the empirical formula only.
- Elements in the empirical formula are arranged as in J. O. Cox, G. Pilcher, "Thermochemistry of Organic and Organometallic Compounds", Academic Press, New York, N.Y., 1970, e.g. $H < C < O < N < S < F < Cl < Br < I < Li < Na < K < Rb < Cs < Be < Mg < Ca < Sr < Ba < B < Al < Ga < In < Tl < Si < Ge < Sn < Pb < P < As < Zn < Cd < Hg < Sb < Se < Te < Bi$.

Average ΔH_f errors for different types of compound are given in Table 6.4, and summarized by element in Table 6.11. Because of the importance of ΔH_f s, the raw data are presented in Table 6.25. Users often want to know where methods do not work well, so the worst predicted ΔH_f are presented for each method. These are shown in Tables 6.1, 6.2 and 6.3.

Summarizing geometry errors proved difficult because of the sheer volume of data. Because bond-lengths are very important, and because of the relatively small amount of data on angles and dihedrals, only average bond-length errors are presented. For bond-types, average errors are given in Table 6.12. This data is further summarized by element in Table 6.16. All the raw data are given in Table 6.52.

Although Ionization Potential (I.P.) and dipole moment data is important, the importance is much less than ΔH_f and geometric data, therefore only summaries are given here, see Tables 6.17, 6.18 and 6.19 for the worst predicted values of I.P.s, and Table 6.20 for a summary by element. For dipole moments, the corresponding Tables are 6.21, 6.22, 6.23, and 6.24.

6.3 Limitations of MNDO

MNDO is the oldest of the three methods surveyed here, and as a direct result, is the least accurate. MNDO has many advantages over earlier semiempirical methods, but as MNDO is such a large improvement over those methods, enumeration of MNDO's good points would be invidious to the other methods. Instead, only the limitations likely to be encountered by users will be mentioned. The principal drawbacks to MNDO are:

1. Sterically crowded molecules are too unstable. Example: neopentane.
2. Four-membered rings are too stable. Example: cubane.
3. The hydrogen bond is virtually non-existent. Example: water dimer.
4. Hypervalent compounds are too unstable. Example: sulfuric acid.
5. Activation barriers are generally too high.
6. Non-classical structures are predicted to be unstable relative to the classical structure. Example: ethyl radical.

7. Oxygenated substituents on aromatic rings are out-of-plane. Example: nitrobenzene.
8. The peroxide bond is systematically too short by about 0.17Å.
9. The C–O–C angle in ethers is too large by about 9°.

6.4 Accuracy of AM1

AM1 is a distinct improvement over MNDO, in that the overall accuracy is considerably improved. Specific improvements are:

1. The strength of the hydrogen bond in the water dimer is 5.5 kcal/mol, in accordance with experiment.
2. Activation barriers for reaction are markedly better than those of MNDO.
3. Hypervalent phosphorus compounds are considerably improved relative to MNDO.
4. In general, errors in ΔH_f obtained using AM1 are about 40% less than those given by MNDO.

Unfortunately, with this improvement a few deficiencies were introduced. The most important of these are:

1. AM1 phosphorus has a spurious and very sharp potential barrier at 3.0 Ångstroms. The effect of this is to distort otherwise symmetric geometries and to introduce spurious activation barriers. A vivid example is given by P_4O_6 , in which the notionally equal P–P bonds are predicted by AM1 to differ by 0.4 Å. This is by far the most severe limitation of AM1.
2. Alkyl groups have a systematic error due to the heat of formation of the CH_2 fragment being too negative by about 2 kcal/mol.
3. Nitro compounds, although considerably improved relative to MNDO, are still systematically too positive in energy.
4. The lowest-energy conformer for ethanol is predicted to have a gauche C–C–O–H angle. High level *ab initio* calculations, and experimental evidence, suggests that the correct geometry is trans. This error was reported by Andreas Jabs, at Martin-Luther-Universitaet Halle-Wittenberg, Halle/Saale, Germany.
5. The peroxide bond is still systematically too short by about 0.17Å.

6.5 Accuracy of PM3

PM3 is a distinct improvement over AM1. Evidence of this is:

1. Hypervalent compounds are predicted with considerably improved accuracy.
2. Overall errors in ΔH_f are reduced by about 40% relative to AM1.

6.6 Faults and Errors in PM3

Although PM3 is more accurate in predicting the ΔH_f of many systems than either AM1 or MNDO, it does have some limitations. The more important of these are:

1. The barrier to rotation in formamide is practically non-existent. In part, this can be corrected by the use of the MMOK option, see Section 5.3.8. Fault discovered by Prof Tim Clark.
2. Minima are predicted for which no experimental systems are known. By inference, these minima are spurious [95].
3. Proton affinities are inaccurate. The distribution of proton affinities is essentially "shotgun." In other words, there is a high, non-systematic scatter from the experimental values compared with all other popular methods [96] (including AM1).
4. CBr_4 is predicted to have a D_{2d} rather than T_d geometry (W. Thiel, private comm.).
5. The lowest-energy conformer for ethanol is predicted to have a gauche C–C–O–H angle. High level *ab initio* calculations, and experimental evidence, suggests that the correct geometry is trans. This error was reported by Andreas Jabs, at Martin-Luther-Universitaet Halle-Wittenberg, Halle/Saale, Germany.
6. Some polycyclic rings are predicted to have butterfly shapes rather than being flat.
7. Almost all sp^3 nitrogen systems are predicted to be pyramidal, in variance with experiment and other methods.
8. The charge on nitrogen atoms are most likely of incorrect sign and of unrealistic magnitude. This was due to faults in the parameterization.
9. H-non-bonded H contacts are too short, at about 1.7Å. They should be more than 2Å. One consequence is that hydrogen bonds are too short by about 0.1Å.
10. The LUMO of methane is of symmetry a_1 instead of the correct symmetry of t_2 .

The tables given here have been carefully checked for accuracy. In many instances the calculated results differ from those of earlier publications. In most cases, the difference has been investigated and found to be due to errors in the earlier work. This does not, of course, mean that the tables are fully accurate, but only that care has been taken to minimize error.

Table 6.1: Worst ΔH_f Errors in MNDO

Formula	Chemical Name	Exp. ΔH_f	Calc. ΔH_f	Diff.
O ₇ Cl ₂	Cl ₂ O ₇	65.00	678.55	613.5
S ₂ F ₁₀	S ₂ F ₁₀	-493.40	50.79	544.2
F ₇ I	Iodine heptafluoride	-229.70	105.16	334.9
F ₅ I	Iodine pentafluoride	-200.84	130.85	331.7
O ₃ FCI	Perchloryl fluoride	-5.12	323.22	328.3
SF ₆	Sulfur hexafluoride	-291.40	29.18	320.6
SOF ₄	SOF ₄	-235.50	33.51	269.0
P ₄ O ₁₀	Phosphorus pentoxide	-694.09	-431.88	262.2
F ₅ Cl	Chlorine pentafluoride	-54.00	204.73	258.7
SF ₅	Sulfur pentafluoride	-217.10	-8.99	208.1
F ₅ Br	Bromine pentafluoride	-102.50	104.84	207.3
SO ₂ F ₂	Sulfuryl fluoride	-181.30	21.90	203.2
SOF ₃	SOF ₃	-185.10	-5.00	180.1
H ₂ SO ₄	Sulfuric acid	-175.70	-13.13	162.6
N	Nitrogen (++)	1133.90	973.15	-160.8
SF ₅	Sulfur pentafluoride (-)	-291.00	-131.66	159.3
C ₂ H ₆ SO ₄	Dimethyl sulfate	-164.10	-5.58	158.5
SO ₃	Sulfur trioxide	-94.60	58.46	153.1
PF ₆	Phosphorus hexafluoride (-)	-522.00	-369.69	152.3
C ₃ H ₈ SO ₂	Methylethyl sulfone	-97.60	47.01	144.6
C ₄ H ₁₀ SO ₂	Diethyl sulfone	-102.50	40.49	143.0
C ₂ H ₆ SO ₂	Dimethyl sulfone	-89.10	53.64	142.7
SOCl ₄	SOCl ₄	-55.70	86.91	142.6
SCl ₆	Sulfur hexachloride	-19.80	118.29	138.1
C ₄ H ₆ SO ₂	Divinyl sulfone	-36.00	101.33	137.3
SF ₄	Sulfur tetrafluoride	-182.40	-46.51	135.9
PF ₅	Phosphorus pentafluoride	-381.10	-248.82	132.3
SO ₂ Cl ₂	Sulfuryl chloride	-86.20	44.58	130.8
SO ₂ Br ₂	Sulfuryl bromide	-59.50	68.33	127.8
SO ₂ F	SO ₂ F	-113.20	13.99	127.2
Al ₂ I ₆	Al ₂ I ₆	-117.00	7.41	124.4
Si ₂ Br ₆	Hexabromodisilane	-182.80	-65.96	116.8
F ₃ Cl	Chlorine trifluoride	-38.00	78.69	116.7
O ₂ Cl	Chlorine dioxide	25.00	135.99	111.0
SBr ₆	Sulfur hexabromide	58.80	166.34	107.5
S ₂ F ₂	SSF ₂	-95.94	7.94	103.9
C ₅ H ₈ N ₄ O ₁	Pentaerythritol tetranitrate	-92.50	1.23	93.7
SO ₂ Cl	SO ₂ Cl	-66.40	25.80	92.2
SF ₃	Sulfur trifluoride	-130.00	-38.53	91.5
Al ₂ Br ₆	Al ₂ Br ₆	-224.00	-132.67	91.3
POF ₃	Phosphorus oxyfluoride	-289.50	-199.54	90.0
SiO ₂	SiO ₂	-99.40	-9.74	89.7
SO ₂ Br	SO ₂ Br	-52.80	36.64	89.4
CH ₅ PO ₃	Methylphosphonic acid	-240.50	-152.09	88.4
CH ₃ POF ₂	Methylphosphonodifluoride	-233.23	-147.22	86.0
SiOBr ₂	SiOBr ₂	-137.40	-51.46	85.9
PSBr ₃	Phosphorus thiobromide	-67.20	18.43	85.6

Table 6.2: Worst ΔH_f Errors in AM1

Formula	Chemical Name	Exp. ΔH_f	Calc. ΔH_f	Diff.
O ₇ Cl ₂	Cl ₂ O ₇	65.00	554.56	489.6
F ₅ I	Iodine pentafluoride	-200.84	94.43	295.3
F ₇ I	Iodine heptafluoride	-229.70	44.36	274.1
O ₃ FCI	Perchloryl fluoride	-5.12	246.46	251.6
P ₄ O ₆	Phosphorus trioxide	-529.23	-321.67	207.6
Be ₃ O ₃	Be ₃ O ₃	-251.90	-45.39	206.5
F ₅ Br	Bromine pentafluoride	-102.50	81.11	183.6
N	Nitrogen (++)	1133.90	972.93	-161.0
F ₅ Cl	Chlorine pentafluoride	-54.00	90.39	144.4
Be ₂ OF ₂	Be(OF) ₂	-287.90	-162.01	125.9
PSBr ₃	Phosphorus thiobromide	-67.20	44.74	111.9
SI ₄	Sulfur tetraiodide	120.20	17.48	-102.7
SBr ₄	Sulfur tetrabromide	53.00	-35.51	-88.5
BeH ₂ O ₂	Beryllium di-hydroxide	-161.80	-76.01	85.8
SI ₃	Sulfur triiodide	100.30	16.33	-84.0
F ₃ Br	Bromine trifluoride	-61.10	21.47	82.6
SCL ₄	Sulfur tetrachloride	-0.70	-82.73	-82.0
O ₂ Cl	Chlorine dioxide	25.00	105.87	80.9
B ₂ H ₄	B ₂ H ₄ D2d	45.70	126.57	80.9
SiO ₂	SiO ₂	-99.40	-18.69	80.7
Be ₂ O	Beryllium oxide	-15.00	65.34	80.3
SiOI	SiOI	-53.30	23.43	76.7
SBr ₃	Sulfur tribromide	50.20	-26.16	-76.4
POI ₃	Phosphorus oxyiodide	-39.70	36.16	75.9
AlOF ₂	AlF ₂ O	-265.00	-191.91	73.1
SCL ₃	Sulfur trichloride	8.80	-60.10	-68.9
SI ₅	Sulfur pentaiodide	130.90	62.25	-68.7
POBr ₃	Phosphorus oxybromide	-97.00	-29.05	68.0
F	Fluoride (-)	-61.00	3.44	64.4
Zn	Zinc (++)	665.10	729.44	64.3
PBr ₅	Phosphorus pentabromide	-11.00	52.99	64.0
SF ₅	Sulfur pentafluoride	-217.10	-280.38	-63.3
P	Phosphorus (++)	775.10	712.09	-63.0
SiOBr ₂	SiOBr ₂	-137.40	-77.39	60.0
COI	COI	63.50	4.72	-58.8
PSCl ₃	Phosphorus thiochloride	-91.00	-32.31	58.7
F ₃ Cl	Chlorine trifluoride	-38.00	20.19	58.2
SI ₂	Sulfur diiodide	81.90	25.68	-56.2
BeF ₂	Beryllium difluoride	-189.70	-134.91	54.8
SCL ₆	Sulfur hexachloride	-19.80	34.76	54.6
AlOF ₂	AlF ₂ O (-)	-311.56	-257.16	54.4
Cl ₄	Carbon tetraiodide	108.20	54.18	-54.0
SOCl ₄	SOCl ₄	-55.70	-2.82	52.9
Cl ₃	Triiodomethyl	117.30	64.59	-52.7
CGe	Germanium carbide	151.00	203.55	52.6
C ₃ H ₆ N ₆	Melamine	12.40	64.25	51.9
BeBr ₂	Beryllium dibromide	-54.80	-106.08	-51.3
H	Hydrogen (+)	365.72	314.91	-50.8
SiOCl ₂	SiOCl ₂	-167.70	-118.35	49.3

Table 6.3: Worst ΔH_f Errors in PM3

Formula	Chemical Name	Exp. ΔH_f	Calc. ΔH_f	Diff.
N	Nitrogen (++)	1133.90	963.89	-170.0
P	Phosphorus (++)	775.10	633.00	-142.1
O ₇ Cl ₂	Cl ₂ O ₇	65.00	-26.27	-91.3
S ₂ F ₁₀	S ₂ F ₁₀	-493.40	-410.71	82.7
COI	COI	63.50	-2.93	-66.4
Al ₂ O ₂	Al ₂ O ₂ (+)	135.60	72.50	-63.1
Al	Al (+)	218.10	279.75	61.7
PSBr ₃	Phosphorus thiobromide	-67.20	-7.84	59.4
C ₂	Carbon, dimer	200.20	258.21	58.0
AlOF ₂	AlF ₂ O	-265.00	-208.55	56.4
SI ₄	Sulfur tetraiodide	120.20	69.22	-51.0
HAIO	AlOH (-)	-55.00	-105.68	-50.7
AlN	Aluminum nitride	125.00	75.45	-49.5
SiOI ₂	SiOI ₂	-99.40	-50.23	49.2
C ₄	Carbon, tetramer	232.00	280.29	48.3
SiOCl ₂	SiOCl ₂	-167.70	-121.23	46.5
SI ₃	Sulfur triiodide	100.30	55.26	-45.0
SiOBr ₂	SiOBr ₂	-137.40	-94.09	43.3
C ₃ InH ₉	Trimethyl indium	40.80	-2.15	-42.9
HPO ₃	HPO ₃	-135.00	-177.27	-42.3
OCl ₂	Chlorine monoxide	25.00	-16.26	-41.3
S ₂ F ₂	SSF ₂	-95.94	-56.11	39.8
InO	Indium oxide	92.50	52.97	-39.5
SbN	Antimony nitride	63.70	102.52	38.8
Al ₂	Al ₂	116.40	80.69	-35.7
Ga ₂ S	Ga ₂ S	5.00	40.20	35.2
C ₈ H ₈	Cubane	148.70	113.66	-35.0
MgN	Magnesium nitride	69.00	103.99	35.0
AlO	AlO (+)	237.30	202.36	-34.9
C ₃ GaH ₉	Trimethyl gallium	-10.80	23.82	34.6
SBr ₄	Sulfur tetrabromide	53.00	18.73	-34.3
Pb ₂	Lead, dimer	79.50	45.33	-34.2
SBr ₃	Sulfur tribromide	50.20	16.38	-33.8
C ₆ AsH ₁₅	Triethylarsine	13.40	-20.44	-33.8
Al	Aluminum, atom	79.49	46.03	-33.5
C ₆ SbH ₁₅	Triethylstibine	10.40	-23.08	-33.5
MgHO	Magnesium hydroxide	-22.00	-55.05	-33.0
All	All	16.24	49.32	33.1
C ₆ H ₁₅ PO ₄	Triethyl phosphate	-284.50	-251.56	32.9
GaO	Gallium oxide	66.80	35.33	-31.5
CH ₅ PO ₃	Methylphosphonic acid	-240.50	-209.04	31.5
POI ₃	Phosphorus oxyiodide	-39.70	-8.28	31.4
HCl ₂	Hydrogen dichloride (-)	-142.00	-111.02	31.0
COBr	COBr	20.50	-10.25	-30.8
PSCl ₃	Phosphorus thiochloride	-91.00	-60.23	30.8
CS	Carbon sulfide	67.00	97.32	30.3
SI ₂	Sulfur diiodide	81.90	51.57	-30.3
SCL ₆	Sulfur hexachloride	-19.80	10.27	30.1
AlCl ₂	AlCl ₂ (-)	-115.00	-144.99	-30.0

Table 6.4: Average Signed and Unsigned Errors in ΔH_f by Compound Type

Elements in compounds	PM3			MNDO			AM1		
	No.	Unsigned	Signed	No.	Unsigned	Signed	No.	Unsigned	Signed
H	2	12.765	-12.765		19.885	-19.165		27.995	-27.995
CH	118	5.375	1.220	116	7.368	0.522	116	5.772	1.008
C	3	38.977	38.977		37.407	37.407		26.770	26.770
HO	5	11.018	5.078		10.018	0.950		7.288	2.168
CO	3	5.767	4.660		13.527	12.760		14.223	14.223
CHO	61	4.539	-2.286		5.401	-2.576		6.024	-2.226
O	3	8.263	3.050		13.203	-3.617		17.327	-14.933
N	3	64.833	-48.507		57.473	-49.693		60.163	-47.150
HN	6	4.742	0.122		6.387	-0.720		5.728	-2.652
CN	3	9.707	9.707		15.893	0.953		7.683	-0.770
CHN	35	6.182	-0.825		7.258	-2.579		6.456	2.429
NO	10	10.365	-5.821		16.068	6.700		13.226	-2.438
CNO	2	9.005	-9.005		38.660	37.640		17.575	17.575
CHNO	45	5.742	0.582		36.746	35.758		14.945	13.195
HNO	2	4.925	-1.035		18.195	-3.695		13.040	-13.040
S	4	6.643	2.658		11.922	11.543		6.835	-0.750
HS	7	5.140	-0.837		15.104	14.930		5.380	1.326
CS	2	19.620	19.620		23.180	23.180		19.180	8.660
CHS	31	4.459	1.871		7.217	-6.547		4.784	-1.152
SO	3	15.077	-1.597		77.143	77.143		13.140	2.833
CSO	1	10.090	10.090		10.930	10.930		4.870	4.870
CHSO	9	10.690	5.068		95.592	95.592		12.962	4.033
HSO	1	12.430	-12.430		162.570	162.570		10.670	-10.670
CHNS	4	14.465	9.760		14.625	7.365		8.655	-0.460
CNS	1	3.990	-3.990		11.850	-11.850		18.210	-18.210
F	2	25.735	4.035		25.595	25.595		43.465	20.975
HF	1	2.390	2.390		5.400	5.400		9.140	-9.140
CF	13	9.562	-3.618		22.062	0.381		17.625	-6.375
CHF	27	6.261	3.570		6.564	-1.221		9.381	-6.841
OF	3	8.493	-1.913		12.593	9.673		4.670	2.337
HOF	1	8.350	-8.350		2.260	2.260		1.710	-1.710
COF	6	11.832	-8.112		12.318	9.748		9.737	-5.893
CHOF	4	7.295	7.295		8.962	8.962		7.258	5.777
CNF	2	2.710	0.590		8.060	-2.910		7.075	-7.075
NOF	4	8.412	4.092		30.710	26.160		12.732	7.323
SF	10	20.535	5.281		162.532	162.532		28.202	-22.926
SOF	6	7.083	-0.657		149.787	149.787		16.598	-15.825
NF	6	6.113	5.607		14.287	-14.287		10.005	1.678
Cl	3	5.417	-2.303		3.963	-3.157		10.803	1.357
HCl	2	16.305	16.305		27.020	27.020		25.050	22.540
CCl	7	5.297	-5.174		6.571	-2.677		8.793	-8.604
CHCl	17	3.782	3.203		2.952	-1.878		2.945	-1.555
HOCl	1	16.500	-16.500		2.120	2.120		3.980	-3.980
COCl	2	2.305	1.195		0.330	-0.330		2.840	2.400

Table 6.5: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	No.	PM3		No.	MNDO		No.	AM1	
		Unsigned	Signed		Unsigned	Signed		Unsigned	Signed
CHOCI	1	7.630	7.630		2.600	2.600		10.270	10.270
OCI	3	52.027	-52.027		243.607	243.607		191.990	188.297
CNCI	1	0.030	-0.030		1.640	1.640		6.970	-6.970
NOCI	2	11.900	-11.900		15.480	-1.110		9.690	1.980
SCI	7	14.544	-5.641		45.359	6.561		42.119	-26.530
SOCI	6	9.723	5.157		74.258	74.258		21.852	3.178
FCI	3	8.510	2.090		131.903	131.903		68.047	68.047
HFCI	1	5.310	5.310		15.820	15.820		23.610	23.610
CHFCI	3	5.530	5.530		2.693	-1.920		2.177	0.343
COFCI	1	8.370	8.370		9.730	9.730		9.980	9.980
OFCI	1	19.630	19.630		328.340	328.340		251.580	251.580
CFCI	3	1.047	0.993		6.867	6.867		8.340	8.340
Br	4	3.978	0.768		9.103	4.562		13.075	2.870
HBr	2	13.815	13.815		14.560	14.560		7.065	-7.065
CBr	4	8.950	7.605		16.637	-11.942		15.900	-12.610
CHBr	9	6.326	6.326		1.870	-1.250		2.129	1.956
OBr	1	9.280	-9.280		5.270	5.270		5.600	5.600
HOBr	1	13.950	-13.950		2.740	-2.740		4.730	-4.730
COBr	2	18.005	-18.005		20.170	-20.170		13.915	-11.695
CHOBBr	2	2.525	2.525		1.455	1.455		11.660	11.660
CNBr	1	10.320	10.320		3.310	-3.310		10.850	-10.850
NOBr	1	13.040	-13.040		17.790	-17.790		1.640	1.640
SBr	7	21.330	-3.764		37.234	0.994		38.696	-29.707
SOBr	6	10.348	-2.612		50.580	50.580		19.877	-2.667
FBr	3	15.953	11.113		99.827	99.827		90.987	90.987
CFBr	2	4.065	-4.065		12.585	12.585		17.075	17.075
CIBr	2	9.530	-9.530		9.395	-3.595		13.870	-13.870
CCIBr	1	4.790	-4.790		5.830	-5.830		5.410	-5.410
I	2	11.970	-6.150		23.190	23.190		24.590	24.590
HI	1	22.500	22.500		9.440	9.440		1.630	1.630
CI	4	4.285	-3.375		38.013	-34.458		28.520	-24.845
CHI	23	6.925	6.925		6.618	-4.881		3.547	-1.143
OI	1	10.900	-10.900		4.820	4.820		4.880	-4.880
COI	2	38.755	-38.755		38.810	-38.810		30.925	-30.925
CHOI	3	3.577	3.577		2.440	1.353		7.020	7.020
CNI	1	9.820	9.820		14.140	-14.140		11.140	-11.140
NOI	1	8.640	-8.640		5.900	-5.900		5.420	5.420
SI	7	23.399	-21.084		37.767	-23.861		56.807	-56.807
SOI	6	13.080	-10.530		28.090	22.373		17.052	-13.598
FI	3	6.237	4.837		226.633	226.633		194.290	194.290
CFI	1	2.520	2.520		12.270	12.270		7.820	7.820
CII	1	6.210	6.210		11.460	-11.460		9.180	-9.180
Brl	1	5.880	5.880		2.540	-2.540		3.770	-3.770
Be	2	7.795	-7.795		13.445	-13.445		13.445	-13.445
BeH	3	16.983	1.323		25.870	-25.870		21.647	-21.647
BeO	3	9.467	4.160		39.423	39.423		108.250	108.250
BeHO	2	10.680	10.680		13.320	13.320		61.625	61.625
BeF	2	5.320	-2.950		6.570	-6.570		36.475	36.475
BeCl	2	4.600	-4.600		7.280	-7.280		13.815	-13.815

Table 6.6: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	No.	PM3		No.	MNDO		No.	AM1	
		Unsigned	Signed		Unsigned	Signed		Unsigned	Signed
BeBr	2	3.755	-2.425		5.960	5.810		39.810	-39.810
BeI	2	5.110	-4.230		4.965	-3.655		8.545	-8.545
BeOF	1	9.430	9.430		1.800	1.800		125.890	125.890
Mg	4	3.050	-0.055						
MgH	1	1.310	1.310						
CMgH	1	1.130	-1.130						
MgO	1	1.950	-1.950						
MgHO	2	20.645	-12.405						
MgN	1	34.990	34.990						
MgS	1	19.640	-19.640						
MgF	3	9.757	-0.030						
MgCl	3	12.677	-5.077						
MgBr	3	6.733	-2.167						
MgI	1	3.480	-3.480						
Al	3	43.607	-2.507		13.057	-3.097		2.357	-0.823
HAl	1	8.060	8.060		16.030	-16.030		9.110	-9.110
CHAl	2	9.985	9.985		17.960	-17.960		6.850	-6.850
AlO	9	19.863	-11.206		25.150	-16.192		10.878	4.160
HAIO	5	16.366	-3.906		14.070	-2.030		17.100	11.544
AlN	1	49.550	-49.550		15.910	15.910		7.090	-7.090
AlS	1	25.700	-25.700		4.520	-4.520		8.260	-8.260
AlF	8	6.566	1.056		16.358	-5.665		16.690	-2.295
AlOF	3	32.110	32.110		47.183	47.183		52.113	52.113
AlCl	7	10.541	-0.890		12.701	-7.470		9.584	-9.584
AlOCl	1	10.780	10.780		14.510	14.510		16.760	16.760
AlFCl	4	6.285	6.285		3.872	-3.872		3.425	-0.970
AlBr	2	6.585	5.665		64.565	64.565		13.305	-11.665
AlI	3	13.247	13.007		65.703	65.703		12.740	12.740
Ga	6	9.708	-3.565						
GaH	1	14.840	14.840						
CGaH	3	24.837	-1.757						
GaO	1	31.470	-31.470						
GaHO	1	4.600	-4.600						
GaF	1	14.040	14.040						
GaCl	3	16.087	2.147						
GaHNCl	1	21.960	21.960						
GaBr	1	19.860	-19.860						
Gal	2	9.890	-9.890						
GaS	1	35.200	35.200						
In	5	11.656	5.184						
InH	1	0.920	0.920						
ClInH	1	42.950	-42.950						
InO	1	39.530	-39.530						
InHO	1	4.800	-4.800						
InS	2	3.835	1.175						
InF	1	13.740	13.740						
InCl	4	10.530	0.115						

Table 6.7: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	No.	PM3		No.	MNDO		No.	AM1	
		Unsigned	Signed		Unsigned	Signed		Unsigned	Signed
InBr	2	5.700	1.790						
InI	2	9.785	0.785						
TI	4	5.697	4.932						
TICl	3	2.043	2.043						
TIBr	2	10.810	0.670						
TII	1	14.370	-14.370						
Si	3	1.910	-1.537		31.283	31.283		11.360	10.427
HSi	9	8.008	-2.457		12.311	-1.613		10.784	-8.509
CHSi	14	7.901	-1.279		10.875	-7.991		10.564	-4.304
SiO	2	9.030	-9.030		25.660	25.660		13.665	13.665
CHSiO	2	2.630	2.630		5.675	-5.675		13.440	13.440
SiF	5	11.780	-10.688		18.246	-4.974		14.322	-11.290
HSiF	3	9.243	9.243		5.803	-4.530		4.973	4.100
CHSiF	1	25.020	25.020		33.750	33.750		24.070	24.070
SiOF	2	15.090	15.090		49.210	49.210		20.190	20.190
SiCl	6	12.740	-7.500		15.953	-2.107		18.155	-17.528
HSiCl	3	5.713	5.713		7.050	-6.103		5.533	-5.533
CHSiCl	5	3.872	3.200		1.956	-1.732		4.112	-4.112
SiOCl	2	37.070	37.070		58.860	58.860		38.960	38.960
SiFCl	1	0.080	0.080		17.460	17.460		7.960	7.960
SiBr	5	11.658	-4.418		47.098	47.098		11.994	6.946
HSiBr	3	3.923	-3.923		14.653	12.820		3.573	-2.060
CHSiBr	1	1.220	1.220		7.600	7.600		3.510	3.510
SiOBr	2	33.775	33.775		64.355	64.355		45.535	45.535
SiI	4	7.035	2.325		47.308	47.308		19.867	19.867
HSiI	3	5.223	5.223		31.810	31.810		13.717	13.717
SiOI	2	34.380	14.790		72.915	72.915		78.720	78.720
CHSiN	1	7.020	-7.020		9.800	-9.800		4.170	-4.170
Ge	2	22.760	-22.760		38.790	38.790		15.170	-15.170
GeH	3	7.113	7.113		12.460	6.960		6.707	-1.613
CGe	1	27.610	27.610		73.430	73.430		52.550	52.550
CGeH	8	10.894	-3.679		15.488	-1.718		15.783	-2.757
GeO	1	5.120	5.120		11.850	11.850		20.240	20.240
CGeHO	3	7.850	-4.903		4.210	1.657		13.433	-3.007
CGeHN	1	12.420	-12.420		3.990	-3.990		1.130	1.130
GeF	3	7.977	4.077		7.130	-0.370		11.073	3.247
GeCl	4	13.197	-8.098		17.237	3.772		10.980	-8.845
CGeHCl	1	3.650	3.650		1.700	1.700		5.110	5.110
GeBr	3	9.987	0.640		22.470	13.550		6.930	-1.777
CGeHBr	1	5.270	5.270		14.820	14.820		12.030	12.030
GeI	2	11.845	11.845		59.935	59.935		19.435	19.435
Sn	1	0.000	0.000		0.000	0.000			
SnH	2	8.835	5.415		4.555	-3.845			
CSnH	10	9.040	-6.168		12.578	-9.052			
SnO	1	0.920	0.920		9.610	9.610			
CSnHN	1	4.900	-4.900		7.030	7.030			
SnS	1	9.230	9.230		1.420	-1.420			

Table 6.8: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	No.	PM3		No.	MNDO		No.	AM1	
		Unsigned	Signed		Unsigned	Signed		Unsigned	Signed
SnF	2	6.520	-1.960		6.395	-5.035			
SnCl	3	6.927	1.307		19.990	-11.543			
CSnHCl	5	12.102	10.402		7.510	3.398			
SnBr	3	13.710	-3.757		20.573	9.207			
CSnHBr	2	6.270	-6.270		4.645	4.645			
SnI	2	11.900	-11.900		1.915	1.915			
CSnHI	1	0.830	-0.830		10.860	10.860			
CGeSnH	1	7.480	7.480		10.050	10.050			
Pb	3	11.753	-11.753		2.530	-2.530			
PbH	2	1.605	-1.605		9.580	-6.190			
CPbH	6	10.560	3.813		11.782	0.872			
PbO	1	4.510	4.510		12.310	12.310			
PbS	1	8.400	8.400		0.460	0.460			
PbF	2	8.065	6.285		6.615	3.215			
PbCl	5	6.108	-2.436		19.620	-11.944			
PbBr	2	7.875	-7.875		9.925	-9.925			
PbI	2	14.375	-14.375		4.920	4.920			
P	3	54.893	-47.047		23.133	-19.727	33.507	-20.653	
HP	4	5.857	0.557		9.925	5.985	7.393	7.393	
CP	1	12.090	12.090		22.120	22.120	23.580	23.580	
CHP	7	7.393	-3.799		19.817	-17.989	7.740	2.511	
PO	5	13.260	-6.000		72.368	65.104	50.902	37.394	
CHPO	6	23.477	14.377		60.502	40.432	9.535	1.468	
HPO	1	42.270	-42.270		29.840	29.840	33.620	-33.620	
NP	1	7.860	7.860		8.870	8.870	7.460	7.460	
PF	7	16.131	-12.180		55.154	49.634	9.651	5.583	
POF	3	15.017	1.430		51.350	49.737	14.397	12.190	
CHPOF	9	5.136	-2.749		61.624	61.624	15.476	-14.242	
PCI	5	17.888	-16.560		26.640	-4.672	18.896	-13.188	
POCI	3	11.387	-1.353		32.397	25.657	13.340	13.067	
CHPOCI	1	4.790	-4.790		48.660	48.660	9.270	9.270	
PSCI	1	30.770	30.770		62.610	62.610	58.690	58.690	
PBr	5	10.678	-7.986		19.434	-2.338	22.112	8.096	
POBr	3	13.797	7.330		35.370	32.903	39.183	39.183	
PSBr	1	59.360	59.360		85.630	85.630	111.940	111.940	
PI	5	6.446	-3.242		11.790	-11.318	17.218	-16.930	
POI	3	15.517	8.697		33.197	33.197	42.190	42.190	
As	5	5.012	-2.380						
AsH	1	3.210	-3.210						
CAsH	3	20.550	-13.810						
CAsHO	3	4.097	-1.397						
AsN	1	23.230	23.230						
AsS	1	2.100	-2.100						
AsF	2	5.085	2.685						
AsCl	2	6.445	-1.265						
AsBr	1	11.300	11.300						
AsI	1	20.380	20.380						

Table 6.9: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	No.	PM3		No.	MNDO		No.	AM1	
		Unsigned	Signed		Unsigned	Signed		Unsigned	Signed
AsO	1	1.110	1.110						
Zn	3	5.487	-5.487		21.647	15.093		23.183	19.710
ZnH	1	2.480	2.480		11.240	-11.240		3.450	-3.450
CZnH	6	6.442	2.768		11.097	11.097		6.045	6.045
ZnS	1	3.180	3.180		30.710	30.710		30.040	30.040
ZnCl	2	5.430	5.430		21.735	21.735		10.370	10.370
Cd	1	0.000	0.000						
CCdH	3	3.400	0.027						
Hg	5	6.276	-0.396		15.554	-13.866		12.496	-11.680
HgH	1	9.140	-9.140		19.420	-19.420		1.270	-1.270
CHgH	9	4.232	1.359		8.867	0.782		4.314	2.794
CHgN	1	2.080	2.080		21.940	-21.940		6.770	-6.770
HgF	2	13.695	5.455		14.400	-11.070		13.845	2.165
HgCl	2	8.760	-6.460		14.185	-14.185		16.320	-16.320
CHgHCl	2	7.125	7.125		3.635	-3.635		0.720	-0.720
HgBr	2	10.365	-10.365		16.805	6.685		25.975	-25.975
CHgHBr	2	1.945	1.375		9.155	9.155		7.175	-7.175
CHgHI	2	2.370	1.910		9.690	9.690		14.870	14.870
HgI	1	0.130	-0.130		25.210	25.210		23.200	23.200
Sb	5	10.850	-0.814						
SbH	1	23.370	23.370						
CSbH	3	15.770	-15.770						
SbO	1	11.770	-11.770						
SbN	1	38.820	38.820						
SbF	1	10.620	10.620						
SbCl	4	10.110	2.200						
SbOCl	1	18.390	18.390						
SbBr	1	9.990	9.990						
InSb	2	9.325	-9.325						
Se	1	0.000	0.000						
SeH	1	15.620	15.620						
CSeH	1	23.690	-23.690						
SeO	2	14.880	14.880						
SeOF	1	12.650	-12.650						
SeF	2	7.860	-7.630						
CSe	1	3.360	3.360						
AsTe	1	6.390	6.390						
ZnTe	1	12.740	-12.740						
Te	1	13.310	13.310						
TeO	1	18.470	-18.470						
TeF	1	10.580	10.580						
Bi	4	9.690	1.055						
CBiH	3	10.690	-10.690						
BiF	1	14.100	14.100						
BiCl	2	16.315	16.315						
BiSe	1	12.550	12.550						
CLiH				2	24.470	-24.470			

Table 6.10: Average Signed and Unsigned Errors in ΔH_f by Compound Type (contd.)

Elements in compounds	PM3		MNDO		AM1			
	No.	Unsigned	Signed	No.	Unsigned	Signed		
BH				7	28.837	3.649	33.640	12.311
CBH				4	9.922	-9.922	5.205	-2.965
BO				5	14.142	-11.806	12.074	-11.682
BHO				9	15.498	13.020	9.312	0.948
CBHO				5	11.010	1.794	7.962	6.702
BN				1	25.290	25.290	18.010	18.010
CBHN				5	20.188	20.188	20.858	20.674
BS				3	21.453	-12.140	12.273	0.713
CBHS				2	4.620	-3.830	4.950	4.950
BF				3	16.427	-6.187	12.660	-12.660
BOF				3	19.023	19.023	5.063	2.230
BHOF				4	18.845	18.845	5.213	0.653
CBHF				2	5.965	5.965	6.430	6.430
CBHOF				2	18.405	18.405	2.325	-2.165
CBHNF				1	40.180	40.180	32.120	32.120
BCI				3	14.110	-3.543	8.630	-8.630
BOCI				2	27.325	10.285	24.580	0.290
CBHOCl				1	12.290	12.290	12.750	12.750
BFCI				2	10.530	10.530	4.635	4.635
BBr				2	19.350	-3.870	10.325	-10.325
BFBBr				1	12.820	12.820	5.910	5.910
BI				1	8.390	-8.390	2.410	2.410
B				1	5.230	-5.230	8.750	8.750
BHN				1	8.920	-8.920	19.970	-19.970
BHOCl				2	13.655	13.655	0.670	-0.520
BHNCI				1	12.660	12.660	20.750	-20.750
CBHP				1	20.220	-20.220	4.150	-4.150
BHPF				1	13.160	-13.160	30.500	-30.500
CBHPF				1	2.340	2.340	30.850	30.850

Table 6.11: Average Errors in ΔH_f by Element

Element	PM3			MNDO			AM1		
	No.	Signed	RMS	No.	Signed	RMS	No.	Signed	RMS
H	613	6.951	9.711	630	13.854	24.420	598	8.498	12.746
Li				2	24.470	24.495			
Be	19	8.595	11.450	19	15.829	23.996	19	45.422	66.926
B				76	16.544	21.183	76	12.833	18.902
C	595	6.962	10.153	597	14.016	24.367	569	8.578	12.354
N	140	9.032	18.044	145	21.136	30.283	144	12.396	19.821
O	305	10.424	15.457	321	32.881	62.721	319	17.858	40.932
F	177	9.673	13.600	184	39.491	83.527	180	20.997	45.279
Mg	21	9.690	13.949						
Al	50	16.124	23.016	50	22.516	32.833	50	13.617	19.861
Si	79	10.116	14.262	79	22.510	33.132	79	14.730	21.824
P	74	14.939	23.657	77	38.352	55.538	77	21.186	35.850
S	129	11.851	17.296	129	48.854	88.441	127	19.129	29.126
Cl	156	10.030	14.834	144	26.118	69.410	131	19.757	53.791
Zn	14	6.026	8.035	13	16.688	22.445	13	12.312	20.774
Ga	21	16.322	19.623						
Ge	34	10.867	12.944	34	18.969	28.055	33	13.481	17.554
As	22	8.626	11.964						
Se	10	11.335	14.110						
Br	98	10.799	14.273	91	24.829	41.901	84	20.173	34.835
Cd	4	2.550	3.482						
In	22	11.801	16.725						
Sn	35	8.664	10.137	35	10.257	13.429			
Sb	20	13.681	17.333						
Te	5	12.298	12.910						
I	96	10.592	15.518	90	28.916	59.613	85	24.643	52.386
Hg	29	5.839	7.653	29	12.410	15.481	29	10.012	13.547
Tl	10	6.491	9.108						
Pb	24	8.580	11.488	24	10.468	13.194			
Bi	11	11.828	14.421						
Totals	1116	9.5	15.0	1050	22.4	48.9	989	14.8	32.2

Table 6.12: Average Errors in Bond-Length by Bond-Type

Elements		PM3			MNDO			AM1		
		No.	Unsigned	Signed	No.	Unsigned	Signed	No.	Unsigned	Signed
H	H	1	0.042	-0.042	1	0.078	-0.078	1	0.064	-0.064
H	Be	2	0.038	-0.038	2	0.059	-0.059	2	0.039	-0.039
H	B				3	0.042	-0.042	3	0.016	-0.006
H	C	68	0.010	0.002	67	0.011	0.006	66	0.015	0.013
H	N	9	0.012	-0.010	8	0.063	0.056	8	0.046	0.029
H	O	8	0.014	-0.013	8	0.014	-0.011	8	0.013	0.008
H	F	3	0.055	-0.029	3	0.425	0.425	3	0.205	0.082
H	Mg	1	0.043	-0.043						
H	Al	1	0.015	0.015	1	0.222	-0.222	1	0.186	-0.186
H	Si	12	0.016	0.011	10	0.112	-0.112	10	0.029	-0.029
H	P	3	0.114	-0.114	3	0.097	-0.009	3	0.091	0.014
H	S	3	0.026	-0.026	3	0.026	-0.026	3	0.005	-0.005
H	Cl	1	0.007	-0.007	1	0.073	0.073	1	0.009	0.009
H	Ga	2	0.027	0.012						
H	Ge	17	0.020	-0.016	17	0.041	-0.041	17	0.019	0.019
H	As	1	0.007	0.007						
H	Se	3	0.013	0.011						
H	Br	1	0.056	0.056	1	0.025	0.025	1	0.006	0.006
H	In	1	0.104	-0.104						
H	Sn	5	0.022	-0.014	5	0.125	-0.125			
H	Sb	1	0.005	-0.005						
H	Te	1	0.017	0.017						
H	I	1	0.068	0.068	1	0.042	-0.042	1	0.022	-0.022
H	Hg	1	0.044	-0.044	1	0.190	-0.190	1	0.069	-0.069
H	Pb	1	0.110	-0.110	1	0.181	-0.181			
Be	O	1	0.027	-0.027	1	0.004	0.004	1	0.071	0.071
Be	F	2	0.007	0.007	2	0.079	0.079	2	0.106	0.106
Be	S	1	0.090	-0.090	1	0.118	-0.118	1	0.057	-0.057
Be	Cl	2	0.047	-0.047	2	0.130	0.130	2	0.042	0.042
Be	Br	1	0.096	-0.096	1	0.102	0.102	1	0.013	-0.013
Be	I	2	0.071	0.019	2	0.012	0.012	2	0.012	-0.007
B	B				5	0.035	0.035	5	0.025	-0.015
B	C				5	0.033	0.008	5	0.016	0.001
B	O				2	0.021	-0.021	2	0.028	-0.028
B	F				1	0.007	-0.007	1	0.020	-0.020
B	P				1	2.386	2.386	1	0.321	0.321
B	S				1	0.118	-0.118	1	0.131	-0.131
C	C	91	0.017	-0.009	90	0.018	0.009	90	0.018	-0.006
C	N	30	0.020	0.015	28	0.023	0.006	28	0.027	-0.003
C	O	31	0.015	0.007	31	0.023	0.017	31	0.029	0.027
C	F	25	0.017	0.009	24	0.019	0.007	24	0.030	0.026
C	Si	4	0.012	0.011	4	0.054	-0.054	4	0.052	-0.052
C	P	6	0.076	-0.064	6	0.101	-0.101	6	0.108	-0.108
C	S	13	0.044	-0.040	12	0.068	-0.068	12	0.076	-0.076
C	Cl	12	0.040	-0.015	12	0.034	0.034	12	0.033	-0.008
C	Zn	5	0.030	0.030	5	0.037	-0.037	5	0.035	-0.004
C	Ge	14	0.024	0.011	14	0.019	-0.007	14	0.048	0.042

Table 6.13: Average Errors in Bond-Length by Bond-Type (contd.)

Elements		PM3			MNDO			AM1		
		No.	Unsigned	Signed	No.	Unsigned	Signed	No.	Unsigned	Signed
C	As	2	0.029	-0.002						
C	Se	11	0.035	-0.020						
C	Br	10	0.016	0.001	10	0.055	-0.049	10	0.040	-0.013
C	Cd	1	0.077	-0.077						
C	Sn	6	0.027	0.008	6	0.078	-0.078			
C	Sb	1	0.007	0.007						
C	Te	1	0.027	-0.027						
C	I	3	0.089	-0.089	3	0.076	-0.076	3	0.068	-0.038
C	Hg	6	0.035	0.016	6	0.102	-0.048	6	0.036	0.001
C	Tl	1	0.007	0.007						
C	Pb	2	0.052	-0.052	2	0.071	-0.071			
C	Bi	1	0.004	-0.004						
N	N	8	0.056	-0.052	8	0.101	-0.098	8	0.093	-0.073
N	O	19	0.028	-0.009	19	0.053	-0.007	19	0.053	-0.017
N	F	2	0.085	-0.085	2	0.136	-0.136	2	0.082	-0.082
N	Si	1	0.108	-0.108	1	0.092	-0.092	1	0.087	-0.087
N	P	1	0.077	-0.077	1	0.093	-0.093	1	0.109	-0.109
N	S	2	0.065	0.020	2	0.058	0.003	2	0.053	-0.053
N	Cl	2	0.099	-0.099	2	0.094	-0.094	2	0.139	-0.139
N	Ga	2	0.294	0.064						
N	Ge	3	0.020	-0.003	3	0.013	0.012	3	0.016	-0.016
N	Se	1	0.053	0.053						
N	Br	1	0.252	-0.252	1	0.271	-0.271	1	0.218	-0.218
N	Cd	1	0.022	-0.022						
N	Te	1	0.045	-0.045						
O	O	3	0.026	-0.005	3	0.090	-0.090	3	0.115	-0.115
O	F	1	0.034	-0.034	1	0.131	-0.131	1	0.058	-0.058
O	Mg	1	0.031	0.031						
O	Al	2	0.028	-0.025	2	0.133	-0.133	2	0.067	-0.067
O	P	5	0.043	0.017	5	0.032	-0.007	5	0.018	0.002
O	S	11	0.039	0.023	11	0.070	0.070	11	0.042	-0.006
O	Cl	3	0.003	-0.002	3	0.120	0.108	3	0.150	0.150
O	Ga	1	0.028	-0.028						
O	Ge	2	0.014	0.014	2	0.035	-0.035	2	0.040	0.038
O	Se	7	0.050	-0.017						
O	In	1	0.021	-0.021						
O	Sn	1	0.006	0.006	1	0.084	-0.084			
O	Te	2	0.093	-0.093						
O	Hg	1	0.001	0.001	1	0.042	0.042	1	0.222	0.222
O	Pb	1	0.016	0.016	1	0.038	-0.038			
F	F	1	0.062	-0.062	1	0.146	-0.146	1	0.015	0.015
F	Mg	2	0.006	-0.002						
F	Al	3	0.006	0.004	3	0.059	-0.059	3	0.073	-0.073
F	Si	5	0.017	0.011	5	0.021	0.014	5	0.038	0.038
F	P	9	0.018	-0.014	8	0.031	0.000	8	0.027	-0.026
F	S	13	0.027	-0.010	13	0.235	0.222	13	0.037	-0.033
F	Cl	4	0.051	0.015	4	0.057	0.057	4	0.043	0.035
F	Zn	1	0.002	-0.002	1	0.062	-0.062	1	0.006	0.006
F	Ga	2	0.010	-0.002						
F	Ge	7	0.016	-0.011	7	0.020	0.002	7	0.035	-0.031

Table 6.14: Average Errors in Bond-Length by Bond-Type (contd.)

Elements		PM3			MNDO			AM1		
		No.	Unsigned	Signed	No.	Unsigned	Signed	No.	Unsigned	Signed
F	As	3	0.010	-0.007						
F	Se	8	0.020	0.010						
F	Br	5	0.033	0.025	5	0.037	0.006	5	0.055	0.055
F	Cd	1	0.004	-0.004						
F	In	1	0.001	0.001						
F	Sb	2	0.402	-0.402						
F	Te	4	0.007	0.001						
F	I	5	0.226	0.219	5	0.236	0.234	5	0.224	0.214
F	Hg	2	0.021	-0.011	2	0.050	-0.050	2	0.028	-0.024
F	Tl	2	0.027	0.022						
F	Pb	2	0.018	-0.018	2	0.050	-0.050			
Mg	Mg	1	3.110	3.110						
Mg	S	1	0.215	0.215						
Mg	Cl	2	0.311	-0.311						
Mg	Br	2	0.010	0.004						
Mg	I	1	0.100	-0.100						
Al	Al	1	0.073	0.073	1	0.175	-0.175	1	0.060	-0.060
Al	Cl	2	0.138	-0.138	2	0.030	-0.025	2	0.240	-0.240
Al	Br	2	0.199	-0.199	2	0.094	-0.094	2	0.028	-0.028
Al	I	1	0.012	-0.012	1	0.174	-0.174	1	0.111	-0.111
Si	Si	3	0.051	0.051	3	0.202	-0.202	3	0.196	-0.109
Si	S	1	0.105	0.105	1	0.070	-0.070	1	0.176	0.176
Si	Cl	3	0.060	-0.018	3	0.046	0.046	3	0.045	-0.007
Si	Ge	1	0.047	0.047	1	0.017	-0.017	1	0.002	0.002
Si	As	1	0.018	0.018						
Si	Se	2	0.103	0.020						
Si	Br	2	0.318	-0.318	2	0.051	0.051	2	0.116	0.116
Si	Sb	1	0.022	-0.022						
Si	I	2	0.231	-0.194	2	0.074	-0.074	2	0.004	-0.004
P	P	3	0.084	-0.084	3	0.189	-0.189	3	0.243	-0.243
P	S	5	0.049	0.022	3	0.104	-0.033	3	0.008	-0.005
P	Cl	3	0.045	-0.020	3	0.045	-0.045	3	0.101	-0.101
P	Ge	2	0.090	-0.090	2	0.150	-0.150	2	0.185	-0.185
P	Se	1	0.062	0.062						
P	Br	3	0.077	-0.077	3	0.085	-0.085	3	0.058	-0.058
S	S	9	0.115	0.084	9	0.454	0.325	9	0.264	0.234
S	Cl	5	0.042	0.036	5	0.064	-0.002	5	0.075	-0.015
S	Ge	2	0.035	-0.004	2	0.079	-0.079	2	0.039	0.039
S	Br	4	0.166	0.166	4	0.129	0.009	4	0.166	0.118
S	Cd	1	0.001	-0.001						
S	Sn	1	0.071	-0.071	1	0.208	-0.208			
S	Te	1	0.234	0.234						
S	Pb	1	0.138	-0.138	1	0.179	-0.179			
Cl	Cl	1	0.049	0.049	1	0.010	0.010	1	0.068	-0.068
Cl	Zn	1	0.002	0.002	1	0.053	0.053	1	0.005	0.005
Cl	Ga	6	0.119	0.049						
Cl	Ge	8	0.061	0.012	8	0.102	0.102	8	0.035	0.001
Cl	As	1	0.002	0.002						
Cl	Se	2	0.006	0.001						
Cl	Br	1	0.040	0.040	1	0.056	-0.056	1	0.072	-0.072

Table 6.15: Average Errors in Bond-Length by Bond-Type (contd.)

Elements		PM3			MNDO			AM1		
		No.	Unsigned	Signed	No.	Unsigned	Signed	No.	Unsigned	Signed
Cl	Cd	1	0.015	0.015						
Cl	In	1	0.002	0.002						
Cl	Sn	4	0.047	0.044	4	0.029	-0.027			
Cl	Sb	3	0.037	0.035						
Cl	Te	1	0.072	0.072						
Cl	I	1	0.135	-0.135	1	0.065	-0.065	1	0.109	-0.109
Cl	Hg	3	0.025	-0.025	3	0.024	0.024	3	0.018	-0.011
Cl	Tl	1	0.004	0.004						
Cl	Pb	3	0.124	0.021	3	0.104	0.028			
Cl	Bi	1	0.006	-0.006						
Zn	Br	1	0.107	-0.107	1	0.031	0.031	1	0.093	-0.093
Zn	I	1	0.003	0.003	1	0.019	-0.019	1	0.055	-0.055
Ga	Ga	1	0.583	0.583						
Ga	Br	4	0.031	0.014						
Ga	I	2	0.077	0.041						
Ge	Ge	1	0.010	-0.010	1	0.121	0.121	1	0.037	-0.037
Ge	Se	1	0.208	-0.208						
Ge	Br	5	0.037	0.031	5	0.056	0.049	5	0.053	-0.053
Ge	Te	1	0.338	-0.338						
Ge	I	2	0.034	-0.034	2	0.013	0.010	2	0.083	-0.083
As	Br	1	0.008	-0.008						
As	I	1	0.041	-0.041						
Se	Se	1	0.040	0.040						
Se	In	1	0.070	-0.070						
Se	Sn	1	0.002	-0.002						
Se	Pb	1	0.041	-0.041						
Br	Br	1	0.160	0.160	1	0.115	-0.115	1	0.099	-0.099
Br	Cd	1	0.034	-0.034						
Br	In	1	0.253	-0.253						
Br	Sn	4	0.041	-0.041	4	0.087	-0.087			
Br	Sb	1	0.019	-0.019						
Br	Te	2	0.075	-0.075						
Br	I	1	0.076	0.076	1	0.135	-0.135	1	0.131	-0.131
Br	Hg	3	0.148	-0.148	3	0.045	-0.020	3	0.146	-0.146
Br	Tl	1	0.059	-0.059						
Br	Pb	2	0.025	-0.007	2	0.103	-0.103			
Br	Bi	1	0.026	-0.026						
Cd	I	1	0.038	0.038						
In	Te	1	0.004	-0.004						
In	I	2	0.012	-0.008						
Sn	Te	1	0.049	0.049						
Sn	I	3	0.055	-0.055	3	0.172	-0.172			
Sb	Sb	1	0.289	-0.289						
Te	Te	1	0.145	0.145						
Te	Pb	1	0.141	0.141						
I	I	1	0.002	0.002	1	0.151	-0.151	1	0.128	-0.128
I	Hg	3	3.441	3.330	3	0.071	-0.071	3	0.063	-0.039
I	Tl	1	0.090	-0.090						
I	Pb	2	0.031	0.006	2	0.181	-0.181			
Pb	Pb	1	0.012	-0.012	1	0.110	-0.110			

Table 6.16: Average Errors in Bond-Lengths for each Element

Element	No.	PM3	No.	MNDO	No.	AM1
Hydrogen	147	0.019	133	0.048	126	0.027
Beryllium	11	0.049	11	0.071	11	0.049
Carbon	344	0.021	320	0.027	311	0.028
Nitrogen	83	0.041	75	0.054	75	0.052
Oxygen	101	0.026	89	0.045	87	0.045
Fluorine	115	0.038	88	0.091	86	0.055
Magnesium	11	0.378				
Aluminum	12	0.071	12	0.105	12	0.104
Silicon	38	0.059	32	0.082	32	0.060
Phosphorus	41	0.057	37	0.080	37	0.079
Sulfur	74	0.063	68	0.159	66	0.087
Chlorine	78	0.060	59	0.062	52	0.063
Zinc	9	0.029	9	0.039	9	0.037
Gallium	20	0.113				
Germanium	66	0.037	64	0.046	64	0.040
Arsenic	10	0.017				
Selenium	40	0.041				
Bromine	61	0.075	47	0.074	41	0.077
Cadmium	7	0.027				
Indium	9	0.053				
Tin	26	0.035	24	0.099		
Antimony	10	0.126				
Tellurium	18	0.080				
Iodine	36	0.371	28	0.117	23	0.099
Mercury	19	0.586	19	0.072	19	0.065
Thallium	6	0.036				
Lead	17	0.064	15	0.109		
Bismuth	3	0.012				
TOTALS:	771	0.058	627	0.062	587	0.048

Table 6.17: Worst Ionization Potential Errors in MNDO

Formula	Chemical Name	Exp. I.P.	Calc. I.P.	Diff.
H ₂ O	Water, fourth I.P.	32.20	40.03	7.8
CCl ₃	Trichloromethyl	8.28	13.09	4.8
SnF ₂	Tin difluoride	8.00	12.60	4.6
NS	Sulfur nitride	8.87	13.26	4.4
CF ₃ Cl	Trifluorochloromethane	10.30	14.13	3.8
SOCl ₃	SOCl ₃	9.63	13.40	3.8
SnH ₄	Tin tetrahydride (stannane)	9.20	12.74	3.5
PF ₃	Phosphorus trifluoride	9.71	13.13	3.4
PS	Phosphorus sulfide	9.00	12.05	3.1
C ₃ H ₇ NO ₂	Alanine	8.10	10.82	2.7
C ₅ H ₁₀	2-Methyl-1-butene	7.40	9.81	2.4
C ₅ H ₈ O ₂	Acetylacetone	8.38	10.78	2.4
PbO ₂	Lead dioxide	8.87	11.06	2.2
SO ₃	Sulfur trioxide	11.00	13.16	2.2
C ₅ H ₁₀	1-Pentene	7.90	9.97	2.1
C ₄ GeH ₁₂	Tetramethylgermanium	9.29	11.36	2.1
CF ₃ I	Trifluoroiodomethane	10.45	12.48	2.0
SiCl ₄	Silicon tetrachloride	11.79	13.82	2.0
C ₄ SnH ₁₂	Tetramethyltin	8.90	10.91	2.0
GeO	Germanium oxide	11.25	9.26	-2.0
P ₄	Phosphorus tetramer	9.54	11.50	2.0
C ₈ H ₁₄	Bicyclo(2.2.2)-octane	9.45	11.40	2.0
GeCl ₄	Germanium tetrachloride	11.88	13.78	1.9
C ₂ Cl ₆	Hexachloroethane	11.20	13.08	1.9
C ₆ H ₁₁ I	Iodocyclohexane	8.91	10.77	1.9
C ₅ H ₁₂	n-Pentane	10.30	12.15	1.8
GeI ₄	Germanium tetraiodide	9.42	11.26	1.8
CHgH ₃ Br	Methylmercuric bromide	9.25	11.09	1.8
C ₈ PbH ₂₀	Tetraethyllead	8.13	9.87	1.7
C ₆ Pb ₂ H ₁₈	Hexamethyldiplumbane	7.41	9.15	1.7
CBr ₄	Carbon tetrabromide	10.30	12.03	1.7
C ₅ H ₁₂	2-Methylbutane	10.30	12.00	1.7
C ₆ GeSnH ₁₈	GeMe ₃ -SnMe ₃	8.20	9.88	1.7
C ₁₀ H ₁₆	Adamantane	9.60	11.27	1.7
C ₂ F ₄ Br ₂	1,2-Dibromotetrafluoroethane	14.44	12.77	-1.7
C ₈ GeH ₂₀	Tetraethylgermanium	9.30	10.97	1.7
PCl ₃	Phosphorus trichloride	10.50	12.17	1.7
S ₂ Cl ₂	ClSSCl	9.40	11.06	1.7
I ₂	Iodine, second I.P.	11.03	12.69	1.7
C ₇ GeH ₁₈	t-Butyltrimethylgermane	8.98	10.64	1.7
SnCl ₂	Tin dichloride	10.10	11.72	1.6
NO ₂ Cl	Nitryl chloride	11.40	13.01	1.6
C ₄ H ₉ I	1-Butyl iodide	9.20	10.80	1.6
C ₇ PbH ₁₈	t-Butyltrimethyllead	7.99	9.60	1.6
Be	Beryllium, atom	9.20	7.60	-1.6
C ₃ H ₉ PO ₃	Trimethyl phosphite	9.00	10.60	1.6
C ₄ ZnH ₁₀	Diethylzinc	8.60	10.20	1.6
C ₈ SnH ₂₀	Tetraethyltin	8.90	10.47	1.6
C ₅ H ₁₀	Cyclopentane	10.50	12.06	1.6
PbCl ₂	Lead dichloride	10.34	11.90	1.6

Table 6.18: Worst Ionization Potential Errors in AM1

Formula	Chemical Name	Exp. I.P.	Calc. I.P.	Diff.
H ₂ O	Water, fourth I.P.	32.20	36.42	4.2
CCl ₃	Trichloromethyl	8.28	12.24	4.0
SOCl ₃	SOCl ₃	9.63	13.36	3.7
SO ₃	Sulfur trioxide	11.00	14.26	3.3
NS	Sulfur nitride	8.87	11.96	3.1
PS	Phosphorus sulfide	9.00	11.95	2.9
CF ₃ Cl	Trifluorochloromethane	10.30	13.24	2.9
C ₂ F ₄ Br ₂	1,2-Dibromotetrafluoroethane	14.44	11.50	-2.9
C ₅ H ₈ O ₂	Acetylacetone	8.38	10.77	2.4
C ₅ H ₁₀	2-Methyl-1-butene	7.40	9.70	2.3
C ₃ H ₇ NO ₂	Alanine	8.10	10.37	2.3
P ₄	Phosphorus tetramer	9.54	11.68	2.1
SF ₄	Sulfur tetrafluoride	12.05	9.92	-2.1
C ₅ H ₁₀	1-Pentene	7.90	10.00	2.1
HF	Hydrogen fluoride	16.06	14.09	-2.0
GeI ₄	Germanium tetraiodide	9.42	11.33	1.9
PCl ₅	Phosphorus pentachloride	10.80	12.70	1.9
GeO	Germanium oxide	11.25	9.40	-1.8
SO ₂	Sulfur dioxide	12.30	10.49	-1.8
SOF ₂	Thionyl fluoride	12.58	10.80	-1.8
C ₆ Ge ₂ H ₁₈	GeMe ₃ -GeMe ₃	8.18	9.92	1.7
C ₄ GeH ₁₂	Tetramethylgermanium	9.29	10.96	1.7
Be	Beryllium, atom	9.20	7.60	-1.6
NO ₂ Cl	Nitryl chloride	11.40	12.94	1.5
CF ₃ I	Trifluoroiodomethane	10.45	11.97	1.5
CHF ₃	Trifluoromethane	14.80	13.32	-1.5
C ₆ H ₁₁ I	Iodocyclohexane	8.91	10.38	1.5
C ₂ I ₂	Diiodoacetylene	9.03	10.50	1.5
GeF ₄	Germanium tetrafluoride	16.06	14.58	-1.5
I ₂	Iodine, second I.P.	11.03	12.48	1.5
SO ₂ F ₂	Sulfuryl fluoride	13.04	14.41	1.4
C ₄ H ₄ N ₂	Pyridazine	9.30	10.67	1.4
C ₂ F ₆	Hexafluoroethane	14.60	13.23	-1.4
N ₂ O ₅	Dinitrogen pentoxide	12.30	13.64	1.3
I ₂	Iodine	9.34	10.67	1.3
CHgH ₃ Br	Methylmercuric bromide	9.25	10.59	1.3
N ₂	Nitrogen	15.60	14.32	-1.3
POCl ₃	Phosphorus oxychloride	11.85	13.12	1.3
H ₂ SiF ₂	Difluorosilane	12.85	11.61	-1.2
C ₄ H ₉ I	1-Butyl iodide	9.20	10.43	1.2
CH ₃ F	Fluoromethane	13.31	12.10	-1.2
C ₂ H ₄ I ₂	1,2-Diiodoethane	9.50	10.71	1.2
SF ₂	Sulfur difluoride	10.20	8.99	-1.2
SiCl ₄	Silicon tetrachloride	11.79	12.98	1.2
C ₄ H ₂ O ₃	Malaic anhydride	10.84	12.02	1.2
C ₃ H ₇ I	1-Iodopropane	9.27	10.43	1.2
CH ₂ F ₂	Difluoromethane	13.17	12.02	-1.2
N ₂ F ₂	trans-Difluorodiazene	13.40	12.25	-1.2
PI ₃	Phosphorus triiodide	9.15	10.30	1.2
B ₂ F ₄	B ₂ F ₄	13.26	12.12	-1.1

Table 6.19: Worst Ionization Potential Errors in PM3

Formula	Chemical Name	Exp. I.P.	Calc. I.P.	Diff.
H ₂ O	Water, fourth I.P.	32.20	36.83	4.6
GaBr ₃	Gallium tribromide	10.94	15.02	4.1
InBr ₃	Indium tribromide	10.30	14.20	3.9
SnH ₄	Tin tetrahydride (stannane)	9.20	12.87	3.7
C ₂ F ₄ Br ₂	1,2-Dibromotetrafluoroethane	14.44	11.03	-3.4
NS	Sulfur nitride	8.87	12.26	3.4
C ₁₀ MgH ₁₀	Dicyclopentadienyl magnesium	11.00	8.36	-2.6
C ₅ H ₈ O ₂	Acetylacetone	8.38	10.89	2.5
C ₅ H ₁₀	2-Methyl-1-butene	7.40	9.85	2.4
SOCl ₃	SOCl ₃	9.63	12.06	2.4
HgBr ₂	Mercury dibromide	10.62	12.91	2.3
C ₅ H ₁₀	1-Pentene	7.90	10.15	2.2
SnF ₂	Tin difluoride	8.00	10.26	2.3
CCl ₃	Trichloromethyl	8.28	10.36	2.1
H ₂ SiF ₂	Difluorosilane	12.85	10.82	-2.0
SnBr ₄	Tin tetrabromide	11.00	13.02	2.0
SO ₃	Sulfur trioxide	11.00	12.91	1.9
PS	Phosphorus sulfide	9.00	10.86	1.9
N ₂	Nitrogen	15.60	13.80	-1.8
C ₄ SnH ₁₂	Tetramethyltin	8.90	10.70	1.8
C ₃ H ₇ NO ₂	Alanine	8.10	9.88	1.8
SO ₂	Sulfur dioxide	12.30	10.55	-1.8
C ₃ SnH ₉ Br	Trimethyltin bromide	9.60	11.35	1.8
ZnBr ₂	Zinc dibromide	10.89	12.62	1.7
P ₂	Phosphorus dimer	10.62	8.91	-1.7
C ₁₈ BiH ₁₅	Triphenylbismuth	7.45	9.16	1.7
HCl	Hydrogen chloride	12.75	11.06	-1.7
PbO ₂	Lead dioxide	8.87	10.55	1.7
SF ₄	Sulfur tetrafluoride	12.05	10.41	-1.6
CF ₃ Cl	Trifluorochloromethane	10.30	11.88	1.6
C ₆ Pb ₂ H ₁₈	Hexamethyldiplumbane	7.41	9.00	1.6
GeO	Germanium oxide	11.25	9.68	-1.6
BeFCl	Beryllium chloride fluoride	13.00	11.44	-1.6
PbF ₂	Lead difluoride	11.84	10.28	-1.6
C ₁₈ SbH ₁₅	Triphenylstibine	8.08	9.64	1.6
SOF ₂	Thionyl fluoride	12.58	11.03	-1.6
C ₄ GeH ₁₂	Tetramethylgermanium	9.29	10.83	1.5
C ₈ H ₁₄	Bicyclo(2.2.2)-octane	9.45	10.94	1.5
N ₂ F ₂	trans-Difluorodiazene	13.40	11.91	-1.5
NF ₃	Nitrogen trifluoride	13.73	12.24	-1.5
CHgH ₃ Br	Methylmercuric bromide	9.25	10.70	1.4
H ₂ SiCl ₂	Dichlorosilane	11.70	10.27	-1.4
C ₃ SbH ₉	Trimethylstibine	8.48	9.90	1.4
C ₆ Ge ₂ H ₁₈	GeMe ₃ -GeMe ₃	8.18	9.59	1.4
NOF	Nitrosyl fluoride	12.94	11.54	-1.4
C ₈ PbH ₂₀	Tetraethyllead	8.13	9.54	1.4
PbTe	Lead telluride	8.04	9.44	1.4
SnBr ₂	Tin dibromide	9.80	11.19	1.4
C ₄ PbH ₁₂	Tetramethyllead	8.90	10.27	1.4
C ₆ GeSnH ₁₈	GeMe ₃ -SnMe ₃	8.20	9.55	1.4

Table 6.20: Average Errors (eV) in Ionization Potential by Element

Element	PM3			MNDO			AM1		
	No.	Signed	RMS	No.	Signed	RMS	No.	Signed	RMS
H	244	0.623	0.848	232	0.818	1.097	221	0.550	0.744
Li				1	0.120	0.120			
Be	8	0.844	0.927	8	0.565	0.788	8	0.586	0.767
B				6	0.765	0.882	6	0.793	0.823
C	253	0.587	0.769	241	0.819	1.041	231	0.575	0.784
N	45	0.616	0.879	46	0.705	1.026	46	0.608	0.843
O	79	0.706	0.983	79	0.876	1.339	77	0.770	1.094
F	67	0.692	0.935	64	0.695	1.104	62	0.754	0.992
Mg	5	0.932	1.271						
Al	2	0.485	0.501	2	0.995	0.998	2	0.520	0.522
Si	11	0.694	0.868	11	0.871	1.050	11	0.680	0.822
P	17	0.774	0.922	17	1.426	1.621	17	0.902	1.192
S	48	0.573	0.900	48	0.833	1.204	47	0.828	1.203
Cl	51	0.691	0.863	47	1.321	1.578	43	0.752	1.158
Zn	5	0.986	1.063	5	1.086	1.120	5	0.492	0.517
Ga	3	1.813	2.457						
Ge	13	0.914	1.014	13	1.437	1.517	12	1.103	1.258
As	5	0.616	0.677						
Se	6	0.517	0.562						
Br	27	1.191	1.620	24	1.136	1.198	20	0.639	0.878
Cd	5	0.386	0.614						
In	2	2.065	2.763						
Sn	12	1.378	1.691	12	1.837	2.127			
Sb	5	1.188	1.250						
Te	2	0.890	1.026						
I	30	0.443	0.537	26	1.292	1.350	25	1.083	1.134
Hg	12	0.727	0.946	12	1.116	1.157	12	0.522	0.673
Tl	2	1.220	1.220						
Pb	12	1.068	1.192	12	1.343	1.433			
Bi	4	0.632	0.914						
Totals:	375	0.7	1.0	343	0.9	1.2	319	0.7	0.9

Table 6.21: Worst Dipole Moment Errors in MNDO

Formula	Chemical Name	Exp. Dipole	Calc. Dipole	Diff.
GeS	Germanium sulfide	2.00	5.48	3.5
H ₂ SiCl ₂	Dichlorosilane	1.18	3.49	2.3
SO ₂	Sulfur dioxide	1.58	3.73	2.1
GeF ₂	Germanium difluoride	2.61	4.58	2.0
GeH ₃ Br	Bromogermane	1.70	3.65	2.0
H ₂ SiBr ₂	Dibromosilane	1.43	3.33	1.9
HSiCl ₃	Trichlorosilane	0.86	2.73	1.9
CGeH ₃ N	Cyanogermane	3.99	2.19	-1.8
C ₂ H ₃ N	Methyl isocyanide	3.85	2.17	-1.7
C ₂ GeH ₃ Br ₃	Tribromovinylgermane	2.47	4.13	1.7
CHgH ₃ I	Methylmercuric iodide	1.30	2.88	1.6
HSiF ₃	Trifluorosilane	1.27	2.80	1.5
GeH ₃ Cl	Chlorogermane	2.10	3.59	1.5
CH ₆ N ₂	Methylhydrazine	1.66	0.23	-1.4
C ₄ B ₂ H ₆	C4B2H6	2.26	0.83	-1.4
PF ₃	Phosphorus trifluoride	1.03	2.44	1.4
SnO	Tin oxide	4.30	2.93	-1.4
GeH ₃ I	Iodogermane	1.81	3.17	1.4
NOF	Nitrosyl fluoride	1.81	0.51	-1.3
C ₂ H ₃ N	Acetonitrile	3.92	2.63	-1.3
CNF	Cyanogen fluoride	2.17	0.89	-1.3
HgCl ₂	Mercury dichloride	1.23	0.00	-1.2
AlF	Aluminum fluoride	1.53	0.31	-1.2
C ₄ GeH ₁₂ O ₃	Trimethoxymethylgermane	1.91	0.70	-1.2
C ₂ HgH ₃ N	Methylmercuric cyanide	4.69	3.53	-1.2
CBH ₃ O	BH3CO	1.80	2.96	1.2
CF ₃ I	Trifluoroiodomethane	1.00	2.13	1.1
CH ₆ SiS	CH3-S-SiH3	1.38	2.47	1.1
C ₃ BH ₁₂ N	Trimethylamine-Borane adduct	4.84	5.87	1.0
C ₂ BH ₆ Cl	dimethylchloroborane	0.86	1.89	1.0
PbO	Lead oxide	4.64	3.62	-1.0
C ₃ SnH ₉ I	Trimethyltin iodide	3.37	2.37	-1.0
HgBr ₂	Mercury dibromide	0.95	0.00	-1.0
C ₂ H ₃ NS	Methyl isothiocyanate	4.03	3.10	-0.9
H ₄ P ₂	P2H4	0.92	0.01	-0.9
C ₃ H ₃ N	Acrylonitrile	3.87	2.97	-0.9
C ₂ NF ₃	Trifluoroacetone	1.26	0.36	-0.9
C ₂ HF	Fluoroacetylene	0.70	1.58	0.9
C ₂ GeH ₄	Germylacetylene	0.14	0.94	0.8
C ₂ GeH ₆ Cl ₂	Dimethylgermanium dichloride	3.14	3.94	0.8
BH	BH	1.27	0.49	-0.8
C ₆ H ₄ F ₂	o-Difluorobenzene	2.59	3.36	0.8
CGeH ₃ Cl ₃	Trichloromethylgermane	2.70	3.48	0.8
H ₃ P	Phosphine	0.58	1.35	0.8
GeH ₃ N ₃	Germylazide	2.58	1.84	-0.7
C ₃ H ₉ P	Trimethylphosphine	1.19	1.93	0.7
CH ₂ PF	CH2=P-F	1.35	2.09	0.7
C ₂ H ₈ Si	Ethylsilane	0.81	0.10	-0.7
GeH ₃ F	Fluorogermane	2.30	1.58	-0.7
C ₂ SnH ₆ Cl ₂	Dimethyltin dichloride	4.41	3.70	-0.7

Table 6.22: Worst Dipole Moment Errors in AM1

Formula	Chemical Name	Exp. Dipole	Calc. Dipole	Diff.
SO ₂	Sulfur dioxide	1.58	4.29	2.7
GeS	Germanium sulfide	2.00	4.25	2.2
C ₄ B ₂ H ₆	C4B2H6	2.26	0.11	-2.1
CHgH ₃ I	Methylmercuric iodide	1.30	3.10	1.8
H ₃ P	Phosphine	0.58	2.29	1.7
CS	Carbon sulfide	0.68	2.24	1.6
NOF	Nitrosyl fluoride	1.81	0.38	-1.4
C ₅ BeH ₆	Cyclopentadienylberyllium hydride	2.08	0.72	-1.4
C ₂ H ₃ NS	Methyl isothiocyanate	4.03	2.73	-1.3
C ₂ NF ₃	Trifluoroacetonitrile	1.26	0.03	-1.2
C ₂ B ₄ H ₆	C2B4H6	1.50	0.27	-1.2
HgCl ₂	Mercury dichloride	1.23	0.00	-1.2
C ₂ H ₃ N	Acetonitrile	3.92	2.89	-1.0
C ₂ H ₃ N	Methyl isocyanide	3.85	2.83	-1.0
C ₂ BH ₆ Cl	dimethylchloroborane	0.86	1.84	1.0
C ₄ GeH ₁₂ O ₃	Trimethoxymethylgermane	1.91	0.93	-1.0
CNF	Cyanogen fluoride	2.17	1.21	-1.0
HgBr ₂	Mercury dibromide	0.95	0.00	-1.0
H ₄ P ₂	P2H4	0.92	0.00	-0.9
CH ₅ P	Methylphosphine	1.10	2.02	0.9
CH ₆ N ₂	Methylhydrazine	1.66	0.77	-0.9
H ₂ S	Hydrogen sulfide	0.97	1.86	0.9
POF ₃	Phosphorus oxyfluoride	1.76	2.64	0.9
C ₃ H ₃ N	Acrylonitrile	3.87	3.00	-0.9
CGeH ₃ N	Cyanogermane	3.99	3.14	-0.9
HI	Hydrogen iodide	0.44	1.27	0.8
CGeH ₃ F ₃	Methyltrifluorogermane	3.80	2.98	-0.8
OBr	BrO	1.61	2.36	0.7
GeH ₃ N ₃	Germylazide	2.58	1.89	-0.7
C ₃ HN	CH.C.CN	3.72	3.04	-0.7
CF ₃ I	Trifluoroiodomethane	1.00	1.67	0.7
O ₃	Ozone	0.53	1.20	0.7
GeO	Germanium oxide	3.28	2.62	-0.7
C ₂ HgH ₅ Cl	Ethylmercuric chloride	2.99	3.65	0.7
CH ₆ SiS	CH3-S-SiH3	1.38	2.03	0.6
OCI	Cl-O	1.24	1.88	0.6
PF ₃	Phosphorus trifluoride	1.03	1.66	0.6
C ₂ HgH ₃ N	Methylmercuric cyanide	4.69	4.06	-0.6
CHN	Hydrogen cyanide	2.98	2.36	-0.6
HOF	Hypofluorous acid	2.23	1.60	-0.6
C ₃ GeH ₁₀	Trimethylgermane	0.67	0.05	-0.6
GeH ₃ Br	Bromogermane	1.70	2.31	0.6
C ₃ H ₄ O	Acrolein	3.12	2.53	-0.6
H ₂ SiCl ₂	Dichlorosilane	1.18	1.76	0.6
C ₂ HBr	Bromoacetylene	0.00	0.57	0.6
CBH ₃ OF ₂	Methoxydifluoroborane	2.62	2.05	-0.6
C ₂ H ₇ P	Dimethylphosphine	1.23	1.79	0.6
CHOF	HCOF	2.02	2.57	0.6
HBr	Hydrogen bromide	0.83	1.38	0.6
HSiCl ₃	Trichlorosilane	0.86	1.42	0.6

Table 6.23: Worst Dipole Moment Errors in PM3

Formula	Chemical Name	Exp. Dipole	Calc. Dipole	Diff.
OCl	Cl-O	1.24	3.67	2.4
SO ₂	Sulfur dioxide	1.58	3.63	2.0
AlF	Aluminum fluoride	1.53	3.29	1.8
OBr	BrO	1.61	3.32	1.7
C ₃ SnH ₉ Br	Trimethyltin bromide	3.45	5.14	1.7
H ₂ SiBr ₂	Dibromosilane	1.43	3.09	1.7
NOF	Nitrosyl fluoride	1.81	0.26	-1.6
GeS	Germanium sulfide	2.00	0.48	-1.5
H ₂ SiCl ₂	Dichlorosilane	1.18	2.69	1.5
HSiCl ₃	Trichlorosilane	0.86	2.38	1.5
CHgH ₃ I	Methylmercuric iodide	1.30	2.75	1.4
SnO	Tin oxide	4.30	2.92	-1.4
GaF	Gallium fluoride	2.50	3.85	1.4
CH ₆ N ₂	Methylhydrazine	1.66	0.32	-1.3
InF	Indium fluoride	3.40	2.08	-1.3
C ₃ SnH ₉ I	Trimethyltin iodide	3.37	4.65	1.3
H ₄ P ₂	P2H4	0.92	2.19	1.3
CSeS	Thiocarbonyl selenide	0.03	1.29	1.3
HgCl ₂	Mercury dichloride	1.23	0.00	-1.2
PF ₃	Phosphorus trifluoride	1.03	2.25	1.2
GeH ₃ Br	Bromogermane	1.70	2.92	1.2
O ₃	Ozone	0.53	1.71	1.2
SeH ₂	Hydrogen selenide	0.24	1.35	1.1
C ₂ GeH ₃ Br ₃	Tribromovinylgermane	2.47	3.53	1.1
C ₃ GeH ₉ Br	Trimethylbromogermane	2.84	3.81	1.0
C ₄ SeH ₄	Selenophene	0.37	1.34	1.0
HgBr ₂	Mercury dibromide	0.95	0.00	-1.0
CGeH ₃ F ₃	Methyltrifluorogermane	3.80	2.85	-0.9
C ₂ NF ₃	Trifluoroacetonitrile	1.26	0.32	-0.9
C ₆ H ₅ I	Iodobenzene	1.70	0.79	-0.9
SeO ₂	Selenium dioxide	2.70	3.61	0.9
GeF ₂	Germanium difluoride	2.61	1.71	-0.9
PbO	Lead oxide	4.64	3.77	-0.9
SeOF ₂	Seleninyl difluoride	2.84	3.70	0.9
FBr	Bromine fluoride	1.42	2.25	0.8
H ₂ S	Hydrogen sulfide	0.97	1.78	0.8
GeH ₃ I	Iodogermane	1.81	2.59	0.8
C ₃ H ₄ O	Acrolein	3.12	2.36	-0.8
C ₃ H ₇ NO	Dimethylformamide	3.82	3.06	-0.8
CS	Carbon sulfide	0.68	1.42	0.7
C ₂ H ₃ N	Acetonitrile	3.92	3.21	-0.7
PbS	Lead sulfide	3.59	2.94	-0.6
C ₃ H ₃ N	Acrylonitrile	3.87	3.25	-0.6
C ₂ H ₃ NS	Methyl isothiocyanate	4.03	3.42	-0.6
C ₂ HgH ₃ N	Methylmercuric cyanide	4.69	4.08	-0.6
CH ₆ SiS	CH ₃ -S-SiH ₃	1.38	1.99	0.6
H ₃ P	Phosphine	0.58	1.18	0.6
C ₂ H ₆ S ₂	2,3-Dithiabutane	1.98	2.57	0.6
C ₂ HgH ₅ Br	Ethylmercuric bromide	2.80	3.39	0.6
CF ₃ I	Trifluoroiodomethane	1.00	1.55	0.6

Table 6.24: Average Errors (Debye) in Dipole Moment by Element

Element	PM3			MNDO			AM1		
	No.	Signed	RMS	No.	Signed	RMS	No.	Signed	RMS
H	160	0.371	0.508	159	0.495	0.678	151	0.368	0.514
Be	1	0.250	0.250	1	0.070	0.070	1	1.360	1.360
B				11	0.686	0.798	11	0.615	0.855
C	144	0.338	0.455	145	0.434	0.585	137	0.359	0.517
N	34	0.401	0.529	35	0.596	0.778	35	0.520	0.640
O	51	0.460	0.703	51	0.410	0.584	49	0.355	0.576
F	42	0.480	0.650	41	0.566	0.730	41	0.371	0.498
Al	1	1.760	1.760	1	1.220	1.220	1	0.530	0.530
Si	11	0.638	0.867	11	1.025	1.259	11	0.335	0.389
P	8	0.456	0.667	8	0.778	0.823	8	0.750	0.881
S	18	0.681	0.825	17	0.637	1.090	15	0.726	1.100
Cl	23	0.505	0.774	21	0.693	0.911	18	0.391	0.504
Zn	4	0.153	0.210	4	0.155	0.212	4	0.150	0.209
Ga	1	1.350	1.350						
Ge	22	0.490	0.649	22	1.060	1.294	22	0.544	0.700
As	7	0.311	0.342						
Se	10	0.695	0.778						
Br	26	0.592	0.778	24	0.505	0.747	23	0.323	0.403
In	2	0.720	0.937						
Sn	9	0.699	0.891	9	0.467	0.636			
Sb	1	0.420	0.420						
Te	1	0.040	0.040						
I	14	0.521	0.668	12	0.654	0.812	11	0.519	0.692
Hg	9	0.631	0.777	9	0.674	0.857	9	0.676	0.866
Tl	3	0.347	0.350						
Pb	4	0.398	0.544	3	0.470	0.621			
Totals:	212	0.4	0.6	197	0.5	0.8	185	0.4	0.6

6.7 Comparison of Calc'd and Observed ΔH_f for MNDO, AM1, and PM3

The following Table lists experimental ΔH_f , in kcal/mol, for 1200 species. In order to save space, only the difference between the calculated and observed ΔH_f is reported. The calculated ΔH_f is, however, easily obtained by adding the difference to the experimental value. Thus, the ΔH_f of methane, CH_4 , calculated using MNDO, is $5.95 + (-17.89) = -11.96$ kcal/mol.

All the results reported in this and the following Table can be easily reproduced using MOPAC. In some cases, care must be taken to ensure that the correct spin state and symmetry are specified. For example, triplet oxygen must be specified using `OPEN(2,2) TRIPLET`, in order for the correct ΔH_f to be obtained. Most open-shell systems will not be indicated as open-shell.

Results for the iodine atom are missing. Although MNDO, AM1, and PM3 all give a zero error, the PM3 result is incorrect. The correct PM3 result for the ΔH_f of iodine atom is -46.87 kcal/mol. This can be easily checked by running an iodine atom using keywords `PM3 OPEN(7,4)`.

Many reference data are of highly questionable accuracy, for example the sulfur halides, and some other reference data refer to highly exotic species, e.g. N^{2+} . Results for calculations on all these systems are reported, as required by the protocols for determining accuracy of methods. If only those systems for which accurate data were used, and the choice of system limited to conventional molecules, then the average errors reported in the earlier Tables would decrease markedly.

Notes on the Table

- The ΔH_f of COI is probably incorrect. Given that the experimental ΔH_f s of CO and I are -26.4 and +25.5 kcal/mol, respectively, and the ΔH_f of COI is 63.6 kcal/mol, then the CO-I bond has a 'strength' of -64.5 kcal/mol. This is unlikely.
- ΔH_f for the elements are included. Isolated atoms are used in the MNDO methods as the standard, and the defined ΔH_f should be exactly zero. For PM3 several elements have negative ΔH_f . This is a result of PM3 incorrectly predicting the atomic configuration.

Table 6.25: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3

Empirical Formula	Chemical Name	ΔH_f		Difference			Ref.
		Exp.	PM3	MNDO	AM1		
H	Hydrogen (+)	365.72	-12.14	-39.05	-50.81	a	
H ₂	Hydrogen	0.00	-13.39	0.72	-5.18	b	
CH	Methylidyne	142.40	4.43	1.18	2.63	c	
CH ₂	Methylene, triplet	92.30	-16.66	-15.01	-11.50	d	
CH ₂	Methylene, singlet	99.80	13.42	7.56	11.05	d	
CH ₃	Methyl (+)	261.00	-4.46	-17.13	-8.66	e	
CH ₃	Methyl radical	34.80	-5.04	-9.01	-3.56	d	
CH ₄	Methane	-17.89	4.86	5.93	9.10	f	
C ₂	Carbon, dimer	200.20	58.01	35.21	16.03	g	
C ₂ H ₂	Acetylene	54.34	-3.65	3.53	0.44	f	
C ₂ H ₃	Vinyl (+)	266.00	-2.09	-0.34	-4.57	h	
C ₂ H ₃	Vinyl	59.60	3.66	4.24	5.15	h	
C ₂ H ₄	Ethylene	12.45	4.16	2.93	4.00	f	
C ₂ H ₄	Ethylene (+)	257.00	-8.31	-18.71	-13.17	i	
C ₂ H ₄	Methylmethylene	90.30	-1.69	-1.95	-2.63	j	
C ₂ H ₅	Ethyl (+) (classical)	216.00	6.44	3.64	0.77	k	
C ₂ H ₅	Ethyl radical	25.00	-7.69	-12.20	-6.89	l	
C ₂ H ₆	Ethane	-20.24	2.08	0.49	2.80	f	
C ₃	Carbon, trimer 196.00	10.63	24.30	16.43	d		
C ₃ H ₃	Propynyl (+)	281.00	-5.70	-15.60	-7.36	i	
C ₃ H ₃	Cyclopropenyl (+)	257.00	12.81	15.52	19.37	i	
C ₃ H ₄	Allene	45.63	1.40	-1.74	0.48	f	
C ₃ H ₄	Cyclopropene	66.20	1.94	2.07	8.58	f	
C ₃ H ₄	Propyne	44.39	-4.20	-3.03	-1.02	f	
C ₃ H ₅	Allyl (+)	226.00	6.66	-4.65	0.18	i	
C ₃ H ₅	Propenyl (+)	237.00	1.20	3.06	-3.36	i	
C ₃ H ₅	Allyl	40.00	-0.41	-4.67	-1.45	h	
C ₃ H ₅	Cyclopropyl (+)	235.00	26.78	23.15	25.57	i	
C ₃ H ₆	Cyclopropane	12.73	3.50	-1.55	5.01	f	
C ₃ H ₆	Propene	4.88	1.49	0.07	1.66	f	
C ₃ H ₇	Propyl (+)	208.00	6.33	4.32	-0.28	i	
C ₃ H ₇	2-Propyl (+)	187.00	10.22	13.65	4.83	k	
C ₃ H ₇	i-Propyl radical	16.80	-11.57	-15.32	-10.20	l	
C ₃ H ₈	Propane	-24.83	1.17	-0.15	0.53	f	
C ₄	Carbon, tetramer 232.00	48.29	52.71	47.85	d		
C ₄ H ₂	Diacetylene	113.00	-10.58	-9.85	-6.96	m	
C ₄ H ₄	Vinylacetylene	72.80	-6.47	-7.26	-4.91	m	
C ₄ H ₆	1-Methylcycloprop-1-ene	58.20	-0.87	-4.55	6.41	f	
C ₄ H ₆	Bicyclobutane	51.90	17.29	12.11	26.15	f	
C ₄ H ₆	Methylenecyclopropane	47.92	-3.45	-10.07	-0.31	f	

Table 6.26: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f		Difference		Ref.
		Exp.	PM3	MNDO	AM1	
C ₄ H ₆	1,2-Butadiene	38.80	-0.85	-5.30	-1.73	f
C ₄ H ₆	1-Butyne	39.50	-3.86	-3.40	-2.05	f
C ₄ H ₆	2-Butyne	34.70	-4.97	-9.85	-2.77	f
C ₄ H ₆	Cyclobutene	37.45	0.17	-6.48	8.26	n
C ₄ H ₆	1,3-Butadiene	26.01	4.98	2.90	3.86	f
C ₄ H ₇	2-Butenyl (+)	200.00	12.51	6.91	6.23	i
C ₄ H ₇	Cyclobutyl (+)	213.00	12.55	8.35	13.11	i
C ₄ H ₈	1-Butene	-0.20	1.93	0.53	0.56	f
C ₄ H ₈	cis-2-Butene	-1.86	-0.65	-2.52	-0.32	f
C ₄ H ₈	Cyclobutane	6.78	-10.62	-18.73	-7.82	f
C ₄ H ₈	Isobutene	-4.30	0.93	2.25	3.10	f
C ₄ H ₈	trans-2-Butene	-3.00	-0.82	-2.14	-0.40	f
C ₄ H ₉	n-Butyl (+)	201.00	6.95	5.83	-1.10	i
C ₄ H ₉	1-Methyl propyl (+)	183.00	7.72	11.29	0.91	i
C ₄ H ₉	Isobutyl	4.50	-10.45	-11.33	-7.21	l
C ₄ H ₉	Isobutyl (+)	176.00	2.70	11.92	-1.26	i
C ₄ H ₁₀	n-Butane	-30.36	1.25	0.60	-0.82	f
C ₄ H ₁₀	Isobutane	-32.41	2.83	5.58	2.99	f
C ₅ H ₅	Cyclopentadienyl (-)	21.30	-5.45	-2.43	3.81	o
C ₅ H ₆	Cyclopentadiene	32.10	-0.40	-0.08	4.90	n
C ₅ H ₈	1,2-Dimethylcyclopropene	46.40	0.26	-7.24	8.13	p
C ₅ H ₈	Methylene cyclobutane	29.10	-9.46	-18.27	-4.03	q
C ₅ H ₈	1,cis-3-Pentadiene	19.10	1.91	0.90	1.78	f
C ₅ H ₈	Cyclopentene	8.30	-5.34	-8.70	-5.39	n
C ₅ H ₈	Bicyclo(2.1.0)-pentane	37.30	0.44	-7.17	8.69	p
C ₅ H ₈	1,4-Pentadiene	25.30	1.22	-0.80	-0.57	f
C ₅ H ₈	Spiropentane	44.30	-1.25	-10.68	6.09	f
C ₅ H ₈	1,trans-3-Pentadiene	18.10	3.11	1.22	2.42	f
C ₅ H ₉	Cyclopentyl (+)	188.00	5.42	6.14	-2.14	r
C ₅ H ₁₀	1-Pentene	-5.30	1.23	0.20	-1.44	f
C ₅ H ₁₀	2-Methyl-1-butene	-8.60	0.67	2.00	1.71	f
C ₅ H ₁₀	2-Methyl-2-butene	-10.10	-2.15	-0.45	0.09	f
C ₅ H ₁₀	3-Methyl-1-butene	-6.60	2.60	4.25	2.16	f
C ₅ H ₁₀	cis-2-Pentene	-7.00	-0.78	-1.86	-2.01	f
C ₅ H ₁₀	cis-Dimethylcyclopropane	1.30	0.05	-3.47	3.40	p
C ₅ H ₁₀	Cyclopentane	-18.30	-5.66	-12.24	-10.56	n
C ₅ H ₁₀	trans-2-Pentene	-7.90	-0.89	-2.46	-1.88	f
C ₅ H ₁₁	1-Pentyl (+)	194.00	8.31	7.79	-1.25	i
C ₅ H ₁₁	2-Pentyl (+)	173.00	11.59	15.93	3.33	i
C ₅ H ₁₁	2-Ethylisopropyl (+)	156.00	15.71	26.87	10.41	i
C ₅ H ₁₁	Neopentyl (+)	188.00	20.50	12.14	17.71	s
C ₅ H ₁₂	2-Methylbutane	-36.80	2.49	6.71	1.34	f
C ₅ H ₁₂	Neopentane	-40.30	4.40	15.63	7.46	f
C ₅ H ₁₂	n-Pentane	-35.10	0.55	0.61	-2.94	f
C ₆ H ₆	Benzene	19.81	3.58	1.44	2.14	f
C ₆ H ₆	Fulvene	47.50	8.64	6.11	15.08	f
C ₆ H ₈	1,3-Cyclohexadiene	25.40	-5.06	-10.98	-7.93	f
C ₆ H ₁₀	2,3-Dimethyl-1,3-butadiene	10.80	3.16	4.28	6.47	f

Table 6.27: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₆ H ₁₀	Cyclohexene	-1.08	-3.87	-8.91	-9.05	f
C ₆ H ₁₀	1,5-Hexadiene	20.10	0.90	-0.54	-2.37	f
C ₆ H ₁₀	1,2-Dimethylcyclobutene	19.80	-3.71	-13.46	7.07	p
C ₆ H ₁₀	Bicyclopropyl	30.90	5.10	-2.33	8.58	f
C ₆ H ₁₁	1-Methylcyclopentyl (+)	165.00	9.42	13.55	2.36	i
C ₆ H ₁₁	Cyclohexyl (+)	177.00	9.00	9.80	-2.85	r
C ₆ H ₁₂	Cyclohexane	-29.49	-1.62	-5.35	-9.13	n
C ₆ H ₁₄	n-Hexane	-39.92	-0.06	0.71	-4.98	f
C ₇ H ₇	Benzyl (+)	212.00	15.34	5.96	10.04	t
C ₇ H ₇	Tropylium (+)	209.00	11.92	-1.41	1.37	u
C ₇ H ₈	Cycloheptatriene	43.20	-0.77	-9.47	-4.98	n
C ₇ H ₈	Norbornadiene	59.70	-1.01	3.10	7.91	p
C ₇ H ₈	Toluene	11.99	2.04	1.52	2.35	f
C ₇ H ₁₁	2-Norbornyl (+)	182.00	26.37	30.99	20.88	i
C ₇ H ₁₂	Norbornane	-12.40	-1.34	1.98	-2.09	v
C ₇ H ₁₆	n-Heptane	-44.85	-0.56	0.91	-6.91	f
C ₈ H ₈	Cubane	148.70	-35.04	-49.76	2.41	f
C ₈ H ₈	Styrene	35.30	3.81	3.62	3.33	f
C ₈ H ₁₀	Ethylbenzene	7.15	2.21	1.50	1.42	f
C ₈ H ₁₀	m-Xylene	4.14	0.54	1.68	2.61	f
C ₈ H ₁₀	o-Xylene	4.56	0.88	3.67	2.78	f
C ₈ H ₁₀	p-Xylene	4.31	0.34	1.34	2.39	f
C ₈ H ₁₄	Bicyclo(2.2.2)-octane	-24.10	-3.79	-2.36	-12.05	v
C ₈ H ₁₈	2,2,3,3-Tetramethylbutane	-53.83	9.38	44.92	17.41	f
C ₈ H ₁₈	n-Octane	-49.86	-0.99	1.20	-8.76	f
C ₉ H ₂₀	3,3-Diethylpentane	-55.41	3.90	28.35	4.03	f
C ₉ H ₂₀	n-Nonane	-54.66	-1.62	1.28	-10.82	f
C ₁₀ H ₈	Azulene	73.50	7.73	-1.50	10.79	f
C ₁₀ H ₈	Naphthalene	36.05	4.51	2.15	4.42	f
C ₁₀ H ₁₆	Adamantane	-31.90	-2.88	5.37	-11.45	w
C ₁₂ H ₁₀	Biphenyl	43.53	4.39	2.39	3.93	f
C ₁₂ H ₁₈	Hexamethylbenzene	-20.75	-6.73	20.80	7.36	f
C ₁₂ H ₂₆	n-Duodecane	-69.24	-3.35	1.68	-16.82	f
C ₁₄ H ₁₀	Anthracene	55.20	6.30	3.47	7.56	f
C ₁₄ H ₁₀	Phenanthrene	49.50	5.36	6.00	7.78	f
HO	Hydroxide (-)	-33.20	15.70	27.42	19.08	o
HO	Hydroxyl radical	9.50	-6.55	-9.00	-8.59	x
H ₂ O	Water	-57.80	4.37	-3.15	-1.45	d
H ₃ O	Hydronium (+)	138.90	20.17	-4.75	4.56	d
CO	Carbon monoxide	-26.42	6.66	20.49	20.72	d
CHO	HCO (+)	199.00	-22.11	-14.14	-11.52	x
CHO	HCO	10.40	-19.68	-10.77	-11.38	d
CH ₂ O	Formaldehyde	-25.95	-8.15	-6.95	-5.56	f
CH ₂ O	Hydroxymethylene (trans)	53.20	-2.75	-6.79	-5.05	j
CH ₂ O	Hydroxymethylene (cis)	58.50	-8.36	-12.19	-11.26	j
CH ₃ O	CH ₂ OH (+)	168.00	-1.73	-12.46	-6.71	i
CH ₃ O	Methoxy (-)	-36.00	-1.96	-3.77	-2.52	o
CH ₃ O	Methoxy	-0.50	-6.33	0.32	-3.22	l

Table 6.28: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
CH ₄ O	Methanol	-48.10	-3.80	-9.28	-8.95	f
C ₂ H ₂ O	Ketene	-11.40	2.17	4.57	5.71	f
C ₂ H ₂ O	HCCOH	36.00	-12.81	-16.72	-11.49	j
C ₂ H ₄ O	Acetaldehyde	-39.73	-4.50	-2.59	-1.87	f
C ₂ H ₄ O	Ethylene oxide	-12.58	4.41	-2.99	3.59	f
C ₂ H ₅ O	Ethoxy (-)	-47.50	2.71	2.16	1.97	o
C ₂ H ₆ O	Ethanol	-56.20	-0.69	-6.83	-6.50	f
C ₂ H ₆ O	Dimethyl ether	-43.99	-4.36	-7.27	-9.22	f
C ₃ H ₆ O	Acetone	-51.99	-1.37	2.54	2.75	f
C ₃ H ₆ O	Propanal	-45.50	-3.86	-2.58	-2.85	f
C ₃ H ₆ O	Trimethylene oxide	-19.25	-7.51	-17.96	-6.35	f
C ₃ H ₈ O	Isopropanol	-65.10	-0.71	-0.05	-4.49	f
C ₃ H ₈ O	Propanol	-61.20	-2.47	-6.38	-9.45	f
C ₄ H ₄ O	Furan	-8.30	4.21	-0.37	11.19	f
C ₄ H ₆ O	2-Butenal	-24.00	-3.17	-2.90	-1.70	f
C ₄ H ₆ O	Divinyl ether	-3.27	2.48	1.22	4.53	f
C ₄ H ₈ O	2-Butanone	-57.02	-0.48	2.91	1.80	y
C ₄ H ₈ O	Butanal	-48.90	-5.86	-3.96	-6.27	f
C ₄ H ₈ O	Tetrahydrofuran	-44.02	-7.37	-15.33	-14.45	f
C ₄ H ₁₀ O	Diethyl ether	-60.30	0.61	0.06	-4.76	f
C ₄ H ₁₀ O	t-Butanol	-74.70	3.36	10.36	3.04	f
C ₅ H ₈ O	Cyclopentanone	-46.03	-9.21	-11.07	-9.37	f
C ₅ H ₁₀ O	Tetrahydropyran	-53.40	-4.09	-8.72	-13.74	f
C ₅ H ₁₂ O	3-Pentanol	-75.20	1.31	1.66	-5.59	f
C ₆ H ₅ O	Phenoxy (-)	-40.50	-3.69	-1.81	-0.55	o
C ₆ H ₆ O	Phenol	-23.04	1.29	-3.72	0.71	f
C ₆ H ₁₀ O	Cyclohexanone	-54.04	-6.20	-6.17	-9.38	f
C ₇ H ₆ O	Benzaldehyde	-8.80	-1.94	-0.86	-0.21	f
C ₇ H ₈ O	Anisole	-17.30	2.65	-0.47	1.35	f
C ₈ H ₁₀ O	Phenetole	-26.30	5.77	3.63	4.85	f
C ₁₀ H ₈ O	1-Naphthol	-5.10	0.85	-1.88	2.80	f
C ₁₀ H ₈ O	2-Naphthol	-10.10	5.41	0.30	6.16	f
O ₂	Oxygen (Singlet)	22.00	-3.63	-9.87	-21.31	z
O ₂	Oxygen (Triplet)	0.00	-4.19	-15.36	-27.08	b
H ₂ O ₂	Hydrogen peroxide	-32.50	-8.30	-5.77	-2.76	d
CO ₂	Carbon dioxide	-94.05	8.98	18.94	14.19	d
CHO ₂	HCOO (-)	-106.60	-4.34	4.95	-2.87	o
CH ₂ O ₂	Formic acid	-90.50	-3.94	-2.11	-6.92	f
C ₂ H ₂ O ₂	trans Glyoxal	-50.70	-13.66	-10.73	-8.05	f
C ₂ H ₃ O ₂	CH ₃ COO (-)	-122.50	2.82	12.44	7.05	o
C ₂ H ₄ O ₂	Acetic acid	-103.30	1.26	2.14	0.27	f
C ₂ H ₄ O ₂	Methyl formate	-83.60	-3.48	-1.97	-7.50	m
C ₂ H ₆ O ₂	Dimethyl peroxide	-30.10	-4.05	1.75	4.32	f
C ₂ H ₆ O ₂	Ethylene glycol	-93.90	-4.70	-12.70	-17.60	f
C ₃ O ₂	Carbon suboxide	-22.40	-1.66	-1.15	7.76	f
C ₃ H ₄ O ₂	beta-Propiolactone	-67.60	-3.08	-3.32	-3.46	f

Table 6.29: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₃ H ₆ O ₂	Propionic acid	-108.40	1.97	2.04	-0.75	f
C ₃ H ₆ O ₂	Methyl acetate	-97.90	3.74	4.18	1.42	m
C ₃ H ₈ O ₂	Dimethoxymethane	-83.30	-6.95	-11.13	-20.06	f
C ₄ H ₆ O ₂	Diacetyl	-78.20	-5.27	-0.58	3.26	f
C ₄ H ₁₀ O ₂	Diethyl peroxide	-46.10	6.05	7.10	9.34	f
C ₅ H ₈ O ₂	Acetylacetone	-90.50	-1.20	6.16	4.86	f
C ₆ H ₄ O ₂	p-Benzoquinone	-29.30	-2.33	-3.65	4.13	f
C ₇ H ₆ O ₂	Benzoic acid	-70.10	3.83	2.31	2.04	f
C ₇ H ₆ O ₂	o-Salicylic acid	-70.10	3.83	2.31	2.04	f
O ₃	Ozone	34.10	16.97	14.38	3.59	d
C ₄ H ₂ O ₃	Malaic anhydride	-95.20	5.04	6.64	18.76	f
C ₄ H ₆ O ₃	Acetic anhydride	-137.10	2.01	4.42	5.35	f
C ₂ H ₂ O ₄	Oxalic acid	-175.00	0.92	-0.14	2.57	f
N	Nitrogen (++)	1133.90	-170.01	-160.75	-160.97	aa
H ₂ N	Amidogen	45.50	-10.09	-8.50	-7.09	d
H ₃ N	Ammonia	-11.00	7.93	4.62	3.71	d
H ₄ N	Ammonium (+)	155.00	-1.60	9.63	-4.43	bb
CN	Cyanide	104.00	23.99	25.27	10.37	d
CHN	Hydrogen cyanide	32.30	0.64	3.00	-1.31	d
CH ₄ N	CH ₂ -NH ₂ (+)	178.00	7.31	8.76	-1.76	l
CH ₄ N	CH ₃ NH (-)	30.50	-8.37	-6.41	3.28	o
CH ₄ N	CH ₃ -NH.	37.00	-9.76	-4.31	-2.90	l
CH ₅ N	Methylamine	-5.50	1.48	-1.00	-0.64	f
C ₂ H ₃ N	Acetonitrile	20.90	2.36	-1.70	-1.65	cc
C ₂ H ₃ N	Methyl isocyanide	35.60	19.06	24.71	14.75	cc
C ₂ H ₅ N	Ethyleneimine (Azirane)	30.20	1.37	-5.15	2.89	f
C ₂ H ₆ N	Me ₂ N (-)	24.70	-16.89	-16.20	-2.38	o
C ₂ H ₇ N	Ethylamine	-11.40	-1.16	-1.88	-3.78	f
C ₂ H ₇ N	Dimethylamine	-4.43	-3.50	-2.25	-1.23	f
C ₃ H ₃ N	Acrylonitrile	44.10	6.02	-0.30	0.82	f
C ₃ H ₅ N	Ethyl cyanide	12.10	6.41	1.62	0.88	m
C ₃ H ₉ N	Isopropylamine	-20.00	1.20	3.60	0.69	f
C ₃ H ₉ N	Trimethylamine	-5.67	-5.25	2.83	3.91	f
C ₃ H ₉ N	n-Propylamine	-16.80	-1.12	-1.47	-5.34	f
C ₄ H ₅ N	Pyrrrole	25.90	1.15	6.48	13.91	f
C ₄ H ₉ N	Pyrrrolidine	-0.80	-11.26	-15.11	-9.70	f
C ₄ H ₁₁ N	t-Butylamine	-28.90	3.63	13.38	7.62	f
C ₅ H ₅ N	Pyridine	34.60	-4.30	-5.86	-2.63	f
C ₆ H ₇ N	Aniline	20.80	0.42	0.83	-0.39	f
C ₆ H ₁₅ N	Triethylamine	-22.06	-4.60	7.26	5.72	f
C ₇ H ₅ N	Phenyl cyanide	51.50	6.89	0.38	1.79	f
NO	Nitrogen oxide	21.58	-6.85	-21.75	-20.41	d
NO	NO (+)	237.00	1.21	-6.42	-8.86	d
CNO	NCO	38.10	-5.77	-1.02	0.77	d
CHNO	Hydrogen isocyanate	-24.30	8.98	13.47	9.11	d
C ₃ H ₇ NO	Dimethylformamide	-45.80	1.13	8.36	8.85	f

Table 6.30: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
NO ₂	Nitrogen dioxide	7.91	-8.95	-12.49	-22.92	d
NO ₂	Nitrogen dioxide (+)	233.00	-24.65	7.60	-11.92	l
HNO ₂	Nitrous acid, trans	-18.80	3.89	-21.89	-20.65	d
CH ₃ NO ₂	Nitromethane	-17.90	1.91	21.17	7.91	f
CH ₃ NO ₂	Methyl nitrite	-15.80	6.63	-20.91	-16.07	f
C ₂ H ₅ NO ₂	Nitroethane	-23.50	2.58	21.35	6.62	f
C ₂ H ₅ NO ₂	Glycine	-93.70	-2.32	-1.33	-7.90	f
C ₃ H ₇ NO ₂	1-Nitropropane	-30.00	3.15	21.81	6.31	f
C ₃ H ₇ NO ₂	2-Nitropropane	-33.20	5.98	26.79	11.48	f
C ₃ H ₇ NO ₂	Alanine	-111.40	10.23	13.35	6.36	f
C ₄ H ₉ NO ₂	1-Nitrobutane	-34.40	2.17	21.47	3.89	f
C ₄ H ₉ NO ₂	2-Nitrobutane	-39.10	7.11	28.90	10.81	f
C ₄ H ₉ NO ₂	2-Nitroisobutane	-42.20	9.85	38.84	18.07	dd
C ₆ H ₅ NO ₂	Nitrobenzene	15.40	-0.97	22.08	9.78	m
C ₇ H ₇ NO ₂	2-Nitrotoluene	9.30	-1.41	20.55	9.49	f
C ₇ H ₇ NO ₂	3-Nitrotoluene	4.10	0.86	25.83	13.37	ee
C ₇ H ₇ NO ₂	4-Nitrotoluene	7.40	-2.83	22.31	9.73	ee
NO ₃	Nitrate radical	17.00	5.81	28.09	16.16	d
NO ₃	Nitrate anion	-74.70	-18.65	7.62	-14.21	aa
HNO ₃	Nitric acid	-32.10	-5.96	14.50	-5.43	d
CH ₃ NO ₃	Methyl nitrate	-29.11	-3.37	16.60	-2.29	f
C ₂ H ₅ NO ₃	Ethyl nitrate	-36.83	0.37	18.84	-0.53	f
C ₂ H ₅ NO ₃	Nitroethanol	-75.10	13.60	29.35	10.43	ff
N ₂	Nitrogen	0.00	17.55	8.26	11.15	b
H ₂ N ₂	Diazene	36.00	1.73	-4.17	-4.47	gg
H ₄ N ₂	Hydrazine	22.80	-2.17	-8.65	-9.15	d
CH ₂ N ₂	Diazomethane	71.00	-10.00	-3.77	-8.40	f
CH ₂ N ₂	N=N-CH ₂ -	79.00	12.65	-6.62	7.80	hh
CH ₆ N ₂	Methylhydrazine	22.60	-4.74	-8.27	-5.31	f
C ₂ N ₂	Cyanogen	73.80	3.64	-7.25	-5.91	f
C ₂ H ₈ N ₂	1,1-Dimethylhydrazine	20.00	-4.95	-1.91	3.91	f
C ₂ H ₈ N ₂	1,2-Dimethylhydrazine	22.00	-6.47	-7.02	-0.53	f
C ₄ N ₂	Dicyanoacetylene	126.50	1.49	-15.16	-6.77	f
C ₄ H ₂ N ₂	Fumaronitrile	81.30	4.63	-6.62	-5.35	f
C ₄ H ₄ N ₂	Pyridazine	66.50	-10.58	-22.96	-11.25	f
C ₄ H ₄ N ₂	Pyrimidine	47.00	-9.10	-12.10	-3.16	f
C ₄ H ₄ N ₂	Pyrazine	46.90	-7.65	-9.22	-2.76	f
C ₆ H ₁₄ N ₂	azo-n-Propane	8.60	-2.93	-5.84	6.03	ii
N ₂ O	Nitrous oxide	19.60	5.76	11.40	8.82	d
CH ₄ N ₂ O	Urea	-58.70	17.70	19.06	14.58	f
C ₂ H ₆ N ₂ O ₂	N-Nitrodimethylamine	-3.20	4.39	25.45	24.81	m
C ₆ H ₆ N ₂ O ₂	Para nitroaniline	16.20	-5.64	20.62	5.22	f
N ₂ O ₃	Dinitrogen trioxide	19.80	3.86	-6.18	1.98	d
N ₂ O ₄	Dinitrogen tetroxide	2.17	6.08	27.69	22.82	d
CH ₂ N ₂ O ₄	Dinitromethane	-13.30	1.26	41.12	16.11	ff
C ₂ H ₄ N ₂ O ₄	1,1-Dinitroethane	-24.10	6.61	47.23	21.41	ff

Table 6.31: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₂ H ₄ N ₂ O ₄	1,2-Dinitroethane	-22.90	3.23	42.77	12.83	ff
C ₃ H ₈ O ₂	Dimethoxymethane	-83.30	-6.95	-11.13	-20.06	f
C ₃ H ₆ N ₂ O ₄	1,1-Dinitropropane	-25.90	3.82	45.21	16.58	f
C ₃ H ₆ N ₂ O ₄	1,3-Dinitropropane	-31.60	4.97	44.41	13.28	ff
C ₃ H ₆ N ₂ O ₄	2,2-Dinitropropane	-27.00	3.90	50.10	21.34	ff
C ₄ H ₈ N ₂ O ₄	1,1-Dinitrobutane	-34.10	6.72	48.81	17.96	ff
C ₄ H ₈ N ₂ O ₄	1,4-Dinitrobutane	-38.90	6.06	45.71	11.13	ff
C ₆ H ₄ N ₂ O ₄	m-Dinitrobenzene	11.30	-2.23	47.24	21.67	f
C ₆ H ₄ N ₂ O ₄	o-Dinitrobenzene	20.20	-11.13	34.49	12.77	ee
C ₆ H ₄ N ₂ O ₄	p-Dinitrobenzene	13.30	-3.32	41.18	19.83	ee
C ₇ H ₆ N ₂ O ₄	2,4-Dinitrotoluene	4.70	-2.84	44.06	21.32	ee
C ₇ H ₆ N ₂ O ₄	2,6-Dinitrotoluene	9.60	-3.37	42.08	20.31	ee
N ₂ O ₅	Dinitrogen pentoxide	2.70	-21.83	31.44	4.16	d
C ₆ H ₄ N ₂ O ₅	4,6-Dinitrophenol	-23.40	-13.92	33.97	15.57	f
N ₃	Azide	99.00	6.94	3.41	8.37	x
HN ₃	Hydrazoic acid	70.30	4.93	2.75	5.52	cc
CHN ₃ O ₆	Trinitromethane	-3.20	-1.61	61.76	28.04	jj
C ₂ H ₃ N ₃ O ₆	1,1,1-Trinitroethane	-12.40	2.27	68.69	33.33	ff
C ₃ H ₅ N ₃ O ₆	1,1,1-Trinitropropane	-18.40	6.81	73.64	33.77	ff
C ₆ H ₃ N ₃ O ₆	1,3,5-Trinitrobenzene	14.90	-1.92	65.36	34.91	ee
C ₇ H ₅ N ₃ O ₆	2,4,6-Trinitrotoluene	12.90	-9.77	61.56	28.28	f
C ₇ H ₅ N ₃ O ₇	2,4,6-Trinitroanisole	-5.80	-19.97	47.45	18.10	f
C ₈ H ₇ N ₃ O ₇	2,4,6-Trinitrophenetole	-20.10	-10.13	56.66	27.80	f
C ₃ H ₅ N ₃ O ₉	Glycerol trinitrate	-64.70	-13.40	57.91	-9.28	f
CH ₂ N ₄	1-H Tetrazole	79.95	6.24	-26.18	29.63	f
CN ₄ O ₈	Tetranitromethane	18.50	-12.24	76.30	34.38	ff
C ₅ H ₈ N ₄ O ₁₂	Pentaerythritol tetranitrate	-92.50	-5.95	93.73	-3.30	f
C ₃ H ₆ N ₆	Melamine	12.40	12.28	9.03	51.85	f
S	S (-)	16.77	3.91	29.84	11.18	d
HS	Hydrogen sulfide	33.30	4.91	4.00	6.82	d
HS	HS (-)	-17.10	1.24	24.01	5.23	d
H ₂ S	Hydrogen sulfide	-4.90	3.98	8.72	6.10	d
CS	Carbon sulfide	67.00	30.32	37.45	27.84	d
CH ₂ S	Thioformaldehyde	24.30	13.27	3.52	5.63	kk
CH ₄ S	Thiomethanol	-5.40	-0.14	-1.90	1.04	f
C ₂ H ₄ S	Thiirane	19.70	9.09	-0.88	11.00	f
C ₂ H ₆ S	Thioethanol	-11.00	2.25	-2.42	0.36	f
C ₂ H ₆ S	Dimethyl thioether	-8.90	-2.08	-8.19	-0.47	f
C ₃ H ₆ S	Thietane	14.60	-7.14	-19.73	-7.41	f
C ₃ H ₈ S	Isopropanthiol	-18.10	3.71	1.83	2.04	d
C ₃ H ₈ S	Methylethylthioether	-14.20	0.10	-8.87	-1.42	ll
C ₃ H ₈ S	1-Propanthiol	-16.20	2.09	-2.00	-1.21	f
C ₄ H ₄ S	Thiophene	27.60	3.07	-1.18	-0.21	f
C ₄ H ₆ S	2,3-Dihydrothiophene	21.80	-8.43	-18.03	-11.67	mm
C ₄ H ₆ S	2,5-Dihydrothiophene	20.90	-6.11	-16.66	-8.29	mm
C ₄ H ₈ S	Tetrahydrothiophene	-8.10	-2.31	-16.11	-8.46	f

Table 6.32: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₄ H ₈ S	Cisdimethylthiirane	2.70	13.19	2.03	15.06	ll
C ₄ H ₈ S	Transdimethylthiirane	0.90	14.20	3.00	15.84	ll
C ₄ H ₁₀ S	Butanethiol	-21.10	1.59	-1.85	-3.16	d
C ₄ H ₁₀ S	Diethyl thioether	-19.89	0.58	-8.29	-2.27	f
C ₄ H ₁₀ S	Methylpropylthioether	-19.50	0.04	-8.34	-2.88	ll
C ₅ H ₁₀ S	Thiacyclohexane	-15.12	-0.67	-12.97	-8.81	f
C ₅ H ₁₀ S	Cyclopentanethiol	-11.40	-3.51	-9.98	-8.97	nn
C ₆ H ₆ S	Thiophenol	26.90	0.70	-3.55	-1.24	f
C ₆ H ₁₂ S	Cyclohexanethiol	-22.96	1.15	-2.92	-7.53	n
C ₇ H ₈ S	Toluenethiol	23.60	1.35	-3.42	-1.29	f
SO	Sulfur monoxide (triplet)	1.20	-14.83	2.95	-12.73	d
CSO	Carbon oxysulfide	-33.85	10.09	10.93	4.87	f
C ₂ H ₄ SO	Thiolacetic acid	-43.50	4.60	2.17	4.28	hh
C ₂ H ₆ SO	Dimethyl sulfoxide	-36.09	-2.66	40.27	-3.22	f
C ₄ H ₁₀ SO	Diethyl sulfoxide	-49.10	2.46	41.41	-2.45	f
SO ₂	Sulfur dioxide	-71.00	20.22	75.42	23.96	d
C ₂ H ₆ SO ₂	Dimethyl sulfone	-89.10	12.71	142.74	18.74	f
C ₃ H ₈ SO ₂	Methylethyl sulfone	-97.60	18.94	144.61	21.75	f
C ₄ H ₆ SO ₂	Divinyl sulfone	-36.00	10.64	137.33	10.55	f
C ₄ H ₁₀ SO ₂	Diethyl sulfone	-102.50	21.56	142.99	21.16	f
SO ₃	Sulfur trioxide	-94.60	-10.18	153.06	-2.73	d
C ₂ H ₆ SO ₃	Dimethyl sulfite	-115.50	-14.58	50.29	-23.82	f
H ₂ SO ₄	Sulfuric acid	-175.70	-12.43	162.57	-10.67	d
C ₂ H ₆ SO ₄	Dimethyl sulfate	-164.10	-8.06	158.52	-10.69	f
CHNS	Hydrogen isothiocyanate	30.00	9.52	13.40	-2.65	d
C ₂ H ₃ NS	Methyl isothiocyanate	27.10	9.03	9.82	0.55	cc
C ₂ H ₃ NS	Methyl thiocyanate	38.33	-9.41	-14.52	-15.58	cc
CH ₄ N ₂ S	Thiourea	-6.00	29.90	20.76	15.84	oo
S ₂	Sulfur dimer	30.80	-2.12	3.96	-6.43	d
H ₂ S ₂	Hydrogen disulfide	3.71	4.93	2.77	4.83	cc
CS ₂	Carbon disulfide	28.00	8.92	8.91	-10.52	d
CH ₄ S ₂	CH ₃ SSH	-0.90	2.62	-3.31	2.84	kk
C ₂ H ₆ S ₂	CH ₃ SCH ₂ SH	0.10	-2.80	-11.28	-3.83	kk
C ₂ H ₆ S ₂	Ethanedithiol-1,2	-2.20	3.35	-4.16	-1.11	d
C ₂ H ₆ S ₂	2,3-Dithiabutane	-5.60	0.78	-9.20	1.41	f
C ₃ H ₈ S ₂	Propane-1,3-dithiol	-7.00	0.58	-4.38	-4.36	nn
C ₄ H ₁₀ S ₂	(C ₂ H ₅ S) ₂	-17.80	6.79	-9.09	1.08	pp
C ₂ N ₂ S ₂	S ₂ (CN) ₂	82.30	-3.99	-11.85	-18.21	f
H ₂ S ₃	Hydrogen trisulfide	7.29	-1.58	1.12	0.49	cc
C ₂ H ₆ S ₃	2,3,4-Trithiapentane	0.00	-6.92	-13.28	-4.71	qq
C ₃ H ₄ S ₃	1,3-Dithiolan-2-thione	22.70	17.61	-11.36	-2.70	d
S ₄	Sulfur tetramer	32.70	14.69	13.13	0.99	aa
H ₂ S ₄	Hydrogen tetrasulfide	10.57	-10.92	64.50	-7.47	cc
H ₂ S ₅	Hydrogen pentasulfide	13.84	-8.42	-0.61	-6.72	cc
S ₈	S ₈	24.00	-5.85	-0.76	-8.74	d
F	Fluoride (-)	-61.00	29.77	43.87	64.44	d

Table 6.33: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
HF	Hydrogen fluoride	-65.14	2.39	5.40	-9.14	d
CF	Fluoromethylidyne	61.00	-6.95	-22.41	-23.04	d
CH ₂ F	Fluoromethyl (+)	200.30	-0.05	-17.55	-19.97	rr
CH ₃ F	Fluoromethane	-56.80	2.98	-4.13	-4.24	ss
CH ₃ F	Trifluoromethane (+)	233.30	-5.14	-10.13	-29.35	rr
C ₂ HF	Fluoroacetylene	30.00	-11.91	-14.37	-14.81	d
C ₂ H ₃ F	Fluoroethylene	-32.50	3.87	-2.05	-1.58	tt
C ₂ H ₄ F	CH ₃ CHF (+)	166.00	6.83	-1.32	-8.97	uu
C ₂ H ₅ F	Fluoroethane	-62.90	2.65	-2.25	-3.44	f
C ₃ H ₇ F	2-Fluoropropane	-69.40	2.55	2.76	-0.42	f
C ₆ H ₅ F	Fluorobenzene	-27.76	7.45	2.43	4.37	f
OF	FO	26.10	-4.93	-4.38	-3.50	vv
HOF	Hypofluorous acid	-20.90	-8.35	2.26	-1.71	ww
COF	COF	-42.30	-12.73	-7.71	-13.67	g
CHOF	HCOF	-90.00	1.17	1.16	-2.96	d
C ₂ H ₃ OF	Acetyl fluoride	-106.40	7.68	9.87	7.59	f
O ₂ F	Fluorine dioxide	3.00	9.87	21.12	5.96	d
C ₇ H ₅ O ₂ F	p-Fluorobenzoic acid	-118.40	9.41	7.30	6.24	f
CNF	Cyanogen fluoride	8.60	-2.12	-10.97	-13.00	d
NOF	Nitrosyl fluoride	-15.70	12.33	-9.10	-10.82	d
NO ₂ F	Fluorine nitrite	-26.00	0.35	26.60	4.61	d
NO ₃ F	Fluorine nitrate	2.50	-8.64	25.43	11.15	d
SF	SF	-4.10	-7.53	4.82	-16.06	d
SOF	SOF	-63.30	-11.15	34.98	-16.32	g
SO ₂ F	SO ₂ F	-113.20	10.67	127.19	2.32	g
F ₂	Fluorine	0.00	-21.70	7.32	-22.49	b
CF ₂	Difluoromethylene	-43.50	-5.61	-21.73	-24.48	g
CF ₂	Difluoromethylene	-45.00	-4.11	-20.23	-22.98	xx
CHF ₂	Difluoromethyl (+)	142.40	3.07	-10.01	-20.56	rr
CH ₂ F ₂	Difluoromethane	-108.13	4.35	-3.67	-7.99	f
CH ₂ F ₂	Difluoromethane (+)	185.20	-4.84	-6.86	-33.65	rr
C ₂ F ₂	Difluoroacetylene	5.00	-16.61	-26.03	-24.64	d
C ₂ H ₂ F ₂	gem-Difluoroethylene	-80.50	7.47	-3.18	-2.22	f
C ₂ H ₃ F ₂	CH ₃ CF ₂ (+)	107.00	15.17	9.54	-1.75	uu
C ₂ H ₄ F ₂	1,1-Difluoroethane	-118.80	6.84	5.32	0.15	f
C ₆ H ₄ F ₂	o-Difluorobenzene	-70.30	7.09	-0.36	3.76	f
C ₆ H ₄ F ₂	m-Difluorobenzene	-74.00	10.58	2.99	6.01	f
C ₆ H ₄ F ₂	p-Difluorobenzene	-73.30	9.92	2.18	5.23	f
OF ₂	Difluorine oxide	5.90	-10.68	12.28	4.55	d
COF ₂	Carbonyl fluoride	-152.70	11.16	14.23	6.55	f
NF ₂	NF ₂ (-)	-29.50	-1.52	-14.15	4.23	yy
NF ₂	NF ₂ .	10.10	1.79	-24.96	-16.51	d
N ₂ F ₂	cis-Difluorodiazene	16.40	11.54	-18.70	4.36	d
N ₂ F ₂	trans-Difluorodiazene	19.40	9.74	-17.06	11.80	d
SF ₂	Sulfur difluoride	-70.90	-20.97	17.93	-31.18	d
SOF ₂	Thionyl fluoride	-130.00	-8.21	84.24	-26.23	d

Table 6.34: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
SO ₂ F ₂	Sulfuryl fluoride	-181.30	-3.03	203.20	-14.61	d
S ₂ F ₂	FSSF	-80.40	6.56	39.11	-6.43	d
S ₂ F ₂	SSF2	-95.94	39.83	103.88	15.80	d
CF ₃	Trifluoromethyl	-112.40	-19.72	-24.69	-30.41	d
CF ₃	Trifluoromethyl (-)	-163.40	-15.44	-15.42	-15.44	yy
CF ₃	Trifluoromethyl (+)	99.30	0.23	1.56	-17.24	rr
CHF ₃	Trifluoromethane	-166.30	4.30	2.45	-6.22	f
CHF ₃	Trifluoromethane (+)	151.90	-2.52	6.76	-30.70	rr
C ₂ HF ₃	Trifluoroethylene	-117.30	-4.22	-13.85	-13.31	f
C ₂ H ₂ F ₃	CF ₃ CH ₂ (+)	114.00	8.30	7.16	0.34	uu
C ₂ H ₂ F ₃	CF ₃ CH ₂ .	-123.60	-7.66	-6.09	-7.83	zz
C ₂ H ₂ F ₃	CH ₂ F ₂ CF ₂ (+)	81.00	11.67	1.37	-11.99	uu
C ₂ H ₃ F ₃	1,1,1-Trifluoroethane	-178.00	5.68	13.59	5.31	f
C ₇ H ₅ F ₃	Trifluoromethylbenzene	-143.20	8.17	15.57	8.71	f
C ₂ HO ₂ F ₃	Trifluoroacetic acid	-255.00	10.92	17.52	12.24	f
NF ₃	Nitrogen trifluoride	-31.60	7.17	-2.66	-8.47	d
C ₂ NF ₃	Trifluoroacetonitrile	-118.40	3.30	5.15	-1.15	d
NOF ₃	F ₃ NO	-39.00	12.33	61.71	24.35	d
SF ₃	Sulfur trifluoride	-130.00	-4.22	91.47	-25.38	g
SOF ₃	SOF ₃	-185.10	8.61	180.10	-21.67	g
CF ₄	Carbon tetrafluoride	-223.30	-1.83	9.08	-2.46	f
C ₂ F ₄	Tetrafluoroethylene	-157.90	-10.33	-17.09	-17.01	f
COF ₄	Trifluoromethyl hypofluorite	-182.80	-4.52	19.41	4.98	d
N ₂ F ₄	Tetrafluorohydrazine	-2.00	4.92	-8.19	14.66	d
SF ₄	Sulfur tetrafluoride	-182.40	-2.92	135.89	-38.97	d
SOF ₄	SOF ₄	-235.50	-0.83	269.01	-18.44	g
C ₆ HF ₅	Pentafluorobenzene	-192.50	3.78	-9.28	0.40	f
SF ₅	Sulfur pentafluoride	-217.10	-15.42	208.11	-63.28	d
SF ₅	Sulfur pentafluoride (-)	-291.00	-12.02	159.34	-34.93	aaa
C ₂ F ₆	Hexafluoroethane	-321.20	3.34	21.52	7.89	d
C ₆ F ₆	Hexafluorobenzene	-228.50	-0.89	-15.06	-2.78	f
C ₂ OF ₆	Dimethyl perfluoroether	-368.95	-12.11	11.41	-11.55	bbb
C ₃ OF ₆	Perfluoroacetone	-325.20	-14.90	3.15	-6.51	ccc
SF ₆	Sulfur hexafluoride	-291.40	-13.19	320.58	-39.41	d
C ₃ F ₈	Perfluoropropane	-426.15	11.70	41.53	22.56	bbb
C ₄ F ₈	Perfluorocyclobutane	-369.50	-9.79	5.66	2.12	f
C ₃ O ₂ F ₈	CF ₃ -O-CF ₂ -O-CF ₃	-520.60	-15.57	18.00	-15.16	bbb
C ₄ F ₁₀	n-Perfluorobutane	-533.89	23.37	66.53	40.56	bbb
S ₂ F ₁₀	S ₂ F ₁₀	-493.40	82.69	544.19	10.58	d
Cl	Chlorine, atom	28.99	0.00	0.00	0.00	a
Cl	Chloride (-)	-55.90	4.67	1.21	18.24	d
HCl	Hydrogen chloride	-22.10	1.63	6.84	-2.51	d
CCl	Chloromethylidyne	111.30	-6.06	-3.87	-10.17	g
CHCl	Chloromethylene	80.00	3.16	0.94	-2.35	d
CH ₃ Cl	Methyl chloride	-20.00	5.30	-2.52	1.03	d
C ₂ HCl	Chloroacetylene	51.10	-4.55	1.44	-3.36	d

Table 6.35: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₂ H ₃ Cl	Chloroethylene	8.60	1.10	-3.69	-2.75	d
C ₂ H ₅ Cl	Chloroethane	-26.80	4.71	-2.00	0.61	aa
HOCl	Hypochlorous acid	-17.80	-16.50	2.12	-3.98	d
COCl	COCl	-15.00	-1.11	-0.60	-0.44	d
C ₇ H ₅ OCl	Benzoyl chloride	-26.10	7.63	2.60	10.27	f
O ₂ Cl	Chlorine dioxide	25.00	-23.55	110.99	80.87	d
CNCl	Cyanogen chloride	31.60	-0.03	1.64	-6.97	d
NOCl	Nitrosyl chloride	12.36	-7.90	-16.59	-7.71	d
NO ₂ Cl	Nitryl chloride	2.90	-15.90	14.37	11.67	d
SCl	SCl	41.80	-13.16	-25.35	-23.63	g
SOCI	SOCI	-17.40	-13.70	1.49	-20.20	g
SO ₂ Cl	SO ₂ Cl	-66.40	9.07	92.20	5.34	g
FCl	Chlorine fluoride	-12.10	-9.60	20.29	1.56	d
HFCI	Hydrogen chloride fluoride(-)	-142.00	5.31	15.82	23.61	aaa
CH ₂ FCI	Fluorochloromethane	-62.60	5.02	-5.40	-2.75	d
COFCI	Carbonyl fluoride chloride	-102.00	8.37	9.73	9.98	d
O ₃ FCI	Perchloryl fluoride	-5.12	19.63	328.34	251.58	d
CHF ₂ Cl	Difluorochloromethane	-115.60	5.87	1.16	1.33	f
F ₃ Cl	Chlorine trifluoride	-38.00	15.90	116.69	58.19	d
CF ₃ Cl	Trifluorochloromethane	-169.20	-0.08	9.56	6.51	d
F ₅ Cl	Chlorine pentafluoride	-54.00	-0.03	258.73	144.39	d
Cl ₂	Chlorine	0.00	-11.58	-10.68	-14.17	b
HCl ₂	Hydrogen dichloride (-)	-142.00	30.98	47.20	47.59	aaa
CCl ₂	Dichloromethylene	57.00	0.43	0.61	-8.57	d
CH ₂ Cl ₂	Dichloromethane	-22.96	5.80	-5.09	-2.91	f
C ₂ H ₂ Cl ₂	gem-Dichloroethylene	0.60	3.36	-3.27	-3.95	d
C ₂ H ₂ Cl ₂	cis-Dichloroethylene	1.00	2.96	-3.67	-4.35	d
C ₂ H ₂ Cl ₂	trans-Dichloroethylene	1.20	2.33	-4.96	-4.67	d
C ₂ H ₄ Cl ₂	1,1-Dichloroethane	-30.90	4.34	-1.68	-0.24	aa
C ₂ H ₄ Cl ₂	1,2-Dichloroethane	-31.00	6.28	-5.43	-2.85	aa
C ₆ H ₄ Cl ₂	o-Dichlorobenzene	7.10	3.96	1.37	2.04	f
C ₆ H ₄ Cl ₂	m-Dichlorobenzene	6.10	4.02	0.38	2.01	f
C ₆ H ₄ Cl ₂	p-Dichlorobenzene	5.30	4.74	1.00	2.59	f
OCl ₂	Chlorine monoxide	25.00	-41.26	6.28	-5.54	d
COCl ₂	Carbonyl chloride	-52.60	3.50	-0.06	5.24	d
O ₇ Cl ₂	Cl ₂ O ₇	65.00	-91.27	613.55	489.56	aa
SCl ₂	Sulfur dichloride	-4.20	-6.75	-19.66	-22.06	d
SOCl ₂	Thionyl chloride	-50.80	3.18	28.61	-13.53	cc
SO ₂ Cl ₂	Sulfuryl chloride	-86.20	5.25	130.78	16.87	d
S ₂ Cl ₂	ClSSCl	-4.00	-3.68	-16.82	-20.58	d
CHFCl ₂	Fluorodichloromethane	-67.70	5.70	-1.52	2.45	ddd
CF ₂ Cl ₂	Difluorodichloromethane	-117.50	1.42	7.30	10.45	d
CCl ₃	Trichloromethyl	21.00	-19.46	-20.55	-25.76	g
CHCl ₃	Chloroform	-24.66	3.76	-4.31	-4.37	ddd
C ₂ HCl ₃	Trichloroethylene	-2.00	-0.37	-4.43	-6.45	d
C ₂ H ₃ Cl ₃	1,1,1-Trichloroethane	-35.50	3.55	4.00	3.53	f

Table 6.36: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f				Ref.
		Exp.	PM3	MNDO	AM1	
SCl ₃	Sulfur trichloride	8.80	-27.93	-41.31	-68.90	g
SOCl ₃	SOCl ₃	-47.50	0.98	49.86	-22.29	g
CFCl ₃	Fluorotrichloromethane	-69.00	1.64	3.74	8.06	d
CCl ₄	Carbon tetrachloride	-22.90	-3.09	-2.62	-5.26	d
C ₂ Cl ₄	Tetrachloroethylene	-2.70	-5.42	-5.33	-9.78	f
SCl ₄	Sulfur tetrachloride	-0.70	-19.13	-32.65	-82.03	g
SOCl ₄	SOCl ₄	-55.70	26.16	142.61	52.88	g
SCl ₅	Sulfur pentachloride	-8.60	1.09	43.63	-23.07	g
C ₂ Cl ₆	Hexachloroethane	-34.50	-2.02	7.10	-1.35	f
C ₆ Cl ₆	Hexachlorobenzene	-8.60	-0.60	5.92	0.66	f
SCl ₆	Sulfur hexachloride	-19.80	30.07	138.09	54.56	g
Br	Bromine, atom	26.74	0.00	0.00	0.00	a
Br	Bromide (-)	-52.30	-3.94	14.80	31.89	d
HBr	Hydrogen bromide	-8.71	14.02	12.35	-1.80	d
HBr	Hydrogen bromide (+)	261.10	13.61	16.77	-12.33	eee
CBr	Bromomethylidyne	125.90	12.48	1.62	0.87	g
CH ₃ Br	Bromomethane	-9.10	7.10	-1.27	2.90	f
C ₂ H ₃ Br	Bromoethylene	18.71	5.05	-2.92	-0.78	f
C ₂ H ₅ Br	Bromoethane	-15.20	3.80	-1.86	2.05	f
C ₃ H ₅ Br	3-Bromopropene	10.90	4.32	-2.10	1.45	fff
C ₃ H ₇ Br	1-Bromopropane	-20.50	3.77	-1.54	0.52	f
C ₃ H ₇ Br	2-Bromopropane	-23.50	2.61	2.79	5.54	ggg
C ₆ H ₅ Br	Bromobenzene	25.20	5.77	-1.38	1.48	f
OBr	BrO	30.06	-9.28	5.27	5.60	eee
HOBr	Hypobromous acid	-20.00	-13.95	-2.74	-4.73	hhh
COBr	COBr	20.50	-30.75	-28.80	-25.61	g
C ₂ H ₃ OBr	Acetyl bromide	-45.60	2.03	2.33	11.23	f
C ₇ H ₅ OBr	Benzoyl bromide	-11.60	3.02	0.58	12.09	f
CNBr	Cyanogen bromide	43.30	10.32	-3.31	-10.85	f
NOBr	Nitrosyl bromide	19.63	-13.04	-17.79	1.64	d
SBr	SBr	56.10	-7.83	-26.72	-25.61	g
SOBr	SOBr	-4.30	-11.92	2.17	-16.18	g
SO ₂ Br	SO ₂ Br	-52.80	9.92	89.44	11.02	g
FBr	Bromine fluoride	-14.00	-7.26	8.18	6.78	d
F ₃ Br	Bromine trifluoride	-61.10	13.97	83.96	82.57	d
CF ₃ Br	Bromotrifluoromethane	-155.10	-2.78	8.47	10.49	d
F ₅ Br	Bromine pentafluoride	-102.50	26.63	207.34	183.61	d
ClBr	Bromine chloride	3.50	-6.70	-12.99	-14.13	d
ClBr	Bromine chloride (+)	260.00	-12.36	5.80	-13.61	aa
CCl ₃ Br	Trichlorobromomethane	-9.30	-4.79	-5.83	-5.41	f
Br ₂	Bromine	7.40	-2.48	-9.08	-12.68	d
Br ₂	Bromine (+)	253.50	9.49	12.53	-7.73	eee
CBr ₂	Dibromomethylene	84.30	20.63	7.77	5.71	g
CH ₂ Br ₂	Dibromomethane	-3.53	11.38	-1.60	2.50	iii
COBr ₂	Carbonyl bromide	-20.10	-5.26	-11.54	2.22	f
SBr ₂	Sulfur dibromide	-3.00	27.85	2.76	0.86	jjj

Table 6.37: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
SOBr ₂	Thionyl bromide	-11.50	-7.15	16.12	-16.55	g
SO ₂ Br ₂	Sulfuryl bromide	-59.50	13.29	127.83	40.61	g
S ₂ Br ₂	S ₂ Br ₂	7.40	14.39	-5.90	-11.83	jjj
C ₂ F ₄ Br ₂	1,2-Dibromotetrafluoroethane	-189.00	-5.35	16.70	23.66	ggg
CBr ₃	Tribromomethyl	64.70	-0.48	-35.85	-37.81	g
CHBr ₃	Bromoform	4.40	13.13	-1.37	1.94	iii
SBr ₃	Sulfur tribromide	50.20	-33.82	-51.28	-76.36	g
SOBr ₃	SOBr ₃	-8.60	-1.46	41.38	-8.43	g
CBr ₄	Carbon tetrabromide	35.10	-2.21	-21.31	-19.21	g
SBr ₄	Sulfur tetrabromide	53.00	-34.27	-42.94	-88.51	g
SOBr ₄	SOBr ₄	-3.30	-18.35	26.54	-26.47	g
SBr ₅	Sulfur pentabromide	55.90	-11.91	23.50	-37.10	g
SBr ₆	Sulfur hexabromide	58.80	19.24	107.54	30.60	g
I	Iodide (-)	-46.50	-18.12	40.09	44.28	d
HI	Hydrogen iodide	6.30	22.50	9.44	1.63	d
CI	Iodomethylidyne	144.80	0.66	7.11	6.17	g
CH ₃ I	Methyl iodide	3.40	6.03	-1.51	2.25	f
C ₂ H ₅ I	Iodoethane	-2.00	4.06	-2.50	0.90	f
C ₃ H ₅ I	Allyl iodide	22.80	4.32	-3.32	-0.42	f
C ₃ H ₅ I	E-1-Iodo-1-propene	22.26	3.12	-7.83	-2.60	n
C ₃ H ₅ I	Z-1-Iodo-1-propene	20.66	8.54	-5.45	-0.21	n
C ₃ H ₇ I	1-Iodopropane	-7.10	4.11	-2.25	-0.78	f
C ₃ H ₇ I	2-Iodopropane	-9.80	4.51	2.26	4.06	f
C ₄ H ₉ I	1-Butyl iodide	-12.00	3.64	-2.09	-2.75	f
C ₄ H ₉ I	2-Iodo-2-methylpropane	-17.22	4.69	10.63	8.87	n
C ₆ H ₅ I	Iodobenzene	39.40	5.27	-7.00	-1.35	n
C ₆ H ₁₁ I	Iodocyclohexane	-11.94	0.17	-5.12	-8.13	n
C ₇ H ₇ I	o-Iodotoluene	31.74	7.03	-4.79	0.02	n
C ₇ H ₇ I	m-Iodotoluene	31.90	3.37	-7.20	-1.51	n
C ₇ H ₇ I	p-Iodotoluene	29.10	6.16	-4.49	1.17	f
C ₇ H ₇ I	Benzyl iodide	25.10	12.45	3.47	5.70	f
OI	IO	41.84	-10.90	4.82	-4.88	eee
COI	COI	63.50	-66.43	-62.47	-58.78	g
C ₂ H ₃ OI	Acetyl iodide	-30.20	0.23	3.21	9.42	n
C ₃ H ₅ OI	1-Iodo-2-propanone	-31.22	5.10	-1.63	0.13	n
C ₇ H ₅ OI	Benzoyl iodide	2.52	5.40	2.48	11.51	n
CNI	Cyanogen iodide	53.68	9.82	-14.14	-11.14	n
NOI	Nitrosyl iodide	26.80	-8.64	-5.90	5.42	d
SI	SI	73.10	-15.07	-26.46	-24.46	g
SOI	SOI	12.70	-13.11	6.95	-13.53	g
SO ₂ I	SO ₂ I	-34.90	7.65	64.62	10.36	g
FI	Iodine fluoride	-22.65	14.65	13.35	13.54	d
CF ₃ I	Trifluoroiodomethane	-140.49	2.52	12.27	7.82	n
F ₅ I	Iodine pentafluoride	-200.84	-2.10	331.69	295.27	d
F ₇ I	Iodine heptafluoride	-229.70	1.96	334.86	274.06	d
CIl	Iodine chloride	4.57	6.21	-11.46	-9.18	ddd

Table 6.38: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
BrI	Iodine bromide	9.77	5.88	-2.54	-3.77	d
I ₂	Iodine	14.92	5.82	6.30	4.91	d
Cl ₂	Diiodomethylene	120.40	1.16	-15.53	1.18	g
CH ₂ I ₂	Diiodomethane	27.00	6.49	-10.20	-5.52	eee
C ₂ H ₂ I ₂	E-1,2-Diiodoethene	49.56	5.39	-14.37	-5.52	n
C ₂ H ₂ I ₂	Z-1,2-Diiodoethene	49.56	13.19	-14.52	-6.09	n
C ₂ H ₄ I ₂	1,2-Diiodoethane	15.96	7.33	-4.46	-0.25	n
C ₃ H ₆ I ₂	1,2-Diiodopropane	8.50	12.24	-1.24	1.19	f
C ₄ H ₈ I ₂	1,2-Diiodobutane	2.70	13.38	3.61	3.49	f
C ₆ H ₄ I ₂	o-Diiodobenzene	60.16	13.57	-15.49	-6.20	n
COI ₂	Carbonyl iodide	9.60	-11.08	-15.15	-3.07	g
SI ₂	Sulfur diiodide	81.90	-30.33	-52.36	-56.22	g
SOI ₂	Thionyl iodide	21.50	-10.40	26.53	-11.32	g
SO ₂ I ₂	Sulfuryl iodide	-26.00	-4.06	51.60	-1.22	g
S ₂ I ₂	S ₂ I ₂	59.00	-13.02	-26.75	-30.81	g
Cl ₃	Triiodomethyl	117.30	-9.78	-68.03	-52.71	g
CHI ₃	Iodoform	50.40	10.21	-18.41	-12.60	iii
SI ₃	Sulfur triiodide	100.30	-45.04	-60.00	-83.97	g
SOI ₃	SOI ₃	40.40	-16.34	1.69	-33.55	g
Cl ₄	Carbon tetraiodide	108.20	-5.54	-61.38	-54.02	g
SI ₄	Sulfur tetraiodide	120.20	-50.98	-49.07	-102.72	g
SOI ₄	SOI ₄	60.00	-26.92	-17.15	-32.33	g
SI ₅	Sulfur pentaiodide	130.90	-1.25	-1.06	-68.65	g
SI ₆	Sulfur hexaiodide	158.90	8.10	48.67	-30.82	g
C ₂ LiH ₅	Ethyl-lithium	13.90		-25.58		f
C ₄ LiH ₉	n-Butyl-lithium	-6.00		-23.36		f
Be	Beryllium, atom	76.96	0.00	0.00	0.00	a
BeH	Beryllium hydride	75.60	7.17	-15.34	-10.43	aa
BeH	Beryllium hydride (+)	276.40	-23.49	-37.86	-38.02	d
BeH ₂	Beryllium dihydride	30.00	20.29	-24.41	-16.49	d
BeO	Beryllium oxide	32.60	20.44	5.97	37.90	d
BeHO	Beryllium hydroxide	-27.40	1.05	1.20	37.46	d
BeH ₂ O ₂	Beryllium di-hydroxide	-161.80	20.31	25.44	85.79	d
BeF	Beryllium fluoride	-41.80	-8.27	-10.92	18.16	aa
BeF ₂	Beryllium difluoride	-189.70	2.37	-2.22	54.79	aa
BeCl	Beryllium chloride	14.51	-8.81	-10.81	-13.97	d
BeCl ₂	Beryllium dichloride	-86.10	-0.39	-3.75	-13.66	d
BeBr	Beryllium bromide	28.70	-6.18	-0.15	-28.34	d
BeBr ₂	Beryllium dibromide	-54.80	1.33	11.77	-51.28	d
BeI	Beryllium iodide	40.60	-9.34	1.31	-1.03	d
BeI ₂	Beryllium diiodide	-15.30	0.88	-8.62	-16.06	d
Be ₂	Beryllium, dimer	152.30	-15.59	-26.89	-26.89	d
Be ₂ O	Beryllium oxide	-15.00	-4.74	40.50	80.34	d
Be ₂ OF ₂	Be(OF) ₂	-287.90	9.43	1.80	125.89	d
Be ₃ O ₃	Be ₃ O ₃	-251.90	-3.22	71.80	206.51	d
Mg	Magnesium, atom	35.00	0.00			a

Table 6.39: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
Mg	Magnesium (+)	213.10	4.76			aa
Mg	Magnesium (++)	561.30	-6.21			aa
MgH	Magnesium hydride	40.50	1.31			d
C ₁₀ MgH ₁₀	Dicyclopentadienyl magnesium	32.60	-1.13			f
MgO	Magnesium oxide	14.00	-1.95			d
MgHO	Magnesium hydroxide	-22.00	-33.05			mmm
MgH ₂ O ₂	Magnesium di-hydroxide	-135.00	8.24			d
MgN	Magnesium nitride	69.00	34.99			d
MgS	Magnesium sulfide	62.00	-19.64			d
MgF	Magnesium fluoride	-56.60	-14.68			d
MgF ₂	Magnesium difluoride	-173.00	12.31			aa
MgCl	Magnesium chloride	-10.30	-22.63			d
MgCl ₂	Magnesium dichloride	-93.80	11.40			d
MgBr	Magnesium bromide	-8.40	-13.35			d
MgBr ₂	Magnesium dibromide	-72.40	4.52			d
MgI ₂	Magnesium diiodide	-41.10	-3.48			aa
Mg ₂	Magnesium, dimer	68.77	1.23			aa
Mg ₂ F ₄	Magnesium difluoride, dimer	-410.70	2.28			d
Mg ₂ Cl ₄	Magnesium dichloride, dimer	-228.10	-4.00			d
Mg ₂ Br ₄	Magnesium dibromide, dimer	-183.50	2.33			d
BH	BH	107.40		-35.34	-25.10	nnn
BH	BH (+)	334.20		-30.29	-46.33	aa
BH ₂	BH ₂	45.70		20.74	30.04	ooo
BH ₃	Borane	23.90		-12.18	2.35	nnn
C ₃ BH ₉	Trimethylborane	-29.30		-10.79	4.48	nnn
C ₆ BH ₁₅	Triethylborane	-37.70		-13.45	-2.07	nnn
C ₉ BH ₂₁	Tripropylborane	-56.60		-6.96	-3.71	ppp
C ₁₂ BH ₂₇	Tributylborane	-68.60		-8.49	-10.56	f
BO	BO	6.00		-7.23	-5.19	nnn
BH ₃ O	H ₂ BOH	-69.40		-8.81	-4.02	ooo
CBH ₃ O	BH ₃ CO	-26.60		-19.58	15.49	nnn
BO ₂	BO ₂	-71.70		3.25	0.98	nnn
BHO ₂	HOBO	-134.30		1.43	-4.18	aa
BHO ₂	OBOH	-134.20		1.32	-4.28	nnn
BH ₃ O ₂	HB(OH) ₂	-153.10		-2.34	-10.62	ooo
C ₂ BH ₇ O ₂	(MeO) ₂ BH	-138.40		-3.46	-2.18	nnn
C ₆ BH ₇ O ₂	PhB(OH) ₂	-151.80		12.78	15.15	ooo
BH ₃ O ₃	B(OH) ₃	-237.60		1.27	-14.06	aa
C ₃ BH ₉ O ₃	Trimethoxyborane	-214.90		4.36	-0.97	nnn
C ₆ BH ₁₅ O ₃	Triethoxyborane	-239.50		14.87	6.02	nnn
BN	BN	114.00		25.29	18.01	d
C ₃ BH ₁₂ N	Trimethylborane-Ammonia adduct	-54.10		1.58	-0.46	nnn
C ₃ BH ₁₂ N	Trimethylamine-Borane adduct	-20.30		26.45	23.98	nnn
C ₆ BH ₁₈ N	Trimethylborane-Trimethylamine	-52.60		9.68	25.33	nnn
C ₆ BH ₁₈ N ₃	B(N(Me) ₂) ₃	-65.90		33.52	31.50	f
BS	Boron sulfide	58.10		13.97	19.48	jjj

Table 6.40: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
C ₂ BH ₉ S	(CH ₃) ₂ S-BH ₃	-10.60		-8.45	1.40	aa
C ₄ BH ₁₃ S	(CH ₃ CH ₂) ₂ S-BH ₃	-30.50		0.79	8.50	aa
BS ₂	BS ₂	28.90		-15.01	-1.71	aa
BF	BF	-29.20		-33.92	-27.36	nnn
BOF	FBO	-145.00		4.82	-4.25	nnn
BH ₂ O ₂ F	FB(OH) ₂	-249.50		7.97	-4.85	nnn
CBH ₃ F ₂	CH ₃ BF ₂	-199.00		7.53	6.13	nnn
C ₂ BH ₃ F ₂	F ₂ BCHCH ₂	-171.00		4.40	6.73	nnn
BOF ₂	F ₂ BO	-200.00		10.14	1.68	ddd
BHOF ₂	F ₂ BOH	-260.70		7.42	-4.27	nnn
BF ₃	Boron trifluoride	-271.60		10.60	-0.59	nnn
C ₂ BH ₆ OF ₃	(CH ₃) ₂ O-BF ₃	-328.20		15.46	-4.49	nnn
C ₄ BH ₁₀ OF ₃	(CH ₃ CH ₂) ₂ O-BF ₃	-343.80		21.35	0.16	aa
C ₃ BH ₉ NF ₃	Trimethylamine-Boron trifluoride	-304.40		40.18	32.12	aa
BCl	Boron chloride	35.70		-26.48	-22.21	nnn
BOCl	BOCl	-75.00		-17.04	-24.29	nnn
C ₂ BH ₆ O ₂ Cl	(MeO) ₂ BCl	-178.20		12.29	12.75	aa
BF ₂ Cl	BClF ₂	-212.70		11.07	4.16	nnn
BFCl ₂	BCl ₂ F	-154.10		9.99	5.11	nnn
BCl ₃	Boron trichloride	-96.40		8.94	-0.65	nnn
BBr	BBr	56.90		-23.22	-19.67	nnn
BF ₂ Br	BBrF ₂	-196.00		12.82	5.91	ddd
BBr ₃	Boron tribromide	-49.10		15.48	-0.98	nnn
BI ₃	Boron triiodide	17.00		-8.39	2.41	nnn
B ₂	Boron dimer	198.40		-5.23	8.75	nnn
B ₂ H ₄	B ₂ H ₄ D2d	45.70		63.43	80.87	qqq
B ₂ H ₄	B ₂ H ₄ (C2v)	45.60		29.53	47.57	qqq
B ₂ H ₆	Diborane	8.50		-10.35	-3.22	aa
B ₂ O ₂	BOBO	-108.60		-29.02	-24.56	nnn
B ₂ O ₂	O-B-B-O	-109.00		-28.62	-24.16	rrr
B ₂ O ₃	B ₂ O ₃	-201.50		2.59	-5.48	nnn
B ₂ H ₂ O ₃	B ₂ O ₃ H ₂	-200.40		57.48	34.16	sss
B ₂ H ₄ O ₄	B ₂ (OH) ₄	-315.00		7.09	-0.48	ooo
C ₂ B ₂ H ₁₁ N	B ₂ H ₅ N(CH ₃) ₂	-27.90		29.71	23.02	f
B ₂ S ₂	S-B-B-S	36.10		-35.38	-15.63	aa
B ₂ F ₄	B ₂ F ₄	-344.00		4.76	-10.03	nnn
B ₂ Cl ₄	B ₂ Cl ₄	-117.10		6.91	-3.03	nnn
B ₃ H ₃ O ₃	B ₃ O ₃ H ₃	-288.80		17.21	2.22	nnn
B ₃ H ₃ O ₆	B ₃ O ₃ (OH) ₃	-543.60		42.53	9.79	nnn
B ₃ H ₆ N ₃	Borazole	-122.20		-8.92	-19.97	nnn
B ₃ H ₂ O ₃ F	B ₃ O ₃ H ₂ F	-382.60		25.76	4.65	nnn
B ₃ HO ₃ F ₂	B ₃ O ₃ HF ₂	-475.80		34.23	7.08	nnn
B ₃ O ₃ F ₃	B ₃ O ₃ F ₃	-567.70		42.11	9.26	nnn
B ₃ H ₂ O ₃ Cl	B ₃ O ₃ H ₂ Cl	-313.80		14.68	0.15	nnn
B ₃ HO ₃ Cl ₂	B ₃ O ₃ HCl ₂	-338.70		12.63	-1.19	nnn
B ₃ O ₃ Cl ₃	B ₃ O ₃ Cl ₃	-390.10		37.61	24.87	nnn

Table 6.41: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
B ₃ H ₃ N ₃ Cl ₃	B3N3Cl3H3	-237.70		12.66	-20.75	nnn
Al	Aluminum, atom	79.49	-33.46	0.00	0.00	a
Al	Al (+)	218.10	61.65	-24.23	-4.77	d
HAl	AlH	62.00	8.06	-16.03	-9.11	d
C ₃ H ₉ Al	Trimethylaluminum	-20.90	15.18	-19.21	-6.58	f
C ₆ H ₁₅ Al	Triethylaluminum	-39.10	4.79	-16.71	-7.12	f
AIO	AIO	16.00	-23.89	-17.79	-7.60	d
AIO	AIO (+)	237.30	-34.94	-14.26	-2.69	d
AIO	AIO (-)	-64.40	13.56	8.62	11.54	d
HAIO	Al-O-H	-43.00	9.64	-18.10	2.93	d
HAIO	AIOH (+)	130.00	6.42	-8.44	27.95	d
HAIO	AIOH (-)	-55.00	-50.68	13.91	10.26	d
HAIO	H-Al=O	-8.00	10.19	-13.71	-13.89	d
AIO ₂	AIO2	-20.60	-16.79	6.18	16.70	d
AIO ₂	AIO2 (-)	-116.00	-1.09	25.51	20.53	d
HAIO ₂	AIO2H	-110.00	4.90	16.19	30.47	d
AlN	Aluminum nitride	125.00	-49.55	15.91	-7.09	d
AlS	Aluminum sulfide	57.10	-25.70	-4.52	-8.26	jjj
AlF	Aluminum fluoride	-63.50	13.37	-20.07	-14.37	d
AlF	AlF (+)	165.40	-5.03	-50.58	-38.41	d
AlOF	AlFO	-139.00	14.29	25.40	28.85	d
AlF ₂	AlF2	-166.00	3.26	-5.15	-3.01	d
AlF ₂	AlF2 (+)	22.00	7.04	-7.83	-3.62	d
AlF ₂	AlF2 (-)	-217.00	-12.60	7.42	6.13	d
AlOF ₂	AlF2O	-265.00	56.45	62.47	73.09	d
AlOF ₂	AlF2O (-)	-311.56	25.59	53.68	54.40	d
AlF ₃	Aluminum trifluoride	-289.00	-2.50	-2.34	3.15	d
AlF ₄	AlF4(-)	-476.00	6.82	35.35	48.30	d
AlCl	Aluminum chloride	-12.30	6.82	-15.55	-8.55	d
AlCl	Aluminum chloride (+)	206.00	-7.27	-31.22	-25.91	d
AlOCl	AlClO	-83.20	10.78	14.51	16.76	d
AlFCl	AlClF	-117.00	0.96	-7.76	-3.54	d
AlFCl	Aluminum chloride fluoride (+)	66.00	8.95	-2.50	-5.25	d
AlF ₂ Cl	AlClF2	-238.80	4.21	-2.86	2.60	d
AlCl ₂	Aluminum dichloride	-67.00	-1.21	-7.58	-7.01	d
AlCl ₂	AlCl2 (+)	115.00	9.33	3.92	-10.82	d
AlCl ₂	AlCl2 (-)	-115.00	-29.99	-15.61	-4.95	d
AlFCl ₂	AlCl2F	-189.00	11.02	-2.37	2.31	d
AlCl ₃	Aluminum trichloride	-139.72	17.63	-0.64	-0.59	d
AlBr ₃	Aluminum tribromide	-98.10	12.25	37.80	1.64	d
AlI	AlI	16.24	33.08	14.96	12.73	d
AlI ₃	Aluminum triiodide	-46.20	6.30	57.74	16.48	d
Al ₂	Al2	116.40	-35.71	14.94	2.30	d
Al ₂ O	Al2O	-34.70	6.12	-36.97	-4.46	d
Al ₂ O	Al2O (+)	155.90	12.54	-46.84	-7.56	d
Al ₂ O ₂	Al2O2	-94.30	6.74	-13.49	18.90	d

Table 6.42: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
Al ₂ O ₂	Al ₂ O ₂ (+)	135.60	-63.10	-56.69	-7.92	d
Al ₂ F ₆	Al ₂ F ₆	-629.45	-1.91	-2.12	-16.53	d
Al ₂ Cl ₆	Al ₂ Cl ₆	-309.70	-1.54	14.39	-9.26	d
Al ₂ Br ₆	Al ₂ Br ₆	-224.00	-0.92	91.33	-24.97	d
Al ₂ I ₆	Al ₂ I ₆	-117.00	-0.36	124.41	9.01	d
Ga	Gallium, atom	65.40	0.00			a
Ga	Gallium (-)	57.00	-14.40			d
Ga	Gallium (+)	205.90	-0.40			aa
Ga	Gallium (++)	680.50	18.43			aa
Ga	Gallium (+++)	1390.20	-2.78			aa
GaH	Gallium hydride	52.70	14.84			aa
C ₃ GaH ₉	Trimethyl gallium	-10.80	34.62			aa
C ₆ GaH ₁₅	Triethylgallium	-14.70	-19.27			ppp
C ₁₂ GaH ₂₇	Tributylgallium	-53.00	-20.62			ppp
GaO	Gallium oxide	66.80	-31.47			aa
GaHO	Gallium hydroxide	-27.40	-4.60			aa
GaF	Gallium fluoride	-60.21	14.04			aa
GaCl	Gallium chloride	-19.10	-2.09			aa
GaCl ₃	Gallium trichloride	-107.00	27.35			aa
GaH ₃ NCl ₃	GaCl ₃ -NH ₃	-150.10	21.96			aa
GaBr	Gallium bromide	-11.70	-19.86			aa
Gal	Gallium iodide	6.91	-19.76			aa
Ga ₂	Gallium, dimer	104.80	-22.24			aa
Ga ₂ S	Ga ₂ S	5.00	35.20			jjj
Ga ₂ Cl ₆	Ga ₂ Cl ₆	-233.10	-18.82			aa
Ga ₂ I ₆	Ga ₂ Br ₆	-76.00	-0.02			aa
In	Indium, atom	58.00	0.00			a
In	Indium (+)	187.30	13.16			aa
In	Indium (++)	624.00	28.94			aa
In	Indium (+++)	1272.00	-15.43			aa
InH	Indium hydride	51.50	0.92			cc
C ₃ InH ₉	Trimethyl indium	40.80	-42.95			f
InO	Indium oxide	92.50	-39.53			aa
InHO	Indium hydroxide	-18.90	-4.80			aa
InS	Indium sulfide	56.90	-2.66			aa
InF	Indium fluoride	-48.60	13.74			cc
InCl	Indium chloride	-18.00	4.68			cc
InCl	Indium chloride (+)	202.90	-0.45			aa
InCl ₃	Indium trichloride	-89.40	16.61			aa
InBr	Indium bromide	-13.60	-3.91			cc
InBr ₃	Indium tribromide	-67.40	7.49			aa
InI	Indium iodide	1.80	-9.00			cc
InI ₃	Indium triiodide	-28.80	10.57			aa
In ₂	Indium, dimer	91.00	-0.75			aa
In ₂ S	Indium sulfide	15.10	5.01			aa
In ₂ Cl ₃	In ₂ Cl ₃	-103.60	-20.38			aa

Table 6.43: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
Tl	Thallium, atom	43.55	0.00			a
Tl	Thallium (+)	185.90	19.15			aa
Tl	Thallium (++)	658.40	-1.53			aa
Tl	Thallium (+++)	1347.80	2.11			aa
TlCl	Thallium chloride	-16.20	2.77			aa
TlCl	Thallium chloride (+)	209.00	2.60			aa
TlBr	Thallium bromide	-9.00	-10.14			aa
TlBr	Thallium bromide (+)	203.30	11.48			aa
TlI	Thallium iodide	1.70	-14.37			aa
Tl ₂ Cl ₂	Thallium chloride dimer	-49.40	0.76			aa
Si	Silicon, atom	108.39	0.00	0.00	0.00	a
HSi	SiH	86.30	8.27	3.92	3.51	aa
H ₂ Si	Silylene (singlet)	61.10	11.69	3.22	6.73	ttt
H ₂ Si	Silylene (triplet)	6.50	-9.43	-2.74	-30.68	uuu
H ₃ Si	Silyl (-)	14.00	-16.80	32.45	-15.82	vvv
H ₃ Si	Silyl (+)	234.10	-10.81	-43.35	-11.21	vvv
H ₃ Si	Silyl	46.40	-3.47	-9.55	-20.08	www
H ₄ Si	Silane	8.20	4.27	-7.02	-4.08	d
CH ₅ Si	Methylsilyl	30.50	-7.27	-11.56	-21.52	www
CH ₆ Si	Methylsilane	-7.80	4.16	-5.86	-3.00	n
C ₂ H ₆ Si	Vinylsilane	-1.90	21.83	8.19	13.51	aa
C ₂ H ₇ Si	Dimethylsilyl	14.30	-10.42	-13.62	-21.81	www
C ₂ H ₈ Si	Ethylsilane	-15.00	4.79	-6.71	-1.52	xxx
C ₂ H ₈ Si	Dimethylsilane	-20.00	-0.79	-9.28	-5.68	n
C ₃ H ₉ Si	Trimethylsilyl	-0.80	-14.25	-16.93	-22.20	www
C ₃ H ₁₀ Si	Trimethylsilane	-37.40	0.10	-6.99	-1.98	n
C ₄ H ₁₂ Si	Diethylsilane	-43.60	10.94	-0.52	7.19	oo
C ₄ H ₁₂ Si	Tetramethylsilane	-55.70	1.95	-3.48	3.54	yyy
C ₅ H ₁₂ Si	1,1-Dimethylsilacyclobutane	-33.70	-1.95	-14.75	-3.54	n
C ₆ H ₁₆ Si	Triethylsilane	-39.50	-15.98	-24.63	-15.29	oo
C ₈ H ₂₀ Si	Tetraethylsilane	-64.40	-13.60	-17.73	-7.53	oo
SiO	Silicon monoxide	-23.90	-2.13	1.26	21.82	g
C ₃ H ₁₀ SiO	Trimethylsilicon hydroxide	-119.40	3.90	-2.55	9.06	f
SiO ₂	Silicon dioxide	-73.00	-15.93	50.06	5.51	d
SiF	Silicon fluoride	1.70	-22.57	-30.47	-27.81	aa
H ₃ SiF	Fluorosilane	-84.50	7.09	-11.95	-1.31	zzz
C ₄ H ₁₂ SiF	SiMe ₄ F (-) C _{3v} symmetry	-147.50	25.02	33.75	24.07	aaa
SiOF	SiOF	-136.10	28.55	56.09	30.12	g
SiF ₂	Silicon difluoride	-141.20	-13.69	-23.70	-13.44	aaaa
H ₂ SiF ₂	Difluorosilane	-189.00	13.82	-3.55	7.52	aaaa
SiOF ₂	SiOF ₂	-231.00	1.63	42.33	10.26	d
SiF ₃	Trifluorosilyl	-245.00	-15.33	-3.88	-22.78	www
HSiF ₃	Trifluorosilane	-287.00	6.82	1.91	6.09	d
SiF ₄	Silicon tetrafluoride	-386.00	-4.58	15.55	3.99	aaaa
SiF ₅	SiF ₅ (-)	-507.10	2.73	17.63	3.59	aaa
SiCl	Silicon chloride	45.30	-15.54	-15.64	-18.03	aa

Table 6.44: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
H ₃ SiCl	Chlorosilane	-32.40	4.77	-11.55	-5.29	aaaa
C ₂ H ₇ SiCl	Chlorodimethylsilane	-69.90	6.14	-3.17	-0.11	oo
C ₃ H ₉ SiCl	Chlorotrimethylsilane	-84.60	2.98	-2.84	-0.25	f
SiOCl	SiOCl	-86.70	27.67	44.15	28.57	g
SiF ₄ Cl	SiF ₄ Cl (-)	-465.30	0.08	17.46	7.96	aaa
SiCl ₂	Silicon dichloride	-40.60	-8.85	-5.71	-6.10	aaaa
H ₂ SiCl ₂	Dichlorosilane	-75.30	5.79	-8.18	-5.44	aaaa
CH ₄ SiCl ₂	Dichloromethylsilane	-96.00	7.08	-1.45	-2.49	f
C ₂ H ₆ SiCl ₂	Dichlorodimethylsilane	-109.50	1.48	-1.76	-5.22	n
SiOCl ₂	SiOCl ₂	-167.70	46.47	73.57	49.35	g
SiCl ₃	Trichlorosilyl	-76.00	-18.86	-13.50	-39.98	www
HSiCl ₃	Trichlorosilane	-119.30	6.58	1.42	-5.87	aaaa
CH ₃ SiCl ₃	Trichloromethylsilane	-131.20	-1.68	0.56	-12.49	n
SiCl ₄	Silicon tetrachloride	-158.40	1.97	10.76	-12.52	d
SiCl ₅	SiCl ₅ (-)	-237.20	-17.47	-19.33	-30.42	aaa
SiBr	Silicon bromide	50.00	-9.03	7.81	-3.02	aa
H ₃ SiBr	Bromosilane	-15.30	-0.73	-2.75	-5.83	bbbb
C ₃ H ₉ SiBr	Trimethylbromosilane	-70.00	1.22	7.60	3.51	f
SiOBr	SiOBr	-71.40	24.24	42.77	31.06	g
SiBr ₂	Silicon dibromide	-9.60	-17.84	20.81	3.63	aaaa
H ₂ SiBr ₂	Dibromosilane	-43.20	-3.98	11.13	-2.62	aaaa
SiOBr ₂	SiOBr ₂	-137.40	43.31	85.94	60.01	g
SiBr ₃	Silicon tribromide	-56.10	-4.67	41.12	-9.60	g
HSiBr ₃	Tribromosilane	-72.50	-7.06	30.08	2.27	bbbb
SiBr ₄	Silicon tetrabromide	-99.30	-8.65	48.91	4.81	aa
SiI	Silicon iodide	76.40	-5.44	15.53	4.11	g
H ₃ SiI	Iodosilane	-0.50	1.34	11.52	5.40	aaaa
SiOI	SiOI	-53.30	-19.59	56.17	76.73	g
SiI ₂	Silicon diiodide	22.00	-3.98	60.11	35.53	aaaa
H ₂ SiI ₂	Diiodosilane	-9.10	5.55	32.05	14.06	d
SiOI ₂	SiOI ₂	-99.40	49.17	89.66	80.71	g
SiI ₃	Silicon triiodide	0.50	6.58	44.85	11.84	g
HSiI ₃	Triiodosilane	-17.80	8.78	51.86	21.69	aaaa
SiI ₄	Silicon tetraiodide	-26.40	12.14	68.74	27.99	d
Si ₂	Silicon dimer	140.90	-5.17	78.91	-1.40	g
Si ₂	Silicon dimer	140.90	-5.17	78.91	-1.40	g
H ₆ Si ₂	Disilane	17.10	0.75	5.28	-0.96	cccc
C ₆ H ₁₈ Si ₂	Hexamethyldisilane	-85.80	2.58	12.00	19.58	n
C ₆ H ₁₈ Si ₂ O	Hexamethyldisiloxane	-185.80	1.36	-8.80	17.82	f
C ₆ H ₁₉ Si ₂ N	Hexamethyldisilazane	-113.90	-7.02	-9.80	-4.17	f
Si ₂ Cl ₆	Hexachlorodisilane	-243.50	13.75	30.78	1.88	cccc
Si ₂ Br ₆	Hexabromodisilane	-182.80	18.10	116.84	38.91	g
Si ₃	Silicon trimer	152.20	0.56	14.94	32.68	g
H ₈ Si ₃	Si ₃ H ₈	28.90	-6.58	3.27	-3.99	aa
Ge	Germanium, atom	89.50	-27.71	0.00	0.00	a
GeH ₄	Germane	21.70	10.36	-8.25	7.32	aa

Table 6.45: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
CGe	Germanium carbide	151.00	27.61	73.43	52.55	aa
C ₃ GeH ₉	Trimethylgermanium	2.20	-14.31	-3.96	-3.52	dddd
C ₃ GeH ₉	Trimethylgermanium (+)	165.00	11.41	-2.37	15.44	dddd
C ₄ GeH ₁₂	Tetramethylgermanium	-32.00	-2.33	-0.82	5.77	dddd
C ₇ GeH ₁₈	t-Butyltrimethylgermane	-55.70	4.35	16.59	9.66	dddd
C ₈ GeH ₂₀	Tetraethylgermanium	-38.60	-17.40	-18.72	-21.93	ppp
C ₁₂ GeH ₂₈	Tetrapropylgermanium	-54.20	-24.25	-21.40	-33.37	ppp
C ₂₄ GeH ₂₀	Tetraphenylgermanium	106.50	2.67	-21.55	-15.34	ppp
GeO	Germanium oxide	-11.04	5.12	11.85	20.24	aa
C ₅ GeH ₁₄ O	Ethoxytrimethylgermane	-87.80	-3.55	-0.60	-2.28	dddd
C ₄ GeH ₁₂ O ₄	Tetramethoxygermanium	-230.40	-15.58	-3.23	-22.38	oo
C ₅ GeH ₁₅ N	Dimethylaminetrimethylgermane	-29.10	-12.42	-3.99	1.13	dddd
GeF	Germanium fluoride	-8.00	4.66	-8.40	-11.74	aa
GeF ₂	Germanium difluoride	-121.00	13.42	-2.85	0.44	aa
GeF ₄	Germanium tetrafluoride	-284.40	-5.85	10.14	21.04	eeee
GeCl	Germanium chloride	37.10	-9.42	-19.66	-17.46	aa
C ₃ GeH ₉ Cl	Trimethylchlorogermane	-63.80	3.65	1.70	5.11	dddd
GeCl ₂	Germanium dichloride	-42.00	10.20	7.45	-1.99	aa
GeCl ₄	Germanium tetrachloride	-118.50	-9.18	-7.27	-20.20	aa
GeCl ₄	Germanium tetrachloride (+)	156.80	-23.99	34.57	4.27	aa
GeBr	Germanium bromide	56.30	-14.02	-13.38	-13.06	aa
C ₃ GeH ₉ Br	Trimethylbromogermane	-53.10	5.27	14.82	12.03	dddd
GeBr ₂	Germanium dibromide	-15.00	10.92	16.33	5.50	aa
GeBr ₄	Germanium tetrabromide	-71.70	5.02	37.70	2.23	aa
GeI ₂	Germanium diiodide	11.20	16.81	49.16	21.63	aa
GeI ₄	Germanium tetraiodide	-13.60	6.88	70.71	17.24	aa
Ge ₂	Germanium, dimer	113.10	-17.81	77.58	-30.34	aa
Ge ₂ H ₆	Digermane	38.80	6.78	8.11	0.32	aa
C ₆ Ge ₂ H ₁₈	GeMe ₃ -GeMe ₃	-62.50	10.43	38.49	21.23	ffff
C ₆ Ge ₂ H ₁₈ O	Bis(trimethylgermanium) oxide	-127.00	4.42	8.80	15.64	dddd
Ge ₃ H ₈	Trigermane	54.20	4.20	21.02	-12.48	aa
Sn	Tin, atom	72.20	0.00	0.00		a
SnH	Tin hydride	49.00	14.25	0.71		gggg
SnH ₄	Tin tetrahydride (stannane)	38.90	-3.42	-8.40		aa
C ₂ SnH ₈	Dimethylstannane	21.00	-8.67	-15.24		aa
C ₃ SnH ₁₀	Trimethyltin hydride	5.00	-4.90	-10.84		aa
C ₄ SnH ₁₂	Diethyltin dihydride	11.00	-6.23	-13.82		aa
C ₄ SnH ₁₂	Tetramethyltin	-4.50	-8.35	-12.38		aa
C ₅ SnH ₁₄	Ethyltrimethyltin	-7.00	-9.49	-13.82		ppp
C ₆ SnH ₁₆	Propyltrimethyltin	-11.20	-10.42	-14.13		ppp
C ₇ SnH ₁₈	t-Butyltrimethyltin	-16.00	-8.22	0.02		ppp
C ₈ SnH ₂₀	Tetraethyltin	-10.90	-16.43	-21.12		aa
C ₉ SnH ₁₄	Phenyltrimethylstannane	26.70	-3.33	-6.80		ppp
C ₂₄ SnH ₂₀	Tetraphenyltin	114.39	14.36	17.61		f
SnO	Tin oxide	3.61	0.92	9.61		aa
C ₅ SnH ₁₅ N	Dimethylaminetrimethyltin	-4.30	-4.90	7.03		dddd

Table 6.46: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference		AM1	Ref.
			PM3	MNDO		
SnS	Tin sulfide	28.50	9.23	-1.42		aa
SnF	Tin fluoride	-9.00	-8.48	-11.43		gggg
SnF ₂	Tin difluoride	-116.00	4.56	1.36		gggg
SnCl	Tin chloride	16.00	-8.43	-21.55		gggg
C ₃ SnH ₉ Cl	Trimethyltin chloride	-46.40	6.45	2.25		dddd
C ₆ SnH ₁₅ Cl	Triethyltin chloride	-46.20	-4.25	-10.28		ppp
SnCl ₂	Tin dichloride	-56.40	0.88	-25.75		gggg
C ₂ SnH ₆ Cl ₂	Dimethyltin dichloride	-71.00	19.30	4.78		hhhh
CSnH ₃ Cl ₃	Methyltin trichloride	-99.80	16.16	11.60		ppp
C ₂ SnH ₅ Cl ₃	Ethyltin trichloride	-102.00	14.35	8.64		ppp
SnCl ₄	Tin tetrachloride	-112.70	11.47	12.67		aa
SnBr	Tin bromide	24.00	-9.45	-7.85		gggg
C ₃ SnH ₉ Br	Trimethyltin bromide	-33.60	-1.42	8.60		iii
C ₁₂ SnH ₂₇ B	Tri-n-butyltin bromide	-65.70	-11.12	0.69		f
SnBr ₂	Tin dibromide	-29.00	-16.75	-9.20		gggg
SnBr ₄	Tin tetrabromide	-75.20	14.93	44.67		aa
SnI	Tin iodide	36.00	-6.57	0.43		gggg
C ₃ SnH ₉ I	Trimethyltin iodide	-19.70	-0.83	10.86		dddd
SnI ₂	Tin diiodide	2.00	-17.23	3.40		gggg
C ₆ GeSnH ₁₈	GeMe ₃ -SnMe ₃	-39.70	7.48	10.05		dddd
Pb	Lead, atom	46.62	0.00	0.00		a
Pb	Lead (+)	219.00	-1.09	-0.64		aa
PbH	Lead hydride	56.50	-2.32	-15.77		ddd
PbH ₄	Lead tetrahydride (plumbane)	59.70	-0.89	3.39		jjjj
C ₃ PbH ₉	Trimethyllead	46.66	15.35	-10.64		ffff
C ₃ PbH ₉	Trimethyllead (+)	200.07	9.83	11.79		ffff
C ₄ PbH ₁₂	Tetramethyllead	32.50	-2.60	-6.45		aa
C ₇ PbH ₁₈	t-Butyltrimethyllead	6.92	7.04	19.50		ffff
C ₈ PbH ₂₀	Tetraethyllead	26.19	-17.64	-15.64		aa
PbO	Lead oxide	16.80	4.51	12.31		x
PbS	Lead sulfide	31.50	8.40	0.46		d
PbF	Lead fluoride	-19.18	-1.78	-3.40		d
PbF ₂	Lead difluoride	-104.00	14.35	9.83		d
PbCl	Lead chloride	3.60	-1.95	-18.89		d
PbCl	Lead chloride (+)	178.20	-3.79	-18.44		d
PbCl ₂	Lead dichloride	-41.60	1.86	-35.40		d
PbCl ₂	Lead dichloride (+)	195.10	-15.62	-6.18		d
PbCl ₄	Lead tetrachloride (est.)	-69.20	7.32	19.19		aa
PbBr	Lead bromide	16.95	-10.00	-10.94		d
PbBr ₂	Lead dibromide	-24.95	-5.75	-8.91		d
PbI	Lead iodide	25.75	-12.30	0.35		d
PbI ₂	Lead diiodide	-0.76	-16.45	9.49		d
Pb ₂	Lead, dimer	79.50	-34.17	-6.95		d
C ₆ Pb ₂ H ₁₈	Hexamethyldiplumbane	38.71	10.90	6.67		ffff
P	Phosphorus (++)	775.10	-142.10	-62.55	-63.01	aa
HP	Phosphinidene	60.60	12.83	27.87	15.04	d

Table 6.47: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
H ₂ P	Phosphino	30.10	-0.80	1.32	4.07	d
H ₃ P	Phosphine	1.30	-1.08	2.63	8.88	d
CP	Carbon phosphide	107.53	12.09	22.12	23.58	d
CHP	Methinophosphine	35.80	10.64	6.40	11.57	d
CH ₅ P	Methylphosphine	-7.00	-2.48	-7.73	5.80	kkkk
C ₂ H ₇ P	Ethylphosphine	-12.00	0.30	-9.12	8.10	kkkk
C ₂ H ₇ P	Dimethylphosphine	-15.00	-4.38	-16.55	3.33	kkkk
C ₃ H ₉ P	Trimethylphosphine	-22.50	-7.36	-25.78	0.46	f
C ₄ H ₁₁ P	Diethylphosphine	-25.00	1.64	-20.08	6.62	kkkk
C ₆ H ₁₅ P	Triethylphosphine	-11.80	-24.95	-53.06	-18.30	f
PO	Phosphorus oxide	-2.90	-16.55	-18.16	-13.58	g
C ₃ H ₉ PO	Trimethylphosphine oxide	-102.20	19.41	59.25	0.59	f
PO ₂	Phosphorus dioxide	-71.00	-5.71	23.93	13.18	g
PO ₃	PO ₃ (-)	-190.00	-7.23	49.56	-13.09	llll
HPO ₃	HPO ₃	-135.00	-42.27	29.84	-33.62	llll
CH ₅ PO ₃	Methylphosphonic acid	-240.50	31.46	88.41	13.74	f
C ₂ H ₇ PO ₃	Ethylphosphonic acid	-239.40	29.75	80.23	10.55	f
C ₃ H ₉ PO ₃	Trimethyl phosphite	-168.30	-14.58	-34.63	-15.01	f
C ₆ H ₁₅ PO ₃	Triethyl phosphite	-195.90	-12.72	-25.58	-9.19	f
C ₆ H ₁₅ PO ₄	Triethyl phosphate	-284.50	32.94	74.91	8.13	f
NP	Phosphorus nitride	25.04	7.86	8.87	7.46	d
PF	Phosphorus fluoride	-20.80	0.37	10.88	1.10	g
POF	POF	-111.80	-12.17	-2.42	0.51	g
C ₂ H ₆ PO ₂ F	Methyl methylphosphonofluoridate	-197.30	-10.32	51.33	-20.81	mmmm
C ₃ H ₈ PO ₂ F	Ethyl methylphosphonofluoridate	-205.80	-6.57	53.80	-18.61	mmmm
C ₄ H ₁₀ PO ₂ F	i-Propyl methylphosphonofluoridate	-214.60	0.71	63.03	-8.76	mmmm
C ₄ H ₁₀ PO ₂ F	n-Propyl methylphosphonofluoridate	-210.21	-7.26	53.80	-20.95	mmmm
C ₅ H ₁₂ PO ₂ F	i-Propyl ethylphosphonofluoridate	-219.75	2.13	59.27	-12.24	mmmm
C ₅ H ₁₂ PO ₂ F	s-Butyl methylphosphonofluoridate	-220.10	-1.15	63.08	-15.66	mmmm
C ₅ H ₁₂ PO ₂ F	n-Butyl methylphosphonofluoridate	-215.14	-7.67	54.02	-22.83	mmmm
C ₆ H ₁₄ PO ₂ F	Neopentyl methylphosphonofluoridate	-224.16	-2.51	70.28	-13.87	mmmm
PF ₂	Phosphorus difluoride	-119.00	-25.44	-19.00	-14.24	g
POF ₂	POF ₂	-213.60	24.67	61.67	39.37	g
CH ₃ POF ₂	Methylphosphonodifluoride	-233.23	7.90	86.01	5.55	mmmm
PF ₃	Phosphorus trifluoride	-229.00	-23.17	-0.32	0.01	d
POF ₃	Phosphorus oxyfluoride	-289.50	-8.21	89.96	-3.31	aa
PF ₄	Phosphorus tetrafluoride	-287.90	-16.05	53.22	6.78	g
PF ₄	Phosphorus tetrafluoride (-)	-325.00	-28.60	18.07	26.20	aaa
PF ₅	Phosphorus pentafluoride	-381.10	-5.83	132.28	1.70	d
PF ₆	Phosphorus hexafluoride (-)	-522.00	13.46	152.31	17.53	aaa
PCl	Phosphorus chloride	25.60	3.32	10.79	-4.29	g
POCl	POCl	-64.70	-11.71	-10.11	-0.41	g
PCl ₂	Phosphorus dichloride	-21.30	-18.80	-28.08	-26.29	g
POCl ₂	POCl ₂	-109.90	15.05	33.91	24.79	g
CH ₃ POCl ₂	Methylphosphonodichloride	-124.10	-4.79	48.66	9.27	mmmm
PCl ₃	Phosphorus trichloride	-69.00	-19.48	-27.45	-20.06	d

Table 6.48: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
POCl ₃	Phosphorus oxychloride	-132.80	-7.40	53.17	14.82	aa
PSCl ₃	Phosphorus thiochloride	-91.00	30.77	62.61	58.69	d
PCl ₄	Phosphorus tetrachloride	-80.50	-22.33	-22.75	-29.57	g
PCl ₅	Phosphorus pentachloride	-86.10	-25.51	44.13	14.27	d
PBr	Phosphorus bromide	43.00	-8.14	-13.58	-15.41	ddd
POBr	POBr	-50.20	-9.70	-3.70	10.62	g
PBr ₂	Phosphorus dibromide	6.70	-6.27	-15.98	-11.11	g
POBr ₂	POBr ₂	-78.30	14.94	34.09	38.98	g
PBr ₃	Phosphorus tribromide	-34.90	6.73	-3.23	11.53	d
POBr ₃	Phosphorus oxybromide	-97.00	16.75	68.32	67.95	d
PSBr ₃	Phosphorus thiobromide	-67.20	59.36	85.63	111.94	d
PBr ₄	Phosphorus tetrabromide	-17.40	-16.14	-21.64	-8.52	g
PBr ₅	Phosphorus pentabromide	-11.00	-16.11	42.74	63.99	g
PI	Phosphorus iodide	54.60	-3.58	-0.85	-3.15	g
POI	POI	-33.40	-10.23	8.79	17.69	g
PI ₂	Phosphorus diiodide	41.30	-4.74	-4.79	-6.36	g
POI ₂	POI ₂	-40.10	4.90	19.49	33.02	g
PI ₃	Phosphorus triiodide	25.10	8.01	1.18	0.72	g
POI ₃	Phosphorus oxyiodide	-39.70	31.42	71.31	75.86	g
PI ₄	Phosphorus tetraiodide	60.20	-15.53	-36.12	-33.68	g
PI ₅	Phosphorus pentaoidide	97.70	-0.37	-16.01	-42.18	g
C ₆ BH ₁₈ P	Trimethylborane-Trimethylphosphine	-68.30		-20.22	-4.15	aa
BH ₃ PF ₃	Phosphorus trifluoride-Borane	-204.10		-13.16	-30.50	aa
C ₃ BH ₉ PF ₃	Trimethylphosphine-Boron trifluoride	-312.60		2.34	30.85	aa
P ₂	Phosphorus dimer	42.80	-10.81	-1.74	-18.23	g
H ₄ P ₂	P ₂ H ₄	5.00	-8.72	-7.88	1.58	aa
P ₄	Phosphorus tetramer	31.10	11.77	5.11	19.28	g
P ₄ O ₆	Phosphorus trioxide	-529.23	18.15	7.98	207.56	d
P ₄ O ₁₀	Phosphorus pentoxide	-694.09	-18.66	262.21	-7.10	d
As	Arsenic, atom	72.30	0.00			a
As	Arsenic (++)	731.30	-11.02			aa
AsH ₃	Arsine	15.88	-3.21			aa
C ₃ AsH ₉	Trimethylarsine	2.80	-17.70			f
C ₆ AsH ₁₅	Triethylarsine	13.40	-33.84			f
C ₁₈ AsH ₁₅	Triphenylarsine	97.60	10.11			ppp
C ₃ AsH ₉ O ₃	Trimethyl arsenite	-131.10	-8.24			f
C ₆ AsH ₁₅ O ₃	Triethyl arsenite	-156.80	1.52			f
C ₉ AsH ₂₁ O ₃	Tri-n-propyl arsenite	-173.70	2.53			f
AsN	Arsenic nitride	46.91	23.23			aa
AsS	Arsenic sulfide	48.50	-2.10			jjj
AsF ₃	Arsenic trifluoride	-187.80	-2.40			aa
AsF ₅	Arsenic pentafluoride	-295.60	7.77			jjj
AsCl	Arsenic chloride	10.50	5.18			g
AsCl ₃	Arsenic trichloride	-62.50	-7.71			aa
AsBr ₃	Arsenic tribromide	-31.57	11.30			jjj
AsI ₃	Arsenic triiodide	6.90	20.38			g

Table 6.49: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f Exp.	Difference			Ref.
			PM3	MNDO	AM1	
As ₂	Arsenic, dimer	53.11	-7.46			aa
As ₃	Arsenic, trimer	62.48	6.53			jjj
As ₄	Arsenic, tetramer	34.39	0.05			aa
As ₄ O ₆	Arsenic trioxide	-289.00	1.11			aa
Zn	Zinc, atom	31.17	0.00	0.00	0.00	a
Zn	Zinc (+)	249.46	-14.09	-9.83	-5.21	aa
Zn	Zinc (++)	665.10	-2.37	55.11	64.34	aa
ZnH	Zinc hydride	54.40	2.48	-11.24	-3.45	a
CZnH ₃	Methylzinc (+)	215.10	13.77	12.59	9.33	aa
C ₂ ZnH ₆	Dimethylzinc	12.67	-4.49	7.18	7.14	aa
C ₂ ZnH ₆	Dimethylzinc (+)	221.70	12.28	31.49	11.85	aa
C ₄ ZnH ₁₀	Diethylzinc	12.10	-6.53	0.83	2.01	aa
C ₆ ZnH ₁₄	Di-n-propylzinc	-2.90	1.38	7.39	5.27	f
C ₈ ZnH ₁₈	Di-n-butylzinc	-11.90	0.20	7.10	0.67	f
ZnS	Zinc sulfide	48.30	3.18	30.71	30.04	jjj
ZnCl ₂	Zinc dichloride	-63.60	10.77	14.89	9.02	aa
ZnCl ₂	Zinc dichloride (+)	207.70	0.09	28.58	11.72	aa
Cd	Cadmium, atom	26.72	0.00			a
C ₂ CdH ₆	Dimethylcadmium	25.80	4.77			f
C ₂ CdH ₆	Dimethylcadmium (+)	223.20	-5.06			aa
C ₄ CdH ₁₀	Diethylcadmium	25.50	0.37			f
Hg	Mercury, atom	14.69	0.00	0.00	0.00	a
Hg	Mercury (+)	256.83	14.70	-34.37	-31.33	aa
Hg	Mercury (++)	690.84	-9.71	-11.56	-5.48	aa
HgH	Mercury hydride	57.20	-9.14	-19.42	-1.27	aa
CHgH ₃	Methylmercury	40.00	5.88	-11.08	-1.86	aa
C ₂ HgH ₆	Dimethylmercury	22.30	6.07	-12.14	5.07	aa
C ₂ HgH ₆	Dimethylmercury (+)	233.90	6.09	-4.07	-4.50	aa
C ₄ HgH ₁₀	Diethylmercury	17.80	-2.07	-5.69	1.93	aa
C ₆ HgH ₁₄	Diisopropylmercury	9.70	-7.06	13.11	6.06	f
C ₆ HgH ₁₄	Di-n-propylmercury	8.20	-2.66	-3.40	-0.48	f
C ₆ HgH ₁₄	Di-n-propylmercury	8.20	-2.66	-3.40	-0.48	f
C ₈ HgH ₁₈	Di-n-butylmercury	-7.80	2.64	3.38	1.89	f
C ₈ HgH ₁₈	Diisobutylmercury	-9.20	-1.14	19.31	9.80	f
C ₁₂ HgH ₁₀	Diphenylmercury	93.80	4.48	7.62	7.24	f
C ₂ HgN ₂	Mercuric cyanide	91.00	2.08	-21.94	-6.77	aa
HgF	Mercury fluoride	1.00	-8.24	-25.47	-11.68	aa
HgF ₂	Mercury difluoride	-70.20	19.15	3.33	16.01	ddd
HgCl	Mercury chloride	20.10	-15.22	-26.40	-22.77	aa
CHgH ₃ Cl	Methylmercuric chloride	-12.50	9.10	-5.51	-0.04	aa
C ₂ HgH ₅ Cl	Ethylmercuric chloride	-15.00	5.15	-1.76	-1.40	aa
HgCl ₂	Mercury dichloride	-34.97	2.30	-1.97	-9.87	ddd
HgBr	Mercury bromide	24.90	-14.19	-10.12	-28.40	ddd
CHgH ₃ Br	Methylmercuric bromide	-4.40	3.32	7.18	-6.61	aa
C ₂ HgH ₅ Br	Ethylmercuric bromide	-7.20	-0.57	11.13	-7.74	aa
HgBr ₂	Mercury dibromide	-20.40	-6.54	23.49	-23.55	ddd

Table 6.50: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f		Difference		Ref.
		Exp.	PM3	MNDO	AM1	
CHgH ₃ I	Methylmercuric iodide	5.21	4.28	8.27	15.94	aa
C ₂ HgH ₅ I	Ethylmercuric iodide	3.30	-0.46	11.11	13.80	aa
HgI ₂	Mercury diiodide	-4.10	-0.13	25.21	23.20	aa
Hg ₂	Mercury, dimer	26.00	-0.11	4.22	2.04	aa
Hg ₂	Mercury, dimer (+)	244.30	-6.86	-27.62	-23.63	aa
Sb	Antimony, atom	63.20	0.00			jjj
Sb	Antimony (+)	263.50	8.49			aa
Sb	Antimony (++)	646.00	-29.16			aa
SbH ₃	Stibine	34.70	23.37			aa
C ₃ SbH ₉	Trimethylstibine	7.70	-8.08			ppp
C ₆ SbH ₁₅	Triethylstibine	10.40	-33.48			ppp
C ₁₈ SbH ₁₅	Triphenylstibine	104.10	-5.75			ppp
SbO	Antimony oxide	47.70	-11.77			aa
SbN	Antimony nitride	63.70	38.82			aa
SbF	Antimony fluoride	-11.30	10.62			aa
SbCl	Antimony chloride	-6.20	19.51			aa
SbOCl	Antimony oxychloride	-25.50	18.39			aa
SbCl ₂	Antimony dichloride	-18.50	-15.82			aa
SbCl ₃	Antimony trichloride	-75.00	2.55			aa
SbCl ₅	Antimony pentachloride	-94.20	2.56			aa
SbBr ₃	Antimony tribromide	-46.50	9.99			aa
InSb	Indium antimonide	82.30	-2.22			aa
Sb ₂	Antimony, dimer	56.30	14.14			aa
InSb ₂	Indium diantimonide	75.00	-16.43			aa
Sb ₄	Antimony, tetramer	49.00	2.46			aa
Se	Selenium, atom	54.30	0.00			a
SeH ₂	Hydrogen selenide	7.10	15.62			aa
C ₄ SeH ₁₀	Diethylselenium	-13.70	-23.69			f
SeO	Selenium oxide	14.80	5.55			g
SeO ₂	Selenium dioxide	-53.90	24.21			cc
SeOF ₂	Seleninyl difluoride	-121.60	-12.65			g
SeF ₄	Selenium tetrafluoride	-194.00	-15.49			g
SeF ₆	Selenium hexafluoride	-267.00	0.23			cc
CSe ₂	Carbon diselenide	61.40	3.36			f
GeTe	Germanium telluride	42.00	25.53			cc
SnTe	Tin telluride	38.40	15.16			cc
AsTe	Arsenic telluride	54.70	6.39			jjj
ZnTe	Zinc telluride	61.00	-12.74			jjj
Te ₂	Tellurium, dimer	38.00	13.31			jjj
Te ₂ O ₂	Tellurium oxide dimer	-26.00	-18.47			jjj
Te ₂ F ₁₀	Te ₂ F ₁₀	-554.40	10.58			jjj
Bi	Bismuth, atom	50.10	0.00			a
Bi	Bismuth (+)	219.10	18.81			aa
Bi	Bismuth (++)	606.10	-17.27			aa
C ₃ BiH ₉	Trimethylbismuth	46.10	-3.07			f
C ₆ BiH ₁₅	Triethylbismuth	51.60	-26.08			f
C ₁₈ BiH ₁₅	Triphenylbismuth	138.60	-2.92			f
BiF	Bismuth fluoride	-7.02	14.10			jjj

Table 6.51: Comparison of Calculated and Observed ΔH_f for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	ΔH_f		Difference		
		Exp.	PM3	MNDO	AM1	Ref.
BiCl	Bismuth chloride	6.00	11.76			jjj
BiCl ₃	Bismuth trichloride	-63.50	20.87			aa
BiSe	Bismuth selenide	42.00	12.55			aa
Bi ₂	Bismuth, dimer	52.50	2.68			aa

References

- a: "CRC Handbook of Chemistry and Physics", 60th Edition, R. C. Weast, (Ed.), CRC Press, Boca Raton, FL, 1980.
- b: Element in its standard state.
- c: R. D. Levin, S. G. Lias, "Ionization Potentials and Appearance Potential Measurements", 1971-1981, Natl. Stand. Ref. Data Ser., Natl. Bur. Stand. 71, (1982), Cat. No. C13.48:71.
- d: M. W. Chase, C. A. Davies, J. R. Downey, D. R. Frurip, R. A. McDonald, A. N. Syverud, JANAF Thermochemical Tables, Third Edition, J. Phys. Chem. Ref. Data 14, Suppl. 1 (1985).
- e: M. T. Bowers, "Gas Phase Ion Chemistry", Vol. 2 (Academic Press, New York, (1979).
- f: J. O. Cox, G. Pilcher, "Thermochemistry of Organic and Organometallic Compounds", Academic Press, New York, N.Y., 1970.
- g: G. Ditter, U. Niemann, Philips J. Res., 37, 1 (1982).
- h: F. P. Lossing, Can J. Chem., 49, 357 (1971).
- i: H. Halim, N. Heinrich, W. Koch, J. Schmidt, G. Frenking, J. Comp. Chem., 7, 93 (1986), references therein.
- j: S. Schroder, W. Thiel, J. Am. Chem. Soc., 107, 4422 (1985). "BEST" ab-initio involving at least MP2 correction to 6-31G* wavefunctions.
- k: J. C. Traeger, R. G. McLoughlin, J. Am. Chem. Soc., 103, 3647 (1981).
- l: J. L. Franklin, J. G. Dillard, H. M. Rosenstock, J. T. Herron, K. Draxl, F. H. Field, Natl. Stand. Ref. Data Ser., Natl. Bur. Stand., No. 26 (1969).
- m: D. R. Stull, E. F. Westrum, Jr., G. C. Sinke, "The Chemical Thermodynamics of Organic Compounds", Wiley, New York, N.Y., 1969.
- n: J. B. Pedley, B. S. Iseard, "CATCH Tables of Silicon Compounds", University of Sussex, 1972
- o: J. E. Bartmess, R. T. Mclver, "Gas Phase Ion Chemistry", Academic Press, New York, 1979, Vol. II.
- p: R. B. Turner, P. Goebel, B. J. Mallon, W. von E. Doering, J. F. Coburn, M. Pomerantz, J. Am. Chem. Soc., 90, 4315 (1968).
- q: W. D. Good, R. T. Moore, A. G. Osborn, D. R. Douslin, J. Chem. Thermodyn., 6, 303 (1974).
- r: J. M. Harris, S. G. Shafer, S. D. Worley, J. Comput. Chem., 3, 208 (1982).

- s: F. P. Lossing, A. Maccoll, *Can. J. Chem.*, 54, 990 (1976).
- t: J. L. Abboud, W. J. Hehre, R. W. Taft, *J. Am. Chem. Soc.*, 94, 6072 (1972).
- u: G. Vincow, H. J. Dauben, F. R. Hunter, W. V. Volland, *J. Am. Chem. Soc.*, 91, 2823 (1969).
- v: P. v. R. Schleyer, J. E. Williams, K. R. Blanchard, *J. Am. Chem. Soc.*, 92, 3277, (1970).
- w: Average value from two determinations: M. Mansson, N. Rapport, E. F. Westrum *J. Am. Chem. Soc.*, 92, 7296 (1970); R. S. Butler, A. S. Carson, P. G. Laye, W. V. Steele, *J. Chem. Thermodyn.* 3, 277 (1971).
- x: M. W. Chase, J. L. Curnutt, A. T. Hu, H. Prophet, A. N. Syverud, L. C. Walke *J. Phys. Chem. Ref. Data*, 3, 311, (1974).
- y: J. Chao, B. J. Zwolinski, *J. Phys. Chem. Ref. Data*, 5, 319 (1976).
- z: G. Herzberg, "Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules", 2nd ed, Van Nostrand, New York, N.Y., 1950.
- aa: D. D. Wagman, W. H. Evans, V. B. Parker, R. H. Schumm, I. Halow, S. M. Bailey, K. L. Churney, R. L. Nuttall, *J. Phys. Chem. Ref. Data Suppl.* 11, 2 (1982).
- bb: P. Kebarle, R. Tamdagni, H. Kirooka, T. B. McMahon, *Int. J. Mass Spectrom. Ion Phys.*, 19 71 (1976).
- cc: D. D. Wagman, W. H. Evans, V. B. Parker, T. Hawlow, S. M. Bailey, R. H. Schumm, *Natl. Bur. Stand. (U.S.), Tech. Note*, No 270-3 (1968) and "Errata in NBS Technical Note 270-8" (1981).
- dd: Yu. K. Knobel, E. A. Miroshchichenko, Yu. A. Lebedev, *Dokl. Akad. Nauk SSSR*, 190, 348 (1970).
- ee: D. Hwang, M. Tamura, T. Yoshida, N. Tanaka, F. Hosoya, *J. Energetic Matls.*, (1989) in press.
- ff: B. I. Istomin, V. Palm, *Reakts. Sposobnosi Org. Soedin.*, 10, 583 (1973).
- gg: C. Willis, F. P. Lossing, R. A. Back, *Can. J. Chem.* 54, 1 (1967). Note: S. N. Foner, R. L. Hudson, *J. Chem. Phys.*, 68, 3162 (1978) give the heat of formation of diazine as 50.7 +/- 2 Kcal/mol.
- hh: S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, R. Walsh, *Chem. Rev.*, 69, 279 (1969).
- ii: P. S. Engel, J. L. Wood, J. A. Sweet, M. L. Margrave, *J. Am. Chem. Soc.* 96, 2381 (1974).
- jj: G. A. Carpenter, M. F. Zimmer, E. E. Baroody, R. A. Robb, *J. Chem. Eng. Data*, 15, 553 (1970).
- kk: J. J. Grabowski, L. Zhang, *J. Am. Chem. Soc.*, 111, 1193 (1989).
- ll: H. Mackle, P. A. G. O'Hare, *Tetrahedron*, 19, 961 (1963).
- mm: J. V. Davis, S. Sunner, *Acta Chem. Scand.*, 16, 1870 (1962).
- nn: M. Mansson, S. Sunner, *Acta Chem. Scand.*, 16, 1863 (1962).
- oo: J. B. Pedley, G. Rylance, "Sussex-N.P.L. Computer Analysed Thermochemical Data: Organic and Organometallic Compounds", Sussex University, 1977.
- pp: H. Mackle, R. T. B. McLean, *Trans. Faraday Soc.*, 59, 669 (1963).
- qq: B. G. Hobrock, R. W. Kiser, *J. Phys. Chem.*, 67, 1283 (1963).

- rr: R. J. Blint, T. B. McMahon, J. L. Beauchamp, *J. Am. Chem. Soc.*, 96, 1269 (1974).
- ss: A. S. Rodgers, J. Chao, R. C. Wilhoit, B. J. Zwolinski, *J. Phys. Chem. Ref. Data*, 3, 117 (1974).
- tt: R. F. Lake, H. Thompson, *Proc. R. Soc. London, Ser. A*, 315, 323 (1970).
- uu: D. P. Ridge, *J. Am. Chem. Soc.*, 97, 5670 (1975).
- vv: J. Berkowitz, P. M. Dehmer, W. A. Chupka, *J. Chem. Phys.*, 59, 925 (1973).
- ww: J. A. Pople, L. A. Curtiss, *J. Chem. Phys.* 90, 2833 (1989). Ref. gives -19.9 Kcal/mol for calculated heat of formation of HOF at 0K.
- xx: J. Vogt, J. L. Beauchamp, *J. Am. Chem. Soc.*, 97, 6682 (1975).
- yy: P. W. Harland, J. L. Franklin, *J. Chem. Phys.*, 61, 1621 (1974).
- zz: E. C. Wu, A. S. Rodgers, *J. Phys. Chem.*; 78, 2315 (1974).
- aaa: J. W. Larson, T. B. McMahon, *J. Am. Chem. Soc.*, 107, 766 (1985).
- bbb: Private Communication by Dr Dave Dixon, DuPont.
- ccc: A. S. Gordon, *Int. J. Chem. Kinet.*, 4, 541 (1972).
- ddd: D. R. Stull, H. Prophet, *Natl. Stand., Ref. Data Ser. (U.S., Natl. Bur. Stand.) NSRDS-NBS 37*, 1971.
- eee: NBS Technical Note 270-3, Jan. (1968).
- fff: D. C. Frost, C. A. McDowell, D. A. Vroom, *J. Chem. Phys.*, 46, 4255 (1967).
- ggg: *Natl. Bur. Stand. Selected Values of Chemical and Thermodynamic Properties*, Rossini, 1st Feb. (1952).
- hhh: R. L. De Kock, B. R. Higginson, D. R. Lloyd, A. Breeze, D. W. J. Cruickshank, D. R. Armstrong, *Mol. Phys.*, 24, 1059 (1972).
- iii: S. A. Kudchadker, A. P. Kudchadker, *J. Phys. Chem. Ref. Data*, 4, 457 (1975).
- jjj: I. Barin, O. Knacke, O. Kubaschewski, "Thermochemical Properties of Inorganic Compounds", Springer-Verlag, Berlin, 1977.
- kkk: "Spectroscopic Properties of Inorganic and Organometallic Compounds", Royal Society of Chemistry, Spottiswoode Ballantyne Ltd., Colchester and London, 1980, Vol. 14.
- lll: C. H. Wu, *J. Chem. Phys.*, 91, 546 (1989).
- mmm: L. Operti, E. C. Tews, T. J. Macmahon, B. S. Freiser, *J. Am. Chem. Soc.*, 111, 9152 (1989).
- nnn: M. D. Harmony, V. W. Laurie, R. L. Kuczkowsky, R. H. Schwendeman, D. A. Ramsay, F. J. Lovas, W. J. Lafferty, A. G. Maki, *J. Phys. Chem. Ref. Data* 8, 3 (1979).
- ooo: M. F. Guest, J. B. Pedley, M. Horn, *J. Chem. Thermodyn.*, 1, 345 (1969).
- ppp: G. M. Kolyakova, I. B. Rabinovich, E. N. Zoria, *Dokl. Akad. Nauk SSSR*, 209, 616 (1973).
- qqq: L. A. Curtiss, J. A. Pople, *J. Chem. Phys.*, 90, 4314 (1989).
- rrr: R. Thomson, F. W. Dalby, *Can J. Phys.*, 47, 1155 (1969).

- sss: L. Barton, S. K. Watson, R. F. Porter, J. Phys. Chem., 69, 3160 (1965).
- ttt: A. J. Vanderwielen, M. A. Ring, H. E. O'Neal, J. Am. Chem. Soc., 97, 993 (1975).
- uuu: M. M. Heaton, J. Chem. Phys., 67, 5396 (1977).
- vvv: M. R. Nimlos, G. B. Ellison, J. Am. Chem. Soc., 108, 6522-6529 (1986).
- www: R. Walsh, Acc. Chem. Res., 14, 246 (1981)
- xxx: W. C. Steele, L. D. Nichols, F. G. A. Stone, J. Am. Chem. Soc., 84, 4441 (1962)
- yyy: W. V. Steele, J. Chem. Thermo., 15, 595 (1983).
- zzz: H. B. Schlegel, J. Phys. Chem., 88, 6254 (1984).
- aaaa: R. Walsh, J. Chem. Soc.; Faraday I, 79, 2233 (1983).
- bbbb: M. Farber, R. D. Srivastava, Chem. Phys. Lett., 60, 216, (1979).
- cccc: P. Potzinger, A. Ritter, J. Krause, Z. Naturforsch., Teil A, 30, 347, (1975).
- dddd: J. C. Baldwin, M. F. Lappert, J. B. Pedley, J. S. Poland, J. Chem. Soc., Dalton Trans., 1943 (1972).
- eeee: CODATA, J. Chem. Thermo., 10, 903 (1978).
- ffff: M. F. Lappert, J. B. Pedley, J. Simpson, T. R. Spalding, J. Organomet. Chem., 29, 195 (1971).
- gggg: "Gmelins Handbuch der anorganischen Chemie", Zinn, Thiel C1, Verlag Chemie, GMBH, Weinheim, 1972.
- hhhh: G. A. Nash, H. A. Skinner, W. F. Stack, Trans. Faraday Soc., 61, 640 (1965).
- iiii: H. A. Skinner, Adv. Organomet. Chem., 2, 49 (1964).
- jjjj: F. E. Saalfeld, H. Svec, J. Inorg. Chem., 2, 46 (1963).
- kkkk: Y. Wada, R. W. Kiser, J. Phys. Chem., 68, 2290 (1964).
- llll: M. Henchman, A. A. Viggiano, J. F. Paulson, A. Freeman, J. Wormhoudt, J. Am. Chem. Soc., 107, 1253 (1985).
- mmmm: E. C. Penski, E. S. Domalski, CRDEC-TR-87063, (1987).

6.8 Comp. of Calc'd and Exp'l Geometries for MNDO, AM1, and PM3

The following Table lists over 1100 experimental bond lengths, angles, and dihedrals or torsion angles, and the errors in the prediction of these quantities for each method.

Reference geometry data are taken from many sources, mainly from X-ray and microwave structures. No differentiation is made between these various sources, although microwave structures are preferred to X-ray. Further, no differentiation is made between various definitions of bond lengths. Thus the bond length corresponding to the bottom of the harmonic well, r_o , is not distinguished from the equilibrium bond length, or from the bond length corresponding to the system in which the zero point energy is included. For semiempirical methods, the average error in bond lengths is large relative to the differences between the bond lengths obtained using the various definitions.

Notes on the Table

- The AM1 results for P_4O_6 and P_4O_{10} are incorrect. The experimental geometry was used to start the calculation. In order to generate the lowest energy geometry, the systems must first be distorted.
- For salicylaloxim, the bonds reported are the internal hydrogen bond and a non-bonding O-N interaction.
- Several 180° angles are reported. Most of these correspond to symmetry-defined angles. In some instances, e.g., $MgCl_2$, the calculated angle is incorrect. The protocols for this Chapter require *all* such results to be reported.
- The PM3 prediction for the bond-length in HgI is in error by over 10\AA . HgI is predicted, by PM3, to consist of a mercury and an iodine atom. The fault can be traced to the incorrect PM3 prediction for the ground state of the iodine atom. Removal of this one error would lower the average errors in Hg bond lengths from 0.586\AA to 0.054\AA , and the average error in I bond lengths from 0.371\AA to 0.091\AA . These, in turn, result in the average bond-length error in PM3 dropping from 0.045\AA . If that were done, then PM3 would become the most accurate method for the prediction of bond-lengths. Again, the protocols involved preclude this being done at this time.

Table 6.52: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
H ₂	Hydrogen	HH	0.741	-0.042	-0.078	-0.064	a
CH ₂	Methylene, singlet	CH	1.110	-0.018	-0.019	-0.007	a
		HCH	102.4	1.3	8.7	8.1	
CH ₂	Methylene, triplet	CH	1.029	0.033	0.023	0.034	a
		HCH	136.0	9.1	13.7	12.4	
CH ₄	Methane	CH	1.094	-0.007	0.010	0.018	b
C ₂	Carbon, dimer	CC	1.242	-0.053	-0.073	-0.078	a
C ₂ H ₂	Acetylene	CC	1.203	-0.013	-0.008	-0.008	b
		CH	1.060	0.004	-0.009	0.001	
C ₂ H ₄	Ethylene	CC	1.339	-0.017	-0.004	-0.013	b
		CH	1.086	0.000	0.003	0.012	
		HCC	121.2	1.9	2.0	1.5	
C ₂ H ₆	Ethane	CC	1.536	-0.032	-0.015	-0.036	b
		CH	1.091	0.007	0.018	0.026	
		HCC	110.9	0.7	0.3	-0.2	
C ₃ H ₄	Allene	CC	1.308	-0.012	-0.002	-0.010	c
		CH	1.087	-0.001	0.003	0.013	
		HCC	120.9	1.3	1.9	1.2	
C ₃ H ₄	Cyclopropene	C2C3	1.509	-0.025	0.003	-0.020	d
		C1C2	1.296	0.018	0.032	0.023	
		C1H	1.072	0.001	-0.010	-0.003	
		HC1C2	149.9	1.6	1.7	2.0	
C ₃ H ₄	Propyne	C2C1	1.206	-0.014	-0.009	-0.009	e
		C1H	1.056	0.008	-0.005	0.004	
		C3C3	1.459	-0.026	-0.014	-0.032	
		C3H	1.105	-0.007	0.006	0.016	
		HCC	111.0	-0.3	0.0	-0.5	
C ₃ H ₆	Cyclopropane	CC	1.510	-0.011	0.016	-0.009	f
		CH	1.089	0.006	0.007	0.015	
C ₃ H ₆	Propene	C=C	1.336	-0.008	0.004	-0.005	g
		C-C	1.501	-0.021	-0.005	-0.025	
		CCC	124.3	-0.9	2.6	0.0	
		C3H	1.085	0.013	0.024	0.033	
		HC3C2	111.2	1.7	1.9	0.7	
		C2H	1.090	0.006	0.006	0.014	
		HC2C1	119.0	1.8	0.3	1.9	
		HC1	1.081	0.005	0.008	0.017	
		HC1C2	121.5	1.1	0.8	0.8	
		C ₃ H ₈	Propane	CC	1.526	-0.014	0.004
CCC	112.4			-0.6	3.0	-0.5	
C2H	1.115			-0.007	0.000	0.007	
HC2C1	109.5			0.3	-0.7	-0.1	
C1H	1.096			0.001	0.014	0.021	
HC1C2	111.8			-0.4	-1.5	-1.3	

Table 6.53: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
C ₄ H ₂	Diacetylene	C1C2	1.205	-0.012	-0.006	-0.006	h
		C2C3	1.376	-0.005	-0.008	-0.020	
C ₄ H ₄	CH ₂ =C=C=CH ₂	CH	1.046	0.019	0.004	0.014	i
		CH	1.083	0.004	0.007	0.017	
		C1C2	1.318	-0.017	-0.008	-0.016	
C ₄ H ₄	Vinylacetylene	C2C3	1.283	-0.016	-0.013	-0.017	j
		C3C4	1.341	-0.009	0.004	-0.005	
		C2C3	1.431	-0.017	-0.014	-0.026	
C ₄ H ₄	Vinylacetylene	C2C3C4	123.1	-0.5	2.3	1.0	j
		C1C2	1.208	-0.015	-0.010	-0.010	
		C1C2	1.498	0.008	0.029	0.012	
C ₄ H ₆	Bicyclobutane	C1C3	1.497	-0.016	0.040	-0.002	
		C2C3C1C4	121.7	-1.7	1.0	0.3	
		C1H	1.071	0.012	0.003	0.008	
C ₄ H ₆	2-Butyne	C2H	1.093	0.003	0.005	0.012	j
		C2C3	1.213	-0.020	-0.013	-0.015	
		C1C2	1.467	-0.035	-0.023	-0.042	
C ₄ H ₆	1,3-Butadiene	CH	1.115	-0.017	-0.004	0.006	l
		HCC	110.7	0.0	0.3	-0.1	
		C1C2	1.344	-0.013	0.000	-0.009	
C ₄ H ₆	1,3-Butadiene	C2C3	1.467	-0.011	-0.002	-0.016	l
		CCC	122.9	-0.5	2.8	0.6	
		C2C3	1.347	-0.019	-0.006	-0.016	
C ₄ H ₈	1-Butene	C1C2	1.508	-0.019	-0.003	-0.024	
		CCC	123.8	-1.2	1.6	-0.4	
		CC	1.548	-0.006	0.001	-0.005	n
C ₄ H ₈	Cyclobutane	C1C2C4C3	153.0	27.0	27.1	26.9	
		CH	1.105	-0.005	0.000	0.005	
		C1C2	1.330	0.003	0.018	0.006	o
C ₄ H ₈	Isobutene	C2C3	1.507	-0.020	0.002	-0.023	
		C1C2C3	122.4	-0.3	-0.5	0.0	
		C1C2	1.533	-0.021	-0.002	-0.026	e
C ₄ H ₁₀	n-Butane	C2C3	1.533	-0.013	0.006	-0.019	
		CCC	112.8	-1.2	2.0	-1.2	
		CC	1.525	-0.005	0.016	-0.011	p
C ₄ H ₁₀	Isobutane	C=C	1.339	-0.011	0.001	-0.008	
		C-C	1.511	-0.022	-0.005	-0.027	
		C-C=C	115.5	7.6	11.1	8.4	
C ₅ H ₈	1,4-Pentadiene C1	C-C-C	113.1	1.3	-0.4	1.2	q
		C=C	1.339	-0.011	0.001	-0.008	
		C-C	1.511	-0.021	-0.005	-0.025	
C ₅ H ₈	1,4-Pentadiene C2	C-C=C	115.5	7.6	11.2	8.3	q
		C-C-C	108.9	1.9	3.7	2.9	
		C=C	1.339	-0.011	0.001	-0.008	
C ₅ H ₈	1,4-Pentadiene Cs	C-C	1.511	-0.021	-0.005	-0.025	
		C-C=C	115.5	7.6	11.1	8.4	
		C-C-C	108.9	2.0	3.7	2.8	
C ₅ H ₁₂	Neopentane	CC	1.539	-0.012	0.015	-0.018	f
		CH	1.120	-0.022	-0.011	-0.004	
		HCC	110.0	1.3	1.7	0.3	

Table 6.54: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
C ₆ H ₆	Benzene	CC	1.399	-0.008	0.008	-0.004	r
		CH	1.084	0.011	0.006	0.016	
C ₆ H ₆	Fulvene	C3C4	1.476	-0.005	0.001	0.000	s
		C2C3	1.355	0.000	0.011	0.008	
		C1C2	1.470	0.008	0.021	0.013	
		C1C6	1.349	-0.018	-0.004	-0.017	
C ₆ H ₁₀	Cyclohexene	C1C2	1.335	-0.001	0.011	0.002	t
		C2C3	1.504	-0.017	0.000	-0.020	
		C3C4	1.515	0.006	0.026	0.002	
		C4C5	1.550	-0.031	-0.011	-0.036	
		C5C4C2C1	21.8	6.0	-0.5	5.4	
C ₆ H ₁₂	Cyclohexane	CC	1.536	-0.015	0.005	-0.021	u
		CCC	111.4	-0.4	3.3	0.0	
		CCCC	46.3	9.7	-0.6	8.7	
H ₂ O	Water	OH	0.957	-0.006	-0.014	0.004	b
		HOH	104.5	3.2	2.3	-1.0	
CO	Carbon monoxide	CO	1.128	0.007	0.035	0.043	v
CH ₂ O	Formaldehyde	CO	1.208	-0.006	0.009	0.019	w
		CH	1.116	-0.025	-0.010	-0.006	
CH ₄ O	Methanol	HCO	121.7	0.1	1.8	0.5	x
		CO	1.425	-0.030	-0.034	-0.014	
		CH	1.094	0.003	0.025	0.025	
		HCO	108.5	3.6	3.8	2.4	
		OH	0.945	0.004	0.002	0.019	
C ₂ H ₂ O	Ketene	COH	107.0	0.5	4.6	0.2	y
		CO	1.161	0.014	0.023	0.032	
		CC	1.314	-0.006	0.005	-0.007	
		CH	1.083	0.001	0.002	0.012	
C ₂ H ₄ O	Acetaldehyde	HCC	118.7	3.3	3.0	2.7	z
		C1-C2	1.501	-0.002	0.016	-0.011	
		C2-O	1.210	0.000	0.011	0.022	
		O-C2-C1	123.9	-0.6	1.1	-0.4	
		C2-H	1.114	-0.012	-0.002	0.000	
C ₂ H ₆ O	Dimethyl ether	C1-C2-H	117.5	-0.7	-3.5	-2.2	aa
		CC	1.410	-0.004	-0.014	0.007	
		COC	111.3	2.8	8.7	1.6	
C ₃ H ₄ O	Acrolein	C3C2	1.335	-0.005	0.008	-0.001	bb
		C2C1	1.478	0.001	0.007	-0.010	
		CCC	121.0	2.6	6.3	2.2	
		CO	1.208	0.003	0.016	0.026	
C ₃ H ₆ O	Acetone	OCC	124.0	0.0	1.5	0.1	cc
		C=O	1.222	-0.006	0.005	0.013	
		C-C	1.507	-0.002	0.020	-0.012	
C ₄ H ₄ O	Furan	C-C=O	121.4	0.9	-0.1	0.9	dd
		CO	1.362	0.016	0.005	0.033	
		CCO	106.6	0.3	1.0	0.0	
		C3C2	1.361	0.012	0.029	0.019	
		CCC	110.7	-0.5	-0.4	-0.6	

Table 6.55: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
O ₂	Oxygen, triplet state	OO	1.216	-0.047	-0.082	-0.131	v
H ₂ O ₂	Hydrogen peroxide	OO	1.475	0.006	-0.180	-0.177	ee
		OH	0.950	-0.005	0.011	0.034	
		HOO	94.8	1.8	12.5	11.5	
		HOOH	119.8	60.4	36.8	1.7	
		HOOH	119.8	60.4	36.8	1.7	ee
		CO ₂	Carbon dioxide	CO	1.162	0.019	0.024
CH ₂ O ₂	Formic acid	C=O	1.202	0.009	0.025	0.028	ff
		C-O	1.343	0.001	0.011	0.014	
		OCO	124.9	-7.7	-4.3	-7.2	
		OH	0.972	-0.019	-0.023	-0.001	
		HOC	106.3	5.5	10.0	4.3	
		CH	1.097	-0.002	0.008	0.006	
		HC-O	124.1	6.3	2.6	6.0	
C ₂ H ₂ O ₂	trans Glyoxal	CO	1.207	0.000	0.013	0.022	bb
		CC	1.525	0.001	0.004	-0.016	
		CCO	121.2	-0.7	0.8	-0.2	
C ₂ H ₄ O ₂	Methyl formate	C=O	1.200	0.008	0.024	0.029	gg
		C-O	1.334	0.022	0.023	0.028	
		O-C=O	125.9	-5.3	-3.9	-6.7	
		O-CH ₃	1.437	-0.024	-0.033	-0.009	
		C-O-C	114.8	4.8	10.8	2.5	
C ₆ H ₄ O ₂	p-Benzoquinone	C1C2	1.477	0.010	0.024	0.002	hh
		C2C3	1.322	0.013	0.027	0.016	
		CCC	121.1	0.5	1.0	0.8	
		CO	1.222	-0.005	0.004	0.014	
O ₃	Ozone	O-O	1.198	0.025	-0.007	-0.038	b
		O-O-O	114.9	-0.8	2.7	6.0	
H ₃ N	Ammonia	NH	1.012	-0.013	-0.005	-0.014	a
		HNH	106.7	1.4	-1.4	2.4	
CN	Cyanide (+)	CN	1.290	0.065	-0.149	-0.164	a
CN	Cyanide	CN	1.175	-0.018	-0.022	-0.027	a
CHN	Hydrogen cyanide	CN	1.154	0.002	0.006	0.006	ii
		CH	1.063	0.007	-0.008	0.006	
CH ₅ N	Methylamine	CN	1.474	-0.005	-0.014	-0.041	jj
		NH	1.011	-0.012	-0.003	-0.010	
		HNC	112.0	-2.2	-2.2	-0.9	
		HNH	105.9	2.8	-0.3	3.0	
C ₂ H ₃ N	Acetonitrile	CC	1.458	-0.018	-0.006	-0.019	kk
		CH	1.104	-0.006	0.006	0.016	
		HCC	109.5	0.9	1.1	0.6	
		CN	1.157	0.002	0.005	0.006	
C ₂ H ₃ N	Methyl isocyanide	CN-	1.424	0.009	0.000	-0.029	kk
		CH	1.101	-0.004	0.014	0.024	
		HCN	109.1	0.6	1.1	1.0	
		-CN	1.166	0.015	0.025	0.015	

Table 6.56: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
C ₂ H ₇ N	Dimethylamine	C-N	1.462	0.012	0.000	-0.024	ll
		N-H	1.019	-0.021	-0.010	-0.016	
		H-N-C	108.9	0.6	0.5	1.5	
		C-N-C	112.2	1.1	5.3	2.1	
C ₃ H ₉ N	Trimethylamine	CN	1.451	0.029	0.014	-0.006	mm
		CNC	110.9	1.3	5.2	2.1	
C ₄ H ₅ N	Pyrrole	CN	1.370	0.027	0.028	0.022	l
		CNC	107.7	2.0	2.0	1.1	
		C3C2	1.382	0.008	0.013	0.020	
		CCC	109.8	-2.8	-2.4	-1.4	
		C4C3	1.417	-0.027	-0.022	-0.015	
NO	Nitrogen oxide	NO	1.151	-0.024	-0.029	-0.037	a
CHNO	Hydrogen isocyanate	NH	0.987	-0.001	0.014	0.000	a
		CN	1.207	0.044	0.043	0.025	
		CNH	128.1	-6.3	-10.3	-2.7	
		CO	1.171	0.011	0.014	0.031	
		OCN	180.0	0.0	0.0	0.0	
CH ₃ NO	Formamide	CN	1.376	0.015	0.013	-0.009	nn
		NH	1.002	-0.012	-0.010	-0.016	
		CH	1.102	-0.001	0.006	0.012	
		CO	1.193	0.027	0.034	0.050	
NO ₂	Nitrogen dioxide	OCN	123.8	-6.1	-2.9	-1.8	a
		NO	1.197	-0.016	-0.023	-0.038	
HNO ₂	Nitrous acid (cis)	ONO	136.0	1.7	-2.8	0.5	a
		N-O	1.460	-0.122	-0.163	-0.169	
HNO ₂	Nitrous acid (trans)	N=O	1.200	-0.025	-0.031	-0.038	a
		ONO	114.0	-0.7	3.1	2.6	
		OH	0.980	-0.020	-0.017	0.003	
		HON	103.0	6.8	16.7	12.5	
		N-O	1.460	-0.077	-0.148	-0.141	
C ₇ H ₇ NO ₂	Salicylaloxime	N=O	1.200	-0.033	-0.034	-0.042	a
		ONO	118.0	-8.9	-4.7	-5.2	
		OH	0.980	-0.029	-0.022	-0.005	
		HON	105.0	-0.4	5.0	2.0	
		N-O	1.460	-0.077	-0.148	-0.141	
C ₇ H ₇ NO ₂	Salicylaloxime	N...H	1.834	0.013	0.450	0.303	oo
		O...N	2.626	0.058	0.350	0.298	
HNO ₃	Nitric acid	N=O	1.206	-0.003	0.005	-0.011	a
		O=N=O	130.0	2.7	-3.5	-0.9	
		N-O	1.405	0.005	-0.065	-0.072	
		OH	0.960	-0.007	0.002	0.022	
		NOH	102.0	7.0	12.0	7.7	
N ₂	Nitrogen	NN	1.094	0.004	0.010	0.012	v
H ₄ N ₂	Hydrazine	NN	1.449	-0.009	-0.052	-0.071	a
		NH	1.022	-0.021	-0.001	-0.008	
		HNN	112.0	-5.5	-4.8	-4.4	
		HNNH	90.0	90.8	90.3	90.2	

Table 6.57: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
C ₂ N ₂	Cyanogen	CN	1.154	0.005	0.008	0.008	b
		CC	1.389	-0.007	-0.011	-0.005	
C ₂ H ₆ N ₂	Dimethyldiazene	NN	1.254	-0.026	-0.032	-0.030	b
		CN	1.474	-0.007	0.000	-0.022	
C ₄ H ₄ N ₂	Pyrimidine	CNN	111.9	7.4	5.0	7.9	pp
		C4C5	1.393	0.003	0.016	0.015	
		N3C4	1.350	0.004	0.003	0.000	
		N3C4C5	121.2	-0.5	0.1	1.1	
		C2N3	1.328	0.029	0.029	0.033	
		C2H	1.082	0.016	0.016	0.028	
		C5H	1.087	0.007	0.000	0.008	
N ₂ O	Nitrous oxide	C6H	1.079	0.017	0.016	0.026	a
		NN	1.128	-0.004	0.000	0.000	
		NO	1.184	0.013	-0.003	-0.009	
H ₂ N ₂ O ₂	NH ₂ -NO ₂	NN	1.427	0.004	-0.020	-0.060	qq
		NO	1.206	0.005	0.003	-0.002	
		NH	1.005	-0.005	0.013	-0.003	
		ONO	130.1	-3.0	-6.0	-6.1	
N ₂ O ₃	Dinitrogen trioxide	NN	2.080	-0.383	-0.550	-0.498	a
		NO	1.120	0.031	0.018	0.008	
		NNO	110.0	6.7	6.1	6.2	
		NO'	1.180	0.022	0.028	0.019	
		O'NO'	134.0	-4.6	-9.1	-9.2	
N ₂ O ₄	Dinitrogen tetroxide	NN	1.750	0.009	-0.135	0.068	a
		NO	1.180	0.015	0.009	-0.008	
		ONN	113.1	1.5	2.2	1.1	
N ₃	Azide	NN	1.181	-0.008	-0.007	-0.004	a
C ₃ H ₃ N ₃	s-Triazine	CN	1.338	0.020	0.019	0.026	rr
		NCN	126.8	-5.2	-3.4	-1.1	
CHN ₃ O ₆	Trinitromethane	C-N	1.505	0.078	0.053	0.055	ss
		N-C-N	110.7	-4.0	-1.1	-3.9	
		N=O	1.219	-0.018	-0.014	-0.027	
H ₂ S	Hydrogen sulfide	HS	1.328	-0.038	-0.029	-0.005	a
		HSH	92.2	1.3	5.8	3.3	
CS	Carbon sulfide	CS	1.534	-0.087	-0.050	-0.105	tt
CH ₂ S	Thioformaldehyde	CS	1.611	-0.072	-0.074	-0.099	nn
		CH	1.093	0.002	0.000	0.013	
		HCS	121.6	4.5	2.4	3.7	
CH ₄ S	Thiomethanol	CS	1.818	-0.018	-0.101	-0.064	uu
		SH	1.329	-0.022	-0.027	-0.008	
		HSC	100.3	-0.4	2.0	-0.8	
		HCSH	180.0	0.0	0.0	0.0	
C ₂ H ₆ S	Dimethyl thioether	CS	1.802	0.000	-0.079	-0.050	ss
		CSC	98.9	3.2	9.4	3.0	
		HCS	106.6	3.8	3.1	2.6	
		H'CS	106.6	0.0	0.0	0.0	

Table 6.58: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.	
				PM3	MNDO	AM1		
C ₄ H ₄ S	Thiophene	CS	1.714	0.011	-0.035	-0.042	vv	
		CCS	92.2	-0.8	1.4	1.6		
		C3C2	1.370	-0.004	0.005	0.007		
		CCC	111.5	0.6	0.4	0.1		
CSO	Carbon oxysulfide	CO	1.159	0.017	0.022	0.042	ww	
		CS	1.559	-0.055	-0.049	-0.101		
SO ₂	Sulfur dioxide	SO	1.432	0.010	0.044	-0.003	a	
		OSO	119.5	-13.4	-12.7	-11.6		
SO ₃	Sulfur trioxide	SO	1.430	-0.046	0.061	-0.080	a	
H ₂ SO ₄	Sulfuric acid	S-O	1.550	0.118	0.079	0.072	a	
		OH	0.970	-0.023	-0.023	-0.015		
		SOH	105.0	12.7	12.2	15.9		
		S=O	1.420	-0.013	0.095	-0.062		
NS	Sulfur nitride	SN	1.495	-0.044	-0.055	-0.105	a	
C ₂ H ₃ NS	Methyl isothiocyanate	CS	1.597	-0.099	-0.087	-0.125	xx	
		C=N	1.192	0.039	0.026	0.029		
		C-N	1.479	-0.039	-0.043	-0.075		
		C-N=C	141.6	-2.0	0.4	-0.4		
C ₂ N ₂ S	Sulfur dicyanide	CN	1.157	0.007	0.007	0.012	yy	
		CS	1.701	-0.037	-0.071	-0.083		
		NCS	170.0	6.4	6.9	6.5		
		CSC	98.4	3.2	5.1	4.8		
S ₂	Sulfur dimer	SS	1.889	-0.032	-0.114	-0.053	a	
H ₂ S ₂	H ₂ S ₂	SS	2.055	-0.022	-0.132	0.052	zz	
		SH	1.327	-0.017	-0.023	-0.001		
		HSS	91.3	12.0	11.2	7.4		
		HSSH	90.5	1.9	9.0	8.9		
CS ₂	Carbon disulfide	CS	1.553	-0.072	-0.061	-0.094	ss	
C ₂ H ₆ S ₂	2,3-Dithiabutane	CS	1.810	-0.006	-0.085	-0.060	aaa	
		SS	2.038	-0.001	-0.102	0.072		
		CSS	102.8	6.4	5.2	3.4		
		CSSC	84.7	5.5	23.1	9.8		
		CS	1.810	-0.006	-0.085	-0.060		ss
		SS	2.038	-0.001	-0.102	0.072		
		SSC	102.8	5.8	5.2	3.2		
		CSSC	84.7	5.5	23.1	9.8		
S ₆	S ₆	SS	2.057	-0.009	-0.110	0.052	bbb	
		SSS	102.2	5.3	2.9	0.3		
		SSSS	74.5	-9.9	-5.1	-0.6		
S ₈	S ₈	SS	2.048	-0.075	-0.113	-0.042	ccc	
		SSS	107.9	8.1	0.4	3.7		
		SSSS	98.6	-11.0	-0.2	-4.4		
HF	Hydrogen fluoride	HF	0.917	0.021	0.039	-0.091	ddd	
CF	Fluoromethyldiyne	CF	1.266	-0.008	-0.003	-0.008	a	
CHF	Fluoromethylene	CH	1.121	-0.021	-0.001	0.006	a	
		CF	1.314	-0.030	-0.029	-0.023		
		FCH	101.6	4.0	9.5	9.0		

Table 6.59: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
CH ₃ F	Fluoromethane	CH	1.098	-0.006	0.019	0.023	eee
		CF	1.382	-0.032	-0.033	-0.007	
C ₂ HF	Fluoroacetylene	FCH	108.5	1.2	0.8	1.0	fff
		H-C	1.053	0.012	-0.004	0.006	
		C-C	1.198	-0.010	-0.005	-0.004	
C ₂ H ₃ F	Fluoroethylene	C-F	1.279	0.020	-0.002	0.018	ggg
		CC	1.333	0.000	0.018	0.007	
		CH(g)	1.076	0.017	0.023	0.028	
		CCH(g)	127.7	-1.6	-4.7	-3.9	
		CH(t)	1.085	0.000	0.002	0.011	
		CCH(t)	123.9	-2.6	-3.0	-3.0	
		CH(c)	1.090	-0.004	-0.003	0.006	
		CCH(c)	121.4	2.3	3.1	1.5	
		CF	1.348	-0.010	-0.024	0.003	
		FCC	121.0	1.1	2.3	2.2	
C ₃ H ₃ F	Fluoroallene	C1C2	1.301	0.009	0.019	0.012	hhh
		C1H	1.083	0.011	0.016	0.022	
		HC1C2	124.3	0.5	-1.4	-1.4	
		C1F	1.360	-0.020	-0.035	-0.007	
		FC1C2	121.9	0.4	1.0	1.8	
		C2C3	1.309	-0.015	-0.006	-0.014	
		C3H	1.086	0.001	0.005	0.015	
		HC3C2	120.8	1.5	2.1	1.5	
CNF	Cyanogen fluoride	CN	1.159	0.000	0.001	0.006	fff
		CF	1.262	0.035	0.011	0.045	
NOF	Nitrosyl fluoride	NF	1.520	-0.153	-0.215	-0.153	a
		NO	1.130	0.032	0.031	0.018	
		FNO	110.2	1.4	3.6	2.1	
CH ₃ SO ₃ F	Methyl fluorosulfate	S=O	1.410	-0.005	0.101	-0.051	iii
		O=S=O	124.4	-0.6	-2.0	-0.5	
		S-F	1.545	-0.009	0.064	-0.038	
		F-S=O	106.8	1.6	0.4	2.3	
		S-O	1.558	0.125	0.075	0.084	
		O-S=O	109.5	0.5	0.3	-1.6	
		C-O	1.420	-0.019	-0.017	-0.012	
		C-O-S	116.5	7.1	9.5	7.1	
F ₂	Fluorine	FF	1.412	-0.062	-0.146	0.015	a
H ₂ F ₂	Hydrogen fluoride dimer	HF	0.920	0.019	0.036	-0.093	a
		H'F	1.870	-0.126	1.199	0.430	
		H'FH	108.0	39.3	36.1	-18.3	
CF ₂	Difluoromethylene	CF	1.300	-0.002	0.004	0.012	a
		FCF	104.9	1.4	3.4	1.1	
OF ₂	F ₂ O	OF	1.412	-0.034	-0.131	-0.058	a
		FOF	103.2	-2.2	5.9	-0.7	
COF ₂	Carbonyl difluoride	CO	1.174	0.025	0.045	0.046	a
		CF	1.312	0.010	0.004	0.016	
		FCO	126.0	-1.4	-1.9	-1.5	

Table 6.60: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
SF ₂	Sulfur difluoride	SF	1.592	-0.032	-0.020	-0.036	a
		FSF	98.2	-1.7	1.4	0.8	
CSF ₂	Thiocarbonyl difluoride	CS	1.589	0.011	-0.014	-0.029	ss
		CF	1.315	0.023	0.006	0.031	
		FCS	126.4	2.5	-0.4	2.4	
SOF ₂	Thionyl fluoride	SO	1.412	0.055	0.068	0.021	a
		SF	1.585	-0.010	0.023	-0.037	
		FSO	106.8	-5.8	-4.0	-4.2	
SO ₂ F ₂	Sulfuryl fluoride	FSF	92.8	1.0	4.2	3.7	a
		SF	1.530	0.017	0.080	-0.015	
		FSF	96.1	-0.8	1.6	1.2	
		SO	1.405	-0.005	0.103	-0.046	
S ₂ F ₂	FSSF	OSO	124.0	2.5	0.9	0.9	a
		SF	1.635	-0.051	-0.063	-0.071	
		SS	1.888	0.121	0.077	0.243	
S ₂ F ₂	SSF2	FSS	108.3	3.7	-1.8	0.8	a
		FSSF	87.9	-0.2	-0.4	1.7	
		SS	1.860	0.059	0.034	-0.027	
		SF	1.598	-0.006	0.003	-0.047	
CHF ₃	Trifluoromethane	FSS	107.5	7.0	1.4	12.4	a
		FSF	92.5	-1.6	4.0	1.0	
		CH	1.098	0.013	0.038	0.032	
NF ₃	Nitrogen trifluoride	CF	1.333	0.013	0.020	0.035	a
		FCH	110.3	2.8	1.2	2.7	
C ₂ NF ₃	Trifluoroacetonitrile	NF	1.371	-0.017	-0.056	-0.011	a
		FNF	102.2	2.8	4.0	0.4	
		CC	1.461	0.026	0.037	0.025	
CF ₄	Carbon tetrafluoride	CF	1.335	0.015	0.020	0.036	a
		CCF	111.4	2.0	0.5	2.1	
		CN	1.153	0.002	0.006	0.006	
		CF	1.321	0.016	0.026	0.037	
C ₂ F ₄	Tetrafluoroethylene	CC	1.311	0.044	0.070	0.057	ggg
		CF	1.319	0.007	-0.001	0.021	
SF ₄	Sulfur tetrafluoride	FCC	123.8	1.3	0.4	1.6	a
		SF	1.545	0.051	0.052	0.001	
		FSF	101.5	19.7	0.6	2.4	
		SF'	1.646	-0.024	0.025	-0.073	
CNSOF ₅	Pentafluoro(isocyanato)sulfur	FSF'	87.8	-6.9	0.6	-0.8	iii
		S-F	1.567	-0.004	0.089	-0.007	
		S-N	1.668	0.085	0.061	0.000	
		N=C	1.234	0.011	0.026	-0.022	
		S-N=C	124.9	27.9	10.7	55.1	
C ₂ F ₆	C2F6	C=O	1.179	-0.005	-0.006	0.015	kkk
		N=C=O	173.8	0.6	-3.9	6.1	
		C-C	1.545	0.062	0.129	0.077	
		C-F	1.314	0.027	0.032	0.049	
		F-C-C	109.8	1.7	0.9	2.3	

Table 6.61: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
SF ₆	Sulfur hexafluoride	SF	1.564	-0.003	0.091	-0.024	a
C ₃ F ₈	Perfluoropropane	C-C	1.551	0.050	0.121	0.063	kkk
		C-C-C	115.7	-4.9	-1.7	-5.6	
		C-F	1.314	0.029	0.033	0.050	
		F-C-C	110.7	1.2	0.6	1.8	
C ₃ O ₂ F ₈	CF ₃ -O-CF ₂ -O-CF ₃	C1-O	1.358	0.035	0.049	0.049	kkk
		O-C2	1.357	0.036	0.046	0.043	
		C1-O-C2	123.3	-4.4	7.6	-1.5	
		C1-F	1.305	0.027	0.038	0.047	
C ₄ F ₁₀	Perfluoro-n-butane	C1-C2	1.554	0.053	0.123	0.063	kkk
		C2-C3	1.554	0.039	0.112	0.051	
		C-C-C	115.0	-3.9	-1.1	-5.1	
		C1-F	1.314	0.028	0.033	0.050	
		F-C1-C2	108.5	2.5	1.6	3.2	
S ₂ F ₁₀	S2F10	S-S	2.210	0.680	3.395	1.819	a
		S-F	1.560	0.021	2.399	-0.006	
C ₅ O ₂ F ₁₂	CF ₃ -O-CF ₂ -O-CF(CF ₃)-O-CF ₃	C1-O	1.355	0.042	0.057	0.057	kkk
		C1-F	1.305	0.026	0.036	0.046	
		F-C1-O	107.0	-4.5	-1.1	-3.2	
HCl	Hydrogen chloride	HCl	1.275	-0.007	0.073	0.009	a
CHCl	Chloromethylene	CH	1.120	-0.020	-0.020	-0.010	a
		CCl	1.689	-0.135	0.050	-0.042	
		CiCH	103.4	12.1	6.0	7.7	
CH ₃ Cl	Chloromethane	CCl	1.781	-0.017	0.014	-0.040	a
		CH	1.096	-0.002	0.006	0.016	
C ₂ HCl	Chloroacetylene	HCCl	110.9	-1.0	-2.8	-2.6	lll
		H-C	1.055	0.009	-0.004	0.005	
		C-C	1.203	-0.009	-0.010	-0.007	
		C-Cl	1.637	-0.065	0.024	-0.039	
OCl	Chlorine monoxide	CiO	1.546	0.002	0.073	0.089	a
		CiN	1.950	-0.186	-0.167	-0.219	
NOCl	NOCl	NO	1.170	-0.014	-0.033	-0.033	a
		CNCl	114.0	4.9	3.5	5.3	
		CiN	1.830	-0.012	-0.020	-0.060	
NO ₂ Cl	NO ₂ Cl	NOCl	1.210	-0.013	-0.019	-0.024	a
		CiF	1.628	-0.045	0.022	0.019	
FCl	Chlorine fluoride	CiF	1.630	0.060	0.105	0.051	a
		CiO	1.460	-0.007	0.270	0.328	
O ₃ FCl	Chlorodifluoromethane	OCiF	95.2	7.6	11.0	3.0	mmm
		CH	1.090	0.018	0.036	0.037	
		CCl	1.740	0.082	0.100	0.070	
		CiCH	107.0	2.9	-2.1	-1.8	
		CF	1.350	-0.004	-0.009	0.019	
		FCCl	110.5	-0.3	-0.3	1.9	
F ₃ Cl	Chlorine trifluoride C _{2v}	FCCiH	120.0	2.5	1.0	2.0	a
		CiF	1.598	0.073	0.101	0.085	
		CiF'	1.698	-0.027	0.001	-0.015	
		FCiF'	87.5	32.5	32.5	32.5	

Table 6.62: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
Cl ₂	Chlorine	ClCl	1.986	0.049	0.010	-0.068	a
CH ₂ Cl ₂	Dichloromethane	CCl	1.772	-0.014	0.014	-0.031	jjj
		CICCl	111.8	-3.9	-0.6	1.1	
		CH	1.103	-0.001	0.000	0.010	
OCl ₂	Cl ₂ O	CIO	1.701	-0.001	-0.018	0.032	a
		CIOCl	110.8	-1.6	2.2	0.3	
COCl ₂	Carbonyl chloride	CO	1.166	0.032	0.034	0.056	a
		CCl	1.746	-0.009	0.014	-0.027	
		CICO	124.3	-0.1	-0.4	-1.0	
SCl ₂	Sulfur dichloride	SCI	2.015	0.016	-0.044	-0.057	a
		CISCI	102.7	-1.1	3.6	3.6	
SOCl ₂	Thionyl chloride	SO	1.443	0.035	0.025	0.007	nnn
		SCI	2.076	0.005	-0.039	-0.069	
		CISO	106.3	-1.6	0.4	1.8	
SO ₂ Cl ₂	Sulfuryl chloride	S=O	1.418	-0.017	0.089	-0.024	iii
		S-Cl	2.012	0.063	0.044	0.032	
		O=S-Cl	108.0	0.7	1.3	1.3	
S ₂ Cl ₂	ClSSCl	SCI	2.057	-0.014	-0.081	-0.098	a
		SS	1.931	0.033	-0.011	-0.014	
		CISSCl	108.2	5.0	0.0	8.9	
CF ₂ Cl ₂	Dichlorodifluoromethane	CCl	1.770	0.038	0.055	0.037	a
		CICCl	108.5	-1.9	-1.7	-1.8	
		CF	1.330	0.015	0.007	0.040	
CHCl ₃	Chloroform	FCCI	109.8	1.6	0.8	2.1	jjj
		CCl	1.782	-0.029	0.000	-0.034	
		CICH	107.5	2.9	1.1	0.2	
CFCl ₃	Trichlorofluoromethane	CF	1.330	0.019	-0.003	0.046	a
		CCl	1.760	0.019	0.046	0.026	
CCl ₄	Carbon tetrachloride	CCl	1.760	-0.013	0.022	0.000	jjj
C ₂ Cl ₆	Hexachloroethane	CC	1.550	-0.038	0.016	0.007	a
		CCl	1.740	0.014	0.050	0.020	
		CICC	109.0	1.2	2.4	0.7	
SCl ₆	Sulfur hexachloride	SCI	2.030	0.110	0.112	0.118	ooo
HBr	Hydrogen bromide	HBr	1.415	0.056	0.025	0.006	ppp
CH ₃ Br	Bromomethane	CBr	1.933	0.018	-0.055	-0.028	ss
		CH	1.086	0.004	0.016	0.024	
		HCBr	107.7	0.7	0.8	1.1	
C ₂ HBr	Bromoacetylene	H-C	1.055	0.010	-0.004	0.006	qqq
		C-C	1.204	-0.016	-0.009	-0.005	
		C-Br	1.792	-0.007	-0.051	-0.049	
C ₂ HOBr	Bromoketene	C-Br	1.880	-0.003	-0.063	-0.042	rrr
		C-H	1.082	0.005	0.001	0.015	
		H-C-Br	124.4	-6.9	-8.9	-8.6	
		C=C	1.316	-0.005	0.006	-0.003	
		Br-C=C	118.4	-2.5	3.0	5.4	
		C=O	1.161	0.011	0.020	0.027	

Table 6.63: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
C ₂ H ₃ OBr	Acetyl bromide	CC	1.516	-0.039	-0.001	-0.026	ss
		CBr	1.973	-0.007	-0.087	-0.026	
		BrCC	111.0	-4.7	2.0	2.3	
		CO	1.183	0.002	0.026	0.042	
		CCO	127.1	7.2	0.2	-3.0	
CNBr	Cyanogen bromide	BrC	1.789	0.007	-0.046	-0.028	a
NOBr	BrNO	CN	1.158	-0.003	0.003	0.006	ss
		BrN	2.140	-0.252	-0.271	-0.218	
FBr	BrF	NO	1.146	0.001	-0.007	-0.011	a
		BrNO	114.5	6.3	4.4	8.0	
F ₃ Br	Bromine trifluoride	BrF	1.756	0.018	-0.029	0.021	a
		BrF'	1.721	0.065	0.036	0.087	
		BrF'	1.806	-0.020	-0.049	0.010	
CF ₃ Br	Trifluorobromomethane	FBBrF'	86.2	33.8	33.8	-4.8	a
		CBr	1.909	0.051	0.029	0.134	
		CF	1.328	0.007	0.019	0.039	
F ₅ Br	Bromine pentafluoride	FCBr	110.3	0.5	0.8	3.4	iii
		BrF(ax)	1.697	0.058	0.069	0.126	
SF ₅ Br	Sulfur bromide pentafluoride	BrF(eq)	1.768	0.006	0.003	0.031	iii
		S-F(ax)	1.520	0.019	0.125	0.022	
		S-F(eq)	1.646	-0.105	0.020	-0.101	
ClBr	Bromine chloride	S-Br	2.133	0.293	0.224	0.461	a
CH ₂ ClBr	Bromochloromethane	BrCl	2.136	0.040	-0.056	-0.072	
Br ₂	Bromine	C-Cl	1.768	-0.050	0.015	-0.035	sss
		C-Br	1.930	0.009	-0.058	-0.014	
		Br-C-Cl	112.3	-9.8	-0.7	1.6	
		C-H	1.074	0.026	0.028	0.038	
		Br-Br	2.283	0.160	-0.115	-0.099	
CH ₂ Br ₂	Dibromomethane	CH	1.079	0.015	0.023	0.031	ss
		HCH	113.6	-2.4	-2.8	-3.5	
		CBr	1.927	-0.015	-0.059	-0.025	
		BrCH	106.5	6.0	2.1	1.7	
		C=C	1.356	-0.013	0.008	0.015	
C ₄ O ₂ Br ₂	1,2-Dibromocyclobutene-3,4-dione	C2-C3	1.518	-0.005	-0.001	0.000	ttt
		C=C-C3	94.4	0.1	-0.9	-1.2	
		C-Br	1.831	-0.004	-0.044	-0.028	
		Br-C=C	133.0	-0.1	2.4	2.5	
		C-O	1.185	0.006	0.020	0.024	
		O-C3-C2	135.0	1.7	1.6	1.1	
		S=O	1.449	-0.002	0.025	0.016	
		S-Br	2.255	0.066	-0.120	-0.048	
		Br-S-O	107.6	-4.2	0.5	3.4	
		Br-S-Br	2.255	0.066	-0.120	-0.048	
C ₂ Br ₄	Tetrabromoethylene	Br-S-Br	2.255	0.066	-0.120	-0.048	ss
		CC	1.362	-0.065	-0.020	-0.018	
		CBr	1.881	-0.043	-0.060	-0.024	
SBr ₆	Sulfur hexabromide	BrCC	122.4	6.2	1.3	0.0	ooo
		SBr	2.190	0.238	0.052	0.108	

Table 6.64: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
HI	Hydrogen iodide	HI	1.609	0.068	-0.042	-0.022	ppp
CH ₃ I	Iodomethane	CH	1.084	0.009	0.020	0.025	ss
		CI	2.132	-0.104	-0.117	-0.082	
		HCH	111.2	-1.2	-2.8	-1.4	
C ₂ HI	Iodoacetylene	H-C	1.056	0.009	-0.004	0.006	uuu
		C-C	1.206	-0.018	-0.007	-0.008	
		C-I	1.989	-0.084	-0.106	-0.077	
		IF	1.906	-0.017	-0.004	-0.025	a
CF ₃ I	Trifluoroiodomethane	CI	2.130	-0.078	-0.005	0.045	a
		CF	1.332	0.008	0.022	0.037	
		FCI	110.6	1.5	1.9	3.5	
F ₅ I	Iodine pentafluoride	IF(ax)	1.844	0.023	0.139	0.087	vvv
		IF(eq)	1.869	0.013	0.088	0.029	
F ₇ I	Iodine heptafluoride	IF(ax)	1.760	1.034	0.731	0.866	www
		IF(eq)	1.860	0.042	0.218	0.114	
ClI	Iodine chloride	ICI	2.327	-0.135	-0.065	-0.109	a
BrI	Iodine bromide	IBr	2.485	0.076	-0.135	-0.131	a
I ₂	Iodine	II	2.666	0.002	-0.151	-0.128	a
BeH	Beryllium hydride (+)	Be-H	1.312	-0.039	-0.065	-0.045	a
BeH	Beryllium hydride	Be-H	1.343	-0.036	-0.053	-0.033	a
BeO	Beryllium oxide	Be-O	1.331	-0.027	0.004	0.071	a
BeS	Beryllium sulfide	Be-S	1.742	-0.090	-0.118	-0.057	a
BeF	Beryllium fluoride	Be-F	1.361	0.012	0.098	0.121	a
BeF ₂	Beryllium difluoride	Be-F	1.400	0.002	0.060	0.090	a
C ₅ BeH ₅ Cl	Cyclopentadienylberyllium chloride	C-H	1.090	-0.006	-0.005	0.000	xxx
		Be-ring	1.485	0.198	0.068	0.262	
		Be-Cl	1.839	-0.068	0.117	0.017	
		Be-Cl	1.770	-0.026	0.143	0.067	a
BeCl ₂	Beryllium dichloride	Be-Cl	1.770	-0.026	0.143	0.067	a
BeBr ₂	Beryllium dibromide	Be-Br	1.910	-0.096	0.102	-0.013	a
BeI	Beryllium iodide	Be-I	2.132	0.090	0.023	0.006	a
BeI ₂	Beryllium diiodide	Be-I	2.120	-0.051	0.001	-0.019	a
MgH	Magnesium hydride	Mg-H	1.730	-0.043			a
MgO	Magnesium oxide	Mg-O	1.749	0.031			a
MgS	Magnesium sulfide	Mg-S	2.143	0.215			a
MgF	Magnesium fluoride	Mg-F	1.750	0.004			a
MgF ₂	Magnesium difluoride	Mg-F	1.771	-0.009			yyy
MgCl	Magnesium chloride	Mg-Cl	2.199	-0.324			a
MgCl ₂	Magnesium dichloride	Mg-Cl	2.186	-0.299			yyy
		Cl-Mg-Cl	180.0	23.7			
MgBr	Magnesium bromide	Mg-Br	2.360	-0.006			a
MgBr ₂	Magnesium dibromide	Mg-Br	2.340	0.013			a
MgI ₂	Magnesium diiodide	Mg-I	2.520	-0.100			a
Mg ₂	Magnesium, dimer	Mg-Mg	3.9	3.1			a
BH	BH	BH	1.236		-0.058	-0.016	xxx
BH ₂	BH ₂	BH	1.180		-0.022	0.016	aaa
		HBH	131.0		-4.6	-3.0	
C ₃ BH ₉	Trimethylboron	B-C	0.850		0.000	0.000	ooo

Table 6.65: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
BO	BO	BO	1.204		-0.034	-0.036	xxx
CBH ₃ O	BH ₃ CO	C-B	1.534		-0.039	-0.004	iii
		B-H	1.222		-0.046	-0.017	
		C-B-H	103.8		3.0	-0.5	
		C-O	1.135		0.028	0.036	
		B-O	1.380		-0.008	-0.020	ooo
C ₃ BH ₉ O ₃	Trimethoxyborane	B-O	1.535		-0.025	-0.029	iii
CBH ₃ S	CH ₃ -B=S	C-H	1.109		0.002	0.003	
		H-C-B	110.3		0.1	-1.7	
		B=S	1.603		-0.118	-0.131	
		B-F	1.330		-0.007	-0.020	bbb
		F-B-F	116.0		-3.7	-3.6	
C ₆ BH ₅ F ₂	PhBF ₂	B-C	1.550		0.007	0.000	
		C-C	1.540		0.055	0.010	ggg
		C-B ₃	1.627		0.023	-0.004	
C ₂ B ₄ H ₆	C ₂ B ₄ H ₆	C-B ₄	1.605		0.072	0.043	
		B ₃ -B ₄	1.721		0.031	0.003	
		B ₄ -B ₆	1.752		0.006	-0.013	
		B ₁ -B ₂	1.757		0.043	-0.029	jjj
		B ₂ -B ₃	1.755		-0.001	-0.058	
B ₆ H ₁₀	B ₆ H ₁₀	BBB	58.7		0.6	1.5	
		BB	1.808		0.096	0.021	
		BBB	58.7		0.6	1.5	
		AIH	1.648	0.015	-0.222	-0.186	a
AIO	AIO	AIO	1.618	0.003	-0.143	-0.068	a
AIF	Aluminum fluoride	AIF	1.654	-0.002	-0.094	-0.096	a
AIF ₃	Aluminum trifluoride	AIF	1.630	0.015	-0.041	-0.052	a
AIF ₄	AIF ₄ (-)	AIF	1.690	-0.002	-0.041	-0.070	a
AICl	Aluminum chloride	AICl	2.130	-0.183	-0.055	-0.294	a
AICl ₃	Aluminum trichloride	AICl	2.060	-0.094	0.005	-0.186	a
AIBr	Aluminum bromide	AIBr	2.295	-0.003	-0.093	-0.031	a
AIBr ₃	Aluminum tribromide	AIBr	2.270	-0.395	-0.095	-0.026	a
All ₃	Aluminum triiodide	All	2.499	-0.012	-0.174	-0.111	a
Al ₂	Al ₂	AlAl	2.467	0.073	-0.175	-0.060	a
Al ₂ O	Al ₂ O	AIO	1.730	-0.053	-0.124	-0.065	a
GaH	Gallium hydride	Ga-H	1.663	-0.015			ppp
GaF	Gallium fluoride	Ga-F	1.774	0.009			ppp
GaF ₃	Gallium trifluoride	Ga-F	1.725	-0.012			yyy
GaCl	Gallium chloride	Ga-Cl	2.202	0.104			ppp
GaH ₃ NCl ₃	Gallium trichloride-ammonia	Ga-N	2.057	0.358			iii
		Ga-Cl	2.142	0.233			
		Cl-Ga-Cl	116.4	4.0			
GaCl ₄	GaCl ₄ (-)	Ga-Cl	2.170	0.046			kkk
GaBr	Gallium bromide	Ga-Cl	2.352	0.042			ppp
GaH ₃ NBr ₃	Gallium tribromide-ammonia	Ga-N	2.081	-0.230			iii
		Ga-Br	2.288	0.048			
		Br-Ga-Br	116.1	-1.7			

Table 6.66: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	PM3	Errors		Ref.
					MNDO	AM1	
Gal	Gallium iodide	Ga-I	2.575	-0.036			ppp
Gal ₃	Gallium triiodide	Ga-I	2.458	0.119			yyy
Ga ₂ O	Gallium(I) oxide	Ga-O	1.824	-0.028			iii
Ga ₂ H ₄ Cl ₂	Ga ₂ Cl ₂ H ₄	Ga-O-Ga	142.9	37.1			
		Ga-Ga	3.241	0.583			III
		Ga-Cl	2.349	0.059			
		Ga-H	1.559	0.039			
Ga ₂ Cl ₆	Ga ₂ Cl ₆	Ga-Cl(b)	2.300	0.061			ooo
		Ga-Cl(t)	2.100	-0.211			
Ga ₂ Br ₆	Ga ₂ Br ₆	Ga-Br(b)	2.450	-0.016			ooo
		Ga-Br(t)	2.250	-0.017			
InH	Indium hydride	In-H	1.838	-0.104			ppp
InF	Indium fluoride	In-F	1.985	0.001			ppp
InCl	Indium chloride	In-Cl	2.401	0.002			ppp
InBr	Indium bromide	In-Br	2.543	-0.253			ppp
InI	Indium iodide	In-I	2.729	-0.019			mmm
InI ₃	Indium triiodide	In-I	2.641	0.004			yyy
In ₂ O	Indium(I) oxide	In-O	2.020	-0.021			iii
		In-O-In	145.0	35.0			
C ₅ TlH ₅	Cyclopentadienyl thallium	Tl-C	2.705	0.007			xxx
TlF	Thallium fluoride (TlF)	Tl-F	2.084	0.050			ooo
TlCl	Thallium chloride	Tl-Cl	2.485	0.004			xxx
TlBr	Thallium bromide	Tl-Br	2.618	-0.059			xxx
TlI	Thallium iodide	Tl-I	2.814	-0.090			xxx
Tl ₂ F ₂	Thallium fluoride dimer	Tl-F	2.290	-0.005			iii
		F-Tl-F	90.0	23.1			
HSi	SiH	SiH	1.520	-0.015	-0.146	-0.066	a
H ₂ Si	Silylene (singlet)	SiH	1.519	-0.006	-0.139	-0.062	nnn
		HSiH	92.1	2.8	5.2	8.9	
H ₄ Si	Silane	SiH	1.481	0.007	-0.105	-0.020	a
CH ₆ Si	Methylsilane	SiH	1.485	0.008	-0.106	-0.022	ss
		HSiH	108.3	-0.1	0.5	-0.4	
		SiC	1.867	0.001	-0.064	-0.058	
		SiC	1.875	0.015	-0.060	-0.046	ss
C ₄ H ₁₂ Si	Tetramethylsilane	SiC	1.875	0.015	-0.060	-0.046	ss
SiN	Silicon nitride	SiN	1.572	-0.108	-0.092	-0.087	a
CH ₆ SiS	CH ₃ -S-SiH ₃	C-S	1.819	-0.019	-0.111	-0.058	ooo
		S-Si	2.134	0.105	-0.070	0.176	
		C-S-Si	98.3	3.9	16.8	10.5	
		Si-H	1.481	0.009	-0.112	-0.020	
		C-H	1.091	0.007	0.016	0.026	
		Si-F	1.602	0.001	-0.004	0.021	ppp
CH ₅ SiF	CH ₃ -SiH ₂ F	Si-C	1.845	0.028	-0.038	-0.043	
		C-Si-F	109.2	-0.1	0.1	2.8	
		Si-H	1.478	0.025	-0.097	-0.019	
		C-Si-H	112.4	-2.3	-1.5	-3.1	
SiF ₂	Difluorosilylene	SiF	1.591	-0.016	-0.013	0.021	a
		FSiF	101.0	-5.7	-4.1	-4.0	

Table 6.67: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
HSiF ₃	Trifluorosilane	SiH	1.447	0.061	-0.072	-0.007	a
		SiF	1.562	0.028	0.031	0.047	
		FSiH	110.6	1.8	3.0	1.2	
SiF ₄	Tetrafluorosilane	SiF	1.552	0.028	0.032	0.052	a
SiCl	Chlorosilylydyne	SiCl	2.063	-0.117	0.009	-0.077	a
SiCl ₂	Dichlorosilylene	ClSiCl	109.7	-7.8	-4.2	-5.3	qqq
CH ₃ SiCl ₃	Trichloromethylsilane	Si-C	1.848	-0.002	-0.053	-0.060	iii
		Si-Cl	2.026	0.038	0.065	0.035	
		C-Si-Cl	110.3	1.0	1.5	0.9	
SiCl ₄	Silicon tetrachloride	SiCl	2.017	0.024	0.063	0.022	a
SiF ₃ Br	Trifluorobromosilane	Si-Br	2.156	-0.282	0.063	0.139	rrr
		Si-F	1.559	0.013	0.026	0.048	
		F-Si-Br	110.4	-2.2	1.8	1.8	
SiBr ₄	Silicon tetrabromide	SiBr	2.150	-0.354	0.040	0.093	a
H ₃ SiI	Iodosilane	SiI	2.437	-0.425	-0.051	-0.003	a
		SiH	1.486	0.006	-0.116	-0.020	
		HSiI	108.5	-0.5	-0.1	1.3	
SiI ₄	Silicon tetraiodide	SiI	2.430	0.037	-0.097	-0.005	a
Si ₂	Silicon dimer	SiSi	2.246	0.050	-0.260	-0.458	a
H ₆ Si ₂	Disilane	SiSi	2.331	0.063	-0.166	0.081	sss
		SiH	1.492	-0.004	-0.118	-0.026	
		HSiSi	110.3	-0.8	2.3	-0.7	
H ₁₀ Si ₅	Cyclopentasilane	Si-Si	2.342	0.040	-0.179	0.050	iii
GeH ₄	Germane	Ge-H	1.527	-0.022	-0.046	0.019	ttt
CGeH ₆	Methylgermane	Ge-C	1.945	0.010	-0.018	0.042	ss
		Ge-H	1.529	-0.024	-0.045	0.016	
		C-H	1.083	0.007	0.024	0.022	
C ₂ GeH ₈	Ethylgermane	Ge-C	1.949	0.008	-0.008	0.055	xxx
		C-C	1.545	-0.052	-0.025	-0.056	
		Ge-C-C	112.2	-7.0	5.0	-1.0	
C ₂ GeH ₈	Dimethylgermane	Ge-H	1.522	-0.016	-0.039	0.023	uuu
		H-Ge-C	109.7	-0.1	0.2	-2.2	
		Ge-C	1.950	0.007	-0.019	0.034	
C ₃ GeH ₁₀	Trimethylgermane	C-Ge-C	110.0	-0.4	0.8	-3.3	xxx
		Ge-H	1.522	-0.017	-0.034	0.021	
		Ge-C	1.947	0.012	-0.013	0.035	
C ₄ GeH ₁₂	Tetramethylgermanium	H-Ge-C	109.3	0.3	-0.5	1.6	vvv
		Ge-C	1.945	0.015	-0.006	0.036	
		C-H	1.120	-0.030	-0.013	-0.014	
GeO	Germanium oxide	Ge-O	1.625	0.012	-0.055	-0.002	ppp
CGeH ₃ N	Cyanogermane	Ge-C	1.919	-0.084	-0.060	-0.031	xxx
		C-N	1.155	0.003	0.011	0.008	
CGeH ₃ NO	Germyl isocyanate	Ge-H	1.532	-0.017	-0.048	0.007	www
		Ge-N	1.831	0.010	-0.001	-0.030	
GeH ₃ N ₃	Germylazide	Ge-H	1.497	0.019	-0.012	0.057	xxx
		Ge-N	1.866	-0.035	0.010	-0.008	

Table 6.68: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
GeS	Germanium sulfide	Ge-S	2.012	-0.039	-0.087	0.020	ppp
CGeH ₃ NS	Germyl isothiocyanate	Ge-H	1.520	-0.002	-0.037	0.021	iii
		Ge-N	1.817	0.015	0.028	-0.010	
		H-Ge-N	106.9	3.6	1.0	1.3	
		C-N	1.144	0.069	0.054	0.054	
		Ge-F	Germanium fluoride	Ge-F	1.750	-0.039	
GeH ₃ F	Fluorogermane	Ge-F	1.734	0.004	0.003	-0.019	iii
		Ge-H	1.523	-0.011	-0.039	0.017	
		F-Ge-H	106.0	3.9	2.6	2.3	
		C-Ge-F	106.0	1.5	1.7	-1.7	
CGeH ₅ F	Methylgermanium fluoride dihydride	Ge-F	1.739	-0.003	0.003	-0.023	ppp
		Ge-C	1.927	0.034	0.004	0.044	
		C-Ge-F	106.0	1.5	1.7	-1.7	
		Ge-H	1.523	-0.012	-0.038	0.015	
		C-Ge-H	113.9	-4.3	-2.5	-3.1	
GeF ₂	Germanium difluoride	Ge-F	1.732	-0.037	-0.017	-0.062	yyy
		F-Ge-F	97.2	14.5	-2.0	-1.7	
C ₂ GeH ₆ F ₂	Dimethylgermanium difluoride	Ge-F	1.739	-0.015	0.004	-0.029	zzz
		F-Ge-F	105.4	-1.7	-3.7	-4.2	
		Ge-C	1.928	0.027	0.008	0.038	
		C-Ge-F	107.4	2.3	2.3	1.8	
CGeH ₃ F ₃	Trifluoromethylgermane	Ge-C	1.904	0.046	0.032	0.065	iii
		Ge-F	1.714	0.000	0.024	-0.007	
		F-Ge-C	113.2	0.0	0.8	1.3	
		Ge-Cl	2.150	0.046	0.098	-0.018	
Ge-H	1.537	-0.033	-0.060	0.009			
H-Ge-H	111.0	-0.9	1.3	1.1			
C ₃ GeH ₉ Cl	Trimethylchlorogermane	Ge-Cl	2.170	0.033	0.088	-0.026	A
		Ge-C	1.940	0.011	-0.010	0.040	
		C-Ge-Cl	106.6	0.5	0.0	1.4	
		Ge-Cl	2.067	0.067	0.189	0.105	
Ge-F	1.688	0.014	0.040	0.015			
F-Ge-F	107.5	0.1	-0.7	-2.0			
GeCl ₂	Germanium dichloride	Ge-Cl	2.186	-0.198	0.027	-0.089	yyy
		Cl-Ge-Cl	100.4	79.6	4.3	13.6	
C ₂ GeH ₆ Cl ₂	Dimethylgermanium dichloride	Ge-Cl	2.143	0.039	0.099	-0.004	B
		Cl-Ge-Cl	105.0	1.9	1.1	3.3	
		Ge-C	1.928	0.014	-0.002	0.054	
		C-Ge-Cl	108.0	0.8	0.7	0.9	
		Ge-Cl	2.114	0.040	0.108	0.015	
CGeH ₃ Cl ₃	Trichloromethylgermane	Ge-Cl	2.135	0.028	0.097	0.001	xxx
		C-Ge-Cl	106.0	4.6	5.6	4.2	
		Ge-Cl	2.113	0.037	0.112	0.024	
GeCl ₄	Germanium tetrachloride	Ge-Cl	2.297	0.035	0.069	-0.041	iii
		Ge-H	1.527	-0.026	-0.051	0.025	
		H-Ge-Br	106.3	1.6	0.7	3.2	
C ₃ GeH ₉ Br	Bromotrimethylgermane	Ge-Br	2.323	0.039	0.051	-0.055	xxx
		Ge-C	1.936	0.012	-0.006	0.055	
		Br-Ge-C	106.3	0.1	0.6	4.2	

Table 6.69: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
GeBr ₂	Germanium dibromide	Ge-Br	2.337	-0.015	-0.016	-0.157	D
		Br-Ge-Br	101.2	11.3	4.2	78.7	
CGeH ₃ Br ₃	Tribromomethylgermane	Ge-C	1.889	0.047	0.036	0.133	iii
		Ge-Br	2.276	0.056	0.075	-0.010	
		C-Ge-Br	111.6	-1.0	-1.4	-5.8	
		C-H	1.120	-0.028	-0.014	-0.015	
GeBr ₄	Germanium tetrabromide	Ge-Br	2.272	0.041	0.067	-0.003	E
GeH ₃ I	Iodogermane	Ge-I	2.508	-0.036	0.023	-0.075	xxx
GeI ₄	Germanium tetraiodide	Ge-I	2.500	-0.032	-0.003	-0.090	ooo
GeH ₆ Si	Germylsilane	Ge-Si	2.357	0.047	-0.017	0.002	xxx
		Ge-H	1.529	-0.031	-0.047	0.020	
		Si-H	1.483	0.005	-0.104	-0.025	
Ge ₂ H ₆	Digermane	GeGe	2.403	-0.010	0.121	-0.037	ss
		GeH	1.541	-0.038	-0.059	0.007	
		HGeH	106.4	4.0	1.4	2.0	
C ₆ Ge ₂ H ₁₈ O	Bis(trimethylgermanium) oxide	Ge-O	1.770	0.016	-0.016	0.079	F
		Ge-O-Ge	141.0	-15.8	38.9	-19.4	
		Ge-C	1.980	-0.005	-0.038	-0.011	
SnH ₄	Tin tetrahydride (stannane)	Sn-H	1.701	0.000	-0.115	G	
CSnH ₆	Methyltin trihydride	Sn-C	2.140	0.047	-0.083	iii	
		Sn-H	1.708	-0.007	-0.122		
		H-Sn-C	109.4	0.7	1.2		
C ₂ SnH ₈	Dimethyltin dihydride	Sn-H	1.680	0.019	-0.092	H	
		Sn-C	2.150	0.027	-0.091		
		H-Sn-C	108.0	1.8	1.4		
		C-Sn-C	104.8	4.5	7.6		
C ₃ SnH ₁₀	Trimethyltin hydride	Sn-H	1.705	-0.008	-0.117	I	
		Sn-C	2.147	0.018	-0.086		
		H-Sn-C	111.5	-1.8	-3.6		
		Sn-C	2.134	0.014	-0.071		
C ₄ SnH ₁₂	Tetramethyltin	Sn-C	2.134	0.014	-0.071	J	
SnO	Tin oxide	SnO	1.833	0.006	-0.084	ppp	
SnS	Tin sulfide	SnS	2.209	-0.071	-0.208	ppp	
SnH ₃ Cl	Tin chloride trihydride	Sn-Cl	2.327	0.069	-0.018	xxx	
SnCl ₂	Tin dichloride	Sn-Cl	2.346	-0.006	-0.075	yyy	
		Cl-Sn-Cl	99.0	-0.1	3.5		
		Sn-Cl	2.327	0.037	-0.020		
		Cl-Sn-Cl	106.2	-0.3	-2.1		
C ₂ SnH ₆ Cl ₂	Dimethyltin dichloride	Sn-C	2.109	0.000	-0.034	H	
		S-Sn-Cl	108.5	-0.2	-0.9		
		Sn-Cl	2.280	0.075	0.004		
		Sn-Cl	2.280	0.075	0.004		
SnH ₃ Br	Tin bromide trihydride	Sn-Br	2.469	-0.016	-0.068	ss	
C ₃ SnH ₉ Br	Trimethyltin bromide	Sn-H	1.767	-0.075	-0.178	L	
		H-Sn-H	112.8	1.2	-0.4		
		Sn-Br	2.490	-0.039	-0.074		
		Sn-C	2.170	-0.056	-0.105		

Table 6.70: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
SnBr ₂	Tin dibromide	Sn-Br	2.512	-0.108	-0.149		yyy
		Br-Sn-Br	100.0	2.5	4.2		
SnBr ₄	Tin tetrabromide	Sn-Br	2.440	0.000	-0.055		M
SnH ₃ I	Tin iodide trihydride	Sn-I	2.674	-0.051	-0.139		xxx
C ₃ SnH ₉ I	Trimethyltin iodide	Sn-I	2.720	-0.052	-0.168		L
SnI ₂	Tin diiodide	Sn-I	2.706	-0.062	-0.210		yyy
		I-Sn-I	103.8	5.7	3.3		
PbH	Lead hydride	Pb-H	1.839	-0.110	-0.181		a
C ₄ PbH ₁₂	Tetramethyllead	Pb-C	2.240	-0.054	-0.069		N
		C-H	1.080	0.013	0.022		
		H-C-Pb	104.6	6.0	4.5		
PbO	Lead oxide	Pb-O	1.920	0.016	-0.038		O
PbS	Lead sulfide	Pb-S	2.290	-0.138	-0.179		O
PbF	Lead fluoride	Pb-F	2.058	-0.030	-0.063		a
PbF ₂	Lead difluoride	Pb-F	2.033	-0.007	-0.038		P
		F-Pb-F	97.2	-8.3	-5.8		
PbCl	Lead chloride	Pb-Cl	2.180	0.203	0.198		a
PbCl ₂	Lead dichloride	Pb-Cl	2.444	0.014	-0.064		Q
		Cl-Pb-Cl	98.3	1.3	2.5		
PbCl ₄	Lead tetrachloride	Pb-Cl	2.430	-0.154	-0.049		a
PbBr	Lead bromide	Pb-Br	2.546	0.017	-0.080		a
PbBr ₂	Lead dibromide	Pb-Br	2.597	-0.032	-0.127		P
		Br-Pb-Br	99.2	3.4	2.7		
PbI	Lead iodide	Pb-I	2.736	0.037	-0.155		a
PbI ₂	Lead diiodide	Pb-I	2.804	-0.026	-0.206		P
		I-Pb-I	99.7	8.2	4.4		
C ₆ Pb ₂ H ₁₈	Hexamethyldiplumbane	Pb-Pb	2.880	-0.012	-0.110		R
		Pb-C	2.250	-0.050	-0.073		
		Pb-Pb-C	109.5	2.8	2.2		
H ₃ P	Phosphine	PH	1.420	-0.096	-0.080	-0.057	a
		HPH	93.8	3.3	2.3	2.6	
CP	Carbon phosphide	CP	1.562	-0.173	-0.145	-0.029	a
CHP	Methinophosphide	CP	1.542	-0.133	-0.114	-0.132	a
		HC	1.067	0.001	-0.010	-0.003	
CH ₅ P	Methylphosphine	CP	1.858	0.007	-0.108	-0.132	ss
		PH	1.423	-0.087	-0.080	-0.059	
		HPC	97.5	2.5	3.3	2.5	
		HPH	93.4	3.5	2.6	2.8	
C ₃ H ₉ P	Trimethylphosphine	CP	1.843	0.029	-0.081	-0.118	ss
		CPC	98.9	1.7	7.9	3.2	
C ₅ H ₅ P	Phosphole (?)	C1C2	1.384	-0.006	0.013	-0.005	ss
		CP	1.733	-0.042	-0.092	-0.133	
		C2C3	1.413	-0.034	-0.016	-0.034	
PO	Phosphorus oxide	PO	1.476	-0.018	-0.053	-0.039	a
NP	Phosphorus nitride	PN	1.491	-0.077	-0.093	-0.109	a
CH ₂ PF	CH ₂ =P-F	P-F	1.598	-0.030	-0.049	-0.052	S
		C=P	1.644	-0.071	-0.064	-0.106	

Table 6.71: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
PF ₃	Phosphorus trifluoride	PF	1.570	-0.012	-0.014	-0.027	a
		FPF	97.8	-2.0	1.1	0.2	
POF ₃	Phosphous oxyfluoride	PF	1.520	0.009	0.034	0.006	a
		FPF	102.5	-2.7	-0.6	-0.8	
PSF ₃	Phosphorus thiofluoride	PO	1.450	0.001	0.036	0.001	a
		PF	1.530	0.009	0.027	0.000	
		FPF	100.3	-5.2	-0.9	-1.8	
PF ₅	Phosphorus pentafluoride	PS	1.870	0.064	0.106	0.004	a
		PF(ax)	1.577	-0.024	0.025	-0.028	
		PF(eq)	1.534	-0.006	0.039	0.001	
PCl ₃	Phosphorus trichloride	PCI	2.039	0.025	-0.050	-0.120	a
		CIPCI	100.3	-0.6	4.9	5.1	
PCl ₅	Phosphorus pentachloride	PCI(ax)	2.190	-0.097	-0.078	-0.115	a
		PCI(eq)	2.040	0.012	-0.007	-0.069	
PBr ₃	Phosphorus tribromide	PBR	2.220	-0.070	-0.131	-0.119	ss
		BrPBr	101.0	0.3	4.7	5.8	
POBr ₃	Phosphorus oxybromide	P=O	1.455	-0.048	0.021	0.025	iii
		P-Br	2.174	-0.059	-0.046	-0.014	
		O=P-Br	114.2	2.3	-0.4	1.0	
PSBr ₃	Phosphorus thiotribromide	P=S	1.894	-0.053	-0.042	-0.011	iii
		P-Br	2.193	-0.102	-0.077	-0.040	
		S=P-Br	116.2	12.0	-1.5	-1.0	
BH ₃ PF ₃	PH ₃ -BF ₃	P-B	1.836		2.386	0.321	xxx
		P-H	1.207		0.132	0.158	
		H-P-H	115.1		-18.8	-16.6	
GeH ₃ PSF ₂	Difluoro(germylthio)phosphine	Ge-S	2.256	0.031	-0.072	0.058	iii
		P-S	2.115	-0.013	-0.163	-0.009	
		Ge-S-P	99.0	26.4	41.3	17.6	
		Ge-H	1.538	-0.032	-0.061	0.006	
		S-Ge-H	110.0	3.5	-3.2	-1.2	
		P-F	1.590	-0.035	-0.032	-0.052	
Ge ₃ H ₉ P	Trigermylphosphine	S-P-F	99.9	13.4	5.4	9.3	iii
		P-Ge	2.306	-0.070	-0.159	-0.204	
		Ge-P-Ge	95.7	13.5	24.3	24.3	
		Ge-H	1.490	0.016	-0.009	0.031	
		P-Ge-H	110.3	5.6	-3.0	-7.8	
P ₂	Phosphorus dimer	PP	1.894	-0.179	-0.200	-0.271	a
Ge ₂ H ₆ P ₂ F ₂	1,1-Difluoro-2,2-digermyldiphosphan	P-F	1.581	-0.032	-0.028	-0.052	iii
		P-P	2.177	-0.061	-0.209	-0.288	
		P-P-F	98.9	8.1	7.1	0.6	
		Ge-P	2.320	-0.110	-0.141	-0.166	
		Ge-P-F	95.7	33.6	31.9	30.5	
		Ge-P-Ge	98.6	12.3	19.8	12.3	
P ₄	Phosphorus tetramer	Ge-H	1.512	0.000	-0.034	0.018	a
		PP	2.210	-0.013	-0.158	-0.170	
P ₄ O ₆	Phosphorus trioxide	PO	1.650	0.058	-0.046	0.020	a
		OPO	99.0	-2.5	-3.0	-2.3	

Table 6.72: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
P ₄ O ₁₀	Phosphorus pentoxide	P-O	1.600	0.090	0.006	0.004	a
		O-P-O	101.0	-4.7	-3.1	14.2	
AsH ₃	Arsine	As-H	1.513	0.007			T
		H-As-H	92.1	2.1			
C ₃ AsN ₃	Arsenic tricyanide	As-C	1.900	-0.031			ooo
		C-As-C	92.0	6.3			
		As-C-N	171.0	4.9			
AsF ₃	Arsenic trifluoride	As-F	1.706	0.000			ooo
		F-As-F	96.2	-0.3			
AsF ₅	Arsenic pentafluoride	As-F(ax)	1.711	-0.026			ss
		As-F(eq)	1.656	0.005			
C ₃ AsF ₉	Triperfluoromethylarsine	As-C	2.053	0.028			ooo
		C-As-C	100.0	-0.5			
AsCl ₃	Arsenic trichloride	As-Cl	2.161	0.002			ooo
		Cl-As-Cl	98.7	1.1			
AsBr ₃	Arsenic tribromide	As-Br	2.323	-0.008			iii
		Br-As-Br	99.8	0.5			
AsI ₃	Arsenic triiodide	As-I	2.550	-0.041			ooo
		I-As-I	100.2	4.6			
AsH ₉ Si ₃	Trisilylarsine	As-Si	2.353	0.018			iii
		Si-As-Si	94.1	-3.3			
		Si-H	1.470	0.024			
C ₂ ZnH ₆	Dimethylzinc	Zn-C	1.930	0.007	-0.046	-0.031	U
		C-Zn-C	180.0	0.0	0.6	-0.1	
C ₄ ZnH ₁₀	Diethylzinc	Zn-C	1.950	0.017	-0.046	-0.022	U
		C-C-Zn	114.5	-15.8	3.9	0.1	
		C-Zn-C	180.0	0.2	0.1	0.0	
C ₆ ZnH ₈	Cyclopentadienylmethylzinc	C-C	1.438	-0.003	0.012	0.000	V
		C-Zn	2.280	0.061	-0.023	0.078	
		Zn-C(H3)	1.903	0.043	-0.025	-0.023	
C ₆ ZnH ₁₄	Di-n-propylzinc	C-C-C	113.6	-0.8	1.8	-1.7	U
		Zn-C	1.952	0.021	-0.044	-0.020	
		Zn-C-C	114.5	-13.6	3.8	0.6	
		C-Zn-C	180.0	0.0	0.0	0.0	
ZnF ₂	Zinc difluoride	Zn-F	1.742	-0.002	-0.062	0.006	yyy
		F-Zn-F	180.0	0.0	0.1	0.0	
ZnCl ₂	Zinc dichloride	Zn-Cl	2.062	0.002	0.053	0.005	yyy
		Cl-Zn-Cl	180.0	0.0	0.0	0.0	
ZnBr ₂	Zinc dibromide	Zn-Br	2.204	-0.107	0.031	-0.093	yyy
		Br-Zn-Br	180.0	0.0	0.0	0.0	
ZnI ₂	Zinc diiodide	Zn-I	2.401	0.003	-0.019	-0.055	yyy
		I-Zn-I	180.0	0.0	-0.1	0.0	
C ₂ CdH ₆	Dimethylcadmium	Cd-C	2.112	-0.077			xxx
C ₄ CdN ₄ S ₄	Cd(NCS) ₃ (SCN) (=)	Cd-N	2.240	-0.022			W
		Cd-S	2.570	-0.001			
		Cd-F	1.970	-0.004			ooo
CdCl ₂	Cadmium dichloride	Cd-Cl	2.210	0.015			ooo

Table 6.73: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
CdBr ₂	Cadmium dibromide	Cd-Br	2.394	-0.034			X
CdI ₂	Cadmium diiodide	Cd-I	2.550	0.038			ooo
HgH	Mercury hydride	Hg-H	1.740	-0.044	-0.190	-0.069	a
HgO	Mercury oxide	Hg-O	1.840	0.001	0.042	0.222	www
C ₂ HgH ₃ N	Methylmercuric cyanide	Hg-CN	2.082	-0.055	-0.148	-0.070	Y
		Hg-CH ₃	2.051	0.056	-0.069	-0.005	
		C-N	1.141	0.016	0.030	0.026	
HgF	Mercury fluoride	Hg-F	1.890	0.010	-0.019	0.003	www
HgF ₂	Mercury difluoride	Hg-F	1.960	-0.032	-0.082	-0.052	www
C ₂ HgF ₆	Ditrifluoromethyl mercury	Hg-C	2.101	-0.003	0.162	0.110	Z
		C-Hg-C	180.0	0.0	0.3	0.2	
HgCl	Mercury chloride	Hg-Cl	2.230	-0.036	0.049	0.010	a
CHgH ₃ Cl	Methylmercuric chloride	Hg-C	2.052	0.060	-0.070	-0.003	xxx
		Hg-Cl	2.285	-0.031	0.006	-0.031	
HgCl ₂	Mercury dichloride	Hg-Cl	2.252	-0.007	0.016	-0.013	AA
HgBr	Mercury bromide	Hg-Br	2.330	-0.116	0.038	-0.101	www
		Hg-C	2.062	0.028	-0.079	-0.012	iii
CHgH ₃ Br	Methylmercuric bromide	Hg-Br	2.405	-0.114	-0.021	-0.148	
		C-H	1.095	-0.008	0.013	0.013	
		Hg-C-H	109.6	-0.1	0.1	-1.8	
HgBr ₂	Mercury dibromide	Hg-Br	2.440	-0.215	-0.077	-0.190	BB
HgI	Mercury iodide	Hg-I	2.490	10.156	-0.024	0.036	www
CHgH ₃ I	Methylmercuric iodide	Hg-C	2.069	0.008	-0.082	-0.015	iii
		Hg-I	2.588	-0.087	-0.109	-0.094	
HgI ₂	Mercury diiodide	Hg-I	2.554	-0.080	-0.079	-0.058	CC
SbH ₃	Stibine	Sb-H	1.707	-0.005			ooo
		H-Sb-H	91.3	1.1			
SbF ₅	Antimony pentafluoride	Sb-F(ax)	2.430	-0.455			ooo
		Sb-F(eq)	2.310	-0.348			
C ₃ SbF ₉	Triperfluoromethylstibine	Sb-C	2.202	0.007			ooo
		C-Sb-C	100.0	-0.8			
SbCl ₃	Antimony trichloride	Sb-Cl	2.323	-0.003			iii
		Cl-Sb-Cl	97.1	-0.1			
SbCl ₅	Antimony pentachloride	Sb-Cl(ax)	2.338	0.035			yyy
		Sb-Cl(eq)	2.277	0.073			
SbBr ₃	Antimony tribromide	Sb-Br	2.490	-0.019			ooo
		Br-Sb-Br	98.0	0.4			
SbH ₉ Si ₃	Trisilylstibine	Sb-Si	2.555	-0.022			iii
		Si-Sb-Si	89.0	5.8			
		Si-H	1.470	0.018			
Sb ₂	Antimony, dimer	Sb-Sb	2.590	-0.289			DD
SeH ₂	Hydrogen selenide	SeH	1.460	0.010			ooo
		HSeH	91.0	2.6			
CSe	Selenium carbide	Se-C	1.676	-0.085			ppp
CSeH ₄	Methylselenium hydride	CSe	1.959	-0.012			ss
		CH	1.088	0.005			
		SeH	1.473	-0.003			

Table 6.74: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.		Errors		Ref.
				PM3	MNDO	AM1	
C ₂ SeH ₆	Ethyl selenol (anti)	Se-H	1.440	0.026			EE
		Se-C	1.962	-0.006			
		H-Se-C	93.5	6.4			
		Se-C-C	108.7	-8.5			
C ₂ SeH ₆	Dimethylselenium	Se-C	1.943	0.005			FF
		C-Se-C	96.2	4.6			
C ₄ SeH ₄	Selenophene	Se-C2	1.855	0.032			xxx
		C5-Se-C2	87.8	0.1			
		C2-C3	1.369	-0.025			
		Se-C2-C3	111.6	-1.5			
		C2-H	1.070	0.014			
C ₄ SeH ₈	Tetrahydroselenophene	Se-C2-H2	121.7	-0.6			xxx
		Se-C2	1.963	-0.004			
		C5-Se-C2	90.7	1.5			
		C2-C3	1.549	-0.043			
SeO	Selenium oxide	Se-O	1.663	-0.063			GG
		SeO ₂	Selenium dioxide	SeO	1.608	-0.003	
SeO ₃	Selenium trioxide	OSeO	113.8	-7.2			iii
		Se-O	1.688	-0.150			
CSeHN	Isoselenocyanic acid (Se=C=N-H)	Se=C	1.719	-0.079			iii
		C=N	1.191	0.005			
		Se=C=N	175.0	0.0			
		N-H	0.990	-0.014			
		C=N-H	140.0	15.1			
CSeS	Carbon sulfide selenide	C-S	1.533	-0.081			xxx
		C-Se	1.695	-0.110			
CSeF ₂	Selenocarbonyl difluoride	C=Se	1.743	0.002			iii
		C-F	1.314	0.006			
		F-C-F	107.5	-2.2			
SeOF ₂	Selenyl fluoride	SeO	1.576	0.049			ss
		SeF	1.730	-0.006			
		FSeO	104.8	-3.4			
		FSeF	92.2	2.3			
SeO ₂ F ₂	Selenoyl fluoride	Se-F	1.685	0.020			iii
		F-Se-F	94.1	1.0			
		Se-O	1.575	-0.017			
		O-Se-O	126.2	3.7			
SeF ₄	Selenium tetrafluoride	SeF	1.770	-0.034			ooo
		FSeF	169.2	-28.9			
		SeF'	1.680	0.028			
		F'SF'	100.5	11.5			
CSeNOF ₅	Pentafluoro(isocyanato)selenium	Se-F	1.677	0.032			iii
		Se-N	1.789	0.053			
		N=C	1.260	0.007			
		Se-N=C	116.9	12.3			
		C=O	1.187	-0.020			
		N=C=O	172.9	-0.6			

Table 6.75: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.	Errors			Ref.
				PM3	MNDO	AM1	
SeF ₆	Selenium hexafluoride	Se-F	1.685	0.005			HH
C ₂ SeF ₆	Diperfluoromethyl selenide	Se-C	1.960	0.047			ooo
		C-Se-C	104.0	-2.8			
SeCl ₂	Selenium dichloride	Se-Cl	2.157	0.007			iii
		Cl-Se-Cl	99.6	0.0			
SeOCl ₂	Selenyl chloride	Se=O	1.614	0.020			iii
		Se-Cl	2.205	-0.005			
		O=Se-Cl	106.0	-3.0			
		Cl-Se-Cl	96.9	1.2			
In ₂ Se	Indium(I) selenide	In-Se	2.650	-0.070			iii
		In-Se-In	113.0	66.3			
SeSi	Silicon selenide	Se-Si	2.058	-0.084			xxx
SeH ₆ Si ₂	Disilyl selenide	Se-Si	2.270	0.123			ooo
		Si-Se-Si	97.0	-0.1			
GeSe	Germanium selenide	Ge-Se	2.135	-0.208			xxx
SnSe	Tin selenide	Sn-Se	2.326	-0.002			xxx
PbSe	Lead selenide	Pb-Se	2.402	-0.041			xxx
SeHPF ₂	Difluorophosphine selenide	P=Se	2.026	0.062			iii
		P-F	1.557	-0.006			
		Se=P-F	116.8	-1.0			
		P-H	1.422	-0.158			
		Se=P-H	118.6	-0.3			
		F-P-F	98.1	-4.3			
		Se-C	1.950	-0.005			
C ₂ Se ₂ H ₆	Me-Se-Se-Me	Se-Se	2.330	0.040			ooo
		Se-Se-C	99.0	1.2			
Se ₂ OF ₁₀	Bis(pentafluoroselenium)oxide	Se-O	1.679	0.048			iii
		Se-O-Se	142.4	-16.4			
		Se-F(eq)	1.683	0.008			
		Se-F(ax)	1.665	0.031			
C ₇ GeSe ₂ H ₁	Tetramethyldiselenagermacyclohexane	Se-C	-10.8	-4.2			II
		Ge-Se	108.7	2.8			
		Ge-Se-C	116.3	-2.1			
		Se-Ge-Se	-114.7	0.5			
		Ge-C'	100.2	-0.4			
		Ge-C''	119.2	-0.5			
TeH ₂	Hydrogen telluride	C'-Ge-C''	-24.2	-4.5			
		Te-H	1.658	0.017			JJ
		H-Te-H	90.3	-2.0			
TeO ₂	Tellurium dioxide	Te-O	1.830	-0.128			iii
CTeNOF ₅	Pentafluoro(isocyanato)tellurium	Te-F	1.826	-0.002			iii
		Te-N	1.859	-0.045			
		Te-N=C	126.5	53.6			
TeF ₆	Tellurium hexafluoride	Te-F	1.815	0.001			iii
TeCl ₄	Tellurium tetrachloride	Te-Cl	2.330	0.072			tt
TeBr ₂	Tellurium dibromide	Te-Br	2.510	0.000			ooo
		Br-Te-Br	98.0	1.5			

Table 6.76: Comparison of Calculated and Observed Geometries for MNDO, AM1, and PM3 (contd.)

Empirical Formula	Chemical Name	Geometric Variable	Exp.		Errors		Ref.
				PM3	MNDO	AM1	
TeBr ₄	Tellurium tetrabromide	Te-Br	2.680	-0.149			tt
In ₂ Te	Indium(I) telluride	In-Te	2.840	-0.004			iii
		In-Te-In	99.0	6.0			
GeTe	Germanium telluride	Ge-Te	2.340	-0.338			ppp
SnTe	Tin telluride	Sn-Te	2.523	0.049			ppp
PbTe	Lead telluride	Pb-Te	2.595	0.141			xxx
C ₉ TeH ₁₃ PS	MeOPh-Te-S-P(S)(OMe) ₂	Te-C	2.114	-0.027			KK
		Te-S	2.444	0.234			
		C-Te-S	94.9	10.8			
		S-P	2.051	0.052			
		P-S-Te	103.3	-13.9			
		P=S	1.933	0.061			
		S=P-S	107.9	-1.4			
Te ₂	Tellurium, dimer	Te-Te	2.560	0.145			ppp
Te ₂ OF ₁₀	Bis(pentafluorotellurium)oxide	Te-O	1.832	-0.057			iii
		Te-O-Te	145.5	10.4			
		Te-Feq	1.820	-0.009			
		Te-Fax	1.799	0.015			
		Fax-Te-Feq	89.9	4.4			
C ₃ BiH ₉	Trimethylbismuth	Bi-C	2.270	-0.004			ooo
		C-Bi-C	96.7	0.3			
BiCl ₃	Bismuth trichloride	Bi-Cl	2.425	-0.006			yyy
		Cl-Bi-Cl	97.3	2.7			
BiBr ₃	Bismuth tribromide	Bi-Br	2.630	-0.026			ooo
		Br-Bi-Br	100.0	-1.4			

bf References

- a: M. W. Chase, C. A. Davies, J. R. Downey, D. R. Frurip, R. A. McDonald, A. N. Syverud, JANAF Thermochemical Tables, Third Edition, J. Phys. Chem. Ref. Data 14, Suppl. 1 (1985).
- b: G. Herzberg, "Molecular Spectra and Molecular Structure III. Electronic Spectra and Electronic Structure of Polyatomic Molecules", Van Nostrand, New York, N.Y., 1966.
- c: A. G. Maki, R. A. Toth, J. Mol. Spectr., 17, 136 (1965).
- d: W. M. Stigliani, V. W. Laurie, J. C. Li, J. Chem. Phys., 62, 1890 (1975).
- e: R. A. Bonham, L. S. Bartell, J. Am. Chem. Soc., 81, 3491 (1959).
- f: O. Bastiansen, F. N. Fritsch, K. Hedberg, Acta. Crystallogr., 17, 538 (1964).
- g: D. R. Lide, D. Christensen, J. Chem. Phys., 35, 1374 (1961).
- h: J. H. Callomon, B. P. Stoicheff, Can. J. Phys., 35, 373 (1957).
- i: A. Almenningen, O. Bastiansen, M. Traetteberg, Acta Chem. Scand., 15, 1557 (1961).
- j: T. Fukuyama, K. Kuchitsu, Y. Morino, Bull. Chem. Soc. Jap., 42, 379 (1969).
- k: K. W. Cox, M. D. Harmony, G. Nelson, K. B. Wiburg, J. Chem. Phys., 50, 1976 (1969).
- l: W. Haugen, M. Traetteberg, Acta Chem. Scand., 20, 1226 (1966).
- m: A. Almenningen, I. M. Anfinson, A. Haaland, Acta Chem. Scand., 24, 43 (1970).
- n: S. Meiboom, L. C. Snyder, J. Chem. Phys., 52, 3857 (1970).
- o: L. R. H. Scharpen, V. W. Laurie, J. Chem. Phys., 39, 1972 (1963).
- p: D. R. Lide, J. Chem. Phys., 33, 1519 (1960).
- q: B. W. McClland, K. Hedberg, J. Am. Chem. Soc., 109, 7304 (1987).
- r: K. Tamagawa, T. Iijima, M. Kimura, J. Mol. Struct., 30, 243 (1976).
- s: P. A. Baron, R. D. Brown, F. R. Burden, P. J. Domaille, J. E. Kent, J. Mol. Spectrosc., 43, 401 (1970).
- t: J. F. Chiang, S. H. Bauer, J. Am. Chem. Soc., 91, 1898 (1969).
- u: O. Bastiansen, L. Fernholt, H. M. Seip, H. Kambara, K. Kuchitsu, J. Mol. Struct., 18, 163 (1973).
- v: G. Herzberg, "Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules", 2nd ed, Van Nostrand, New York, N.Y., 1950.
- w: K. Takagi, T. Oka, J. Phys. Soc. Jap., 18, 1174 (1963).
- x: R. M. Lees, J. G. Baker, J. Chem. Phys., 48, 5299 (1968).
- y: A. P. Cox, L. F. Thomas, J. Sheridan, Spectrochim. Acta, 15, 542 (1959).
- aa: U. Blukis, P. H. Kasai, R. J. Myers, J. Chem. Phys., 38, 2753 (1963).
- bb: K. Kuchitsu, T. Fukuyama, Y. Marino, J. Mol. Struct., 1, 463 (1968).

- cc: R. Nelson, L. Pierce, *J. Mol. Spectrosc.* 18, 344 (1965).
- dd: B. Bak, D. Christensen, W. B. Dixon, L. Hansen-Nygaard, J. Rastrup-Anderson, M. Shottlander, *J. Mol. Spectrosc.*, 9, 124 (1962).
- ee: W. C. Oelfke, W. Gordy, *J. Chem. Phys.*, 51, 5336 (1969).
- ff: G. H. Kwei, R. F. Curl, *J. Chem. Phys.*, 32, 1592 (1960).
- gg: R. F. Curl, Jr. *J. Chem. Phys.*, 30, 1529 (1959).
- hh: J. Trotter, *Acta Crystallogr.*, 13, 86 (1960).
- ii: D. G. Lister, J. K. Tyler, J. H. Hog, N. W. Larson, *J. Mol. Struct.*, 23, 253 (1974).
- jj: D. R. Lide, *J. Chem. Phys.*, 27, 343 (1957).
- kk: C. C. Costain, *J. Chem. Phys.*, 29, 864 (1958).
- ll: J. E. Wollrab, V. W. Laurie, *J. Chem. Phys.*, 48, 5058 (1968).
- mm: J. E. Wollrab, V. W. Laurie, *J. Chem. Phys.*, 51, 1580 (1969).
- nn: D. R. Johnson, F. J. Lovas, W. H. Kirchhoff, *J. Phys. Chem. Ref. Data*, 1, 1011 (1972).
- oo: J. O. Morley, *J. Chem. Soc. Perkin Trans. II*, (1987), (in press).
- pp: S. Cradock, P. B. Liescheski, D. W. H. Rankin, H. E. Robertson, *J. Am. Chem. Soc.*, 110, 2758 (1988).
- qq: J. K. Tyler, *J. Mol. Spectrosc.*, 11, 39 (1963).
- rr: J. E. Lancaster, B. J. Stoicheff, *Can. J. Phys.*, 34, 1016 (1965).
- ss: J. H. Callomon, E. Hirota, K. Kuchitsu, W. J. Lafferty, A. G. Maki, C. S. Pote, "Structure Data on Free Polyatomic Molecules", *Landolt-Bornstein, New Series, Group II, Vol. 7*, Springer, Berlin (1976).
- tt: L.E. Sutton, "Tables of Interatomic Distances and Configurations in Molecules and Ions", *Special Publication No. 11 + 18, Chem. Soc., London* (1958), (1965).
- uu: R. W. Kilb, *J. Chem. Phys.*, 23, 1736 (1955).
- vv: B. Bak, D. Christensen, L. Hansen-Nygaard, J. Rastrup-Anderson, *J. Mol. Spectrosc.*, 7, 58 (1961).
- ww: J. K. G. Watson, *J. Mol. Spectrosc.*, 48, 479 (1973).
- xx: D. W. W. Anderson, D. W. H. Rankin, A. Robertson, *J. Mol. Struct.*, 14 385 (1972).
- yy: P. Rosmus, H. Stafast, H. Bock, *Chem. Phys. Lett.*, 34, 275 (1975).
- zz: G. Winnewisser, M. Winnewisser, W. Gordy, *J. Chem. Phys.*, 49, 3465 (1968).
- aaa: R. Sutter, H. Driezler, F. Z. Rudolph, *Naturforschung A*, 20 1676 (1965).
- bbb: J. Donohue, A. Caron, E. Goldish, *J. Am. Chem. Soc.*, 83, 3748 (1961).
- ccc: A. Caron, J. Donohue, *Acta Crystallogr.*, 18, 562 (1965).
- ddd: G. A. Kuipers, D. F. Smith, A. N. Nielsen, *J. Chem. Phys.*, 25, 275 (1956).

- eee: J. L. Duncan, *J. Mol. Struct.*, 6, 447 (1970).
- fff: J. K. Tyler, J. Sheridan, *Trans. Faraday Soc.*, 59, 2661 (1963).
- ggg: J. L. Carlos, R. R. Karl, S. H. Bauer, *J. Chem. Soc., Faraday Trans. 2*, 177 (1974).
- hhh: T. Ogata, K. Fujii, M. Yoshikawa, F. Hirota, *J. Am. Chem. Soc.*, 109, 7639 (1987).
- iii: J. H. Callomon, E. Hirota, T. Iijima, K. Kuchitsu, W. J. Lafferty, "Structure Data on Free Polyatomic Molecules", *Landolt-Bornstein, New Series, Group II, Vol. 15*, Springer, Berlin (1977).
- jjj: A. S. Rodgers, J. Chao, R. C. Wilhoit, B. J. Zwolinski, *J. Phys. Chem. Ref. Data*, 3, 117 (1974).
- kkk: Private Communication by Dr Dave Dixon, DuPont.
- lll: H. Jones, M. Takami, J. Sheridan, *Z. Naturforsch.*, A33, 156 (1978).
- mmm: S. S. Chen, A. C. Wilhoit, B. J. Zwolinski, *J. Phys. Chem. Ref. Data* 5, 571 (1975).
- nnn: I. Hargittai, *Acta Chem. (Budapest)*, 49, 351 (1969).
- ooo: "Structural Inorganic Chemistry", A. F. Wells, Clarendon Press, Oxford, 1984
- ppp: K. P. Huber, G. Herzberg, "Molecular Spectra and Molecular Structure", IV. "Constants for Diatomic Molecules", Van Nostrand, Reinhold, New York, (1979).
- qqq: H. Jones, J. Sheridan, O. Stiefvater, *Z. Naturforsch.*, A32, 866 (1977).
- rrr: N. P. C. Westwood, W. Lewis-Bevan, M. C. L. Gerry, *J. Mol. Spectr.*, 136, 93 (1989).
- sss: Y. Niide, I. Ohkoshi, *J. Mol. Spectr.*, 136, 17 (1989).
- ttt: K. Hagen, B. Lunelli, *J. Phys. Chem.*, 93, 1326 (1989).
- uuu: U. Anderson, N. Heineking, H. Dreizler, *J. Mol. Struct.* 137, 296 (1989).
- vvv: N. N. Greenwood, A. Earnshaw, "Chemistry of the Elements", Pergamon Press, Oxford, p.976 (1984).
- www: D. R. Stull, H. Prophet, *Natl. Stand., Ref. Data Ser. (U.S., Natl. Bur. Stand.) NSRDS-NBS 37*, 1971.
- xxx: M. D. Harmony, V. W. Laurie, R. L. Kuczkowsky, R. H. Schwendeman, D. A. Ramsay, F. J. Lovas, W. J. Lafferty, A. G. Maki, *J. Phys. Chem. Ref. Data* 8, 3 (1979).
- yyy: M. Hargittai, *Coord. Chem. Rev.*, 91, 35 (1988).
- zzz: L. J. Guggenberger, R. E. Rundle, *J. Am. Chem. Soc.*, 86, 5344 (1964).
- aaaa: G. Herzberg, J. W. C. Johns, *Proc. R. Soc. London, Ser. A.*, 298, 142 (1967).
- bbbb: D. Christen, D. G. Lister, J. Sheridom, *J. Chem. Soc., Faraday Trans. 2*, 70, 1953 (1974).
- cccc: S. G. W. Ginn, J. K. Kennedy, J. Overend, *J. Chem. Phys.*, 48, 1571 (1968).
- dddd: L. A. Curtiss, J. A. Pople, *J. Chem. Phys.*, 90, 4314 (1989).
- eeee: C. H. Chang, R. F. Porter, S. H. Bauer, *Inorg. Chem.*, 8, 1689 (1969).
- ffff: W. Harshbarger, G. Lee, R. F. Porter, S. H. Bauer, *Inorg. Chem.* 8, 1683 (1969).

- gggg: R. A. Beaudet, R. L. Poynter, *J. Chem. Phys.*, 53, 1899 (1970).
- hhhh: D. Schwock, A. B. Burg, R. A. Beaudet, *Inorg. Chem.* 16, 3219 (1977).
- iiii: E. Switkes, I. R. Epstein, J. A. Tossell, R. M. Stevens, W. N. Lipscomb, *J. Am. Chem. Soc.*, 92, 3837 (1970).
- jjjj: nI. R. Epstein, J. A. Tossell, E. Switkes, I. R. M. Stevens, W. N. Lipscomb, *Inorg. Chem.*, 10, 171, (1971).
- llll: M. J. Goode, A. J. Downs, C. R. Pulham, D. W. H. Rankin, H. E. Robertson, *J. Chem. Soc. Chem. Comm.*, 768 (1988).
- mmmm: S. N. Vempati, W. E. Jones, *J. Mol. Spect.*, 127, 232 (1988).
- nnnn: J. H. Meadows, H. F. Schaefer III, *J. Am. Chem. Soc.*, 98, 4383 (1976).
- oooo: J. Nakagawa, Y. Shiko, M. Hayashi, *J. Mol. Spect.*, 122, 1 (1987).
- pppp: M. Hayashi, S. Kaminaka, M. Fujitake, S. Miyazaki, *J. Mol. Spect.*, 135, 289 (1989).
- qqqq: R. W. Davis, M. C. L. Gerry, *J. Mol. Spectrosc.* 60, 117 (1976).
- rrrr: A. P. Cox, I. C. Ewart, T. R. Gayton, *J. Mol. Spect.*, 125, 76 (1987).
- ssss: K. C. Shotton, A. G. Lee, W. J. Jones, *J. Raman Spectrosc.*, 1, 243 (1973).
- tttt: R. A. Eades, D. A. Dixon, *J. Chem. Phys.*, 72, 3309 (1980).
- uuuu: J. Durig, M. M. Chen, Y. S. Li, *J. Phys. Chem.*, 77, 227 (1973), and E. C. Thomas, V. W. Laurie, *J. Chem. Phys.*, 50, 3512 (1969).
- vvvv: J. L. Hencher, F. J. Mistoe, *Can. J. Chem.*, 53, 3542 (1975).
- wwww: J. D. Murdoch, D. W. H. Rankin, B. Beagley, *J. Mol. Struct.*, 31, 291, (1976).
- xxxx: P. Groner, G. M. Attia, A. B. Mohamad, J. F. Sullivan, Y. S. Li, J. R. Durig, *J. Chem. Phys.*, 91, 1434 (1989).
- yyyy: S. Saito, Y. Endo, E. Hirota, *J. Mol. Spect.*, 116, 499 (1986).
- zzzz: J. E. Drake, R. T. Hemmings, J. L. Hencher, F. M. Mustoe, Q. Shem, *J. Chem. Soc., Dalton Trans.*, 394, (1976).
- A: J. Durig, K. L. Hellams, *J. Mol. Struct.*, 29, 349 (1975).
- B: J. E. Drake, H. L. Hencher, Q. Shem, *Can. J. Chem.*, 55, 1104 (1977).
- C: Y. Morino, Y. Nakamura, T. Iijima, *J. Chem. Phys.*, 32, 643 (1960).
- D: G. Schultz, J. Tremmel, I. Hargittai, N. D. Kagramanov, A. K. Maltsev, O. M. Nefedov, *J. Mol. Struct.*, 82, 107 (1982).
- E: G. G. B. Souza, J. D. Wieser, *J. Mol. Struct.*, 25, 442 (1975).
- F: L. V. Vilkov, N. A. Tarensenko, *Zh. Strukt. Khim.*, 10, 102 (1969).
- G: G. R. Wilkinson, M. K. Wilson, *J. Chem. Phys.*, 25, 784 (1956).
- H: B. Beagley, K. McAloon, J. M. Freeman, *Acta Crystallogr., Sect. B*, B30, 444 (1974).

- I: H. C. Clark, S. G. Furnival, J. T. Kwon, *Can. J. Chem.*, 41, 2889 (1963).
- J: H. Fujii, M. Kimura, *Bull. Chem. Soc. Japan*, 44, 2643 (1970)
- K: H. Fujii, M. Kimura, *Bull. Chem. Soc. Japan*, 43, 1933 (1970)
- L: H. A. Skinner, L. E. Sutton, *Trans. Farad. Soc.*, 44, 164 (1944).
- M: E. Coop, L. E. Sutton, *J. Chem. Soc.*, 1269 (1938).
- N: T. Oyamada, T. Iijima, M. Kimura, *Bull. Chem. Soc. Japan*, 44, 2638 (1971).
- O: F. J. Lovas, E. Tiemann, *J. Phys. Chem. Ref. Data* 3, 609 (1974).
- P: A. V. Demidov, A. G. Gershikov, E. Z. Zasorin, V. P. Spiridonov, A. A. Ivanov, *J. Struct. Chem.*, 24, 9 (1983).
- Q: I. Hargittai, *J. Mol. Struct.*, 100, 129 (1983).
- R: H. A. Skinner, L. E. Sutton, *Trans. Farad. Soc.*, 36, 1209 (1940).
- S: O. I. Osman, B. J. Whitaker, N. P. C. Simmons, D. R. M. Walton, J. F. Nixon, H. W. Krojo, *J. Mol. Spect.*, 137, 373 (1989).
- T: G. A. McRae, M. C. L. Gerry, M. Wong, I. Ozier, E. A. Cohen, *J. Mol. Spect.*, 123, 321 (1987).
- U: A. Almenningen, T. U. Halgaker, A. Haaland, S. Samdal, *Acta Chem. Scand.*, Part A. A36, 159 (1982).
- V: A. Haaland, S. Samdal, R. Seip, *J. Organomet. Chem.*, 153, 187 (1978).
- W: K. Ozutsumi, T. Takamuka, S. Ishiguko, H. Ohtaki, *Bull. Chem. Soc. Japan*, 62, 1875 (1989).
- X: V. M. Petrov, A. N. Utkin, G. V. Girichev, A. A. Ivanov, *Zh. Strukt. Khim.*, 26, 52 (1985).
- Y: N. Iwasaki, *Bull. Chem. Soc. Japan*, 49, 2735 (1976).
- Z: H. Oberhammer, *J. Mol. Struct.*, 48, 389 (1978).
- AA: K. Kashiwabara, S. Konaka, N. M. Kimura, *Bull. Chem. Soc. Japan*, 46, 410 (1973).
- BB: P. W. Allen, L. E. Sutton, *Acta Crystallogr.* 3, 46 (1950).
- CC: V. P. Spiridonov, A. G. Gershikov, B. S. Butayev, *J. Mol. Struct.*, 52, 53 (1979).
- DD: K. Balasubramanian, J. Li, *J. Mol. Spect.*, 135, 169 (1989).
- EE: J. Nakagawa, H. Okutani, M. Hayashi, *J. Mol. Spectrosc.*, 94, 410 (1982).
- FF: J. F. Beecher, *J. Mol. Spectrosc.*, 21, 414 (1966).
- GG: E. H. Fink, K. D. Setzer, *J. Mol. Spect.*, 125, 66 (1987).
- HH: L. S. Bartell, A. Jin, *J. Mol. Struct.*, 118, 47 (1984).
- II: S. Tomoda, M. Shimoda, M. Sanami, Y. Takeuchi, Y. Litaka, *J. Chem. Soc. Chem. Comm.*, 1304 (1989).
- JJ: N. K. Moncur, P. D. Willson, T. H. Edwards, *J. Mol. Spectrosc.*, 52, 380 (1974).
- KK: S. Husebye, K. Maartmann-Moe, O. Mikalsen, *Acta Chem. Scand.*, 43, 868 (1989)

Chapter 7

Program

The logic within MOPAC is best understood by use of flow-diagrams.

There are two main sequences, geometric and electronic. These join only at one common subroutine COMPFG. It is possible, therefore, to understand the geometric or electronic sections in isolation, without having studied the other section.

7.1 Main geometric sequence

The main geometric sequence in MOPAC is given in Figure 7.1.

7.2 Main electronic flow

The main conventional electronic sequence in MOPAC is given in Figure 7.2.

The LMO method is similar to the conventional method. The main changes are:

- HQRII and DIAG are replaced by DIAGG (the annihilation operation).
- MECI, MECIP, etc, and INTERP are not available.
- Extra subroutines, used in creating the initial LMOs, are present.

7.3 Control within MOPAC

Almost all the control information is passed *via* the single datum “KEYWRD”, a string of 241 characters, which is read in at the start of the job.

Each subroutine is made independent, as far as possible, even at the expense of extra code or calculation. Thus, for example, the SCF criterion is set in subroutine ITER, and nowhere else. Similarly, subroutine DERIV has exclusive control of the step size in the finite-difference calculation of the energy derivatives. If the default values are to be reset, then the new value is supplied in KEYWRD, and extracted *via* INDEX and READA. The flow of control is decided by the presence of various keywords in KEYWRD.

When a subroutine is called, it assumes that all data required for its operation are available in either modules or arguments. Normally no check is made as to the validity of the data received. All data are “owned” by one, and only one, subroutine. Ownership means the implied permission and ability to change the data. Thus STATE “owns” the number of atomic orbitals, in that it calculates this number, and stores it in the variable NORBS. Many subroutines use NORBS, but none of them is allowed to change it. For obvious reasons no exceptions should be

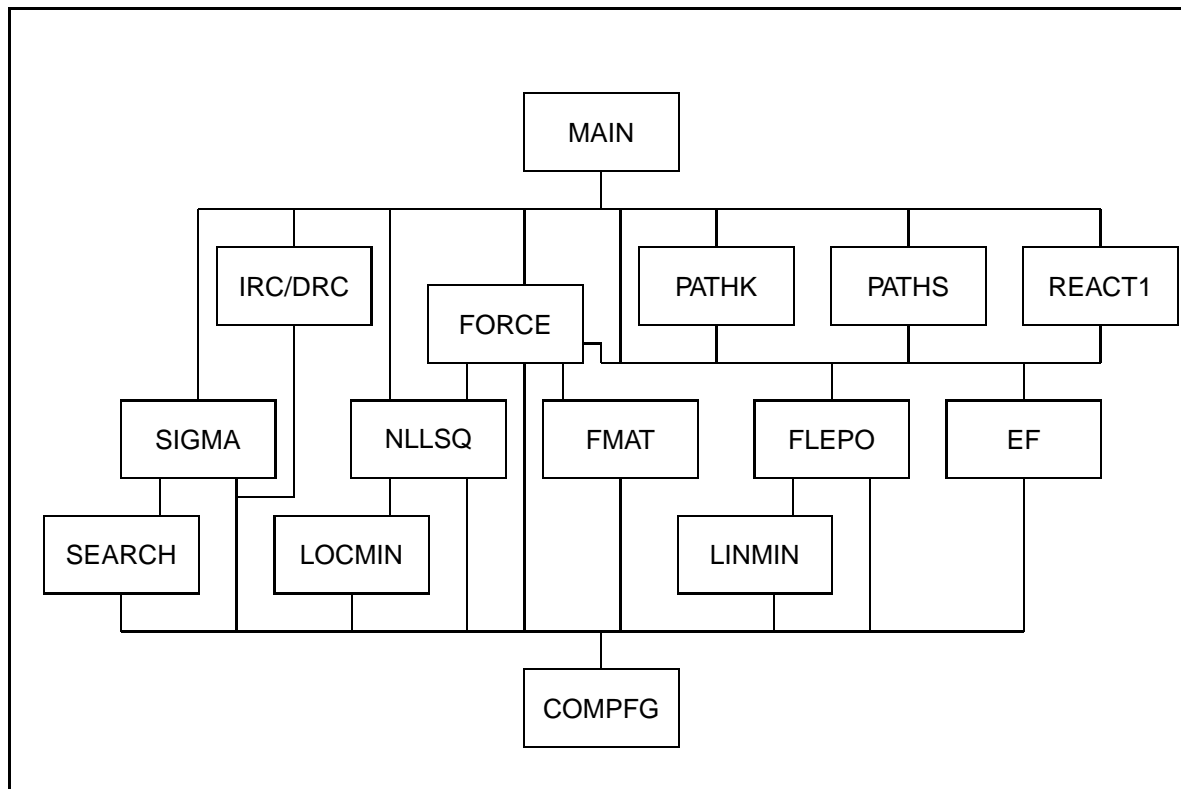


Figure 7.1: Diagram of Main Geometric Sequence in MOPAC

made to this rule. To illustrate the usefulness of this convention, consider the eigenvectors, C and $CBETA$. These are owned by ITER. Before ITER is called, C and $CBETA$ are not calculated, after ITER has been called C and $CBETA$ are known, so any subroutine which needs to use the eigenvectors can do so in the certain knowledge that they exist.

Any variables which are only used within a subroutine are not passed outside the subroutine unless an overriding reason exists. This is found in PULAY and CNVG, among others where arrays used to hold spin-dependent data are used, and these cannot conveniently be defined within the subroutines. In these examples, the relevant arrays are “owned” by ITER.

A general subroutine, of which ITER is a good example, handles three kinds of data: First, data which the subroutine is going to work on, for example the one and two electron matrices; second, data necessary to manipulate the first set of data, such as the number of atomic orbitals; third, the calculated quantities, here the electronic energy, and the density and Fock matrices.

Reference data are entered into a subroutine by way of modules. This is to emphasize their peripheral role. Thus the number of orbitals, while essential to ITER, is not central to the task it has to perform, and is passed through a module.

Data the subroutine is going to work on are passed via the argument list. Thus the one and two electron matrices, which are the main reason for ITER’s existence, are entered as two of the four arguments. As ITER does not own these matrices it can use them but may not change their contents. The other argument is EE , the electronic energy. EE is owned by ITER even though it first appears before ITER is called.

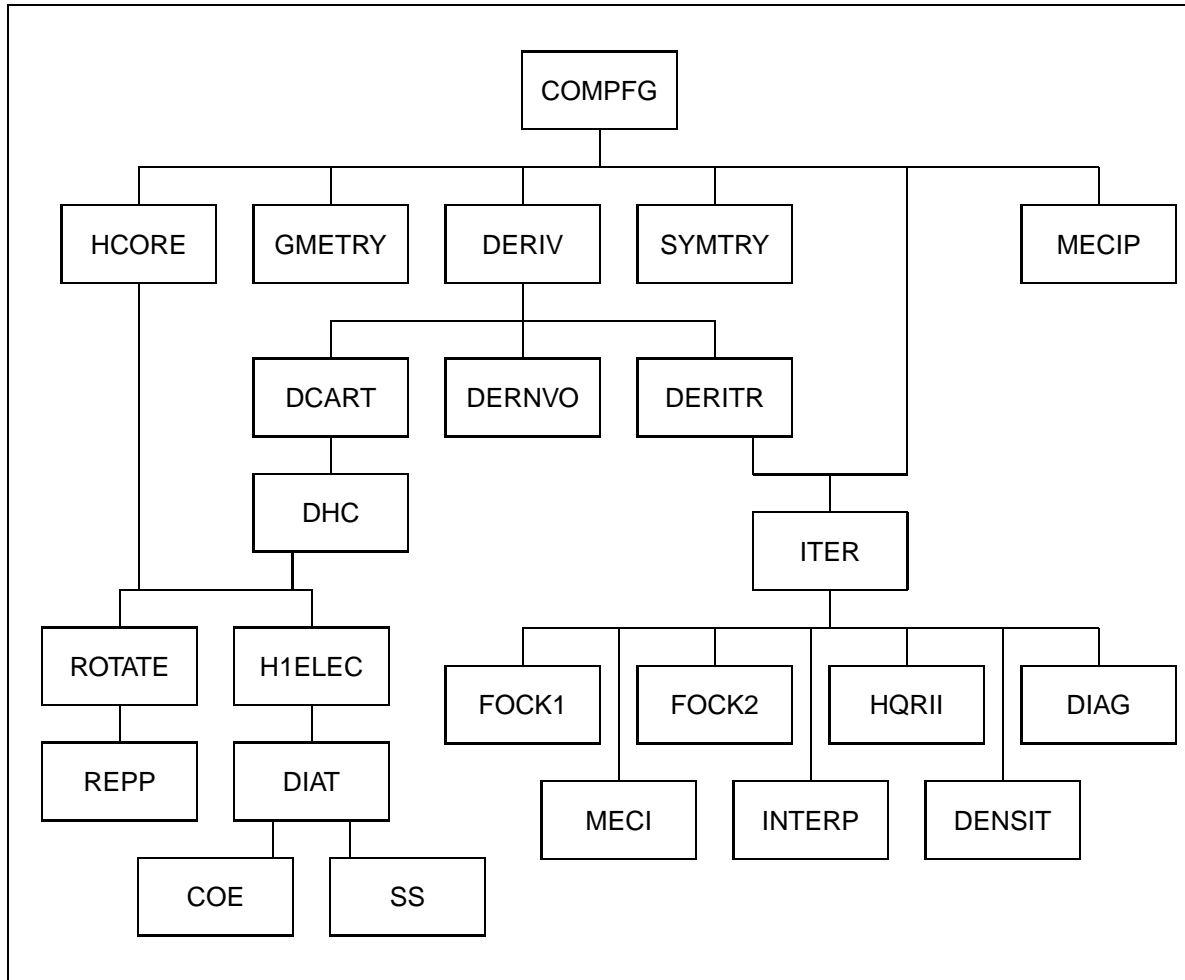


Figure 7.2: Diagram of Main Conventional Electronic Sequence in MOPAC

7.4 Arrays.

7.4.1 Array Specification.

In MOPAC, most arrays are created dynamically. The main exceptions are a few module arrays: `GEO`, `COORD`, `LABELS`, `NAT`, `NA`, `NB`, and `NC`. These arrays are only used at the start of the run, and once the job is fully running, they are abandoned. These arrays have the same names as dynamic arrays, but, because they are used only at the very start of the run, there is no ambiguity. The size of these arrays is defined by `NUMATM` in the included file `sizes.F90`. This quantity, `NUMATM` is equivalent to the MOPAC 93 parameters `MAXHEV` and `MAXLIT`. By default, MOPAC can run systems of up to 20,000 atoms. Although this is large, the small number of static arrays means that the memory required for a small system is still small—less than 8 megabytes.

Most of the memory is assigned dynamically. Once the data are read in—using the static arrays—the dynamic memory arrays are created as needed.

7.5 Names and Storage of Variables in MOPAC

Most subroutines in MOPAC contain the module `common_system`. This module contains variables that are central to the transmission of information throughout the program.

`IKST(100)`

Because MOPAC can calculate many quantities, such as the number of atoms, heat of formation, etc., a method has to exist to allow these quantities to be transmitted throughout the program. This is accomplished by two arrays, `IKST` and `RKST`. Each element of these arrays is reserved for use as a storage site for a scalar. `IKST` holds integer scalars and `RKST` holds real scalars. Table 7.1 summarizes these quantities. At the start of a calculation, every element of `IKST` is set to -99999, and every element of `RKST` is set to -99999.99. As a result of this, if an element of these arrays is examined, it is immediately obvious whether or not it has been set. Thus, until the number of atoms is known, `IKST(2) (NUMAT)`, is -99999.

1. `NATOMS` The number of atoms supplied by the data set, including dummy atoms and translational vectors. `NATOMS` is set in the subroutine that reads in the geometry (`GETGEO`, `GETGEG`, or `GETPDB`).

With one exception, once `NATOMS` is determined, it is not changed. In `FORCE`, only real atoms are used, therefore, for convenience, the geometry is re-defined so that any dummy atoms are deleted.

2. `NUMAT` The number of real atoms. Here, “real” atoms are the sparkles (100% ionized atoms) and normal atoms. The difference between `NATOMS` and `NUMAT` is the number of dummy atoms plus any translational vectors.
3. `NELECS` The total number of valence electrons in the system. This is calculated by adding up the number of valence electrons on each atom, and subtracting the net charge on the system.
4. `NALPHA` In a UHF calculation, `NALPHA` is the number of α electrons in the system. By default, the number of α electrons is the same as the number of β electrons (for closed-shell systems), or one more than the number of β electrons (for odd-electron systems). If a spin is defined (`TRIPLET`, `QUARTET`, etc.), then the number of α electrons is two to four more than the number of β electrons. For example, if `QUINTET` is specified, then `NALPHA` would be four more than `NBETA`. The value of `NALPHA` is set in `STATE`.

In an RHF calculation, `NALPHA=0`.

Table 7.1: Constants used in MOPAC

RKST	IKST	No.	RKST	IKST	No.	RKST	IKST	No.
ESCF	NATOMS	1	EOUTER	ICOMPF	34			67
ENUCLR	NUMAT	2	THRESH	ISAFE	35			68
EE	NELECS	3	RSOLV	LIMSCF	36			69
ATHEAT	NALPHA	4		NA1	37			70
FRACT	NBETA	5	CORHYB	IGROUP	38			71
EMIN	NCLOSE	6	ESCFL	NCLASS	39			72
COSINE	NOPEN	7	VOLUME	NIRRED	40			73
GNORM	NORBS	8	FEPSI	NSPA	41			74
TIMEO	ISCF	9	RDS	NPS	42			75
EZQ	NSCF	10	DISEX2	NPS2	43			76
EEQ	ID	11	CUTOF1	IOLDCV	44			77
AREA	NVAR	12	CUTOF2	NSTATE	45			78
EDIEL	MOLNUM	13	CUTOFS	NUMPTS	46			79
GVW	LATOM	14	CUTOFP		47			80
GVWS	LPARAM	15	FNSQ	LENABC	48			81
ELC1	ITRY	16	CLOWER	MESP	49			82
TLEFT	ITYPE	17	CUPPER	NUMRED	50			83
TDUMP	MFLAG	18	DLM	NORRED	51			84
ELECT	ITERQ	19	XNORM	NELRED	52			85
STEP	IFLEPO	20	DIFE	NRES	53			86
TVEC(1,1)	MPACK	21	PRESSURE	NUMCAL	54			87
TVEC(2,1)	N2ELEC	22		NIP	55			88
TVEC(3,1)	NMOS	23		IDIAGG	56			89
TVEC(1,2)	LAB	24		MDRCRS	57			90
TVEC(2,2)	NELEC	25			58			91
TVEC(3,2)		26			59			92
TVEC(1,3)		27		IPAD2	60			93
TVEC(2,3)		28		IPAD4	61			94
TVEC(3,3)	MAXTXT	29		NL1	62			95
RJKAB1	LAST	30		NL2	63		ICROS	96
WTMOL	NDEP	31		NL3	64		MORB	97
VERSON	LABSIZ	32			65		MOP/MOZ	98
TOTIME	LM61	33			66			99
							DERIV	100

5. NBETA In a UHF calculation, NBETA is the number of β electrons in the system. In an RHF calculation, NBETA=0. The value of NBETA is set in STATE.
6. NCLOSE In an RHF calculation, NCLOSE is the number of doubly-occupied M.O.s. In a UHF calculation, NCLOSE=0. The value of NCLOSE is set in STATE.
7. NOPEN In an RHF calculation, NOPEN is the number of the highest M.O. that would be fractionally occupied in the SCF calculation. The "fraction" can range from 0.0 to 2.0. For an isolated carbon atom, the fractional occupancy (two p electrons in three atomic orbitals) would be 0.6667, and NCLOSE would be 1, and NOPEN would be 4. By default, the fractional occupancy of the SOMO in odd-electron systems is 1.0. In a UHF calculation, NOPEN=0. The value of NCLOSE is set in STATE.
8. NORBS The number of atomic orbitals in a system. Each hydrogen atom contributes one atomic orbital, every other atom contributes four, except atoms with d -orbitals, which contribute nine, and sparkles, which do not contribute any. NORBS is set in STATE.
9. ISCF At the end of a calculation, ISCF=1 if an SCF exists, otherwise ISCF=2. Set in ISITSC.
10. NSCF The number of SCF calculations done. Each time an SCF calculation is done, NSCF is incremented by one.
11. ID The number of dimensions for polymeric systems. For molecules, ID=0; for polymers, ID=1; for layer systems, ID=2; and for solids, ID=3. Set in GMETRY.
12. NVAR The number of geometric variables. Normally, this is the number of coordinates flagged by a "1" in the input data set. In a FORCE calculation, or one that implies a FORCE calculation, such as an IRC, then NVAR=3*NUMAT.
13. MOLNUM The number of the system. Each new molecule run is given a new number, with the first molecule having MOLNUM=1. If several geometric operations are done on one system, they all share the same MOLNUM, however, if OLDGEO is used, then the system is considered to be "new". The only exception is when an IRC is run, in which case the value of MOLNUM is increased between the FORCE and the IRC calculations. This is needed, because the system is re-oriented before the IRC is run.
14. LATOM In a reaction coordinate, LATOM is the atom number of the atom that moves. This coordinate is indicated in the input data set by the "-1" flag. The atom can be real or dummy. Set in READMO. See also LPARAM.
15. LPARAM In a reaction coordinate, LPARAM is the coordinate (first, second, or third) of the atom that moves. Set in READMO. See also LATOM.
16. ITRY The number of iterations allowed in attempting to generate an SCF. By default, ITRY=200, but this can be changed by use of ITRY= n .
17. ITYPE The type of calculation to be run. If an MNDO calculation, ITYPE=4; AM1, ITYPE=2; PM3, ITYPE=3, MINDO/3, ITYPE=4; MNDO- d , ITYPE=5.
18. MFLAG Used by the Tomasi method.
19. ITERQ Used by the Tomasi method
20. IFLEPO A flag to indicate how the geometry optimization terminated. Initialized in RMOPAC and normally set in EF, FLEPO, NLLSQ, and POWSQ. Set in ITER when the SCF fails. The various meanings of IFLEPO are defined in DATA statements in WRITMN.

21. MPACK The number of array elements in the density, one-electron, and Fock matrices. In conventional work, this is $(\text{NORBS} * (\text{NORBS} + 1)) / 2$ and is set in STATE, in MOZYME work, MPACK is determined from the number of atoms that interact with each other. For large systems, MPACK is normally much smaller than the value in conventional work.
22. N2ELEC The number of two-electron integrals. The number of integrals (0, 1, 16, or 81) contributed by an atom depends on the number of orbitals (0, 1, 4, or 9) on that atom. The NDDO approximation indicates that for atom pairs, the number of integrals is $(n(n + 1)) * (m(m + 1)) / 4$, where n and m are the number of orbitals on each atom. In MOZYME, the number is reduced when two atoms are far apart. See p. 163 for details.
23. NMOS The number of M.O.s involved in the active space in a C.I. calculation. Set in STATE only.
24. LAB The number of microstates or configurations included in the C.I. calculation. Set in MECI only.
25. NELEC The number of electrons involved in the C.I. calculation. Set in MECI only.
29. MAXTXT Maximum number of characters in the description of an atom. (An atom label can include a short description of up to 13 letters, for example "N(634 ASN 46)")
30. LAST LAST=0 if the current SCF is *not* the last SCF. Only on the last SCF, when LAST=1, are the symmetries of the M.O.s determined. If the system is RHF open shell, then in WRITMN, LAST is set to 3, and MECI is called. This prints the symmetries of the states.
31. NDEP The number of symmetry-dependent coordinates. Set in GETGEG, GETSYM, and MAKSYM.
32. LABSIZ The number of states to be calculated in a C.I. (usually less than LAB, until the last C.I., when all states are calculated.)
33. LM61 The number of matrix elements in the one-center part of the density, one-electron, or Fock matrices. Each hydrogen contributes 1 element, each $s - p$ atom contributes 10 $((4 * (4 + 1)) / 2)$, and each $s - p - d$ atom contributes 45 elements $((9 * (9 + 1)) / 2)$.
34. ICOMPF Used by the Tomasi method.
35. ISAFE A flag to indicate whether some large arrays can be ignored. If ISAFE=1 then these arrays are used. If ISAFE=0, then the amount of memory used is reduced, but some functionalities, such as the Camp-King converger, will not work. Normally, this is not important. ISAFE is set at compile time, and can be re-set at run time by SAFE or UNSAFE.
36. LIMSCF In the SCF, if LIMSCF=1, then the SCF can be stopped before the usual criteria are satisfied, if the energy is obviously lower or obviously higher. This saves time, in that the geometry is still not near to the stationary point. LIMSCF=0 if the usual SCF criteria must be met. This is necessary in FORCE and transition state calculations.
37. NA1 A flag to indicate the nature of the coordinates. If NA1=1 then all coordinates are Cartesian. Default: NA1=0.
38. IGROUP The number of the point-group of the system.
39. NCLASS The number of classes necessary to define the point-group. For I_h , for example, four operations (four classes) are necessary: E , C_5 , C_3 , and the inversion operation.
40. NIRRED The number of irreducible representations in the point-group.

41. NSPA Used by the COSMO method.
42. NPS Used by the COSMO method.
43. NPS2 Used by the COSMO method.
45. NSTATE The number of states used in describing the electronic state of a system. Normally 1, the number can increase when degenerate states are involved. Thus in CH_4^+ , without Jahn-Teller distortion, the value of NSTATE would be 3 (representing the 2T_u state). Note: For degenerate open shell systems, OPEN(n,m) will normally be needed.
46. NUMPTS Used by the COSMO method.
48. LENABC Estimated maximum number of surface points in the COSMO method.
49. MESP Size of the ESP array in the electrostatic potential calculation, set in "sizes.h".
50. NUMRED The number of atoms involved in the SCF when the RAPID option is used. The number of atoms that move plus the number of atoms attached to them.
51. NORRED The number of LMOs involved in the SCF when the RAPID option is used.
52. NELRED The number of electrons involved in the SCF when the RAPID option is used.
53. NRES The number of residues identified in a protein when RESIDUES or RESEQ is specified.
54. NUMCAL For a given calculation (input geometry), the number of the type of calculation. For example, an initial SCF (NUMCAL=1), followed by a FORCE calculation (NUMCAL=2).
56. IDIAGG Used in the annihilation operation in MOZYME, IDIAGG increments every time DIAGG is called.
57. MDRCRS Controls whether the restart file in a DRC or IRC run should be formatted or unformatted. Set in MOPAC from sizes.h. If MDRCRS=1, then a formatted restart file is used. This can be read easily, but uses more space and is less precise than unformatted (MDRCRS=0).
60. IPAD2 The estimated maximum average number of atoms in a LMO. The default value is usually sufficient, but it can be reset by NLMO=nnn.
61. IPAD4 The estimated maximum average number of atomic orbitals in a LMO. Set as a function of IPAD2.
96. ICROS The type of calculation: 1 if an intersystem crossing, 0 otherwise.
97. MORB The maximum number of orbitals on any atom: 1, 4, or 9.
98. unnamed The type of method used in solving the SCF equations. 0 indicates conventional (MOPAC type), 1 indicates LMO (MOZYME type).
100. unnamed The type of derivative calculation: 1 = MOZYME -1 = remove contributions to derivative arising from moving atoms; 2 = MOZYME 0 = calculate MOZYME derivatives *de novo*; 3 = MOZYME +1 = add in contributions to derivatives arising from moving atoms; 4 = MOPAC RHF; 5 = MOPAC UHF.

RKST(100)

1. ESCF The calculated heat of formation in kcal.mol⁻¹.
2. ENUCLR The nuclear energy in eV.
4. ATHEAT The reference heat of atomization plus ionization, in eV.
5. FRACT The fractional orbital occupancy, in electrons. Range: 0.0 to 2.0.
6. EMIN Within any geometry optimization calculation, the lowest heat of formation calculated. Used in deciding when to exit the SCF.
7. COSINE The angle between the current and previous gradient vector. Range: -1.0 to +1.0.
8. GNORM The scalar of the gradient vector. GNORM=MOD(GRAD).
9. TIME0 The internal zero of time for determining CPU usage.
10. EZQ The nuclear - solvent interaction term in the Tomasi model.
11. EEQ The electronic - solvent interaction term in the Tomasi model.
12. AREA The surface area of a molecule, in Å², calculated by COSMO.
14. GVW Used by the Tomasi method.
15. GVWS Used by the Tomasi method.
16. ELC1 Used by the Tomasi method.
17. TLEFT The allowed amount of CPU time left for the calculation. Default: one hour.
18. TDUMP The interval between checkpoints (dumps of the calculation that can be used by RESTART). Default: one hour.
19. ELECT The electronic energy in eV.
20. STEP The step size in a reaction path.
21. TVEC A three by three matrix representing the translation vector. The first vector is TVEC(1,1):TVEC(3,1).
30. RJKAB1 The two-electron integral over M.O.s used in calculating the correction to the energy in the half-electron approximation.
31. WTMOL The molecular weight in AMU. Calculated in MOPAC.
32. VERNON The version number of MOPAC, e.g., 2000.0.
33. TOTIME The total time, in seconds, for the calculation, including time used in previous runs, if RESTART is used.
34. EOUTER Nuclear energy in eV due to atom pairs not present in P, H, or F. Calculated in HCORZ, and used by WRITMN only.
35. THRESH In MOZYME, the threshold value for intensity of a LMO on an atom. An atom with intensity below THRESH would either (a) not be added to a LMO (in DIAGG2), or (b) would be deleted from the LMO (in TIDY).

38. CORHYB The correction to the MOZYME energy due to the use of the dipole approximation. (for interactions arising from atoms separated by more than CUTOF2 but less than CUTOF1, that is, atoms pairs not present in P, H, or F.
39. ESCFL In an intersystem crossing, ESCFL is the ΔH_f of the lower state. Note $ESCFL < ESCF$.
40. VOLUME The volume of the system in \AA^3 , calculated by COSMO.
41. FEPSI Used by COSMO
42. RDS Used by COSMO.
43. DISEX2 Used by COSMO. DISEX2 is needed in both INITSN and CONSTS.
44. CUTOF1 The cutoff distance for dipolar electrostatic terms. Beyond CUTOF1, only point-charge terms are used. Default: 30 Ångstroms.
45. CUTOF2 The cutoff distance for NDDO two-electron, two-center terms. Beyond CUTOF2, only dipole and point charge terms are used.
46. CUTOFS The distance beyond which overlap integrals are *not* calculated. By default, CUTOFS=7Å.
47. CUTOFP In polymers, all electrostatic terms involving atoms separated by more than CUTOFP are calculated as if they were separated by
48. CUTOFP Needed in order to ensure that equivalent atoms are correctly predicted.
49. FNSQ
50. CLOWER In a polymer calculation, CLOWER marks the onset, in Ångstroms, of the truncation function for electrostatic interactions. Set in function TRUNK
51. CUPPER In a polymer calculation, CUPPER marks the upper bound, in Ångstroms, of the truncation function for electrostatic interactions. Set in function TRUNK
52. DLM Used by intersystem crossing.
53. XNORM Used by intersystem crossing.
54. DIFE In an intersystem crossing, DIFE is the square of the difference in energy of the two states involved.
55. PRESSURE In a solid-state calculation, PRESSURE is the stress (in kcal/mol) that the system is under.

Files used in MOPAC

All input-output channels in MOPAC are assigned numbers. To allow the values of these numbers to be easily changed, the module `common_system` contains symbolics for the various channels used. These symbolics are set in GETDAT and can only be changed by editing this subroutine.

Table 7.2: Channels Used in MOPAC

Input/output Channel No.	Symbol	Name	Type	Formatted?
2		<filename>.dat	Input	Yes
3	ISCR	Scratch	Internal working	No
4	ISSETUP	SETUP	Setup	Yes
25	IR	Scratch	Input (during run)	Yes
26	IW	<filename>.out	Output	Yes
9	IRES	<filename>.res	Restart	No
10	IDEN	<filename>.den	Density	No
11	ILOG	<filename>.log	Log	Yes
12	IARC	<filename>.arc	Archive	Yes
13	IGPT	<filename>.gpt	Graphics	No
14	IEXT	User defined	New parameters	Yes
15	IESR	<filename>.esr	ESP Restart	No
16	ISYB	<filename>.syb	SYBYL	Yes
17	ISOL	<filename>.sol	SOL map in MEP	No
18	IBRZ	<filename>.brz	Brillouin Zone	No
17	IPDB	<filename>.pdb	PDB file	No
20	IESR	<filename>.ump	Grid map	Yes
21	IESP	<filename>.esp	Electrostatic map	Yes
22	IS	<filename>.mep	Tomasi MEP map	Yes

Chapter 8

Error messages produced by MOPAC

MOPAC produces several hundred messages, all of which are intended to be self-explanatory. However, when an error occurs it is useful to have more information than is given in the standard messages.

The following alphabetical list gives more complete definitions of the messages printed.

1SCF SPECIFIED WITH PATH. ... (FATAL)

The pair of options, 1SCF with a path calculation, is not allowed, except in a RESTART calculation.

A SINGLE ATOM HAS NO VIBRATIONAL MODES (FATAL)

An attempt has been made to calculate the vibrations of a single atom. The smallest system that can have vibrations is a diatomic molecule.

A SYMMETRY FUNCTION IS USED TO DEFINE A NON-EXISTENT ATOM (FATAL)

Symmetry functions can only be used in the definition of atoms or dummy atoms. Check the dependent atom numbers in the symmetry data.

ALL CONVERGERS ARE NOW FORCED ON

The default SCF convergers have not worked. A new set of convergers will be tried. This is often caused by faulty data, so the data should be checked to see if anything is wrong. This sometimes happens naturally, particularly with exotic systems.

AN ATTEMPT WAS MADE TO PRINT ...

To avoid printing very large matrices in MOZYME, an upper limit (200) has been set on the size of matrix that can be printed. If the array requested is larger than this limit, only part of the array will be printed.

AN UNOPTIMIZABLE GEOMETRIC PARAMETER ...

When internal coordinates are supplied, six coordinates cannot be optimized. These are the three coordinates of atom 1, the angle and dihedral on atom 2 and the dihedral on atom 3. An attempt has been made to optimize one of these. This is usually indicative of a typographic error, but might simply be an oversight. Either way, the error will be corrected and the calculation will not be stopped here.

ANALYTIC C.I. DERIVATIVES DO NOT WORK ... (FATAL)

The analytical C.I. derivative calculation failed. Add NOANCI or UHF and re-run.

ANALYTICAL DERIVATIVES TOO INACCURATE ... (FATAL)

The analytical C.I. derivative calculation failed. Add NOANCI or UHF and re-run.

AT LEAST ONE ATOM DEFINED BY CVB ... (FATAL)

The atoms defined in CVD are limited to the real atoms. If any dummy atoms are used, these should *not* be included in the atom numbering used.

AT LEAST ONE ATOM HAS A ZERO MASS (FATAL)

In a FORCE calculation, the mass of an atom is zero. To correct this, , see p. 89.

At this point, both reactants ...

In a SADDLE calculation, both reactants and products are on the same side of the transition state. Options at this point are:

- If it is near to the transition state (the gradient norm has been dropping for reactants or products, or the "DISTANCE A-B" is small, e.g. less than 0.2), refine the geometry using TS.
- If it is not near to the transition state, identify geometries on both sides of the transition state (these will be separated in the output by the message "REACTANTS AND PRODUCTS SWAPPED AROUND"), and use these to start a new SADDLE calculation. Add BAR=0.02.
- If CPU time is not important, add BAR=0.02 to the original data set and re-run.

Atom *nn* is Cartesian ... (FATAL)

An attempt has been made to use symmetry function 19 (the bond length is a multiple of the reference bond length) using an atom whose position is defined using Cartesian coordinates. Correct the error and re-run.

Atom *nn* is internal ... (FATAL)

An attempt has been made to use symmetry function 18 (the "z" coordinate is set equal to minus the reference "y" coordinate) using an atom whose position is defined using internal coordinates. Correct the error and re-run.

ATOM NUMBER *nn* IS ILLDEFINED (FATAL)

The rules for definition of atom connectivity are:

1. Atom 1 has no connectivity.
2. Atom 2 can be Cartesian or internal. If internal, it must be defined with connectivity (1,0,0).
3. Atom 3 can be Cartesian or internal. If internal, it must be defined with connectivity (2,1,0) or (1,2,0).
4. All other atoms can be Cartesian or internal. If internal, they must be defined in terms of already-defined atoms: these atoms must all be different. Thus atom 9 might be connected to atom 5, make an angle with atom 6, and have a dihedral with atom 7. If the dihedral was with atom 5, then the geometry definition would be faulty.

If any of these rules is broken, a fatal error message is printed, and the calculation stopped.

ATOMIC MASS OF ATOM *nn* TOO SMALL (FATAL)

In a DRC calculation, all atoms must have masses greater than 0.1 times the mass of a hydrogen atom. If sparkles are used, these have no default mass. Atom masses are specified as numbers after the symbol, e.g. H1.008 or +12.00 (for a "+" sparkle having a mass of 12.00.)

ATOMIC NUMBER *nn* IS NOT AVAILABLE ... (FATAL)

An element has been used for which parameters are not available. Only if a typographic error has been made can this be rectified. This check is not exhaustive, in that even if the elements are acceptable there are some combinations of elements within MINDO/3 that are not allowed. This is a fatal error message.

ATOMIC NUMBER OF nn ? (FATAL)

An atom has been specified with a negative or zero atomic number. This is normally caused by forgetting to specify an atomic number or symbol. This is a fatal error message.

ATOMS nn AND nn ARE SEPARATED BY $nn.nnnn$ ANGSTROMS (FATAL)

Two genuine atoms (not dummies) are separated by a very small distance. This can occur when a complicated geometry is being optimized, in which case the user may wish to continue. This can be done by using the keyword GEO-OK. More often, however, this message indicates a mistake, and the calculation is, by default, stopped.

ATTEMPT TO GO DOWNHILL IS UNSUCCESSFUL ...

A quite rare message, produced by Bartel's gradient norm minimization. Bartel's method attempts to minimize the gradient norm by searching the gradient space for a minimum. Apparently a minimum has been found, but not recognized as such. The program has searched in all $3N - 6$ directions, and found no way down, but the criteria for a minimum have not been satisfied. No advice is available for getting round this error.

AVAILABLE MEMORY IN GREENS FUNCTION TOO SMALL (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

BOTH GEOMETRIES ARE IDENTICAL (FATAL)

The SADDLE technique uses two geometries, one for the reactant(s) and one for the product(s). These must be different. Correct fault and re-run.

BOTH SYSTEMS ARE ON THE SAME SIDE ...

A non-fatal message, but still cause for concern. During a SADDLE calculation the two geometries involved are on opposite sides of the transition state. This situation is verified at every point by calculating the cosine of the angle between the two gradient vectors. For as long as it is negative, then the two geometries are on opposite sides of the TS. If, however, the cosine becomes positive, then the assumption is made that one moiety has fallen over the TS and is now below the other geometry. That is, it is now further from the TS than the other, temporarily fixed, geometry. To correct this, identify geometries corresponding to points on each side of the TS. (Two geometries on the output separated by the message 'SWAPPING ...') and make up a new data-file using these geometries. This corresponds to points on the reaction path near to the TS. Run a new job using these two geometries, but with BAR set to a third or a quarter of its original value, e.g. BAR=0.05. This normally allows the TS to be located.

C.I. IS OF SIZE LESS THAN ROOT SPECIFIED (FATAL)

The value of n in ROOT= n is less than the size of the C.I.

See p. 143 for the sizes of various C.I. matrices, implied by C.I.= m .

C.I. NOT ALLOWED WITH UHF (FATAL)

There is no UHF configuration interaction calculation in MOPAC. Either remove the keyword that implies C.I. or the word UHF.

CALCULATION ABANDONED AT THIS POINT (FATAL)

A particularly annoying message! In order to define an atom's position, the three atoms used in the connectivity table must not accidentally fall into a straight line. This can happen during a geometry optimization or gradient minimization. If they do, and if the angle made by the atom being defined is not zero or 180 degrees, then its position becomes ill-defined. This is not desirable, and the calculation will stop in order to

allow corrective action to be taken. Note that if the three atoms are in an exactly straight line, this message will not be triggered.

Cannot open *filename.out*! **(FATAL)**

The program cannot open the output file. Possible causes are (a) the file already exists, but is owned by another user, (b) the subdirectory is "read only", (c) there is no space left on the partition (unlikely).

Cannot write density matrix to *filename.den* **(FATAL)**

The program cannot open the density restart file. Possible causes are (a) the file already exists, but is owned by another user, (b) the subdirectory is "read only", (c) there is no space left on the partition (unlikely).

CARTESIAN CALCULATION NOT ALLOWED WITH ... **(FATAL)**

XYZ is not allowed when geometries are specified using GAUSSIAN format. To allow XYZ to be used, first do a OSCF calculation to convert GAUSSIAN format geometry into the MOPAC format.

CHARGE ON ATOM *N* UNREASONABLE **(FATAL)**

The range of allowed charges for an atom is limited. Allowed charges are: Group I: +1; II: +1, +2; III: +1, +2, +3; IV: -1, +1; V: +1; VI: -1, -2; VII: -1. Any charges outside these ranges are considered unreasonable.

Correct the geometry and re-run.

Connectivity of atom *NN* ...

The rules for definition of atom connectivity are:

1. Atom 1 has no connectivity.
2. Atom 2 can be Cartesian or internal. If internal, it must be defined with connectivity (1,0,0).
3. Atom 3 can be Cartesian or internal. If internal, it must be defined with connectivity (2,1,0) or (1,2,0).
4. All other atoms can be Cartesian or internal. If internal, they must be defined in terms of already-defined atoms: these atoms must all be different. Thus atom 9 might be connected to atom 5, make an angle with atom 6, and have a dihedral with atom 7. If the dihedral was with atom 5, then the geometry definition would be faulty.

If any of these rules is broken, a fatal error message is printed, and the calculation stopped.

COORDINATES MUST BE CARTESIAN **(FATAL)**

In a DRC calculation, in which an initial velocity is to be used, the geometry must be supplied in Cartesian coordinates, in order for the velocity to be meaningful. If it is essential that internal coordinates be used for the geometry, add LET, and re-run. The velocity vector, however, must still be in Cartesian coordinates.

CROSS requires ROOT=2 or higher **(FATAL)**

In an intersystem crossing calculation, the crossing is defined in terms of the higher energy state. Because of this, ROOT=*n* *must* be used, and *n* *must* be more than 1.

CUTOF1 WAS SET SMALLER THAN CUTOF2

In a MOZYME calculation in which CUTOFs were used, CUTOF1 (the transition from dipole plus monopole to monopole only) was set smaller than the transition from NDDO to dipole plus monopole. CUTOF1 would be automatically increased to a reasonable value and the calculation allowed to continue.

DATA ARE NOT AVAILABLE FOR ELEMENT NO. *N* **(FATAL)**

Parameters are not available for the element with atomic number *N*. If new parameters are available, these can be supplied to MOPAC by use of , see p. 43.

Data for TOMASI Model missing or faulty (FATAL)

When TOM is specified, extra data are required after the geometry in order to define the solvent, unless H₂O, CCL₄, or CHCL₃ is used.

DEGENERATE LEVELS DETECTED IN MECI ... (FATAL)

If only some M.O.s of a degenerate manifold are used in a MECI calculation, the results will be nonsense. To prevent such calculations, the message DEGENERATE LEVELS ... is printed, and the job stopped. To continue, specify GEO-OK. See Space Quantization and GEO-OK.

DENSITY FILE MISSING OR CORRUPT (FATAL)

In a run involving OLDENS, the old density matrix, in <filename>.den, is either missing or corrupt. Either generate a new <filename>.den file using 1SCF and DENOUT or do not use OLDENS.

Density Restart File missing or corrupt (FATAL)

In a run involving OLDENS, the old density matrix, in <filename>.den, is either missing or corrupt. Either generate a new <filename>.den file using 1SCF and DENOUT or do not use OLDENS.

DIPOLE CONSTRAINTS NOT USED

An attempt had been made to run an ESP calculation, with dipole constraints, on an ionized system. This is not allowed, the keyword DIPOLE will be ignored and the calculation allowed to proceed.

DUE TO A PROGRAM BUG, THE FIRST THREE ... (FATAL)

Due to a problem caused by the definition of internal coordinates, the first three atoms must not form a straight line.

ECHO is not allowed at this point (FATAL)

ECHO can only be used at the start of a run. If the run has several geometries, and at least one is in GAUSSIAN format, then ECHO would cause an infinite loop to be created. Remove ECHO and re-run.

EIGENVECTOR FOLLOWING IS NOT RECOMMENDED ... (FATAL)

If internal coordinates are used, the maximum number of variables is $3N-6$. If Cartesian coordinates are used, up to $3N$ variables can be used. If GEO-OK is present, any number of variables can be used.

EITHER ADD 'LET' OR ... (FATAL)

A FORCE calculation is *only* meaningful if the geometry is at a stationary point. Either add LET, to run the current geometry, or refine the geometry and re-run.

ELEMENT NOT FOUND (FATAL)

When an external file is used to redefine MNDO, AM1, PM3, or MNDO-*d* parameters, the chemical symbols used must correspond to known elements. Any that do not will trigger this fatal message.

ELEMENT NOT RECOGNIZED (FATAL)

When a Gaussian data set is supplied, the chemical symbols used must correspond to known elements. Any that do not will trigger this fatal message. Correct the data set and re-run.

ERROR TOO MANY NEIGHBORS (FATAL)

An extraordinarily difficult error to make. An atom has more than 200 neighboring atoms, in an ESP or PMEP calculation.

ERROR DETECTED DURING READ ... (FATAL)

In a PATH calculation involving a RESTART, the RESTART file is faulty—damaged or corrupt. To correct this, start over again.

ERROR DETECTED IN SUBROUTINE CHECK ... (FATAL)

In a MOZYME calculation, the LMOs are periodically re-normalized. This involves making small adjustments to the LMOs. If large changes are necessary, something has gone wrong in the LMO procedure. Try restarting the job, but do *not* use OLDENS. If this does not work, there is a problem in the program.

ERROR DURING READ AT ATOM NUMBER ... (FATAL)

Something is wrong with the geometry data. In order to help find the error, the geometry already read in is printed. The error lies either on the last line of the geometry printed, or on the next (unprinted) line. This is a fatal error.

Error in BLAS (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

ERROR IN CVB KEYWORD (FATAL)

An error is present in the CVB keyword used in defining specific bonds in a Lewis structure. See p. 38 for details of this keyword.

Error in EF (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in GETGEO (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in GETMEM (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in GREENF (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in MOLDAT. (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in PATHS (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in PMP (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

ERROR IN READ OF FIRST THREE LINES (FATAL)

The data-set has a severe error in the first three lines. This is a very unusual error, and indicates that the data-set is likely to be severely in error, or that MOPAC has not been installed correctly.

Error in READMO (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in TOM (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in UPDATE (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

Error in USAGE (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

ERROR!! MODE IS LARGER ... (FATAL)

In a system of N variables, a request has been made to follow the M th mode, where $M > N$. Correct the data and re-run.

ERROR: PERMANENT ARRAY (FATAL)

This error can *only* be caused by a programming error. Please contact Dr Stewart as soon as possible.

ERRORS DETECTED IN CONNECTIVITY (FATAL)

The connectivity in the MOPAC internal coordinate Z-matrix is faulty. For a description of the connectivity rules, see p. 89.

EXCITED USED WITH TRIPLET (FATAL)

EXCITED implies the first singlet excited state. This cannot be used if TRIPLET is requested. Correct the data set and re-run.

EXTERNAL PARAMETERS FILE MISSING OR EMPTY (FATAL)

EXTERNAL=*text* has been specified, but the file *text* is either missing or empty.

FAILED IN SEARCH, SEARCH CONTINUING

Not a fatal error. The McIver-Komornicki gradient minimization involves use of a line-search to find the lowest gradient. This message is merely advice. However, if SIGMA takes a long time, consider doing something else, such as using TS or NLLSQ, or refining the geometry a bit before resubmitting it to SIGMA.

FAILED TO ACHIEVE SCF.

The SCF calculation failed to go to completion; an unwanted and depressing message that unfortunately appears every so often.

To date three unconditional convergers have appeared in the literature: the SHIFT technique, Pulay's method, and the Camp-King converger. It would not be fair to the authors to condemn their methods. In MOPAC all sorts of weird and wonderful systems are calculated, systems the authors of the convergers never dreamed of. MOPAC uses a combination of all three convergers at times. Normally only a quadratic damper is used.

If this message appears, suspect first that the calculation might be faulty, then, if you feel confident, use PL to monitor a single SCF. Based on the SCF results either increase the number of allowed iterations (default: 200) or use PULAY, or Camp-King, or a mixture.

If nothing works, then consider slackening the SCF criterion. This will allow heats of formation to be calculated with reasonable precision, but the gradients are likely to be imprecise.

Fatal error in reading from channel 9 (FATAL)

A fatal error has occurred during an attempted restart of a DRC or IRC calculation. Likely causes are:

- The restart file does not exist.
- The restart file is from a different type of job.
- The restart file was written in FORMATTED or UNFORMATTED code, and an attempt was made to read it in in the other code. This is most likely to happen if MOPAC has been recompiled between runs.

Fatal error in trying to open RESTART file (FATAL)

In a FORCE calculation, the RESTART file is missing. Remove **RESTART** and re-run.

FAULT DETECTED IN INTERNAL COORDINATES (FATAL)

The nature of the fault in the internal coordinates is described in the output file immediately before this message. Correct fault and re-run.

FAULT IN READ OF AB INITIO DERIVATIVES (FATAL)

When AIDER is used, the *ab initio* derivatives must be supplied after the Z-matrix. Correct fault and re-run.

FAULTY LINE: *text* (FATAL)

An error was detected during the read of an EXTERNAL parameter set. An unrecognized parameter type was used.

See p. 43 for details of how to define parameters.

FILE *file.den* is missing (FATAL)

An attempt has been made to read in an old density matrix (*fime.den*), but this file apparently does not exist in the subdirectory. Correct the error and re-run.

First atom must not be H if keyword H2O used (FATAL)

In TOM calculations where H2O is used, the first atom must not be hydrogen. Hydrogen atoms are assigned values that depend on the type of atom they are attached to. Re-arrange the system so that the positions of hydrogen atoms are defined using other atoms.

GAUSSIAN INPUT REQUIRES STAND-ALONE JOB (FATAL)

Because of the way Gaussian geometries are recognized, only one such geometry is permitted in any given run, unless AIGIN is used. To correct this fault, either add AIGIN or break the run into parts, and run each part as a separate job.

Geometry cannot be optimized when TOM is used (FATAL)

In its current form, the gradients in TOM calculations are incorrect. The best strategy here is to optimize the geometry first, possibly using **COSMO**, then do a single SCF using TOM.

GEOMETRY CONTAINS FAULTS ... (FATAL)

Errors in the geometry were detected when the Lewis structure was constructed in a MOZYME calculation. Correct the errors using information in the output, or add GEO-OK, and re-run.

Geometry in PLATO is unrecognizable! (FATAL)

This error occurs when a geometry appears to be cubic, but does not belong to any of the cubic point groups. Check the geometry to verify that it is what is wanted. If it is, then add NOSYM and re-run; this will prevent the symmetry routines being used.

Geometry is apparently cubic ... (FATAL)

A severe error. The geometry has confused the symmetry recognition subroutines. Most likely, the geometry is nonsense. Examine the geometry printed after this message, and take corrective action. If the geometry is correct, add NOSYM, to disable the symmetry features.

GEOMETRY IS FAULTY (FATAL)

The nature of the fault in the geometry is described in the output file immediately before this message. Correct fault and re-run.

GEOMETRY TOO UNSTABLE FOR EXTRAPOLATION ...

In a reaction path calculation the initial geometry for a point is calculated by quadratic extrapolation using the previous three points.

If a quadratic fit is likely to lead to an inferior geometry, then the geometry of the last point calculated will be used. The total effect is to slow down the calculation, but no user action is recommended.

GNORM HAS BEEN SET TOO LOW ...

By default, the lowest value for GNORM is 0.01. To override this, add LET. There is no routine need to reduce the GNORM below 0.01, and if LET is used, the geometry optimization procedures are modified. Because of this, LET should not be used routinely.

GRADIENT IS TOO LARGE TO ALLOW ... **(FATAL)**

Before a FORCE calculation can be performed the gradient norm must be so small that the third and higher order components of energy in the force field are negligible. If, in the system under examination, the gradient norm is too large, a warning message will be printed and the calculation stopped, unless LET has been specified. In some cases the **FORCE** calculation may be run only to decide if a state is a ground state or a transition state, in which case the results have only two interpretations. Under these circumstances, LET may be warranted.

GRADIENT IS VERY LARGE ...

In a calculation of the thermodynamic properties of the system, if the rotation and translation vibrations are non-zero, as would be the case if the gradient norm was significant, then these 'vibrations' would interfere with the low-lying genuine vibrations. The criteria for THERMO are much more stringent than for a vibrational frequency calculation, as it is the lowest few genuine vibrations that determine the internal vibrational energy, entropy, etc.

GREENS FUNCTION IS NOT ALLOWED WITH EXTENDED SYSTEMS **(FATAL)**

The program has not been set up to allow Greens function corrections to polymers and other extended systems. No correction is possible, this kind of calculation cannot be run.

Hydrogen atom N is ... **(FATAL)**

In a MOZYME calculation, hydrogen atom N is unreasonably far from any heavy atom. Check the geometry and re-run.

Hydrogen atoms must be defined using bond-lengths when keyword H20 is used **(FATAL)**

In TOM calculations where H20 is used, hydrogen atoms are assigned values that depend on the type of atom they are attached to. Re-arrange the system so that the positions of hydrogen atoms are defined in terms of their bond-lengths.

HYDROGENS MUST BE BONDED TO NON ... **(FATAL)**

In a MOZYME calculation, hydrogen atoms should be bonded to non-hydrogen atoms. If molecular hydrogen is present, add LET, otherwise remove the molecular hydrogen.

If QPMEP is used, then PMEPE must also be present **(FATAL)**

QPMEP is a keyword that modifies a PMEPE (Parametric Molecular Electrostatic Potential) calculation. On its own, QPMEP will do nothing useful.

ILLEGAL ATOMIC NUMBER **(FATAL)**

An element has been specified by an atomic number which is not in the range 1 to 107. Check the data: the first datum on one of the lines is faulty. Most likely line 4 is faulty.

IMPOSSIBLE NUMBER OF CLOSED SHELL ELECTRONS **(FATAL)**

The keywords used imply that the number of closed shells (doubly-occupied levels) is less than zero! Correct the error in the data set, and re-run.

IMPOSSIBLE NUMBER OF OPEN SHELL ELECTRONS **(FATAL)**

The keyword OPEN($n1, n2$) has been used, but for an even-electron system $n1$ was specified as odd or for

an odd-electron system $n1$ was specified as even. Either way, there is a conflict which the user must resolve.

IMPOSSIBLE OPTION REQUESTED (FATAL)

A general catch-all. This message will be printed if two incompatible options are used, such as both MINDO/3 and AM1 being specified. Check the keywords, and resolve the conflict.

IMPOSSIBLE VALUE OF DELTA S (FATAL)

The keywords used imply that either the number of α or the number of β electrons is negative! Correct error and re-run.

INPUT FILE MISSING OR EMPTY (FATAL)

The data set is either empty or does not exist, or MOPAC has not been installed correctly. Correct error and re-run.

INSUFFICIENT DATA ON DISK FILES FOR A FORCE (FATAL)

A FORCE calculation has been attempted using RESTART, however, the <filename>.res file is either corrupt or does not exist. The best course of action would be to start over from the beginning—that is, remove RESTART, and re-run the job.

JOB STOPPED BY OPERATOR

Any MOPAC calculation, for which the shut command works, can be stopped by a user who issues the command shut <filename>, from the directory which contains <filename>.dat.

MOPAC will then stop the calculation at the first convenient point, usually after the current cycle has finished. A restart file will be written and the job ended. The message will be printed as soon as it is detected, which would be the next time the timer routine is accessed.

KEYWORD AIDER SPECIFIED, BUT NOT PRESENT AFTER Z-MATRIX. JOB STOPPED (FATAL)

When AIDER is used, the *ab initio* derivatives must be supplied after the Z-matrix. Correct fault and re-run.

KEYWORD ANALYT CANNOT BE USED HERE (FATAL)

Analytical derivatives (which use STO n Gaussian orbitals) cannot be used with RHF open shell derivatives calculated using Liotard's method. To correct this, either add NOANCI or remove ANALYT.

LINE OF KEYWORDS DOES NOT HAVE ENOUGH SPACES FOR PARSING. PLEASE CORRECT LINE. (FATAL)

Every keyword must be preceded by a space. This applies to the first keyword. As supplied, the keywords line does not have a space before the first keyword, and there is no space to move the keywords around in order to put a space there.

Delete or abbreviated keywords so that there are unused spaces.

MAX. NUMBER OF ATOMS ALLOWED: ... (FATAL)

To correct this, edit the file sizes.h to increase the value of NUMATM, re-run "make" and re-run the job.

MAXIMUM ALLOWED NUMBER OF ANIONS EXCEEDED (FATAL)

In a routine MOZYME calculation, the maximum number of anions allowed is 200. If there are more than 200, add LET, however, first check that the data set is correct.

MAXIMUM ALLOWED NUMBER OF CATIONS EXCEEDED (FATAL)

In a routine MOZYME calculation, the maximum number of cations allowed is 200. If there are more than 200, add LET, however, first check that the data set is correct.

Table 8.1: Names of Parameters for use with EXTERNAL=<file>

Uss	Betas	Gp2	GSD
Upp	Betap	Hsp	GPD
Udd	Betad	AM1	GDD
Zs	Gss	Expc	FN1
Zp	Gsp	Gaus	FN2
Zd	Gpp	Alp	FN3

MICROSTATES SPECIFIED BY KEYWORDS BUT MISSING FROM DATA (FATAL)

If MICROS= n is present, then after the geometry and symmetry data, if any, there should be a line with the word MICRO followed by n microstates.

See p. 56 for examples .

MISSING VAN DER WAALS RADIUS *Chemical – symbol* (FATAL)

In the COSMO method and in generating Lewis structures, van der Waals radii are used. If a VDW radius is missing, it can be supplied by use of , see p. 83.

MIXED PARAMETER SETS. USE ''PARASOK'' TO CONTINUE (FATAL)

A calculation using mixed parameter sets has been attempted. By default, MNDO parameters are used if the element has not been parameterized for the specified method. Use of such mixed parameter sets is not usually recommended. To allow such sets to be used, add PARASOK.

MODIFY SUBROUTINE NAMES (FATAL)

The MOZYME function is designed to work with proteins of up to 2,000 residues, about 28,000 atoms. Larger systems can be run, provided no attempt is made to work out the residue sequence. This means that RESIDUES, PDBOUT, and RESEQ cannot be used. If larger systems need to be run, modify MAXRES in "atomrs.F" and in "names.F".

MORE THAN 3N-6 COORDINATES OPTIMIZED! (FATAL)

In an EF calculation, more than $3N-6$ coordinates are flagged for optimization. By implication, at least one root of the Hessian matrix is exactly zero. Since EF involves using the inverse of an approximate Hessian, the method is intrinsically unstable. However, for most systems, the geometry optimizes completely before the Hessian is accurate enough for the instability to cause problems.

To correct this fault, either reduce the number of coordinates being optimized, or add GEO-OK.

MORE THAN ONE GEOMETRY OPTION HAS BEEN SPECIFIED (FATAL)

The keywords indicate that two or more geometric operations have been requested for one system. Only one operation (e.g. geometry optimization by TS or by NLLSQ) is allowed for any given system.

MULLIKEN POPULATION NOT AVAILABLE WITH UHF (FATAL)

The requested operation, a Mulliken population analysis using a UHF wavefunction, has not been written in the program. This type of calculation will not run.

NAME NOT FOUND (FATAL)

Various atomic parameters can be modified in MOPAC by use of EXTERNAL=. These comprise the symbols given in Table 8.1.

Thus to change the Uss of hydrogen to -13.6 the line USS H -13.6 could be used. If an attempt is made to modify any other parameters, then an error message is printed, and the calculation terminated.

NEGATIVE SYMBOLICS MUST BE PRECEDED BY THE POSITIVE EQUIVALENT (FATAL)

When specifying GAUSSIAN Z-matrix geometries, a negative symbolic must be related to an already defined positive symbolic. Correct the data set and re-run.

NEWGEO CAN ONLY BE RUN IF THERE ARE NO DUMMY ATOMS, OR IF 'RESEQ' IS USED (FATAL)

In a protein, NEWGEO converts backbone atoms to Cartesian, and side-chain atoms to internal coordinates. For this function to work correctly, there must not be any dummy atoms. Dummy atoms can be removed by RESEQ, or by a OSCF + INT run.

NLLSQ USED WITH REACTION PATH; (FATAL)

The capability of using NLLSQ with a reaction path is not available within MOPAC. As an alternative, use TS.

NO ATOMS IN SYSTEM (FATAL)

The system provided does not contain any atoms! Check the data-set. A common error is to have a blank line before the keyword line. There should be exactly three lines before the Z-matrix, unless '+' is used.

NO DUMMY ATOMS ALLOWED BEFORE (FATAL)

In a FORCE calculation on a polymer or solid, no dummy atoms are allowed. This is a program limitation. Modify the data-set (use OSCF and INT or XYZ to get rid of the dummy atoms), and re-run.

NO PEPTIDE LINKAGES FOUND. CHECK DATA SET (FATAL)

At least one of the following keywords: RESIDUES, PDBOUT, and RESEQ was used on a system that had no peptide linkages. Modify the data set and re-run.

NO POINTS SUPPLIED FOR REACTION PATH (FATAL)

A reaction path calculation is indicated by a '-1' in the optimization flags. If STEP=*n* and POINT=*m* are present, then the reaction path is defined by *n* and *m*. If these keywords are not present, the reaction path must be specified by numbers after the Z-matrix and symmetry data (if any). See p. 224 for an example.

NO RESTART FILE EXISTS! (FATAL)

An attempt has been made to restart a job, but the <filename>.res file does not exist. The easiest correction is to remove RESTART and re-run.

NOANCI MUST BE USED FOR RHF OPEN-SHELL SYSTEMS THAT INVOLVE TRANSLATION VECTORS (FATAL)

Liotard's analytical RHF open shell derivatives have not been extended to allow polymers or solids to be calculated. An alternative is to use UHF.

NONET SPECIFIED WITH ODD NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

When NONET is specified, the system must have an even number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

NUMBER OF DOUBLY FILLED PLUS PARTLY FILLED LEVELS GREATER THAN TOTAL NUMBER OF ORBITALS (FATAL)

The keywords used here imply a system that is larger than that used. Correct data set (probably by changing the keywords) and re-run.

NUMBER OF ELECTRONS IN M.O.s BELOW ACTIVE SPACE IS LESS THAN ZERO (FATAL)

In a C.I. calculation, the active space extends below the lowest energy level. Correct data set (probably by changing the keywords) and re-run.

NUMBER OF M.O.s IN ACTIVE SPACE EXCEEDS MAXIMUM ALLOWED SIZE OF ACTIVE SPACE (FATAL)

In a C.I. calculation, the number of M.O.s in the active space is greater than that allowed by the program. Modify the data set (probably by changing the keywords) and re-run.

NUMBER OF M.O.s REQUESTED IN C.I. IS GREATER THAN THE NUMBER OF ORBITALS (FATAL)

In a C.I. calculation, the active space requested is greater than the number of orbitals in the system. Correct data set (probably by changing the keywords) and re-run.

NUMBER OF OCCUPIED ORBITALS IS INCORRECT IN MAKVEC (FATAL)

During the construction of the starting LMOs in a MOZYME calculation, more occupied LMOs were found than expected. This is a programming problem. Please inform Dr Stewart as soon as possible.

(FATAL)

NUMBER OF OPEN-SHELLS ALLOWED IN C.I. IS LESS THAN THAT SPECIFIED BY OTHER KEYWORDS The size of the active space in a C.I. calculation implied by $C.I.=(n,m)$ is less than that implied by other keywords, e.g., SEXTET. Correct data set by changing the keywords and re-run.

NUMBER OF PARTICLES, nm GREATER THAN ...

When user-defined microstates are not used, the MECI will calculate all possible microstates that satisfy the space and spin constraints imposed. This is done in PERM, which permutes N electrons in M levels. If N is greater than M , then no possible permutation is valid. This is not a fatal error—the program will continue to run, but no C.I. will be done.

NUMBER OF PERMUTATIONS TOO GREAT, LIMIT $nnnn$ (FATAL)

Unless the file `meci.h` is changed, the number of permutations of alpha or beta microstates is limited to $4 \times \text{MAXCI}$ or 4800. Thus if 3 alpha electrons are permuted among 5 M.O.s, that will generate $10 = 5!/(3!2!)$ alpha microstates, which is an allowed number. However if 7 alpha electrons are permuted among 15 M.O.s, then 6435 alpha microstates result and the arrays defined will be insufficient. To correct this error, increase MAXCI in `meci.h` and recompile.

NUMERICAL PROBLEMS IN BRACKETING LAMBDA

Although this is not a deadly error, it does indicate that there are potential problems in optimizing geometries. If the run finishes correctly, don't worry. If the geometry is not optimized sufficiently, try one or more of the following strategies:

- Use LET. This allows more of the potential energy surface to be sampled, thus giving more information to the Hessian.
- Tighten up the SCF criterion. Try RELSCF=0.1 or RELSCF=0.01.
- If the calculation involves an open shell RHF, consider running it with UHF.
- Carefully examine the data set—is there any possibility that it is faulty?
- Go to a different coordinate system. If Cartesian, go to internal coordinates, and *vice versa*.

OCCUPIED C VECTOR DAMAGED IN SELMOS (FATAL)

In a MOZYME calculation, when RAPID is used (in a partial geometry optimization), the occupied M.O. coefficients were found to be corrupt. Please inform Dr Stewart as soon as possible.

OCCUPIED IC VECTOR DAMAGED IN SELMOS (FATAL)

In a MOZYME calculation, when RAPID is used (in a partial geometry optimization), the list of atoms in the occupied M.O. was found to be corrupt. Please inform Dr Stewart as soon as possible.

OCTET SPECIFIED WITH EVEN NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

When OCTET is specified, the system must have an odd number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

OLDENS FILE FOR *file.den* IS CORRUPT (FATAL)

Although a file called *file.den* exists, its contents do not match the data set supplied for the calculation. The easiest option is to delete OLDENS and re-run. Note: OLDENS cannot be used to switch from a conventional to a MOZYME run, or *vice versa*.

OLDGEO used and previous geometry had no atoms. (FATAL)

For OLDGEO to work, an earlier calculation in the same data set must exist, and must contain at least one atom. Correct fault and re-run.

OMIN MUST BE BETWEEN 0 AND 1 (FATAL)

Keyword OMIN=*n* has been used, with an unreasonable value for *n*. Modify *n* and re-run.

ONLY *N* POINTS ALLOWED IN REACTION COORDINATE (FATAL)

The maximum number of points on a reaction coordinate is three times the maximum number of atoms allowed. To increase this number, edit the file *sizes.h* to increase the value of NUMATM, re-run "make" and re-run the job.

ONLY DIHEDRAL SYMBOLICS CAN BE PRECEDED BY A '-' SIGN (FATAL)

Symmetry relationships are allowed when a geometry is read in in Gaussian format, however the range of relationships is limited to setting bond lengths equal, setting bond angles equal, setting dihedrals equal, and setting dihedrals equal to the negative of a reference dihedral. An attempt has been made to set a bond length or a bond angle to the negative of a reference. This is not allowed. Correct the data set and re-run.

Only H, C, N, O, F, and S allowed (FATAL)

The TOMASI model has parameters for H, C, N, O, F, and S only. This model will not work for systems that contain other elements. Try using COSMO?

ONLY ONE REACTION COORDINATE PERMITTED (FATAL)

In a reaction coordinate calculation, only one optimization flag can be set to "-1", all others must either be "0" or "1". In order for two optimization flags to be set to "-1", a grid calculation must be run. This involves keywords STEP1=*n.nn* and STEP2=*m.mm*. For details on how to do this, see p. 76.

PANIC: Temporary array being destroyed ...

Although this is not a fatal error, it does indicate a severe programming error in the program. Each time a temporary array is destroyed, a check is made to verify that the correct amount of memory is freed up. If the check fails, this message will be printed. Please inform Dr Stewart as soon as possible.

PARAMETERS FOR SOME ELEMENTS MISSING (FATAL)

An attempt has been made to run a calculation on a system that contains atoms for which there are no parameters. Check that the method you are using (p. 86). If this bug occurs during a port of MOPAC, it is likely to be caused by the BLOCKDATA file not being used by the compiler. One way around this fault is to paste the file *block.F* at the end of the file *mopac.F*, and then delete *block.F*. This usually corrects this problem.

PERMANENT INTEGER ARRAY *name* DAMAGED (FATAL)

All permanent arrays are separated by a single datum, that acts as a sensor. If this datum is changed, there

is a high probability that part or all of the array has been overwritten, as the result of an error in programming. Please inform Dr Stewart as soon as possible.

PERMANENT REAL ARRAY *name* DAMAGED (FATAL)

All permanent arrays are separated by a single datum, that acts as a sensor. If this datum is changed, there is a high probability that part or all of the array has been overwritten, as the result of an error in programming. Please inform Dr Stewart as soon as possible.

POLAR does not work with open-shell RHF (FATAL)

The POLAR method only works with RHF closed shell systems. If only the polarizability is needed, use STATIC instead of POLAR.

POLAR does not work with UHF (FATAL)

The POLAR calculation only works with Restricted Hartree Fock calculations, both closed shell and open shell, and both ground and excited states. If only the polarizability is needed, use STATIC; this uses external fields and works with UHF, however it is not as precise as POLAR.

POLAR has not been implemented with MINDO (FATAL)

POLAR only works with the NDDO methods, MNDO, AM1, PM3, and MNDO-*d*. If only the polarizability is needed, use STATIC; this uses external fields and works with MINDO/3, however it is not as precise as POLAR.

PROBLEM IN SYMR (FATAL)

There is a problem in subroutine SYMR. The system has symmetry, but small distortions are preventing the symmetry operations from being done correctly—the subroutine has become confused. To correct this, add , p. 60.

QUARTET SPECIFIED WITH EVEN NUMBER (FATAL)

When QUARTET is specified, the system must have an odd number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

QUINTET SPECIFIED WITH ODD NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

When QUINTET is specified, the system must have an even number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

Ran out of storage for *arrayname* (FATAL)

The amount of dynamic memory reserved for storing arrays is less than that required by the program. This is a programming bug. The strategy to follow to identify the fault is:

- Re-run the calculation, but add SIZES. This will print out the amount of memory reserved for each array, and the arrays created and destroyed, and the arrays that exist when the run stopped.
- Determine whether the error is in the integer or real arrays.
- Check the total memory reserved for the permanent arrays. This should be slightly greater than the total actually used. If it is less, then a change needs to be made to `setupr.F` or `setupi.F`, to include the missing array.
- Check the total memory reserved for the temporary arrays. This should be slightly greater than the total actually used. If it is less, identify which array(s) exist at the end of the calculation, for which memory was not reserved at the start. Make the appropriate correction to `tmpi.F` (if the array is integer), `tmpmr.F` (if real, and conventional), or `tmpzr.F` (if real, and MOZYME).

RESEQ cannot be used with XYZ (FATAL)

The pair of keywords RESEQ and XYZ is not supported. However, the effect of these two keywords can be reproduced in a two stage calculation by using RESEQ in the first calculation and OLDGEO, XYZ, and OSCF in the second calculation.

RESTART FILE EXISTS, BUT IS CORRUPT (FATAL)

In a BFGS run involving RESTART, the file <filename>.res, is corrupt. Remove RESTART and re-run.

RESTART FILE EXISTS, BUT IS FAULTY (FATAL)

In a FORCE calculation, involving RESTART, the file <filename>.res, is corrupt. Remove RESTART and re-run.

Restart file is corrupt! (FATAL)

In a run involving RESTART, the file <filename>.res, is corrupt. Remove RESTART and re-run.

ROOT REQUESTED DOES NOT EXIST IN C.I. (FATAL)

A specific excited state has been specified, but it does not exist in the set of states calculated. Correct the data set and re-run.

RSOLV MUST NOT BE NEGATIVE (FATAL)

In COSMO calculations, the radius of a solvent molecule must be positive. Correct data set and re-run.

Run stopped because keyword LEWIS was used

LEWIS is only used when a description of the Lewis structure is wanted. This can be done very rapidly, provided no quantum mechanical calculations are run. Therefore, when LEWIS is used, the calculation is stopped immediately after the Lewis structure is printed. If the calculation is to be continued, remove LEWIS.

SEPTET SPECIFIED WITH ODD NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

When SEPTET is specified, the system must have an even number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

SETUP FILE MISSING, EMPTY OR CORRUPT (FATAL)

In a run involving SETUP, the SETUP file is either missing or is corrupt. Either create a valid SETUP file or remove SETUP from the data set, and re-run.

SEXTET SPECIFIED WITH EVEN NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

When SEXTET is specified, the system must have an odd number of electrons. Check (a) the system, and (b) the charge (if any). Correct data set and re-run.

SIGMA USED WITH REACTION PATH; (FATAL)

The only geometry options allowed with reaction paths are EF and the default BFGS optimizers. Delete SIGMA and re-run.

Size of active space allowed: N (FATAL)

The largest active space in a C.I. calculation is N . If a larger size is needed, edit `meci.h` to increase the value of `NMECI`, re-run "make", and re-run the job.

SOME ELEMENTS HAVE BEEN SPECIFIED FOR WHICH NO PARAMETERS ARE AVAILABLE ... (FATAL)

Parameters are not available for the element with atomic number N . If new parameters are available, these can be supplied to MOPAC by use of , see p. 43.

Something disastrous has happened (**FATAL**)

In a MOZYME calculation, something has gone wrong in an SCF calculation. The effect is to increase the heat of formation suddenly by over 200 kcal.mol⁻¹. The job will automatically shut down, and can subsequently be restarted. In the unlikely event that the shutdown sequence fails to start, the geometry will be printed, and the calculation stopped.

Re-start the calculation, either using RESTART or the final geometry printed, but do *not* use OLDENS.

SPECIFIED SPIN COMPONENT NOT SPANNED BY ACTIVE SPACE (**FATAL**)

A spin state has been specified, but the active space is too small to allow that state to exist. Correct the data set and re-run.

Storage needed for next step: *NNN* (**FATAL**)

The amount of dynamic memory reserved for storing arrays is less than that required by the program. This is a programming bug. The strategy to follow to identify the fault is:

- Re-run the calculation, but add SIZES. This will print out the amount of memory reserved for each array, and the arrays created and destroyed, and the arrays that exist when the run stopped.
- Determine whether the error is in the integer or real arrays.
- Check the total memory reserved for the permanent arrays. This should be slightly greater than the total actually used. If it is less, then a change needs to be made to *setupr.F* or *setupi.F*, to include the missing array.
- Check the total memory reserved for the temporary arrays. This should be slightly greater than the total actually used. If it is less, identify which array(s) exist at the end of the calculation, for which memory was not reserved at the start. Make the appropriate correction to *tmpi.F* (if the array is integer), *tmpmr.F* (if real, and conventional), or *tmpzr.F* (if real, and MOZYME).

STRUCTURE UNRECOGNIZABLE

A problem was encountered while trying to identify the amino acid residues in a protein. Although not deadly, this represents a fault in the program. Please inform Dr Stewart as soon as possible.

SYMMETRY SPECIFIED, BUT CANNOT BE USED IN DRC

This is self explanatory. The DRC requires all geometric constraints to be lifted. Any symmetry constraints will first be applied, to symmetrize the geometry, and then removed to allow the calculation to proceed.

SYSTEM DOES NOT APPEAR TO BE OPTIMIZABLE

This is a gradient norm minimization message. These routines will only work if the nearest minimum to the supplied geometry in gradient-norm space is a transition state or a ground state. Gradient norm space can be visualized as the space of the scalar of the derivative of the energy space with respect to geometry. To a first approximation, there are twice as many minima in gradient norm space as there are in energy space.

It is unlikely that there exists any simple way to refine a geometry that results in this message. While it is appreciated that a large amount of effort has probably already been expended in getting to this point, users should steel themselves to writing off the whole geometry. It is not recommended that a minor change be made to the geometry and the job re-runned.

Try using TS or SIGMA instead of POWSQ.

SYSTEM SPECIFIED WITH ODD NUMBER OF ELECTRONS, CORRECT FAULT (**FATAL**)

When EXCITED or BIRADICAL is used, the system *must* be a singlet—*must* have an even number of electrons. Correct the data set and re-run.

SYSTEMS WITH TV CANNOT BE RUN WITH 'INT' (FATAL)

The effect of INT is to convert the geometry into internal coordinates. This is done in two steps: first, the geometry is converted into Cartesian coordinates, this removes any dummy atoms; then it is converted into internal coordinates. An unwanted, but logical, consequence of this is that translation vectors are made useless. To achieve the effect of INT, remove the translation vectors, run the calculation with OSCF, then add the translation vectors back in 'by hand'. Note that translation vectors can be Cartesian or internal.

TEMPERATURE RANGE STARTS TOO LOW, ...

The thermodynamics calculation assumes that the statistical summations can be replaced by integrals. This assumption is only valid above 100K, so the lower temperature bound is set to 100, and the calculation continued.

The COSMO option cannot be used ... (FATAL)

Due to a program limitation, sparkles cannot be used in a COSMO calculation. There is no recovery—try using real atoms instead of sparkles.

The data set contains alternative location indicators. Keyword ALT_A must be used (FATAL)

The PDB format allows for alternate locations. According to this format, atoms are defined by four characters (characters 13-16), after which comes the alternate location indicator. Usually, this (the 17th character) is a space, however, in atoms that have positional disorder the 17th character will be a letter, e.g. 'A', 'B', 'C', etc. A valid geometry will contain only one of the alternative locations, and this can be identified in the data set by use of `ALT_A=letter`. For a full definition of the PDB format, see the WWW sites: <http://pdb.pdb.bnl.gov/pdb-docs/atoms.html> and http://www.pdb.bnl.gov/pdb-docs/Format.doc/Contents_Guide_21.html.

The data set contains alternative location indicators. Keyword ALT_R must be used (FATAL)

The PDB format allows for alternate residues. According to this format, a given residue site can be occupied by one of two or more possible residues. The indicator is character 27, and by default is a space. If there are alternative residues possible, these are indicated by a letter, e.g. 'A', 'B', 'C', etc. A valid geometry will contain only one of the alternative residues, and this can be identified in the data set by use of `ALT_R=letter` or by `ALT_R=` .

THE FOLLOWING SYMBOL HAS BEEN DEFINED MORE THAN ONCE: *symbol* (FATAL)

A symbolic parameter in a Gaussian Z-matrix geometry has been defined more than once. Remove the excess definitions and re-run.

THE FOLLOWING SYMBOL WAS NOT USED: *symbol* (FATAL)

A symbolic parameter in a Gaussian Z-matrix geometry has been defined, but was not present in the Z-matrix. Remove the definition and re-run.

THE GEOMETRY DATA-SET CONTAINED ERRORS (FATAL)

The geometry was not in MOPAC Z-matrix format, or in Cartesian format, or in Gaussian format. The output immediately before this error message describes the errors detected. Correct the geometry and re-run.

THE GREENS FUNCTION IS LIMITED TO 200 ORBITALS (FATAL)

The Greens function correction to the I.P.'s calculation is limited to systems of 200 orbitals. Larger systems cannot be run. This is a program limitation.

THE HAMILTONIAN REQUESTED IS NOT AVAILABLE IN THIS PROGRAM (FATAL)

This is a bug that occurs only during a port of MOPAC. It is probably caused by the BLOCKDATA file not being used by the compiler. One way around this fault is to paste the file `block.F` at the end of the file `mopac.F`, and then delete `block.F`. This usually corrects this problem.

The input data file does not exist **(FATAL)**

This is a bug that occurs only during a port of MOPAC. The first step in trying to remove this bug is to check that the command:

```
% mopac.exe test
```

will run the data set `test.dat`. Once this works, that is, once it does not generate this error message, check out `mopac.csh`.

THE SCF CALCULATION FAILED. **(FATAL)**

The SCF failed to form. To correct this, add PL and try using different convergers. In order, try `SHIFT= n` , `$n=50$` is a good starting value, then PULAY, then CAMP. Quite often, the fault lies in the geometry. Use the information from the output generated by PL as a guide.

THE STATE REQUIRED IS NOT PRESENT IN THE SET OF CONFIGURATIONS AVAILABLE **(FATAL)**

A specific excited state has been specified, but it does not exist in the set of states calculated. Correct the data set and re-run.

There are more than 200 atoms in moiety **(FATAL)**

A non-proteinaceous moiety in a protein apparently contains more than N atoms. If this is true, then edit "ligand.F" to change the value of NATOMR in subroutine MOIETY, re-run "make", and re-run the job.

There are more than 200 atoms in residue M **(FATAL)**

A residue apparently contains more than 200 atoms. A likely cause is if the residue contains covalent bonds to another part of the protein (common cross links, e.g. N-S, O-S, and S-S, are automatically broken when residues are being analyzed). Check the atom list printed after the error message to identify the atoms that are considered part of the residue, modify the data-set and re-run.

THERE ARE NO VARIABLES IN THE SADDLE **(FATAL)**

For a SADDLE calculation, the two geometries must be optimizable, that is, at least one coordinate must have the optimization flag set to "1". ...

There are too many keywords **(FATAL)**

Although each line of keywords can have up to 120 characters, the maximum number of characters plus spaces for all keywords is limited to 241. Reduce the number of keywords, or abbreviate those that can be shortened.

There is a bug in MOZYME **(FATAL)**

This message should never appear! When NEWGEO was used to convert a protein geometry to mixed coordinates, an absurd set of coordinates was detected. Please inform Dr Stewart as soon as possible.

THERE IS A FAULT IN GEOCHK **(FATAL)**

During the check of the Lewis structure, the number of electrons calculated from the Lewis structure does not equal the number of electrons from the valence shell of the atoms plus any net charge. This can only happen if there is a program bug. Please inform Dr Stewart as soon as possible.

THERE IS A FAULT IN RESEQ **(FATAL)**

During an attempt to re-sequence the atoms in a protein, the number of atoms in the new system differed from the number in the original system. This can only happen if there is a program bug. Please inform Dr Stewart as soon as possible.

There is a mistake in SUTUPI or TMPI (FATAL)

The amount of memory needed in annihilating matrix elements in a MOZYME calculation is estimated. Sometimes this estimate is low, and more memory is needed. Fix this problem by following the instructions printed after the error message. Please inform Dr Stewart as soon as possible.

There is a mistake in SUTUPR or TMPZR (FATAL)

The amount of memory needed in annihilating matrix elements in a MOZYME calculation is estimated. Sometimes this estimate is low, and more memory is needed. Fix this problem by following the instructions printed after the error message. Please inform Dr Stewart as soon as possible.

THERE IS A RISK OF INFINITE LOOPING ...

The SCF criterion has been reset by the user, and the new value is so small that the SCF test may never be satisfied. This is a case of user beware!

THERE MUST BE EXACTLY THREE VELOCITY DATA PER LINE (FATAL)

The format for the initial velocity in a DRC calculation is three numbers per line, corresponding to the x , y , and z speeds in $\text{cm}\cdot\text{sec}^{-1}$. Correct the data set and re-run.

THIS MESSAGE SHOULD NEVER APPEAR, CONSULT A PROGRAMMER!

This message should never appear; a fault has been introduced into MOPAC, most probably as a result of a programming error. If this message appears in the basic version of MOPAC (a version ending in 00), please contact JJPS as I would be most interested in how this was achieved.

THREE ATOMS BEING USED TO DEFINE ... (FATAL)

If the Cartesian coordinates of an atom depend on the dihedral angle it makes with three other atoms, and those three atoms fall in an almost straight line, then a small change in the Cartesian coordinates of one of those three atoms can cause a large change in its position. Normally, the connectivity will automatically be changed to prevent this happening, however, if there is no obvious way to correct the problem, this message will be printed. When that happens, the data should be changed to make the geometric specification of the atom in question less ambiguous. Note that neither LET nor GEO-OK will allow the calculation to proceed.

TIME UP

The time defined on the keywords line or 3,600 seconds, if no time was specified, is likely to be exceeded if another cycle of calculation were to be performed. A controlled termination of the run would follow this message. The job may terminate earlier than expected: this is ordinarily due to one of the recently completed cycles taking unusually long, and the safety margin has been increased to allow for the possibility that the next cycle might also run for much longer than expected.

TOO MANY CONFIGURATIONS (FATAL)

The size of the C.I. matrix requested is larger than that allowed by MOPAC. Either reduce the size requested, or increase the allowed size. To do that, increase the value of MAXCI in the file `meci.h`, and recompile.

TOO MANY ITERATIONS IN LAMDA BISECT (FATAL)

During a run involving EF, the search procedure failed. This was most likely due to a faulty geometry. Check the geometry carefully, make corrections, and re-run.

TRIPLET SPECIFIED WITH ODD NUMBER OF ELECTRONS, CORRECT FAULT (FATAL)

If TRIPLET has been specified the number of electrons must be even. Check the charge on the system, the empirical formula, and whether TRIPLET was intended.

TRUST RADIUS NOW LESS THAN ...

When EF is used, the calculated trust radius has become too small. An easy fix is to add LET, another option is to specify RMIN=-10.

TS FAILED TO LOCATE TRANSITION STATE (FATAL)

The geometry is almost certainly faulty. Locating transition states is still more art than science. Modify geometry, and re-run.

TS SEARCH AND BFGS UPDATE WILL NOT WORK (FATAL)

In EF, a transition state optimization has been requested. The rarely used option IUPD=2 has been specified. This option is not allowed for transition states. Remove IUPD=2 and re-run.

TS SEARCH REQUIRE BETTER THAN DIAGONAL HESSIAN (FATAL)

When using TS, do not also use IGTHERS=0. Remove IGTHERS=0 and re-run.

TWO ADJACENT POINTS ARE IDENTICAL: (FATAL)

In a reaction path, adjacent points must be different. A common mistake is to have the first point (this comes from the Z-matrix) and the second point (this comes from the extra data after the Z-matrix and symmetry data (if any)) the same. Correct the data set (the fault is in the reaction coordinate data after the the Z-matrix and symmetry data (if any)).

TWO ATOMS ARE COINCIDENT. A FATAL ERROR. (FATAL)

In the input geometry, two atoms have the same coordinates. Correct the error and re-run.

UNABLE TO ACHIEVE SELF-CONSISTENCY

See the error-message: "FAILED TO ACHIEVE SCF"

UNACCEPTABLE VALUE FOR NO. OF ORBITALS ON ATOM (FATAL)

Allowed values for the number of orbitals per atom are 1, 4, and 9. Correct the EXTERNAL file, and re-run.

UNDEFINED SYMMETRY FUNCTION USED (FATAL)

Symmetry operations are restricted to those defined, i.e., in the range 1-19. Any other symmetry operations will trip this fatal message.

UNKNOWN RESIDUE: *N*

The 20 common amino acid residues are automatically identified. Other residues can be made known to MOPAC by use of the keyword XENO=*text*. The numbers printed after this message are the numbers of C, N, O, and S atoms found in the residue.

Inspect the residue, and identify which of the 20 common residues it is related to. In XENO specify the number of extra atoms needed to make up the numbers in the unknown residue. Re-run the calculation.

UNRECOGNIZED ELEMENT NAME (FATAL)

In the geometric specification a chemical symbol which does not correspond to any known element has been used. The error lies in the first datum on a line of geometric data.

UNRECOGNIZED HESS OPTION (FATAL)

The allowed values for *n* in HESS=*n* are 0, 1, and 2. Correct the keyword and re-run.

UNRECOGNIZED KEY-WORDS. (FATAL)

Check these keywords. Correct any misspellings and re-run, or, if the keywords are DEBUG keywords, add DEBUG and re-run.

UNRECOGNIZED SPECIES: *text*

While reading a PDB file, an unrecognized chemical symbol has been used. Recognized symbols are those for any element that can be handled by MOPAC, and the symbols 'X' (for hydrogen) and 'L' (for a lone pair).

The job will not be stopped by this error message, but the *text* should be examined to see if it is a real element, but with an unconventional symbol. If it is, add `PDB(code)`, where *code* is defined on p. 64. Even if *text* does not apply to an element, `PDB(code)` can still be used, but set the atomic number to zero. This will prevent the error message from recurring.

UPPER BOUND OF ACTIVE SPACE IS GREATER THAN THE NUMBER OF ORBITALS (FATAL)

The keywords used here imply a system that is larger than that used. Correct data set (probably by changing the keywords) and re-run.

Use AIGIN to allow more geometries to be used (FATAL)

Only one GAUSSIAN geometry is allowed in a run, unless each GAUSSIAN geometry is identified by AIGIN on the keyword line. Add AIGIN to each GAUSSIAN geometry keyword line, and re-run.

USE EITHER SAFE OR UNSAFE, BUT NOT BOTH (FATAL)

MOPAC can be compiled so as to either minimize memory demand or to maximize reliability. Depending on which option was used, the alternative option can be selected at run time by using SAFE or UNSAFE.

See p. 72 and p. 83 for more detail.

VALUE OF NPPA NOT ALLOWED: IT MUST BE *NNN* (FATAL)

This error is caused by a program bug. Please make a report to Dr. Stewart.

VAN DER WAALS (FATAL)

In the ESP method, van der Waals radii are used. Only the following elements are allowed: H, B, C, N, O, F, P, S, Cl, Br, I.

In the PMEP method, van der Waals radii are also used. Only the elements up to $Z=17$ and bromine are allowed.

VECPRT CAN ONLY PRINT ARRAYS OF SIZE .LT. 200 (FATAL)

To minimize output, the size of arrays that can be printed in a MOZYME calculation is limited to 200. Bigger arrays are truncated so that only the sub-array of size 200 is printed. If larger arrays are to be printed, edit "vecprt.F" to change MAXARR to the desired value, re-run "make", and re-run the job.

VIRTUAL C VECTOR DAMAGED IN SELMOS (FATAL)

In a MOZYME calculation, when RAPID is used (in a partial geometry optimization), the virtual M.O. coefficients were found to be corrupt. Please inform Dr Stewart as soon as possible.

VIRTUAL IC VECTOR DAMAGED IN SELMOS (FATAL)

In a MOZYME calculation, when RAPID is used (in a partial geometry optimization), the list of atoms in the virtual M.O. was found to be corrupt. Please inform Dr Stewart as soon as possible.

WARNING -- GEOMETRY IS NOT AT A STATIONARY POINT

Under certain circumstances the gradient norm can drop to zero, but the derivatives of the energy with respect to Cartesian coordinates might be quite large. If this happens, this error message will be printed.

To avoid this message, make sure that the geometry can be optimized, given the optimization flags you have chosen. In particular, if $(3N-6)$ optimization flags are set, and there are no dummy atoms, then it is unlikely that this message will be generated.

WARNING -- SPARKLES ARE NOT TREATED CORRECTLY

When ENPART is used on a system involving sparkles, the energy terms due to the sparkles are not printed correctly.

WARNING

Don't pay too much attention to this message. Thermodynamics calculations require a higher precision than vibrational frequency calculations. In particular, the gradient norm should be very small. However, it is frequently not practical to reduce the gradient norm further, and to date no-one has determined just how slack the gradient criterion can be before unacceptable errors appear in the thermodynamic quantities. The 0.4 gradient norm is only a suggestion.

WARNING! NO BACKBONE ATOMS IDENTIFIED

When NEWGEO is used to generate a mixed coordinate Z -matrix for a protein, this message will be generated if no backbone atoms (the $-\text{NH}-\text{CH}-\text{CO}-$ of a peptide) are found. Check that at least one "CO-NH" linkage exists.

Chapter 9

Criteria

MOPAC uses various criteria which control the precision of its stages. These criteria are chosen as the best compromise between speed and acceptable errors in the results. The user can override the default settings by use of keywords; however, care should be exercised as increasing a criterion can introduce the potential for infinite loops, and decreasing a criterion can result in unacceptably imprecise results. These are usually characterized by 'noise' in a reaction path, or large values for the trivial vibrations in a force calculation.

9.1 SCF criterion

SCFCRT

Defined in: ITER

Default value: 0.0001 kcal/mol

Basic test: Change in energy in kcal/mol on successive iterations is less than SCFCRT.

Exceptions:

- If PRECISE is specified, SCFCRT=0.000001
- If a polarization calculation, SCFCRT=1.D-11
- If a FORCE calculation, SCFCRT=0.0000001
- If SCFCRT=*n.nnn* is specified SCFCRT=*n.nnn*
- If a BFGS optimization, SCFCRT becomes a function of the difference between the current energy and the lowest energy of previous SCFs.

Secondary tests:

1. Change in density matrix elements on two successive iterations must be less than 0.001
2. Change in energy in eV on three successive iterations must be less than $10 \times \text{SCFCRT}$.

9.2 Geometric optimization criteria

GNORM "RMS GRADIENT = *n.nn* IS LESS THAN CUTOFF = *m.mm*"

Defined in: EF

Default value: 1.0

Basic test: The calculated GNORM is less than that defined by GNORM.

TOLERX “Test on X Satisfied”

Defined in: FLEPO

Default value: 0.0001 Ångstroms

Basic test: The projected change in geometry is less than TOLERX Ångstroms.

Exception: If GNORM is specified, the TOLERX test is not used.

DELHOF “Herbert’s Test Satisfied”

Defined in: FLEPO

Default value: 0.001

Basic test: The projected change in geometry is less than DELHOF kcals/mol.

Exception: If GNORM is specified, the DELHOF test is not used.

TOLERG “Test on Gradient Satisfied”

Defined in: FLEPO

Default value: 1.0

Basic test: The gradient norm in kcals/mol/Ångstrom is less than TOLERG multiplied by the square root of the number of coordinates to be optimized.

Exceptions:

- If GNORM=*n.nnn* is specified, TOLERG=*n.nnn* divided by the square root of the number of coordinates to be optimized, and the secondary tests are not done. If LET is not specified, *n.nnn* is reset to 0.01, if it was smaller than 0.01.
- If PRECISE is specified, TOLERG=0.2
- If a SADDLE calculation, TOLERG is made a function of the last gradient norm.

TOLERF “Heat of Formation Test Satisfied”

Defined in: FLEPO

Default value: 0.002 kcal/mol

Basic test: The calculated heats of formation on two successive cycles differ by less than TOLERF.

Exception: If GNORM is specified, the TOLERF test is not used.

Secondary tests: For the TOLERG, TOLERF, and TOLERX tests, a second test in which no individual component of the gradient should be larger than TOLERG must be satisfied.

Other tests: If, after the TOLERG, TOLERF, or TOLERX test has been satisfied three consecutive times the heat of formation has dropped by less than 0.3kcal/mol, then the optimization is stopped.

TOL2**Defined in:** POWSQ**Default value:** 0.4**Basic test:** The absolute value of the largest component of the gradient is less than TOL2**Exceptions:**

- If bf PRECISE is specified, TOL2=0.01
- If GNORM=*n.nn* is specified, TOL2=*n.nn*
- If LET is not specified, TOL2 is reset to 0.01, if *n.nn* was smaller than 0.01.

TOLS1**Defined in:** NLLSQ**Default value:** 0.000 000 000 001**Basic test:** The square of the ratio of the projected change in the geometry to the actual geometry is less than TOLS1.**Name:** <none>**Defined in:** NLLSQ**Default value:** 0.2**Basic test:** Every component of the gradient is less than 0.2.

Chapter 10

Debugging

There are three potential sources of difficulty in using MOPAC, each of which requires special attention. There can be problems with data, due to errors in the data, or MOPAC may be called upon to do calculations for which it was not designed. There are intrinsic errors in MOPAC which extensive testing has not yet revealed, but which a user's novel calculation uncovers. Finally there can be bugs introduced by the user modifying MOPAC, either to make it compatible with the host computer, or to implement local features.

For whatever reason, the user may need to have access to more information than the normal keywords can provide, and a second set, specifically for debugging, is provided. These keywords give information about the working of individual subroutines, and do not affect the course of the calculation.

10.1 Keywords for use in debugging

A full list of keywords for debugging subroutines:

1ELEC The one-electron matrix. See Note 1.

CARTAB Print all the character tables.

CHARST Calculation of state symmetry characters.

COMPFG Heat of Formation.

DCART Cartesian derivatives.

DEBUG See Note 2.

DEBUGPULAY Pulay matrix, vector, and error-function. See Note 3.

DENSITY Every density matrix. See Note 1.

DERI1 Details of DERI1 calculation.

DERI2 Details of DERI2 calculation.

DERITR Details of DERITR calculation.

DERIV All gradients, and other data in DERIV.

DERNVO Details of DERNVO calculation.

DFORCE Print Force Matrix.

DIAGG2 Details of the annihilation operation.

DIIS Details of DIIS calculation.

EIGS All eigenvalues.

FLEP0 Details of BFGS minimization.

FMAT See Note 1.

FOCK Every Fock matrix.

GEOCHK Details of the check on geometry.

GROUP Determination of symmetry group.

HCORE The one electron matrix, and two electron integrals.

HYBRID Construction of hybrid atomic orbitals.

ITER Values of variables and constants in ITER.

LARGE Increases amount of output generated by other keywords.

LEWIS Print the connectivity and Lewis structure.

LINMIN Function values, step sizes at all points in the line minimization (LINMIN or SEARCH).

MAKVEC Construction of the initial LMOs.

MOLDAT Molecular data, number of orbitals, "U" values, etc.

MOLSYM Calculation of molecular symmetry.

MECI C.I. matrices, M.O. indices, etc.

PL Differences between density matrix elements in ITER. See Note 4.

SIZES Print sizes of all arrays created and destroyed. Note 5.

SYM0IR Symmetry characters for vibrational frequencies.

SYMTRZ Symmetry determination (point group and IR's).

TIDY Tidy up of the LMOS (removing unused space).

TIMES Times of stages within ITER.

VECTORS All eigenvectors on every iteration. See Note 1.

Notes

1. These keywords are activated by the keyword `DEBUG`. Thus if `DEBUG` and `FOCK` are both specified, every Fock matrix on every iteration will be printed.
2. `DEBUG` is not intended to increase the output, but does allow other keywords to have a special meaning.
3. `PULAY` is already a keyword, so `DEBUGPULAY` was an obvious alternative.
4. `PL` initiates the output of the value of the largest difference between any two density matrix elements on two consecutive iterations. This is very useful when investigating options for increasing the rate of convergence of the SCF calculation.
5. `SIZES` is very useful in monitoring the sizes and order of creation and annihilation of arrays.

Suggested procedure for locating bugs

Users are supplied with the source code for MOPAC, and, while the original code is fairly bug-free, after it has been modified there is a possibility that bugs may have been introduced. In these circumstances the author of the changes is obviously responsible for removing the offending bug, and the following ideas might prove useful in this context.

First of all, and most important, before any modifications are done a back-up copy of the standard MOPAC should be made. This will prove invaluable in pinpointing deviations from the standard working. This point cannot be over-emphasized—*make a back-up before modifying MOPAC!*

Clearly, a bug can occur almost anywhere, and a logical search sequence is necessary in order to minimize the time taken to locate it.

If possible, perform the debugging with a small molecule, in order to save time (debugging is, of necessity, time consuming) and to minimize output.

The two sets of subroutines in MOPAC, those involved with the electronics and those involved in the geometrics, are kept strictly separate, so the first question to be answered is which set contains the bug. If the heats of formation, derivatives, I.P.s, and charges, etc., are correct, the bug lies in the geometrics; if faulty, in the electronics.

Bugs discovered in the electronics subroutines

Use formaldehyde for this test. Use keywords `1SCF`, `DEBUG`, and any others necessary.

The main steps are:

1. Check the starting one-electron matrix and two-electron integral string, using the keyword `HCORE`. It is normally sufficient to verify that the two hydrogen atoms are equivalent, and that the π system involves only p_z on oxygen and carbon. Note that numerical values are not checked, but only relative values.

If an error is found, use `MOLDAT` to verify the orbital character, etc.

If faulty the error lies in `READMO`, `GETGEO` or `MOLDAT`.

Otherwise the error lies in `HCORE`, `H1ELEC` or `ROTATE`.

If the starting matrices are correct, go on to step (2).

2. Check the density or Fock matrix on every iteration, with the words `FOCK` or `DENSITY`. Check the equivalence of the two hydrogen atoms, and the π system, as in (1).

If an error is found, check the first Fock matrix. If faulty, the bug lies in `ITER`, probably in the Fock subroutines `FOCK1` or `FOCK2`, or in the (guessed) density matrix (`MOLDAT`). An exception is in the UHF closed-shell

calculation, where a small asymmetry is introduced to initiate the separation of the α and β UHF wavefunctions.

If no error is found, check the second Fock matrix. If faulty, the error lies in the density matrix DENSIT, or the diagonalization RSP.

If the Fock matrix is acceptable, check all the Fock matrices. If the error starts in iterations 2 to 4, the error probably lies in CNVG, if after that, in PULAY, if used.

If SCF is achieved, and the heat of formation is faulty, check HELECT. If C.I. was used check MECI.

If the derivatives are faulty, use DCART to verify the Cartesian derivatives. If these are faulty, check DCART and DHC. If they are correct, or not calculated, check the DERIV finite difference calculation. If the wavefunction is non-variationally optimized, check DERNVO.

Bugs discovered in the geometric subroutines

If the geometric calculation is faulty, use FLEPO or PRNT=5 to monitor the optimization, DERIV may also be useful here.

For the FORCE calculation, DCART or DERIV are useful for variationally optimized functions, COMPPG for non-variationally optimized functions.

For reaction paths, verify that FLEPO is working correctly; if so, then PATHS is faulty.

For saddle-point calculations, verify that FLEPO is working correctly; if so, then REACT1 is faulty.

Keep in mind the fact that MOPAC is a large calculation, and, while intended to be versatile, many combinations of options have not been tested. If a bug is found in the original code, please communicate details to Dr. James J. P. Stewart, Stewart Computational Chemistry, 15210 Paddington Circle, Colorado Springs, CO 80921-2512; E-mail: jstewart@fujitsu.com.

Appendix A

Reference Heats of formation

Test MNDO, PM3 and AM1 compounds

In order to verify that MOPAC is working correctly, a large number of tests need to be done. It is obviously impractical to ask users to test all the options in MOPAC. However, users must be able to verify the basic working of MOPAC, and to do this the following tests for the elements have been provided.

Each element can be tested by making up a data-file using estimated geometries and running that file using MOPAC. The optimized geometries should give rise to heats of formation as shown in Table A.1. Any difference greater than 0.1 kcal/mole indicates a serious error in the program.

Caveats

1. Geometry definitions must be correct.
2. Heats of formation may be too high for certain compounds. This is due to a poor starting geometry trapping the system in an excited state. (Affects ICI at times)

Table A.1: Reference Heats of Formation

Element	Test Compound	Heat of Formation				
		MINDO/3	MNDO	AM1	PM3	MNDO- <i>d</i>
Hydrogen	CH ₄	-6.3	-12.0	-8.8	-13.0	
Lithium	LiH		+23.2			
Beryllium	BeO		+38.6		+53.0	
Boron	BF ₃	-270.2	-261.0	-272.1*		
Carbon	CH ₄	-6.3	-12.0	-8.8	-13.0	
Nitrogen	NH ₃	-9.1	-6.4	-7.3	-3.1	
Oxygen	CO ₂	-95.7	-75.1	-79.9	-85.1	
Fluorine	CF ₄	-223.9	-214.2	-225.8	225.1	
Magnesium	MgF ₂				160.7	
Aluminium	AlF		-83.6	-77.9	-50.1	-65.7
Silicon	SiH	+82.9	+90.2	+84.5	+94.6	+89.3
Phosphorus	PH ₃	+2.5	+3.9	+10.2	+0.2	+5.0
Sulfur	H ₂ S	-2.6	+3.8	+1.2	-0.9	-11.7
Chlorine	HCl	-21.1	-15.3	-24.6	-20.5	
Zinc	ZnMe ₂		+19.9	+19.8	8.2	
Gallium	GaCl ₃				-79.7	
Germanium	GeF		-16.4	-19.7	-3.3	
Arsenic	AsH ₃			+3.0	+12.7	
Selenium	SeCl ₂			-42.2	-38.0	
Bromine	HBr		+3.6	-10.5	+5.3	+2.8
Cadmium	CdCl ₂				-48.6	
Indium	InCl ₃				-72.8	
Tin	SnF		-20.4		-17.5	
Antimony	SbCl ₃			-74.8	-72.5	
Tellurium	TeH ₂			+12.6	+23.8	
Iodine	ICl		-6.7	-4.6	+10.8	+0.5
Mercury	HgCl ₂		-36.9	-44.8	-32.7	
Thallium	TlCl				-13.4	
Lead	PbF		-22.6		-21.0	
Bismuth	BiCl ₃				-42.6	

* Not an exhaustive test of AM1 boron.

Appendix B

Bibliography

Elements

MINDO/3

The oldest (1975) of the methods used in MOPAC [97] is MINDO/3 [2], which was parameterized for H, B, C, N, O, F, Si, P, S, and Cl [2]. Some faults were found in the P–O, P–F and P–Cl parameters [51], but these were *not* corrected in MOPAC (see Section 3.4.1).

MNDO

Later, in 1977 MNDO [1] appeared. Initially, parameters were available for C, H, N, and O only [1]. Results were much better [98] than for MINDO/3, and MNDO was soon extended to B [99], F [100], Al [99], Si [101], P [102], S [103], Cl [104], Zn [105], Ge [106], Br [107], Sn [108], I [109], Hg [110], and Pb [111].

AM1

Many of the faults in MNDO were corrected in AM1 [3], which also was initially available only for H, C, N, and O [3]. AM1 has since been extended to B [112], F [113], Al [114], Si [115], P [116], S [117], Cl [113], Zn [118], Ge [119], Br [113], I [113], and Hg [120]. Four unpublished elements available under AM1 are As, Se, Sb, and Te.

PM3

In 1985, PM3 [4] was developed. Initially, 12 elements were available: H, C, N, O, F, Al, Si, P, S, Cl, Br, and I [121]. This set was expanded in 1991 to include Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi [122]. In 1993 lithium was parameterized [123].

MNDO-*d*

When *d* orbitals are added to the basis set [6], the accuracy of the method rises. At present (1998), parameters are available for Al, Si, P, S, Cl, Br, and I [6].

Geometry

By default, geometries are optimized using Baker's EF routine [7]. If this is not desired, then the Broyden [19]-Fletcher [20]-Goldfarb [21]-Shanno [22] method can be used. This replaces the older Davidon [124]-Fletcher [125]-Powell [126] method. The Powell line search in both the BFGS and DFP methods has been upgraded with Thiel's FSTMIN technique [44].

The energy minimum for intersystem crossing can be calculated using the method of Anglada and Bofill [28]. Reaction paths can be followed, but sometimes (unavoidable) numerical instability causes difficulty [127].

Various methods can be used to locate transition states. In the region of a transition state, TS [7] is the best method to use. Other options are NLLSQ [8] and SIGMA [9, 10].

However, getting to this region can often be difficult. One effective strategy is to use the reaction path option. A more costly method is the SADDLE technique [18]. Once the transition state is located, the geometry must be refined

All systems can be characterized by determining the number of imaginary vibrations using FORCE [73]. Ground state systems should have none, and transition states should have exactly one imaginary frequency. The normal modes can then be used to calculate heat capacities and entropies and other thermodynamic quantities [50].

Molecular dynamics can be followed via DRC [29].

Electronics

SCF

Semiempirical methods use approximations to the Roothaan [52]-Hall [53] equations. The four approximate methods are MINDO/3 [2], MNDO [1], AM1 [3], and PM3 [4]. For RHF open shell systems, a Multi-Electron-Configuration-Interaction procedure [39] is available. SCF convergence can be assisted by an inexpensive SHIFT technique [49], by a slightly more complicated Direct Inversion of the Iterative Subspace method (DIIS) [48], or by an expensive but sophisticated interpolation procedure [128].

For large systems, an alternative to conventional methods is to use localized molecular orbitals [129]. This is usually more efficient, particularly for geometry optimization [130].

Gradients

For variationally-optimized wavefunctions, two derivative methods are available. The default is to use finite difference, but under user request (ANALYT), analytical derivatives can be used [16]. When analytical derivatives are used, STO-6G [17] Gaussian functions are used instead of Slater orbitals.

When non-variationally optimized wavefunctions are used, the derivatives are much more complicated [43].

General

Fundamental constants are taken from the CODATA report [71]. A good introduction to MOPAC can be found in Tim Clark's book [131].

Results

The SCF M.O.s, which diagonalize the Fock matrix, can be localized [38] to give M.O.s which can be identified with the conventional picture of two-electron bonds and lone pairs. The localization scheme is faster at the semiempirical level than the Edmiston-Ruedenberg [62] or Boys [61] methods. Associated with each conventional M.O. is a bond-index [23], which represents the contribution to the bond-order matrix due to each M.O. Bond orders and valencies can be displayed by use of BONDS [15]. Other phenomena relating to bonding can also be calculated [24, 25, 26]. An alternative to the normal Coulson density matrix is the Mulliken [41, 42] matrix, invoked by MULLIK.

Dipole moments for ions are coordinate system dependent, so by definition [132], the center of mass is used. Higher terms, e.g., polarizability, and first and second hyperpolarizability, can be calculated [47] by POLAR.

Ionization potentials [59] can be corrected using Green's Functions [34, 35, 36, 33, 37, 32]

Solvent and Electrostatics

Solvent phenomena can be studied. The COSMO technique [30], unlike the self-consistent reaction fields [74],

allows geometries to be optimized. Although the Miertus-Scrocco-Tomasi model [75, 76] cannot optimize geometries, is more sophisticated in that it allows cavitation effects. This model has been modified [77, 78, 79, 80, 81] to allow NDDO methods to be used. In this, optimized VdW radii [83, 84] are used to construct [82] a cavity.

The free energy of hydration is computed as the addition of three contributions:

1. The electrostatic term, which is computed from the linear free energy response theory [75, 76, 77, 78, 79, 80, 81].
2. The cavitation contribution, which is computed from Pierotti's scaled particle theory [85].
3. The van der Waals terms, which is computed using a linear relation with the solute accessible surface, and optimized "hardness" parameters [83, 84].

In addition to the free energy of hydration a "solvent-adapted" wavefunction is obtained. Such a wavefunction can be used to determine changes in solute properties due to the solvent [86, 87, 88, 89].

Electrostatic potentials can be used with the MST method both by deorthogonalizing the wavefunction [90, 91, 40] and by keeping the wavefunction orthogonal [92, 93]

Other ESP methods available are the Merz-Bessler technique [31] and the Ford-Wang procedure [45, 46]. The Ford-Wang is much faster and more accurate than the Merz-Bessler method, but is limited to AM1 calculations on systems containing H, C, N, O, F, and Cl, only.

References

- [1] M. J. S. Dewar and W. Thiel. Ground States of Molecules, 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.*, 99:4899–4907, 1977.
- [2] R. C. Bingham, M. J. S. Dewar, and D. H. Lo. Ground States of Molecules. XXV. MINDO/3. An Improved Version of the MINDO Semiempirical SCF-MO Method. *J. Am. Chem. Soc.*, 97:1285–1293, 1975.
- [3] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. AM1: A New General Purpose Quantum Mechanical Model. *J. Am. Chem. Soc.*, 107:3902–3909, 1985.
- [4] J. J. P. Stewart. Optimization of Parameters for Semi-Empirical Methods I-Method. *J. Comp. Chem.*, 10:209–220, 1989.
- [5] W. Thiel and A. Voityuk. Extension of the MNDO Formalism to d orbitals: Integral Approximations and Preliminary Numerical Results. *Theor. Chim. Acta*, 81:391–404, 1992.
- [6] W. Thiel and A. Voityuk. Extension of MNDO to d-Orbitals - Parameters and Results for the 2nd Row Elements and for the Zinc Group. *J. Phys. Chem.*, 100:616–629, 1996.
- [7] J. Baker. An Algorithm for the Location of Transition States. *J. Comp. Chem.*, 7:385, 1986.
- [8] R. H. Bartels. *Report CNA-44*. University of Texas and Center for Numerical Analysis, 1972.
- [9] A. Komornicki and J. W. Mclver. Rapid Geometry Optimization for Semiempirical Molecular Orbital Methods. *Chem. Phys. Lett*, 10:303, 1971.
- [10] A. Komornicki and J. W. Mclver. Structure of Transition States in Organic Reactions. General Theory and an Application to the Cyclobutene-Butadiene Isomerization using a Semiempirical Molecular Orbital Method. *J. Am. Chem. Soc.*, 94:2625–2633, 1972.
- [11] A. Klamt. On the Calculation of UV/Vis-Spectra in Solution. *J. Phys. Chem.*, 99:2224–2235, 1995.
- [12] F. J. Luque, Y. Zhang, C. Alemn, M. Bachs, J. Gao, and M. Orozco. Solvent Effects in Chloroform Solution: Parameterization of the MST/SCRF Continuum Model. *J. Phys. Chem.*, 100:4269–4276, 1996.
- [13] F. J. Luque, M. Bachs, C Alemn, and M. Orozco. Extension of MST/SCRF Method to Organic Solvents: Ab initio and Semiempirical Parameterization for Neutral Solutes in CCl₄. *J. Comp. Chem.*, 17:806–820, 1996.
- [14] M. J. Frisch, J. S. Binkley, H. B. Schlegel, K. Raghavachari, C. F. Melius, R. L. Martin, J. J. P. Stewart, F. W. Bobrowicz, C. M. Rohlfing, R. L. Kahn, D. J. DeFrees, R. Seeger, R. A. Whiteside, D. J. Fox, E. M. Fleuder, and J. A. Pople. *Gaussian 86*. Carnegie-Mellon Quantum Chemistry Publishing Unit, Pittsburgh PA, 1984.
- [15] D. R. Armstrong, P. G. Perkins, and J. J. P. Stewart. Bond Indices and Valency. *J. Chem. Soc., Dalton*, pages 838–840, 1973.

- [16] M. J. S. Dewar and Y. Yamaguchi. Analytical First Derivatives of the energy in MNDO. *Computers and Chemistry*, 2:25–29, 1978.
- [17] R. F. Stewart. Small Gaussian Expansions of Slater-Type Orbitals. *J. Chem. Phys.*, 52:431–438, 1970.
- [18] M. J. S. Dewar, E. F. Healy, and J. J. P. Stewart. Location of Transition States in Reaction Mechanisms. *J. Chem. Soc., Faraday Trans. II*, 80:227–233, 1984.
- [19] C. G. Broyden. The Convergence of a Class of Double-Rank Minimization Algorithms 2, The New Algorithm. *J. of the Inst. for Math. and Applications*, 6:222–231, 1970.
- [20] R. Fletcher. A New Approach to Variable-Metric Algorithms. *Computer Journal*, 13:317–322, 1970.
- [21] D. Goldfarb. A Family of Variable-Metric Algorithms Derived by Variational Means. *Mathematics of Computation*, 24:23–26, 1970.
- [22] D. F. Shanno. Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24:647–656, 1970.
- [23] M. S. Gopinathen, P. Siddarth, and C. Ravimohan. Valency and Molecular Structure. *Theor. Chim. Acta*, 70:303–322, 1986.
- [24] M. A. Natiello, H. F. Reale, and J. A. Medrano. A Generalized Formalism of the Quantum Theory of Valence and Bonding. *J. Comput. Chem.*, 6:108, 1985.
- [25] J. A. Medrano, H. F. Reale, and R. C. Bochicchio. On the Theory of Valence and Bonding: A Further Generalization. *J. Mol. Struct. (Theochem)*, 135:117, 1986.
- [26] J. A. Medrano and R. C. Bochicchio. The UHF Extension of the Quantum Theory of Valence and Bonding. *J. Mol. Struct. (Theochem)*, 59:463–470, 1989.
- [27] H. F. Reale and R. C. Bochicchio. Spin Populations for Molecular Systems. *J. Phys. B*, 24:2937–2946, 1991.
- [28] Josep M. Anglada and Josep M. Bofill. A Reduced-Restricted-Quasi-Newton-Raphson Method for Locating and Optimizing Energy Crossing Points Between Two Potential Energy Surfaces. *J. Comp. Chem.*, 18:992–1003, 1997.
- [29] K. M. Dieter and J. J. P. Stewart. Calculation of Molecular Frequencies using Molecular Trajectories. *J. Mol. Struct. (Theochem)*, 163:143–149, 1988.
- [30] A. Klamt and G. Schüumann. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin Transactions 2*, pages 799–805, 1993.
- [31] B. H. Besler, Jr. K. M. Merz, and P. A. Kollman. Atomic Charges Derived from Semiempirical Methods. *J. Comp. Chem*, 11:431–439, 1990.
- [32] D. Danovich and Y. Apeloig. Ionization Energies of Triazines and Tetrazines. Application of Green's Function Method Coupling with Semiempirical Molecular Orbital Calculations. *J. Chem. Soc., Perkin Trans. 2*, pages 1865–1873, 1991.
- [33] D. Danovich, V. K. Turchaninov, and V. G. Zakrzewski. OVGf AM1 Calculations on the Ionization Energies of Pyridine Derivatives. *J. Mol. Struct. (Theochem)*, 209:77–87, 1990.

- [34] L.S. Cederbaum and W. Domcke. *Adv. Chem. Phys.*, 36:206, 1977.
- [35] W. von Niessen, J. Schirmer, and L.S. Cederbaum. Computational Methods for the One-Particle Green's Function. *Comput. Phys. Rep.*, 1:57, 1984.
- [36] D. Danovich, V. Zakrzewski, and E. Domnina. Ionization Energies of Azines from Green-Function Method in Semiempirical AM1 Method. *J. Mol. Struct. (Theochem)*, 188:159–166, 1989.
- [37] D. Danovich, V. Zakrzewski, and E. Domnina. AM1 Outer Valence Greens-Function Ionization Energies of the Azoles. *J. Mol. Struct. (Theochem)*, 187:297–306, 1989.
- [38] P. G. Perkins and J. J. P. Stewart. A New Rapid Method for Orbital Localization. *J. Chem. Soc., Faraday II*, 78:285–296, 1982.
- [39] D. R. Armstrong, R. Fortune, and P. G. Perkins. Molecular Orbital Theory for the Excited States of Transition Metal Complexes. *J. Chem. Soc., Faraday II*, 68:1839–1846, 1972.
- [40] F. J. Luque and M. Orozco. Reliability of the AM1 Wavefunction to Compute Molecular Electrostatic Potentials. *Chem. Phys. Lett.*, 168:269–275, 1990.
- [41] R. S. Mulliken. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.*, 23:1833–1841, 1955.
- [42] R. S. Mulliken. Criteria for the Construction of Good Self-Consistent Field Molecular Orbital Wavefunctions and the Significance of LCAO-MO Population Analysis. *J. Chem. Phys.*, 36:3428, 1962.
- [43] M. J. S. Dewar and D. A. Liotard. An Efficient Procedure for Calculating the Molecular Gradient using SCF-CI Semiempirical Wavefunctions with a Limited Number of Configurations. *J. Mol. Struct. (Theochem)*, 206:123–133, 1990.
- [44] W. Thiel. Fast Semiempirical Geometry Optimizations. *J. Mol. Struct.*, 163:415–429, 1988.
- [45] G. P. Ford and B. Wang. New Approach to the Rapid Semiempirical Calculation of Molecular Electrostatic Potentials Based on the AM1 Wave Function: Comparison with Ab Initio HP/6-31G* Results. *J. Comp. Chem.*, 14:1101–1111, 1993.
- [46] B. Wang and G. Ford. Atomic Charges Derived from a Fast and Accurate Method for Electrostatic Potential based on Modified AM1 Calculations. *J. Comp. Chem.*, 15:200–207, 1994.
- [47] H. A. Kurtz, J. J. P. Stewart, and K. M. Dieter. Calculation of the Nonlinear Optical Properties of Molecules. *J. Comp. Chem.*, 11:82–87, 1990.
- [48] P. Pulay. Convergence Acceleration of Iterative Sequences. The Case of SCF Iteration. *Chem. Phys. Lett.*, 73:393–398, 1980.
- [49] A. V. Mitin. The Dynamic Level Shift method for Improving the Convergence of the SCF Procedure. *J. Comp. Chem.*, 9:107–110, 1988.
- [50] M. J. S. Dewar and G. P. Ford. Ground States of Molecules. 44 MINDO/3 Calculations of Absolute Heat Capacities and Entropies of Molecules without Internal Rotations. *J. Am. Chem. Soc.*, 99:7822–7829, 1977.
- [51] G. Frenking, F. Marschner, and H. Goetz. MINDO3-Berechnungen von Phosphororganischen Verbindungen. IV. Parametrisierung und Berechnung von Verbindungen mit P-O, P-F, und P-Cl Bindungen. *Phosphorus and Sulfur*, 8:337–342, 1980.

- [52] C. C. J. Roothaan. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.*, 23:69–89, 1951.
- [53] G. G. Hall. The Molecular Orbital Theory of Chemical Valency VIII: A Method for Calculating Ionization Potentials. *Proc. Royal Soc. London*, A205:541–542, 1951.
- [54] L. Oleari, L. DiSipio, and G. DeMichells. The Evaluation of the One-Centre Integrals in the Semi-Empirical Molecular Orbital Theory. *Mol. Physics*, 10:97–109, 1966.
- [55] M. J. S. Dewar and W. Thiel. A Semiempirical Model for the Two-Center Repulsion Integrals in the NDDO Formalism. *Theor. Chim. Acta*, 46:89–104, 1977.
- [56] M. J. S. Dewar and N. L. Sabelli. *Proc. R. Soc. London, Ser. A*, 264:431, 1961.
- [57] M. J. S. Dewar and N. L. Sabelli. *J. Phys. Chem.*, 66:2310, 1962.
- [58] G. Klopman. *J. Am. Chem. Soc.*, 86:4550, 1964.
- [59] T. Koopmans. Ordering of Wave Functions and Eigenenergies to the Individual Electrons of an Atom. *Physica*, 1:104–113, 1933.
- [60] K. B. Wiberg. Application of the Pople-Santry-Segal CNDO Method to the Cyclopropylcarbanyl and Cyclobutyl Cation and to Bicyclobutane. *Tetrahedron*, 24:1083–1096, 1968.
- [61] S. F. Boys. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.*, 32:296–299, 1960.
- [62] C. Edmiston and K. Ruedenberg. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.*, 35:457–465, 1963.
- [63] W. Von Niessen. Density localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.*, 56:4290, 1972.
- [64] J. S. Griffith. *The Theory of Transition-Metal Ions*. Cambridge University Press, 1961.
- [65] F. L. M. A. H. de Laat. *Unrestricted Hartree-Fock Molecular Orbital Calculations on Transition-Metal Complexes*. Technische Hogeschool Eindhoven, 1968.
- [66] J. C. Slater. Cohesion in Monovalent Metals. *Phys. Rev.*, 35:509–529, 1930.
- [67] J. C. Slater. *Phys. Rev.*, 34:1293, 1959.
- [68] M. Born and T. von Kármán. Vibrations in Space Gratings (Molecular Frequencies). *Z. Physik*, 13:297–309, 1912.
- [69] A. F. Block. *Z. Physik*, 52:555, 1928.
- [70] J. J. P. Stewart. MOSOL, MOPAC for Solid-State Physics. *Quant. Chem. Prog. Exch.*, 5:62–63, 1985.
- [71] E. R. Cohen and B. N. Taylor. *The 1986 Adjustments of the Fundamental Physical Constants*. Pergamon, Elmsford, NY, 1986.
- [72] W. Klyne and V. Prelog. Description of Steric Relationships Across Single Bonds. *Experientia*, 16:521–568, 1960.
- [73] P. Pulay. *Modern Theoretical Chemistry*, volume 4. Plenum, 1977.

- [74] H. S. Rzepa, M. Y. Yi, M. M. Karelson, and M. C. Zerner. Geometry Optimization at the Semiempirical Self-Consistent-Reaction-Field Level using the AMPAC and MOPAC Programs. *J. Chem. Soc. Perkin Trans. II*, 5:635–637, 1991.
- [75] S. Miertus, E. Scrocco, and J. Tomasi. Electrostatic Interaction of a Solute with a Continuum. A Direct Utilization of ab initio Molecular Potentials for the Prevision of Solvent Effects. *Chem. Phys.*, 55:117, 1981.
- [76] S. Miertus and J. Tomasi. Approximate Evaluations of the Electrostatic Free Energy and Internal Energy Changes in Solution Processes. *Chem. Phys.*, 65:239, 1982.
- [77] F. J. Luque, M. J. Negre, and M. Orozco. An AM1-SCRF Approach to the Study of Changes in Molecular Properties Induced by Solvent. *J. Phys. Chem.*, 97:4386–4391, 1993.
- [78] J. Gao, F. J. Luque, and M. Orozco. Induced Dipole Moment and Atomic Charges based on Average Electrostatic Potentials in Aqueous Solution. *J. Chem. Phys.*, 94:2975–2982, 1993.
- [79] M. Negre, M. Orozco, and F. J. Luque. A New Strategy for the Representation of Environment Effects in Semi-Empirical Calculations based on Dewars Hamiltonians. *Chem. Phys. Lett.*, 196:27–36, 1992.
- [80] F. J. Luque, M. Bachs, and M. Orozco. An Optimized AM1-MST Method for the SCRF Representation of Solvated Systems. *J. Comp. Chem.*, 15:847–857, 1994.
- [81] M. Orozco, M. Bachs, and F. J. Luque. Development of Optimized MST/SCRF Methods for Semiempirical Calculations. The MNDO and PM3 Hamiltonians. *J. Comp. Chem.*, 16:563–585, 1995.
- [82] J. L. Pascual-Ahuir, E. Silla, J. Tomasi, and R. Bonaccorsi. Electrostatic Interactions of a Solute with a Continuum. Improved Description of the Cavity and of the Surface Cavity Bound Charge Distribution. *J. Comp. Chem.*, 8:778, 1987.
- [83] M. Orozco, W. L. Jorgensen, and F. J. Luque. Comparison of 6-31G* based MST/SCRF and FEP Evaluations of the Free-Energies of Hydration for Small Neutral Molecules. *J. Comp. Chem.*, 14:1498–1503, 1993.
- [84] M. Bachs, F. J. Luque, and M. Orozco. Optimization of Solute Cavities and van der Waals parameters in ab initio MST-SCRF Calculations of Neutral Molecules. *J. Comp. Chem.*, 14:446–454, 1994.
- [85] R. A. Pierotti. A Scaled Particle Theory of Aqueous and Nonaqueous Solutions. *Chem. Rev.*, 76:717, 1976.
- [86] F. J. Luque, M. Orozco, P. K. Bhadane, and S. R. Gadre. SCRF Calculation of the Effect of Hydration on the Topology of the Molecular Electrostatic Potential. *J. Phys. Chem.*, 97:9380–9384, 1993.
- [87] F. J. Luque and M. Orozco. Theoretical Study of N-methylacetamide in Vacuum and Aqueous Solution - Implications for the Peptide Bond Isomerization. *J. Org. Chem.*, 58:6397–6405, 1993.
- [88] F. J. Luque and M. Orozco. Reactivity of Planar and Twisted Amides in Vacuum and Aqueous Environments: an ab initio Study. *J. Chem. Soc., Perkin Trans. II*, pages 683–690, 1993.
- [89] F. J. Luque, M. Orozco, P. K. Bhadane, and S. R. Gadre. Effect of Solvation on the Shapes, Sizes and Anisotropies of Polyatomic Anions via Molecular Electrostatic Potential Topography: an ab initio Reaction Field Approach. *J. Chem. Phys.*, 100:6718–6726, 1994.
- [90] F. J. Luque, F. Illas, and M. Orozco. Comparative Study of the Molecular Electrostatic Potential Obtained from Different Wave- functions - Reliability of the Semiempirical MNDO Wavefunction. *J. Comp. Chem.*, 11:416–430, 1990.

- [91] C. Alemán, F. J. Luque, and M. Orozco. Suitability of the PM3-Derived Molecular Electrostatic Potentials. *J. Comp. Chem.*, 14:799–808, 1993.
- [92] G. G. Ferenczy, C. A. Reynolds, and W. G. Richards. Semiempirical AM1 Electrostatic Potentials and AM1 Electrostatic Potential Derived Charges - A Comparison with ab initio Values. *J. Comp. Chem.*, 11:159–169, 1990.
- [93] C. Alhambra, F. J. Luque, and M. Orozco. Comparison of NDDO and Quasi-ab initio Approaches to Compute Semiempirical Molecular Electrostatic Potentials. *J. Comp. Chem.*, 15:12–22, 1994.
- [94] M. Orozco and F. J. Luque. Optimization of the Cavity Size for ab initio MST-SCRF Calculations of Monovalent Ions. *Chem. Phys.*, 182:237–248, 1994.
- [95] G. Klebe. The Use of Crystal Data Together with Other Experimental and Computational Results to Discuss Structure-Reactivity and Activity Relationships. *Struct. Chem.*, 1:597–616, 1990.
- [96] J. L. Ozment and A. M. Schmiedekamp. Proton Affinities of Molecules containing Nitrogen and Oxygen: Comparing Ab Initio and Semi-Empirical methods to Experiments. *Int. J. Quant. Chem.*, 43:783–800, 1992.
- [97] J. J. P. Stewart. MOPAC: A General Molecular Orbital Package. *Quant. Chem. Prog. Exch.*, 10:86, 1990.
- [98] M. J. S. Dewar and W. Thiel. Ground States of Molecules, 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen, and Oxygen. *J. Am. Chem. Soc.*, 99:4907–4917, 1977.
- [99] L. P. Davis, R. M. Guidry, J. R. Williams, M. J. S. Dewar, and H. S. Rzepa. MNDO Calculations for Molecules Containing Aluminum and Boron. *J. Comp. Chem.*, 2:433–445, 1981.
- [100] M. J. S. Dewar and H. S. Rzepa. Ground States of Molecules. 40. MNDO Results for Molecules Containing Fluorine. *J. Am. Chem. Soc.*, 100:58–67, 1978.
- [101] M. J. S. Dewar, J. Friedheim, G. Grady, E. F. Healy, and J. J. P. Stewart. Revised MNDO Parameters for Silicon. *Organometallics*, 5:375–379, 1986.
- [102] M. J. S. Dewar, M. L. McKee, and H. S. Rzepa. MNDO Parameters for Third Period Elements. *J. Am. Chem. Soc.*, 100:3607–3607, 1978.
- [103] M. J. S. Dewar and C. H. Reynolds. An Improved Set of MNDO Parameters for Sulfur. *J. Comp. Chem.*, 7:140–143, 1986.
- [104] M. J. S. Dewar and H. S. Rzepa. Ground States of Molecules. 53. MNDO Calculations for Molecules Containing Chlorine. *J. Comp. Chem.*, 4:158–169, 1983.
- [105] M. J. S. Dewar and K. M. Merz. MNDO Calculations for Compounds Containing Zinc. *Organometallics*, 5:1494–1496, 1986.
- [106] M. J. S. Dewar, G. L. Grady, and E. F. Healy. MNDO Calculations for Compounds Containing Germanium. *Organometallics*, 6:186–189, 1987.
- [107] M. J. S. Dewar and E. F. Healy. Ground States of Molecules. 64. MNDO Calculations for Compounds Containing Bromine. *J. Comp. Chem.*, 4:542–551, 1983.
- [108] M. J. S. Dewar, G. L. Grady, and J. J. P. Stewart. MNDO Calculations for Compounds Containing Tin. *J. Am. Chem. Soc.*, 106:6771–6773, 1984.

- [109] M. J. S. Dewar, E. F. Healy, and J. J. P. Stewart. Ground States of Molecules. 67. MNDO Calculations for Compounds Containing Iodine. *J. Comp. Chem.*, 5:358–362, 1984.
- [110] M. J. S. Dewar, G. L. Gready, K. M. Merz, and J. J. P. Stewart. MNDO Calculations for Compounds Containing Mercury. *Organometallics*, 4:1964–1966, 1985.
- [111] M. J. S. Dewar, M. K. Holloway, G. L. Grady, and J. J. P. Stewart. MNDO Calculations for Compounds Containing Lead. *Organometallics*, 4:1973–1980, 1985.
- [112] M. J. S. Dewar, C. Jie, and E. G. Zoebisch. AM1 Calculations for Compounds Containing Boron. *Organometallics*, 7:513–521, 1988.
- [113] M. J. S. Dewar and E. G. Zoebisch. Extension of AM1 to the Halogens. *J. Mol. Struct. (Theochem)*, 180:1–21, 1988.
- [114] M. J. S. Dewar and A. J. Holder. AM1 Parameters for Aluminum. *Organometallics*, 9:508–511, 1990.
- [115] M. J. S. Dewar and C. Jie. AM1 Calculations for Compounds Containing Silicon. *Organometallics*, 6:1486–1490, 1987.
- [116] M. J. S. Dewar and C. Jie. AM1 Parameters for Phosphorus. *J. Mol. Struct. (Theochem)*, 187:1–13, 1989.
- [117] M. J. S. Dewar and Y-C Yuan. AM1 parameters for Sulfur. *Inorganic Chemistry*, 29:3881–3890, 1990.
- [118] M. J. S. Dewar and K. M. Merz. AM1 Parameters for Zinc. *Organometallics*, 7:522–524, 1988.
- [119] M. J. S. Dewar and C. Jie. AM1 Calculations for Compounds Containing Germanium. *Organometallics*, 8:1544–1547, 1989.
- [120] M. J. S. Dewar and C. Jie. AM1 Calculations for Compounds Containing Mercury. *Organometallics*, 8:1547–1549, 1989.
- [121] J. J. P. Stewart. Optimization of Parameters for Semiempirical Methods II. Applications. *J. Comp. Chem.*, 10:221–264, 1989.
- [122] J. J. P. Stewart. Optimization of Parameters for Semi-Empirical Methods III- Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi. *J. Comp. Chem.*, 12:320–341, 1991.
- [123] E. Anders, R. Koch, and P. Freunsch. Optimization and Application of Lithium Parameters for PM3. *J. Comp. Chem.*, 14:1301–1312, 1993.
- [124] W. C. Davidon. Variance algorithm for minimization. *Comput. J.*, 10:406–410, 1968.
- [125] R. Fletcher. Function minimization without evaluating derivatives - a review. *Comput. J.*, 8:33–41, 1965.
- [126] R. Fletcher and M. J. D. Powell. A Rapidly Convergent Descent Method for Minimization. *Comput. J.*, 6:163–168, 1963.
- [127] D. B. Boyd, D. W. Smith, J. J. P. Stewart, and E. Wimmer. Numerical Sensitivity of Trajectories across Conformational Energy Hypersurfaces from Geometry Optimized Molecular Orbital Calculations: AM1, MNDO, and MINDO/3. *J. Comp. Chem.*, 9:387–398, 1988.
- [128] R. N. Camp and H. F. King. An Interpolation Method for Forcing SCF Convergence. *J. Chem. Phys.*, 75:268–274, 1981.

-
- [129] J. J. P. Stewart. Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations. *Int. J. Quant. Chem.*, 58:133–146, 1996.
- [130] J. J. P. Stewart. Calculation of the Geometry of a Small Protein using Semiempirical Methods. *J. Mol. Struct. (Theochem)*, pages 195–205, 1997.
- [131] T. Clark. *A Handbook of Computational Chemistry*. Wiley, 1985.
- [132] A. D. Buckingham. Molecular Quadrupole Moments. *Quarterly Reviews*, pages 183–214.

Index

- * (in data sets), 19
- , 127
- , 127
- ΔH_f , 1, **221**
 - definition, 115
 - average errors in, 252
 - criteria, 378
 - effect of SYMMETRY on, 76
 - molecular standards, 385
 - precision, 207
 - worked example, 160
- +
 - keyword, 30
 - sparkle, 127
- ++, 127
- &, 29
- “T” - Optimization flag, 235
- 1SCF SPECIFIED WITH ... , 353
- 3-J symbol, 142

- A SINGLE ATOM HAS ... , 353
- A SYMMETRY FUNCTION ... , 353
- Ab initio
 - C_v , 218
- ab initio*, 1
 - band structures, 196
 - derivatives, 122
 - geometry, 123
 - total energies, 218
- Abelian groups, 203
- Accuracy of methods, 251
- Active space
 - in C.I., **141**, 205
 - increasing the size of, 35
- AD, 111
- AIDER, 31
- AIGIN, 31
- AIGOUT, 31
- ALL CONVERGERS ARE ... , 353
- ALLBONDS, 31

- ALLVEC, 31
- ALT_A=A, 31
- ALT_R=A, 32
- AM, 111
- AM1, 1, **32**, 357
 - core-core terms, 114
 - difference from MNDO, 114
 - accuracy, 253
 - accuracy of, 251
 - elements in, 86
- Amino acids
 - symbols for, 70
- AN ATTEMPT WAS MADE ... , 353
- AN UNOPTIMIZABLE GEO ... , 353
- ANALYT, **32**, 210
- ANALYTIC C.I. DERIV ... , 353
- Analytical derivatives, 210
- ANALYTICAL DERIVATIVES ... , 353
- Angular momentum, 207
- Aniline, dimer, 210
- Applied electric fields, 239
- AQ, 111
- Aromatic rings, 163
- Array specification, 344
- Asterisks
 - in data sets, 19
- AT LEAST ONE ATOM ... , 354
- AT LEAST ONE ATOM HAS ... , 354
- At this point, both reactants ... , 354
- ATHEAT, 137
- Atom *nn* is Cartesian ... , 354
- Atom *nn* is internal ... , 354
- ATOM NUMBER *nn* IS ILL ... , 354
- Atomic
 - charge, from ESP, 42
 - electron density dist., 110
 - masses, 89
- ATOMIC MASS OF ... , 354
- ATOMIC NUMBER *nn* IS ... , 354

- ATOMIC NUMBER OF *NN* ?, 355
- Atoms
- making fully ionic, 83
- ATOMS *nm* AND *nn* ARE SEP ... , 355
- ATTEMPT TO GO DOWN ... , 355
- Aufbau principle, 141
- AUTOSYM, 32, 201
- AVAILABLE MEMORY IN ... , 355
- Band structure
- electronic, 195
 - geometry requirements, 195
- BAR, 32, 355
- Bartel's method, **59**, 355
- BCC, 196
- use of, 100
- Benzene, 53
- Besler, B. H.**, 42
- BFGS, 33
- BIGCYCLES, 33
- BIGSCF, 33
- BIRADICAL, **33**, 125
- Biradicaloid character, **33**, 125
- Bischof, Peter**, 201
- BLOCK data, 39, **156**
- Boltzmann constant, 208
- Bond order, 116
- BONDS, 34
- Bonds
- capped, 88
- Born-von Kármán
- periodic boundary conditions, 194
- BOTH GEOMETRIES ARE ... , 355
- BOTH SYSTEMS ARE ON ... , 355
- Boyd, Donald B.**, 211
- Buckminsterfullerene, 203
- Bugs
- how to report, 4
 - locating, 383
- Bulk modulus, 63
- C_p , 218
- C.I., 35
- description, 138
- C.I. IS OF SIZE LESS THAN ... , 355
- C.I. NOT ALLOWED WITH UHF, 355
- CALCULATION ABANDONED ... , 356
- Camp, 36
- Camp-King converger, 125
- Cannot open *filename.out*, 356
- Cannot write density ... , 356
- Canonical M.O.s, 41
- Capped bonds, 88, **128**
- CARTESIAN CALCULATION ... , 356
- Cb, 88, 128
- CHARGE, 36
- determining using MOZ, 36
 - use of, 99
- CHARGE ON ATOM *N* ... , 356
- Charged sites in proteins, 176
- CHEM3D, 92
- Chirality, 211
- CIS, 36
- CISD, 36
- CISDT, 36
- Clebsch-Gordon 3-J symbol, 142
- Cluster model, 194
- CNDO, 106
- CNDO/2, 156
- CNVG, **125**, 342
- Code
- contributed, 3
- COMPFG, 36
- use of LARGE, 52
- Compounds
- 'normal'
 - definition, 162
- Computer speed
- comparing, 33
- Configuration interaction, 33, 34, 40, 42, 53, 59–61, 67, 68, 72, 73, 82, 138–155
- Connectivity of atom *NN* ... , 356
- Constants, 43
- physical, 208
- Construction of acyclic π bonds, 175
- Construction of cyclic π bonds, 176
- Construction of hybrid orbitals, 171
- Construction of lone pairs, 175
- Construction of sigma orbitals, 174
- Construction of starting LMOs, 170
- Contributed code, 3
- criteria for, 3
- Contributors to MOPAC, 17

- Conventional SCF
 N^3 dependence, 160
- Convergence tests
 SCF, 211
- Cooling systems, 41
- Coordinates, 21
 Cartesian, 52, 90, **92**
 suppressing, 61
 conversion between formats, 93
 dynamic reaction, 47
 fractional unit cell, 197
 Gaussian, 91
 example, **92**, 122
 in FORCE calcn., 44
 internal, 52
 internal to Cartesian, 90
 intrinsic reaction, 51
 normal, 205
 prolate spheroidal, 130
 reaction, 226
 use of SYMMETRY, 76
- COORDINATES MUST BE ... , 356
- Core-core repulsion, 113
- COSMO, 42, **241**
- Coulomb integrals, 194
- Cpu time
 specification, 80
- Crambin, 93
- Criteria, 207
 for contributed code, 3
 SCF, 377
- CROSS, 37
- CROSS requires ROOT=2 ... , 356
- Crossing, bands in solid state, 195
- Cubic Groups, 202
- CUTOF1 WAS SET SMALL ... , 356
- CUTOF1=*n.nn*, 37
- CUTOF2=*n.nn*, 37
- Cutoff, **163**
- CUTOFP=*n.nn*, 37
- CUTOFS=*n.nn*, 37
- CVB(*text*), 38
- CYCLE, 187
- CYCLES, 38
- Cyclobutadiene
 need for C.I., 138
- Cyclohexane, 123
- D_{8d} missing, 201
- Danovich, David**, 201
- Data
 for ethyl radical (SADDLE calcn.), 236
 for formic acid, 94
 commas in, 85
 comments in, 19
- Diamond
 construction of, 198
 examples, 97
 files, 19
 for cyclohexane, 123
 for diamond, **100**
 for ethane, 92
 for ethene, 19
 for polytetrahydrofuran, 100
 free-format, 85
 overview of, 21
 tabs in, 85
- DATA ARE NOT AVAIL ... , 356
- Data for TOMASI Model miss ... , 357
- DC-EFISH, 240
- DCART, **38**, 212
 use of LARGE, 52
- DDMAX, 38
- DDMIN, 38
- DEBUG, 39
- Definitions
 Boltzmann constant, 208
 velocity of light, 208
- Degenerate groups, 203
- DEGENERATE LEVELS ... , 357
- Degenerate M.O.s, **140**, 205
- DELHOF, 377
- Delocalizability, 76
- Delocalized π systems, 163, 196
- DENOUT, 39
- DENSITY, **39**
- DENSITY FILE MISSING ... , 357
- Density matrix
 decomposition into σ , π , and δ , 64
 iterating, 159
 reconstruction in MECI, 146
 starting, 157

- Density of states
 - electronic, 195
- Density Restart File missing . . . , 357
- DEP, **39**, 43
- DERIV, 341
- Derivatives, 211–213, *see also* Gradient
 - “double-sided”, 213
 - “single-sided”, 213
 - ab initio*, 122
 - analytical, **210**
- DERNVO
 - use of LARGE, 52
- Dewar’s half-electron method, 34
- DFP, 39
- DIAGG2, 39
- Diagonalization, 180
- Diamond, 198
 - data for, 100
- Dielectric energy, 244
- Dihedral angles
 - ambiguities in, **238**
 - coherency, 211
- DIIS, 125
- DIPOLE, 40
- Dipole, 110
 - induced, 239
 - made from sparkles, 128
 - transition, 146
- DIPOLE CONSTRAINTS . . . , 357
- Dipole moment
 - definition, 116
 - for ions, 116
- DIPX, 40
- DIPY, 40
- DIPZ, 40
- DISEX, 40
- DMAX, 40
- DOUBLET, 40
- DRAW, 238
- DRC, 41, 47, 80, 226–236, **238**
 - conservation of momentum, 228
 - definition, 226
 - dummy atoms in, 229
 - keyword options, 231
 - use of KINETIC, 52
 - use of LARGE, 52
 - use of VELOCITY, 83
 - use of X-PRIORITY, 84
- DUE TO A PROGRAM . . . , 357
- Dummy atoms, 85, 88
 - in DRC, 229
- DUMP, 39, **41**
- Dynamic reaction coordinates, 47
- ECHO, 41
- ECHO is not allowed at this point, 357
- Education, 4
- EF, 40, **41**, 68
 - criterion, 377
 - effect of LET on, 53
 - use of, 102
- Effective mass, 217
- EFISH, 66
- EHEAT, 137
- EIGEN, 41
- Eigenvalues, 158
- Eigenvector
 - following, 187
- EIGENVECTOR FOLLOWING . . . , 357
- Eigenvectors, **158**
 - from MOZYME, 41
 - printing, 83
- EISOL, 137
- EITHER ADD ‘LET’ OR . . . , 357
- Electric fields
 - applied, 239
- Electro-optic Pockels effect, 240
- Electron
 - effective mass, 196
- 1ELECTRON, 30
- Electronic density of states, 195
- Electronic energy, 115
 - of atoms, 115
 - of microstates, 143
- Electrooptic effects, 66
- Electrophilic Delocalizability, 76
- Electrostatic Potential
 - how to calculate, 65
 - PMEP model, 245
- Electrostatic potential, 42
 - Orozco-Luque model, 247
- ELEMENT NOT FOUND, 357

- ELEMENT NOT RECOGNIZED, 357
- Elements
- making fully ionic, 83
 - specification of, 86
- Enantiomers, 211
- Energy, 218
- Energy considerations, 162
- ENPART, 42
- Entropy, 80, 218–223
- EPS, 42
- Equipartition of energy, 219
- Error
- messages, 353–375
 - trapping, 30
- ERROR TOO MANY ... , 357
- ERROR DETECTED DURING ... , 357
- ERROR DETECTED IN SUB ... , 358
- ERROR DURING READ AT ... , 358
- ERROR IN READ OF FIRST ... , 358
- Error message
- 1SCF SPECIFIED WITH ... , 353
 - A SINGLE ATOM HAS ... , 353
 - A SYMMETRY FUN ... , 353
 - AN UNOPTIMIZABLE ... , 353
 - ANALYTIC C.I. DERIV ... , 353
 - ANALYTICAL DERIV ... , 353
 - AT LEAST ONE ATOM ... , 354
 - Atom *nn* is Cartesian ... , 354
 - Atom *nn* is internal ... , 354
 - ATOM NUMBER ... , 354
 - ATOMIC MASS OF ATOM ... , 354
 - ATOMIC NUMBER ... , 354, 355
 - ATOMS *nn* AND *nn* ... , 355
 - ATTEMPTS TO GO ... , 355
 - AVAILABLE MEMORY ... , 355
 - BOTH GEOMETRIES ARE ... , 355
 - BOTH SYSTEMS ARE ... , 355
 - C.I. IS OF SIZE LESS ... , 355
 - C.I. NOT ALLOWED ... , 355
 - CALCULATION ABAND ... , 356
 - Cannot open *filename.out*, 356
 - Cannot write density matrix ... , 356
 - CARTESIAN CALC ... , 356
 - CHARGE ON ATOM *N* ... , 356
 - Connectivity of atom *NN* ... , 356
 - COORDINATES MUST ... , 356
 - CROSS requires ROOT=2 ... , 356
 - DATA ARE NOT AVAIL ... , 356
 - Data for TOMASI Model ... , 357
 - DEGENERATE LEVELS ... , 357
 - DENSITY FILE MISSING ... , 357
 - Density Restart File missing ... , 357
 - DUE TO A PROGRAM ... , 357
 - ECHO is not allowed ... , 357
 - EIGENVECTOR FOLLOW ... , 357
 - EITHER ADD 'LET' ... , 357
 - ELEMENT NOT FOUND, 357
 - ELEMENT NOT REC ... , 357
 - ERROR DETECTED ... , 357
 - ERROR DETECTED IN ... , 358
 - ERROR DURING READ ... , 358
 - Error in BLAS, 358
 - Error in EF, 358
 - Error in GETGEO, 358
 - Error in GETMEM, 358
 - Error in GREENF, 358
 - Error in MOLDAT., 358
 - Error in PATHS, 358
 - Error in PMP, 358
 - ERROR IN READ OF ... , 358
 - Error in READMO, 358
 - Error in TOM, 358
 - Error in UPDATE, 358
 - Error in USAGE, 358
 - ERROR TOO MANY ... , 357
 - ERROR: PERM ... , 359
 - ERRORS DETECTED ... , 359
 - EXCITED USED ... , 359
 - EXTERNAL PARA ... , 359
 - FAILED IN SEARCH ... , 359
 - FAILED TO ACHIEVE SCF, 359
 - Fatal error in try ... , 359
 - FAULT DETECTED IN INT ... , 359
 - FAULT IN READ OF AB ... , 360
 - FAULTY LINE: *text*, 360
 - FILE *file.den* is missing, 360
 - First atom must not be H ... , 360
 - GAUSSIAN INPUT REQ ... , 360
 - Geometry cannot be ... , 360
 - GEOMETRY CONTAINS ... , 360
 - Geometry in PLATO ... , 360
 - Geometry is apparently cubic ... , 360

- GEOMETRY IS FAULTY, 360
GEOMETRY TOO UN ... , 360
GRADIENT IS TOO LA ... , 361
GRADIENT IS VERY ... , 361
GREENS FUNCTION IS ... , 361
Hydrogen atom N is ... , 361
Hydrogen atoms must be ... , 361
HYDROGENS MUST BE ... , 361
If QPMEP is used, ... , 361
ILLEGAL ATOMIC ... , 361
IMPOSSIBLE NUMBER ... , 361, 362
IMPOSSIBLE OPTION ... , 362
IMPOSSIBLE VALUE OF ... , 362
INPUT FILE MISSING OR ... , 362
INSUFFICIENT DATA ON ... , 362
JOB STOPPED BY ... , 362
KEYWORD AIDER SPECIF ... , 362
KEYWORD ANALYT ... , 362
LINE OF KEYWORDS ... , 362
MAX. NUMBER OF AT ... , 362
MAXIMUM ALLOWED ... , 362
MICROSTATES SPEC ... , 363
MISSING VAN DER ... , 363
MIXED PARAMETER ... , 363
MODIFY SUBROUTINE ... , 363
MORE THAN 3N-6 COORD ... , 363
MORE THAN ONE GEO ... , 363
MULLIKEN POPULATION ... , 363
NAME NOT FOUND, 363
NEGATIVE SYMBOLICS ... , 364
NEWGEO CAN ONLY ... , 364
NLLSQ USED WITH ... , 364
NO ATOMS IN SYSTEM, 364
NO DUMMY ATOMS ... , 364
NO PEPTIDE LINK ... , 364
NO POINTS SUPPLIED ... , 364
NO RESTART FILE EXISTS, 364
NOANCI MUST BE USED ... , 364
NONET SPECIFIED ... , 364
NUMBER OF DOUBLY ... , 364
NUMBER OF ELEC ... , 364
NUMBER OF M.O.s in ... , 365
NUMBER OF M.O.s REQ ... , 365
NUMBER OF OCCUPIED ... , 365
NUMBER OF OPEN-S ... , 365
NUMBER OF PART ... , 365
NUMBER OF PERM ... , 365
NUMERICAL PROBLEMS ... , 365
OCCUPIED C VECTOR ... , 365
OCCUPIED IC VECTOR ... , 365
OCTET SPECIFIED ... , 366
OLDENS FILE ... , 366
OLDGEO used and ... , 366
OMIN MUST BE ... , 366
ONLY N POINTS ... , 366
ONLY DIHEDRAL SYM ... , 366
Only H, C, N, O, F, ... , 366
ONLY ONE REACTION ... , 366
PARAMETERS FOR ... , 366
PERMANENT INTEGER ... , 367
PERMANENT REAL ARRAY, 367
POLAR does not work ... , 367
PROBLEM IN SYMR, 367
QUARTET SPECIFIED ... , 367
QUINTET SPECIFIED ... , 367
Ran out of storage ... , 367
RESTART FILE EXISTS, ... , 368
Restart file is corrupt, 368
ROOT REQUESTED ... , 368
RSOLV MUST NOT BE ... , 368
SEPTET SPECIFIED ... , 368
SETUP FILE MISSING, ... , 368
SEXTET SPECIFIED ... , 368
SIGMA USED ... , 368
SINGLE ATOM HAS ... , 353
Size of active ... , 368
SOME ELEMENTS HAVE ... , 368
Something disaster ... , 369
SPECIFIED SPIN ... , 369
Storage needed for ... , 369
SYMMETRY FUNCTION ... , 353
SYMMETRY SPECIFIED ... , 369
SYSTEM DOES NOT ... , 369
SYSTEM SPECIFIED ... , 369
SYSTEMS WITH TV ... , 370
TEMPERATURE ... , 370
The COSMO option ... , 370
The data set contains ... , 370
THE FOLLOWING SYM ... , 370
THE GEOMETRY DATA- ... , 370
THE GREENS FUNCT ... , 370
THE HAMILTONIAN ... , 370

- The input data file ... , 371
THE SCF CALCULATION ... , 371
THE STATE REQUIRED ... , 371
There are more ... , 371
THERE ARE NO VAR ... , 371
There are too ... , 371
There is a bug in MOZYME, 371
THERE IS A FAULT ... , 371
There is a mistake ... , 372
THERE IS A RISK ... , 372
THERE MUST BE ... , 372
THIS MESSAGE ... , 372
THREE ATOMS BEING ... , 372
TIME UP, 372
TOO MANY CONFIG ... , 372
TOO MANY ITER ... , 372
TRIPLET SPECIFIED ... , 372
TS FAILED ... , 373
TS SEARCH AND ... , 373
TS SEARCH REQUIRE ... , 373
TWO ADJACENT POINTS ... , 373
TWO ATOMS ARE ... , 373
UNABLE TO ACHIEVE ... , 373
UNACCEPTABLE VALUE ... , 373
UNDEFINED SYM ... , 373
UNKNOWN RESIDUE: *N*, 373
UNOPTIMIZABLE ... , 353
UNRECOGNIZED ELE ... , 373
UNRECOGNIZED HESS ... , 373
UNRECOGNIZED KEY ... , 373
UPPER BOUND OF ACT ... , 374
Use AIGIN to allow more ... , 374
USE EITHER SAFE ... , 374
VALUE OF NPPA NOT ... , 374
VAN DER WAALS, 374
VIRTUAL C VECTOR ... , 374
VIRTUAL IC VECTOR ... , 374
WARNING, 375
WARNING – GEOM ... , 374
WARNING – GEOM ... , 209
ERROR: PERMANENT ARRAY, 359
ERRORS DETECTED IN CON ... , 359
ESP, 40, **42**
ESPRST, 42
ESR, 42
Ethane, 92
barrier in, 226
Ethyl radical, 236
Ethylene, twisted
 need for C.I., 138
Euler matrices, 202
Examples
 of data-sets, 97
Excimers, 154
EXCITED, 42
Excited electronic state, 99
EXCITED USED WITH TRIPLET, 359
Exponents
 transition dipole, 146
EXTERNAL, 39, **43**
External
 PM3(tm), 86
EXTERNAL PARAMETERS ... , 359
FAILED TO ACHIEVE SCF, 359
Fatal error in trying to open ... , 359
FAULT DETECTED IN INT ... , 359
FAULT IN READ OF ... , 360
FAULTY LINE: *text*, 360
Fd3m, 197
FIELD, 43
FILE *file.den* is missing, 360
FILL, 44
First atom must not be H ... , 360
Flat potential surfaces, 210
FLEPO, 44
Fluorescence, 153
FMAT, 44
 use of LARGE, 52
Fock matrix, 124
 iterating, 159
 starting, 157
 symmetrization in solid state calcs., 197
FORCE, 44
 effect of LET on, 52
 use of, 99
 use of LARGE, 52
 use of PRECISE, 67
 use with IRC, 228
 use with MODE, 57
Force constant, 1, **44**, 233
Force constants, 223

- internal, 44
- Ford, George**, 245
- Formaldehyde, data set, 99
- Formic acid, 94
- Fractional
 - M.O. occupancy, **62**, 140
 - unit cell coordinates, 197
- Frame
 - description of, 215
- Franck-Condon, 153
- Frequencies
 - imaginary, 238
 - isotopic substitution, 51
- Frequency-dependent NLO, 240
- Fundamental Physical Constants, 208
- Γ point
 - in solid state, 194
- Gas constant, R, 208
- Gaussian
 - coordinates, 91
 - orbitals, 129
 - to MOPAC format, 30
 - wavefunctions, 210
- GAUSSIAN INPUT REQ . . . , 360
- GEO-OK, 45
- GEOCHK, 45
- Geometry
 - BFGS optimizer, 187
 - criterion
 - modifying, 45
 - flags for, 92
 - specification, 85
 - using old, 61
- Geometry cannot be optimized . . . , 360
- GEOMETRY CONTAINS . . . , 360
- Geometry in PLATO . . . , 360
- Geometry is apparently cubic . . . , 360
- GEOMETRY IS FAULTY, 360
- GEOMETRY TOO UNSTABLE . . . , 360
- Gibbs free energy, 221
- GNMIN, 45
- GNORM, 45, 67, **209**
 - effect of LET on, 53
 - use with IRC/DRC, 229
- GNORM HAS BEEN SET . . . , 361
- Gordon, Mark**, 226
- G_{p2} , 109
- G_{pp} , 109
- Gradient, *see also* Derivatives
 - minimization
 - using NLLSQ, 59
 - using SIGMA, 73
 - using TS, 82
- GRADIENT IS TOO LARGE . . . , 361
- GRADIENT IS VERY LARGE . . . , 361
- GRADIENTS, **45**, 191
- Gradients, **211**
- GRAPH, 46
- GREENF, 46
- GREENS FUNCTION IS NOT . . . , 361
- Grid Calculation, **226**
- Grid map, 89
 - use of MAX, 53
 - use of POINT n , 65
 - use of STEP n , 76
- Groups, 201–207
- GSE , used in MECI, 144
- G_{sp} , 109
- G_{ss} , 109
- H-PRIORITY, **47**, 235
- H2O , 47
- Half-electron, 138
- Half-life, 41
- Hardness, 76
- Harmonic oscillator, 218
- Harmonic, third, generation, 240
- Heat capacity, 80, **218**
- Heat of atomization, 115
- Heat of Formation, 1, **221**
 - definition, 115
 - average errors in, 252
 - criteria, 378
 - effect of SYMMETRY on, 76
 - molecular standards, 385
 - precision, 207
 - worked example, 160
- Heating systems, 41
- HESS, 48
- Hessian, 44, 68, 73, 187, 213
 - in BFGS optimizer, 189

- Hilltop, escaping from, 238
Hirano, Tsuneo, 218
HOMO, 42
Hook's law calculation, 225
 H_{sp} , 109
HYBRID, 50
Hybrid orbitals
 construction of, 171
Hybridization, 64
Hydrogen atom N is ... , 361
Hydrogen molecule
 use of GEO-OK, 45
HYDROGENS MUST BE ... , 361
HYPERFINE, 50
Hyperpolarizability, 66, 239
Hypervalent systems
 allowed, 163

Identification of ions, 176
If QPMEP is used, then ... , 361
ILLEGAL ATOMIC NUMBER, 361
IMPOSSIBLE NUMBER OF ... , 361–362
IMPOSSIBLE OPTION REQ ... , 362
IMPOSSIBLE VALUE ... , 362
INDEX, 341
Infinite groups, 202
INPUT FILE MISSING ... , 362
INSUFFICIENT DATA ON ... , 362
INT, 51
Interatomic distances
 suppressing, 59
Internal
 coordinate
 force constants, 44, 223
 coordinate definition, 89
 energy, 80
Intrinsic reaction coordinates, 51
Ionization potential, 115
Ions, 1
 dipole moment for, 388
 identification of, 176
IRC, 51, 228–236
 example of, 230
 example of restart, 230
 keywords for, 229
 use of ISOTOPE, 51
 use of LARGE, 52
 use of RESTART, 71
Irreducible representations, 205
ISOTOPE, 51
 use with IRC, 228
Isotopes, 1, 44
 specification of, 86, **89**
ITER, 341
ITRY, 51
IUPD, 51

Jahn-Teller, 76
 distortion, 56
 theorem, 155
Jensen, Frank, 187
JOB STOPPED BY OPERATOR, 362

 $K=(n_1, n_2)$, 51
Kerr effect, 240
KEYWORD AIDER ... , 362
KEYWORD ANALYT ... , 362
Keywords
 debugging, 381
 definition of, 29
 full list of, 21
 too many words, 29, 30
 use of SETUP, 72
KEYWRD, 341
KINETIC, **52**, 231
Kinetic energy
 damping, 227
King, 52
Klamt, Andreas, 241
Klyne and Prelog, 211
Koopmans' theorem, 115
Kramer's degeneracy, 142
Kurtz, Henry, A., 66
Kurtz, Henry, A., 239

Laguerre polynomials, 129
LARGE, 52
LET, **52**, 210
LEWIS, 53
Lewis structure
 checking, 166
 printing, 53
 unusual, 38

- LINE OF KEYWORDS . . . , 362
Line search, 189–191
LINMIN, 53
Liotard, Daniel, 155
Liquids, 227
LMOs
 construction of starting, 170
 properties of starting, 177
 storage considerations, 178
LOCALIZE, 53
Localized MOs
 theory of, 160
Localized orbitals, 120
Locating transition states, 191
Log file, 59
Lone pairs, 116–117
 construction of, 175
LUMO, 42
Luque, F. J., 247

Magnetic component of spin, 58
MAKPOL, **196**
Mass
 effective, 217
 reduced, 215
Mass of atoms, 89
MAX, 53
MAX. NUMBER OF ATOMS . . . , 362
MAXIMUM ALLOWED . . . , 362
Mclver-Komornicki method, 359
MECI, 53, 138–155
 Increasing number of states, 143
 print State vectors, 52
 selecting specific roots, 71
 use of OPEN, 62
 use with MS, 58
 user-defined microstates, 54
MEMORY, 54
Memory
 how to minimize, 94
 management, 163
MEP, 247
MEP=1, 54
MEP=2, 54
MERS, 197
 use of, 102

Merz, K. M., 42
Messages, see Error messages
Methane
 cation, **56**, 76
Methane, data set, 99
MICROS, 54
Microstates, **141**, 205
 number used in C.I., 143
MICROSTATES SPECIFIED . . . , 363
MINDO/3, 1, **56**, 354
 accuracy of, 251
 Coulomb integrals, 112
 exchange integrals, 112
MINMEP, 57
MISSING VAN DER WAALS . . . , 363
MIXED PARAMETER SETS. . . , 363
MMOK, **57**, 122
MNDO, 1, 57, **156**, 357
 accuracy, 252
 accuracy of, 251
 elements in, 86
MNDO-*d*, 357
 elements in, 86
MNDOD, **57**
MODE, **57**, 187
MODIFY SUBROUTINE NAMES, 363
Modulus
 bulk, 63
MOLDAT , 57
Molecular dynamics, 226
Molecular mechanics, 1
Molecular mechanics correction
 to peptide linkage, **57**, 60
Molecular orbitals, 1, **158**
Molecular size, 193
Moments of inertia, 219
Monopole, 110
MOPAC, 57
 geometric structure, 341
 accuracy, 251
 as an educational tool, 4
 contributors, 17
 criteria, 377
 definition, 1
 development, 3
 electronic structure, 341

- error messages, 353
- how to cite, 16
- programming policy, 341
- to Gaussian format, 30
- MORE THAN 3N-6 COORD ... , 363
- MORE THAN ONE GEO ... , 363
- MOZYME, 57
 - limitations, 57
 - patent, 2
 - theory, 160
- MS, 58
- MST Solvation Model, 247
- MST solvation model, 247
- MULLIK, 46, **58**
- Mulliken
 - electronegativity, 76
- MULLIKEN POPULATION ... , 363
- Mulliken populations, 118
- Multipoles, 110

- N-H
 - core-core terms, 114
- N-methyl acetamide, 57, **122**
- NAME NOT FOUND, 363
- Narcissistic reactions, 191
- NEGATIVE SYMBOLICS ... , 364
- Neglect of overlap, 107
- NEWGEO , 59
- NEWGEO CAN ONLY ... , 364
- NLLSQ, **59**, 238, 359
- NLLSQ USED WITH ... , 364
- NLMO=*nnn*, 59
- NLO effects, 66
- NO ATOMS IN SYSTEM, 364
- NO DUMMY ATOMS ... , 364
- NO PEPTIDE LINKAGES ... , 364
- NO POINTS SUPPLIED ... , 364
- NO RESTART FILE EXISTS, 364
- NOANCI, 59
- NOANCI MUST BE USED ... , 364
- NOINTER, 59
- NOLOG, 59
- NOMM, 60
- Non-linear molecules, 220
- Non-Linear Optics, 66, 239
- Non-primitive translations, **197**

- NONET, 60
- NONET SPECIFIED ... , 364
- NONR, **60**, 187
- NORBS, 341
- NOREOR, 60
- Normal coordinates, 205
 - calculation of, 214–218
- NOSYM, 60
- NOTHIEL, **60**, 189
- NOXYZ, 61
- NSPA, 61
- NSURF, 61
- Nuclear energy, 115
- Nucleophilic Delocalizability, 76
- NUMBER OF DOUBLY ... , 364
- NUMBER OF ELECTRONS ... , 364
- NUMBER OF M.O.s IN ... , 365
- NUMBER OF M.O.s REQ ... , 365
- NUMBER OF OCCUPIED ... , 365
- NUMBER OF OPEN-SHELLS ... , 365
- NUMBER OF PARTICLES ... , 365
- NUMBER OF PERM ... , 365
- NUMERICAL PROBLEMS ... , 365

- O-H
 - core-core terms, 114
- OCCUPIED C VECTOR ... , 365
- OCCUPIED IC VECTOR ... , 365
- OCTET, 61
- OCTET SPECIFIED ... , 366
- Odd-electron systems, 35
- O_h^7 , 197
- Old parameter sets, 86
- OLDENS, 61
- OLDENS FILE FOR *file.den* ... , 366
- OLDFPC, 61
- OLDGEO, 61
- OLDGEO used and ... , 366
- Oleari, 109
- Oligomers, 196
- OMIN, 62, 187
- OMIN MUST BE ... , 366
- One electron
 - energy due to, 42
- One-electron integrals, **113**, 157
- ONLY *N* POINTS ... , 366

- ONLY DIHEDRAL SYMBOL . . . , 366
Only H, C, N, O, F, and S allowed, 366
ONLY ONE REACTION . . . , 366
OPEN, **62**, 140, 361
 use of, 99
OPTBACK, 62
Optical
 rectification, 240
OPTRES, 63
Orbital
 energies in C.I., 144
Orientation
 for symmetry, 203
Orozco, Modesto, 247
ORT, 63
Oscillations
 damping, 73, **124**
Oscillator, 146
Overlap
 neglect of, 107
 integrals, 129–131
 between two hydrogen atoms, 156

P=nn, 63
Paired electron excitations, 64
PANIC: Temporary array . . . , 366
Parameter sets
 old, 86
PARAMETERS FOR SOME . . . , 366
Parameters, altering, 43
PARASOK, **63**
Partition function, 80, **218**
Path calculations, 224
PBT, 225
PDB
 disorder in, 31
 format, **92**
 keyword, 63
 unrecognized symbols in, 64
PDB(*text*)
 keyword, 64
PDBOUT, 64
PECI, 64
Peptide linkage
 correction to barrier, **57**, 60, 122
Periodic boundary conditions, **194**

PERMANENT INTEGER . . . , 367
PERMANENT REAL ARRAY . . . , 367
Phenoxy radical, 42
Phosphorescence, 154
Phosphorus
 AM1 fault, 253
Photoemission, 153
Photoexcitation energy, 153
PI, 64
 π bonds, acyclic
 construction of, 175
 π bonds, cyclic
 construction of, 176
 π systems
 delocalized, 163
Pierotti's scaled particle theory, 247
PINOUT, 64
PL, 64
Platonic solids, 203
PM3, 1, **65**, 357
 difference from MNDO, 114
 accuracy, 253
 fault in peptide linkage, 122
PM3(tm), 86
 specification of parameters, 87
PMEP, 65, **245**
PMEPR, 65
Pockel's Effect, 66
Pockels effect, 240
POINR, 66
Point Group D_{8d} missing, 201
Point-group
 dynamic Jahn-Teller, 155
 theory, 201
POINT1, 65
POINT2, 65
POLAR, 66
 effect of LET on, 53
POLAR does not work with . . . , 367
Polarizability, 66, 75, 146, 153, **239**
 atomic transition, 146
Polyacetylene, 196
Polyethylene, 102
Polymers, 1
 example, 101
 stretching, 225

- Polytetrahydrofuran, 101
Pople, 107
POPS, 66
port.dat, 97
Porting, 38
Positron equivalent, **108**, 206
POTWRT, 66
POWSQ, 67
PRECISE, 67, 378, 379
 use in FORCE calcn., 45
Precision, 207
 increasing, 67
 of SCF, 72
 specifying high precision, 45
Pressure
 applied, 63
PROBLEM IN SYMR, 367
Proteins
 charged sites in, 176
 geometry optimization
 caution, 58
 residues in, 69
PRTMEP, 67
PULAY, **67**, 342
Pulay's DIIS, 125
Pyridine, 142

Q.N., 153
QPMEP, 67
Quadrupole, 110
Quantum numbers
 in overlap, 129
 of states, 153
QUARTET, 67
QUARTET SPECIFIED ... , 367
Quartet states, 142
QUINTET, 68
QUINTET SPECIFIED ... , 367

Radicals, 1, **138**
Ran out of storage for *arrayname*, 367
RAPID, 68
RE-LOCAL, 68
Reaction coordinate
 specification, 89
 use of POINT, 66
 use of STEP, 76
Reaction path calculation, 224
Reaction paths, 224
READA, 341
RECALC, 68
Red-shift, 153
Reduced mass, **215**, 233
Reference data, 342
Refractive index, 240
RELSCF, 69
RELTHR, 69
REORTH, 69
RESEQ, 70
RESIDUES, 69
Resonance integrals, 113
RESTART, 19, **70**
 use with IRC/DRC, 229
RESTART FILE EXISTS, ... , 368
Restart file is corrupt, 368
Restart files, 41
RMAX, 71, 187
RMIN, 71, 187
ROOT, 71
 use of, 99
ROOT REQUESTED DOES ... , 368
Roothaan, 108
Rotation barrier calculation, 224
Rotation barriers, 124, **226**
Rotational invariance, 33, **108**, 121
RSCAL, 71, 187
RSOLV, 71
RSOLV MUST NOT BE ... , 368
Run stopped because keyword LEWIS was used, 368
Russell-Saunders, **56**, 207

S, 221
 S_{N^2} , 120
 reaction, 224
SADDLE, 32, **71**
 description, 236–239
SAFE, 72
SAS, **244**
SCALE, 72
SCF, 44, 61, 67, 73, **114**
 configurations, 140
 convergence techniques, 124
 criterion, 341, **377**

- modifying, 72
- failure to achieve, 125
- test, 211
- worked example, 155
- 0SCF, **30**, 93
- 1SCF
 - worked example, 155
- 1SCF, 30
- SCFCRT, 67, **72**, 209, 377
- Schrödinger, 120
- SCINCR, 72
- SCRf Method, 247
- Second Harmonic Generation, 66
- Secular determinant
 - in MECI, 144
- SELCON, 211
- Self Consistent Reaction Field method, 247
- SEPTET, 72
- SEPTET SPECIFIED . . . , 368
- SETUP, 72
- SETUP FILE MISSING . . . , 368
- SEXTET, 73
- SEXTET SPECIFIED . . . , 368
- Sextet states, 142
- SHG, 66
- SHIFT, **73**, 359
- Shift operators, 150
- SIGMA, **73**, 238, 359
- Sigma Orbitals
 - construction of, 174
- SIGMA USED WITH . . . , 368
- Simple harmonic motion, 215
- SINGLE ATOM HAS NO VIB . . . , 353
- Single Electron Excitations, 36
- Single-sided derivatives, 213
- SINGLET, 73
- Singlet, 42
- Size of active space allowed: N , 368
- Size of molecule, 193
- SIZES, 74
- Slater determinant, 141–142, 205
- Slater orbitals, 129
 - overlap of, 156
- SLOG, 75
- SLOG=*n.nn*, 75
- SLOPE, 75
- SNAP, 75
- Solid state
 - data sets, 198
 - description, 193
- Solvation Model
 - MST, 247
- Solvation model
 - COSMO, 241
 - MST model, 247
- Solvent Accessible Surface, **244**
- SOME ELEMENTS HAVE . . . , 368
- Something disastrous . . . , 369
- sp³ hybrid atomic orbitals, 173
- Space group
 - operations, **197**
 - theory, 197
- Space quantization, **56**, 138
- Sparkles
 - description, 127
- SPECIFIED SPIN COMP . . . , 369
- Spherical harmonics, 129
- SPIN, 75
- Spin
 - density, unpaired, 42, **155**
 - operators, 150
 - quantization, 56, 138, **149**
 - states, 40
- Standard Enthalpy of Formation, 221
- Standard Gibbs Free-Energy of Formation, 223
- Standard Internal Energy of Formation, 223
- STATE, 341
- State
 - energies, 146
 - function, 42, **151**
 - vectors, 146
 - printing, 52, 53
- States
 - arising from C.I., 149
 - electronic, 205
 - mixtures of, 155
 - symmetry of, 205
 - vibration, 205
- STATIC, 75
- Stationary point, 44
- STEP, 76
- STEP2, 76

- Stewart, Anna C.**, 17
STO-6G, 210
STO3G, 76
Storage needed for next . . . , 369
STRUCTURE UNRECOGNIZABLE, 369
SUPER, 76
Superdelocalizability, 76
sym.force, 201
SYMAVG, 76
SYMMETRY, 76
Symmetry
 functions
 defined, 76
 in group theory, 201
 not used in FORCE calcn., 44
 orientation of molecules, 203
 use in reaction paths, 226
 use of sub-group, 60
SYMMETRY FUNCTION IS . . . , 353
SYMMETRY SPECIFIED . . . , 369
SYSTEM DOES NOT . . . , 369
SYSTEM SPECIFIED . . . , 369
SYSTEMS WITH TV . . . , 370

T-PRIORITY, 80
T= n , 80
Taylor expansion, 187
TEMPERATURE RANGE . . . , 370
The COSMO option cannot be . . . , 370
The data set contains alt . . . , 370
THE FOLLOWING SYMBOL . . . , 370
THE GEOMETRY DATA-SET . . . , 370
THE GREENS FUNCTION IS . . . , 370
THE HAMILTONIAN REQ . . . , 370
The input data file does not exist, 371
THE SCF CALCULATION . . . , 371
THE STATE REQUIRED IS . . . , 371
There are more than N . . . , 371
THERE ARE NO VARIABLES, 371
There are too many keywords, 371
There is a bug in MOZYME, 371
THERE IS A FAULT IN . . . , 371
THERE IS A FAULT IN RESEQ, 371
There is a mistake . . . , 372
THERE IS A RISK OF INF . . . , 372
THERE MUST BE EXACTLY . . . , 372

THERMO, 44, **80**
Thermochemistry, 218–223
THG, 66
Thiel, Walter, 60
Thiel, Walter, 190
Third Harmonic Generation, 66, 240
THIS MESSAGE SHOULD . . . , 372
THREE ATOMS BEING . . . , 372
Three Electron Excitations, 36
THRESH, 81
TIDY, 81
TIME UP, 372
Time-dependent phenomena, 226
Time-reversal, 33
TIMES, 81
Title of calculation, 21
TOL2, 377
TOLERG, 377
TOLERX, 377
TOLS1, 377
TOM, 81
TOM, 248
TOO MANY CONFIGURATIONS, 372
TOO MANY ITERATIONS . . . , 372
TRANS, 82
Transition
 dipole moments, 146
Transition state
 location, 1, **71**
 use with DRC, 227
Transition states
 locating, 191
Translation vector
 requirements, 194
 specification of, 90
Travel, 217
TRIPLET, **82**, 125, 140
Triplet, 42
TRIPLET SPECIFIED . . . , 372
TRUST RADIUS NOW LESS . . . , 373
TS, **82**, 238, 359
TS FAILED TO LOCATE . . . , 373
TS SEARCH AND BFGS . . . , 373
TS SEARCH REQUIRE . . . , 373
TWO ADJACENT POINTS . . . , 373
TWO ATOMS ARE COIN . . . , 373

- Two electron
 energy due to, 42
 excitations, 36, 64
 integrals, 109
- UHF, **82**, 125
- UMNDO, 201
- UNABLE TO ACHIEVE ... , 373
- UNACCEPTABLE VALUE ... , 373
- UNDEFINED SYMMETRY ... , 373
- UNKNOWN RESIDUE: *N*, 373
- UNOPTIMIZABLE GEO ... , 353
- Unpaired spin density, 42, **155**
- Unpolarizable core
 used in C.I., 144
- UNRECOGNIZED ELEMENT ... , 373
- UNRECOGNIZED HESS OPTION, 373
- UNRECOGNIZED KEY-WORDS., 373
- UNRECOGNIZED SPECIES: *text*, 374
- Unrestricted Hartree-Fock, 82
- UNSAFE, 83
- UPPER BOUND OF ACTIVE ... , 374
- Use AIGIN to allow more ... , 374
- USE EITHER SAFE OR ... , 374
- Valency, 31, 34, **116**
- VALUE OF NPPA NOT ... , 374
- VAN DER WAALS, 374
- Van der Waals, 72, **215**
 in MST theory, 247
 radii, changing, 83
- VDW, 83
- VECPRT CAN ONLY PRINT ARRAYS OF SIZE .LT.
 200, 374
- VECTORS, 83
- VELOCITY, 83
- Vibrational analysis, 215
- Vibrations, 1
- VIRTUAL C VECTOR ... , 374
- VIRTUAL IC VECTOR ... , 374
- Von Niessen, 120
- Wang, Bingze**, 245
- WARNING, 375
- WARNING – GEOMETRY ... , 374
- WARNING – SPARKLES ARE ... , 375
- WARNING NO BACKBONE ATOMS ... , 375
- Water, dimer, 210
- Wavefunctions
 non-variational, 212
 variational, 212
- Wiberg indices, 116
- Wigner's symbol, 142
- WILLIAMS, 84
- WINMOPAC, 84
- X-PRIORITY, 84
- XENO, 84
- XX, 85, 88
- XYZ, **85**
- ZDO, 106

*
This
Manual
was type-
set using the
L^AT_EX Document
Preparation Sys-
tem written by Leslie
Lamport. L^AT_EX is based
on the T_EX program writ-
ten by Donald Knuth. The most
common font used is ten-point ro-
man.

v v v

v v v

v v v

v v v

v v v

v v v

v v v

v v v

v v v

v v v

v v v

v v v

XX

XX