

General Description of 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](#)], AM1 [[3](#)], PM3 [[4](#)], PM6, [RM1](#), 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 unique, 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, MOPAC was an obvious contender.

James J. P. Stewart	E-mail: MrMOPAC@OpenMOPAC.net
Stewart Computational Chemistry	Tel: (USA) + 719-488-9416
15210 Paddington Circle	SKYPE: Jimmy.Stewart2
Colorado Springs, CO 90821	© 2007

MOPAC2007

[Description](#)

A brief description of what MOPAC is

[Cooperative Development](#)

How to help with the development of the program

[Graphical User Interfaces](#)

Graphical programs to help with input and output

[Teaching](#)

How MOPAC can be used in teaching

[Testing and Validation](#)

[Installing the executable](#)

For users who download the executable and have received a license key

[Installing and compiling source](#)

For users who have source code for research.

[Running](#)

[Structure of the program](#)

[Debugging](#)

[Porting](#)

[Validation](#)

[Control within MOPAC](#)

Keywords used in MOPAC2009

<u>&</u>	Turn next line into keywords
<u>±</u>	Add another line of keywords
<u>0SCF</u>	Read in data, then stop
<u>1ELECTRON</u>	Print final one-electron matrix
<u>1SCF</u>	Do one scf and then stop
<u>AIDER</u>	Read in ab-initio derivatives
<u>AIGIN</u>	Geometry must be in gaussian format
<u>AIGOUT</u>	In arc file, include ab-initio geometry
<u>ALLBONDS</u>	Print final bond-order matrix, including bonds to hydrogen
<u>ALLVEC</u>	Print all vectors (keywords vectors also needed)
<u>ALT_A=A</u>	In PDB files with alternative atoms, select atoms A
<u>ALT_R=A</u>	In PDB files with alternative residues, select residues A
<u>AUTOSYM</u>	Symmetry to be imposed automatically
<u>AUX</u>	Output auxiliary information for use by other programs
<u>AM1</u>	Use the AM1 hamiltonian
<u>BAR=n.nn</u>	reduce bar length by a maximum of n.nn%
<u>BCC (Unique)</u>	Solid is body-centered cubic (used by BZ)
<u>BFGS</u>	Optimize geometries using bfgs procedure
<u>BIGCYCLES=n</u>	Do a maximum of n big steps
<u>BIRADICAL</u>	System has two unpaired electrons
<u>BONDS</u>	Print final bond-order matrix
<u>CAMP</u>	Use Camp-King converger in SCF
<u>CARTAB</u>	Print point-group character table
<u>C.I.=n</u>	A multi-electron configuration interaction specified
<u>C.I.=(n,m)</u>	
<u>CHAINS(text)</u>	In a protein, explicitly define the letters of chains.
<u>CHARGE=n</u>	Charge on system = n (e.g. NH4 = +1)
<u>CHARGES</u>	Print net charge on system, and all charges in the system
<u>CHARST</u>	Print details of working in CHARST
<u>CIS</u>	C.I. uses 1 electron excitations only
<u>CISD</u>	C.I. uses 1 and electron excitations
<u>CISDT</u>	C.I. uses 1, 2 and 3 electron excitations
<u>COMPFG</u>	Print heat of formation calculated in COMPFG
<u>COSCCH</u>	Add in COSMO charge corrections
<u>COSWRT</u>	Write details of the solvent accessible surface to a file
<u>CUTOFP=n.nn</u>	Madelung distance cutoff is n.nn Angstroms
<u>CUTOFF=n.nn</u>	In MOZYME, the interatomic distance where the NDDO approximation stops
<u>CYCLES=n</u>	Do a maximum of n steps
<u>CVB</u>	In MOZYME. add and remove specific bonds to allow a Lewis or PDB structure.
<u>DAMP=n.nn</u>	n MOZYME. damp SCF oscillations using a factor of n.nn
<u>DATA=text</u>	Input data set is re-defined to <i>text</i>
<u>DCART</u>	Print part of working in DCART
<u>DDMAX=n.nn</u>	See EF code
<u>DDMIN=n.nn</u>	Minimum trust radius in a EF/TS calculation
<u>DEBUG</u>	Debug option turned on
<u>DEBUGPULAY</u>	Print working in PULAY
<u>DENOUT, DENOUTF</u>	Density matrix output
<u>DENSITY</u>	Print final density matrix

<u>DERI1</u>	Print part of working in DERI1
<u>DERI2</u>	Print part of working in DERI2
<u>DERITR</u>	Print part of working in DERIT
<u>DERIV</u>	Print part of working in DERIV
<u>DERNVO</u>	Print part of working in DERNVO
<u>DFORCE</u>	Force calculation specified, also print force matrix.
<u>DFP</u>	Use Davidson-Fletcher-Powell method to optimize geometries
<u>DISEX=<i>n.nn</i></u>	Distance for interactions in fine grid in COSMO
<u>DMAX=<i>n.nn</i></u>	Maximum stepsize in eigenvector following
<u>DOUBLET</u>	Doublet state required
<u>DRC</u>	Dynamic reaction coordinate calculation
<u>DRC=<i>n.nnn</i></u>	
<u>DUMP=<i>nn.nn</i></u>	Write restart files every n seconds
<u>ECHO</u>	Data are echoed back before calculation starts
<u>EF</u>	Use ef routine for minimum search
<u>EIGEN</u>	Print canonical eigenvectors instead of LMOs in MOZYME calculations
<u>EIGS</u>	Print all eigenvalues in ITER
<u>ENPART</u>	Partition energy into components
<u>EPS=<i>n.nn</i></u>	Dielectric constant in COSMO calculation
<u>ESP</u>	Electrostatic potential calculation
<u>ESPRST</u>	Restart of electrostatic potential
<u>ESR</u>	Calculate RHF spin density
<u>EXCITED</u>	Optimize first excited singlet state
<u>EXTERNAL=<i>name</i></u>	Read parameters off disk
<u>FIELD=(<i>n.nn,m.mm,l.ll</i>)</u>	An external electric field is to be used
<u>FILL=<i>n</i></u>	In RHF open and closed shell, force M.O. n to be filled
<u>FLEPO</u>	Print details of geometry optimization
<u>FMAT</u>	Print details of working in FMAT
<u>FOCK</u>	Print last Fock matrix
<u>FREQCY</u>	Print symmetrized Hessian in a FORCE calculation
<u>FORCE, _FORCETS</u>	Calculate vibrational frequencies
<u>GEO-CHECK</u>	Apply extra safety checks
<u>GEO-OK</u>	Override some safety checks
<u>GEO_REF=<<i>text</i>></u>	Use native structure as reference
<u>GNORM=<i>n.nn</i></u>	Exit when gradient norm drops below n .n kcal/mol/Angstrom
<u>GRADIENTS</u>	Print all gradients
<u>GRAPH</u>	Generate unformatted file for graphics
<u>GRAPHF</u>	Generate formatted file for graphics suitable for Jmol and MOPETE .
<u>HCORE</u>	Print all parameters used, the one-electron matrix, and two-electron integrals
<u>HESSIAN</u>	Print Hessian from geometry optimization
<u>HESS=<i>n</i></u>	Options for calculating Hessian matrices in EF
<u>H-PRIORITY</u>	Heat of formation takes priority in DRC
<u>H-PRIORITY=<i>n.nn</i></u>	
<u>HYPERFINE</u>	Hyperfine coupling constants to be calculated
<u>INT</u>	Make all coordinates internal coordinates
<u>INVERT</u>	Reverse all optimization flags
<u>IRC</u>	Intrinsic reaction coordinate calculation
<u>IRC=<i>n</i></u>	
<u>ISOTOPE</u>	Force matrix written to disk (channel 9)
<u>ITER</u>	Print details of working in ITER
<u>ITRY=<i>nn</i></u>	Set limit of number of SCF iterations to n
<u>IUPD=<i>n</i></u>	Mode of Hessian update in eigenvector following

<u>KINETIC=<i>n.nnn</i></u>	Excess kinetic energy added to DRC calculation
<u>KING</u>	Use Camp-King converger for SCF
<u>LARGE</u>	Print expanded output
<u>LBFGS</u>	Use the low-memory version of the BFGS optimizer
<u>LET</u>	Override certain safety checks
<u>LEWIS</u>	Print the Lewis structure
<u>LINMIN</u>	Print details of line minimization
<u>LOCALIZE</u>	Print localized orbitals. These are also called Natural Bond Orbitals or NBO
<u>LOG</u>	Generate a log file
<u>MECI</u>	Print details of MECI calculation
<u>MERS=(<i>n</i>₁,<i>n</i>₂,<i>n</i>₃)</u>	Keyword for BZ
<u>MICROS=<i>n</i></u>	Use specific microstates in the C.I.
<u>MINI</u>	Reduce the amount of geometry data printed
<u>MINMEP</u>	Minimize MEP minima in the plane defined
<u>MMOK</u>	Use molecular mechanics correction to CONH bonds
<u>MNDO</u>	Use the MNDO hamiltonian
<u>MNDOD</u>	Use the MNDO-d hamiltonian
<u>MODE=<i>n</i></u>	In EF, follow Hessian mode no. <i>n</i>
<u>MOLDAT</u>	Print details of working in MOLDAT
<u>MOLSYM</u>	Print details of working in MOLSYM
<u>MOPAC</u>	Use old MOPAC definition for 2nd and 3rd atoms
<u>MOZYME</u>	Use the Localized Molecular Orbital method to speed up the SCF
<u>MS=<i>n</i></u>	In MECI, magnetic component of spin
<u>MULLIK</u>	Print the Mulliken population analysis
<u>N**2</u>	In excited state COSMO calculations, set the value of N**2
<u>NLLSQ</u>	Minimize gradients using NLLSQ
<u>NOANCI</u>	Do not use analytical C.I. derivatives
<u>NOLOG</u>	Suppress log file trail, where possible
<u>NOMM</u>	Do not use molecular mechanics correction to CONH bonds
<u>NONET</u>	NONET state required
<u>NONR</u>	Do not use Newton-Raphson method in EF
<u>NOOPT, _NOOPT-X</u>	Do not optimize the coordinates of all atoms of type X
<u>NOREOR</u>	In symmetry work, use supplied orientation
<u>NOSYM</u>	Point-group symmetry set to C1
<u>NOTHIEL</u>	Do not use Thiel's FSTMIN technique
<u>NOXYZ</u>	Do not print Cartesian coordinates
<u>NSPA=<i>n</i></u>	Sets number of geometric segments in COSMO
<u>NSURF</u>	Number of surfaces in an ESP calculation
<u>OCTET</u>	Octet state required
<u>OLDCAV</u>	In COSMO, use the old Solvent Accessible Surface calculation
<u>OLDENS</u>	Read initial density matrix off disk
<u>OLDFPC</u>	Use the old fundamental physical constants
<u>OLDGEO</u>	Previous geometry to be used
<u>OMIN=<i>n.nn</i></u>	In TS, minimum allowed overlap of eigenvectors
<u>OPEN(<i>n</i>₁,<i>n</i>₂)</u>	Open-shell RHF calculation requested
<u>OPT, _OPT-X</u>	Optimize the coordinates of all atoms of type X
<u>P=<i>n.nn</i></u>	An applied pressure of <i>n.nn</i> Newtons/m ² to be used
<u>PDB</u>	Input geometry is in protein data bank format
<u>PDB=(<i>text</i>)</u>	User defined chemical symbols in protein data base
<u>PDBOUT</u>	Output geometry in pdb format
<u>PECI</u>	C.I. involves paired excitations only
<u>PI</u>	

<u>pKa</u>	Resolve density matrix into s, p, and d components
<u>PL</u>	Print the pKa for ionizable hydrogen atoms attached to oxygen atoms
<u>PM3</u>	Monitor convergence of density matrix in ITER
<u>PM6</u>	Use the MNDO-PM3 Hamiltonian
<u>PM6-DH+</u>	Use the PM6 Hamiltonian
<u>PM6-DH2</u>	Use the PM6 Hamiltonian with corrections for dispersion and hydrogen-bonding
<u>PMEP</u>	Use the PM6 Hamiltonian with corrections for dispersion and hydrogen-bonding
<u>PMEPR</u>	Complete semiempirical MEP calculation
<u>POINT=n</u>	Complete semiempirical MEP in a plane to be defined
<u>POINT1=n</u>	Number of points in reaction path
<u>POINT2=n</u>	Number of points in first direction in grid calculation
<u>POLAR</u>	Number of points in second direction in grid calculation
<u>POTWRT</u>	Calculate first, second and third order polarizabilities
<u>POWSQ</u>	In ESP, write out electrostatic potential to unit 21
<u>PRECISE</u>	Print details of working in POWSQ
<u>PRESSURE</u>	Criteria to be increased by 100 times
<u>PRNT=n</u>	Apply pressure or tension to a solid or polymer
<u>PRTINT</u>	Print details of geometry optimization in EF
<u>PRTMEP</u>	Print interatomic distances
<u>PULAY</u>	MEP contour data output to <filename>.mep
<u>QPMEP</u>	Use Pulay's converger to obtain a SCF
<u>QUARTET</u>	Charges derived from Wang-Ford type AM1 MEP
<u>QUINTET</u>	Quartet state required
<u>RAPID</u>	Quintet state required
<u>RECALC=n</u>	In MOZYME geometry optimizations, only use atoms being optimized in the SCF
<u>RE-LOCAL, RE-LOCAL=n</u>	In EF, recalculate Hessian every n steps
<u>RELSCF</u>	During and at end of MOZYME calculation, re-localize the LMOs
<u>REORTHOG</u>	Default SCF criterion multiplied by n
<u>RESEQ</u>	In MOZYME, re-orthogonalize LMO's each 10 SCF calculations.
<u>RESIDUES</u>	Re-arrange the atoms to match the PDB convention
<u>RESTART</u>	Label each atom in a polypeptide with the amino acid residue
<u>RHF</u>	Calculation restarted
<u>RM1</u>	Use Restricted Hartree-Fock methods
<u>RMAX=n.nn</u>	Use the RM1 Hamiltonian
<u>RMIN=n.nn</u>	In TS, maximum allowed ratio for energy change
<u>ROOT=n</u>	In TS, minimum allowed ratio for energy change
<u>RSCAL</u>	Root n to be optimized in a C.I. calculation
<u>RSOLV=n.nn</u>	In EF, scale p-RFO to trust radius
<u>SADDLE</u>	Effective radius of solvent in COSMO
<u>SCALE</u>	Optimize transition state
<u>SCFCRT=n.nn</u>	Scaling factor for van der waals distance in ESP
<u>SCINCR=n.nn</u>	Default SCF criterion replaced by the value supplied
<u>SEPTET</u>	Increment between layers in ESP
<u>SETPI</u>	Septet state required
<u>SETUP</u>	In MOZYME, some p bonds are explicitly set by the user
<u>SEXTET</u>	Extra keywords to be read from setup file
<u>SHIFT=n.nn</u>	Sextet state required
<u>SHUT <file></u>	a damping factor of n defined to start SCF
	Send a command to MOPAC to make a restart and density file, then stop.

<u>SIGMA</u>	Minimize gradients using SIGMA
<u>SINGLET</u>	Singlet state required
<u>SITE=(text)</u>	Define ionization state of residues in proteins
<u>SLOG=n.nn</u>	In BFGS optimization, use fixed step of length n .nn
<u>SLOPE</u>	Multiplier used to scale MNDO charges
<u>SNAP</u>	Increase precision of symmetry angles
<u>SPARKLE</u>	Use sparkles instead of atoms with basis sets
<u>SPIN</u>	Print final UHF spin matrix
<u>START_RES(text)</u>	Define starting residue numbers in a protein, if different from the default
<u>STATIC</u>	Calculate Polarizability using electric fields
<u>STEP</u>	Step size in path
<u>STEP1=n.nnn</u>	Step size n for first coordinate in grid calculation
<u>STEP2=n.nnn</u>	Step size n for second coordinate in grid calculation
<u>STO3G</u>	Deorthogonalize orbitals in STO-3G basis
<u>SUPER</u>	Print superdelocalizabilities
<u>SYBYL</u>	Output a file for use by Tripos's SYBYL program
<u>SYMAVG</u>	Average symmetry equivalent ESP charges
<u>SYMOIR</u>	Print characters of eigenvectors and print number of I.R.s
<u>SYMTRZ</u>	Print details of working in subroutine SYMTRZ.
<u>SYMMETRY</u>	Impose symmetry conditions
<u>T=n[M,H,D]</u>	A time of n seconds requested
<u>THERMO</u>	
<u>THERMO(nnn)</u>	Perform a thermodynamics calculation
<u>THERMO(nnn,mmm)</u>	
<u>THERMO(nnn,mmm,lll)</u>	
<u>TIMES</u>	Print times of various stages
<u>T-PRIORITY</u>	
<u>T-PRIORITY=n.nn</u>	Time takes priority in DRC
<u>TRANS</u>	
<u>TRANS=n</u>	The system is a transition state (used in thermodynamics calculation)
<u>TRIPLET</u>	Triplet state required
<u>TS</u>	Using EF routine for TS search
<u>UHF</u>	Use the Unrestricted Hartree-Fock method
<u>VDW(text)</u>	Van der waals radius for atoms in COSMO defined by user
<u>VDWM(text)</u>	Van der waals radius for atoms in MOZYME defined by user
<u>VECTORS</u>	Print final eigenvectors
<u>VELOCITY</u>	Supply the initial velocity vector in a DRC calculation
<u>WILLIAMS</u>	Use Williams surface
<u>X-PRIORITY=n.nn</u>	Geometry changes take priority in DRC
<u>XENO</u>	Allow non-standard residues in proteins to be labeled.
<u>XYZ</u>	Do all geometric operations in Cartesian coordinates
<u>Z=n</u>	Number of mers in a cluster

Geometry Specification

Geometries are defined in terms of the atoms involved and their coordinates. If you are new to MOPAC, then the fastest way to understand the general layout is to look at an [example of a simple data set](#).

The data is in a "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 Gaussian format.

-
- Definition of Elements for [MNDO](#), [AM1](#), [PM3](#), [RM1](#), and [PM6](#)
 - Definition of [Isotopes](#)
 - [Atom labels](#)
 - [Internal Coordinate Definition](#)
 - [Constraints](#)
 - [Cartesian Coordinate Definition](#)
 - [Mixed Cartesian and internal coordinate definition](#)
 - [Translation Vector](#)
 - [Optimization flags](#)
 - [Applying constraints, freezing coordinates, and fixing distances, etc.](#)
 - [Specification of Gaussian Z-matrices](#)
 - [Conversion Between Various Formats](#)
 - [Notes on proteins and other macromolecules](#)

Accuracy

Accuracy of PM6

A frequently asked question is, "Which method in MOPAC is the most accurate?" Although apparently a simple question, there is no simple answer. Several years of work went into developing the PM6 method, and that work made extensive use of the experience gained during the development of the earlier methods. So my answer would be, "PM6." But if you were to accept that answer, you would also need to trust that what I said was also accurate. When there is the opportunity to see an unbroken chain of logic from raw experimental data to a summary of a statistical analysis of the errors in various properties for various methods, that is almost infinitely preferable to hearing someone say, "trust me."

[Raw data with references](#) on the basic set of ~7600 species used in the analysis.

A [statistical package can be downloaded](#) that will allow a wide range of comparisons of AM1, PM3, and PM6.

A comparison of various [hydrogen bonded structures predicted by PM6](#) with reference X-Ray structures

The ability of PM6 to reproduce [bond-alternation](#)

Users might be interested in a very quick summary of [advantages and disadvantages of the various methods](#).

Average Unsigned Errors in PM6 Predictions

Quantity	PM6	PM5	PM3	AM1	Units
DH_f	8.01	22.19	18.20	22.86	Kcal/mol
Bond Lengths	0.091	0.123	0.104	0.130	Angstroms
Angles	7.86	9.55	8.50	8.77	Degrees
Dipoles	0.85	1.12	0.72	0.67	Debye
I.P.s	0.50	0.50	0.68	0.63	eV

Other Tables of Results

Quantity	Average Unsigned Error
DH_f (PM6 and other semiempirical)*	(Depends on set used)
Bond-lengths (PM6 and other semiempirical)	(Depends on set used)
DH_f (PM6 and ab-initio)*	4.44 kcal/mol
Relative conformer energies	2.35 kcal/mol
Nitrogen pyramidalization	5.0 degrees
Hydrogen bond energies	1.70 kcal/mol
Intermolecular interactions using PM6-DH2	0.37 kcal/mol
Intermolecular interactions using PM6-DH+	
Water Dimerization energies	1.04 kcal/mol (one value)
Polarizability volume (Å³)	2.1%
Vibrational frequencies (cm⁻¹)	14%
Entropy	3.16 cal/(degree.mole)
pKa	0.31
Cp (heat capacity)	1.48 cal/(degree.mole)

* DH_f is the heat required to form one mole of the gaseous compound from its elements in their standard state, at 298K.

Known severe errors in PM6

Type of error	Example
Relative energies of neutral and Zwitterionic species	-NH₂ + -OH <=> -NH₃⁺ + -O⁻
Geometry of Fe(CO) ₅ is predicted to be C _{4v} , should be D _{3h}	Ref. Fe(CO)₅ and PM6 Fe(CO)₅
Ferrocene is predicted to be C _{2v} not D _{5d} or D _{5h} .	
Fe(III)X ₆ , 6A1g, is distorted from the octahedral symmetry	FeF ₆
Oxalic acid is twisted, should be planar	Oxalic acid
Heat of reaction: Oxalic acid = CO ₂ + C(OH) ₂ is in error by ~30 kcal/mol	Oxalic acid heat of reaction
R-(R-O-C=O)-NH ₂ -R is stable, should decompose to R-COO-R + R-NH ₂	amine-carboxylic acid
Biphenyl torsion is predicted to be 57 degrees, should be 45 degrees	
N-(Benzylidene)aniline is predicted to be cis, should be trans	
Iodine - non-bonded oxygen too close, 2Angstroms, should be ~3Angstroms	
H(+) is in error by -54 kcal/mol	These species are unlikely to be found in biochemical systems. The faults cannot easily be corrected by re-parameterization.
Methylene triplet is predicted to be linear, should be bent	
Ag ₇ decomposes into Ag + 3 Ag ₂	
SiO ₂ is predicted to be bent, it should be linear. The O-Si-O angle in silica is predicted to be too large	All forms of crystal SiO ₂ , except hexagonal beta tridymite
Non-bonding Br - N interactions are bonding.	HCN - BrCH₃
Non-bonding S - N interactions are bonding.	Me₂C=S - N(Me)₃
Non-bonding S - O interactions are bonding.	Me₂O - S=CH₂
Non-bonding S - Cl(-) interactions are bonding.	H₂C=S - Cl(-)
Non-bonding S - S interactions are bonding.	
Non-bonding Se - Se interactions are bonding.	
Non-bonding Se - I interactions are bonding.	
Non-bonding Te - Cl interactions are bonding.	
Non-bonding Br - O interactions are bonding.	
Non-bonding Br - Br interactions are bonding.	
Non-bonding I - O interactions are bonding.	
Non-bonding I - I interactions are bonding.	
Non-bonding N - I interactions are bonding.	
Non-bonding Cl(-) -H interactions are bonding.	
Pyramidal enolate to planar (CH ₃ -CO(-)=CH ₂) energy difference is too large (14, should be 6 kcal/mol)	
BrF ₃ is predicted to be D _{3h} should be C _{2v}	
PrF ₅ is predicted to be D _{3h} should be C _{4v}	
TiH ₄ collapses	
[Re ₂ Cl ₈] ⁻ has two bridging Cl atoms	
(Please report errors as soon as they are found)	

Known errors in MOPAC2009

By default, [UHF CH₄ + F](#) generates an electronic excited state

Accuracy of operations within MOPAC2009

[Geometry optimization](#)

[Precision levels](#)

[Keyword "PRECISE"](#)

[Reasons for low precision](#)

[Setting the Gradient Norm \(GNORM\)](#)

SCF criteria:

[Test for self-consistency \(SELCON and SCFCRT\)](#)

[Setting the absolute SCF criterion \(SCFCRT\)](#)

[Setting the relative SCF Criterion \(RELSCF\)](#)

Monitoring SCF convergence using [PL](#)

Features of MOPAC2009

Relationship to MOPAC2007

MOPAC2009 is best regarded as MOPAC2007 plus the MOZYME function.

Relationship to MOPAC6

MOPAC2007 is an updated version of MOPAC6, the last public domain version of MOPAC. MOPAC6 had been written in FORTRAN-77, and had become unwieldy. Maintenance had become increasingly difficult, and the decision was made to completely re-write the program in FORTRAN-90/95. During this process several major changes were made. The method MINDO/3 was deleted as were several infrequently-used functionalities. The structure of the program was reorganized, so that maintenance and debugging would be easier. A consistent style was imposed on the whole code. This meant that once the overall structure was understood, individual routines could be read by developers with relative ease.

Reliability

Despite all the work that has gone into MOPAC2007, it has relatively few features that were not present in MOPAC6; the focus of the effort has not been to add new features, rather it has been to produce a highly-debugged, robust program. The goal is that all the functions implied by the keywords should work as described. During the beta-test, from September 2006 to the end of January 2007, several hundred bugs were found and fixed. Towards the end of the beta-test the rate at which bugs were being found dropped to about two to three a day, and most of those that were reported were not too serious. After MOPAC2007 was released in late January, over 500 licenses were issued and, as a result of this large number of users/groups all busily using the new program, the steady trickle of bug-reports continued. The policy is that whenever a bug is reported all work stops until the fault is corrected. In practice, this meant that bugs were fixed with about two days, so MOPAC2007 is constantly being maintained in a state of "No Known Bugs" (NKB). It is expected that this trickle of bug reports will continue for several months, during which time the program will get steadily more robust.

New features in MOPAC2007

PM6: This is the most recent NDDO method in MOPAC.

Features of MOPAC6

Geometry optimization

Methods

SCF Procedures

Derived Properties

Types of Species

Symmetry theory

Electric fields

Electrostatic Potential

Solvent effects

Electronic Excited states

Vibrations

Thermodynamics

Dynamics
Intrinsic Reaction
Coordinate
Reaction Paths
Saddle-Point Location

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.

NDDO Methods

The following NDDO methods are supported: MNDO, AM1, PM3, RM1, and PM6. For a list of elements available, see [Elements available](#).

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 a and b 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 a and b spin molecular orbitals, the associated a and b spin density matrices can be printed, as can the a-b spin density matrix.

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 DH_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 (b and g), Electric Field Second Harmonic Generation (EFISH), optical rectification, electrooptic Pockel's Effect, second and third harmonic generation. Where appropriate, averages are printed.

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-p* 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 (the degree of elastic behavior) 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 DH_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.

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.

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.

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.

Solvent effects

The effect of solvents can be modeled by Andreas Klamt's COSMO technique. In the COSMO method, both ground and excited state systems in solution can be modeled, and the geometries of solvated systems optimized.

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.

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 in a [FORCE](#) calculation

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.

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 DH_f can be monitored.

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 decent from the starting geometry to a stationary point.

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 DH_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.

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.)

Examples of Data-Sets and Output Files for MOPAC2007 Calculations

Please send requests for examples to MrMOPAC@OpenMOPAC.net
They will be added to this page as quickly as possible.

Getting started

[Geometry Optimization Data Set for Formic Acid](#)
[Output file for Formic Acid](#)
[Data Set for a single-point calculation of Formic Acid](#)

Transition state location

[Strategy](#)

MOZYME (Giant Molecules)

[Overview](#) and [Introduction](#)
[Simple MOZYME example](#)
[Eigenvectors and eigenvalues in MOZYME](#)
[Identifying charged sites](#)

[Generating a PDB file from a MOPAC data set](#)
[Automatically correcting net charge on system](#)
[Data set for Bacteriorhodopsin](#)
[Data set for a short section of DNA](#)
[Constrained Geometry Optimization of Crambin, 1CMB structure](#)

C.I. Calculations

[States resulting from one-electron excitations](#)

Vibrational Frequencies

Simple example: Formaldehyde, [data set](#), [output](#)

Use of SYMMETRY

In simplifying the geometry optimization of benzene
[Benzene data set using symmetry](#)
[Output from benzene calculation](#)

Solid State Calculations

[How to construct a data-set](#)

Generating data for Graphical User Interfaces

Use keyword [GRAPH](#) or [GRAPHF](#)
(for formatted data)

Polarizability Calculations

Simple example: p-nitroaniline, [data set](#), [output](#)

Theory used in MOPAC2009

[Semiempirical Theory](#)
[An SCF Calculation](#)
[Solid state capability](#)
[Multi-Electron Configuration Interaction](#)
[Geometry Optimization](#)
[Normal Coordinate Calculation](#)
[Dynamic and Intrinsic Reaction Coordinates](#)
[Point Group Theory](#)
[Level of Precision with MOPAC](#)
[Miscellaneous Electronic](#)
[Miscellaneous](#)

Semiempirical theory

[Approximations used in Semempirical Methods](#)
[Parameters used in Semiempirical Methods](#)
[Basic Roothaan-Hall Equations](#)
[Neglect of diatomic overlap integral](#)
[Neglect of three and four center integrals](#)
[One-center two-electron integrals](#)
[NDDO two-electron two-center integrals](#)
[Final Assembly of Two-Electron Two-Center](#)

[Integrals](#)

[The one-center one-electron integral \$H_{mm}\$](#)
[The two-center one-electron integral \$H_{mn}\$](#)
[Core-core repulsion integrals](#)
[The Self-Consistent Field Calculation](#)
[Calculation of \$DH_f\$](#)

Solid state capability

[The Cluster](#)
[The Madelung Problem](#)
[Piezoelectric Effect](#)
[Refractive Index](#)
[Derivatives](#)
[Charge Balanced Unit Cells](#)
[Geometry Specification for Band Structure](#)
[Calculations](#)
[Electronic Band Structure](#)

An SCF Calculation

[An SCF Calculation](#)
[Starting density matrix](#)
[Assembly of the starting Fock matrix](#)
[Diagonalization of the Fock matrix](#)
[Exercises involving eigenvectors](#)
[Iterating density matrix](#)
[Iterating Fock matrix](#)
[Calculation of heat of formation](#)

Multi-Electron Configuration Interaction

[Starting electronic configuration](#)
[Microstates](#)
[Permutations](#)
[Energy of microstates](#)
[Zero of energy used in MECI](#)
[Construction of secular determinant](#)
[States arising from various calculations](#)
[Spin angular momentum](#)

[Electronic Density of States](#)
[Brillouin Zone: Generation of Band Structures](#)
[Space-group operations](#)

[Calculation of spin-states](#)
[Choice of State to be Optimized](#)
[Quantum Numbers](#)
[Polarizability](#)
[Franck-Condon considerations](#)
[Definition of some C.I. Keywords](#)
[Degenerate States](#)

Geometry Optimization

[EigenFollowing](#)
[The BFGS function optimizer](#)
[Optimization of one unknown](#)
[Considerations in Geometry Optimization](#)

Normal Coordinate Calculation

[Calculation of Vibrational Frequencies](#)
[Calculation of the Hessian matrix](#)
[Mechanism of the frame in FORCE calculation](#)
[Vibrational Analysis](#)
[Reduced masses](#)
[Effective masses](#)
[Travel](#)
[Force Constants](#)
[Transition dipole](#)

Dynamic and Intrinsic Reaction Coordinates

[Description](#)
[Equations used](#)
[Specifying temperature in a DRC simulation](#)
[IRC](#)
[General description of the DRC and IRC](#)
[Option to allow only extrema to be output](#)
[Keywords for use with the IRC and DRC](#)
[Examples of DRC/IRC data](#)
[Output format for IRC and DRC](#)
[Test of DRC](#)

Point Group Theory

[Limitations](#)
[Representation of Point Groups](#)
[Identification of Point-Groups](#)
[Tolerance](#)
[Orientation of the Abelian groups \$C_{2v}\$ and \$D_{2h}\$](#)
[Molecular Orbitals](#)
[Normal Coordinates States](#)
[Atomic energy levels and States](#)

Level of Precision with MOPAC

[Fundamental Physical Constants](#)
[Various precision levels](#)
[Reasons for low precision](#)

Miscellaneous Electronic

[Koopmans' Theorem](#)
[Dipole moments](#)
[Bond Orders](#)
[Mulliken populations](#)

Miscellaneous

[Correction to the Peptide Linkage](#)
["Size" of a molecule](#)
[Gradients](#)
[Reaction paths](#)

[Localized orbitals](#)

[Localized M.O.s](#)

[Localization Theory](#)

[Oscillator Strength](#)

[Convergence in SCF Calculation](#)

[Causes of failure to achieve an SCF](#)

[Use of C.I. in Reaction Path Calculations](#)

[Sparkles](#)

[Capped Bonds](#)

[Algebraic Derivation of Overlap Integrals over](#)

[Slater Orbitals](#)

[Energies of Isolated Atoms](#)

[Reaction paths](#)

[Polarizability and Hyperpolarizability Calculation](#)

[Time-Dependent Hartree-Fock](#)

[COSMO](#)

[Parametric Molecular Electrostatic Potential](#)

[\(PMEP\)](#)

[Grid Calculation](#)

[Thermochemistry](#)

[Use of SADDLE Calculation](#)

[Gibbs free energy](#)

Bibliography

- [Elements](#)
- [Geometry](#)
- [Electronics](#)
- [General](#)

Elements

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 Be[150], B [99], F [100], Al [99], Si [101], P [102], S [103], Cl [104], K [135], Zn [105], Ge [106], Br [107], Sn [108], I [109], Hg [110], and Pb [111]. In 2004, the remaining main-group elements Na, Mg, K, Ca, Ga, As, Se, Rb, Sr, In, Sb, Te, Cs, Ba, Tl and Bi were published [138]. Earlier versions of MOPAC used parameters for Na [134], and K [135], but these have been replaced by newer parameters that are intended to be more accurate.

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], Mo [133], I [113], and Hg [120]. In 2004, the remaining main-group elements Li, Be, Na, Mg, K, Ca, Ga, As, Se, Rb, Sr, In, Sn, Sb, Te, Cs, Ba, Pb, and Bi were published [138]. Sparkle models for La, Ce, Pr[142], Nd[141], Pm, Sm, Eu[137], Gd[137], Tb[137], Dy[139], Ho[140], Er, Tm[143], Yb[144], and Lu and Th[145] have been developed.

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], and in 2004, the remaining main-group elements B, Na, K, Ca, Rb, Sr, Cs, and Ba were published [138].

RM1

A parameter set was developed to give improved results for systems containing the elements H, C, N, O, P, S, F, Cl, Br, and I only. For details, see *Rocha, G.B., R.O. Freire, A.M. Simas, and J.J.P. Stewart., RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. J. Comp. Chem., 27(10): 1101-1111 (2006).*

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].

PM6

PM6 parameters[148] are available for all elements from H to Bi, except the lanthanides from Ce to Yb. All the lanthanides[149] can be used, as their 3+ sparkles. To use them, add keyword [SPARKLE](#). The sparkle versions of Ln give good geometries.

All methods

Eight point-charges, called "sparkles" are provided, these have charges of +3, +2, +1, +1/2, -1/2, -1, -2, and -3. They were developed for use as counterions. No journal reference exists, the only [description of the sparkles](#) is in

this manual.

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 MNDO [1], AM1 [3], PM3 [4], RM1*, and PM6 [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].

*: Rocha, G.B., R.O. Freire, A.M. Simas, and J.J.P. Stewart., RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comp. Chem.*, **27**(10): 1101-1111 (2006).

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 [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.

Troubleshooting or Frequently Asked Questions

MOPAC2009

Downloading, installing, and getting started

License Issues

How to do specific types of calculations

Faults in MOPAC

MOPAC2009

- [What is MOPAC2009?](#)
- [How is MOPAC2009 related to MOPAC2007? Is MOZYME in MOPAC2009?](#)
- [Is MOPAC2009 public domain software?](#)
- [Does MOPAC2009 use any network connectivity?](#)
- [What forms of MOPAC2009 manual are there?](#)
- [Can I get the source code for MOPAC2009 so I can compile it?](#)
- [Why was my application for a license key refused with the message: "Remote host said: 554 refused mailfrom because of SPF policy"](#)
- [Are there plans to develop a parallel version of MOPAC2009?](#)

Downloading, installing and getting started

- [I've requested a license key, but nothing happened. What has gone wrong?](#)
- [When MOPAC is run, it gives the message "Password invalid," what's wrong?](#)
- [Why does MOPAC gives the message "permission denied" on my Linux machine.](#)
- [I can't put MOPAC into the default location.](#)
- [Why will MOPAC not run even though the password was installed correctly?](#)
- [My Linux OS is 64 bits. MOPAC does not run. Why?](#)
- [I'm completely new to MOPAC, how do I get started?](#)
- [How do I run MOPAC on several machines?](#)
- [How do I visualize results using Jmol?](#)
- [When I install a new version of MOPAC, why does the Version Number not change?](#)

License Issues

- [Who can I share my license key with?](#)
- [My license is for a year. What happens when it expires?](#)

How to do specific types of calculations

- [What are the defaults if no keywords are used?](#)
- [How do I orientate the system in a special way?](#)
- [Can the output be sent to a different directory?](#)
- [How do I run a batch job?](#)
- [How do I get the atomic orbital coefficients of the HOMO or LUMO?](#)
- [Can MOZYME calculations give eigenvalues?](#)
- [How do I freeze an atom's position during a geometry optimization?](#)
- [When should I use C.I. and when should I use UHF?](#)

- [In a DRC calculation, what is the relationship of kinetic energy and temperature?](#)
- [Why does MOPAC sometimes have difficulties with geometry optimization?](#)
- [How do I do a geometry optimization and force constant calculation in one run?](#)
- [How can the re-orientation of a molecule be prevented in a polarizability calculation?](#)
- [How do I run multiple jobs, with a pre-set maximum number of jobs running at any one time?](#)
- [Can you help me with my research?](#)

Faults in MOPAC

- [What does, "Internal Error-Unexpected error file "E:\forrtl\build\qwnbuild\qwgswc.c", line 1051" mean?](#)
- [Why does the gradient sometimes increase suddenly during a geometry optimization?](#)
- [Why do net charges not add to exact integer values?](#)
- [Why is the sign of the gradient different from that of the forces in, e.g., Gaussian 03?](#)
- [Why cannot the precision of the gradients in the AUX file be increased?](#)
- [I've found a bug in MOPAC. How do I report it to get it fixed?](#)

MOPAC2009

How is MOPAC2009 related to MOPAC2007? Is MOZYME in MOPAC2009?

MOPAC2009 contains the MOZYME function. With that one exception, MOPAC2009 is identical to MOPAC2007

MOPAC2007 was a re-write of MOPAC 6, and included the new method PM6. It is owned by SCC. The main differences between MOPAC2007 and MOPAC2002 are:

MOPAC2002 contains MOZYME, the Localized Molecular Orbital method for rapidly calculating the electronic structures of large molecules. MOPAC2007 did not contain MOZYME.

MOPAC2002 can run MINDO/3, intersystem crossing, and the NDDO method PM5. MOPAC2007 did not contain these functions.

For MOPAC2002, the practical upper limit for 1SCF is about 20000 atoms. For MOPAC2007 it was about 1000 atoms.

Is MOPAC2009 public domain software?

No. MOPAC2009 is owned by Stewart Computational Chemistry. Executables are made available at no cost to academic users, and source code will be made available to selected groups.

Does MOPAC2009 use any network connectivity?

MOPAC2009 does not use any network connectivity.

If you chose to make the executable available over a network, then you have control of that functionality, but the program itself does not have any built-in network functionality.

MOPAC2009 is designed to be completely "clean" - it does not modify the Register (in WINDOWS) or any parts of the kernel (in Linux or UNIX). This is because MOPAC2009 is intended for use in a wide range of environments and many are very sensitive about security.

What forms of MOPAC2009 manual are there?

The MOPAC2009 Manual exists in only one form, the on-line manual you are looking at just now. There are no PDF or hard-copy manuals. This was a deliberate policy decision: Writing and maintaining more than one manual would involve a lot of work, and the on-line manual should be sufficient for most users. If you need hard copy - print the individual pages involved.

Can I get the source code for MOPAC2009 so I can compile it?

In general, the answer is, "No!" Currently, MOPAC2009 is made available in executable form for use on Windows, Macintosh, and Linux platforms only.

Only in exceptional circumstances will source code be made available for this type of work. Most executables must originate from OpenMOPAC.net in order to ensure that all executables are up-to-date and have undergone quality assurance testing.

If a specific compilation is needed, and are the resources required are provided to SCC, and attempt will be made to do the port. The resources required are: a computer, running the target operating system, and an installed and working FORTRAN 90 compiler. If the port is successful, and MOPAC2009 is validated for that OS, then it will be maintained on OpenMOPAC.net along with the other operating systems it currently runs on. Because this procedure obviously involves a significant expense, before any action is taken, please discuss this with SCC.

Why is my application refused with the message: "Remote host said: 554 refused mailfrom because of SPF policy"

This message will be sent if the ISP for OpenMOPAC.net has reason to believe that your E-mail address might not be legitimate. See http://www.openspf.org/FAQ/What_is_SPF for more information. If your application for a license key was refused because of that message, please send an E-mail to MrMOPAC@OpenMOPAC.net.

Are there plans to develop a parallel version of MOPAC2009?

No. With the addition of the MOZYME function, MOPAC2009 is much faster than MOPAC2007 for modeling large systems. The [MOZYME](#) function is quite complicated, and parallelizing it would be difficult. A lot of effort had been expended in parallelizing earlier versions of MOPAC, and with few exceptions, these attempts had not been successful. Only very specific calculations would benefit from parallelization. These include the generation of the Hessian for vibrational frequencies and the effect of applied electric fields in the [STATIC](#) calculation of polarizability.

Downloading, installing and getting started

I've requested a license key, but nothing happened. What has gone wrong?

When you request a license key, an E-mail message will be automatically sent to the address you used in the form. If you do not get an acknowledgment message after submitting the form, then there is a problem with the E-mail address you provided. No E-mail help can be given on this, since messages cannot be sent to you.

When MOPAC is run, it gives the message "Password invalid," what's wrong?

MOPAC will only work after the password has been correctly recognized. The "Password invalid" message is given when both:

- (A) The password has not been installed, and
- (B) The name of the data set being run is not <password>.mop

To correct this fault, make a data set called <password>.mop, e.g., if your password is 1234567a12345678 then make a data set called 1234567a12345678.mop, and open it with the MOPAC executable. MOPAC should produce a black screen with white text. Press <return> two times, followed by a space the the word "yes" (without the quotation marks). then press <return> once more. Exit from MOPAC. The password should now be installed.

Why does MOPAC gives the message "permission denied" on my Linux machine.

MOPAC needs to read the license key, which is stored in the same subdirectory as the MOPAC2009.exe executable. If the directory or the license key are protected against reading, then the "permission denied" message will be seen. To correct this, the subdirectory and file must have read and write permission. Use the "chmod" command to set the permissions, the simplest is "chmod 777 /opt/mopac/". If the "chmod" command cannot be

used, put MOPAC in a subdirectory you have control over, see the next question..

I can't put MOPAC into the default location.

By default, the MOPAC executable should be put in the Windows folder C:/Program Files/MOPAC or the Linux or Macintosh subdirectory /opt/mopac. If necessary, the executable can be put somewhere else. If that is done, then the location of the MOPAC2009 license key must be defined using the environmental variable MOPAC_LICENSE. If MOPAC_LICENSE is not defined, then MOPAC behaves as if it was defined as "C:\Program Files\MOPAC\" or "/opt/mopac/"

Windows:

(start) => Settings => Control Panel => System => Advanced => Environmental Variables => User variables for <user> => New => Variable name = MOPAC_LICENSE Variable value = C:\Program Files\MOPAC\ => OK => OK => OK

Linux:

Edit .bashrc to add the line(s):

```
export MOPAC_LICENSE=/home/mydirectory/
alias mopac='/home/mydirectory/MOPAC2009.exe'
```

or

Edit .cshrc to add the line(s)

```
setenv MOPAC_LICENSE /home/mydirectory/
alias mopac '/home/mydirectory/MOPAC2009.exe'
```

Macintosh:

Edit .profile to add the line(s):

```
export MOPAC_LICENSE=/home/mydirectory/
alias mopac='/home/mydirectory/MOPAC2009.exe'
```

A re-boot might be necessary in order for the environmental variable to be recognized.

Why will MOPAC not run even though the password was installed correctly?.

When MOPAC2009 is run, the message 'File "C:\Program Files\mopac\password for MOPAC2009" is currently inaccessible' appears. The license key is correctly installed, but MOPAC2009 does not have permission to read the license key file. Change the permissions on the file "C:\Program Files\mopac\password for MOPAC2009" so that you can open it using, e.g., notepad.

My Linux OS is 64 bits. MOPAC does not run. Why?

Most likely you will need to install the 32-bit glibc and gcc libraries.

I'm completely new to MOPAC, how do I get started?

Start by running a simple calculation. Open the page "[Geometry Optimization Data Set for Formic Acid](#)". This describes the MOPAC data set for HCOOH. Make up a file for this data set either by (A) downloading the data set, (B) highlighting the data set, and copying it into a file called "formic_acid.mop", or (C) open notepad and typing up the data set. Save it as "formic_acid.mop". Run the data set using MOPAC2009 by following the instructions in [RUNNING MOPAC](#). To see what the results mean, look at [Formic Acid output](#).

Most users prefer to view their data sets and output files using a [Graphical User Interface](#). Several are available at no cost, but of course they are less powerful than the commercial programs.

How do I run MOPAC on several machines?

A single calculation cannot be run on more than one machine. The code in MOPAC only allows a single processor to be used for a single calculation. However, if several calculations are to be run, then each calculation could be started on a different machine. That is, MOPAC jobs can be run in coarse-grain parallel mode; this is 100% efficient. In order for MOPAC to work on a machine, the license key must exist on that machine. Copy the folder that contains the license key from the original computer to the same position in each of the other computers in the cluster.

If Windows, then copy "[C:\Program Files\MOPAC](#)" from the original computer to each of the other computers in the cluster.

If Linux or Macintosh, then copy `"/opt/mopac"` from the original computer to each of the other computers in the cluster.

Administrative point: If all the machines are permanently connected together, then delete all copies of MOPAC2009.exe or MOPAC2009.out, except one. Use the single remaining executable on all the machines. That way, if you need to update the executable you only need to update the single file. Note, however, that the license key file must exist on each machine.

How do I visualize Molecular Orbitals using Jmol?

First, download Jmol, version 11.1.28 or higher, and MOPAC2009, version 7.101 or higher. Open the Jmol package and put the contents into a folder (call this folder "whateverdirectoryjmolisin") Run a MOPAC job with keyword GRAPHF. This will generate a file ending in .mgf (MOPAC Graphics File) Then:

EITHER:

At the DOS prompt, open the .mgf file using the command: `whateverdirectoryjmolisin\jmol.jar filename.mgf` Within Jmol, do a left-click so that you can rotate the molecule, then do a right click. Navigate Surfaces => Molecular Orbitals

OR

In Windows:
Create a utility called Jmol_MOPAC.cmd The utility contains one line: `whateverdirectoryjmolisin\jmol.jar %1`
Right-click on a .mgf file, and browse to Jmol_MOPAC.cmd
Put a check in the option "Always use the selected program to open this type of file"
Click on "OK"
Within Jmol, do a left-click so that you can rotate the molecule, then do a right click. Navigate Surfaces => Molecular Orbitals
After this is done, a double click on a .mgf will automatically open it.

The result should look like this: [CH2O](#)

When I install a new version of MOPAC, why does the Version Number not change?

Every version of MOPAC has a unique version number. The first few lines of the output should look like this:

```
*****
** Site#: nnnn          <TextTextTextTextTextText>          Version 7.263W **
*****
** Cite this work as: MOPAC2009, James J. P. Stewart, Stewart Computational **
** Chemistry, Version 7.263W web: HTTP://OpenMOPAC.net   Days remaining: nnn **
*****
```

The version number is near the end of line 2 of the output, and should be of form n.nnnW (for Windows

platforms) or n.nnnL (for Linux platforms). The decimal part of the version refers to the Julian date when the executable was made. In this example, the version was built on the 263rd day of the year 2007. When a new executable is downloaded, the ZIP file will have the version number in the name of the file. Before opening a distribution ZIP file, look at the version number in the file name - make sure it is correct. In some instances, the version number printed by the executable will be higher than that in the ZIP file - this is not a fault - it just indicates a minor, undocumented, bug-fix has occurred since the last official release.

A problem sometimes occurs when a new version is installed but after it is installed and MOPAC is run, the resulting output shows that the old version is still being used. To fix this problem, first delete the old executable before opening the ZIP file containing the new MOPAC. Then try to run MOPAC2009. If MOPAC2009 still executes, then you have two or more old copies of the executable of MOPAC on your machine, and when you run MOPAC, it's one of these old versions that's running.

To correct this problem, delete every copy of the MOPAC2009 executable until it will no longer run. Then install the new executable from the ZIP file.

License Issues

Who can I share my license key with?

Each academic license key is valid for all users at the educational site. Everyone at the site can install MOPAC2009 on their computers at the site and at home. Please do NOT request separate license keys, because issuing each key requires some work. The keys may NOT be freely distributed off-site.

My license is for a year. What happens when it expires?

The original plan was that users were encouraged to stay up to date with their copy of MOPAC. To encourage this, each new version would re-set the expiry date to one year from the date of creation of the new program. But the number of errors reported and found during testing has been less than expected (not enough testing?) so the re-set of the expiry date has been discontinued. Users are still encouraged to stay up-to-date with their copy of MOPAC2009 even though all recent changes have only been minor bug-fixes or enhancements. By staying up-to-date you can be sure that all the bugs that have already been reported have been fixed.

All licenses are for at least 8 months. Long before they start to run out, a new policy will be in place.

How to do specific types of calculations

What are the defaults if no keywords are used?

If no keywords are supplied, then the following options are used:

PM6	(Default method)
RHF	(Not UHF)
Singlet or doublet	(Default spin states)
Geometry optimization	(Not a single point)
T=2D	(Default time: 2 CPU days)
DUMP=2H	(Checkpoint files (.res and .den) written every two hours)

How do I orientate the system in a special way?

The simplest way to set up the system so that it is orientated the way you want it, is to use [Cartesian coordinates](#). An alternative is to use three dummy atoms at the start of an [internal coordinate](#) Z-matrix, one at the origin, one along the "x" axis, and one along the "y" axis:

```

XX  0.0 0  0.0 0  0.0 0
XX  1.0 0  90.0 0  0.0 0  1 0 0
XX  1.0 0  90.0 0  90.0 0  2 1 0

```

and then use internal coordinates to position the real atoms relative to the dummy atoms. This is not as easy as

using Cartesian coordinates.

Can the output be sent to a different directory?

The output is always sent to the directory that the data set was in. To send the output to a different directory, use the following procedure:

(A) Copy the data files to the directory where you want the output to go to. If the data sets were in directory A and you want the output to be in directory B, copy the data sets from A to B. Do not move the data sets - copy them - so there are two sets of data files, one in A and one in directory B.

(B) Run the data sets in directory B using MOPAC.

(C) Delete the data sets in B. Now the output of the data sets in directory A will be in directory B.

How do I run a batch job?

To run a batch job, make up a command file, e.g. MOPAC2009_Batch_job.cmd, and paste in the following lines:

```
REM
REM  How to run several MOPAC2009 jobs with one command.
REM
REM  Use the program "MOPAC2009 for all windows" version of MOPAC.
REM  Do NOT use the program "MOPAC2009 for WINDOWS XP" - if you do
REM  it will request a finish dialogue ("Program finished with exit code 0.
REM  Exit Window?")
REM
REM  Use commands of the following type to run all the jobs.
REM
call MOPAC2009.exe job1
call MOPAC2009.exe job2
call MOPAC2009.exe job3
call MOPAC2009.exe job4
```

How do I get the atomic orbital coefficients of the HOMO or LUMO?

Add keyword [VECTORS](#) to the data set. From the output, identify the number of occupied levels. The atomic orbital coefficients of the HOMO are then given by the eigenvector with that number on the line "Root No." in the "EIGENVECTORS" array. The LUMO is the next higher eigenvector. For example, if formaldehyde (CH₂O) is being calculated, then there are 6 doubly occupied levels. The sixth eigenvector is the HOMO and the seventh eigenvector is the LUMO.

Can MOZYME calculations give eigenvalues and eigenvectors?

Yes, if [EIGEN](#) and [VECTORS](#) are both present.. The operation to generate eigenvalues and eigenvectors uses matrix algebra, and therefore will need a lot of CPU time and memory for large systems such as proteins. Do NOT use this pair of keywords as the default.

How do I freeze an atom's position during a geometry optimization?

There are two ways to do this, depending on what is meant by "freeze."

To freeze the Cartesian coordinate of an atom.

To do this, the atom's position must be defined using Cartesian coordinates.

Consider the following entry in a Z- matrix::

```
H -1.31 1  0.12 1  1.89 1
```

The atom's position is defined using Cartesian coordinates, because there is no connectivity given. In this entry, the hydrogen atom is located at Cartesian coordinate: (-1.31, 0.12, 1.89), and the three "1"s on this line mean "Optimize the X, Y, and Z coordinates of the atom." If you want to freeze the position of the atom, change the three "1"s to zeros, e.g.:

```
H -1.31 0  0.12 0  1.89 0
```

To freeze a bond-length, angle, or dihedral:

Consider the following entry in a Z-matrix::

```
H 1.092  1 120.615  1 179.979  1 10 9 11
```

In this entry, the hydrogen atom is 1.092 Angstroms from atom 10, makes an angle of 120.615 degrees with atom 9, and a dihedral of 179.979 degrees with atom 11.

The three "1"s on this line mean "Optimize the bond length, angle, and dihedral." If you want to fix the bond-length at 1.092 Angstroms, then the entry would be:

```
H 1.092  0 120.615  1 179.979  1 10 9 11
```

To optimize only the angle, the entry would be:

```
H 1.092  0 120.615  1 179.979  0 10 9 11
```

When a bond length, angle, or dihedral angle is frozen, the atom must be in internal coordinates. This does not mean that the other atoms in the system must also be in internal coordinates - they can be in internal or Cartesian coordinates, as the user chooses. The only requirement is that the atoms used in the connectivity are specified before the atom whose coordinates are frozen is specified.

For more information, see [Optimization Flags](#).

When should I use C.I. and when should I use UHF?

The following generalizations should be used as a guide: Use C.I. when modeling excited states or when there is a need for spin quantization. Use UHF when optimizing geometries, doing reaction paths, calculating normal modes, etc.

Considerations: UHF is normally significantly faster than SCF-C.I., and the gradients are mathematically more stable. So where the geometry is important - ground state or transition state - vibrational frequencies, DRC or IRC, unless there is an overriding reason to use ROHF (RHF-C.I.), it's best to avoid it. In organic chemistry, particularly in biochemistry, UHF is best for modeling odd-electron systems, i.e., radicals. In inorganic chemistry, RHF-C.I. is often the preferred method, particularly when the electronic state is important, thus in octahedral chromium-III complexes, the ground state, $^4A_{2g}$, should be defined using the OPEN(3,3) keyword. RHF-C.I. also allows high and low spin to be modeled.

In a DRC calculation, what is the relationship of kinetic energy and temperature?

For a system at the equilibrium geometry, the kinetic energy in a DRC calculation is equal to the enthalpy or internal energy of the system, or $DH_f(T) - DH_f(0)$, where "T" is the temperature of the system.

Normally, the system is not at the equilibrium geometry, so the previous statement can be generalized to: The kinetic plus potential energy of a DRC system is equal to $DH_f(T) - DH_f(0)$, where "T" is the temperature of the system.

So, if you know the heat of formation of the system (that is, the heat of formation of the optimized geometry) and the total energy in a DRC calculation (Heat of formation plus potential plus kinetic energy) then the internal energy can be readily calculated as the difference of these two quantities.

The temperature of the system can then be worked out by running a [THERMO](#) calculation, and finding what temperature corresponds to the enthalpy of the system in the DRC. Consider formaldehyde: If, in a DRC calculation using PM6, the potential plus kinetic energies equals 2.403 kcal/mol, then the temperature of the system would be 298K. If a THERMO calculation were to be run on formaldehyde, the enthalpy at 298K would be 2.403 kcal/mol.

Why does MOPAC sometimes have difficulties with geometry optimization?

Sometimes a geometry optimization goes wrong, and MOPAC starts to give messages of the type:

NUMERICAL PROBLEMS IN BRACKETING LAMDA

The commonest cause of this is the "big ring" problem - the data set is in internal coordinates and the system contains large rings. Small changes in internal coordinate angles or dihedrals make large changes in the distances

of some pairs of atoms that should be bonded together. If a calculation gives this message, the easiest way to correct it is to add keyword "[XYZ](#)". Do not use this keyword if you are using keyword [SYMMETRY](#) or if any coordinates are not marked for optimization.

How do I do a geometry optimization and force constant calculation in one run?

Before a force calculation is run, the geometry must be optimized. This is most easily done in two steps: The geometry is optimized in one calculation, then the [FORCE](#) calculation is run using the optimized geometry, in a second calculation. Both calculations are run in one job. This is best illustrated by the following example. In this, the geometry of formaldehyde is first optimized, then the force calculation run. Keyword [OLDGEO](#) instructs the FORCE calculation to use the geometry from the previous calculation.

```
SYMMETRY precise
Formaldehyde (Geometry optimization)
```

```
O 0.0 0      0.0 0      0.0 0
C 1.2 1      0.0 0      0.0 0      1 0 0
H 1.0 1      122.0 1      0.0 0      2 1 0
H 1.0 0      122.0 0      180.0 0      2 1 3
```

```
3 1 4
3 2 4
```

```
oldgeo force
Formaldehyde (using optimized geometry from the previous step)
Last line of data set.
```

How can the re-orientation of a molecule be prevented in a polarizability calculation?

By default, a molecule in a polarizability calculation is rotated so that it is oriented along the principal moments of inertia. To prevent this re-orientation being done, add keyword [LET](#).

How do I run multiple jobs, with a pre-set maximum number of jobs running at any one time?

Write a command script (.cmd in windows, .csh in Linux and Mac) to run each job in turn. For example, to run jobs "test_1.mop", "test_2.mop", and "test_3.mop", the following Windows script might be used:

```
"C:/Program Files/MOPAC/MOPAC2009.exe" test_1.mop
"C:/Program Files/MOPAC/MOPAC2009.exe" test_2.mop
"C:/Program Files/MOPAC/MOPAC2009.exe" test_3.mop
```

This would run test_1, then when it finished, it would run test_2, then when that finished it would run test_3.

A utility "Wait.exe" has been written. It runs continuously but uses very little CPU time. While Wait.exe is running, it blocks the next instruction, here a call to run a MOPAC job, until the number of MOPAC jobs running has dropped below a pre-set limit. For [Windows](#), the batch file would look like this:

```
echo off
call M:\utility\MOPAC.cmd Test_1
call M:\utility\MOPAC.cmd Test_2
call M:\utility\MOPAC.cmd Test_3
```

and the MOPAC.cmd would look like this:

```
M:\utility\wait.exe G:\ MOP 2
call start /b /low M:\utility\Not_XP\mopac2009.exe %1
```

In this example, Test_1 and Test_2 would start running, but Test_2 would be prevented from running until either of the MOPAC jobs that were currently running finished.

For [Mac](#), the batch file would look like this:

```
~/utilities/run_MOPAC.csh "Test_1"
~/utilities/run_MOPAC.csh "Test_2"
~/utilities/run_MOPAC.csh "Test_3"
```

where run_MOPAC.csh contains just one line:

```
~/bin/wait.exe "/Users/name" MOPAC 2; nice -n 20 ~/bin/MOPAC2009.exe "$1" &
```

Of course, change the paths to suit local conditions. For more details on Wait.exe, run it at the command prompt, without arguments.

Can you help me with my research?

No. A small amount of research is done, but that work is restricted to investigating how well or how badly MOPAC performs in certain types of simulations. Almost all the work at SCC is directed towards making MOPAC2009 a better tool for research. If your request has a side-effect of improving MOPAC, including improving the manual, then help might be provided, otherwise if you really want help, contract work can be done, but be warned this is expensive because it takes time away from the primary task! The best analogy is that the objective of SCC is to make a good television, not to produce the show that is seen on the set.

Faults in MOPAC

What does, "Internal Error-Unexpected error file "E:\forrtl\build\qwnbuild\qwgswc.c", line 1051" mean?

This error is caused by a fault in Microsoft's FORTRAN compiler. It affects QuickWin programs only. There are two workarounds:

First, if the output file is okay, ignore the error message. When you get the message, "Program Terminated with exit code 0," that indicates a normal exit. Click on the "yes" button.

Second, replace your MOPAC executable with the "all windows" executable. The fault you mention will not occur with the "all windows" form.

Why does the gradient sometimes increase suddenly during a geometry optimization?

Sometimes, particularly during a geometry optimization, the gradient will unexpectedly increase then drop back down again. Why does this happen? The accurate answer is, I don't know. There are two possibilities: (A) The geometry optimizer changed the geometry by a large amount in an energetically unfavorable direction. This resulted in both the gradient and DH_f increasing. The information obtained from that faulty step then allowed the geometry optimizer to block that specific direction for future steps. (B) There is a bug in the calculation of gradients. If the bug is in the gradient calculation, it would be very difficult to find.

If the gradient and DH_f both rise then fall back, and the rest of the geometry optimization looks okay, then ignore the fluctuation.

Why do net charges not add to exact integer values?

In general, when the net atomic charges on all atoms are added, the result is not an exact integer, but instead is in error by an integer multiple of $10^{(-6)}$. This is a consequence of the way number are represented in files.

All numbers printed in MOPAC are ill-defined in the last decimal. Consider a regular triatomic system with net charge of +1. Using double precision (16 significant figure precision) the charge on each atom would be:

```
0.3333333333333333
0.3333333333333333
0.3333333333333333
```

In the NET ATOMIC CHARGES section, this would be printed as:

```
0.333333
0.333333
0.333333
```

for a total charge of 0.999999, i.e., an error of -0.000001, or -1×10^{-6} for 6 figure precision. If the charges were printed to 8 significant figures, the error would be -1×10^{-8} , and if to 3 figures, the error would be -1×10^{-3} .

Using the standard rules for addition of real numbers of finite precision, an error of a small integer times 10^{-6} for a net charge for a neutral system of n atoms would be completely reasonable. Conversely, an error in the net charge of n times 10^{-6} would be impossible.

Why is the sign of the gradient different from that of the forces in, e.g., Gaussian 03?

The sign of the forces is the negative of the sign of the gradient. Gradient is dE/dx , force is $-dE/dx$. Gradient is the slope of the potential energy surface, force is in the downhill direction.

Why cannot the precision of the gradients in the AUX file be increased?

They can. See PRECISION in [AUX](#).

I've found a bug in MOPAC. How do I report it to get it fixed?

First, make sure that your copy of MOPAC is up-to-date. If it's not, then download a new copy and check that the bug is still present. If it is, then please send an E-mail to MrMOPAC@OpenMOPAC.net describing the bug, and include a data set that illustrates the bug, and output from that data set, as attachments

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.

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 HAS A ZERO MASS (FATAL)

In a [FORCE](#) calculation, the mass of an atom is zero. To correct this, [give the mass explicitly](#).

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=n.nn](#).

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 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.

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 (TS). 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.
[The sizes of various C.I. matrices](#), implied by C.I.=*m* appear elsewhere.

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 [USCF](#) 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 *NW*...

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.

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 [EXTERNAL=](#).

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](#).

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 [SCF](#) 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 THRE... (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, RM1, 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 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 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.

ERRORS DETECTED IN CONNECTIVITY (FATAL)

The [connectivity in the MOPAC internal coordinate Z-matrix](#) is faulty.

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 [NLLSO](#), 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.

FILE "C:\program files\mopac\password for MOPAC2007" is currently inaccessible. Correct this fault before continuing (FATAL)

MOPAC2007 attempted to read the license key file. The file exists, but MOPAC is unable to read it. Check the permissions and ownership of the file, and make sure that the file can be read. (Try opening it with notepad. If it cannot be opened, then MOPAC can't read it.)

FILE file.den is missing (FATAL)

An attempt has been made to read in an old density matrix (*fine.den*), but this file apparently does not exist in the subdirectory. Correct the error and re-run.

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 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 NOT properly optimized

See "TRUST RADIUS NOW LESS THAN "
 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.

If QPMEP is used, then PMEP must also be present (FATAL)

[QPMEP](#) is a keyword that modifies a [PMEP](#) (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\(*n*₁,*n*₂\)](#) has been used, but for an even-electron system *n*₁ was specified as odd or for an odd-electron system *n*₁ 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 [PM3](#) 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.

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.

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.

MISSING VAN DER WAALS RADIUS *Chemical-symbol* (FATAL)

In the `COSMO` method, van der Waals radii are used. If a VDW radius is missing, it can be supplied by use of [VDW\(text\)](#).

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=name](#). These comprise the symbols given in the following Table.

Table: 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

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.

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 [QSCF](#) and [INT](#) or [XYZ](#) to get rid of the dummy atoms), and re-run.

NO POINTS SUPPLIED FOR REACTION PATH (FATAL)

A reaction path calculation is indicated by a '-' 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).

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 OPEN-SHELLS ALLOWED IN C.I. IS LESS THAN THAT SPECIFIED BY OTHER KEYWORDS (FATAL)

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, nn 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

✕ `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*.

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.

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 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.nnn](#) and [STEP2=n.nnn](#).

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 [has been parameterized](#). 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.

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`.

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 [NOSYM](#).

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.

RESTART FILE EXISTS, BUT IS CORRUPT (FATAL)

In a [BEGS](#) 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.

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`, recompile, 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 [EXTERNAL=](#).

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.

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 [POWSO](#).

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 [OSCE](#), 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.nn`](#), `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 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 172,800 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 override the [default options](#) by adding [LET](#) and [DDMIN=0.0001](#), another option is to specify [RMIN=-10](#). In CAChe, the equivalent

message is "Geometry NOT properly optimized!"

This warning most often occurs when COSMO is being used, because the finite mesh used in defining the solvent accessible surface is, of its nature, finite. This introduces unavoidable artifacts into the PES, which, in turn, causes the trust radius to become very small.

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 [HESS=0](#). Remove `HESS=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."

Unable to allocate memory in subroutine ...(FATAL)

An error occurred while trying to create an array. The commonest reason for this error is that the array is very large, and there is not enough memory allowed. There may be limitations imposed by the operating system. Windows has limits in the order of 250Mb to 2Gb. Try increasing the memory available to the process to 2 Gb (real plus virtual). If that does not work, try avoiding the function that causes the error. If MOZYME is available, try using that. If the array is greater than 2^{31} elements, the array size will be inaccurate (negative), in which case there is nothing that can be done.

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.

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.

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](#).

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.

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

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.

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. McIver. Rapid Geometry Optimization for Semiempirical Molecular Orbital Methods. *Chem. Phys. Lett*, 10:303, 1971.
- 10** A. Komornicki and J. W. McIver. 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. Aleman, 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. Aleman, 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 ÎLevel Shiftâ 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 Cyclopropylcarbiny 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. Zebisch. AM1 Calculations for Compounds Containing Boron. *Organometallics*, 7:513-521, 1988.
- 113** M. J. S. Dewar and E. G. Zebisch. 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. Freunscht. 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", *Chemical Society (Great Britain), Quarterly Reviews*, London. Vol. 13, pages 183-214, (1959).
- 133** Alexander A. Voityuk and Notker Rösch ; AM1/d Parameters for Molybdenum, *The Journal of Physical Chemistry A*; 2000; 104(17); 4089-4094.
- 134** *Angew. Chem. Int. Ed. (Engl.)* 29, 10442 (1990).
- 135** *J. C. S. Chem. Comm.* 765, (1992).
- 136** M. J. S. Dewar and J. A. Hashmall, C. G. Venier, *J. Am. Chem. Soc.*, 90, 1953 (1968); M.J. S. Dewar and N. Trinajstić, *J. Chem. Soc., A*, 1220 (1971).
- 137** G. B. Rocha, R. O. Freire, N. B. da Costa Jr., G. F. de Sá and A. M. Simas, *Sparkle Model for the Calculation of Lanthanide Complexes: AM1 Parameters for Eu(III), Gd(III), and Tb(III)*, *Inorg. Chem.*, 44, 3299:3310 (2004).
- 138** J. J. P. Stewart, *Optimization of Parameters for Semiempirical Methods IV: Extension of MNDO, AM1, and PM3 to more Main Group Elements*, *J. Mol. Model.* 10, 155-164 (2004).
- 139** Nivan B. da Costa Jr., Ricardo O. Freire, Gerd B. Rocha, Alfredo M. Simas, *Sparkle model for the AM1 calculation of dysprosium (III) complexes*, *Inorg. Chem. Comm.* 8, 831:835 (2005).
- 140** Nivan B. da Costa Jr., Ricardo O. Freire, Gerd B. Rocha, Alfredo M. Simas, *Sparkle/AM1 modeling of holmium (III) complexes*, *Polyhedron* (in press).
- 141** Cristiano C. Bastos, Ricardo O. Freire, Gerd B. Rocha, Alfredo M. Simas, *Sparkle model for AM1 calculation of neodymium(III) coordination compounds*, *Journal of Photochemistry and Photobiology* (in press)
- 142** Ricardo O. Freire, Nivan B. da Costa Jr., Gerd B. Rocha, Alfredo M. Simas, Modeling lanthanide coordination compounds: Sparkle/AM1 parameters for praseodymium (III), *Journal of Organometallic Chemistry*, 690, 4099–4102 (2005)
- 143** Ricardo O. Freire, Gerd B. Rocha, Alfredo M. Simas, *Modeling rare earth complexes: Sparkle/AM1 parameters for thulium (III)*, *Chem. Phys. Lett.* 411, 61–65 (2005).
- 144** Ricardo O. Freire, Nivan B. da Costa Jr., Gerd B. Rocha, Alfredo M. Simas, *Modeling Lanthanide Complexes: Sparkle/AM1 Parameters for Ytterbium (III)*
- 145** Rocha, G.B., *D.Sc. Thesis, Departamento de Química Fundamental da Universidade Federal de Pernambuco*,

Brazil (2002). Departamento de Química Fundamental, Universidade Federal de Pernambuco, 50.590-470, Recife, PE, Brazil, "Development and parametrization of semiempirical models for the calculation of organic molecules and complexes of lanthanide and actinide ions."

146 Rocha, G.B., R.O. Freire, A.M. Simas, and J.J.P. Stewart., *RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I*. *J. Comp. Chem.*, **27**(10): 1101-1111 (2006).

147 MOPAC2009, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, [HTTP://OpenMOPAC.net](http://OpenMOPAC.net) (2008).

148 Stewart J. J. P., *Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements* *J. Mol. Modeling* 13, 1173-1213 (2007).

149 R.O.Freire and A.M.Simas, to be submitted

150 M. J. S. Dewar and H. S. Rzepa, "Ground States of Molecules. 39. MNDO Results for Molecules Containing Beryllium" *J. Am. Chem. Soc* **100**, 777-784 (1978).

General Description of MOPAC2007

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](#)], AM1 [[3](#)], PM3 [[4](#)], PM6, [RM1](#), 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 unique, 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, MOPAC was an obvious contender.

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 that 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 functionality, 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, 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, E-mail: MrMOPAC@OpenMOPAC.net.

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.

[List](#) of readily available and supported [G.U.I.s](#).

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. DH_f . See [Semiempirical Theory](#).

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

Various time-dependent phenomena can be modeled. These Dynamic Reaction Coordinates include following a normal coordinate vibration, modeling a chemical reaction path, and running dynamics trajectories. A time-independent mass-weighted path known as the Intrinsic Reaction Coordinate can be generated for reactions.

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.

Installing MOPAC2009 instructions

To install the MOPAC2007 executable, please read and follow the instructions in the file "Installation instructions.txt" in the ZIP file you downloaded from OpenMOPAC.txt.

The instructions for installing MOPAC2009 depend on the platform MOPAC2009 is to be run on, and may change from time to time. So instead of giving instructions here in the manual, it is safer to put the instructions with the executable.

Installing and compiling MOPAC2009 instructions

The distribution of MOPAC2009 source is restricted to sites that have signed an End User License Agreement. After the EULA has been received at SCC, source will be sent electronically to the site. Please indicate which of the two platforms, WINDOWS or Linux, that you want to compile MOPAC2007 on.

All the files are in seven folders:

Folder	Contents
MOPAC	mopac.f90, to_screen.f90, jdate.f90, a Makefile, and CVF build files for MOPAC2007
Subroutines	CVF build files for compiling all the subroutines
src_subroutines	Source code for all the subroutines
Modules	CVF build files for compiling all the modules
src_modules	Source code for all the modules (These files end in "_C.f90")
Interfaces	CVF build files for compiling all the interfaces
src_interfaces	Source code for all the interfaces(These files end in "_I.f90")

WINDOWS

The development compiler used at SCC is Compaq Visual Fortran, and the following instructions assume that CVF is being used. If you are using a different compiler, then use the instructions as a guide.

The dependencies are as follows:

- MOPAC depends on Subroutines, Modules and Interfaces.
- Subroutines depends on Modules and Interfaces.
- Interfaces depends on Modules
- Modules does not depend on anything.

To compile, open the CVF window, and compile using build => build MOPAC2009.exe. This will put the Release executable in M:/utility. If you want it to go somewhere else, change the entry in Project => settings => Link => Output file name.

Linux

First, copy all the files in directories src_subroutines, src_modules, and src_interfaces into the directory MOPAC. The original directory structure was intended for use on a WINDOWS platform, and the use of separate directories was convenient there. On Linux platforms, the Makefile utility does not need separate directories for each group of source files.

Read over the Makefile, and make any changes necessary. The most probable change would be to replace the FORTRAN compiler, "g95" with a different compiler.

When the command, "make MOPAC2009.out" is issued from within the directory MOPAC, the Makefile will build MOPAC2009.out.

Running MOPAC

General

Data sets can have any name, but for convenience they should end in ".mop", ".dat", or ".arc". Examples of valid data set names are: H2O.mop, water.mop, water.dat, water.arc, "water dimer.mop". Other names can be used, such as water.mopac, water.archive. If suffixes other than ".mop", ".dat", or ".arc" are used, then the resulting files may have names other than those desired.

From the desktop in Windows:

This is the recommended sequence for getting MOPAC up and running on a Windows XP platform. It will install the license key and set up the association of MOPAC2007 with the data set in one step.

Make up a data-set with the name of the set being that of the license-key. Thus if the key is "1234567a1234567" then create a file (using e.g. notepad) with the name "1234567a1234567.mop". The file can be empty, it's only the name that is important.

In Windows Explorer, navigate to the file, and right-click on it. Use "Open with" and browse to C:\Program files\mopac. Select MOPAC2009.exe and click on "open". If you want to, put a check in "always use this program to open this kind of file" to associate files ending in ".mop" with MOPAC2009, then click on "ok". MOPAC2009 will start by displaying the End User License Agreement. After reviewing it, if it is acceptable, type " yes" (with the extra space) when prompted. You will probably need to scroll down to see this part.

From then on, whenever you want to run a MOPAC job, either double click on the data set (if you chose to make the association) or right-click and use "Open with" and select MOPAC2007.

From the DOS prompt:

In Windows systems, this is referred to as the "command prompt." Although running from the command prompt is straightforward, by preparing a suitable icon the ease of use of the command prompt can be improved. The steps involved in preparing the icon are as follows:

Paste a shortcut to the command prompt on the desktop.

Open the icon by double-clicking.

Right-click on the blue bar at the top and select "properties."

In "Font" chose "Lucida Console", bold, size14.

In "Layout" chose "Screen buffer size" Width: 107; Height: 6000; "Window size" Width: 107; Height 50;

"Window position" -4, -4, uncheck "let system position window"

In "Colors" chose "Screen background" as white (255,255,255), "Screen text" as black (0,0,0).

Click on "OK", then chose "Modify shortcut that started this window."

Double click on the icon to get the DOS prompt.

At the dos prompt, C:\>, type the path plus executable name followed by the file to be run. If the path contains spaces, then enclose the path plus executable name in quotation marks, thus:

```
C:\> "C:\program files\mopac\mopac2007.exe" water
```

where "water" is the name of the data set, with or without a suffix, to be run. If the location of the MOPAC executable is put in the path, then the command can be shortened to just the executable followed by the data set,

e.g.:

```
C:\> mopac2007 water
```

If the data set contains spaces, then enclose it in quotation marks, thus:

```
C:\> mopac2009 "benzoic acid"
```

Linux Version

For ease of use, put an alias to the location of the MOPAC executable in the startup command sequence (the .bashrc or .cshrc)

Once that is done, the procedure for running a MOPAC job is the same as running from the DOS prompt.

Windows Version

Before the Windows version can be used, make a MOPAC executable icon on the desktop, by copying the executable (use Windows Explorer) and pasting a shortcut on the desktop.

In Windows Explorer, highlight the file(s) to be run. Drag them to the icon for MOPAC, and drop them there. Up to 50 files can be "dragged and dropped" at a time. The jobs will run as if in a batch queue. The progress of each job will be displayed in the MOPAC window. The current job can be paused, by clicking on the "State - pause" option, and resumed by clicking on the "State - resume" option. The "Shut" command (to stop the current calculation by making it appear to run out of time) can be given by clicking on the "Shutdown" key.

To familiarize yourself with the system, the following operations might be useful.

1. Run some test molecules, and verify that MOPAC is producing "acceptable" results. Representative examples are:

Benzene	A simple organic compound.
Ferrocene	A simple organometallic compound.
Polyethylene	A simple polymer
Stretching polyethylene	Example of reaction path
Diamond	A simple solid
Chromium acetate	A complex with a metal-metal bond
[Fe(H₂O)₆]²⁺	A low-spin open-shell d ⁵ transition metal complex
[Mn(H₂O)₆]²⁺	A high-spin open-shell d ⁵ transition metal complex
Fe₂(C₄Me₃)₂(CO)₃	A complicated organometallic complex
Ni₂(Cp)₃(CO)₂	Another complicated organometallic complex

2. Make some simple modifications to the data-files supplied in order to test your understanding of the data format
 3. When satisfied that MOPAC is working, and that data files can be made, begin production runs.

Restarts should be user transparent. If MOPAC does make any restart files, do not change them (it would be hard to do anyhow, as they're in machine code), as they will be used when you run a RESTART job.

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.

- [Main geometric sequence](#)
- [Main electronic flow](#)
- [Control within MOPAC](#)

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 [HSCORE](#). It is normally sufficient to verify that the two hydrogen atoms are equivalent, and that the p 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 `HSCORE`, `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 p 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 a and b 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, [COMPFG](#) 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: MrMOPAC@OpenMOPAC.net.

Porting MOPAC to other platforms

MOPAC is designed to be as platform independent as possible. The code is written in as near pure FORTRAN90/95 as possible, and where possible simple constructions are used. This is deliberately done to make it easier to port the program to other machines.

To validate a port, the data sets in the [validation suite](#) should be run. If all tests are passed, within the tolerances given in the tests, then the ported program can be called a valid version of MOPAC. Insofar as is practical, the mode of submission of a MOPAC job should be preserved, e.g.,

```
(prompt) MOPAC <data-set> [<queue-options>...]
```

Any changes which do not violate the FORTRAN-90 conventions, and which users believe would be generally desirable, may be sent to the author.

The main difficulty in modifying the program so that it runs on other platforms is likely to be in the channel assignments. In an attempt to make this easier, all channel assignments are defined in subroutines `getdat.f90`, `chanel_C.f90` and `init_filenames.f90`. MOPAC uses the UNIX instructions `iargc` and `getarg`. These read in the number of arguments supplied with the MOPAC command, and if this is greater than zero, it then reads in the name of the job. This is stored in `molkst_C` module string variable `JOBNAM`.

Other possible trouble-spots are (a) the CPU timer (in `second.F90`), and the date function `fdate` (in `readmo.f90` and `writmo.f90`) and (b) the `RESTART` command.

In order to make porting easier, all channel numbers are represented by symbols. The default assignments are shown in the Table.

Table: Files used by MOPAC

Channel	Symbol	Name	Input?	Output?	Contents
4	iend	<filename>.end	Yes		Shut command
7	from_data_set	<filename>.mop*	Yes		Data-set
9	ires	<filename>.res	Yes	Yes	Restart file
10	iden	<filename>.den	Yes	Yes	Density file
11	ilog	<filename>.log		Yes	Log file
12	iarc	<filename>.arc		Yes	Archive file
13	igpt	<filename>.gpt		Yes	Graphics file
14		User-defined	Yes		New parameters
15	iesr	<filename>.esr		Yes	ESP Restart
16	isyb	<filename>.syb		Yes	SYBYL file
18	ibrz	<filename>.brz		Yes	Brillouin Zone
20	iump	<filename>.ump		Yes	Data for grid
21	iesp	<filename>.esp		Yes	ESP data
22		SCRATCH			Used by GEOUTG
25	ir	SCRATCH			Internal working
26	iw	<filename>.out		Yes	Results file

All the data are read in on channel 2, and immediately written to a SCRATCH channel, 5. For the rest of the job, data are read off channel 5 as if it were the ordinary input channel. By doing this, the data-set is used only at the very start of a job, and is released immediately afterwards.

*: Alternatives: <filename>.dat, <filename>.arc, <filename>

If, for whatever reason, a channel number must be changed, then the change should be made in `chanel_C`. Of course, if the channel numbers are to be changed, care should be taken to avoid channel numbers which are already in use.

Algorithmic validation

A powerful device to validate a program is by examination of the computer code involved. Reading the program is not practical: MOPAC2009 is over 120,000 lines long. Instead various software validation tools can be used. The more important of these are:

- **FORCHECK**: Checks conformance with the ANSI FORTRAN 90 Standard. MOPAC has known exceptions to the standard, but all of these have been independently checked.
- **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.

Control within MOPAC

Almost all the control information is passed *via* the single datum "KEYWRD", a string of 80 to 240 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 common blocks 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 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 the common blocks. 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 common block.

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.

&

&

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 also 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 T-PRIORITY=0.5
LET GEO-OK VELOCITY KINETIC=5.0
```


+

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.

0SCF

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 0SCF 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](#).

1ELECTRON

The final one-electron matrix is printed out. This matrix is composed of atomic orbitals; the array element between orbitals f_l and f_s 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 f_l and f_s on the same atom are calculated from the electron-nuclear attraction energy, and also from the U_{ll} value if $l = s$.

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

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 [An SCF Calculation](#) 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 [Frank Condon Considerations](#)). This allows the vertical transition energies to be calculated (the Franck-Condon transitions).

AIDER

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

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

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

When keyword [BONDS](#) is used in a [MOZYME](#) calculation, only those bonds with bond orders greater than 0.01 and that do not involve hydrogen atoms are printed. This is to reduce the size of the output. When `ALLBONDS` is used, all bonds with bond orders greater than 0.001 are printed. This includes bonds involving hydrogen atoms.

This prints the rotationally invariant bond order 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 [BONDS](#).

Valencies are calculated from the atomic terms only and are defined as the sum of the bonds the atom makes with other atoms.

ALLVEC

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

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

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).

AUTOSYM

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.

AUTOSYM only adds symmetry to coordinates that can be optimized. If a coordinate has an optimization flag of "0", it will not be looked at by AUTOSYM.

Consider benzene: If someone built the Z-matrix in internal coordinates, they would likely mark the C-C and C-H distances as optimizable, and lock the angles and dihedrals. If they used AUTOSYM, then it would symmetry-relate all the C-C distances and all of the C-H distances, but leave the angles and dihedrals untouched.

AUX[([*n*] [,] [COMP] [,] [PRECISION=*m*] [,] [MOS=*n*])]

When AUX is used then auxiliary information is output to a file called <file>.aux. When AUX(*n*) is used, output for the **changing geometry** is re-directed to output channel *n*, *n* is limited to the range 0 - 99. Thus AUX(6) would re-direct the changing geometry to standard output unit 6. Useful values are *n* = 0 and *n* = 6. Avoid using units 2 to 4, and 9 to 30, as these are used by MOPAC. The <file>.aux files are intended for use by Graphical User Interfaces (GUI's), but the files can be converted using a simple Windows utility, [Read_AUX](#), into ASCII tables for use by programs such as EXCEL.

When COMP is present, all large blocks of data are printed in compressed form.

To increase precision by *m* digits (*m* limited to the range 1 to 9), use PRECISION=*m*.

By default, only the 10 highest occupied and 10 lowest unoccupied M.O.s are printed. To change the default, use MOS=*n*, where *n* is the number of occupied and virtual M.O.s to be printed. If *n* is greater than the number of M.O.s, it will be re-set to all the M.O.s, so a convenient way to print all the M.O.s is to use MOS=99999. To prevent printing of M.O.'s, use MOS=-99999.

The commas are optional.

If the re-direction unit "*n*" is used together with any of COMP, PRECISION, or MOS, then "*n*" must come first, followed by the other options.

Only data considered to be important to GUI programs will be included in the .aux file. A consequence of this is that some data might be missing. If such data can be identified, and a message sent to MrMOPAC@OpenMOPAC.net, then the data will be added to the .aux file.

The auxiliary information is a very compact form of the results of a calculation in a primitive mark-up language. This file is intended for use by other programs. An example of the auxiliary file for water is given below. No explanation is necessary: the file is intended to be self-descriptive.

```
START OF MOPAC FILE
#####
#                                     #
#      Start of Input data          #
#                                     #
#####
MOPAC_VERSION=MOPAC2007.7.150W
DATE="Mon Jun  4 12:15:37 2007"
METHOD=PM6
TITLE=" Water"
KEYWORDS=" SYMMETRY"
ATOM_EL[0003]=
  O  H  H
ATOM_X:ANGSTROMS[0009]=
  0.0000    0.0000    0.0000
  0.6018    0.7647    0.0000
  0.6018   -0.7647    0.0000
AO_ATOMINDEX[0006]=
  1  1  1  1  2  3
ATOM_SYMTYPE[0006]=
  S PX PY PZ  S  S
AO_ZETA[0006]=
  5.4218  2.2710  2.2710  2.2710  1.2686  1.2686
ATOM_PQN[0006]=
  2  2  2  2  1  1
NUM_ELECTRONS=0008
EMPIRICAL_FORMULA="H2 O"
#####
#                                     #
#      Geometry optimization        #
#                                     #
#####
HEAT_OF_FORM_UPDATED:KCAL/MOL=-0.540108D+02
GRADIENT_UPDATED:KCAL/MOL/ANG=+0.210682D+02
ATOM_X_UPDATED:ANGSTROMS[0009]=
  0.0000    0.0000    0.0000
  0.6018    0.7647    0.0000
```

```

    0.6018    -0.7647    0.0000
HEAT_OF_FORM_UPDATED:KCAL/MOL=-0.542600D+02
GRADIENT_UPDATED:KCAL/MOL/ANG=+0.893385D+01
ATOM_X_UPDATED:ANGSTROMS[0009]=
    0.0000    0.0000    0.0000
    0.5693    0.7558    0.0000
    0.5693    -0.7558    0.0000
HEAT_OF_FORM_UPDATED:KCAL/MOL=-0.543026D+02
GRADIENT_UPDATED:KCAL/MOL/ANG=+0.225094D+01
ATOM_X_UPDATED:ANGSTROMS[0009]=
    0.0000    0.0000    0.0000
    0.5641    0.7630    0.0000
    0.5641    -0.7630    0.0000
HEAT_OF_FORM_UPDATED:KCAL/MOL=-0.543063D+02
GRADIENT_UPDATED:KCAL/MOL/ANG=+0.144932D+00
ATOM_X_UPDATED:ANGSTROMS[0009]=
    0.0000    0.0000    0.0000
    0.5610    0.7655    0.0000
    0.5610    -0.7655    0.0000
#####
#
#           Final SCF results           #
#
#####
HEAT_OF_FORMATION:KCAL/MOL=-0.543063D+02
ENERGY_ELECTRONIC:EV=-0.458386D+03
ENERGY_NUCLEAR:EV=+0.139308D+03
POINT_GROUP=C2v
DIPOLE:DEBYE=+0.206870D+01
AREA:SQUARE  ANGSTROMS=+0.424512D+02
VOLUME:CUBIC  ANGSTROMS=+0.251444D+02
ATOM_X_OPT:ANGSTROMS[0009]=
    0.0000    0.0000    0.0000
    0.5610    0.7655    0.0000
    0.5610    -0.7655    0.0000
ATOM_CHARGES[0003]=
-0.61867 +0.30933 +0.30933
OVERLAP_MATRIX[000021]=
# Lower half triangle only
    1.0000    0.0000    1.0000    0.0000    0.0000    1.0000    0.0000    0.0000
0.0000    1.0000
    0.1618    0.2330    0.3180    0.0000    1.0000    0.1618    0.2330    -0.3180
0.0000    0.2333
    1.0000
EIGENVECTORS[000036]=
    -0.8288    -0.2451    0.0000    0.0000    -0.3556    -0.3556    0.0000    0.0000
0.8129    0.0000
    0.4118    -0.4118    0.4394    -0.8417    0.0000    0.0000    -0.2220    -0.2220
0.0000    0.0000
    0.0000    1.0000    0.0000    0.0000    0.3463    0.4812    0.0000    0.0000    -
0.5694    -0.5694
    0.0000    0.0000    0.5824    0.0000    -0.5748    0.5748
TOTAL_DENSITY_MATRIX[000021]=
# Lower half triangle only
    1.7601    -0.3333    1.5370    0.0000    0.0000    1.3216    0.0000    0.0000
0.0000    2.0000
    0.3944    0.5480    0.6695    0.0000    0.6907    0.3944    0.5480    -0.6695
0.0000    0.0123
    0.6907
M.O.SYMMETRY_LABELS[0006]=
    1a1    1b2    2a1    1b1    3a1    2b2
EIGENVALUES[0006]=
    -30.402    -18.983    -14.111    -11.905    4.067    6.038
MOLECULAR_ORBITAL_OCCUPANCIES[00006]=
    2.0000    2.0000    2.0000    2.0000    0.0000    0.0000

```

When COMP is present, the overlap and density matrices, and eigenvectors, are printed in compressed form.
Compare the previous uncompressed data with the following compressed data to

see how the compression is done.

```

OVERLAP_INDICES[013]=
  1      3      6      10      15      21      18      13      20      17      12      16      11
OVERLAP_COEFFICIENTS[013]=
  1.0000  1.0000  1.0000  1.0000  1.0000  1.0000  -0.3180  0.3180  0.2336  0.2332
  0.2332  0.1619  0.1618
SET_OF_MOS=
  1      6
MO_INDICES[004]=
  1      6      5      2
MO_COEFFICIENTS[004]=
  0.8287  0.3558  0.3557  0.2452
MO_INDICES[003]=
  3      5      6
MO_COEFFICIENTS[003]=
  0.8129  0.4118  -0.4118
MO_INDICES[004]=
  2      1      5      6
MO_COEFFICIENTS[004]=
  0.8416  -0.4396  0.2219  0.2219
MO_INDICES[001]=
  4
MO_COEFFICIENTS[001]=
  1.0000
MO_INDICES[004]=
  5      6      2      1
MO_COEFFICIENTS[004]=
  0.5694  0.5693  -0.4812  -0.3465
MO_INDICES[003]=
  3      6      5
MO_COEFFICIENTS[003]=
  0.5824  0.5749  -0.5748
DENSITY_MATRIX_INDICES[014]=
  10      1      3      6      21      15      13      18      12      17      16      11      2      20
DENSITY_MATRIX_COEFFICIENTS[014]=
  2.0000  1.7599  1.5369  1.3217  0.6908  0.6908  0.6695  -0.6695  0.5480  0.5480
  0.3945  0.3945  -0.3334  0.0124

```

When COMP and MOS are present, a user-defined set of M.O.s is printed. The default is the 10 highest M.O.s and 10 lowest virtual M.O.s.
 If [LARGE](#) is present, all the M.O.s are printed.

AM1

The AM1 [\[3\]](#) method is to be used. By default PM6 is run.

BAR=*n . nn*

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 0.01, or 1 percent. This represents a maximum rate of reduction of the bar of 1 percent per step. Alternative values that might be considered are `BAR=0.005` or `BAR=0.001`, although other values may be used. See also [SADDLE](#).

If CPU time is not a major consideration, use `BAR=0.001`.

BCC

BCC is not used by MOPAC at all, but instead is used to inform program BZ about the nature of the solid being studied.

BCC should be specified if all odd unit cells are omitted. This applies only to layer and solid systems.

BFGS

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*

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.

In a [DRC](#), BIGCYCLES=*n* will run until *n* complete oscillations have occurred. BIGCYCLES=1 would allow a single complete vibration to be generated.

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](#).









BIRADICAL

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

BIRADICAL is limited to restricted Hartree Fock systems, so [UHF](#) should not be present.

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 the Figure); 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, $(1/2)^{1/2}(Y_1 - Y_2)$, and a triplet state $(1/2)^{1/2}(Y_1 + Y_2)$. 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.

Figure: Configurations used in BIRADICAL Calculation

	Microstate 1	Microstate 2	Microstate 3	Microstate 4
LUMO				
HOMO				

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 * 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: BIR

BONDS

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 RHF calculations, the total density matrix (alpha plus beta density matrices) is perfectly duodempotent, that is, the square of the density matrix, P , is exactly two times the density matrix, or $P*P = 2*P$ (see [Bond Orders](#)), and valencies will be correct. In UHF and non-variationally optimized wavefunctions the calculated valency will be incorrect, the degree of error being proportional to the non-duodempotency of the total density matrix. (In UHF work, the individual alpha and beta density matrices are idempotent, that is, $P(\alpha)*P(\alpha) = 1.0*P(\alpha)$ and $P(\beta)*P(\beta) = 1.0*P(\beta)$, but, in general, the sum of these two matrices is not duodempotent, i.e., $(P(\alpha)+P(\beta))*(P(\alpha)+P(\beta)) \neq 2.0*(P(\alpha)+P(\beta))$.)

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.

CAMP

The [Camp-King](#) converger is to be used. This is a very powerful, but CPU intensive, SCF converger. The Camp-King converger does *not* work with [MOZYME](#)

CARTAB

Print part of the character table for the appropriate point group. Only the operations essential for defining the character table are printed, thus, for example, the complete character table for Td is

Point Group Td					
Td	E	6s _d	8C ₃	6S ₄	3C ₂
A ₁	+1	+1	+1	+1	+1
A ₂	+1	-1	+1	-1	+1
E	+2	0	-1	0	+2
T ₁	+3	-1	0	+1	-1
T ₂	+3	+1	0	-1	-1

The character table printed is abbreviated:

Abbreviated Point Group Td				
Td	E	s _d	C ₃	C ₂ (z)
A ₁	+1	+1	+1	+1
A ₂	+1	-1	+1	+1
E	+2	0	-1	+2
T ₁	+3	-1	0	-1
T ₂	+3	+1	0	-1

All missing operations can readily be generated by multiplying two or more operations together, thus in Td, the S₄ operation can be generated by multiplying a s_d by C₂.

This keyword is intended for use in debugging and for teaching.

C.I.=n

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 [136] 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. Normally, configuration interaction (see the [MECI](#)) is invoked if the keywords which imply a C.I. calculation are used ([BIRADICAL](#) or [EXCITED](#)). Note that [UHF](#) should not be present.

Table: Number of Configurations used in MECI

	No. of Electrons		Confgs	No. of Electrons		Confgs
	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 [Choice of State to be Optimized](#).

If a change of spin is defined, then larger numbers of M.O.s can be used up to a maximum of 22. If more than 22 M.O.s are needed, then `meCI_C.F90` would need to be changed. In `meCI_C.F90`, 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), 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.

If microstates are supplied, using [MICROS](#), then the states used are precisely controlled. This requires more work, because each microstate has to be explicitly defined, but it is very flexible and much more certain than when other keywords such as [CISDT](#) are used.

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

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

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). If [OPEN\(\$n_1\$, \$n_2\$ \)](#) is present, then the number of electrons may be increased.

Table: 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.

CHAINS(*text*)

CHAINS is used with [PDBOUT](#).

Protein chains are normally labeled A, then B, then C, etc. If this sequence is correct, then keyword CHAINS is not needed. If a different sequence is to be used, put that sequence in place of *text* in CHAINS(*text*) .

For example, in PDB entry 2V66, the chains are B, C, D, and E, not A, B, C, D; for this system, the keyword would be: CHAINS(BCDE) .

CHAINS is also useful if there is a break in the reported sequence. Thus if a protein consists of one chain with three breaks where residues are missing, then, in order for the PDB file to have the correct chain-letters, keyword CHAINS(AAAAA) would be necessary

See also [START_RES\(text\)](#) to define residue numbers.

CHARGE=*n*

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. Examples of various charged systems are given in the Table below.

MOZYME-specific definition of CHARGE:

If CHARGE=*n* is not present, the charge required by the Lewis structure will be used.
If CHARGE=*n* is present, including CHARGE=0, that charge will be used, unless it is in conflict with the Lewis structure, in which case the job will be stopped.
If both GEO-OK and CHARGE=*n* are present, that charge will be used, unless it is in conflict with the Lewis structure, in which case the charge required by the Lewis structure will be used. (This option is not useful, it is equivalent to CHARGE=*n* being absent)

In cases where the charge supplied is in conflict with that required by the Lewis structure, the Lewis structure will be printed. That, together with the printed charged sites, will allow the correct charge to be determined.

Table: Use of CHARGE=*n*

Ion	Keyword	Ion	Keyword
NH ₄ ⁺	CHARGE=1	CH ₃ COO ⁻	CHARGE=- 1
C ₂ H ₅ ⁺	CHARGE=1	(COO) ⁼	CHARGE=- 2
SO ₄ ⁼	CHARGE=- 2	PO ₄ ⁻³	CHARGE=- 3
HSO ₄ ⁻	CHARGE=- 1	H ₂ PO ₄ ⁻	CHARGE=- 1

CHARGES

In a [MOZYME](#) calculation, all ionized sites in the system will be identified and printed. The run will then be stopped. This operation is useful during the preparation of a data set, in that it is very easy for even experienced users to have incorrect charges on large systems, particularly complicated organic species such as enzymes and organometallics.

The method for identifying charged atoms is very powerful. So although you might be quite certain that you have the charges correct, if **CHARGES** indicates that you have made a mistake, please check the system using the information printed. In 100% of the cases where users have complained that MOPAC reported faulty charges, **CHARGES** identified errors in the user's structure.

In MOPAC, a common Lewis structure involves hypervalent systems being represented as ions. Thus sulfuric acid would be represented as $[\text{SO}_2]^{2+} + 2[\text{OH}]^-$. Be aware that "salt bridges" can occur, e.g. $-\text{COOH} \cdots \text{H}_2\text{N}-$ can convert to $-\text{[COO]}^- \cdots [\text{H}_2\text{N}]^+$. Such structures should *not* be regarded as faults.

The choice of which nitrogen atoms in the heterocycle in histidine have attached hydrogen atoms is often not immediately obvious. But in many cases, the choice is important - the wrong choice will always result in a high-energy structure, and be accompanied by severe rotation of the heterocycle, as it tries to relieve the stress. **CHARGES** cannot help with identifying the right choice.

Related key-words: [LEWIS](#), [CVB](#), [METAL](#), [VDWM](#), [CHARGE](#), and [MOZYME](#)

See also: [Lewis Structures](#), [MOZYME introduction](#)

CHARST

Print details of the working of CHARST. CHARST calculates the symmetry characters of the state functions, that is:

$$c = \langle F | \text{Operator} | F \rangle$$

where F is a state function, that is, a linear combination of microstates, Y . Each microstate is an antisymmetrized product of molecular orbitals, and each M.O. is a linear combination of atomic orbitals, which, in turn, are represented by Slater orbitals. Only information on the first state is printed; the quantities printed are:

Symmetry operation in CHARST: the 3x3 Euler rotation matrix representing the operation. This is a rotation, mirror plane, or a product of the two.

Transform of M.O.s: The result of the operation on the M.O.s. The effect of the operation is to convert each M.O. of the active space into a linear combination of the M.O.s of the active space.

State Transform for State i under Operation j : As the name suggests, this is the result of operating on state $|F_i\rangle$ with operator $|\text{Operator}_j$, i.e.:

$$|\text{Operator}_j | F_i \rangle$$

The state is then converted into a linear combination of states.

For CHARST to work [DEBUG](#) must also be present. Adding [SYMOIR](#) will give the characters of the operations for all states.

See [States](#) for more information

CIS

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

In a configuration interaction involving an active space in which n levels are doubly occupied and m levels are empty, the number of states considered in the C.I. would be $(1 + 2.n.m)$. By [Koopmans' theorem](#), the ground microstate does not interact with any microstates resulting from one-electron excitation; this means that the lowest state resulting from the C.I. consists of the doubly occupied microstate only. It is included in the C.I. so that the relative energies are relative to the ground state. Because the C.I. cannot affect the lowest root, a CIS ground-state calculation should not be used in runs that affect the geometry, e.g. geometry optimization or vibrational frequency calculations.

CISD

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.)

The number of states considered in a CISD C.I. involving n doubly-occupied M.O.s and m empty M.O.s is:

$$\text{No. of States} = 1 + 2.n.m + (n.m)^2 + (n(n-1).m(m-1))/2$$

This represents:

Ground state

+ (one a electron excited + one b electron excited)

+ (one a electron and one b electron excited)

+ (two a electrons excited + two b electrons excited)

The first term represents the ground state, the second term represents number of one-electron excitations, and the third and fourth terms represent the number of two-electron excitations.

CISDT

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.)

The number of states considered in a CISDT C.I. involving n doubly-occupied M.O.s and m empty M.O.s is:

$$\text{No. of States} = 1 + 2nm + (nm)^2 + (n(n-1).m(m-1))/2 + (n^2(n-1).m^2(m-1))/2 + (n(n-1)(n-2).m(m-1)(m-2))/6$$

This represents:

Ground state

+ (one a electron excited + one b electron excited)

+ (one a electron and one b electron excited)

+ (two a electrons excited + two b electrons excited)

+ (one a electron and two b electrons excited + two a electrons and one b electron excited)

+ (three a electrons excited + three b electrons excited)

The first term represents the ground state, the second term represents number of one-electron excitations, the third and fourth terms represent the number of two-electron excitations, and the fifth and sixth terms represent the number of three-electron excitations.

COMPFG

When `COMPFG` is specified, the DH_f is printed whenever `COMPFG` is called. 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.

COSCCH

In a COSMO calculation, where the [COSWRT](#) is used, COSCCH allows for bond dipole and bond quadrupole corrections before the final COSMO calculation. Quadrupoles require a charge dq on the bond between atoms i1, and i2. The syntax is dq: i1 i2 dq.

The list of dq correction charges is followed by a list of correction charges on all atoms.

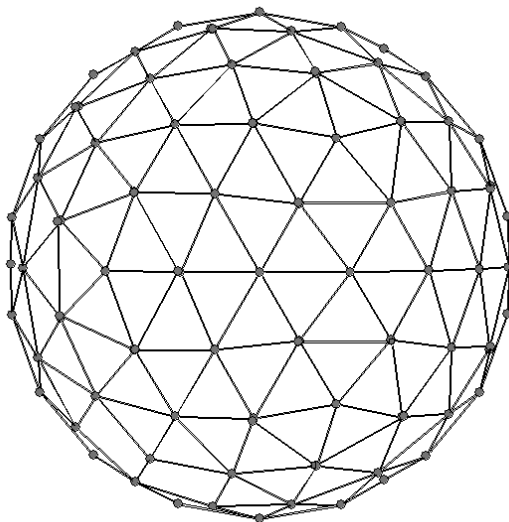
This keyword is intended for use by Dr Andreas Klamt's software only.

COSWRT

Print out details of the [COSMO](#) surface to a file called <file>.cos. Quantities printed include the solvent accessible surface area, per atom, and the segment that make up the SAS.

The points on the SAS are printed; these can then be used with a graphics package to give a pictorial representation of the SAS.

As an example, when [NSPA](#)=122 is used with a single atom, the set of points represented by the following picture is generated:



For a sigma profile, corrections to the bond dipole and quadrupole can be supplied by [COSCCH](#).

CUTOFP=*n*.*nn*

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.

CUTOFF=*n . nn*

[MOZYME](#) calculations run faster when some NDDO interactions are ignored. By default, all NDDO interactions are included between atoms separated by 12 Ångstroms or less, and all polarization functions between atoms separated by 30 Ångstroms or less. For rough work, such as optimizing a structure that is far from a minimum, if `CUTOFF=n . nn` is used, then only point charge interactions beyond *n.nn* Ångstroms are used. A suitable value for *n.nn* is 6.0. This will give a large increase in speed.

If `CUTOFF=6` is used, then [GNORM=20](#) is recommended. Once the geometry is optimized using `GNORM=20`, the `CUTOFF` should be increased (9 is a good value for general use) or removed, and `GNORM` reduced to, e.g., `GNORM=10`.

For the rare occasions when finer control is needed, use `CUTOF1` and `CUTOF2`.

CUTOF1=*n . nn*

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.

CUTOF2=*n . nn*

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.

CYCLES=*n*

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 (A) cycles of optimization in [EF](#), [SIGMA](#), [NLLSQ](#), and [BFGS](#), (B) steps in a [FORCE](#) calculation, and (C) 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](#).

CVB(*text*)

Sometimes in a [MOZYME](#) calculation, a Lewis structure cannot be constructed from the topology of the system supplied. When that happens, a warning message will be printed and the run stopped. Unwanted bonds can be removed using CVB. The format is:

```
CVB(n1:-n2[;n3:-n4[;n5:-n6....]])
```

where the bonds to be removed are n1-n2, n3-n4, n5-n6, etc. Thus if bonds were to be removed between atoms 4 and 5, and between atoms 665 and 670, then `CVB(4:-5;665:-670)` would be used. If bonds are needed, then don't use the minus sign. Thus if bonds were needed between atoms 4 and 5, and between atoms 665 and 670, then `CVB(4:5,665:670)` would be used. CVB is also useful when [PDBOUT](#) is present. A topology sometimes has an unwanted bond between two chains. This prevents some of the residues being identified. To correct this fault, remove the unwanted bond using CVB. If there are a several unwanted bonds, and these bonds involve a specific element, then consider the alternative of using [VDWM](#) to change the Van der Waals radius for that element.

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

See also [SETPI](#) to explicitly assign p bonds, and [atom labels](#) to explicitly assign charges.

Related keywords:

See also [LEWIS](#), [CHARGES](#), [METAL](#), [VDWM](#), [CHARGE](#), and [MOZYME](#)

See also: [Lewis Structures](#), [MOZYME introduction](#)

DAMP

In a [MOZYME](#) calculation, damp SCF oscillations using a factor of *n.nn*. *n.nn* should be in the range $1.0 < n.nn < 0.0$. Values in the region of 0.5 are usually successful. For solids, the default factor is 0.5, so *n.nn* should be less than 0.5.

DATA= *text*

Normally, MOPAC reads data from the data-file supplied. When the name of the data set cannot be changed, the data set to be used can be provided in the fixed data set by use of DATA=<text>. Thus if the name of the fixed data set is "input.dat", so the only job that can be run is:

MOPAC input.dat

and the job that is to be run is the data set "toluene.mop" in folder "F:\my mopac jobs" then the file input.dat would contain the single line:

DATA=F:\my mopac jobs\toluene.dat

When the MOPAC job is run, input.dat would be read in, then because "DATA=F:\my mopac jobs\toluene.dat" was present, the file toluene.dat would be read in. The results would be put into files of the type input.<suffix>.

DCART

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 .

Only the derivatives of the closed-shell SCF wavefunction are printed, that is, any post-hoc corrections such as the effects of pressure, dispersion, etc. are not included at this point. For the equivalent open-shell RHF gradients, see [DERNVO](#).

DDMAX=*n . nn*

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*

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

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](#).

DEBUGPULAY

Print some of the working in subroutine PULAY.

DENOUT , DENOUTF

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](#)). The density matrix file is unformatted. To output the formatted file, use DENOUTF.

In both formatted and unformatted density files, the layout of the data is: Number of atomic orbitals, number of atoms, lower half triangle of density matrix elements. When the density file is used in a subsequent calculation, the number of atomic orbitals and number of atoms is compared with that expected. If they are different, an error message is printed, and the run is stopped.

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](#)). The formatted option, DENOUTF , is not available .

The opposite of DENOUT is [OLDENS](#)

DENSITY

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.

DERI1

Print some of the working in subroutine DERI1.

DERI2

Print some of the working in subroutine DERI2.

DERITR

Print some of the working in subroutine DERITR.

DERIV

Print some of the working in subroutine DERIV. Quantities printed are:

(A) The geometry using in the derivative calculation.

(B) The gradients of the coordinates flagged for optimization. Units: kcal/mol/Angstrom and kcal/mol/radian (for angles and dihedrals)

These are the values that will be passed back from subroutine DERIV.

(C) The values of the coordinates flagged for optimization. Units: Angstroms and radians.

Some of the working in subroutines within DERIV can be printed, see [DCART](#) and [DERNVO](#),

DERNVO

Print some of the working in subroutine DERNVO. DERNVO works out contributions to the gradients of the Cartesian coordinates arising from the open-shell RHF wavefunction. Quantities printed are:

- * CARTESIAN DERIVATIVES DUE TO FROZEN CORE: This is RO-UHF equivalent to the material printed by [DCART](#).
- * RESIDUAL ERROR: If the Cartesian derivatives are calculated correctly, then the sum of the gradients in "x", "y", and "z" will be zero. The difference from zero is a measure of the error in the Cartesian derivatives due to the frozen core.
- * CARTESIAN DERIVATIVES DUE TO RELAXING CORE: This is the contribution to the Cartesian derivatives arising from the C.I. terms only.
- * RESIDUAL ERROR: If the Cartesian derivatives due to the relaxing core are calculated correctly, then the sum of the gradients in "x", "y", and "z" will be zero. The difference from zero is a measure of the error in the Cartesian derivatives due to the relaxing core.
- * CARTESIAN DERIVATIVES FROM ANALYTICAL C.I. CALCULATION The sum of the frozen plus relaxing contributions (the total derivative due to the wave-function)
- * RESIDUAL ERROR: The errors in the calculation of the sum of the frozen plus relaxing contributions to the total derivative due to the wave-function. Ideally, these should be zero.

DFORCE

Print some of the working in the FORCE calculation.

DFP

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.

DISEX=*n . nn*

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.

D_{MAX}=*n . nn*

In the EF routine [7], the starting value for the trust radius is set to *n.nn*. After the first geometry optimization cycle, D_{MAX} is changed by _{EF}, and can increase by a factor of 2.0 on every cycle. If D_{MAX} is *not* set, the default of 0.20 is used. (Try to avoid using this keyword.)

DOUBLET

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

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 [Dynamic and Intrinsic Reaction Coordinates](#) for more details.

By default, steps of 0.1fs are used. This default can be changed using keywords:

To print reaction coordinate steps that are separated by constant time, use [T-PRIORITY](#) or [T-PRIORITY=*n.nn*](#).

To print reaction coordinate steps that are separated by constant energy, use [H-PRIORITY](#) or [H-PRIORITY=*n.nn*](#).

To print reaction coordinate steps that are separated by constant distance, use [X-PRIORITY](#) or [X-PRIORITY=*n.nn*](#).

Use [LARGE](#) to control print of internal coordinates and Cartesian coordinates.

A useful keyword for limiting the DRC is [BIGCYCLES=*n*](#). `BIGCYCLES=n` will run until *n* complete oscillations have occurred. Normal modes of vibration can be generated using `BIGCYCLES=1` and [IRC=*n*](#), where *n* is the *n*'th normal mode of interest. Such a normal mode would involve one complete oscillation, from equilibrium to turning point back to equilibrium, then up the other side of the potential well to the other turning point, then back to the equilibrium position, at which point the DRC is stopped.

Most of the trajectories mapped out by `BIGCYCLES=1 IRC=n DRC` are anharmonic, that is, they reflect the asymmetry of the Morse potential. They are, however, accurate within the semiempirical approximation used. Thus if a vibrational frequency is *m* cm⁻¹, the duration of one complete vibration will be 1/(*m.c*) seconds. For example, if a normal mode has an energy (frequency) of 1000 cm⁻¹, each complete vibration would take 33.4 femtoseconds.

$$\mathbf{DRC}=\mathbf{\textit{n . nnn}}$$

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*

Restart files are written automatically at two 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

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

EF

The Eigenvector Following routine [7] is the default geometry optimization method. For systems with more than 2000 variables, the L-BFGS optimizer is the default. To prevent the L-BFGS optimizer being used, add keyword EF. That is, EF is provided to allow explicit definition of the optimizer. Alternative geometry optimizers are the BFGS [19,20,21,22] and the low-memory method, L-BFGS . See also [DDMAX](#), [DDMIN](#), [DMAX](#), [HESS](#), [IUPD](#), [LET](#), [MODE=](#), [NONR](#), [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

[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).

EIGS

Print all eigenvalues, in eV, on every iteration.

ENPART

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.

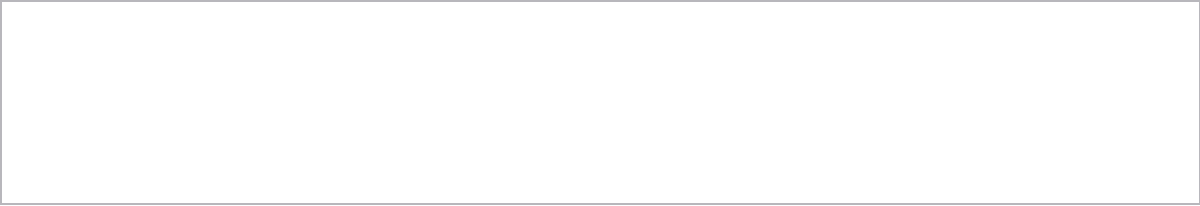
Atoms can be grouped into sets, up to 9 sets are allowed. The format is ENPART(n1, n2,n3...) where n1, n2, n3, etc., are the number of atoms in each set. All atoms in each set must be together in the geometry file. For example, to do an energy partition on the water molecule, use atoms O, H, H, O, H, H, and use ENPART(3,3). In this, the first three atoms are the first water molecule.

Here is an example of energy partitioning for the HF dimer. In order to work out the HF - HF interaction energy, keyword ENPART(2,2) was used.

TOTAL ENERGY PARTITIONING
ALL ENERGIES ARE IN ELECTRON VOLTS

ONE-CENTER TERMS
E-E: ELECTRON-ELECTRON REPULSION
E-N: ELECTRON-NUCLEAR ATTRACTION

ATOM		E-E	E-N	(E-E + E-N)
F	1	333.3949	-800.5397	-467.1449
H	2	1.8754	-8.1040	-6.2286
F	3	333.3956	-800.5406	-467.1450
H	4	1.8754	-8.1039	-6.2285

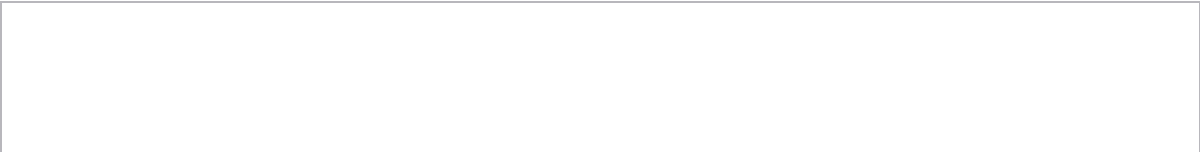


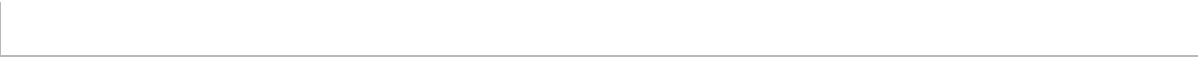
TWO-CENTER TERMS

J: RESONANCE ENERGY
K: EXCHANGE ENERGY
E-E: ELECTRON-ELECTRON REPULSION
E-N: ELECTRON-NUCLEAR ATTRACTION
N-N: NUCLEAR-NUCLEAR REPULSION
C: COULOMBIC INTERACTION = E-E + E-N + N-N
EE: TOTAL OF ELECTRONIC AND NUCLEAR ENERGIES

ATOM PAIR		J	K	E-E	E-N	N-N	C	EE
H 2 F 1		-12.1007	-4.8374	51.3351	-121.4430	75.1795	5.0717	-11.8664
F 3 F 1		0.0000	-0.0002	223.2326	-429.5659	206.6544	0.3210	0.3208
F 3 H 2		-0.0012	-0.0001	23.5576	-55.3454	31.4388	-0.3490	-0.3503
H 4 F 1		-0.0012	-0.0001	23.5496	-55.3269	31.4284	-0.3489	-0.3502
H 4 H 2		-0.0015	-0.0001	2.2758	-6.3168	4.3835	0.3425	0.3409
H 4 F 3		-12.0999	-4.8372	51.3339	-121.4406	75.1776	5.0709	-11.8662

System is to be split into 2 parts
Part No. of atoms Atoms
1 2 1 to 2
2 2 3 to 4
Part 1 self - energy (ONE-CENTER(E-E + E-N) + EE): -485.2398
Part 2 self - energy (ONE-CENTER(E-E + E-N) + EE): -485.2398
Part 2 - part 1 interaction energy (EE): -0.0388 [Note on interaction energies](#)





*** SUMMARY OF ENERGY PARTITION ***		

ONE-CENTER TERMS		
ELECTRON-NUCLEAR (ONE-ELECTRON)	-1617.2882	EV
ELECTRON-ELECTRON (TWO-ELECTRON)	670.5412	EV
TOTAL OF ONE-CENTER TERMS	-946.7470	EV

TWO-CENTER TERMS		
RESONANCE ENERGY	-24.2044	EV
EXCHANGE ENERGY	-9.6751	EV
EXCHANGE + RESONANCE ENERGY:	-33.8795	EV
ELECTRON-ELECTRON REPULSION	375.2846	EV
ELECTRON-NUCLEAR ATTRACTION	-789.4387	EV
NUCLEAR-NUCLEAR REPULSION	424.2622	EV
TOTAL ELECTROSTATIC INTERACTION	10.1081	EV
GRAND TOTAL OF TWO-CENTER TERMS	-23.7714	EV

ETOT (EONE + ETWO)	-970.5184	EV

EPS=*n*.*nn*

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. A [test](#) is provided to allow COSMO to be validated.

To alter the Van der Waals radius, use [VDW](#).

See also [RSOLV](#), [DISEX](#), [COSWRT](#), and [NSPA](#).

ESP

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

ESPRST restarts a stopped [ESP](#) calculation. Do not use with [RESTART](#).

ESR

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.

Spin density for state Y_j , calculated as spin, S_A , on atom A, is given in terms of contributions from the M.O.s of the active space y_i , $y_i = S_{ij}$ as:

$$S_A = S_i^A j_{li}^2 S_i S_i,$$

where the S_i are the contributions of spin from each M.O., expressed as a sum over the microstates of the C.I., $Y_j = S c_{kj} Y_j$:

$$S_i = S_k (O_i^{ak} - O_i^{bk}) c_{kj},$$

Where O_i^{aj} is the alpha or "spin up" occupancy of M.O. y_i , in microstate Y_j . Although j_{li}^2 is obligate positive, S_i can be positive or negative, therefore S_A can be positive or negative, although it is unlikely to be very negative, and the sum over all atoms must equal $2M_S(Z)$

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

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.

EXCITED is limited to restricted Hartree Fock systems, so [UHF](#) should not be present.

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

Normally, all semiempirical parameters are taken from the module 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. If the file-name contains spaces, then enclose the filename in quotation marks, e.g. `EXTERNAL="E:\Test Data\Parameter Set"`. When MOPAC is used with a GUI, e.g. Chem-3D, always use the full path name for the external file, i.e. use, e.g., `EXTERNAL="E:\Test Data\Parameter Set"` and *not*, e.g., `EXTERNAL="Parameter Set"`.

`<filename>` contains 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, F0SD, G2SD, ALP, FNnm, 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 the Figure:

Figure: Use of `EXTERNAL=name`

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			
Line 13:	BETAD	TC	-
5.72433498			
Line 14:	F0SD	TC	
5.43488602			
Line 15:	G2SD	TC	
1.10687502			
Line 16:	XFAC_Ag	I	2.0
Line 17:	ALPB_C	Rb	2.0

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.

$$\mathbf{FIELD} = (n.nn, m.mm, l.ll)$$

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. The potential arising from the field is zero at the origin of Cartesian space, i.e. $V=0$ at $(x=0.0, y=0.0, z=0.0)$. At any other point, (x,y,z) , the effect of $\mathbf{FIELD} = (n.nn, m.mm, l.ll)$ is to produce a potential equal to $V = (x.n.nn + y.m.mm + z.l.ll)$ volts.

The effect on a molecule depends on the orientation of the molecule relative to the field. Thus a hydrogen molecule would be polarized slightly by a field parallel to the axis, but unaffected by a field perpendicular to the axis. The polarization would be to produce a small positive charge on the hydrogen in the (relatively) negative part of the field, and a small negative charge in the positive part of the field. Any polarization would, of course, result in the energy being lowered.

Because molecules have no net charge, the effect of an applied field is independent of the location of the molecule. That is, if the field is defined by $\mathbf{FIELD} = (1.0, 0.0, 0.0)$, the energy of a molecule centered at $(+100, 0, 0)$ would be the same as that at $(-100, 0, 0)$ and the same as that at $(0, 0, 0)$. Put another way, the energy would be unaffected by motion in any direction, including in the direction of the field.

The geometries of small molecules are only slightly affected by fields in the order of 1 volt per Angstrom. This is because such a molecule is, of its nature, small. The change in potential from one end of such a molecule to the other would only be a few volts. Larger systems are affected much more - a system that extends over 20Å, say, would experience a potential change of 20 volts, if the field were along that axis. Macroscopic changes, such as Zener breakdown, are the result of ions being accelerated by the field. When these ions collide with neutral molecules, the molecules become ionized, e.g. $e^- + N_2 \Rightarrow 2e^- + N_2^+$.

On the other hand, an ion will be strongly affected by such a field. For example, F^- at coordinates 1.00, 0.00, 0.00, would be stabilized by 1.0eV if a field were to be applied by specifying $\mathbf{FIELD} = (1.0, 0.0, 0.0)$. This is a direct consequence of the net charge on the system interacting with the electric field: a charge of $+x$ at a point in space where the potential is y volts would have an energy due to the potential of $x.y$ eV. This is the origin of the term electron-volt.

A useful exercise is to monitor the behavior of an ion in an electric field, using the [DRC](#):

Let the field be $\mathbf{FIELD} = (1, 0, 0)$, let the ion be F^- , and have a mass of 19 amu. The data set:

```
DRC LARGE T-PRIO=0.1 CHARGE=-1 FIELD=(1,0,0) GNORM=0 CYCLES=75
Fluoride ion
In An Electric Field
F 0.0 0 0.0 0 0.0 0
```

would be suitable here. After about 63fs, the ion would have moved through 1 Angstrom, and the potential energy would have dropped by 23.06 kcal/mol. Conservation of energy requires that the kinetic energy equals 23.06 kcal/mol. Using $E = (1/2)MV^2$, implies that the velocity is:

$$V = (2 \cdot 4.184 \cdot 10^{10} \cdot 23.06 / 19) \text{ cm/sec} = 3.19 \cdot 10^5 \text{ cm/sec} = 3.19 \text{ km/sec.}$$

The $4.184 \cdot 10^{10}$ is the [conversion](#) from kcal/mol to ergs per mol.

All other quantities, such as the acceleration ($5.06 \cdot 10^{18} \text{ cm/sec}^2$ or $5.06 \cdot 10^{13} \text{ km/sec}^2$), follow simply. Used in this way, the [DRC](#) is useful for modeling ion implantation.

FILL=*n*

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 using [C.I.=*n*](#) and [ROOT=*m*](#) instead.

FLEPO

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

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

FOCK

The final Fock matrix is to be printed. If [DEBUG](#) is present, then every Fock matrix is to be printed.

FREQCY

Print out the symmetrized Hessian matrix in a [FORCE](#) calculation.

Symmetry is used to accelerate the construction of the Hessian matrix, and also to symmetrize the Hessian, once it is constructed. The effect of symmetrization is to make the resulting eigenvectors (normal coordinates) irreducible representations of the associated point group. If this is not done, then there is the potential for the normal coordinates to be unrecognizable as irreducible representations.

To turn off the use of symmetry, add [NOSYM](#).

FORCE and FORCETS

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.

Calculating the Hessian for a large system takes a long time, and often the only reason for running a `FORCE` calculation is to verify that the system is a transition state. To speed up this calculation, `FORCETS` is provided. The `FORCETS` calculation builds a Hessian for the atoms involved in the transition state. This will be used in generating vibrations for the transition state. If the system is a genuine transition state, then there will be one imaginary vibration, indicated in the output as a "negative" vibration. Its value will be within a few percent of the value that would be obtained if a full calculation were done. The imaginary vibration should involve the atom(s) that move during the reaction. All other vibrations should be positive, but their value is not useful, because they would involve atoms other than those in the transition state.

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 [LET](#) and [TRANS](#). 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 - CHECK

Normally, a calculation would run without any problems. In some jobs, the results might initially appear to be correct, but on further examination, certain subtle faults could be detected. These faults can automatically be found by adding `GEO-CHECK`, in which case the job would be stopped if such errors were detected. The types of errors that could be detected are:

Dynamic Jahn-Teller Effect

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. An example would be methane, if a C.I. were set up involving two M.O.s. The M.O.s are three fold degenerate, therefore such a C.I. would be invalid, although the effect of the C.I. would be very small. A more important example is found in the octahedral transition metal complexes in which dynamic Jahn-Teller effects can occur, e.g. $[\text{Ti}^{\text{III}}\text{F}_6]^{3-}$, a d^1 system, in which the single "d" electron is shared equally by the three t_{2g} molecular orbitals. If the C.I. involves some, but not all, components of the degenerate set of M.O.s, then the system will Jahn-Teller distort, and the octahedral symmetry of the wave-function will be lost. To detect such a condition, add `GEO-CHECK`. This is the most important use of this keyword.

If a transition state is being calculated in internal coordinates and all the coordinates are flagged for optimization, sometimes the internal coordinate gradients can all be very small and the Cartesian coordinate gradients be large. This is a very rare condition. If it occurs, there will be a warning message printed. To kill such a job, add `GEO-CHECK`.

If the geometry is specified in internal coordinates and large rings exist and the [BFGS](#) optimizer used, the optimization might be slowed down and the geometry might not be optimized correctly. To avoid wasting CPU time, `GEO-CHECK` should be specified.

GEO-OK

Normally the program will stop if certain errors occur. If `GEO-OK` is present, the job will continue, but there is an increased probability that the results will not be meaningful. The errors that are affected are:

If two atoms are within 0.1 Ångstroms of each other.

In practice, most jobs that terminate due to these checks contain errors in data, so caution should be exercised if `GEO-OK` is used.

In [MOZYME](#) calculations, if the supplied charge is incorrect and `GEO-OK` is present, it will be replaced by the correct charge, and the calculation allowed to continue.

GEO_REF=" *text* "

GEO_REF has two uses - to improve X-ray geometries and to help locate a transition state.

Improving X-ray geometries

The accuracy of native structures, X-ray and NMR, can be increased by performing a limited geometry optimization, starting with the conditioned native structure.

The optimized geometry is a hybrid of the PM6 geometry, assuming that PM6 is used, and the native geometry. The hybrid heat of formation, H' , is given as:

$$H' = DH_f(\text{PM6}) + cS(X_i - X_i(0))^2$$

That is, at each geometry, a restraining potential is added to the PM6 heat of formation. The proportionality constant " c " is 3 kcal/mol/Å².

The effect of the restraining potential is to reduce PM6 errors in the secondary, tertiary, and quaternary structures. It has very little effect on the primary structure, e.g., bond lengths and angles.

The effect of the restraining potential can be illustrated using crambin, 1CBN. The difference in heat of formation of crambin, after preconditioning (optimizing the positions of the hydrogen atoms), and after complete geometry optimization, is ~360 kcal mol⁻¹. The RMS change in geometry is about 0.7 Ångstroms. If the geometry is optimized using GEO_REF, then the heat of formation drops by ~310 kcal mol⁻¹, and the RMS geometry change is ~0.1 Ångstroms. That is, by allowing the X-ray geometry to change by 0.1 Ångstroms, about 86% of the strain energy in the X-ray structure is removed. Put another way, the resulting geometry is seven times more accurate, in terms of chemistry, than the X-ray structure.

"text" is the complete path to the reference X-ray geometry. A Windows example would be

GEO_REF="M:\data_sets\CRAMBIN_1CBN_X-ray.mop", a Linux, Unix, and Mac example would be

GEO_REF="/Users/jstewart/data_sets/CRAMBIN_1CBN_X-ray.mop" To change the proportionality constant from 3, put the new value after the reference data set, e.g., if a value of 10 is wanted, use:

GEO_REF="M:\data_sets\CRAMBIN_1CBN_X-ray.mop"10 OR

GEO_REF="/Users/jstewart/data_sets/CRAMBIN_1CBN_X-ray.mop"10

Locating transition states

GEO_REF can also be used for moving a reactant or product geometry in the direction of the transition state. Consider two data sets, reactant.mop and product.mop in folder M:\, in which the heat of formation of the optimized product geometry is lower than that of the optimized reactant geometry. The product geometry can be moved in the direction of the transition state by using keyword GEO_REF="M:\reactant.mop" The .arc file can then be edited to give a new data set, product_on_slope.mop. The reactant geometry can then be moved in the direction of this new geometry by using keyword GEO_REF="M:\product_on_slope.mop" Again, edit the .arc file to give reactant_on_slope.mop. Why was the product moved first? Because by moving it towards the reactant geometry, its heat of formation would rise in proportion to the distance to the reactant geometry. When the reactant geometry is moved towards the product geometry on the slope, the distance from the starting reactant geometry to the product geometry on the slope is less, so the rise in energy would be less.

If keyword TS is present in a GEO_REF calculation, the optimization is not run. Instead, the two geometries are averaged, and the result written to a new file <file>.new.

For the current exercise, a good approximation to the transition state can then be generated from the data set reactant_on_slope.mop by using keywords TS and GEO_REF="M:\product_on_slope.mop"

GNORM=*n . nn*

The geometry optimization termination criteria (see [Criteria](#)) 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 kcal/mol/Ångstrom, 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. The keyword GNORM defines the criterion for an optimized geometry, GNORM is the value calculated during a geometry optimization, and is printed in the output at the end of each cycle. When GNORM, drops below the level set by GNORM=n.nn, the geometry optimization will terminate.

GRADIENTS

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 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

Information needed to generate electron density contour maps can be written to an unformatted file, <name>.gpt, 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.

GRAPHF

Information needed by program [MOPETE](#) and [Jmol](#) can be written to a formatted file, <name>.mgf, by using the keyword GRAPHF.

For an example of a Jmol display of formaldehyde, download [CH2O.zip](#), extract CH2O.mgf, then open it using Jmol.

The layout of the data generated by GRAPHF is as follows::

- Block 1, 1 line: Number of atoms (5 characters), plain text: "MOPAC-Graphical data", and version number
- Block 2, 1 line per atom: Atom number (4 characters), Cartesian coordinates (3 sets of 12 characters), net atomic charge (9 characters)
- Block 3, 1 line per atom: Orbital exponents for "s", "p", and "d" Slater orbitals. (3 sets of 11 characters)
- Block 4, All the molecular orbital coefficients in the order M.O. 1, M.O.2, etc. (5 data per line, 15 characters per datum, FORTRAN format: 5d15.8) The number of values in this block is equal to the square of the number of orbitals.
- Block 5, Lower half of the inverse-square-root of overlap matrix. The number of values in this block is equal to (number of orbitals*(number of orbitals + 1))/2.
- Block 6, Eigenvalues (9 characters) and occupancies (2 characters) of the M.O.s, and a counter (ignore)
- Block 7, The word "Keywords:" followed by all the keywords used by MOPAC

Here is an example of the formatted graphical output for formaldehyde:

```
      4 MOPAC-Graphical data Version 2007.7.156W
      8      0.0000000      0.0000000      0.0000000      -0.4225
      6      1.2108153      0.0000000      0.0000000      0.2364
      1      1.7927832      0.9304938      0.0000000      0.0931
      1      1.7927832      -0.9304938      0.0000000      0.0931
      5.4217510      2.2709600      0.0000000
      2.0475580      1.7028410      0.0000000
      1.2686410      0.0000000      0.0000000
      1.2686410      0.0000000      0.0000000
      0.72990985D+00 0.34171335D+00 -0.19374693D-15 0.54377670D-15 0.45669941D+00
-0.33435297D+00 -0.27755576D-15 0.13737285D-14 0.12267301D+00 0.12267301D+00
      0.45424800D+00 -0.23179137D+00 0.31658703D-16 -0.26435458D-14 -0.60169277D+00
-0.33466696D+00 0.66613381D-15 -0.67319691D-14 -0.36461990D+00 -0.36461990D+00
-0.24096393D-15 -0.18185797D-16 -0.50244609D+00 0.12998172D-13 -0.33007180D-15
-0.10049199D-15 -0.68168942D+00 0.17933349D-13 -0.37606346D+00 0.37606346D+00
-0.43340632D+00 0.68895964D+00 0.37047188D-15 -0.84851588D-13 -0.43845362D-01
-0.47440245D+00 0.41633363D-15 -0.64902343D-13 -0.23506752D+00 -0.23506752D+00
      0.44516789D-13 -0.72063092D-13 -0.21470630D-13 -0.81949389D+00 0.12158810D-13
      0.54042566D-13 -0.15068204D-13 -0.57308792D+00 0.23952170D-13 0.23876692D-13
      0.35735304D-15 -0.34520997D-15 0.84565001D+00 -0.22228335D-13 0.24980018D-15
-0.34694470D-15 -0.25935442D+00 0.62764233D-14 -0.32985705D+00 0.32985705D+00
      0.16389881D-14 -0.35679314D-14 0.15146189D-13 0.57308792D+00 0.13409761D-13
      0.10193538D-13 -0.22014771D-13 -0.81949389D+00 -0.79041107D-14 -0.86303860D-14
      0.11479452D+00 0.39476332D+00 0.44741771D-14 0.51532251D-14 -0.65360634D+00
      0.87827921D-01 -0.14821477D-13 -0.16346726D-13 0.44501076D+00 0.44501076D+00
-0.24468662D+00 -0.44609560D+00 -0.14069759D-13 0.35732089D-14 0.16364179D-01
-0.73718023D+00 0.52423343D-13 -0.11413266D-13 0.31418521D+00 0.31418521D+00
      0.20976276D-13 0.42237047D-13 -0.18006664D+00 0.50732138D-14 -0.16209256D-13
      0.57287508D-13 0.68413070D+00 -0.18474934D-13 -0.49977054D+00 0.49977054D+00
      0.10196267D+01
      0.40054833D-01 0.10920846D+01
      0.85750152D-15 0.11888902D-14 0.10190736D+01
      0.71659852D-15 0.12796171D-14 0.16599177D-14 0.10170846D+01
-0.45740611D-01 -0.17793295D+00 -0.40404331D-14 -0.49341528D-14 0.12185877D+01
      0.11293286D+00 0.19305307D+00 -0.36900138D-14 -0.44426269D-14 0.52644247D-01
      0.11269900D+01
-0.72340703D-14 -0.77045179D-14 -0.12640479D+00 -0.67769893D-14 0.58981014D-14
-0.72147181D-15 0.12120691D+01
```

GRAPH

-0.48273961D-14-0.59693754D-14-0.47569905D-14-0.10798284D+00 0.10678472D-13
0.37625447D-14-0.31733256D-14 0.10170846D+01
-0.14856782D-01-0.17687973D-01 0.24197372D-01 0.47201684D-15-0.26124232D+00
-0.16391581D+00-0.30192832D+00-0.53944178D-14 0.12216511D+01
-0.14856782D-01-0.17687973D-01-0.24197372D-01-0.33803516D-14-0.26124232D+00
-0.16391581D+00 0.30192832D+00-0.19963219D-13-0.72366597D-01 0.12216511D+01
-32.0037 2 1
-23.5573 2 2
-17.1753 2 3
-15.1402 2 4
-14.7591 2 5
-10.2036 2 6
0.0321 0 7
4.0856 0 8
4.4599 0 9
4.7427 0 10
Keywords: PM6 SYMMETRY GRAPHF

HCORE

Print the final one-electron matrix in conventional SCF work. If [DEBUG](#) is present, print every one electron matrix.

HESSIAN

Print the Hessian in kcal/mol/Angstrom^2 from a geometry optimization. Keyword [EE](#) must also be present.

HESSIAN is useful in diagnostic work and as a source of a rough Hessian for other applications.

HESS=*n*

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 kcal/mol/Å² for bond-lengths, 500 kcal/mol/degrees² for angles, and 200 kcal/mol/degrees² for dihedrals. If Cartesian coordinates are used, all diagonal elements are set to 200 kcal/mol/Å².

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 [Hessian Matrix](#), 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.)

H-PRIORITY

In a [DRC](#) calculation, results will be printed whenever the calculated heat of formation changes by 0.1 kcal/mol.
Abbreviation: H-PRIO. See [Dynamic and Intrinsic Reaction Coordinates](#) for more details.

H-PRIORITY=*n . nn*

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

HYPERFINE

The quantity:

$$(1/3)(P_s^a - P_s^b + 2(y_s^a)^2)$$

is printed for all atoms in a radical when `HYPERFINE` is specified for a UHF system. The y_s^a refers to the highest occupied alpha-spin molecular orbital. This quantity is of use in predicting hyperfine coupling coefficients.

INT

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.

Notes

- If `INT` is used, the optimization flags are not changed. Therefore, before running a system with `INT` all the optimization flags should be set to "1", except for atoms 1, 2, and 3. For atom 1, all optimization flags should be set to "0", for atom 2, the second and third flags should be "0", and for atom 3 the third flag should be "0".
- Be very careful if some atoms have optimization flags set to "0". If the geometry supplied is in Cartesian coordinates, and an atom, say atom 10, has all three optimization flags set to "0", then that means "Do not change the Cartesian coordinates of atom 10." If that atom is converted to internal coordinates, the optimization flags will still be zero, but now the definition changes to "Do not change the bond length, angle, and dihedral of atom 10, with reference to the atoms used for the connectivity." Now atom 10 will move if the atom it is attached to moves. This is fundamentally different from the Cartesian coordinate definition.
- Any dummy atoms are automatically deleted if `INT` is present. This is because the first step in running `INT` is to convert the system to Cartesian coordinates. This involved deleting any dummy atoms. The next step involves conversion from Cartesian to internal coordinates.
- `INT` is a rapid and efficient way of deleting atoms from a molecule that is defined using internal coordinates. Re-label the atoms to be deleted as dummy atoms, and run `INT` with [0SCF](#).

See also [XYZ](#).

INVERT

When `INVERT` is present, all optimization flags are reversed. That is, if an optimization flag is "1", it will be set to "0", and if it is "0", it will be set to "1"

Refinement of a transition state for a large system runs faster when only the atoms near to the transition state are used in gradient minimization, and the rest are used in energy minimization. `INVERT` is useful for calculations of this type. The optimization flags for all atoms in the gradient minimization are set to "1", the rest are set to "0". In one calculation, a normal gradient minimization is run, in a second calculation a geometry optimization is run, with keyword `INVERT` present.

The two calculations can be combined into one job. For example, the first data set, for the gradient minimization, can be followed by a second data set consisting of keywords "[OLDGEO](#)" and "`INVERT`" plus any other keywords needed for the job.

Alternatively, make two complete data sets, one for gradient minimization and one for energy minimization. Run these data sets on different computers. Then edit the resulting ARC files to make up a new data set. This saves time, because the jobs run twice as fast.

IRC

An Intrinsic Reaction Coordinate calculation is to be run. All kinetic energy is shed at every point in the calculation. See [Dynamic and Intrinsic Reaction Coordinates](#) for more details.

By default, steps of 0.05Angstroms are used. This default can be changed using keywords:

To print reaction coordinate steps that are separated by constant energy, use [H-PRIORITY](#) or [H-PRIORITY=*n.nn*](#).

To print reaction coordinate steps that are separated by constant distance, use [X-PRIORITY](#) or [X-PRIORITY=*n.nn*](#).

Use [LARGE](#) to control print of internal coordinates and Cartesian coordinates.

IRC=*n*

An [Intrinsic Reaction Coordinate](#) calculation to be run. The following options are provided:

IRC=1 and IRC=1*

This option is intended for use with a transition state. The path from the transition state to either the reactants or the products is generated. The path generated depends on the coefficients of the eigenvectors of the normal mode. There is no easy a priori way of determining what the phase of these coefficients is, so try running IRC=1, and see if it goes the correct way. If not, use IRC=-1. If IRC=1* is used, then the entire [reaction path](#), from reactants or products, through the transition state, to products or reactants, is mapped. If DRC is present, then the system is given an initial perturbation in the positive normal coordinate, then the DRC is calculated.

IRC=-1 and IRC=-1*

This option is intended for use with any a transition state. The path from the transition state to either the products or the reactants is generated. The path generated depends on the coefficients of the eigenvectors of the normal mode. There is no easy a priori way of determining what the phase of these coefficients is, so try running IRC=-1, and see if it goes the correct way. If not, use IRC=1. If IRC=-1* is used, then the entire reaction path, from products or reactants through the transition state to reactants or products, is mapped. If DRC is present, then the system is given an initial perturbation in the negative normal coordinate, then the DRC is calculated.

IRC=+*n* or IRC=*n*

Normally used with [DRC](#), this option perturbs the system along the *n*'th normal mode. One quantum of kinetic energy is added, the value of the quantum being that of the associated vibrational energy. This option allows the exact normal mode to be mapped out. IRC=*n* with DRC can be used with both ground and transition state systems. If IRC=1 and DRC is used with transition states, it would map out the dynamic path from the transition state to reactants or products.

IRC=-*n*

The same as IRC=+*n*, except the initial perturbation is in the opposite direction. Except for transition states, this option can be duplicated by one of the other options described here.

When studying reaction paths involving transition states, one option, although implied logically from the above description of the IRC, should not be used. The option IRC=*n*, where *n* is not equal to 1 or -1, is meaningless. To see why, consider the option IRC=1. That maps the reaction path, and uses the lowest transition state normal mode, i.e. the mode with the imaginary frequency. Any other mode would simply return to the transition state, in other words, it would do nothing useful.

ISOTOPE

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.

ITER

Print details of the working in ITER.

ITRY=*nn*

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*

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 [Description of EF and TS](#).

KINETIC=n . nnn

In a DRC calculation *n.nnn* kcal/mol of excess kinetic energy is added to the system. At the start of a DRC calculation, the atoms are initially at rest. So kinetic energy cannot be added to the system. When KINETIC=n.nn is used, the DRC is first run for a few steps until the atoms have a significant velocity. This is when the kinetic energy amounts to 0.2 kcal/mol. When that happens, the kinetic energy is increased suddenly by *n.nn* kcal/mol. The next few steps will show errors in the output as the energy is added to the system. The use of 0.2 kcal/mol is arbitrary - any small number could be used.

If a system starts off with an energy of -100 kcal/mol, and KINETIC=10 is used, the DRC will be run with a total (kinetic plus potential) of -90 kcal/mol.

See [VELOCITY](#) and [Dynamic and Intrinsic Reaction Coordinates](#) for more details.

KING

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

LARGE

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](#), and [MECI](#).

When [COMPFG](#) 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`.

LBFGS

Optimize the geometry using the L-BFGS-B function minimizer. This is based on the BFGS optimizer, but it does not store the inverse Hessian, instead it is calculated as needed. Because of this, the L-BFGS-B method uses very little storage, and is therefore suitable for optimizing very large systems. It is not as efficient as the other optimizers, but for large systems, it is often the only method that can be used.

The L-BFGS optimizer is the default if 2000 or more variables are to be optimized. For such large systems, L-BFGS is more efficient than EF. If L-BFGS is not wanted, add keyword [EF](#)

LET

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 DH_f to rise. Obviously, the DH_f should decrease on every step, but sometimes, particularly with very flat potential energy surfaces, motion in response to the gradients might not result in a decrease in energy. By using LET, small increases in DH_f are allowed. This will allow the gradient minimum to be reached. This is a well-defined point, whereas the minimum in energy is not, so in this case the use of LET is justified. Note that the use of LET will not result in motion to a different minimum.

LEWIS

Print the topography and Lewis structure. When `LEWIS` is present, the run is stopped after the structure is printed.

This keyword is useful for checking that a Lewis structure exists. Quite complicated systems can be supplied, such as buckyball (C_{60}), carbon nanotube (buckytube), graphite, phthalocyanines, and diamond.

A Lewis structure consists of single, double and triple bonds, lone pairs, anionic and cationic sites, and positively and negatively charged atoms (ions). The method used in MOPAC is quite powerful, and except for rare cases, gives a reasonable structure. The exceptions are:

- In buckytubes, the assumption is made that a graphitic lattice, rather than an extended conjugated poly-acetylene type p bonding structure exists. In some cases, the only unionized structure involves using an extended conjugated poly-acetylene motif.
- (To be added, as necessary)

When large organic molecules are being studied, a preliminary calculation to generate the Lewis structure should be run before doing serious work. This will identify all ionized sites in the system. *Check that these sites are correct!* See also [CHARGES](#)

If only the topography is important, then also add keyword [USCF](#). This will stop the run after the topography, but before the Lewis structure is printed. This option is useful when using keyword [CVB](#) to correct faults in the topology.

Related key-words: [CVB](#), [METAL](#), [VDWM](#), [CHARGE](#), [CHARGES](#), and [MOZYME](#)

See also: [Lewis Structures](#), [MOZYME introduction](#)

LINMIN

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

The occupied eigenvectors are transformed into a localized [38] set of M.O.s by a series of 2 by 2 rotations which maximize $\langle y^4 \rangle$. The value of $1/\langle y^4 \rangle$ is a direct measure of the number of centers involved in the MO: Thus, the value of $1/\langle y^4 \rangle$ is 2.0 for H_2 , 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.

Localized orbitals are also known as Natural Bond Orbitals or NBO.

LOG

By default, the LOG file is suppressed. If LOG is specified, a log file will be generated.

In some calculations, an abbreviated log file is generated. When LOG is specified, a full log file will be created.

MECI

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 keywords such as [C.I.=n](#) and [OPEN\(n,n\)](#).

$$\mathbf{MERS} = (n_1, n_2, n_3)$$

Information on a solid is written to a file for use by program BZ, if `MERS` is present on the key-word line. The three integers, n_1 , n_2 , and n_3 define the number of fundamental unit cells in each dimension in a cluster. See [Description of BZ](#) for more information on the values of n_1 , n_2 , and n_3 . and [MAKPOL](#) for how to construct a data set for use by BZ

The definition of a 'mer' might be of interest. In a polymer or other extended system, the fundamental repeating unit is the mer. For some simple polymers, such as polyethylene, the mer can also be identified with the monomer, in that both the mer and the monomer have the same formula. In other polymers, the mer is not the same as the monomer--condensation polymers are a good example, where the monomer loses water on forming the polymer. In order to be general, therefore, the unit of a polymer is not the monomer, but instead is the mer.

MICROS=n

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.

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 the Table.

Table: 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

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 the Table 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 (t₂)³(t₂⁻)¹(t₂^{*})¹(t₂^{*})⁰.

The format is defined as 20I1 so that spaces can be used for empty M.O.s.

MINI

When very large molecules such as enzymes are modeled, the output can become very large. To reduce this, use `MINI`. When `MINI` is present, only atoms whose coordinates are flagged with a "1" will be printed. Because the flag "1" is used in deciding which atoms to print, the old optimization flag "1" should be replaced by a "2". For example, if a transition state optimization calculation is run, then the atoms involved would be flagged with optimization symbol "2", and all the nearby atoms flagged with a "1". More distant atoms would be flagged with "0".

When this system is run, only atoms flagged with "1" or "2" will be printed, and only atoms flagged with a "2" will be moved during the transition state optimization.

`MINI` is useful in following reaction paths, [IRC](#), and [DRC](#) trajectories.

This is a new keyword. Please send a note of any anomalous behavior.

MINMEP

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

MMOK

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

Use the MNDO Hamiltonian.

MNDOD

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

MODE=*n*

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

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

MOLSYM

Print details of the working in MOLSYM, the routine that works out the symmetry point-group of the molecule. Point-groups are identified using a set of 20 integers. These are 0 if the associated operation is absent, 1 if the operation is present. The operations are:

Operation Number	Operation
1	C ₂ (X)
2	C ₂ (Y)
3	C ₂ (Z)
4	s(XY)
5	s(XZ)
6	s(YZ)
7	inversion
8	C ₃
9	C ₄
10	C ₃
11	C ₆
12	C ₃
13	C ₈
14	S ₄
15	S ₆
16	S ₈
17	S ₁₀
18	S ₁₂
19	1 if cubic
20	1 if infinite

A tetrahedral molecule, for example, would have the pattern:

Operation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Value	0	0	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0

This indicates that operations C₂(Z), s(XZ), s(YZ), C₃, and S₄ are present, and that the system is cubic.

The pattern of operations is unique for each point-group, and is used by subroutine cartab to identify the point-group.

MOPAC

The keyword `MOPAC` is provided in order to allow compatibility with earlier versions of MOPAC which did 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

The keyword `MOZYME` replaces the standard SCF procedure with a localized molecular orbital (LMO) method. `MOZYME` was developed to allow very large organic compounds, specifically enzymes, to be easily calculated. The time required for a SCF calculation increases approximately linearly with the size of the system, see [literature on MOZYME](#).

Notes, warnings and cautions concerning MOZYME calculations

Although a job that uses the `MOZYME` technique should give results that are the same as conventional SCF calculations, in practice there are differences. Most of these differences are small, but in some jobs the differences between `MOZYME` and conventional SCF calculations, particularly the calculation of DH_f , can be large. A single point calculation using `MOZYME` and conventional methods would produce essentially the same DH_f , and for the purposes of this discussion the results of a single SCF calculated by both methods can be regarded as being identical. The problem with different DH_f occurs when multiple SCF calculations are performed, this is the situation in a geometry optimization or reaction path calculation. In such calculations, the LMOs that result from an SCF calculation are used as starting LMOs in the next SCF calculation. In the first SCF calculation, the starting LMOs are exact - they form rigorously orthonormal sets, one for the occupied and one for the virtual sets. At the end of the SCF, small errors arising from truncation of the LMOs and incomplete pairwise rotations give rise to a small degradation of the orthonormal nature of the LMOs. In a single SCF, this degradation is unimportant, but when many SCF calculations are done, the loss of orthonormality increases steadily. This manifests itself as an error in the calculated DH_f and to a much smaller extent in the gradients, and therefore, by implication, in the geometry. The loss of orthonormality could be corrected by re-orthogonalizing the LMOs, but the CPU cost of this is great, and re-orthogonalization is not done by default, although it can be done if desired using [REORTHOG](#). Fortunately, a very simple procedure exists to completely correct this error: After any long run involving many `MOZYME` SCF calculations, use the final geometry generated as the starting point for a `1SCF` calculation, and then use the DH_f from that calculation. This strategy should be used:

(A) In global optimizations.

(B) In transition state location runs.

(C) At the end of IRC runs.

Do not use [OLDENS](#) as that would re-use the now-inaccurate sets of LMOs, and thus defeat the purpose of doing the `1SCF` calculation. As mentioned above, the errors in the gradient are small, so the geometry is essentially unaffected by the loss of orthonormality. However, it is still a good practice to optimize geometries in three or more separate runs, if only to provide an opportunity to check that the calculation is proceeding as intended.

During geometry optimizations, the error in DH_f caused by the deterioration of the LMOs can result in the energy rising near the end of the run. If this happens, the lowest energy structure will be output, instead of the last structure calculated.

By default, the M.O.s printed are LMOs. If canonical M.O.s are needed, use keyword [EIGEN](#). `EIGEN` uses a large amount of memory and might not work if the system is too large. Even if it does work, it might take a lot of CPU time, so `EIGEN` should only be used with `1SCF`.

Memory considerations:

With 1Gb of RAM, systems of up to 10,000 atoms can be run without paging. Above about 11,000 atoms, paging becomes severe, and jobs take longer than necessary. With 2Gb of RAM, systems of up to 18,500 atoms can be run without significant paging. Above that, the interatomic distance array cannot be created. For a system of 19,000 atoms, the interatomic distance array would use 4×19000^2 bytes, or 1.4Gb. When `MOZYME` is used for large systems, the machine becomes very slow for other activities.

If polarizabilities are required, use [STATIC](#). If keyword [POLAR](#) is used, then [STATIC](#) will be used instead.

The total charge on the system must be correct. If it's not, the charge will be changed to that predicted by the Lewis structure. To prevent the charge being changed, use [CHARGE=n](#), even in cases where the system is neutral, i.e., CHARGE=0. If CHARGE=n is not present, then the charge predicted by the Lewis structure will be used.

Limitations of MOZYME

- 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. Ions that are definitely open-shell, such as Cr(III), cannot be run.
- The results are not so precise, so for runs that need high precision (such as [FORCE](#) calculations), MOZYME should not be used.
- For large systems, the recommended geometry optimizer is [LBFGS](#). This is a modified BFGS optimizer designed to minimize memory usage, and is the default for systems of 2000 parameters or more. If LBFGS 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.
- Electrostatic Potentials cannot be calculated, that is [ESP](#) cannot be used, however the Parametric Molecular Electrostatic Potential, [PMEP](#), can be used.

Recommended use of MOZYME

MOZYME can be used for simple geometric calculations, such as geometry optimization, transition state location, and intrinsic 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 keyword [RAPID](#) should be considered.

For calculation of vibrational frequencies, frequency-dependent polarizability, 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.

[Examplex of calculations involving proteins](#)

MS=*n*

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

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.

$$N^{**2}=n.nn$$

In excited state [COSMO](#) calculations, the refractive index, squared, is set to *n.nn*. By default, a value of 2.0 is used. The effect of N^{**2} is to modify the dielectric energy.

NLLSQ

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.

NOANCI

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, because it evaluates the gradient using finite differences involving complete SCF calculations. In contrast, 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 [GRADIENTS](#).

NOANCI can also be used for closed-shell RHF calculations when there is a need to avoid the standard diatomic derivatives. Thus, when [PM6-DH2](#) is used, there are known small errors in the gradient calculation that prevent the exact energy minimum being reached. When these errors are avoided by using NOANCI, the exact energy minimum is obtained.

NOLOG

In some calculations, a log file is generated. If this is not wanted, it can be suppressed completely by `NOLOG`.

NOMM

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

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

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.

NOOPT, NOOPT-X

Turn off all optimization flags for atoms of type X. This keyword is useful for partial geometry optimizations. If "-X" is missing, turn off all optimization flags. If [OPT-X](#) is present, optimization flags can be turned on after NOOPT has been run .

NOREOR

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.

In a `FORCE` calculation, `NOREOR` will prevent the molecules being reoriented to line up the moment of inertia axes with the Cartesian axes.

NOSYM

By default, symmetry point-groups are automatically assigned, and used in identifying the irreducible representations of molecular orbitals, vibrations, and states, and in accelerating the construction of the Hessian in **FORCE** calculations.

Sometimes a system almost has the symmetry of a point group higher than C_1 . Minor deviations from exact symmetry are allowed, but these sometimes 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 any reason to suspect that the point-group symmetry features are causing problems.

A second type of symmetry can be used by specifying **SYMMETRY**. This is distinct from the default of using the point-group of the system to label irreducible representations and to accelerate **FORCE** calculations. When **SYMMETRY** is used, the symmetry relations between coordinates are defined by the user. These can be used for defining the point-group of a molecule, for example the D_{6h} of benzene, but no point-group theory is used when **SYMMETRY** is specified.

NOTHIEL

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

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.

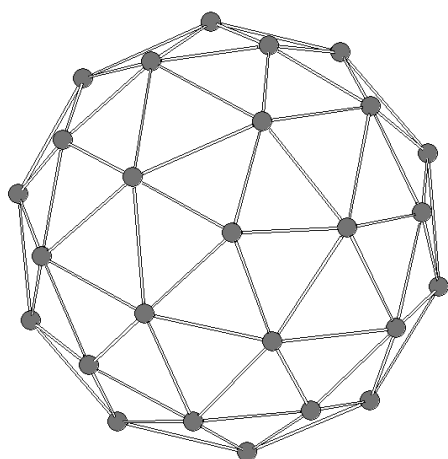
NSPA=*n*

Used for the COSMO method (see [EPS](#)) to set the number of geometrical segments per atom 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 $3^i \times j^2 \times 10 + 2$ should be used (e.g., 92, 122, or 162). See also [DVFILL](#).

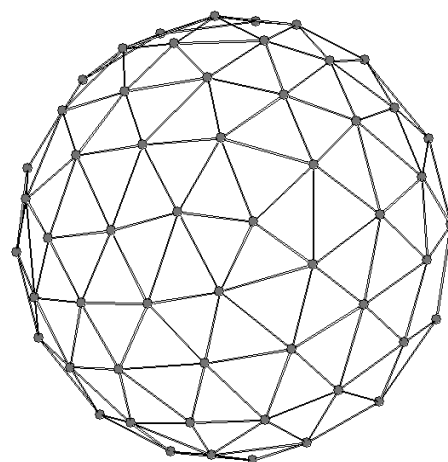
Examples of NSPA

In these examples, the surface of a single atom is outlined in terms of geometric segments. All segments are triangles, although some lines indicating edges in the following diagrams are missing.

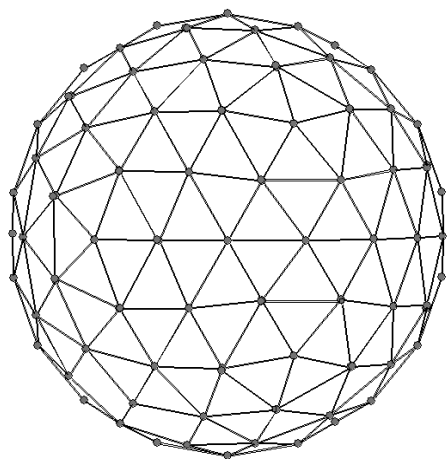
NSPA=42



NSPA=92



NSPA=122



NSURF

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

OCTET

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.\(n1,n2\)](#) 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.

OLDCAV

In a [COSMO](#) calculation, use the old method for calculating the solvent accessible surface (SAS). For more details on this method, please see: www.cosmologic.de

OLDENS

A density matrix produced by an earlier run of MOPAC 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.

The density matrix contains information on the electron density distribution only. It does not contain any information on the geometry.

The opposite of OLDENS is [DENOUT](#)

See also [RESTART](#).

OLDFPC

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

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 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*

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)

RHF description: The M.O. occupancy during the SCF calculation can be defined in terms of doubly occupied, empty, and fractionally occupied M.O.s. By default, RHF SCF calculations are run using doubly occupied M.O.s, with, at most, one singly occupied M.O. For some systems, the symmetry of the M.O.s can be preserved only if two or more M.O.s have fractional occupancies. Such fractional occupancies can be defined using 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; The ratio of n_1 to n_2 is restricted to $2 > n_1/n_2 > 0$. Example: if OPEN(3 , 4) were used then the occupancy near the HOMO-LUMO gap would be ...2, 2, 2, 0.75, 0.75, 0.75, 0.75, 0, 0, 0, ...

The two p M.O.s in molecular oxygen would not be degenerate if a closed-shell SCF were run. In order to maintain the degeneracy, OPEN(2 , 2) would need to be used.

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 O_2 , or if there is a very small band-gap (such as in metal clusters).

Examples of OPEN(n_1 , n_2) are given in the Table. 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 [136] in a MECI calculation when OPEN(n_1 , n_2) is used. See also C.I. = n.

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.

UHF description: The M.O. occupancy during the SCF calculation can be defined in terms of singly occupied, empty, and fractionally occupied M.O.s. By default, UHF SCF calculations are run using singly occupied M.O.s. When OPEN(n_1 , n_2) is present, the highest n_2 a molecular orbitals are each given a population of n_1/n_2 electrons. To use OPEN(n_1 , n_2) in a UHF calculation, UHF must also be present. Because fractional electrons are used, small errors in energy, on the order of 1-2 kcal/mol, are introduced when OPEN(n_1 , n_2) is used. These errors are *not* corrected by C.I., however they are small and should not affect the results significantly.

The main use of OPEN(n_1 , n_2) in UHF calculations is to mimic the dynamic Jahn-Teller effect in transition metal complexes. Without OPEN(n_1 , n_2) an octahedral complex system that could undergo Jahn-Teller distortion would be distorted. However, in many X-ray structures, systems that should Jahn-Teller distort do not in fact distort and instead retain high symmetry, typically O_h . This high symmetry can be regarded as a dynamic J-T effect.

Transition metal complexes are, as their name suggests, complicated, and correctly defining the various J-T states is also complicated. The following set of examples is provided to illustrate how to define keywords for various octahedral systems. These systems are assumed to have a three-fold degenerate set of "d" M.O.s, with dominant contributions from the d -orbitals d_{xy} , d_{yz} , d_{xz} , that can be related to the t_{2g} set of point group O_h , and a two-fold degenerate set, composed mainly of the other two d -orbitals; these are equivalent to the e_g set of point group O_h .

d^0 : No keyword needed - closed shell, i.e., $^1A_{1g}$.

d^1 : keywords: "UHF OPEN(1 , 3)" One a d electron in a t_{2g} set of M.O.s. The resulting state is $^2T_{2g}$.

d^2 : keywords: "UHF OPEN(2 , 3) MS=1" Two a d electrons in a t_{2g} set of M.O.s. The resulting state is $^3T_{1g}$.

d^3 : keywords: "UHF MS=1.5" The state is $^4A_{2g}$.

d^4 : keywords for low spin: "UHF OPEN(1 , 3) MS=-1" Because OPEN(n_1 , n_2), applies only to a electrons, and because there are four electrons in a t_{2g} set of M.O.s, the only way to have a fractional population of the a set of M.O.s is to have the b set filled, i.e., a^1b^3 . By using "MS=-1" the number of b electrons is defined as being two

more than the number of a electrons The resulting state is T_{1g} .

d^4 : keywords for high spin: "UHF OPEN(1,2) MS=2" In high-spin d^4 complexes, there are three a electrons in the t_{2g} set of M.O.s, and a single a electron in the two-fold degenerate e_g M.O. "MS=2" results in there being four more a than b electrons. The resulting state is 5E_g .

d^5 : keywords for low spin: "UHF OPEN(2,3) MS=-0.5" See d^4 low spin. This gives rise to a $^2T_{2g}$ state.

d^5 : keywords for high spin: "UHF MS=2.5" The state is $^6A_{1g}$.

d^6 : No keyword needed - closed shell, i.e., $^1A_{1g}$.

d^6 : keywords for high spin: "UHF OPEN(1,3) MS=-2" The resulting state is $^5T_{2g} = A_{2g}$ (from the b^2 of e_g symmetry)* A_{2g} (from the b^3 of t_{2g} symmetry)* T_{2g} (from the a^1 of t_{2g} symmetry).

d^7 : keywords for low spin: "UHF OPEN(1,2)" As with d^1 , the extra electron has, by default, a spin, and so the keyword "MS=0.5" is unnecessary. The resulting state is 2E_g .

d^7 : keywords for high spin: "UHF OPEN(1,2) MS=-1.5" With seven d-electrons and MS=-1.5 there are three more a than a electrons, i.e., a^2b^5 . The resulting state is $^4T_{1g}$.

d^8 : keywords: "UHF MS=1" The resulting state is $^3A_{2g}$.

d^9 : keywords: "UHF OPEN(1,2)" The resulting state is 2E_g .

d^{10} : No keyword needed - closed shell, i.e., $^1A_{1g}$.

A similar exercise could be done with the T_d complexes. As this is simple, albeit tedious, it is left as an exercise.

Table: Use of OPEN(n , m)					
UHF Keywords	RHF Keywords	Example	Number of M.O.s	No. of Electrons	State
MS=1	OPEN(2 , 2)	Twisted Ethylene	2	2	3A_2
OPEN(1 , 2)	OPEN(1 , 2)	O_2^+	2	1	2P_g
MS=-0.5 OPEN(1 , 3)	OPEN(5 , 3)	CH_4^+	3	5	2T_2
MS=1.5	OPEN(3 , 3)	$[Cr^{III}F_6]^{3-}$	3	3	$^4A_{2g}$
MS=2.5	OPEN(5 , 5)	$[Mn^{II}(H_2O)_6]^{2+}$	5	5	6A_1
MS=-0.5 OPEN(2 , 3)	OPEN(5 , 3)	$[Fe^{III}(CN)_6]^{3-}$	3	5	$^2T_{2g}$

OPT-X

Turn on all optimization flags for atoms of type X. This keyword is useful for partial geometry optimizations. If "-X" is missing, turn on all optimization flags. If [NOOPT-X](#) is present, optimization flags can be turned off before OPT is run .

P=n . nn

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 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 (Vol(0)) to the unit cell volume under a pressure of P=n . nn (Vol(n . nn) from:

$$B=n.nn*Vol(0)/(Vol(0)-Vol(n.nn))$$

Useful conversion factors are:

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

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

PDB

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*)

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

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.

If the data set includes comments (lines that start with an asterisk (*), see [example](#)), these will be printed in the PDB format. This is very useful in preserving the integrity of the PDB file.

If residues are missing, or if multiple proteins are present, keyword `START_RES(text)` should be used to indicate this. By default, each chain, or section of chain, will be given a different letter, starting with "A". This can cause problems, particularly if a chain has a gap in it: by default, MOPAC would give the two chain sections different letters. Chain letters can be explicitly assigned by using `CHAIN(text)`.

If hydrogen atoms have been added, these can be put in their correct place in the PDB file by using `RESEQ`.

If either `OSCF` or `RESEQ` is present, the PDB file written will not contain any non-PDB text, i.e. it will only contain PDB data.

PECI

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

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 s, p, and d bonds. The resulting "density matrix" is composed of the following basis-functions: s - s , p - s , p - p , d - s , d - p , d - d . The on-diagonal terms give the hybridization state, so that an sp^2 hybridized system would be represented as s - s : 1.0, p - s : 2.0, p - p : 1.0.

pKa

When pK_a is specified, the pK_a of hydrogen atoms attached to oxygen atoms is calculated and printed. For a large set of [compounds](#), the average unsigned error is about 0.31 pK_a units.

pK_a is calculated using the O-H distance calculated using PM6, and a charge calculated using a method specifically designed to reproduce the charge for pK_a .

Do not use COSMO in optimizing the geometry. Use PM6 only. When the pK_a calculation is done on the optimized PM6 geometry, the COSMO technique is switched on temporarily in order to get a better charge on the hydrogen atom(s). Once the pK_a calculation finishes, the COSMO is switched off again.

Known problem: If an ionizable hydrogen is within hydrogen-bonding distance of an atom with a large negative charge, as in the ene-ol form of acetylacetone, then the predicted pK_a will be much too negative. This is a consequence of the induced extra positive charge on that hydrogen.

PL

A tool for monitoring the behavior of SCF calculations is useful when they take too long, or even fail altogether. When keyword `PL` is present, the energy of the system and the rate of change of the electronic structure can be monitored iteration to iteration. This keyword is useful particularly when trying out various combinations of convergers such as [PULAY](#), [KING](#), [SHIFT](#), and [DAMP](#) to find the method that works best.

The rate of change in the electronic structure is given by the quantity PLS in the output. For example, in the line:

```
ITERATION      7 PLS=  0.168E-01  0.395E-06 ENERGY      -34.585270 DELTAE      -2.2653063
```

the alpha wavefunction changed by 0.0168 between iteration 6 and iteration 7. At the same time, the beta wavefunction changed by 0.000000395. The change in energy over these iterations is -2.265 kcal/mol.

If the calculation uses a restricted Hartree-Fock method, then the line would look like this:

```
ITERATION      7 PLS=  0.900E-02  0.000E+00 ENERGY      46.712559 DELTAE      -0.3203785
```

Now the change in total wavefunction between iteration 6 and 7 would be 0.009. The second number, here 0.000E+00, is not used and should be ignored.

The precise meaning of PLS depends on whether the calculation uses [MOZYME](#) or the conventional, default, method.

Definition of PLS for conventional SCF calculations

In conventional methods, PLS is the largest change in any density matrix element on two successive iterations in the SCF calculation. At self-consistency, this change drops to zero. For UHF calculations, both the alpha and beta density matrices are used, therefore two numbers are printed. The total density matrix is used in RHF calculations, and consequently the second number is not used.

Definition of PLS for MOZYME calculations

MOZYME calculations do not use a normal density matrix so the definition used above cannot be used. Instead, the largest Fock matrix element connecting an occupied LMO with any virtual LMO is used. This can be written as:

$$\text{PLS} = |\langle y^{\text{occ}} | F | y^{\text{vir}} \rangle|$$

Although this definition is fundamentally different from that in conventional work, in that data from only one iteration is used and the units are now electron volts and not electrons, the meaning is the same - at self-consistency all Fock matrix elements connecting occupied and virtual LMOs are zero. Therefore, from a user's perspective, the significance of PLS in both conventional and MOZYME calculations is the same: it is a measure of how far the system is from self-consistency.

PM3

The PM3 method [4] is to be used.

The PM3(tm) method is not in MOPAC, but it is so similar to PM3 that PM3(tm) calculations can be run using the EXTERNAL option. To run PM3(tm) calculations, use keywords PM3 and EXTERNAL=<filename>. The contents of the file <filename> would be the normal PM3(tm) parameters. An example of such a parameter set is as follows:

Table: Example of PM3(tm) parameter set (for Ti)

atnum	22
n	4. 0
uss	- 26. 45829779
upp	- 21. 17197024
udd	- 36. 34653108
betas	- 23. 30450933
betap	- 1. 07045796
betad	- 2. 69874461
zs	1. 04638304
zp	1. 25955299
zd	1. 31534641
alpha	1. 57689752
gss	12. 64676054
gpp	6. 92904760
gdd	13. 01302063
gsd	4. 28866787
hsd	1. 35838275
a1	- 0. 09046372
b1	4. 25904322
c1	1. 25471803
a2	- 0. 01221641
b2	4. 07610129
c2	2. 79140274
hform	112. 30000000
zval	4. 00000000

PM6

The PM6 method is to be used. This keyword is redundant, because the default method in MOPAC2009 is PM6.

PM6-DH+ [(n , m , k : l)]

Uses the PM6-DH+ procedure of M. Korth, ["Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields,"](#) J. Chem. Theory Comput., 2010, 6 (12), pp 3808–3816.

Abstract: Computational modeling of biological systems is a rapidly evolving field that calls for methods that are able to allow for extensive sampling with systems consisting of thousands of atoms. Semiempirical quantum chemical (SE) methods are a promising tool to aid with this, but the rather bad performance of standard SE methods for non-covalent interactions is clearly a limiting factor. Enhancing SE methods with empirical corrections for dispersion and hydrogen-bonding interactions was found to be a big improvement, but for the hydrogen-bonding corrections the drawback of breaking down in the case of substantial changes to the hydrogen bond, e.g., proton transfer, posed a serious limitation for its general applicability. This work presents a further improved hydrogen bonding correction that can be generally included in parameter fitting procedures, as it does not suffer from the conceptual flaws of previous approaches: hydrogen bonds are now treated as an interaction term between electronegative acceptor and donor atoms, “weighted” by a function of the position of H atoms between them, and multiplied with a damping function to correct the short- and long-range behavior. The performance of the new approach is evaluated for PM6, AM1, OM3, and SCC-DFTB as well as several force-field (FF) methods for a number of standard benchmark sets with hydrogen-bonded systems. The new approach is found to reach the same accuracy as the second-generation hydrogen-bonding correction with less parameters, while it avoids among other issues the conceptual problem with electronic structure changes. SE methods augmented this way reach the accuracy of DFT-D approaches for a large number of cases investigated, while still being about 3 orders of magnitude faster. Moreover, the new correction scheme is transferable also to FF methods that were shown to have serious problems with hydrogen-bonding interactions.

The PM6-DH+ method was parameterized to reproduce interaction energies for geometries obtained from high-level quantum mechanical calculations. See [accuracy](#).

The PM6-DH+ procedure corrects binding errors in the PM6 method. It can be used with geometry optimization or with a single point ([1SCE](#)) calculation. Normally, two or three calculations would be needed to get the binding energy.

Proposed strategy for calculating binding energy

- (A) Optimize the geometry of the two species (call these R_1 and R_2) separately using PM6-DH+.
- (B) Optimize the geometry of the adduct R_1 non-covalently bound to R_2 , using PM6-DH+.
- (C) Locate the heats of formation (DH_f) of each of the three systems, R_1 , R_2 , and R_1 bound to R_2 .

The binding or intermolecular interaction energy is then given by $DH_f(R_1 - R_2) - DH_f(R_1) - DH_f(R_2)$. Ignore the quantities: "DH Dispersion contribution" and "DH H-bond contribution," these are only of use if you want to know the component contributions.

Notes

Frozen geometries were used during the development of PM6-DH+. By contrast, in the proposed strategy, fully optimized geometries are used. No inconsistency is involved - by sketching a simple Born cycle, it becomes apparent that any errors arising from optimizing the PM6-DH+ parameters using frozen geometries and using those same parameters when calculating the binding energy using the above strategy would be very small; they would contribute only second order perturbation effects, and would be completely negligible.

Without the optional arguments, PM6-DH+ will apply to all atoms. With optional arguments, PM6-DH+ will apply only to the interaction of one part of the system with a second part. The optional argument defines atoms used in one of the parts. Atom numbers are specified by numbers separated by commas, and ranges of the type 645-670 or 645:670. To specify atoms 600, 610, 611, 612, and 630 the keyword would be PM6-DH+(600,610:612,630) or PM6-DH+(600,630,610-612)

PM6-DH2 [(n , m , k : l)]

Use the PM6-DH2 procedure of: M. Korth, M. Pitonák, J. Rezáč, and P. Hobza, "A *Transferable H-bonding Correction For Semiempirical Quantum-Chemical Methods*", J. Chem. Theory Comp. **2010**, 6, 344–352 and J. Rezáč, J. Fanfrlik, D. Salahub and P. Hobza, " *Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes* " J. Chem. Theory Comp. **2009**, 5, 1749-1760. See: [Abstract](#).

The PM6-DH2 method was parameterized to reproduce interaction energies for geometries obtained from high-level quantum mechanical calculations, see [accuracy](#). While it is possible to use it for geometry optimization, please be aware of the following limitation and that the method should be used with extra care. Users are recommended to optimize the geometry with the PM6 or PM6-D methods and calculate the final energy or interaction energy using PM6-DH2.

KNOWN LIMITATION OF PM6-DH2: *The gradients obtained by the current implementation of the -H correction do not include the term containing the derivative of atomic charges with change of the coordinates. The structure obtained by minimization using this gradient is not the exact minimum of PM6-DH2 energy. This error is negligible in the case of weaker H-bonds, but in case of a strongly bound formic acid dimer, the error is 0.30 kcal/mol. This error can be avoided by using [NOANCI](#), but then the calculations will then take much longer.*

The PM6-DH2 procedure corrects binding errors in the PM6 method. It can be used with geometry optimization or with a single point ([1SCF](#)) calculation. Normally, two or three calculations would be needed to get the binding energy.

Proposed strategy for calculating binding energy

- (A) Optimize the geometry of the two species (call these R_1 and R_2) separately using PM6, i.e., not using PM6-DH2.
- (B) Optimize the geometry of the adduct R_1 non-covalently bound to R_2 , using PM6.
- (C) Calculate the heats of formation (DH_f) of each of the three systems, R_1 , R_2 , and $R_1 - R_2$, using 1SCF and PM6-DH2.

The binding or intermolecular interaction energy is then given by $DH_f(R_1 - R_2) - DH_f(R_1) - DH_f(R_2)$. Ignore the quantities: "DH Dispersion contribution" and "DH H-bond contribution," these are only of use if you want to know the component contributions.

The rationale for this strategy is as follows: Geometry optimization using PM6-DH2 has small errors. These errors cannot easily be corrected (they can be corrected by using [NOANCI](#), but the jobs would then take a long time), however by using PM6 a well-defined stationary point can be easily obtained. By optimizing the various structures separately, the energy penalty that would result from the geometry change when the species form the noncovalent complex is avoided. The binding energy is thus the energy released when the two species form the complex in the gas phase; it is not the energy of interaction of R_1 with R_2 .

Notes

Frozen geometries were used during the development of PM6-DH2. By contrast, in the proposed

strategy, fully optimized geometries are used. No inconsistency is involved - by sketching a simple Born cycle, it becomes apparent that any errors arising from optimizing the PM6-DH2 parameters using frozen geometries and using those same parameters when calculating the binding energy using the above strategy would be very small; they would contribute only second order perturbation effects, and would be completely negligible.

Without the optional arguments, PM6-DH2 will apply to all atoms. With optional arguments, PM6-DH2 will apply only to the interaction of one part of the system with a second part. The optional argument defines atoms used in one of the parts. Atom numbers are specified by numbers separated by commas, and ranges of the type 645-670 or 645:670. To specify atoms 600, 610, 611, 612, and 630 the keyword would be PM6-DH2(600,610:612,630) or PM6-DH2(600,630,610-612)

If only the dispersion contribution is wanted, replace PM6-DH2 with PM6-D2. If only the hydrogen-bonding term is wanted, replace PM6-DH2 with PM6-H2.

Consider the dimer of formic acid: The binding energy would be the DH_f of the formic acid dimer minus the DH_f of two well-separated formic acid molecules. Here, two calculations are needed. For this work, the accurate geometries from the [S22 database](#) are used.

First calculation:

PM6-DH2(1:5) 1scf

C	-1.888896000	1	-0.179692000	1	0.000000000	1
O	-1.493280000	1	1.073689000	1	0.000000000	1
O	-1.170435000	1	-1.166590000	1	0.000000000	1
H	-2.979488000	1	-0.258829000	1	0.000000000	1
H	-0.498833000	1	1.107195000	1	0.000000000	1
C	1.888896000	1	0.179692000	1	0.000000000	1
O	1.493280000	1	-1.073689000	1	0.000000000	1
O	1.170435000	1	1.166590000	1	0.000000000	1
H	2.979488000	1	0.258829000	1	0.000000000	1
H	0.498833000	1	-1.107195000	1	0.000000000	1

Second calculation:

PM6-DH2(1:5) 1scf

C	-1.888896000	1	-0.179692000	1	99.000000000	1
O	-1.493280000	1	1.073689000	1	99.000000000	1
O	-1.170435000	1	-1.166590000	1	99.000000000	1
H	-2.979488000	1	-0.258829000	1	99.000000000	1
H	-0.498833000	1	1.107195000	1	99.000000000	1
C	1.888896000	1	0.179692000	1	0.000000000	1
O	1.493280000	1	-1.073689000	1	0.000000000	1
O	1.170435000	1	1.166590000	1	0.000000000	1
H	2.979488000	1	0.258829000	1	0.000000000	1
H	0.498833000	1	-1.107195000	1	0.000000000	1

Consider a substrate docked in an enzyme. The binding energy would be the DH_f of the docked assembly minus the DH_f of the isolated substrate minus the DH_f of the isolated enzyme. As *in vivo* enzymes exist in aqueous media, solvation will be necessary.

PMEP

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 `esplot` to use) *must* be present.

PMEPR

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 the Table. Other keywords, e.g., `MINMEP` (to print the minima) or `PRTMEP` (to write data for `esplot` to use) *must* be present.

Table: Data Required by PMEPR

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

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.

Examples of extra data:

To generate data for the plane through atoms 10, 11, and 20, use as extra data: "1 10 11 20 0" (do not include the quotation marks.)

POINT=*n*

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

POINT1=*n*

In a [GRID](#) calculation, the number of points to be calculated in the first direction is given by POINT1=*n*. '*n*' can be any number greater than 1. default: 11.

POINT2=*n*

In a [GRID](#) calculation, the number of points to be calculated in the second direction is given by POINT2=*n*. '*n*' can be any number greater than 1, default: 11;

POLAR

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 (a, b, and g) at user-defined frequencies.

See [units](#).

In 2004, the polarizability volume reported was modified by the use of [additive corrections](#).

Examples of POLAR keyword

To calculate the NLO quantities a, b and g 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 lscf 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.

POTWRT

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

Line 1 : Number of points

Line2 on: Potential (volts), x, y, z coordinate (in Angstroms) (4 numbers) Format: "1x,4e16.7"

POWSQ

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

PRECISE

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*](#) or [RELSCF=*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.

$$P=n \cdot nn$$

The effect of an applied pressure on a solid can be simulated by $P=n \cdot nn$, where $n \cdot nn$ is the pressure in Newtons per square meter, or $P=n \cdot nn \text{ GPa}$, where $n \cdot nn$ is the pressure in Gigapascals.

Typical bulk moduli are in the order of 10 - 700 GPa (gigapascals), and typical pressures should be $n \cdot nn = 1.D^9$ to $n \cdot nn = 1.D^{10} \text{ NM}^{-2}$. In general, solids are subjected to compression only, not tension, therefore the sign of $n \cdot nn$ should normally be positive, e.g. $P=2.D9$ or $P=2.0 \text{ GPa}$, however, for mechanically strong solids, negative pressures, even quite large pressures, can be used.

As the pressure rises, the calculated heat of formation will become more positive because energy must be used in generating the volume of the solid under pressure. Geometry optimization minimizes the total heat of formation = DH_f of the solid + energy due to volume. The final geometry is then the optimized geometry at equilibrium with that pressure.

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 $P=n \cdot nn$ ($\text{Vol}(n \cdot nn)$) from:

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

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 $n \cdot nn$ Newtons. Suitable values for $P=n \cdot nn$ are in the order of 10^{13} Newtons. In general, polymers are subjected to tension only, not compression, therefore the sign of $n \cdot nn$ should normally be negative, e.g. $P=-2.D13$.

Useful conversion factors are:

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

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

so a unit cell of 100 \AA^3 under 1 GPa would have an energy equivalent of $14.39 \text{ kcal.mol}^{-1}$.

PRNT=n

PRNT=n causes details of the working in EF to be printed. The amount of detail increases on going from PRNT=1 to PRNT=5.

PRTINT

By default, the interatomic distances are not printed. If you do want them to be printed, specify `PRTINT`.

PRTMEP

Print grid of electrostatic points calculated by P_{MEP}.

PULAY

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 envisaged, 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.

PULAY does *not* work with [MOZYME](#)

The combined set of convergers very seldom fails.

QPMEP

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

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

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

When only part of a geometry is being modified, the speed of a [MOZYME](#) SCF 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 ". `RAPID` should not be used if many atoms are to be optimized and those atoms are not all in one part of the molecule. For example, `RAPID` should not be used with [NOOPT](#) [OPT-H](#).

Because of limitations in `RAPID`, the termination criterion should be quite large - say [GNORM](#)=5 or more. Also, after a `RAPID` calculation, a conventional `MOZYME` calculation should be run, in order to get a more correct heat of formation.

If the following conditions are used, the optimization will run smoother:

- 1. Use Cartesian coordinates.
- 2. Before using `RAPID`, a useful step is to optimize the whole system.

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.

As an example of the effect of using `RAPID`, for a decapeptide containing 143 atoms, with the coordinates of the atoms of the last residue being optimized, the times required for 10 cycles of geometry optimization are as follows:

Method	Time (s)
Using conventional MOPAC	19.4
Using MOZYME	5.5
Using MOZYME with RAPID	2.8

RECALC= n

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.

RE-LOCAL RE-LOCAL=*n*

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.

In rare instances, the LMOs might become too delocalized. They can be re-localized on each *n* SCF calculations by using RE-LOCAL=*n* .

RELSCF

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](#).

REORTH

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.

RESEQ

According to the PDB format, atoms in a protein are specified starting with the nitrogen of the -NH_2 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.

Care should be exercised in using RESEQ in that there is the possibility of atoms being re-arranged in unexpected ways. Consider phenylalanine, for example. In it, the sequence of atoms is uniquely defined except for C_{d1} and C_{d2} . This means that either one of these atoms might be chosen at random to be C_{d1} , and the choice might be different from that used in the starting data set. The effect of this is that any attempt to calculate the RMS difference between two structures, one of which has been re-sequenced, is likely to give a nonsense result, even if hydrogen atoms are ignored in the RMS calculation.

Also, if the protein has gaps where residues are missing, RESEQ might put the fragments into the wrong order, particularly if the residues in the protein have been re-arranged as the result of earlier operations. To ensure that the order of fragments is correct, use [CVB](#) to make dummy bonds that bridge the gaps. Suitable atoms are the N terminus of one residue and the carboxyl of the other residue.

RESIDUES

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 '(*nnnnn* *TYP**frrr*)'

where *nnnnn* 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 the Table). 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; *r**r**r* is the residue number. A modified residue can still be recognized if *XENO=**text* is used. The residue nearest to the NH₂ end of the protein is No. 1, the next is 2, and so on.

When RESIDUES is present, the net charge on each residue will be printed in the normal output; this is useful in identifying ionized residues. If gradients are printed, then gradient norms for each residue are also printed.

Table: 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 O H 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.

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. If [RESEQ](#) is used, then RESIDUES is automatically run.

RESTART

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.

The restart file contains all the geometry data: current geometry, cycle number, gradient and Hessian information, if calculated, etc. It does not contain information on the atoms, so the restart data set has to have a normal geometry just like a non-restart data set. This is used only for supplying the atom-types and number of atoms..

Example: If the following data set is run, it will stop after three geometry optimization cycles, and output a message indicating that restart files have been created.

```
cycles=3
Ethanol
C2H5OH
O
C 1.5 1 0 0 0 0 1
C 1.5 1 120 1 0 0 2 1
H 1.0 1 120 1 90 1 1 2 3
H 1.0 1 109 1 120 1 2 1 3
H 1.0 1 109 1 240 1 2 1 3
H 1.0 1 109 1 0 1 3 2 1
H 1.0 1 109 1 120 1 3 2 1
H 1.0 1 109 1 240 1 3 2 1
```

By adding the word "RESTART" to the original data set, and re-submitting the job, the calculation can be continued. In this example, "cycles=3" has been deleted; this allows the job to run to completion.

```
restart
Ethanol
C2H5OH
O
C 1.5 1 0 0 0 0 1
C 1.5 1 120 1 0 0 2 1
H 1.0 1 120 1 90 1 1 2 3
H 1.0 1 109 1 120 1 2 1 3
H 1.0 1 109 1 240 1 2 1 3
H 1.0 1 109 1 0 1 3 2 1
H 1.0 1 109 1 120 1 3 2 1
H 1.0 1 109 1 240 1 3 2 1
```

It is not necessary to change the geometric data to reflect the new geometry, although this can be done, if desired. When `RESTART` is used, the `<file>.res` file contains all the geometric data, and this will immediately over-write the geometry supplied by the data set. Note, however, that the data set must have a geometry because the `<file>.res` file does not contain the atom types. 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.

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.

See also [OLDENS](#).

RHF

The Restricted Hartree-Fock Hamiltonian is to be used.

This is the default for even-electron systems and for odd-electron systems that use keywords such as [MECI](#) that imply an RHF calculation. For odd electron (radical) systems, the [UHF](#) method is used by default, because geometry optimization using UHF runs faster than when `RHF` is used.

If RHF methods are to be used on odd electron systems, add keyword `RHF`.

RHF methods are needed if configuration interaction is used, see [MECI](#).

RM1

The [RM1](#) method(1) is to be used.

(1) Rocha, G.B., R.O. Freire, A.M. Simas, and J.J.P. Stewart., *RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. J. Comp. Chem.*, **27**(10): 1101-1111 (2006).

$$\mathbf{RMAX} = \mathbf{n . nn}$$

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*

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

Two forms are provided: `ROOT=n`, where n is an integer, and `ROOT=n<text>`, where n is an integer and `<text>` is a text string.

`ROOT=n`

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). If the state selected is degenerate, all components of that state will be selected.

`ROOT=n<text>`

The n 'th root of a C.I. calculation that has the symmetry `<text>` 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 that has the symmetry `<text>` will be selected. Thus, in an octahedral system, `ROOT=3T2g` and `SINGLET` will select the third singlet T2g root. If `ROOT=3T2g` is used without any spin-state being specified, then the third T2g root will be used, regardless of spin. If the state selected is degenerate, all components of that state will be selected.

See also [C.I.=n](#).

To see how these different forms behave, consider the following states of a d^6 transition metal complex:

STATE	Q.N.	Spin	Symmetry
1	1	TRIPLET	T2g
4	1	SINGLET	A1g
5	1	SINGLET	T2g
8	1	TRIPLET	T1g
11	1	QUINTET	T1g
14	1	SEPTET	A1g
15	1	QUINTET	T2g
18	2	TRIPLET	T1g
21	3	TRIPLET	T1g
24	2	SINGLET	T2g
27	2	TRIPLET	T2g
30	1	TRIPLET	A2g
31	1	SINGLET	T1g
34	3	TRIPLET	T2g
37	4	TRIPLET	T1g
40	1	TRIPLET	Eg
42	2	SINGLET	T1g
45	1	SINGLET	Eg
47	1	QUINTET	Eg

`ROOT=14` would select the 14th state, the 1^7A_{1g} state. No spin being specified, `ROOT` applies to the STATE column.

`ROOT=7` and `QUINTET` would select the 47th and 48th states, the 1^5E_g state. This is the 7th quintet state, the states 1-6 being T1g, T1g, T1g, T2g, T2g, and T2g. That these states are degenerate is not important, because `ROOT=n` specifies the n 'th state, without regard to symmetry.

`ROOT=2T2g` and `TRIPLET` would select the 27th, 28th, and 29th states, the 2^3T_{2g} state. This is the preferred method of specifying states. When `ROOT=n<text>` is used then the state specified will not change if the state moves up or down the list of states. If the system has no symmetry, `ROOT=nA` can be used.

When a geometry is to be optimized, symmetry should be used, if present. This is particularly important in octahedral transition metal complexes. If the state has orbital degeneracy, e.g. if it is of type E, T, G, or H, then Jahn-Teller effects might cause a loss of symmetry. High symmetry is automatically detected, so, if present, it will be conserved. However, during a normal unconstrained geometry optimization, minor excursions from high symmetry are allowed, and these might confuse the high-symmetry detector. To prevent this, use symmetry. In the case of a simple octahedral complex, MX_6 , the data set might look like this:

```
ROOT=1T2g QUINTET OPEN(5,5) MECI SYMMETRY
Generic octahedral complex
```

```
M 0.0 0 0 0 0 0 0 0 0
X 2.0 1 0 0 0 0 1 0 0
X 2.0 0 90 0 0 0 1 2 0
X 2.0 0 90 0 90 0 1 2 3
X 2.0 0 90 0 180 0 1 2 3
X 2.0 0 90 0 -90 0 1 2 3
X 2.0 0 180 0 0 0 1 2 3
```

```
2 1 3 4 5 6 7
```

RSCAL

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*

Used by the [COSMO](#) method to set the effective radius of the solvent molecule to *n.nn*. The default value (1.3) 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.

Dr Klamt suggests: For COSMO-RS, use a solvent radius which is independent of the solvent, e.g. `R_solv` = `R_hydrogen` = 1.3 Angstrom. Do not use a large solvent radius for large solvents, because the solvent is not probing the solute globally but locally. It is the local curvature which is relevant here, and a general value of 1.3 Angstrom is normally reasonable.

SADDLE

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 if [SYMMETRY](#) is present, it will only be used in setting up the starting geometry, after that it will be ignored. See also [BAR=n.nn](#).

Example

```
*
*
* Locate the approximate transition state for the reaction
*
* 2CH4 -> H2 + C2H6
*
* For this calculation to work, the C2H6 system had to be put into the
* D2h geometry.
*
T=4H bar=0.04 geo-ok SADDLE mndo

C 0.000000 0 0.000000 0 0.000000 0 0 0 0 0.0704
C 2.521764 1 0.000000 0 0.000000 0 1 0 0 0.0704
H 1.104340 1 144.595726 1 0.000000 0 1 2 0 -0.0178
H 1.104167 1 90.121334 1 -125.283960 1 1 2 3 -0.0175
H 1.104166 1 90.116566 1 125.235484 1 1 2 3 -0.0176
H 1.103818 1 76.622615 1 0.975117 1 2 1 3 -0.0173
H 1.104407 1 122.428917 1 120.148896 1 2 1 3 -0.0177
H 1.104394 1 122.390158 1 240.200743 1 2 1 3 -0.0178
H 1.103576 1 35.239589 1 -0.019577 1 1 2 3 -0.0175
H 1.103620 1 33.058707 1 0.000000 0 2 1 3 -0.0176

C 0.000000 0 0.000000 0 0.000000 0 0 0 0 0.0164
C 1.521764 1 0.000000 0 0.000000 0 1 0 0 0.0164
H 1.208966 1 111.093608 1 0.000000 0 1 2 0 -0.0055
H 1.108971 1 111.085646 1 -120.010199 1 1 2 3 -0.0055
H 1.108949 1 111.082331 1 119.998535 1 1 2 3 -0.0055
H 1.108982 1 111.087018 1 .998485 1 2 1 3 -0.0055
H 1.109025 1 111.084646 1 120.004584 1 2 1 3 -0.0055
H 1.109003 1 111.081348 1 240.005897 1 2 1 3 -0.0055
H 2.0 1 90 1 0 1 1 2 3
H 1.0 1 90 1 70 1 9 1 2

*
* The resulting gradient is large, because the geometry is near but not at the transition
* state.
* To refine the transition state, a TS calculation must be run.
```

When things go wrong

If, in a SADDLE calculation, you get the message,

```
" BOTH SYSTEMS ARE ON THE SAME SIDE OF THE TRANSITION STATE -
GEOMETRIES OF THE SYSTEMS ON EACH SIDE OF THE T.S. ARE AS FOLLOWS"
```

There are two ways to proceed.

(A) If the gradient is small (less than about 10) at the point where this message is printed, then find the geometry that has the smallest gradient near to the end of the run, and use that as the starting geometry for a transition state refinement. If the geometry refines correctly ([PRECISE](#) might be needed in order to get rid of methyl rotations, etc.), and a FORCE calculation verifies that it is a transition state, then the calculation has been successful.

(B) If the gradient is large, the examine the output to find two adjacent points that are on opposite sides of the transition state. Geometries on one side (call it the left side) will be indicated by the message "FOR POINT nn FIRST STRUCTURE", points on the other side will be indicated by "FOR POINT nn SECOND STRUCTURE"

Check that the direction cosine for both is greater than 0.9.

Edit the output file to delete everything except the two geometries.

Reduce the value of the BAR by a factor of about 10. If BAR was not used previously, then start with BAR=0.05.

SADDLE

Run the new SADDLE calculation

In general, in a SADDLE calculation, avoid using extra keywords such as PRECISE.

SCALE

`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 . nn*

The default SCF criterion, 0.0001 kcal/mol, 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. See also [RELSCF](#)

SCINCR=*n . nn*

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

SEPTET

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.

SETPI

In some instances, the default [Lewis structure](#) used in a [MOZYME](#) calculation is not correct. The correct structure can be selected by explicitly supplying one or more p bonds. To do this, add keyword `SETPI`. After the data set, add the atom-pairs that define one or more p bonds. For example, a hexagon of carbon atoms with four hydrogen atoms and two oxygen atoms in the 1 and 4 positions, i.e. $C_6H_4O_2$, can either be the neutral molecule *para*-benzoquinone or the di-anion of hydroquinone. The default Lewis structure is the neutral system. If the di-anion is wanted, then the three aromatic p bonds need to be explicitly defined, thus:

```

Mozyne setpi charge=-2
Di-anion of hydroquinone

C      0.00000000  0      0.00000000  0      0.00000000  0
C      1.42891759  1      0.00000000  0      0.00000000  0      1      0      0
C      1.39123760  1      121.7152363  1      0.00000000  0      2      1      0
C      1.42884933  1      121.6862338  1      0.0522686  1      3      2      1
C      1.42731168  1      116.6123504  1      -0.0326433  1      4      3      2
C      1.39298952  1      121.7002715  1      0.0291866  1      5      4      3
O      1.28854675  1      121.6283321  1      179.9905509  1      1      2      3
O      1.28863750  1      121.5839611  1      179.9914815  1      4      3      2
H      1.07825816  1      117.6731944  1      -179.9642741  1      2      1      3
H      1.07822008  1      120.6215632  1      -179.9762574  1      3      2      1
H      1.07805743  1      117.7610380  1      -179.9848142  1      5      4      3
H      1.07809519  1      120.5409971  1      179.9757231  1      6      5      4

2 3
1 6
4 5

```

There is no need to supply all the p bonds, only those necessary to resolve any ambiguities. Other examples of the need for `SETPI` include:

Distinguishing the anion of p-hydroxy N-methyl pyridine from its quinone type tautomer.
 Ensuring that *meso*-tetra(*para*-N-methylpyridinato)-porphyrin has the correct charge.

See also [CVB](#) to explicitly make or break bonds, and [atom labels](#) to explicitly assign charges.

SETUP

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.

SEXTET

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.\(n1-n2\)](#) 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*

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 in an attempt to optimize convergence. This can result in a positive or zero shift of the virtual M.O. energy levels.

The SHUT command

The SHUT command is not a keyword, instead it is a small command script that can be used to send a message to a running MOPAC job to instruct it to shut down in a tidy manner, and to generate restart and density files. When MOPAC reads the message, it increases the apparent CPU time by 10,000,000 seconds (over 100 days). This exceeds any reasonable job time, so MOPAC behaves as if it had run out of time.

How SHUT works

When a MOPAC job is running, it periodically checks a file called <name>.end. At the start of the run, <name>.end is empty, that is, it is a file with nothing in it. While <name>.end has nothing in it, the job will continue normally. If there is any text at all in <name>.end the next time MOPAC checks the file it will increase the apparent CPU time by 10^8 seconds. SHUT simply puts some text into <name>.end. The two scripts given below.

For WINDOWS: Copy the script below and put it into a file called shut.cmd. Put shut.cmd into a folder in the PATH. shut.cmd is used at the command prompt in the folder where the job is running. To use it to shut down a job called <file>.mop, issue the command "shut <file>" , e.g.:

```
M:\> shut crambin
```

would issue the SHUTDOWN command to the running job crambin.

Start of WINDOWS shut.cmd script:

```
echo off
if exist %1.dat goto dat
if exist %1.mop goto mop
if exist %1.arc goto arc
echo The file %1.mop, %1.dat, or %1.arc does not exist in this folder
goto end
:dat
copy %1.dat %1.end
goto end1
:arc
copy %1.arc %1.end
goto end1
:mop
copy %1.mop %1.end
:end1
echo shutdown command issued to %1
:end
```

For Mac and Linux: Copy the script below and put it into a file called shut.csh. Change permissions on shut.csh to make it executable (chmod +x shut.csh) Put shut.csh into a directory in the PATH, or alias it in the .bashrc or .cshrc scripts. shut.csh is used at the command prompt in the directory where the job is running. To use it to shut down a job called <file>.mop, issue the command "shut <file>" , e.g.:

```
~ > shut crambin
```

or

```
~ > shut crambin.mop
```

would issue the SHUTDOWN command to the running job crambin.

Start of Linux and Mac file shut.csh:

```
#!/bin/sh
#
# SHUTDOWN command
# Remove last four characters of file.
#
file=$1
if [ `expr "$file" : ".*.mop"` -gt 0 ]; then file=${file:0:`expr $file : '.*'` -4}; fi
if [ `expr "$file" : ".*.out"` -gt 0 ]; then file=${file:0:`expr $file : '.*'` -4}; fi
if [ `expr "$file" : ".*.dat"` -gt 0 ]; then file=${file:0:`expr $file : '.*'` -4}; fi
if [ `expr "$file" : ".*.arc"` -gt 0 ]; then file=${file:0:`expr $file : '.*'` -4}; fi
echo Shutdown > $file.end
```

The second form, ~ > shut crambin.mop, is useful if the name is long and has been supplied to shut using "copy and paste."

SIGMA

When refining transition states, and [TS](#) is *not* wanted, the McIver-Komornicki gradient norm minimization [[9,10](#)] routines, POWSQ and SEARCH, can 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 where TS does not work, [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

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 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](#).

If the [UHF](#) method is used, then if `SINGLET` is present, the system will be checked to verify that the number of electrons is even. If it is, then `SINGLET` will be ignored, otherwise the calculation will be stopped, and an error message printed.

SITE=(text)

Quite often, after hydrogen atoms have been added to a protein, the ionization state of various sites is not what is wanted. The ionization state of various sites in a protein can be explicitly defined by `SITE=(text)`. The *text* in this case can be one or more of the entries in the table below. Entries should be separated by a comma or by a semicolon.

To neutralize all sites in a protein, use `SITE=(COOH,NH2,ARG,HIS)`. To ionize all sites, use `SITE=(COO,NH3,ARG(+),His(+))`. No specific tautomer is defined - if it's not the one wanted, make the change outside MOPAC. Because atoms are added or removed, the job cannot continue. Edit the resulting `<file>.out`, `<file>.arc`, or `<file>.pdb` to make a new data set..

Text	Effect
COOH	Add a hydrogen atom to -COO groups
COO	Remove the hydrogen atom from -COOH groups
NH3	Add a hydrogen atom to -NH ₂ groups
NH2	Remove a hydrogen atom from -NH ₃ groups
Arg(+)	Add a hydrogen atom to -Arg-
Arg	Remove a hydrogen atom from -Arg(+)-
His(+)	Add a hydrogen atom to -His-
His	Remove a hydrogen atom from -His(+)-

SLOG=*n . nn*

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

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

SNAP

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 $\arccos(-1/3)$ accurate to 15 or 16 figures. The tolerance of SNAP is quite strict, so 109.47 would NOT be recognized.

SPARKLE

The lanthanides can be represented by their fully ionized 3+ sparkles. That is, they have no basis set, and therefore cannot have a charge different from +3.000. To use these sparkles, add keyword `SPARKLE`.

The geometries of the lanthanides are reproduced with good accuracy, but the heats of formation and electronic properties are not accurate.

To specify AM1 Dysprosium 3+ ion, use keywords [AM1](#) and `SPARKLE`, and specify Dy at the appropriate point in the Z-matrix.

SPIN

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.

START_RES(*text*)

Some PDB files represent two or more proteins, for example there are four proteins in one hemoglobin. At the end of each such protein there is a TER entry, indicating the termination of the protein. In other proteins, some residues are missing. Missing residues are indicated in the PDB file by a discontinuity in the residue number. When writing PDB formats using [PDBOUT](#), information on the various proteins and on missing residues can be indicated by the `START_RES(text)` keyword. The format of the keyword is as follows:

```
START_RES [=](n1[ , - ][n2[ , - ][n3[ , - ]... ])
```

n_n is the number of the first residue in a section of contiguous residues in the PDB file. If two contiguous sections are separated by missing residues, then the two n_n are separated by a minus sign. If the two contiguous sections are on different proteins, then the two n_n are separated by a single space.

Thus if a PDB file consists of two proteins, A and B, and A has residues 30 31 32 33 34 38 39 40, and B has residues 44 45 46 49 50 51 55 56 57, then the keyword would be:

```
START_RES(30-38 44-49-55) or START_RES=(30-38 44-49-55)
```

In the PDB file written by MOPAC, a "TER" line would be added after the last atom of protein A, residue 8 (residue 38 in the output), and before the first atom of protein B, residue 9 (residue 44 in the output) .

If the PDB file consists of one protein only, and there is a gap in the residue sequence, for example suppose the residue numbers were: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, then the `START_RES` keyword would have two entries, e.g., `START_RES(1-9)`, the first entry being the residue number of the first contiguous section.

See also [CHAINS\(*text*\)](#) to define chain-letters.

STATIC

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 DH_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`).

In 2004, the polarizability volume reported was modified by the use of [additive corrections](#).

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

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*

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*

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

STO3G

In an ESP calculation `STO3G` means "Use the STO-3G basis set to de-orthogonalize the semiempirical orbitals".

SUPER

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 Schuurmann Molecular Orbital shift alpha quantities.

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

In the MOPAC output, the various quantities are defined as follows:

Mulliken electronegativity = $-a$, where a = the average of the HOMO and LUMO energies, i.e., $a = \frac{1}{2}(e_{\text{homo}} + e_{\text{lumo}})$

Parr & Pople absolute hardness = $\frac{1}{2}(e_{\text{homo}} - e_{\text{lumo}})$

Schuurmann MO shift alpha = a .

$E_{\text{homo}} = e_{\text{homo}}$ = energy of the Highest Occupied Molecular Orbital, in eV

$E_{\text{lumo}} = e_{\text{lumo}}$ = energy of the Lowest Unoccupied Molecular Orbital, in eV

$D^N(r)$ and $D^E(r)$ printed in the MOPAC output are described in (B) as follows: The nucleophilic and electrophilic delocalizabilities $D^N(r)$ and $D^E(r)$ of a reactant's centre r according to Fukui et al. [1] and the respective total sums D^N and D^E can be defined within all-valence electron schemes [2] according to

$$D^N(r) = 2 \sum_k^{\text{vac}} \sum_{\sigma(r)} \frac{c_{\sigma k}^2}{\alpha - \epsilon_k} \quad D^N = \sum_r D^N(r)$$

$$D^E(r) = 2 \sum_i^{\text{occ}} \sum_{\sigma(r)} \frac{c_{\sigma i}^2}{\epsilon_i - \alpha} \quad D^E = \sum_r D^E(r)$$

$q(r) - Z(r)$ is the charge density on each atom, i.e., the negative of the number of electrons on each atom. In formaldehyde, the oxygen has a partial charge, $q(r)$, that is slightly negative, say -0.4. The core charge, $Z(r)$, is 6, so the charge density is -6.4. That is, there are 6.4 valence electrons on the oxygen atom.

$\pi^S(r)$ is the self-polarizability $p^S(r)$ of an atom r . It was introduced as a reactivity measure for p electron systems by Coulson and Longuet-Higgins [3]; the all-valence electron formula can be defined according to:

$$\pi^S(r) = -4 \sum_i^{\text{occ}} \sum_k^{\text{vac}} \sum_{\sigma(r)} \frac{c_{\sigma i}^2 c_{\sigma k}^2}{\epsilon_k - \epsilon_i} \quad \pi^S = \sum_r \pi^S(r)$$

The dimension of $p^S(r)$ is charge²/energy, which can be transformed to a volume dimension by multiplying with (length)²; for convenience the Bohr radius was chosen as the conversion factor in order to arrive at the same order of magnitude for $p^S(r)$ and p^S for a_{vol} .

The quantities below the line "a n homo-1 homo lumo lumo+1" are the percentage (%) contribution of the electron density on the various atoms to the four molecular orbitals.

[1] Fukui, K., Kato, H. and Yonezawa, T., *Buff. Chem. SOC. Jup.* 27, 423 -427 (1961).

[2] Schuurmann, G., *Environ. Toxicof. Chem.*, 9,417 -428 (1990).

- [3] Coulson, C.A. and Longuet-Higgins, H.C., *Proc. Roy. SOC. (London) A* 192, 16-32 (1947).

SYBYL

Output for use by Tripos's SYBYL program can be generated by use of SYBYL. This requires a minimum of at least one SCF calculation to be done, in order to calculate the heat of formation, partial charges, HOMO and LUMO energies. A normal output is also required, so SYBYL should *not* be used with keywords that prevent normal output, i.e., it should not be used with [FORCE](#), [STEP](#), [DRC](#), etc.

SYMAVG

Used by the [PMEP](#), SYMAVG will average charges that should have the same value by symmetry. Thus, to generate symmetry-equivalent charges, use [OPMEP](#) and SYMAVG.

SYMOIR

When irreducible representations are being assigned to eigenvectors (M.O.s, state functions, and normal modes), print the characters of the eigenvectors for the symmetry operations of the point group, and print the number of irreducible representations of each class. Useful in diagnostics and in teaching.

SYMTRZ

Print details of working in subroutine SYMTRZ.

Each operation is represented by a 3 by 3 Eulerian matrix. The layout is *not* very user friendly!

SYMMETRY

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 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.

SYMMETRY does not involve any point-group symmetry or symmetry theory. When SYMMETRY is used, the symmetry relations between coordinates are defined by the user. These can be used for defining the point-group of a molecule, for example the D_{6h} of benzene, but no point-group theory is used when SYMMETRY is specified.

By default, symmetry point-groups are automatically assigned, and used in identifying the irreducible representations of molecular orbitals, vibrations, and states, and in accelerating the construction of the Hessian in [FORCE](#) calculations. To disable this function, use [NOSYM](#).

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 below.

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 1. 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 below) 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.

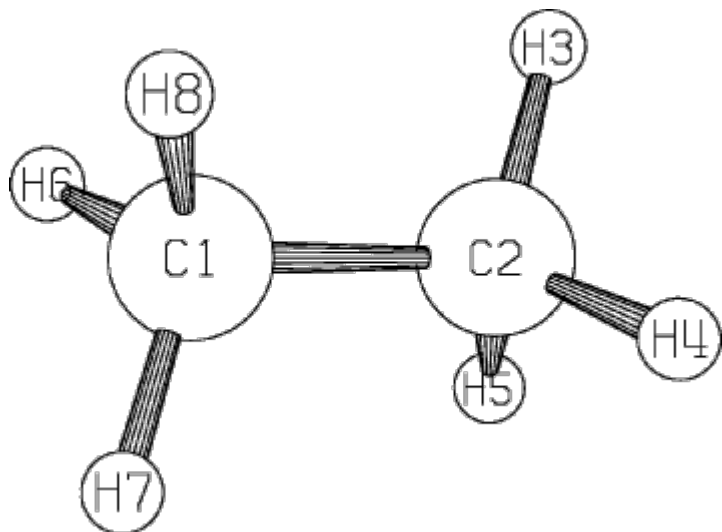
Methane, specified using Cartesian coordinates, can be described with one unknown, the C-H bond length, as shown in Figure 2.

Table 1: 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° - reference dihedral

- 5 Dihedral angle varies as $90^\circ +$ reference dihedral
- 6 Dihedral angle varies as $120^\circ -$ reference dihedral
- 7 Dihedral angle varies as $120^\circ +$ reference dihedral
- 8 Dihedral angle varies as $180^\circ -$ reference dihedral
- 9 Dihedral angle varies as $180^\circ +$ reference dihedral
- 10 Dihedral angle varies as $240^\circ -$ reference dihedral
- 11 Dihedral angle varies as $240^\circ +$ reference dihedral
- 12 Dihedral angle varies as $270^\circ -$ reference dihedral
- 13 Dihedral angle varies as $270^\circ +$ reference dihedral
- 14 Dihedral angle varies as the negative of the reference dihedral
- 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 $180^\circ -$ reference bond angle
- 18 (not used)
- 19 Bond length is a multiple of the reference bond length

Figure 1: Ethane, showing use of SYMMETRY



SYMMETRY						NA NB NC		
ETHANE, D3D								
C	1.53	1				1		
H	1.10	1	110	1		2	1	
H	1.10	0	110	0	120	0	2	1
H	1.10	0	110	0	240	0	2	1
H	1.10	0	110	0	60	0	1	2
H	1.10	0	110	0	180	0	1	2
H	1.10	0	110	0	300	0	1	2
0	0.00	0	0	0	0	0	0	0
3,	1,	4,	5,	6,	7,	8,		
3,	2,	4,	5,	6,	7,	8,		

Table 2: 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

Figure 2: Example of Cartesian SYMMETRY functions

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	

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 two days, i.e., 172,800 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 two days, 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, T=*nn.nn*D, in days, and T=*nn.nn*W, in weeks, for very long jobs.

A job that's running can be instructed to run out of time by issuing the [SHUT](#) command. If that is done, then, at the first opportunity, MOPAC will write out a RESTART file called <file>.res and a DENSITY file, <file>.den. These can then be read in to a subsequent job using [RESTART](#) and [OLDENS](#).

THERMO

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 the [Note on Thermochemistry](#) 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\(nnn\)](#), [THERMO\(nnn,mmm\)](#), and [THERMO\(nnn,mmm,lll\)](#). See also [TRANS](#)

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

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*)

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* , *III*) OR THERMO= (*nnn* , *mmm* , *III*)

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

TIMES

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.

T-PRIORITY

In a DRC calculation, results will be printed whenever the calculated time changes by 0.1 femtoseconds. Abbreviation, T-PRIO. See [Dynamic and Intrinsic Reaction Coordinates](#) for more details.

T - PRIORITY = *n . nn*

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

TRANS

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*

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

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:

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

is calculated. See keywords `C.I.=n` and `OPEN(n1,n2)`.

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

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

The unrestricted Hartree-Fock Hamiltonian is to be used.

This is the default for odd-electron (radical) systems because geometry optimization using UHF runs faster than when RHF is used. This is because the open-shell RHF gradient calculation is relatively slow.

If UHF methods are to be used on even electron systems, add keyword UHF.

A limitation of the UHF method is that the resulting wavefunction is not spin quantized, that is, although $\langle S_z \rangle$ has a well-defined value, e.g. 0, 1/2, 1, 3/2, etc., $\langle S^2 \rangle$ is normally not quantized. For a system with an $\langle S_z \rangle$ of 1/2, the $\langle S^2 \rangle$ should be $3/4 = 1/2 * (1/2 + 1)$, but is usually higher. This is a consequence of spin contamination. Spin contamination is not present if UHF is not used.

VDW(*text*)

In a COSMO calculation (keyword [ESP=78.4](#)), 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).

By default, the COSMO VdW radius is 117% of the default VdW radius.

Where radii specific for various methods are available, they are used. Radii available are:

Table 1: Default radii (Angstroms) used in COSMO

I	II		III	IV	V	VI	VII
H 1.08							
Li 1.80	Be		B	C 1.53	N 1.48	O 1.36	F 1.30
Na 2.30	Mg		Al 2.05	Si 2.10	P 1.75	S 1.70	Cl 1.65
K 2.30	Ca 2.75	Transition Metals (none)	Ga	Ge	As	Se	Br 1.80
Rb	Sr		In	Sn	Sb	Te	I 2.05
Cs	Ba		Tl	Pb	Bi		

VDWM(*text*)

In [MOZYME](#), the [LEWIS](#) structure is constructed from the topology of the system. Atoms are considered to be connected if they are within 10% of the addition of the two radii, or within twice the heavy-atom radius if one of the atoms is hydrogen. Values for all radii used are shown in the Table below.

The criterion of whether atoms are connected or not can be changed by modifying the radii. using the `VDWM` keyword. The format of the command is:

```
VDWM(<chemical symbol1>=n.nn;<chemical symbol2>=n.nn;<chemical symbol3>=n.nn;...)
```

For example, if a hydrogen atom should be regarded as connected to a nitrogen, but the N-H distance is 1.62 Angstroms, then `VDWM(N=0.85)` would override the default values of the van der Waals radii of nitrogen (0.75), and result in the hydrogen being connected (the N-H distance of 1.62 being less than 2*0.85).

If a fluorine atom should be connected to a bromine, but the F-Br distance is 2.1 Angstroms, i.e., it is too large for the two atoms to be regarded as being connected ($2.1 > 1.1(0.71 + 1.14)$), then either the F radius, or the Br radius, or both, could be increased. One possible change would be to use keyword `VDWM(F=0.75;Br=1.2)`

Entries for different elements should be separated by a colon, a comma, or a semi-colon.

Table 1: Default radii (Angstroms) used in MOZYME

I	II	Transition Metals										III	IV	V	VI	VII	VIII
H 0.37																	He 0.32
Li 1.34	Be 0.90											B 0.82	C 0.77	N 0.75	O 0.73	F 0.71	Ne 0.69
Na 1.54	Mg 1.30											Al 1.18	Si 1.11	P 1.06	S 1.02	Cl 0.99	Ar 0.97
K 1.96	Ca 1.74	Sc 1.44	Ti 1.36	V 1.25	Cr 1.27	Mn 1.39	Fe 1.25	Co 1.26	Ni 1.21	Cu 1.38	Zn 1.31	Ga 1.26	Ge 1.22	As 1.19	Se 1.16	Br 1.14	Kr 1.10
Rb 2.11	Sr 1.92	Y 1.62	Zr 1.48	Nb 1.37	Mo 1.45	Tc 1.56	Ru 1.26	Rh 1.35	Pd 1.31	Ag 1.53	Cd 1.48	In 1.44	Sn 1.41	Sb 1.38	Te 1.35	I 1.33	Xe 1.30
Cs 2.25	Ba 1.98	La 1.69	Hf 1.50	Ta 1.38	W 1.46	Re 1.59	Os 1.28	Ir 1.37	Pt 1.28	Au 1.44	Hg 1.49	Tl 1.48	Pb 1.47	Bi 1.46			

Related key-words: [LEWIS](#), [CVB](#), [SETPL](#), [METAL](#), [CHARGE](#), [CHARGES](#), and [MOZYME](#)
See also: [Lewis Structures](#), [MOZYME introduction](#)

VECTORS

The eigenvalues, in eV, and 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.

If [MOZYME](#) is used, then the localized molecular orbitals will be printed. If the eigenvectors are wanted, then [EIGEN](#) must also be present.

VELOCITY

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

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.

X-PRIORITY=*n . nn*

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

XENO

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. If an unknown residue is detected, then a message of the type: "Unknown residue: 19 atom 296 C:40 N: 5 O: 6 S: 2 see keyword 'XENO'" will be printed.

`XENO`, from the Greek *xenos*, *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=( nC, nN, nO, nS, name [ ; nC, nN, nO, nS, name ] [ ; nC, nN, nO, nS, name ] . . . ) e.g.  
XENO=( 38, 3, 4, 2, HEM; 4, 3, 2, 0, RES )
```

Up to ten 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. If *name* has exactly three letters, then the residue name will be replaced by *name*.

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



Not counting hydrogens, the empirical formula for a lysine fragment is C_6N_2O (See below and [Table of Residues](#)). For residue 216, the empirical formula is $C_{26}N_2O$, so the extra fragment, the Schiff base, accounts for C_{20} . 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)`. In the output or PDB file generated, residue 216 would be identified as LYS. If `XENO=(20, 0, 0, 0, RET)` were used, residue 216 would be identified as RET.

Quite often the side-chain cannot be related to any of the 20 standard amino acids. In that case, the simplest option is to define the un-modified residue as being glycine (GLY). If that is done, then the `XENO` keyword could be constructed by subtracting two carbon atoms, one nitrogen and one oxygen atom. Thus if the message was "Unknown residue: 19 atom 296 C:40 N: 5 O: 6 S: 2 see keyword 'XENO'", the keyword would be `XENO=(38, 4, 5, 2, HEME)`. If the side-chain can be related to a standard amino acid, then the appropriate number of atoms should be deleted. The following table gives the patterns for the various standard amino acids.

Number of atoms in each residue for use in working out the XENO keyword

Amino acid	Number C	Number N	Number O	Number S	Amino acid	Number C	Number N	Number O	Number S
Glycine	2	1	1	0	Glutamic acid	5	1	3	0
Alanine	3	1	1	0	Glutamine	5	2	2	0
Valine	5	1	1	0	Arginine	6	4	1	0
Leucine	6	1	1	0	Histidine	6	3	1	0
Isoleucine	6	1	1	0	Phenylalanine	9	1	1	0
Serine	3	1	2	0	Cysteine	3	1	1	1
Threonine	4	1	2	0	Tryptophan	11	2	1	0
Aspartic acid	4	1	3	0	Tyrosine	9	1	1	0
Asparagine	4	2	1	0	Methionine	5	1	1	1
Lysine	6	2	1	0	Proline	5	1	1	0

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. Residues with other atoms will not be recognized.

If atoms are missing from a particular residue, the "missing" atoms can be defined using negative numbers. Thus, if a residue should be lysine, but the terminal amino group is missing, then it can be replaced using `XENO(0,-1,0,0,Lysine)`. However, if, as a result of missing atoms, the residue corresponds to a known residue, then the residue will be incorrectly recognized as a known residue. For example, if the residue should be lysine, but the

group $-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}_2$, is missing, so that the residue is $-\text{NH}-\text{CH}(\text{CH}_3)-\text{CO}$, the keyword XENO(-3,-1,0,0,Lysine) will not replace it, because the damaged residue corresponds to alanine, and would be recognized as such. This error is uncorrectable!

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

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 system involving several rings, or big rings, is optimized using internal coordinates, the geometry optimizers often have difficulty. This is because the effect of a small change in an angle can be a large change in the interatomic distance of two bonded atoms. This is particularly important in enzymes and other large molecules. This problem is completely solved if Cartesian coordinates are used. Therefore, if problems are encountered with geometry optimizations, particularly if internal coordinates are used, keyword `xyz` should be used.

Notes

- If `xyz` is used, the optimization flags are not changed. Therefore, before running a system with `xyz` all the optimization flags should be set to "1". For atoms 1, 2, and 3, the optimization flags are forced to 1 if all the other optimization flags are set to 1. To be safe, use `xyz` with [OSCF](#), then edit the resulting ARC file to delete `xyz` and set the optimization flags by hand.
- Be very careful if some atoms have optimization flags set to "0". If the geometry supplied is in internal coordinates, and an atom, say atom 10, has all three optimization flags set to "0", then that means "Do not change the bond length, angle, and dihedral of atom 10, with reference to the atoms used for the connectivity." The atom can still move if the atom it is attached to moves. If that atom is converted to Cartesian coordinates, the optimization flags will still be zero, but now the definition changes to "Do not change the Cartesian coordinates of atom 10." This is very different from the internal coordinate definition.
- Any dummy atoms are automatically deleted if `xyz` is present. Dummy atoms are only meaningful if internal coordinates are used.

See also [INT](#)

$Z=n$

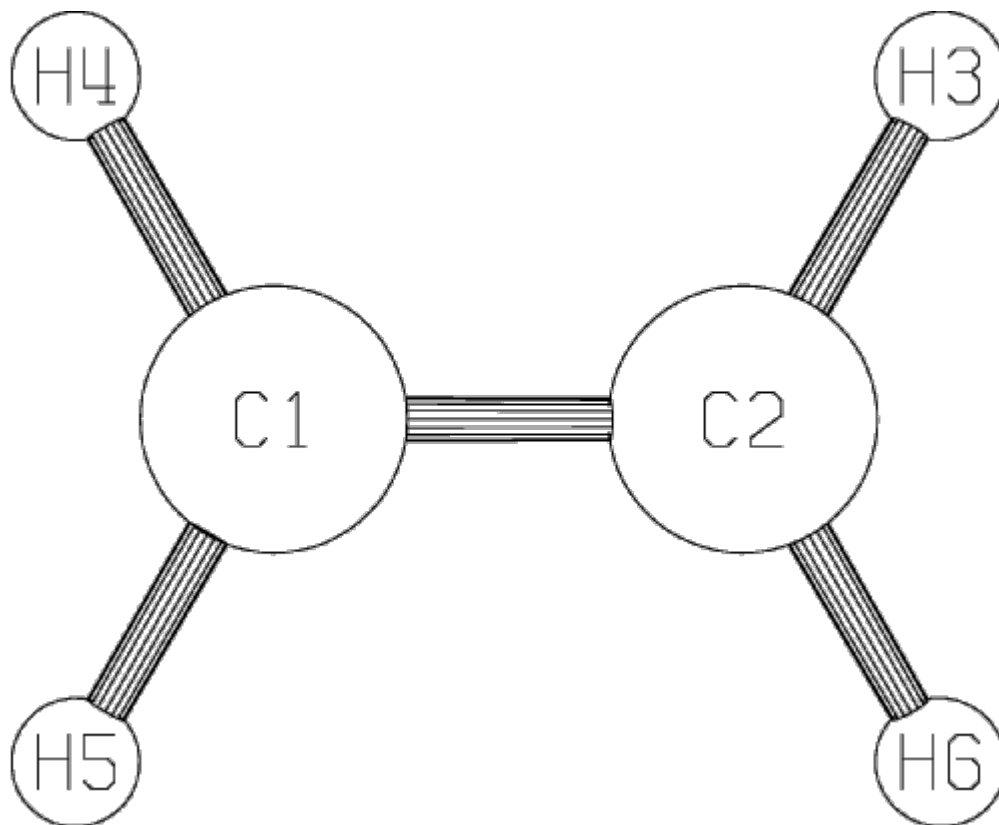
By default, the number of mers, Z , in a unit cell is worked out from the empirical formula. Thus, if a unit cell has the empirical formula $C_{24}H_{24}$, then $Z=24$. In those cases where the empirical formula should not be used, e.g., benzene, C_6H_6 , the number of mers can be explicitly defined by use of $Z=n$, here $Z=4$. Be careful to distinguish between the number of mers in a unit cell and the number in a cluster - $Z=n$ applies to the number in the unit cell.

If $Z=n$ should be used, but is not used, then the heat of formation per formula unit will be in error by a factor, in the case of crystalline benzene, the DH_f per formula unit would be 1/6 what it should be.

Example of Data for Ethylene

Ethylene is a simple organic molecule, of formula C_2H_4 , shown in the figure below.

Figure 1: Ethylene



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 2.

Figure 2: Data set for Ethylene

```

Line 1: VECTORS LOCAL PI PM6
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:
```

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 difficult 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. Comments are most often put before the first line of the data set (before Line 1), in which case they are printed in the .arc file and in the PDB file if [PDBOUT](#) is used, however, they can be anywhere, including in or after the Z-matrix, but these comments will not be printed in the .arc file. Comments are ignored by the calculation. They are useful when one or more atoms are to be excluded from a specific calculation. By adding an asterisk to the start of an atom-line, the atom is excluded; by removing its asterisk, an excluded atom can then be put back in to the calculation easily.

Figure 3: Data set for Acetylene, with comments

```
*****
*
*   Example of a MOPAC data set
*
*****
Line  1: VECTORS LOCAL PI PM6
Line  2: EXAMPLE OF MOPAC DATA-SET FOR ACETYLENE
Line  3: The only geometric constraint is that the system must be planar
Line  4: C
Line  5: C   1.3 1
*
*   The supplied C-C distance is 1.3 A, the C-H distance is 1.1
*   Angstroms, and the H-C-C angle is 120 degrees.
*   During optimization, the angle will open to ~180 degrees.
*
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
*
*   The next two lines are ignored
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:
*   Comments can be anywhere!
```

Elements Available in MNDO

I	II	Transition Metals										III	IV	V	VI	VII	VIII
H																	
Li	Be											B	C	N	O	F	
Na	Mg											Al	Si	P	S	Cl	
K	Ca										Zn	Ga	Ge	As	Se	Br	
Rb	Sr										Cd	In	Sn	Sb	Te	I	
Cs	Ba										Hg	Tl	Pb	Bi			

In addition to the elements, other symbols used in geometry definition are:

XX	A dummy atom for assisting with geometry specification
Tv	A translation vector for use with polymers, layer systems, and solids
	See also MAKPOL .
Cb	The "capped bond" atom
+3	A "sparkle" with a charge of +3
++	A "sparkle" with a charge of +2
+	A "sparkle" with a charge of +1
Fr	A "sparkle" with a charge of +1/2
At	A "sparkle" with a charge of -1/2
-	A "sparkle" with a charge of -1
--	A "sparkle" with a charge of -2
-3	A "sparkle" with a charge of -3

Elements +, ++, -, and -- are the "[Sparkles](#)"; Tv is the translation vector for polymer calculations. The half-integer entities Fr and At are intended for use in simulating crystal fields for [transition metal complex](#) work, and should only be used in even numbers.

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.

Elements Available in AM1

I	II	Transition Metals										III	IV	V	VI	VII	VIII
H																	
Li	Be												C	N	O	F	
Na	Mg											Al	Si	P	S	Cl	
K	Ca										Zn	Ga	Ge	As	Se	Br	
Rb	Sr				Mo						Cd	In	Sn	Sb	Te	I	
Cs	Ba	*									Hg	Tl	Pb	Bi			

Lanthanides:

La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

The lanthanides are represented by the Ln3+ sparkles. These give good geometries, but do not have a basis set, so partial charges are not possible.

In addition to the elements, other symbols used in geometry definition are:

XX	A dummy atom for assisting with geometry specification
Tv	A translation vector for use with polymers, layer systems, and solids
	See also MAKPOL .
Cb	The "capped bond" atom
+3	A "sparkle" with a charge of +3
++	A "sparkle" with a charge of +2
+	A "sparkle" with a charge of +1
Fr	A "sparkle" with a charge of +1/2
At	A "sparkle" with a charge of -1/2
-	A "sparkle" with a charge of -1
--	A "sparkle" with a charge of -2
-3	A "sparkle" with a charge of -3

Elements +, ++, -, and -- are the "[Sparkles](#)"; Tv is the translation vector for polymer calculations. The half-integer entities Fr and At are intended for use in simulating crystal fields for [transition metal complex](#) work, and should only be used in even numbers.

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.

Elements Available in PM3

I	II	Transition Metals										III	IV	V	VI	VII	VIII
H																	
Li	Be											B	C	N	O	F	
Na	Mg											Al	Si	P	S	Cl	
K	Ca										Zn	Ga	Ge	As	Se	Br	
Rb	Sr										Cd	In	Sn	Sb	Te	I	
Cs	Ba	Ln ³⁺									Hg	Tl	Pb	Bi			

Lanthanides:

La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

The lanthanides are represented by the Ln³⁺ sparkles. These give good [geometries](#), but do not have a basis set, so partial charges are not possible.

In addition to the elements, other symbols used in geometry definition are:

XX	A dummy atom for assisting with geometry specification
Tv	A translation vector for use with polymers, layer systems, and solids
See also MAKPOL .	
Cb	The "capped bond" atom
+3	A "sparkle" with a charge of +3
++	A "sparkle" with a charge of +2
+	A "sparkle" with a charge of +1
Fr	A "sparkle" with a charge of +1/2
At	A "sparkle" with a charge of -1/2
-	A "sparkle" with a charge of -1
--	A "sparkle" with a charge of -2
-3	A "sparkle" with a charge of -3

Elements +, ++, -, and -- are the "[Sparkles](#)"; Tv is the translation vector for polymer calculations. The half-integer entities Fr and At are intended for use in simulating crystal fields for [transition metal complex](#) work, and should only be used in even numbers.

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.

Elements Available in RM1

I	II	Transition Metals										III	IV	V	VI	VII	VIII
H																	
													C	N	O	F	
														P	S	Cl	
																Br	
																I	

In addition to the elements, other symbols used in geometry definition are:

XX	A dummy atom for assisting with geometry specification
Tv	A translation vector for use with polymers, layer systems, and solids
	See also MAKPOL .
Cb	The "capped bond" atom
+3	A "sparkle" with a charge of +3
++	A "sparkle" with a charge of +2
+	A "sparkle" with a charge of +1
-	A "sparkle" with a charge of -1
--	A "sparkle" with a charge of -2
-3	A "sparkle" with a charge of -3

Elements +, ++, -, and -- are the "[Sparkles](#)"; Tv is the translation vector for polymer calculations.

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.

Elements Available in PM6

I	II	Transition Metals												III	IV	V	VI	VII	VIII												
H																								He							
Li	Be																									B	C	N	O	F	Ne
Na	Mg																									Al	Si	P	S	Cl	Ar
K	Ca	Sc		Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr													
Rb	Sr	Y		Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe													
Cs	Ba	La	Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi																

[Lanthanides](#)[§]:

La *	Ce *	Pr *	Nd *	Pm *	Sm *	Eu *	Gd *	Th *	Dy *	Ho *	Er *	Tm *	Yb *	Lu *
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

*: Available as 3+ sparkles, if keyword SPARKLE is used. Hyperlink is to individual example data sets.
§: Hyperlink to a ZIP file containing example data sets for all the Lanthanides

In addition to the elements, other symbols used in geometry definition are:

XX	A dummy atom for assisting with geometry specification
Tv	A translation vector for use with polymers, layer systems, and solids See also MAKPOL .
Cb	The "capped bond" atom
+3	A "sparkle" with a charge of +3
++	A "sparkle" with a charge of +2
+	A "sparkle" with a charge of +1
Fr	A "sparkle" with a charge of +1/2
At	A "sparkle" with a charge of -1/2
-	A "sparkle" with a charge of -1
--	A "sparkle" with a charge of -2
-3	A "sparkle" with a charge of -3

Elements +, ++, -, and -- are the "[Sparkles](#)"; Tv is the translation vector for polymer calculations. The half-integer entities Fr and At are intended for use in simulating crystal fields for [transition metal complex](#) work, and should only be used in even numbers.

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.

Isotope Specification

By default, atomic masses are those that represent the normal isotopic mixture, so, for example, the atomic weight of chlorine is 35.45. If specific isotopes are needed, the isotopic mass would be given immediately after the atomic symbol. Thus if an atom in a molecule was a labeled C-13, then the symbol for the atom would be: C13.0034.

Because the three isotopes of hydrogen are so important, they have been given special symbols. If "heavy hydrogen" - deuterium - is wanted, then use the symbol "D", if tritium is wanted, use the symbol "T"

For example, the geometry specification for heavy water could be written as:

D
O 1.0 1
D 1.0 1 120 1 0 0 2 1

or as

H2.01355
O 1.0 1
H2.01355 1.0 1 120 1 0 0 2 1

Description of Atom Labels

The most simple definition of an atom is to use either its chemical symbol or its atomic number, thus:

```
Mg 1.23 1 1.23 1 1.23 1 or
12 1.23 1 1.23 1 1.23 1
```

If [isotopes](#) are used, then the atomic number option cannot be used.

In addition to the symbol or atomic number, a label can be added to an atom. The format of the label is: "*(text)*" (without the quotation marks), that is, "open curved bracket, *text*, closed curved bracket" The label should be immediately after the chemical symbol or after the isotopic mass, if present.

text can be up to 38 characters long, but much shorter descriptions should normally be used, otherwise the layout of the output becomes very messy. *text* can be any description that does not involve any of the symbols ")", or "+", or "-" if [MOZYME](#) is used.

Examples of atom labels:

```
Mg(At center of porphyrin ring) 1.23 1 1.23 1 1.23 1
C( 17 PHE* 2) 1.19 1 4.59 1 -0.18 1
C14.0( 17 PHE* 2) 1.19 1 4.59 1 -0.18 1
```

If [RESIDUES](#) is present, the resulting ARC file will include labels on each atom.

Atom Charges

Charges on individual charges for use in MOZYME can be specified by the symbols "+" and "-" in the atom labels. When the symbol "+" is present, the atom is given a unit positive charge, when "-" is present, the atom is given a unit negative charge. Avoid using these keywords unless necessary. Thus if the ethyl group, C₂H₅, is run with [CHARGE=1](#), then the CH₂ carbon will automatically be given a positive charge, if [CHARGE=-1](#) is present, then it will be given a negative charge.

See also [SETPI](#) to explicitly assign p bonds, also [CVB](#) to explicitly make or break bonds.

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 [Torsion or Dihedral Angle Coherency](#). 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 on geometric parameters in Z-matrices

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 0 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](#).

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.

Here is an example of a MOPAC data-set for formaldehyde, in which all coordinates are Cartesian.

PM3

Formaldehyde

All coordinates are Cartesian

O	0.00	0	0.00	0	0.00	0
C	1.21	1	0.00	0	0.00	0
H	1.79	1	0.93	1	0.00	0
H	1.79	1	-0.93	1	0.00	0

the "1" after the "0.93" indicates that the y coordinate of the first hydrogen atom is to be optimized.

Mixed Cartesian and Internal Coordinate Definition

With the exception of the first atom, real or dummy, atom positions may be defined using either Cartesian or internal coordinates. The coordinates of an atom are defined using internal coordinates if a connectivity is supplied. If the connectivity is missing, the atom is defined in terms of Cartesian coordinates. Atom 1, because it does not have a connectivity, is always defined in Cartesian coordinates. Although these are usually (0.0, 0.0, 0.0) there is no a priori reason for this, and atom 1 can be put at any user-defined position.

An atom's position can be defined using internal coordinates involving atoms whose positions have not yet been defined if those atoms are defined using Cartesian coordinates. This is different from the normal restriction of internal coordinates, where an atom's position can only be defined using atoms that have already been defined. Put another way, if the position of atom 100 is defined using atoms 110, 120, and 60, then the positions of atoms 110 and 120 must be defined using Cartesian coordinates. Atom 60 can be in Cartesian or internal, since its position has already been defined.

In the following example, the geometry of a tripeptide is defined using Cartesian coordinates for atom 1 and all the atoms in the second residue. The rest of the atoms in residue 1 and all the atoms in residue 3 are defined using internal coordinates.

Note that the first three atoms of the third residue are defined using atoms whose positions are given in Cartesian coordinates. This is a valid definition - Cartesian and internal coordinates can be mixed together in any way desired.

Example of a mixed coordinate definition:

```
Gradients lscf
Example of mixed coordinates. A tripeptide with residues 1 and 3 in internal coordinates
residue 2 is in Cartesian coordinates
N      0.00000000 +0      0.00000000 +0      0.00000000 +0
C      1.45099200 +1      0.00000000 +0      0.00000000 +0      1      0      0
C      1.51835900 +1      114.2160460 +1      0.00000000 +0      2      1      0
O      1.22627600 +1      119.2494590 +1      150.0224060 +1      3      2      1
C      1.53154900 +1      110.2108510 +1      -122.9442610 +1      2      1      3
S      1.81497400 +1      108.2495690 +1      -116.9760990 +1      5      2      1
H      1.09144500 +1      107.8966510 +1      117.4730020 +1      2      1      3
H      1.09439500 +1      110.6252580 +1      125.3849350 +1      5      2      1
H      1.09375800 +1      111.5846500 +1      4.2686790 +1      5      2      1
H      1.37249300 +1      102.4190520 +1      -179.3933780 +1      6      5      2
H      1.01203000 +1      116.9993300 +1      127.1027860 +1      1      2      3
H      1.01203000 +1      116.9993300 +1      -127.1027860 +1      1      2      3
N      1.41779186 +1      2.3818640 +1      -0.5930817 +1
C      1.96005201 +1      3.7395708 +1      -0.5873182 +1
C      0.87144992 +1      4.6925379 +1      -0.0859199 +1
O      -0.30894710 +1      4.3852105 +1      -0.2277412 +1
C      2.58984680 +1      4.0757744 +1      -1.9613244 +1
C      3.95504011 +1      3.4174209 +1      -2.2109945 +1
C      5.13263384 +1      4.0760823 +1      -1.8023387 +1
C      6.39390379 +1      3.5338821 +1      -2.0746443 +1
C      6.51251534 +1      2.3040175 +1      -2.7409980 +1
C      5.35236761 +1      1.6339502 +1      -3.1585724 +1
C      4.08327785 +1      2.1864831 +1      -2.9055825 +1
H      2.75184646 +1      3.8479695 +1      0.1426395 +1
H      2.74129210 +1      5.1566073 +1      -1.9998683 +1
H      1.90797747 +1      3.8313152 +1      -2.7790034 +1
H      5.07815939 +1      5.0182815 +1      -1.2765870 +1
H      7.28256125 +1      4.0711027 +1      -1.8000485 +1
H      3.20599873 +1      1.6658896 +1      -3.2578962 +1
H      5.45626342 +1      0.6951745 +1      -3.6940611 +1
H      7.49097179 +1      1.8927662 +1      -2.9377954 +1
H      0.46717507 +1      2.2471723 +1      -0.9238687 +1
N      1.33611900 +1      118.9784180 +1      152.6871430 +1      15      14      13
C      1.45816400 +1      122.1765280 +1      173.0360570 +1      33      15      14
C      1.54029400 +1      106.1916250 +1      -98.1274360 +1      34      33      15
O      1.22473500 +1      118.4172840 +1      -38.7329790 +1      35      34      33
C      1.54046400 +1      112.6361960 +1      139.8205900 +1      34      33      15
C      1.53533100 +1      115.3620840 +1      -28.7237370 +1      37      34      33
C      1.52273000 +1      109.3021640 +1      175.4610420 +1      38      37      34
O      1.24543100 +1      114.2017740 +1      -65.1675870 +1      39      38      37
O      1.25543200 +1      119.1622320 +1      113.3463330 +1      39      38      37
H      1.08372800 +1      109.7991700 +1      24.1081960 +1      34      33      15
H      1.09211200 +1      107.8205160 +1      -148.0829980 +1      37      34      33
H      1.09379600 +1      109.2718540 +1      94.4693220 +1      37      34      33
H      1.08910800 +1      109.7719310 +1      58.7340750 +1      38      37      34
```

Mixed Cartesian and Internal Coordinate Definition

H	1.08916200	+1	111.1381130	+1	-61.8510410	+1	38	37	34
H	0.90000000	+1	120.0000000	+1	-179.9999990	+1	40	39	38
H	1.01192900	+1	119.8483270	+1	10.8020530	+1	33	15	14
H	1.34588100	+1	121.3449610	+1	139.3147470	+1	35	34	33

Translation Vector for defining Polymers, Layer systems, and Solids

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).

Translation vectors can be defined in internal coordinates or Cartesian coordinates. Both definitions give rise to the same result, but the methods of definition are very different. Before proceeding, remember that atom positions are defined as being in internal coordinates if there is a connectivity. If there is no connectivity, the atom position is defined in Cartesian coordinates. The utility [MAKPOL](#) is provided to simplify construction of MOPAC data sets for solids.

The number of translation vectors determines the type of infinite solid. A polymer has one T_V , a layer system has two T_V , and a solid has three T_V .

Internal coordinate definition of translation vector

If a T_V has a connectivity, then its position is defined using internal coordinates. The translation vector is then given by the difference in position of the atom the T_V is connected to and the position defined by T_V . For example, consider polyethylene:

Polyethylene, $-[C_{12}H_{24}]-$

C	0.00000000	+0	0.00000000	+0	0.00000000	+0				-0.2638
C	1.53502951	+1	0.00000000	+0	0.00000000	+0	1	0	0	-0.2639
C	1.53434080	+1	111.4738515	+1	0.00000000	+0	2	1	0	-0.2638
C	1.53486713	+1	111.4515730	+1	-179.5035734	+1	3	2	1	-0.2638
C	1.53436383	+1	111.4544128	+1	-179.5609499	+1	4	3	2	-0.2639
C	1.53487336	+1	111.4657224	+1	-179.9694611	+1	5	4	3	-0.2638
C	1.53438258	+1	111.4286979	+1	-179.7424610	+1	6	5	4	-0.2638
C	1.53483201	+1	111.4669973	+1	179.6187779	+1	7	6	5	-0.2638
C	1.53442061	+1	111.4159036	+1	179.6200571	+1	8	7	6	-0.2639
C	1.53492245	+1	111.4448809	+1	179.5385773	+1	9	8	7	-0.2638
C	1.53443267	+1	111.3905812	+1	179.4842656	+1	10	9	8	-0.2639
C	1.53474400	+1	111.3843156	+1	-179.9610941	+1	11	10	9	-0.2638
H	1.10720431	+1	109.8104572	+1	-57.5906521	+1	1	2	3	0.1319
H	1.10698586	+1	109.8681806	+1	116.0168453	+1	1	2	13	0.1319
H	1.10709315	+1	109.8700918	+1	122.0438631	+1	2	1	3	0.1319
H	1.10713575	+1	109.8236237	+1	115.9314480	+1	2	1	15	0.1319
H	1.10712021	+1	109.8652470	+1	-122.0100491	+1	3	2	4	0.1319
H	1.10709847	+1	109.8813887	+1	-116.0033729	+1	3	2	17	0.1319
H	1.10711978	+1	109.8682857	+1	122.0383223	+1	4	3	5	0.1319
H	1.10711511	+1	109.8397525	+1	115.9441858	+1	4	3	19	0.1319
H	1.10713850	+1	109.8677152	+1	-122.0280078	+1	5	4	6	0.1319
H	1.10712347	+1	109.8680744	+1	-115.9603262	+1	5	4	21	0.1319
H	1.10712096	+1	109.8603331	+1	122.0225998	+1	6	5	7	0.1319
H	1.10706606	+1	109.8553502	+1	115.9629566	+1	6	5	23	0.1319
H	1.10712955	+1	109.8765480	+1	-121.9985907	+1	7	6	8	0.1319
H	1.10711605	+1	109.8542406	+1	-115.9994488	+1	7	6	25	0.1319
H	1.10707238	+1	109.8604415	+1	122.0413124	+1	8	7	9	0.1319
H	1.10709630	+1	109.8633232	+1	115.9604117	+1	8	7	27	0.1319
H	1.10708893	+1	109.8924995	+1	-122.0076508	+1	9	8	10	0.1319
H	1.10711581	+1	109.8477424	+1	-116.0065534	+1	9	8	29	0.1319
H	1.10704898	+1	109.8675671	+1	122.0217258	+1	10	9	11	0.1319
H	1.10708168	+1	109.8627808	+1	115.9892025	+1	10	9	31	0.1319
H	1.10704508	+1	109.9042865	+1	-122.0607012	+1	11	10	12	0.1319
H	1.10715046	+1	109.8566042	+1	-115.9742091	+1	11	10	33	0.1319
H	1.10698039	+1	109.8868070	+1	-58.2346895	+1	12	11	10	0.1319
H	1.10741358	+1	109.7138737	+1	116.0025351	+1	12	11	35	0.1320
T_V	15.21694969	+1	34.2865646	+1	-0.0375057	+1	1	2	3	

The 'bond length', 15.22 Ångstroms, is the distance of T_V from atom 1. The direction is given by the angle (here 34 degrees) and dihedral (-0.04 degrees). So atom 1, on translation, would be moved to the position of T_V . All other atoms would then be moved the same way.

But this way of defining T_V has a severe drawback - a small change in the angle or dihedral of T_V would produce a large change in position of the T_V . An easy way around this is to put a dummy atom at the position of T_V , and then define T_V in terms of atom 1 and the dummy atom. To illustrate this, consider polyethylene again. By adding a dummy atom, the end of the data set now looks like this:

H	1.10698039	+1	109.8868070	+1	-58.2346895	+1	12	11	10	0.1320
H	1.10741358	+1	109.7138737	+1	116.0025351	+1	12	11	35	0.1320
XX	1.53482200	+1	111.5010360	+1	122.0188610	+1	12	11	36	
Tv	15.21694969	+1	0.0000000	+0	0.0000000	+0	1	37	35	

Now the angle and dihedral of Tv can be defined using the dummy atom. The angle Tv - atom 1 - dummy atom can be fixed at zero degrees, and marked "not to be optimized." The dihedral is not important, but needs to be given, so select any other atom for the connectivity and lock the dihedral at zero degrees. This gives a perfectly general and robust method of Tv in terms of internal coordinates.

Cartesian coordinate definition of translation vector

If the connectivity is missing Tv is defined as Cartesian coordinates. The absolute position of Tv defines the motion of all atoms. Consider polyethylene again. In Cartesian coordinates, the data set would be:

Polyethylene, -[C12H24]-

C	0.00000000	+1	0.00000000	+1	0.00000000	+1				0.0000
C	1.53502951	+1	0.00000000	+1	0.00000000	+1				0.0000
C	2.09671572	+1	1.4278341	+1	0.00000000	+1				0.0000
C	3.63153284	+1	1.4272569	+1	-0.0123771	+1				0.0000
C	4.19338646	+1	2.8550080	+1	-0.0014355	+1				0.0000
C	5.72820359	+1	2.8547421	+1	-0.0145707	+1				0.0000
C	6.28919628	+1	4.2827881	+1	0.0028023	+1				0.0000
C	7.82402390	+1	4.2834120	+1	-0.0008140	+1				0.0000
C	8.38373432	+1	5.7120857	+1	0.0070968	+1				0.0000
C	9.91863588	+1	5.7134285	+1	0.0149906	+1				0.0000
C	10.47706263	+1	7.1426306	+1	0.0100363	+1				0.0000
C	12.01179057	+1	7.1438002	+1	0.0169597	+1				0.0000
H	-0.37524221	+1	0.5583029	+1	0.8794274	+1				0.0000
H	-0.37621723	+1	0.5451135	+1	-0.8869777	+1				0.0000
H	1.91131794	+1	-0.5524188	+1	0.8825507	+1				0.0000
H	1.91048784	+1	-0.5523063	+1	-0.8830279	+1				0.0000
H	1.72795936	+1	1.9771690	+1	0.8876740	+1				0.0000
H	1.71422443	+1	1.9828794	+1	-0.8782324	+1				0.0000
H	4.01464282	+1	0.8652321	+1	0.8611611	+1				0.0000
H	3.99989284	+1	0.8845857	+1	-0.9042980	+1				0.0000
H	3.82488823	+1	3.3975418	+1	0.8905409	+1				0.0000
H	3.81002536	+1	3.4169220	+1	-0.8749396	+1				0.0000
H	6.11169319	+1	2.2888993	+1	0.8563337	+1				0.0000
H	6.09647982	+1	2.3164462	+1	-0.9091127	+1				0.0000
H	5.91538732	+1	4.8232426	+1	0.8938195	+1				0.0000
H	5.91095296	+1	4.8460271	+1	-0.8720690	+1				0.0000
H	8.20242684	+1	3.7254711	+1	0.8773217	+1				0.0000
H	8.19832236	+1	3.7375404	+1	-0.8882750	+1				0.0000
H	8.00284989	+1	6.2609926	+1	0.8898619	+1				0.0000
H	8.01178944	+1	6.2664053	+1	-0.8761314	+1				0.0000
H	10.29078179	+1	5.1665370	+1	0.9026699	+1				0.0000
H	10.29979525	+1	5.1573344	+1	-0.8631362	+1				0.0000
H	10.09560328	+1	7.6993624	+1	0.8875824	+1				0.0000
H	10.10444119	+1	7.6888972	+1	-0.8779548	+1				0.0000
H	12.38474913	+1	6.6007161	+1	0.9065479	+1				0.0000
H	12.38973517	+1	6.5834579	+1	-0.8602742	+1				0.0000
Tv	12.57289793	+1	8.5719164	+0	0.0067071	+0				

Now Tv is defined as X=12.57, Y=8.57, and Z=0.01. On translation, every atom would be displaced by this amount. Although it is the default, atom 1 does not need to be at the origin.

The definition of Tv, Cartesian or internal, does not depend on the definitions of any of the other atoms, that is, some atoms can be in Cartesian coordinates and some in internal coordinates. The definitions gives here refer to the Tv only.

For polymers, internal coordinates are easier than Cartesian coordinates. For solids, Cartesian coordinates are recommended, unless the system has high symmetry, in which case internal coordinates are preferred, as that allows extensive use of symmetry functions to reduce the number of geometric parameters to be optimized. In most high symmetry cases, only one or two geometric parameters need to be optimized.

Optimization flags

Each internal coordinate is followed by an flag, to indicate the action to be taken.

Flag Action

1 Optimization Flags

0 Do not optimize the coordinate.

-1 Reaction coordinate, or grid index.

T Monitor turning points in DRC

Consider the data set for formaldehyde:

SYMMETRY
Formaldehyde

```
O  0.0  0      0.0  0      0.0  0
C  1.2  1      0.0  0      0.0  0  1  0  0
H  1.1  1  120.0  1      0.0  0  2  1  0
H  1.1  0  120.0  0  180.0  0  2  1  3

  3  1      4
  3  2      4
```

The optimization flags for the first atom, oxygen, are all zero, because the system is in internal coordinates and translating the molecule would not alter its energy at all.

For atom 2, carbon, the first coordinate, the C=O bond length, 1.2Å, is flagged for optimization, the other two coordinates are not flagged for optimization, because there is no angle or dihedral.

Atom 3, hydrogen, has both the C-H bond length, 1.1Å, and the H-C=O angle, 120 degrees, flagged for optimization.

Atom 4, also a hydrogen, has no coordinates marked for optimization, because the C-H and H-C=O parameters are [symmetry-defined](#) to be the same as those of atom 3, and the dihedral is exactly 180 degrees.

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 [Dynamic and Intrinsic Reaction Coordinates](#) for more detail.

Applying geometry constraints, freezing coordinates and fixing distances, etc

A wide range of methods exist that allow constraints to be applied to a geometry. The most important are:

- [Cartesian coordinates can be fixed by setting the appropriate optimization flags to zero.](#)
- [Internal coordinates can be fixed by setting the appropriate optimization flags to zero.](#)
- [Cartesian coordinates can be set equal using symmetry](#)
- [Internal coordinates can be set equal by using symmetry](#)
- [An interatomic distance can be set to a multiple of another interatomic distance](#)
- [An atom can be frozen in a DRC](#)
- [Individual internal coordinates can be fixed in a Cartesian geometry optimization](#)
- [The overall structure can be biased in favor of a reference geometry](#)
- [An overall structure can be biased \(moved\) towards the structure on the other side of a barrier.](#)

Cartesian coordinates can be fixed by setting the appropriate optimization flags to zero.

Any coordinate, internal or Cartesian, can be optimized if the optimization flag is set to "1" and the coordinate is not otherwise defined, for example by [SYMMETRY](#). This flexibility applies to all atoms, including atoms 1, 2, and 3. If the coordinates of atom 1 are flagged to be optimized, and atom 1 has no connectivity (the usual situation), then it is defined in Cartesian coordinates and its position can change during a geometry optimization. If atom 2 has a connectivity, then setting the angle or dihedral is meaningless, and these will be turned off before the calculation starts. The same applied to the dihedral of atom 3, if it, too, is defined in internal coordinates.

In this example, an acetylene molecule is oriented along the "z" axis, and the second carbon atom is fixed in space at coordinates (0.0, 0.0, 1.0). This calculation is most easily set up using Cartesian rather than internal coordinates. Acetylene is perfectly linear, so the "x" and "y" coordinates can be fixed at 0.0. On optimizing this system, carbon atom 1 would move down along the "z" axis, in order from the C-C distance to increase from 1.0 Ångstroms.

```
PM6
Acetylene oriented along the "z" axis,
Atom 2 is fixed at coordinates (0.0, 0.0, 1.0)
C      0.00000000 +0    0.00000000 +0    0.00000000 +1
C      0.00000000 +0    0.00000000 +0    1.00000000 +0
H      0.00000000 +0    0.00000000 +0    2.00000000 +1
H      0.00000000 +0    0.00000000 +0   -1.00000000 +1
```

Freezing Cartesian coordinates is useful when positioning charged species in an electric field, for example a fluoride ion at coordinates (0.0, 0.0, 1.0) when a field of 1V per Ångstrom is applied along the "z" direction. Less common is the use of Cartesian coordinates in small systems.

Internal coordinates can be fixed by setting the appropriate optimization flags to zero.

Consider chloromethane, CH₃Cl, in which two constraints are applied: the C-Cl distance is fixed at 1.7 Ångstroms, and one C-H distance is to be fixed at 1.1 Ångstroms. The data set, in internal coordinates, might look like this:

```

PM6
Methyl chloride, C-Cl distance defined as 1.7 Angstroms,
One C-H distance defined as 1.1 Angstroms
Cl      0.0 +0      0.00000000 +0      0.00000000 +0
C       1.7 +0      0.00000000 +0      0.00000000 +0      1      0      0
H       1.0 +1     109.2843786 +1      0.00000000 +0      2      1      0
H       1.1 +0     109.2843786 +1     120.00000000 +1      2      1      3
H       1.0 +1     109.2843786 +1    -120.00000000 +1      2      1      3

```

In this system, there are only two constraints although there are eight optimization flags set to zero. This is because six of the optimization flags are for the three trivial translations (atom 1) and three trivial rotations (atoms 2 and 3)

Cartesian coordinates can be set equal using symmetry

This is a rarely used option. An example is given in Figure 2 of [SYMMETRY](#).

Internal coordinates can be set equal by using symmetry

Many small molecules have symmetry in that various bond lengths or angles or dihedrals are equal, or are simply related. For an example, see Figure 1 of [SYMMETRY](#). If symmetry is used, and a dependent coordinate is flagged for optimization, the flag will be turned off before the calculation is started.

An interatomic distance can be set to a multiple of another interatomic distance

In some solids, particularly the binary solids such as NaCl or CsCl, the number of independent geometric variables is small. In the cases just mentioned, the number is precisely one. By setting up a symmetry-constrained calculation involving only one variable, the computational effort required for geometry optimization can be reduced considerably. An example of the steps involved in setting up such a calculation are as follows:

- First, set up the Large Unit Cell (LUC) system for NaCl in internal coordinates, see [Ty](#).. For a 3 by 3 by 3 LUC constructed using [MAKPOL](#), the resulting [data set](#) has 108 NaCl units (Z=4).
- Run a 0SCF calculation with [AUTOSYM](#).. Edit the resulting arc file to make a new data set. During this process, change every angle optimization flag to zero - all the angles are symmetry-defined by the space-group, and therefore should not be optimized.
- There are five interatomic distances that are flagged for optimization, but all of them are related to the distance between atoms 1 and 2, that is, to the bond-length of atom 2. This distance, 2.82Å, can be used to symmetry-define all the other distances using symmetry function 19. Starting with the second unique bond-length, 3.988Å, divide that distance by the reference distance to give a ratio of 1.4142. This is then used in constructing the symmetry function:
2 19 1.4142 3.
This function is put at the start of the symmetry functions, just after the Z-matrix. The optimization flag for the bond-length of atom 3 can then set to zero, as it is now symmetry defined to that of atom 2. The same process is then done for the remaining three distances. In each case, the multiplier is an easily recognizable number: $2^{1/2}$, $3^{1/2}$, 2, and 6.
- The number of optimization flags in the new [symmetry-defined data-set](#) having now been reduced to exactly one, geometry optimization can be completed in one line-search.

An atom can be frozen in a DRC

In a DRC, atoms move in response to the forces acting on them. For a given force, the heavier atoms move less than light atoms. If, for any reason, an atom needs to be frozen, set its apparent weight to a large number. For example, if a carbon atom should be fixed, replace the symbol "C" by "C9999999" or "C1.d12" or other mass. The frozen atom will still move, but not as much as an ordinary atom. To verify that the frozen atom is in fact not

moving, use [LARGE=-1](#)

Individual internal coordinates can be fixed in a Cartesian geometry optimization

Because large molecule geometries are normally defined in Cartesian coordinates, it might seem difficult to lock individual bond-lengths or angles so that their values cannot be modified during the optimization. The answer is to define individual atoms using internal coordinates. Suppose that, in a large molecule, hydrogen atom 3000 needs to be defined as being 1.1 Å from oxygen atom 2999. Then using a graphics package, locate H3000 and O2999, and two other atoms that could be used in defining the angle and dihedral in an internal coordinate connectivity. Suitable atoms might be the atom O2999 is connected to (e.g. C2967) and the next atom (e.g. N2960). Of course, any atoms could be used provided their atom-number is lower than the atom being defined, i.e. here lower than 3000. Once the four atoms (here H3000, O2999, C2967, and N2960) are identified, work out the angle and dihedral. Use these data in constructing the internal coordinate definition for H3000. If the H3000-O2999-C2967 angle was 120°, and the H3000-O2999-C2967-N2960 torsion was 180°, then the position of the hydrogen atom with a fixed O-H distance would be:

```
H 1.1 0 120 1 180 1 2999 2976 2960
```

If the angle needed to be fixed, then the line would be:

```
H 1.1 1 120 0 180 1 2999 2976 2960
```

and so on. There is no restriction on the number of internal coordinates that can be used. Indeed, entire hetero groups can be defined as being rigid, but free to move within a cavity. In such a case, the first atom in the hetero group would be defined in either unconstrained internal or Cartesian coordinates, the next atom would be in internal coordinates and have its bond-length fixed, the next atom would be in internal coordinates and have a frozen bond length and angle, and all the remaining atoms in the hetero group would be in internal coordinates and have all coordinates frozen.

In a small molecule, in this example: toluene, if the molecule is defined in Cartesian coordinates and a C-C bond-length is to be frozen, then a suitable data set would be:

```
PM6
Toluene
Cartesian definition, but with one C-C distance frozen
C      -0.00997463 +1    -0.0015231 +1    -0.0000930 +1
C      2.79787104 +1    -0.0096765 +1    -0.0012704 +1
C      0.69635100 +1     1.2055923 +1     0.0410044 +1
C      0.68930681 +1    -1.2109956 +1    -0.0416784 +1
C      2.09215830 +1     1.2048435 +1     0.0406317 +1
C      2.08705695 +1    -1.2186939 +1    -0.0423481 +1
H     -1.09719779 +1     0.0021777 +1     0.0002893 +1
H      0.15466169 +1     2.1496180 +1     0.0735817 +1
H      0.14334405 +1    -2.1525875 +1    -0.0736970 +1
H      2.63556968 +1     2.1465054 +1     0.0727064 +1
H      2.62170344 +1    -2.1652386 +1    -0.0748682 +1
C      1.5 0 120 1 180 1 2 5 3
H      4.68822161 +1     0.4932505 +1     0.8985801 +1
H      4.72735404 +1    -1.0061093 +1    -0.0329544 +1
H      4.68721477 +1     0.5478562 +1    -0.8700774 +1
```

A GUI is essential for working out the bond-length, angle, and dihedral to be used.

What to when the atoms to be used in the connectivity occur after the atom whose bond length or angle is to be locked

Quite often, the geometry does not allow a bond-length or angle to be defined, because the atoms that would be used in the connectivity are defined after the atom of interest. Fortunately, there is an easy way to solve this problem: the order of occurrence of Cartesian atoms is not important, so simply move the atoms to be used in the connectivity to before the atom whose bond length or angle is to be defined. Alternatively, move the atom whose bond length or angle is to be defined to the end of the data set.

The overall structure can be biased in favor of a reference geometry

X-ray structures can be improved (!) by combining the accuracy of prediction of bond-lengths and angles with the accurate X-ray tertiary structure. This is done using the [GEO_REF="<text>"](#) option. To use this option, first prepare a data-set starting with the X-ray structure. Resolve any structural or positional disorder, add hydrogen atoms as necessary, then optimize the positions of the hydrogen atoms. The result would be a data-set representing a chemically sensible structure in which all the non-hydrogen atoms are based on the X-ray structure. Copy this data set and give the copy a name such as ref.mop. In the original data set, add keywords

GEO_REF="/Users/<myname>/ref.mop" OR GEO_REF="M:/<pathname>/ref.mop". The default bias is 3 kcal/mol/Ångstrom. This can be changed by adding the bias after the closing quotation mark, as in GEO_REF="/Users/<myname>/ref.mop"10.

An overall structure can be biased (moved) towards the structure on the other side of a barrier.

Locating transition states has traditionally been difficult. However, if data sets representing the reactants and products are available, the [GEO_REF="<text>"](#) option can be used for moving the reactant geometry towards that of the product, and *vice-versa*. Let the reactant geometry be R.mop and the product geometry be P.mop. The product geometry can be moved in the direction of the reactant geometry by adding the keyword

GEO_REF="/Users/<myname>/R.mop" to the product geometry data set. When that constrained optimization is run, the result is a geometry that is nearer to the reactant geometry than the starting geometry. If the constraint were subsequently removed, and the geometry re-optimized, the geometry would slide back downhill in the direction of the original product geometry. For more information on locating transition states, see [protein calculations](#).

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 (see Figure), 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.

Figure: Example of Gaussian Z-matrix geometry specification

```

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      d4213      120.0
Line 17
Line 18      d300      300.0
Line 19

```

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.

3. Gaussian Cartesian coordinates are not supported.
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.

Conversion Between Various Formats

MOPAC can accept any of the following formats: Cartesian, MOPAC internal coordinates, a mixture of Cartesian and MOPAC internal coordinates, 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](#), 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. 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.

Preparing protein data sets

Protein specific features

Constructing data-sets for polypeptides, proteins and protein-like molecules is difficult. Most of the sources of protein structures, such as the [Protein Data Bank](#), do not include hydrogen atoms, or, if they are present, there are errors in where they are put. These errors usually result in sites being ionized when they should be neutral, and *vice versa*. If hydrogen atoms are not present, add them using any of the readily available utilities.

Before a meaningful calculation on a protein can be done, some checks need to be run to ensure that the system is chemically sensible. These checks are:

- Ensure that the ionized sites, if any, are correct.
- Ensure that the total ionization is in the range -3 to +3.
- Ensure that the system is not a radical (unless it should be a radical)

The sequence for running these checks is as follows:

1. Add keyword [MOZYME](#) and [CHARGE=500](#), and run the data set. It will run for a very short time, then quit. Look at the end of the output file. It will say that the charge is incorrect, obviously, but more important, it will also say which sites are charged. Using a GUI, look at each of these sites in turn, and check that they should or should not be charged. Make whatever changes are necessary.
2. Make sure that the net charge is in the range -3 to +3. Edit the keyword CHARGE=n to use the correct net charge.
3. If the system is a protein, a useful next step is to re-sequence the atoms to put them into the standard PDB format. To do this, run the data set with keyword [RESEQ](#). This will also identify all the 20 common amino acid residues. If any residues are not correct, edit the data set to correct them. If you are using an unusual residue, its formula can be added using keyword [XENO](#).
4. Before optimizing everything, first optimize the positions of all hydrogen atoms. To do this, set all optimization flags to zero, and do an optimization using [OPT-H](#).
5. To optimize the entire geometry, use Cartesian coordinates. The MOZYME technique is very efficient for geometry optimization. Once the geometry has been optimized, a single-point calculation can be done using conventional methods, if desired. This can be used to verify that MOZYME gave the correct heat of formation.

Bond alternation (PM6 only)

A sensitive test of the ability of computational chemistry methods to reproduce the electronic structure of molecular systems is provided by comparing the the degree of bond alternation predicted by a method with that observed experimentally. Bond alternation is the difference in bond-lengths of two adjacent bonds that, except for electronic effects, would be expected to have equal lengths. For example, in the hypothetical poly-acetylene, all C-C bonds would be expected to be equal, because they are in the same chemical environment, but electronic effects (double - single bonds alternating) might cause bond-length alternation.

For this test, a set of known species is used, this set covers the range from no bond alternation to large bond alternation. All reference systems are based on X-ray structures, with most of these coming from the [Cambridge Structural Database](#).

RHF PM6 severely over-estimates bond-alternation, and therefore was not used here, instead UHF PM6 was used. With the exception of spin-free systems, such as benzene, UHF methods do not give spin-quantized states, instead the states are mixtures of states. This, however, does not appear to affect the results: the degree of bond-alternation predicted by UHF PM6 appears to be similar to that observed.

Species	X-ray structure			UHF PM6 structure			
	Short	Long	Diff	Short	Long	Diff	
GAYTAB	1.361	1.463	0.102	1.347	1.486	0.139	The table given here shows the degree of bond-alternation found in the X-ray structure and that predicted by UHF PM6. Specific bonds used in determining bond alternation are:
EDIMEI	1.334	1.500	0.166	1.337	1.483	0.146	
FLCBOC	1.358	1.434	0.076	1.331	1.469	0.138	GAYTAB: the bond connecting two rings and the adjacent annulene bond.
SELQIK	1.400	1.418	0.018	1.376	1.447	0.071	
SENQEI	1.365	1.460	0.095	1.393	1.434	0.041	SELQIK, SENQEI, and WALJEZ: the bonds at the <i>meso</i> position
Nonaphyrin (WALJEZ)	1.381	1.400	0.019	1.411	1.424	0.013	
[18]-Annulene	1.401	1.410	0.009	1.410	1.410	0.000	[18]-Annulene: outer rim bonds.
[22]-Annulene	-	-	-	1.405	1.405	0.000	
[26]-Annulene	-	-	-	1.405	1.405	0.000	[22]-Annulene, [26]-Annulene, and [30]-Annulene: the two bonds at the middle of the straight polyacetylene section.
[30]-Annulene	-	-	-	1.405	1.405	0.000	
Polyacene	-	-	-	1.416	1.417	0.001	Polyacene and polyacetylene: any pair of adjacent C-C bonds in the backbone.
Polyacetylene	-	-	-	1.405	1.406	0.001	
Pentacene (PENCEN02)	1.406	1.407	0.001	1.403	1.405	0.002	Pentacene: the middle bonds in the direction of the long axis.
Azulene	-	-	-	1.398	1.398	0.000	
Azulene	-	-	-	1.414	1.414	0.000	Azulene: the two pairs of bonds that would be symmetry related if azulene had C _{2v} symmetry.
Naphthalene	1.368	1.406	0.038	1.372	1.423	0.051	
Triphenylene	1.421	1.460	0.039	1.421	1.460	0.039	Naphthalene: the C1-C2 and C2-C3 distances.
Triangular-4-phenylene	1.344	1.492	0.148	1.340	1.524	0.184	
Perfluoro Benzotricyclobutane (PFTEBZ11)	1.386	1.393	0.007	1.468	1.342	-0.126	Triphenylene, triangular-4-phenylene, perfluoro benzotricyclobutane, and SAJSOL02: adjacent bonds in the central benzene ring.
SAJSOL02	1.386	1.400	0.014	1.380	1.429	0.049	

When spin quantization is important, RHF-CI methods can be used. Here is a guide to the size of the active space to be used in the C.I.: If the system is small, such as benzene, use a C.I. equal to the number of double-bonds. For larger systems, try using a C.I. of 5 (C.I.=5) then a larger C.I., e.g., C.I.=6. If there is no significant change in the results, then stop. Otherwise increase the size of the C.I. and repeat until geometric changes are negligible.

A useful keyword here is MECI, this allows the low-lying electronic excited states to be printed.

There is no obvious reason for the large error in bond-alternation in perfluoro benzo[1,2-c]tricyclobutane.

More diagnostic or discriminatory examples would be welcomed.

Comparison of Methods in MOPAC2009

Method	No. of elements	Advantages over other methods	Disadvantages
MNDO	37	None	Low accuracy
MNDO/d	9	None	Used MNDO for non-MNDO/d elements
AM1	37	Lanthanides as sparkles	
PM3	37	Lanthanides as sparkles	
PM5	51	Good torsion angles in biphenyls	Not published. No good statistics on accuracy
RM1	10	Most accurate dipoles and I.P.s	Limited range of elements
PM6	70	Most accurate DH _f , geometries, good H-bonds	Zwitterions too stable, dipoles of low accuracy,
		Under active development.	non-bonded interactions often too strong

This page needs input from users. Notes on advantages and disadvantages and criticisms of this page are welcomed.

Average Errors in Heat of Formation (kcal/mol)

Element	PM6	No. in set	PM5	No. in set	PM3	No. in set	AM1	No. in set
1 Hydrogen	7.29	3039	20.13	3003	17.09	2340	21.12	2270
2 Helium	0.00	1	0.00	1	0.00	1	0.00	1
3 Lithium	7.98	83	15.31	83	18.02	83	18.84	82
4 Beryllium	5.92	34	29.06	34	29.58	34	18.51	34
5 Boron	6.44	122	11.15	122	11.84	120	13.80	5
6 Carbon	7.31	2828	20.17	2795	15.06	2155	19.42	2123
7 Nitrogen	8.22	1067	27.17	1051	20.96	761	24.23	744
8 Oxygen	8.42	1758	21.32	1733	20.13	1244	27.68	1229
9 Fluorine	8.49	497	23.98	491	21.25	350	37.40	334
10 Neon	0.00	1	0.00	1	0.00	1	0.00	1
11 Sodium	5.72	40	8.63	40	9.47	39	10.77	38
12 Magnesium	9.84	66	12.07	66	17.94	66	18.71	66
13 Aluminium	7.61	75	17.49	75	19.15	75	18.99	75
14 Silicon	6.51	98	10.34	98	12.80	96	17.00	95
15 Phosphorus	8.20	110	16.39	110	17.36	98	20.06	95
16 Sulfur	8.81	427	24.77	426	18.44	330	26.38	323
17 Chlorine	8.28	670	29.24	655	18.71	390	23.06	383
18 Argon	0.00	1	0.00	1	0.00	1	0.00	1
19 Potassium	6.53	43	12.74	43	9.36	42	28.38	41
20 Calcium	11.87	43	28.68	43	43.44	43	63.20	43
21 Scandium	10.33	52	25.79	52	-	-	-	-
22 Titanium	10.20	85	15.18	85	-	-	-	-
23 Vanadium	14.29	59	40.53	59	-	-	-	-
24 Chromium	14.09	60	35.86	60	-	-	-	-
25 Manganese	12.77	44	24.33	44	-	-	-	-
26 Iron	18.31	76	28.97	76	-	-	-	-
27 Cobalt	15.51	42	19.89	42	-	-	-	-

28 Nickel	15.10	51	35.63	51	-	-	-	-
29 Copper	13.00	47	38.36	47	-	-	-	-
30 Zinc	5.56	54	17.84	54	32.93	54	37.06	54
31 Gallium	7.51	47	29.12	47	37.58	47	46.87	47
32 Germanium	9.83	67	12.20	67	15.86	67	19.12	67
33 Arsenic	6.94	49	15.22	49	16.68	49	17.34	49
34 Selenium	4.40	25	39.58	25	39.71	25	32.00	25
35 Bromine	7.37	330	23.87	326	25.04	199	28.22	199
36 Krypton	0.00	1	0.00	1	0.00	1	0.00	1
37 Rubidium	10.91	24	16.57	24	21.47	24	29.33	23
38 Strontium	7.72	38	52.46	38	103.16	38	57.21	38
39 Yttrium	13.28	51	23.23	51	-	-	-	-
40 Zirconium	11.18	46	23.47	46	-	-	-	-
41 Niobium	8.57	51	114.82	51	-	-	-	-
42 Molybdenum	13.41	70	47.78	70	-	-	35.77	69
43 Technetium	15.14	50	45.69	50	-	-	-	-
44 Ruthenium	13.87	56	47.65	56	-	-	-	-
45 Rhodium	20.92	32	36.67	32	-	-	-	-
46 Palladium	11.65	47	19.17	47	-	-	-	-
47 Silver	4.67	14	31.04	14	-	-	-	-
48 Cadmium	3.49	38	34.66	38	61.92	38	-	-
49 Indium	7.33	54	31.53	54	29.83	54	32.16	54
50 Tin	7.14	77	16.83	77	17.10	77	20.21	77
51 Antimony	5.41	58	30.98	58	34.61	58	35.00	58
52 Tellurium	8.20	45	35.66	45	46.80	45	22.91	45
53 Iodine	7.23	279	26.28	274	25.90	176	36.55	175
54 Xenon	0.00	1	0.00	1	0.00	1	0.00	1
55 Cesium	6.89	40	37.01	40	35.22	40	55.33	39
56 Barium	12.12	37	98.20	37	154.65	37	161.09	37
57 Lanthanum	10.37	37	-	-	-	-	-	-
71 Lutetium	7.68	24	-	-	-	-	-	-

72 Hafnium	8.52	37	21.77	37	-	-	-	-
73 Tantalum	14.37	36	118.90	36	-	-	-	-
74 Tungsten	7.38	28	55.02	28	-	-	-	-
75 Rhenium	10.40	57	90.09	57	-	-	-	-
76 Osmium	6.46	19	24.87	19	-	-	-	-
77 Iridium	10.21	25	23.46	25	-	-	-	-
78 Platinum	11.61	77	20.55	77	-	-	-	-
79 Gold	12.82	32	44.68	32	-	-	-	-
80 Mercury	5.94	51	16.39	51	17.67	51	19.75	51
81 Thallium	10.42	44	32.63	44	73.96	45	73.18	45
82 Lead	7.92	44	18.08	44	14.18	44	16.71	44
83 Bismuth	7.74	53	99.88	53	28.95	53	119.23	53
All elements, avoiding double counting	8.01	4492	22.19	4431	18.20	3189	22.86	3106
Same as above, but with double counting	8.17	13764	24.15	13562	19.98	9492	25.35	9234

Average Errors in Bond Lengths (Angstroms)

Element	PM6	No. in set	PM5	No. in set	PM3	No. in set	AM1	No. in set
1 Hydrogen	0.044	238	0.061	237	0.032	217	0.035	181
2 Helium	0.251	6	0.459	6	0.182	4	0.655	5
3 Lithium	0.175	111	0.191	110	0.167	110	0.171	105
4 Beryllium	0.076	42	0.131	42	0.067	42	0.085	42
5 Boron	0.027	116	0.043	116	0.066	122	-	-
6 Carbon	0.057	1191	0.078	1163	0.051	634	0.063	628
7 Nitrogen	0.090	663	0.146	653	0.124	259	0.163	253
8 Oxygen	0.095	1163	0.122	1132	0.103	577	0.117	571
9 Fluorine	0.063	396	0.080	391	0.069	251	0.101	228
10 Neon	0.353	5	0.182	2	0.062	1	0.030	1
11 Sodium	0.229	33	0.200	33	0.208	30	0.140	29
12 Magnesium	0.089	106	0.067	106	0.167	105	0.073	106
13 Aluminium	0.045	77	0.120	76	0.098	70	0.138	70
14 Silicon	0.039	97	0.061	99	0.074	95	0.077	90
15 Phosphorus	0.039	141	0.138	134	0.073	92	0.083	87
16 Sulfur	0.094	359	0.133	347	0.091	207	0.134	200
17 Chlorine	0.069	672	0.103	647	0.095	284	0.130	285
18 Argon	0.258	4	0.303	1	-	-	-	-
19 Potassium	0.139	46	0.135	47	0.148	47	0.281	46
20 Calcium	0.133	67	0.177	69	0.151	67	0.102	60
21 Scandium	0.053	90	0.152	92	-	-	-	-
22 Titanium	0.078	140	0.127	132	-	-	-	-
23 Vanadium	0.090	168	0.098	171	-	-	-	-
24 Chromium	0.080	89	0.083	90	-	-	-	-
25 Manganese	0.083	107	0.106	108	-	-	-	-
26 Iron	0.102	117	0.071	118	-	-	-	-
27 Cobalt	0.107	100	0.109	101	-	-	-	-

28 Nickel	0.065	133	0.115	129	-	-	-	-
29 Copper	0.174	130	0.189	134	-	-	-	-
30 Zinc	0.076	77	0.084	77	0.098	77	0.142	76
31 Gallium	0.048	80	0.105	81	0.192	81	0.135	81
32 Germanium	0.038	131	0.044	132	0.056	133	0.068	133
33 Arsenic	0.073	72	0.070	71	0.080	72	0.099	72
34 Selenium	0.056	56	0.097	56	0.071	54	0.061	54
35 Bromine	0.104	358	0.101	347	0.146	182	0.136	184
36 Krypton	0.059	6	0.417	3	0.623	3	0.602	3
37 Rubidium	0.413	36	0.498	37	0.176	34	0.230	36
38 Strontium	0.087	56	0.199	55	0.128	32	0.242	47
39 Yttrium	0.132	69	0.075	73	-	-	-	-
40 Zirconium	0.063	65	0.081	65	-	-	-	-
41 Niobium	0.060	88	0.196	87	-	-	-	-
42 Molybdenum	0.104	89	0.145	86	-	-	0.095	84
43 Technetium	0.078	84	0.129	83	-	-	-	-
44 Ruthenium	0.073	113	0.143	111	-	-	-	-
45 Rhodium	0.162	68	0.192	66	-	-	-	-
46 Palladium	0.080	120	0.073	121	-	-	-	-
47 Silver	0.151	41	0.272	39	-	-	-	-
48 Cadmium	0.159	54	0.179	55	0.121	50	-	-
49 Indium	0.039	77	0.085	77	0.155	77	0.102	77
50 Tin	0.073	96	0.065	96	0.078	96	0.087	94
51 Antimony	0.060	92	0.170	92	0.083	91	0.135	92
52 Tellurium	0.070	80	0.162	79	0.123	77	0.122	79
53 Iodine	0.144	286	0.134	275	0.146	145	0.175	141
54 Xenon	0.620	8	0.584	4	0.472	2	0.793	6
55 Cesium	0.258	40	0.335	43	0.372	25	0.358	43
56 Barium	0.202	51	0.228	47	0.207	48	0.261	51
57 Lanthanum	0.253	47	-	-	-	-	-	-
71 Lutetium	0.050	60	-	-	-	-	-	-

72 Hafnium	0.071	42	0.223	43	-	-	-	-
73 Tantalum	0.074	59	0.211	59	-	-	-	-
74 Tungsten	0.141	57	0.137	57	-	-	-	-
75 Rhenium	0.068	108	0.156	105	-	-	-	-
76 Osmium	0.072	50	0.068	49	-	-	-	-
77 Iridium	0.169	71	0.094	66	-	-	-	-
78 Platinum	0.057	140	0.053	138	-	-	-	-
79 Gold	0.158	84	0.143	76	-	-	-	-
80 Mercury	0.143	64	0.110	64	0.135	63	0.139	64
81 Thallium	0.202	59	0.248	55	0.208	45	0.268	43
82 Lead	0.140	53	0.167	53	0.121	53	0.125	51
83 Bismuth	0.142	81	0.616	75	0.225	82	0.682	75
All elements, avoiding double counting	0.091	5164	0.123	5014	0.104	2493	0.130	2402
Same as above, but with double counting	0.089	9975	0.121	9684	0.104	4736	0.129	4573

Average Errors in Bond Angles (Degrees)

Element	PM6	No. in set	PM5	No. in set	PM3	No. in set	AM1	No. in set
3 Lithium	7.79	28	6.82	28	3.44	28	9.65	28
4 Beryllium	6.61	14	6.44	14	6.91	14	5.94	14
5 Boron	3.27	31	4.41	31	4.61	32	-	-
6 Carbon	2.50	134	2.79	134	2.75	131	2.25	131
7 Nitrogen	7.32	37	8.01	37	6.79	35	7.80	31
8 Oxygen	12.14	59	11.12	58	8.97	55	9.31	43
9 Fluorine	8.32	3	16.18	3	19.27	3	24.87	2
11 Sodium	21.00	4	2.87	4	3.42	4	5.26	4
12 Magnesium	8.44	24	7.28	24	13.80	24	6.90	24
13 Aluminium	4.05	20	5.26	20	7.22	19	4.24	19
14 Silicon	5.25	35	3.37	35	2.81	34	2.93	34
15 Phosphorus	3.24	35	4.40	35	5.88	36	4.96	36
16 Sulfur	5.23	46	5.64	45	5.39	41	5.05	41
17 Chlorine	3.65	5	19.47	5	10.09	5	14.87	5
19 Potassium	17.90	11	10.27	11	14.80	11	13.69	11
20 Calcium	14.99	16	11.35	16	17.03	16	18.50	16
21 Scandium	7.98	32	8.10	32	-	-	-	-
22 Titanium	7.86	39	12.13	39	-	-	-	-
23 Vanadium	7.46	44	7.24	44	-	-	-	-
24 Chromium	3.77	19	4.97	19	-	-	-	-
25 Manganese	6.02	26	7.04	26	-	-	-	-
26 Iron	11.21	30	4.91	30	-	-	-	-
27 Cobalt	10.68	29	7.29	29	-	-	-	-
28 Nickel	10.44	48	23.56	50	-	-	-	-
29 Copper	10.77	44	20.00	44	-	-	-	-
30 Zinc	10.92	27	14.41	27	8.30	27	14.31	27
31 Gallium	4.43	18	10.86	18	13.56	18	11.05	18

32 Germanium	4.58	52	5.37	52	8.82	53	5.73	53
33 Arsenic	6.29	36	6.52	36	6.55	36	5.08	36
34 Selenium	7.27	24	16.16	24	12.38	23	5.40	23
35 Bromine	12.64	4	20.03	4	19.10	3	3.22	3
37 Rubidium	9.69	11	10.20	11	21.38	11	6.37	11
38 Strontium	18.16	25	32.91	25	33.60	25	30.80	25
39 Yttrium	12.29	34	8.98	34	-	-	-	-
40 Zirconium	10.36	12	7.68	12	-	-	-	-
41 Niobium	6.54	23	8.49	23	-	-	-	-
42 Molybdenum	8.15	27	12.96	27	-	-	8.25	27
43 Technetium	4.96	22	10.41	22	-	-	-	-
44 Ruthenium	6.93	34	7.56	34	-	-	-	-
45 Rhodium	10.66	22	17.00	22	-	-	-	-
46 Palladium	9.19	46	5.80	46	-	-	-	-
47 Silver	23.36	9	24.67	9	-	-	-	-
48 Cadmium	15.23	10	13.52	10	20.03	10	-	-
49 Indium	4.47	17	7.21	17	5.27	17	5.02	17
50 Tin	3.06	34	4.09	34	3.74	34	8.53	34
51 Antimony	6.49	41	12.24	41	6.28	41	4.98	41
52 Tellurium	4.85	25	7.00	25	4.41	25	5.81	25
53 Iodine	8.33	1	12.55	1	20.57	1	4.58	1
55 Cesium	15.50	12	8.52	12	19.21	12	11.34	12
56 Barium	28.65	10	28.43	10	34.72	10	43.21	10
57 Lanthanum	9.25	14	-	-	-	-	-	-
71 Lutetium	7.08	26	-	-	-	-	-	-
72 Hafnium	5.64	10	6.28	10	-	-	-	-
73 Tantalum	9.88	15	8.56	15	-	-	-	-
74 Tungsten	10.90	9	7.58	9	-	-	-	-
75 Rhenium	7.39	32	8.28	32	-	-	-	-
76 Osmium	12.67	10	7.31	11	-	-	-	-

77 Iridium	7.86	18	17.26	18	-	-	-	-
78 Platinum	5.92	72	6.85	72	-	-	-	-
79 Gold	13.59	16	12.58	16	-	-	-	-
80 Mercury	20.20	15	20.99	15	18.37	15	21.81	15
81 Thallium	5.73	10	10.28	10	15.10	10	20.25	10
82 Lead	4.33	20	5.24	20	4.81	20	3.49	19
83 Bismuth	8.01	25	21.74	25	8.20	25	34.36	25
All elements	7.86	1681	9.55	1642	8.50	904	8.77	871

Average Errors in Dipole Moment (Debye)

Element	PM6	No. in set	PM5	No. in set	PM3	No. in set	AM1	No. in set
1 Hydrogen	0.62	266	0.80	265	0.64	222	0.50	204
3 Lithium	0.78	16	0.95	16	0.79	16	0.52	16
4 Beryllium	1.63	1	1.49	1	0.27	1	0.53	1
5 Boron	0.66	17	0.66	17	0.73	17	-	-
6 Carbon	0.51	219	0.62	218	0.41	176	0.42	165
7 Nitrogen	0.61	48	0.66	48	0.46	40	0.55	39
8 Oxygen	0.99	198	1.27	196	1.05	74	0.74	75
9 Fluorine	0.80	124	1.11	121	0.59	63	0.69	59
11 Sodium	1.34	6	0.80	6	1.97	6	1.26	6
13 Aluminium	0.33	1	1.50	1	1.76	1	0.53	1
14 Silicon	0.21	11	1.09	11	0.72	11	0.29	11
15 Phosphorus	0.83	14	0.79	14	0.37	10	0.87	10
16 Sulfur	0.62	28	1.01	28	0.74	21	0.70	21
17 Chlorine	0.99	103	1.27	100	0.77	47	0.84	43
19 Potassium	0.44	4	0.34	4	1.30	4	0.58	4
20 Calcium	0.73	4	1.12	4	1.23	4	0.33	4
21 Scandium	1.11	9	1.51	9	-	-	-	-
22 Titanium	1.02	8	0.90	8	-	-	-	-
23 Vanadium	0.82	8	1.70	8	-	-	-	-
24 Chromium	1.98	9	0.71	9	-	-	-	-
25 Manganese	1.06	11	0.64	11	-	-	-	-
26 Iron	1.61	14	1.14	14	-	-	-	-
27 Cobalt	1.04	6	1.55	6	-	-	-	-
28 Nickel	1.40	15	1.26	15	-	-	-	-
29 Copper	1.11	10	1.16	10	-	-	-	-
30 Zinc	0.21	4	0.18	4	0.16	4	0.16	4
31 Gallium	0.20	1	1.81	1	1.35	1	0.64	1

32 Germanium	0.63	23	0.63	23	0.55	23	0.59	23
33 Arsenic	0.37	6	0.99	6	0.35	6	0.37	6
34 Selenium	0.66	10	0.94	10	0.61	10	0.80	10
35 Bromine	0.90	88	1.34	87	1.01	37	0.50	39
37 Rubidium	1.84	6	2.43	6	1.65	6	0.44	6
38 Strontium	1.64	6	1.31	6	2.55	6	1.51	6
39 Yttrium	1.70	8	3.83	8	-	-	-	-
40 Zirconium	0.94	8	0.68	8	-	-	-	-
41 Niobium	0.91	10	1.96	10	-	-	-	-
42 Molybdenum	1.09	8	1.94	8	-	-	1.48	8
43 Technetium	1.74	13	0.87	13	-	-	-	-
44 Ruthenium	1.13	12	2.27	12	-	-	-	-
45 Rhodium	1.09	6	1.53	6	-	-	-	-
46 Palladium	0.97	8	0.81	8	-	-	-	-
47 Silver	1.98	9	5.31	9	-	-	-	-
48 Cadmium	0.42	2	2.22	2	0.67	2	-	-
49 Indium	0.47	3	0.78	3	0.75	3	1.36	3
50 Tin	0.28	13	0.41	13	0.88	13	0.81	13
51 Antimony	0.55	5	0.77	5	0.48	5	0.61	5
52 Tellurium	0.47	2	0.75	2	0.31	2	1.35	2
53 Iodine	1.03	77	1.54	77	1.48	28	1.22	30
55 Cesium	1.25	9	3.47	9	1.89	9	0.87	9
56 Barium	1.77	11	1.29	11	1.93	11	1.11	11
57 Lanthanum	1.23	8	-	-	-	-	-	-
72 Hafnium	0.63	6	1.28	6	-	-	-	-
73 Tantalum	0.97	5	0.58	5	-	-	-	-
74 Tungsten	0.92	14	1.77	14	-	-	-	-
75 Rhenium	0.76	13	0.98	13	-	-	-	-
76 Osmium	0.63	8	0.45	8	-	-	-	-
77 Iridium	0.96	8	1.18	8	-	-	-	-
78 Platinum	1.07	8	0.79	8	-	-	-	-

79 Gold	0.78	14	1.14	14	-	-	-	-
80 Mercury	0.63	9	0.77	9	0.63	9	0.67	9
81 Thallium	0.89	3	1.35	3	0.45	3	2.43	3
82 Lead	0.73	6	0.76	6	0.41	6	0.82	6
83 Bismuth	0.42	8	3.21	8	1.14	8	3.40	8
All elements, avoiding double counting	0.85	569	1.12	561	0.72	313	0.67	302
Same as above, but with double counting	0.83	1608	1.09	1589	0.73	905	0.65	861

Average Errors in Ionization Potential (eV)

Element	PM6	No. in set	PM5	No. in set	PM3	No. in set	AM1	No. in set
1 Hydrogen	0.43	226	0.40	226	0.60	226	0.52	217
3 Lithium	0.89	12	0.88	12	1.29	12	0.59	12
4 Beryllium	0.52	7	0.29	7	0.93	7	0.45	7
5 Boron	0.31	11	0.34	11	1.01	11	-	-
6 Carbon	0.41	230	0.39	230	0.54	230	0.54	227
7 Nitrogen	0.55	43	0.45	43	0.53	43	0.48	42
8 Oxygen	0.62	72	0.56	72	0.63	72	0.69	69
9 Fluorine	0.64	67	0.65	67	0.74	67	0.85	65
11 Sodium	0.34	5	0.34	5	1.43	5	0.51	4
12 Magnesium	0.97	4	1.05	4	1.10	4	1.41	4
13 Aluminium	0.62	3	0.29	3	0.40	3	0.69	3
14 Silicon	0.43	11	0.81	11	0.70	11	0.68	11
15 Phosphorus	0.49	13	0.47	13	0.64	13	0.56	13
16 Sulfur	0.52	46	0.51	46	0.48	46	0.62	46
17 Chlorine	0.48	62	0.58	62	0.57	60	0.61	57
19 Potassium	0.23	4	0.50	4	0.54	4	0.34	3
20 Calcium	0.74	1	1.24	1	0.52	1	0.41	1
21 Scandium	3.73	1	4.94	1	-	-	-	-
22 Titanium	0.09	1	1.72	1	-	-	-	-
30 Zinc	0.32	5	0.35	5	0.99	5	0.49	5
31 Gallium	0.52	3	0.73	3	1.28	3	1.16	3
32 Germanium	0.70	13	0.49	13	0.93	13	1.05	13
33 Arsenic	0.69	5	0.31	5	0.62	5	0.79	5
34 Selenium	0.38	10	0.29	10	0.47	10	1.22	10
35 Bromine	0.28	33	0.39	33	1.20	33	0.49	32
37 Rubidium	0.18	3	0.39	3	0.93	3	0.22	3
38 Strontium	0.63	1	0.38	1	0.14	1	0.26	1

48 Cadmium	0.33	5	0.46	5	0.39	5	-	-
49 Indium	0.63	2	0.86	2	2.06	2	0.83	2
50 Tin	0.70	14	0.48	14	1.22	14	0.44	14
51 Antimony	0.44	5	0.90	5	1.16	5	0.54	5
52 Tellurium	0.43	3	0.20	3	0.25	3	0.70	3
53 Iodine	0.47	29	0.46	29	0.48	29	0.89	29
55 Cesium	0.58	4	0.71	4	1.37	4	1.11	4
56 Barium	0.08	1	0.97	1	0.08	1	0.75	1
80 Mercury	0.51	12	0.43	12	0.74	12	0.49	12
81 Thallium	0.30	3	0.46	3	0.80	3	0.53	3
82 Lead	0.56	13	0.47	13	0.93	13	0.65	13
83 Bismuth	0.98	5	1.28	5	0.72	5	1.66	5
All elements, avoiding double counting	0.50	385	0.50	385	0.68	383	0.63	367
Same as above, but with double counting	0.48	988	0.48	988	0.66	984	0.61	944

PM6-DH2 - Accuracy

PM6-DH2 gives much improved interaction energies compared to PM6. Whereas in PM6, the average unsigned error for interaction energies for the complexes in the S22 set is 3.27 Kcal/mol, for PM6-DH2, the equivalent average unsigned error is 0.37 Kcal/mol. For details of PM6-DH2, see: M. Korth, M. Pitonak, J. Rezac, and P. Hobza, "A Transferable H-bonding Correction For Semiempirical Quantum-Chemical Methods", J. Chem. Theory Comp. **2010**, 6, 344-352

Some representative complexes are shown in the Table.

Table of Representative Complexes

Complex	CCSD(T)/CBS	PM6	Error	PM6-DH2	Error
A-T WC	-16.37	-9.06	7.31	-16.45	-0.08
G-C WC	-31.40	-21.35	10.05	-30.75	+0.65
A-T Stack	-12.23	-4.94	7.39	-11.67	+0.54
G-C Stack	-19.02	-13.07	5.93	-19.66	-0.64
Water dimer	-5.02	-3.94	1.08	-4.90	+0.12
Formic acid dimer	-18.61	-11.14	7.47	-18.65	-0.04
Formamide dimer	-15.96	-12.55	3.41	-15.87	+0.10
Lys ⁺ - Glu-	-108.40	-98.83	9.57	-104.63	+3.77
Arg ⁺ - Asp-	-110.80	-100.83	9.97	-109.80	+1.01

Calculations were done using the high-quality geometries from the [S22 database](#), see: Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, 8 (17), 1985-1993

Reference interaction energies (CCSD(T) from the [CCSD(T)/complete basis set (CBS) limit] of the dispersion-bonded complexes.

Worked example:

A. PM6-DH2 Formic acid dimer $DH_f = -188.158$

B. PM6-DH2 two separated formic acid molecules $DH_f = -169.504$

Interaction energy = B - A = -18.654

Note: Heats of formation predicted by PM6-DH2 should not be compared to reference heats of formation. This is because PM6-DH2 is designed for predicting energies of interaction. If heats of formation are wanted, use PM6, for which the predicted DH_f can be compared to experimental values.

See also PM6-DH+ [accuracy](#).

Fault in PM6 in Prediction of Oxalic Acid geometry

PM6 predicts the twisted conformation of oxalic acid (HOOC-COOH) to be 2.9 kcal/mol lower in energy than the planar form, and B3LYP predicts the twisted form to be 1.4 kcal/mol higher than the planar form. There is no reason to doubt the B3LYP result, therefore PM6 has a definite error.

PM6 predicts [crystalline oxalic acid](#) to be composed of planar oxalic acid molecules, in agreement with X-Ray results.

Fault in PM6 in Prediction of Oxalic Acid Heat of Reaction

PM6 predicts the heat of reaction of oxalic acid (HOOC-COOH) decomposing to CO₂ + C(OH)₂ to be in error by 29.6 kcal/mol. There is no reason to doubt the B88PW91 result, therefore PM6 has a definite error.

Reaction				
	Oxalic acid	→ CO ₂	+ C(OH) ₂	DH _r
Ref.	-175.0	-94.1	-41.4	39.5
PM6	-156.3	-84.8	-61.6	9.9

The main source of the error is likely to be in C(OH)₂. If that is the case, then it is also likely that divalent carbon is in error.

Reference and PM6 heats of formation (kcal/mol)

Molecule	Ref.	PM6	Diff
Oxalic acid	-175.0	-156.3	18.7
CO ₂	-94.1	-84.8	9.3
C(OH) ₂	-41.4	-61.6	-20.2

Fault in PM6 in Prediction that $\text{R-COO-R} + \text{R-NH}_2$ forms a stable system

PM6 predicts that an organic acid and an amine should react to form a stable C-N bond. The fault does not appear with simple compounds, but can be shown in a [more complicated system](#).

In this fault, R-NH_2 reacts with R-COO-R to give $\text{R-NH}_2\text{-C(R)OOR}$. The NH_2 is formally positive, and the C=O becomes $\text{C-O}(-)$.

Fault in PM6 in Prediction of Br-N non-bonding interaction

PM6 predicts that HCN and CH₃Br should combine to form a linear HCN-BrCH₃ complex. B3LYP and other high-level methods predict the N - Br distance to be greater than 3.5 Angstroms.

Fault in PM6 in Prediction of S-N non-bonding interaction

PM6 predicts that $\text{Me}_2\text{C}=\text{S}$ and $\text{N}(\text{Me})_3$ should combine to form a $\text{Me}_2\text{C}=\text{S}-\text{N}(\text{Me})_3$ complex, with the C-S and S-N distances approximately the same. B3LYP and other high-level methods predict the N-S distance to be greater than 3.5 Angstroms.

Fault in PM6 in Prediction of S-O non-bonding interaction

PM6 predicts that Me_2O and $\text{S}=\text{CH}_2$ should combine to form a $\text{Me}_2\text{O}-\text{S}=\text{CH}_2$ complex, with the C-S and S-O distances 1.6 and 2.1 Angstroms, respectively. B3LYP and other high-level methods predict the O-S distance to be greater than 4.5 Angstroms.

Fault in PM6 in Prediction of S-Cl(-) non-bonding interaction

PM6 predicts that Cl(-) and S=CH₂ should combine to form a Cl(-)-S=CH₂ complex, with the C-S and S-Cl(-) distances 1.6 and 2.3 Angstroms, respectively. B3LYP and other high-level methods predict the Cl-S distance to be much larger.

By default, UHF CH₄ plus F generates an excited state

In general, SCF calculation involving halogen atoms that are far from neutral organic molecules give rise to electronic excited states. An example would be a fluorine atom about 10 Angstroms from a methane molecule. In this system (see data set below), the fluorine atom should have one unpaired extra alpha electron, and methane should be neutral. Instead, the SCF results in the fluorine atom having one unpaired beta electron and methane has two unpaired alpha electrons. What went wrong? The problem was traced to the initial guess for the density matrix. In the first few iterations of the SCF calculation, fluorine has a unit negative charge. The odd alpha electron goes on to the methane. As the SCF forms, a second alpha electron goes on to the methane; this results in the excited state forming.

To correct errors of this type, add [PULAY](#) to the keyword list.

This error occurs so rarely that a bug-fix has not been made.

Data set that shows this fault:

```
1scf
CH4 plus F

F      -0.13005907 +1      0.05852702 +1      0.00195838 +1      0.0000
C      11.69496462 +1     -5.34383705 +1     -1.18161572 +1     -0.4405
H      10.09784740 +1     -4.44034777 +1      1.15461967 +1     -0.0002
H      12.10011696 +1     -6.26901783 +1     -0.86283618 +1      0.1468
H      10.80424362 +1     -5.32647020 +1     -1.75399119 +1      0.1468
H      12.17920911 +1     -4.43724484 +1     -0.92583745 +1      0.1470
```

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 kcal/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 kcal/mol/Ångstrom is less than TOLERG multiplied by the square root of the number of coordinates to be optimized.

Exceptions:

- If [GNORM](#) 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 [PRECISE](#) is specified, TOL2=0.01
- If [GNORM](#) 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.

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 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 the Figure.

Figure: Example of Spurious Stationary Point

```

Line 1:  GRADIENTS 1SCF DEBUG DCART
Line 2:  EXAMPLE OF INTERNAL COORDINATE
DERIVATIVES ZERO
Line 3:  AND CARTESIAN DERIVATIVES
LARGE
Line 4:  O
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

```

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, the

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.

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 DH_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 DH_f errors can result.

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.

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.

SCF Test in MOPAC

An exact definition of the SCF test used in MOPAC is not easy to write because of the large number of qualifying criteria used, so in the following description some details have been left out. For most systems, the SCF criterion can be described as follows:

First, two quantities need to be defined

SCFCRT: This is the primary quantity used in deciding if a self-consistent field exists. For most operations ([1SCF](#), geometry optimization, reaction path, etc.) the value of SCFCRT is 10^{-4} , but for gradient minimization and [FORCE](#) calculations a higher precision, 10^{-7} , is needed. For other calculations, the precision can be increased to 10^{-6} using [PRECISE](#). The value of SCFCRT can also be set using [SCFCRT](#) or modified using [RELSCF](#). The value of SCFCRT is a constant in any calculation.

SELCON: For all "difficult" calculations ([FORCE](#), gradient minimization, RHF open shell, etc.) SELCON is set to SCFCRT. For normal ground state calculations SELCON is set to 100 times SCFCRT and possibly higher depending on the gradient. The value of SELCON can vary from one SCF calculation to the next.

For an SCF field to exist, the following test must be passed twice consecutively: The change in DH_f on going from one iteration to the next must be less than SELCON kcal mol⁻¹, and the largest change in density matrix element values must be less than $0.05(SELCON)^{1/2}$.

PL

A tool for monitoring the behavior of SCF calculations is useful when they take too long, or even fail altogether. When keyword `PL` is present, the energy of the system and the rate of change of the electronic structure can be monitored iteration to iteration. This keyword is useful particularly when trying out various combinations of convergers such as [PULAY](#), [KING](#), [SHIFT](#), and [DAMP](#) to find the method that works best.

The rate of change in the electronic structure is given by the quantity PLS in the output. For example, in the line:

```
ITERATION      7 PLS=  0.168E-01  0.395E-06 ENERGY      -34.585270 DELTAE      -2.2653063
```

the alpha wavefunction changed by 0.0168 between iteration 6 and iteration 7. At the same time, the beta wavefunction changed by 0.000000395. The change in energy over these iterations is -2.265 kcal/mol.

If the calculation uses a restricted Hartree-Fock method, then the line would look like this:

```
ITERATION      7 PLS=  0.900E-02  0.000E+00 ENERGY      46.712559 DELTAE      -0.3203785
```

Now the change in total wavefunction between iteration 6 and 7 would be 0.009. The second number, here 0.000E+00, is not used and should be ignored.

The precise meaning of PLS depends on whether the calculation uses [MOZYME](#) or the conventional, default, method.

Definition of PLS for conventional SCF calculations

In conventional methods, PLS is the largest change in any density matrix element on two successive iterations in the SCF calculation. At self-consistency, this change drops to zero. For UHF calculations, both the alpha and beta density matrices are used, therefore two numbers are printed. The total density matrix is used in RHF calculations, and consequently the second number is not used.

Definition of PLS for MOZYME calculations

MOZYME calculations do not use a normal density matrix so the definition used above cannot be used. Instead, the largest Fock matrix element connecting an occupied LMO with any virtual LMO is used. This can be written as:

$$\text{PLS} = |\langle y^{\text{occ}} | F | y^{\text{vir}} \rangle|$$

Although this definition is fundamentally different from that in conventional work, in that data from only one iteration is used and the units are now electron volts and not electrons, the meaning is the same - at self-consistency all Fock matrix elements connecting occupied and virtual LMOs are zero. Therefore, from a user's perspective, the significance of PLS in both conventional and MOZYME calculations is the same: it is a measure of how far the system is from self-consistency.

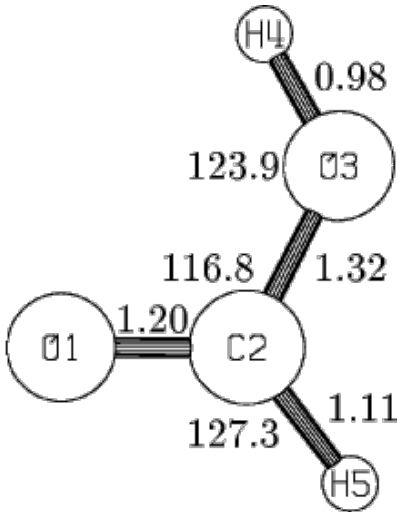
Examples of Coordinate Definitions: Formic acid

A MOPAC data set normally consists of one line of keywords, two lines of user-defined text, then the Z-matrix. In this example, normal internal coordinates are used for all atoms. After the Z-matrix there should be either a blank line or a line of zero's (as shown here).

AM1
Formic acid
Example of normal geometry definition

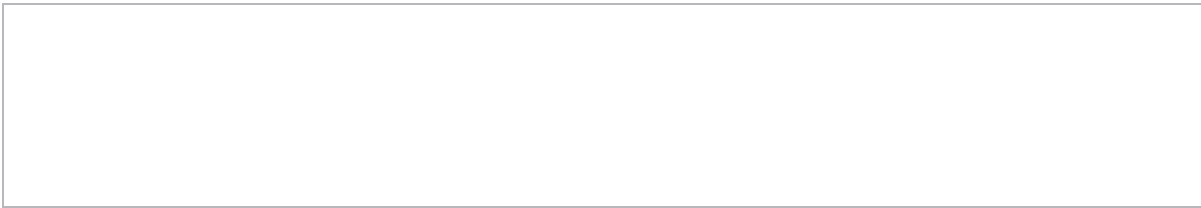
O	0.0	0	0.0	0	0.0	0	0	0	0
C	1.20	1	0.0	0	0.0	0	0	0	0
O	1.32	1	116.8	1	0.0	0	2	1	0
H	0.98	1	123.9	1	0.0	0	3	2	1
H	1.11	1	127.3	1	180.0	0	2	1	3
0	0.00	0	0.0	0	0.0	0	0	0	0

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.



Description of a simple MOPAC2007 Output File

(move the cursor over the file to see what the parts are)

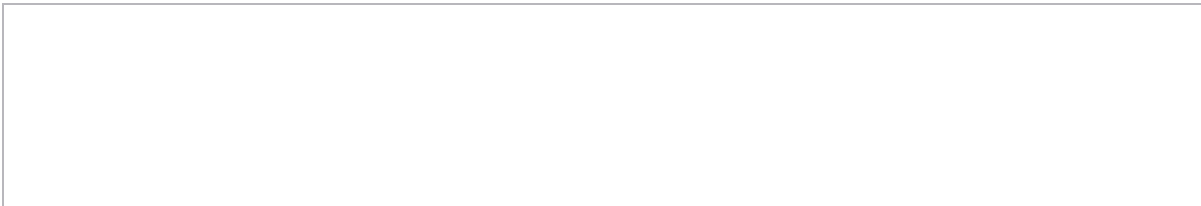


```
*****
** Site#: 1321 For non-commercial use only Version 7.060W **
*****
** Cite this work as: MOPAC2007, James J. P. Stewart, Stewart Computational **
** Chemistry, Version 7.060W web: HTTP://OpenMOPAC.net Days remaining: 366 **
*****
**
** MOPAC2007 **
**
*****
```

AM1 CALCULATION RESULTS

```
*****
* CALCULATION DONE: Thu Mar 1 16:17:07 2007 *
* AM1 - The AM1 Hamiltonian to be used
* T= - A TIME OF 172800.0 SECONDS REQUESTED
* DUMP=N - RESTART FILE WRITTEN EVERY 7200.000 SECONDS
*****
```

AM1
Formic acid
Example of normal geometry definition



ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS)	BOND ANGLE (DEGREES)	TWIST ANGLE (DEGREES)			
		NA:I	NB:NA:I	NC:NB:NA:I	NA	NB	NC
1	O	0.0000000	0.000000	0.000000			
2	C	1.2000000 *	0.000000	0.000000			
3	O	1.3200000 *	116.800000 *	0.000000	2	1	0
4	H	0.9800000 *	123.900000 *	0.000000	3	2	1
5	H	1.1100000 *	127.300000 *	180.000000	2	1	3

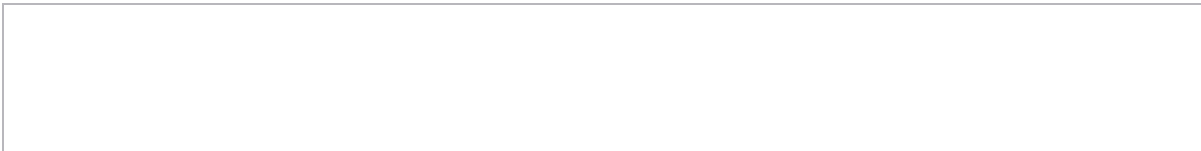
CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
1	O	0.0000	0.0000	0.0000
2	C	1.2000	0.0000	0.0000
3	O	1.7952	1.1782	0.0000
4	H	1.3156	2.0328	0.0000
5	H	1.8726	-0.8830	0.0000

H: (AM1): M.J.S. DEWAR ET AL, J. AM. CHEM. SOC. 107 3902-3909 (1985)
C: (AM1): M.J.S. DEWAR ET AL, J. AM. CHEM. SOC. 107 3902-3909 (1985)
O: (AM1): M.J.S. DEWAR ET AL, J. AM. CHEM. SOC. 107 3902-3909 (1985)

Empirical Formula: C H2 O2
MOLECULAR POINT GROUP : Cs

RHF CALCULATION, NO. OF DOUBLY OCCUPIED LEVELS = 9



DIAGONAL MATRIX USED AS START HESSIAN

CYCLE:	1	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	118.472	HEAT:	-91.15276
CYCLE:	2	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	64.573	HEAT:	-93.61392
CYCLE:	3	TIME:	0.016	TIME LEFT:	2.00D	GRAD.:	23.477	HEAT:	-95.72593
CYCLE:	4	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	16.802	HEAT:	-96.77744
CYCLE:	5	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	14.051	HEAT:	-97.31996
CYCLE:	6	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	4.873	HEAT:	-97.39856
CYCLE:	7	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	1.076	HEAT:	-97.41413
CYCLE:	8	TIME:	0.000	TIME LEFT:	2.00D	GRAD.:	0.467	HEAT:	-97.41479

RMS GRADIENT = 0.46676 IS LESS THAN CUTOFF = 1.00000

AM1
Formic acid
Example of normal geometry definition

GEOMETRY OPTIMISED USING EIGENVECTOR FOLLOWING (EF).
SCF FIELD WAS ACHIEVED

AM1 CALCULATION
MOPAC2007 (Version: 7.060W)
Thu Mar 1 16:17:07 2007
No. of days left = 366

FINAL HEAT OF FORMATION = -97.41479 KCAL = -407.58348 KJ

TOTAL ENERGY	=	-797.12666	EV	
ELECTRONIC ENERGY	=	-1697.04783	EV	POINT GROUP: Cs
CORE-CORE REPULSION	=	899.92117	EV	
COSMO AREA	=	71.42	SQUARE ANGSTROMS	
COSMO VOLUME	=	52.65	CUBIC ANGSTROMS	
IONIZATION POTENTIAL	=	11.81903		
NO. OF FILLED LEVELS	=	9		
MOLECULAR WEIGHT	=	46.026		

MOLECULAR DIMENSIONS (Angstroms)

Atom		Atom		Distance
H	5	H	4	2.86513
O	3	O	1	2.10425
C	2	H	4	0.00000

SCF CALCULATIONS	=	10
COMPUTATION TIME	=	0.016 SECONDS

ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS) NA:I	BOND ANGLE (DEGREES) NB:NA:I	TWIST ANGLE (DEGREES) NC:NB:NA:I	NA	NB	NC
1	O	0.0000000	0.000000	0.000000			
2	C	1.2304588 *	0.000000	0.000000			
3	O	1.3568332 *	117.681304 *	0.000000	2	1	0
4	H	0.9712524 *	110.582833 *	0.000000	3	2	1
5	H	1.1033197 *	129.967961 *	180.000000	2	1	3

Empirical Formula: C H2 O2

MOLECULAR POINT GROUP : Cs

EIGENVALUES

-40.62125 -36.77351 -24.87659 -18.96303 -18.12673 -16.19373 -14.67936 -12.58306
-11.81903 0.95713 2.28141 3.84205 5.09596 6.44101

NET ATOMIC CHARGES AND DIPOLE CONTRIBUTIONS

ATOM NO.	TYPE	CHARGE	No. of ELECS.	s-Pop	p-Pop
1	O	-0.357283	6.3573	1.91393	4.44335
2	C	0.260327	3.7397	1.26671	2.47296
3	O	-0.324044	6.3240	1.86042	4.46363
4	H	0.241808	0.7582	0.75819	
5	H	0.179193	0.8208	0.82081	



DIPOLE	X	Y	Z	TOTAL
POINT-CHG.	1.722	-0.361	0.000	1.759
HYBRID	-0.243	0.298	0.000	0.384
SUM	1.479	-0.063	0.000	1.480

CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
1	O	0.0000	0.0000	0.0000
2	C	1.2305	0.0000	0.0000
3	O	1.8608	1.2015	0.0000
4	H	1.2142	1.9263	0.0000
5	H	1.9392	-0.8456	0.0000

ATOMIC ORBITAL ELECTRON POPULATIONS

1.91393	1.13249	1.87986	1.43100	1.26671	0.88857	0.86899	0.71541
1.86042	1.42153	1.18851	1.85359	0.75819	0.82081		

TOTAL CPU TIME: 0.03 SECONDS

== MOPAC DONE ==

Examples of a single-point calculation of Formic acid

By default, MOPAC will attempt to optimize the geometry. If this is not wanted, then two methods are provided for doing a single SCF calculation and then stopping.

The following data sets illustrate these two methods.

```
AM1 1SCF
Formic acid
Example of a single-point calculation
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
```

```
AM1
Formic acid
Example of a single-point calculation
O
C      1.20 0
O      1.32 0 116.8 0      0.0 0      2 1
H      0.98 0 123.9 0      0.0 0      3 2 1
H      1.11 0 127.3 0 180.0 0      2 1 3
O      0.00 0      0.0 0      0.0 0      0 0 0
```

The recommended way to do a single SCF calculation is by explicitly using the keyword [1SCF](#). This will cause MOPAC to do one SCF calculation and then stop. No derivatives will be calculated; if you want them, add [GRADIENTS](#).

An alternative is to set all the optimization flags to zero. This is not the preferred option because the flags might all be set to zero when the data set was created, and if you don't notice that, the results might not be what you expected. By using 1SCF, the user has to make a conscious decision to explicitly request that a single SCF be done.

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.

General strategy (works for almost all reactions, but is slower than the special cases below, when they are applicable).

The [SADDLE](#) calculation will work for most systems. This uses the following sequence:

- Optimize the reactant geometry.
- Using the same atoms in the same sequence, optimize the product geometry.
- Do a geometry optimization of the product geometry using the reactant geometry as reference, this uses [GEO_REF](#). This gives rise to a geometry on the product side of the transition state that is between the true product geometry and the transition state, that is, it is in the slope that leads up to the transition state.
- Do a geometry optimization of the reactant geometry using the new product geometry as reference, again using [GEO_REF](#). This gives rise to a geometry on the reactant side nearer to the transition state than the true reactant geometry.
- Using these two geometries, one on the reactant side of the transition state and one on the product side, run the [SADDLE](#) calculation. For most calculations, this will terminate in a point near to the transition state.
- 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 [BAR=n.nn](#), e.g., [BAR=0.03](#), and re-run the calculation. If CPU time is not important, run the original data set with [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.

For worked examples, see:

[Ester hydrolysis Saddle calculation](#)

[1,5 Hexadiene Cope re-arrangement](#)

[Acetone ene-ol - keto tautomerization](#)

[Retro-ene elimination of isochoricismic acid to give pyruvic plus salicylic acids](#)

[Prototype of aspartate protease catalyzed hydrolysis of peptide](#)

Transition states for some special types of reaction can be located more easily using different computational techniques. The more important of these are:

For narcissistic reactions (reactions in which the reactants and products are the same, e.g. the inversion of ammonia.

- Use geometric constraints, e.g. [SYMMETRY](#), to lock the geometry in the symmetry of the potential transition state.
- Minimize the DH_f .
- Verify that the system is a transition state. If it has more than one negative force constant, use another method.

(For the Cope rearrangement of 1,5 hexadiene, even though the transition state is symmetric, it is faster to use the general [SADDLE](#) method described above.)

For a bond making-bond breaking reaction (e.g., an S_N^2 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 DH_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 a worked example, see [SN2](#)

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 [path option](#) 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 a worked example involving the rotation of the C2-C3 bond in butane, see [Barriers](#).

For bond making or bond breaking reactions involving exactly one change in covalent bonding

- Identify the reaction coordinate (the bond that makes or breaks)
- Use a [path calculation](#) to drive the reaction.
- The geometry of the highest point on the reaction path should then be used to start a `TS` calculation.
- For a worked example ($\text{CH}_3\text{-Br} + \text{F}(-) \Rightarrow \text{CH}_3\text{-F} + \text{Br}(-)$) see [SN2_Path](#)

For bond making or bond breaking reactions involving exactly one bond being made and one bond being broken

The [GRID](#) calculation can be used. In this, two coordinates that are important in the reaction are chosen, and the Potential Energy Surface for the optimized geometry is generated for all reasonable values of these coordinates. Visual inspection of the resulting map can usually provide a guide to the transition state. Once the approximate transition state is determined, it can be further refined using [TS](#).

Introduction to MOZYME

[Introduction](#)

[The model must be realistic](#)

[Identifying charged sites](#)

[Identifying ionized residues](#)

[Ramachandran angles](#)

Introduction

MOZYME is a technique that allows very large systems, in particular proteins and sections of DNA, to be modeled in a reasonable time. Although in principle, it is similar to conventional methods in that the underlying computational model is PM6 or AM1, etc., in practice there are several differences. Most of these differences are a result of the large size of the systems. In this page, some of the considerations will be discussed.

The model must be realistic

Most important of all, is the continuing requirement that in quantum chemical calculations each data set must represent a realistic system. In Molecular Mechanics, hydrogen atoms are often ignored or combined into the heavy atom that they are bonded to, to give the "combined atom" model. In quantum mechanical calculations, every hydrogen atom must be explicitly defined.

Some residues are usually described as being ionized, thus Arg is usually represented as $[\text{Arg}]^+$, and Asp as $[\text{Asp}]^-$. But if there is a large difference in numbers of anionic and cationic residues, the net charge on the system can be unrealistically large. Neutralizing individual residues requires making decisions that are hard to justify. To avoid having to make subjective decisions, a well-defined starting structure would be to use the completely neutral system. This is somewhat artificial in that there are almost certainly ions present in the system, but usually, during geometry optimization, protons will migrate from carboxylic acid groups to form carboxylate anions, and some Arg and Lys will become cations.

An example of a faulty starting geometry would be DNA in which every phosphate group was singly ionized. If the geometry of such a system was calculated, the results would be nonsense - the very large net negative charge would cause the DNA to spontaneously untwist, ruining the structure. Only when the phosphate groups are neutralized can the characteristic twist be reproduced.

Identifying charged sites

All ionized sites can be identified by keyword [CHARGES](#). If this keyword is used, the charged sites and net charge will be printed, and the calculation stopped. This is a quick and reliable way of finding ionized sites. If an attempt is made to run a system with the wrong charge, either because a mistake was made in the structure, or because the assumed net charge was incorrect, the error will be detected and the job stopped. However, if [GEO-OK](#) is present, the supplied charge will be replaced with the calculated charge, and the job will be run using the calculated charge.

Identifying ionized residues

When [RESIDUES](#) is present, the amino-acid residues in the system are identified, and in the output the net charge on each residue is printed. If the charge on a residue is larger than 0.5, the residue will be labeled "CATION", if it is less than -0.5, it will be labeled "ANION".

Ramachandran angles

When [RESIDUES](#) is present, the three Ramachandran angles, ϕ , ψ , and ω , are printed for each residue.

Simple MOZYME example

MOZYME is a method for speeding up the process of solving the SCF equations. It does not affect any geometric processes. To use the MOZYME function, add keyword MOZYME to the data set. As an example, look at the data set for the geometry optimization of formic acid:

Data-set:

```
MOZYME
Optimize the geometry of formic acid

O      0.0  0      0.0  0      0.0  0      0  0  0
C      1.20  1      0.0  0      0.0  0      0  0  0
O      1.32  1  116.8  1      0.0  0      2  1  0
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
```

The difference between the output from a conventional matrix-algebra method and the Localized Molecular Orbital method can be easily seen by comparing the output of MOPAC calculations in which the keyword MOZYME is present and when it is not present. The main difference is the absence of ionization potentials in the MOZYME calculation: these are functions of eigenvectors and MOZYME does not generate eigenvectors, only molecular orbitals.

MOZYME example: Print Eigenvectors

The [MOZYME](#) method uses localized molecular orbitals. When keyword [VECTORS](#) is used, MOZYME will print out the LMOs in a highly compressed form. If eigenvectors are required, the occupied and virtual LMOs must be converted into eigenvectors. The keywords for this are [EIGEN](#) and [VECTORS](#) or [ALLVECS](#), if all the eigenvectors are needed. Be careful if all the eigenvectors are requested - the output can be large.

Data-set:

```

MOZYME Eigen Vectors
Optimize the geometry of formic acid

O      0.0  0      0.0  0      0.0  0      0  0  0
C      1.20  1      0.0  0      0.0  0      0  0  0
O      1.32  1    116.8  1      0.0  0      2  1  0
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

```

To see the difference between the MOZYME eigenvectors and conventional, run the data set with and without MOZYME and look at the eigenvectors in the output file. In MOZYME, the occupied set of eigenvectors and eigenvalues are printed, then the virtual set. This reflects the different methods of generating them.

Because the calculation of the eigenvectors involves matrix algebra, the computational effort rises as the cube of the size of the system, and can rapidly become very great. For this reason, do not request eigen vectors unless absolutely necessary.

MOZYME: Identifying Charged Sites

Finding all the charged sites in a large molecule is tedious, and if done by hand, is prone to errors. If the charge calculated by MOZYME does not match the charge supplied in the data set (either zero, if charge=*n* is absent, or *n*, if charge=*n* is supplied), then the calculated Lewis structure will be printed along with a list of the charged sites found when the Lewis structure was generated. In some cases a charged atom will be different to that expected. Thus in a charged arginine residue, the ionized site might be expected to be a nitrogen, but another valid Lewis structure puts the charge on the carbon atom that is bonded to the three nitrogen atoms.

Checking all the ionized sites in a protein is important before any modeling work is started.

Sometimes [MOZYME](#) generates an incorrect Lewis structure. For example, if the system C₆H₄O₂ is supplied, this being a hexagon of carbon atoms with an oxygen atom at positions 1 and 4, hydrogen atoms being at the other positions, MOZYME will generate the Lewis structure for *para*-benzoquinone. But the user might have wanted to calculate the di-anion of *para*-hydroquinone. Since this has two negative charges, the keyword [CHARGE=-2](#) would have been used. If an incorrect charge, and by implication, an incorrect Lewis structure is generated by MOZYME, the fault can be corrected using two methods.

Use of atom labels

A unit charge can be assigned to individual atoms using [atom labels](#). In the case of the di-anion of *para*-hydroquinone, both oxygen atoms could be given the label "(-)" to indicate that they are anions. This would result in the correct Lewis structure for the di-anion of *para*-hydroquinone being generated.

Another less critical but still informative example is the anion of N-methyl *para*-hydroxy pyridine. Its default Lewis structure results in a neutral nitrogen and neutral oxygen. If the more conventional Zwitterionic form is wanted, then the nitrogen atom would need to be labelled "(+)" and the oxygen atom "(-)". What would happen if only one of these labels was used? If only oxygen was labeled, then oxygen would have a negative charge, and the carbon atom it's bonded to would be assigned a positive charge. If only nitrogen was labeled, then the Lewis structure would be generated as usual, but at the end, the odd electron on nitrogen would be converted into a lone pair, and the assigned positive charge would be cancelled out.

User-modified topology

Individual bonds can be added or removed (using [CVB](#)) in order to create a topology that would give rise to the desired Lewis structure. This option will most commonly be used when metal complexes are present. Thus if a zinc atom has four ligands, it will be given a formal charge of -2 and an oxidation state of +2. Although this is completely valid as a Lewis structure, some users might be concerned at this unusual formal charge. By selectively removing Zn-ligand bonds, the more conventional Lewis structure can be generated.

In the unlikely situation that the desired Lewis structure cannot be generated, please send details to SCC.

MOZYME example: PDBOUT

A PDB file can be generated from a normal MOPAC Z-matrix by using [PDBOUT](#). Any comments in the MOPAC data set (lines starting with an asterisk, '*') are put at the start of the pdb file. If 0SCF is also present, a file with the suffix ".pdb" will be made, otherwise the PDB file will be embedded in the .arc and .out files.

If a residue is not one of the standard 20 residues, it will be labeled "UNK", for unknown. If it is a modified standard residue, for example the retinal group in rhodopsin is a modified lysine residue, then keyword [XENO](#) can be used to allow it to be identified.

Data-set:

```
*REMARK This text will be put into the .pdb file
*REMARK
*REMARK
  0scf pdbout

N      0.00000000 +1      0.00000000 +1      0.00000000 +1
C      1.45099200 +1      0.00000000 +1      0.00000000 +1
C      2.07379016 +1      1.3847514 +1      0.00000000 +1
O      3.16479845 +1      1.5510512 +1      0.5346009 +1
C      1.98010532 +1     -0.7816083 +1     -1.2061384 +1
S      2.91021593 +1     -2.2138976 +1     -0.5916694 +1
H      1.78639414 +1     -0.4791530 +1      0.9215042 +1
H      2.66987764 +1     -0.1661407 +1     -1.7918980 +1
H      1.16732537 +1     -1.1139738 +1     -1.8582339 +1
H      3.33442776 +1     -2.8105177 +1     -1.7526284 +1
H     -0.45944146 +1     -0.5439661 +1     -0.7191795 +1
H     -0.45944146 +1     -0.5439661 +1      0.7191795 +1
N      1.41779186 +1      2.3818640 +1     -0.5930817 +1
C      1.96005201 +1      3.7395708 +1     -0.5873182 +1
C      0.87144992 +1      4.6925379 +1     -0.0859199 +1
O     -0.30894710 +1      4.3852105 +1     -0.2277412 +1
C      2.58984680 +1      4.0757744 +1     -1.9613244 +1
C      3.95504011 +1      3.4174209 +1     -2.2109945 +1
C      5.13263384 +1      4.0760823 +1     -1.8023387 +1
C      6.39390379 +1      3.5338821 +1     -2.0746443 +1
C      6.51251534 +1      2.3040175 +1     -2.7409980 +1
C      5.35236761 +1      1.6339502 +1     -3.1585724 +1
C      4.08327785 +1      2.1864831 +1     -2.9055825 +1
H      2.75184646 +1      3.8479695 +1      0.1426395 +1
H      2.74129210 +1      5.1566073 +1     -1.9998683 +1
H      1.90797747 +1      3.8313152 +1     -2.7790034 +1
H      5.07815939 +1      5.0182815 +1     -1.2765870 +1
H      7.28256125 +1      4.0711027 +1     -1.8000485 +1
H      3.20599873 +1      1.6658896 +1     -3.2578962 +1
H      5.45626342 +1      0.6951745 +1     -3.6940611 +1
H      7.49097179 +1      1.8927662 +1     -2.9377954 +1
H      0.46717507 +1      2.2471723 +1     -0.9238687 +1
N      1.23320241 +1      5.8266104 +1      0.5208984 +1
C      0.26492485 +1      6.8517935 +1      0.8919516 +1
C      0.31272594 +1      7.9119232 +1     -0.2244470 +1
O      1.39993820 +1      8.1980713 +1     -0.7103118 +1
C      0.56056668 +1      7.4759401 +1      2.2689288 +1
C      1.27996141 +1      6.5571211 +1      3.2666683 +1
C      1.60234572 +1      7.3416902 +1      4.5312735 +1
O      2.41760942 +1      8.2681242 +1      4.3634539 +1
O      1.02819708 +1      7.0347336 +1      5.6046982 +1
H     -0.71623918 +1      6.4059436 +1      1.0059751 +1
H     -0.39373680 +1      7.7519156 +1      2.7226369 +1
H      1.15143293 +1      8.3863594 +1      2.1332694 +1
H      2.23164513 +1      6.2323719 +1      2.8483475 +1
H      0.67739509 +1      5.6796776 +1      3.4974943 +1
H      2.66377974 +1      8.7825627 +1      5.0596961 +1
H      2.20037691 +1      6.1233889 +1      0.4986194 +1
H     -0.81037727 +1      8.4883333 +1     -0.6911228 +1
```

ff

MOZYME example: Automatically correcting faulty charge

When [GEO-OK](#) is present in a [MOZYME](#) calculation, any errors in net charge are automatically corrected. In this example, CH₃-NH₃ is obviously a cation (methyl ammonium cation), but in the data set shown here, the charge is incorrectly given as -1. Because GEO-OK is present, when the job is run the error is detected and corrected, and the output and ARC files will be correct.

Data-set:

```
mozyme CHARGE=-1 GEO-OK
Methylammonium ion, charge is wrong, but since GEO-OK is present it will be corrected

C      0.00000000 +0      0.00000000 +0      0.00000000 +0
N      1.50605484 +1      0.00000000 +0      0.00000000 +0      1      0      0
H      1.03644234 +1     110.9040994 +1      0.00000000 +0      2      1      0
H      1.03643101 +1     110.9036359 +1     120.0082541 +1      2      1      3
H      1.03641712 +1     110.8984303 +1     119.9977109 +1      2      1      4
H      1.11206587 +1     111.8774978 +1      60.0148557 +1      1      2      3
H      1.11205744 +1     111.8759622 +1     119.9971123 +1      1      2      6
H      1.11209052 +1     111.8816803 +1     120.0004127 +1      1      2      7
```

MOZYME example: Constrained optimization of Crambin biased towards 1CBN

X-ray structures frequently have errors that can be corrected by a constrained geometry optimization using keyword [GEO REF](#). In this example, the 1CBN X-ray structure is improved using PM6. Two data sets are needed. One data set, "Crambin_X-ray_geometry.mop" holds the X-ray geometry. This is the reference that will be used by the constrained optimization. The other data set, "Constrained_Crambin_1CBN_optimization.mop" holds the geometry that will be optimized. In this case, the geometry is identical to the X-ray structure.

[Reference data set, data set, and final ARC file.](#)

States resulting from one-electron excitations

This job illustrates how to calculate the energies of electronic states of formic acid resulting from one-electron excitation.

Data-set:

```
bonds
First, optimize the geometry of formic acid

O      0.0  0      0.0  0      0.0  0      0  0  0
C      1.20  1      0.0  0      0.0  0      0  0  0
O      1.32  1    116.8  1      0.0  0      2  1  0
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

cis c.i.=4 meci oldgeo
Now calculate the excited states of formic acid
```

There are two calculations in this data set. In the first calculation, the geometry of formic acid is optimized. No keywords are needed for this operation, but because a blank line is not easy to see, the keyword [BONDS](#) has been added. BONDS is not necessary, but is used here only to show where the keyword line is.

The second calculation is the Configuration Interaction (C.I.) calculation. This uses the optimized geometry of formic acid from the previous calculation (using [OLDGEO](#)). Four C.I. keywords are used here:

[CIS](#): Use only the ground state and those states that can be created by excitation of one electron.

[C.I.=4](#): Use an active space of four molecular orbitals. Since this keyword is unqualified, these four M.O.s are the HOMO, HOMO-1, LUMO, and LUMO+1. That is, the two highest occupied and two lowest unoccupied M.O.s. The C.I. thus involves four electrons.

[MECI](#): Print out the microstates and the States.

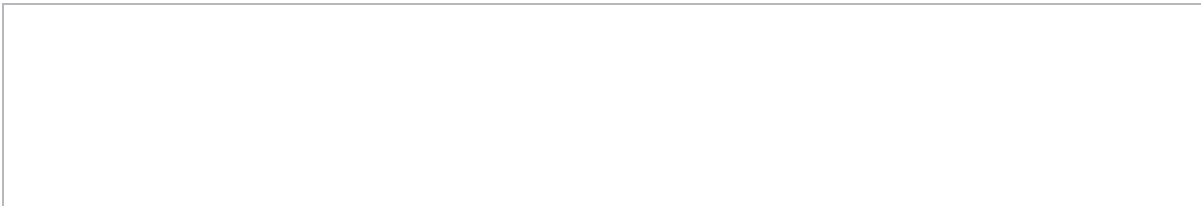
Other keywords, such as [SINGLET](#) are unnecessary here, because no geometric operations are being done.

In the output, there are a total of nine states: one for the ground state and eight resulting from one-electron excitations. In most C.I. calculations, the ground state would be stabilized by C.I., but in this case Koopmans' theorem applies. This says that no state resulting from one-electron excitation from a ground state configuration will interact with the ground state. The ground state is included in the CIS calculation, to indicate its position in the energy spectrum.

The methods in MOPAC are not accurate for predicting excitation energies. If excited state energies are needed, use ZINDO or other C.I. method.

Description of a simple vibrational calculation output file

(move the cursor over the file to see what the parts are)



```
*****
** Site#:      10                For non-commercial use only          Version 7.101W **
*****
** Cite this work as: MOPAC2007, James J. P. Stewart, Stewart Computational **
** Chemistry, Version 7.101W web: HTTP://OpenMOPAC.net    Days remaining: 363 **
*****
**
**                                MOPAC2007                            **
**
*****
```

PM6 CALCULATION RESULTS

```
*****
*  CALCULATION DONE:                               Wed Apr 11 11:13:27 2007  *
*  SYMMETRY - SYMMETRY CONDITIONS TO BE IMPOSED
*  T=      - A TIME OF 172800.0 SECONDS REQUESTED
*  DUMP=N   - RESTART FILE WRITTEN EVERY 7200.000 SECONDS
*****
```

PARAMETER DEPENDENCE DATA

REFERENCE ATOM	FUNCTION NO.	DEPENDENT ATOM(S)
3	1	4
3	2	4

DESCRIPTIONS OF THE FUNCTIONS USED

1	BOND LENGTH	IS SET EQUAL TO THE REFERENCE BOND LENGTH
2	BOND ANGLE	IS SET EQUAL TO THE REFERENCE BOND ANGLE

SYMMETRY
Formaldehyde

ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS) NA:I	BOND ANGLE (DEGREES) NB:NA:I	TWIST ANGLE (DEGREES) NC:NB:NA:I	NA	NB	NC
1	O	0.0000000	0.000000	0.000000			
2	C	1.2108153	* 0.000000	0.000000	1	0	0
3	H	1.0974996	* 122.023501	* 0.000000	2	1	0
4	H	1.0974996	122.023501	180.000000	2	1	3

CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
1	O	0.0000	0.0000	0.0000
2	C	1.2108	0.0000	0.0000
3	H	1.7928	0.9305	0.0000
4	H	1.7928	-0.9305	0.0000

```
H (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!
C (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!
O (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!
```

Empirical Formula: C H2 O

MOLECULAR POINT GROUP : C2v

RHF CALCULATION, NO. OF DOUBLY OCCUPIED LEVELS = 6

SYMMETRY
Formaldehyde

GRADIENTS WERE INITIALLY ACCEPTABLY SMALL
SCF FIELD WAS ACHIEVED

PM6 CALCULATION
MOPAC2007 (Version: 7.101W)
Wed Apr 11 11:13:27 2007
No. of days left = 363

FINAL HEAT OF FORMATION = -20.69775 KCAL = -86.59937 KJ

TOTAL ENERGY = -440.23219 EV
ELECTRONIC ENERGY = -825.76440 EV POINT GROUP: C2v
CORE-CORE REPULSION = 385.53221 EV
COSMO AREA = 60.76 SQUARE ANGSTROMS
COSMO VOLUME = 42.60 CUBIC ANGSTROMS
IONIZATION POTENTIAL = 10.20363
NO. OF FILLED LEVELS = 6
MOLECULAR WEIGHT = 30.026

MOLECULAR DIMENSIONS (Angstroms)

Atom	Atom	Distance
H 3	O 1	2.01987
H 4	O 1	1.65176
O 1	C 2	0.00000

SCF CALCULATIONS = 1
COMPUTATION TIME = 0.000 SECONDS

ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS)	BOND ANGLE (DEGREES)	TWIST ANGLE (DEGREES)	NA	NB	NC
		NA:I	NB:NA:I	NC:NB:NA:I			
1	O	0.0000000	0.000000	0.000000			
2	C	1.2108153 *	0.000000	0.000000	1	0	0
3	H	1.0974996 *	122.023501 *	0.000000	2	1	0
4	H	1.0974996	122.023501	180.000000	2	1	3

Empirical Formula: C H2 O

MOLECULAR POINT GROUP : C2v

EIGENVALUES

-32.00372 -23.55732 -17.17527 -15.14025 -14.75913 -10.20363 0.03208 4.08556
4.45994 4.74270

NET ATOMIC CHARGES AND DIPOLE CONTRIBUTIONS

ATOM NO.	TYPE	CHARGE	No. of ELECS.	s-Pop	p-Pop
1	O	-0.422510	6.4225	1.85390	4.56861
2	C	0.236385	3.7636	1.14506	2.61855
3	H	0.093063	0.9069	0.90694	
4	H	0.093063	0.9069	0.90694	
DIPOLE	X	Y	Z	TOTAL	
POINT-CHG.	2.978	0.000	0.000	2.978	
HYBRID	-0.160	0.000	0.000	0.160	
SUM	2.818	0.000	0.000	2.818	

CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
1	O	0.0000	0.0000	0.0000
2	C	1.2108	0.0000	0.0000
3	H	1.7928	0.9305	0.0000
4	H	1.7928	-0.9305	0.0000

ATOMIC ORBITAL ELECTRON POPULATIONS

1.85390	1.29031	1.93516	1.34314	1.14506	0.89773	1.06397	0.65686
0.90694	0.90694						

TOTAL CPU TIME: 0.02 SECONDS

== MOPAC DONE ==

START OF FORCE CALCULATION OUTPUT

** Site#: 10 For non-commercial use only Version 7.101W **

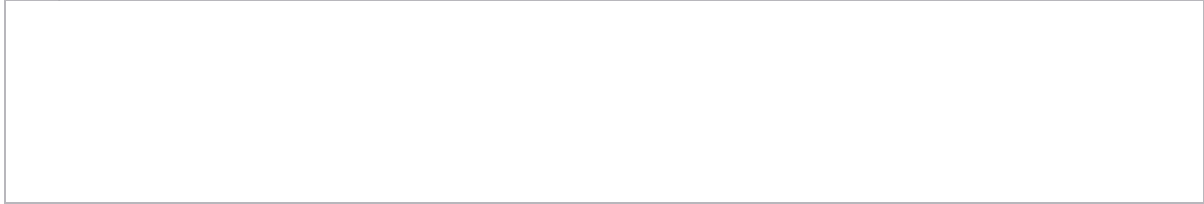
** Cite this work as: MOPAC2007, James J. P. Stewart, Stewart Computational **
** Chemistry, Version 7.101W web: HTTP://OpenMOPAC.net Days remaining: 363 **

**
** MOPAC2007 **
**

PM6 CALCULATION RESULTS

* CALCULATION DONE: Wed Apr 11 11:13:27 2007 *
* OLDGEO - PREVIOUS GEOMETRY TO BE USED
* FORCE - FORCE CALCULATION SPECIFIED
* T= - A TIME OF 172800.0 SECONDS REQUESTED
* DUMP=N - RESTART FILE WRITTEN EVERY 7200.000 SECONDS

oldgeo force



ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS) NA:I	BOND ANGLE (DEGREES) NB:NA:I	TWIST ANGLE (DEGREES) NC:NB:NA:I	NA	NB	NC
1	O	0.0000000	0.000000	0.000000			
2	C	1.2108153 *	0.000000	0.000000	1	0	0
3	H	1.0974996 *	122.023501 *	0.000000	2	1	0
4	H	1.0974996	122.023501	180.000000	2	1	3

CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
1	O	0.0000	0.0000	0.0000
2	C	1.2108	0.0000	0.0000
3	H	1.7928	0.9305	0.0000
4	H	1.7928	-0.9305	0.0000

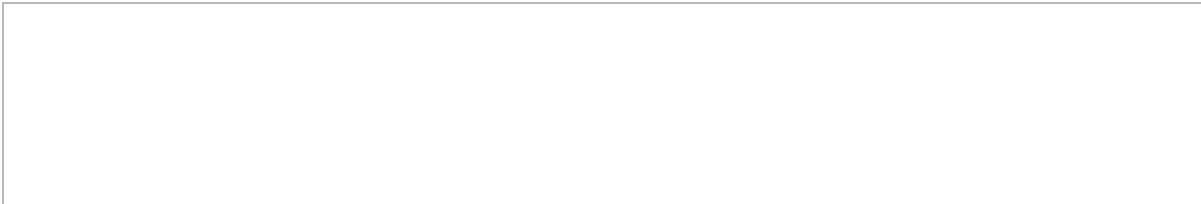
H (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!
C (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!
O (PM6): J. J. P. Stewart J. Mol. Mod. (to be submitted)!

Empirical Formula: C H2 O

MOLECULAR POINT GROUP : C2v

RHF CALCULATION, NO. OF DOUBLY OCCUPIED LEVELS = 6

HEAT OF FORMATION = -20.697747 KCALS/MOLE



INTERNAL COORDINATE DERIVATIVES

NUMBER	ATOM	BOND	ANGLE	DIHEDRAL
1	O			
2	C	-0.044282		
3	H	0.000000	0.000000	
4	H	-0.001510	0.000000	0.000000

GRADIENT NORM = 0.12762

MOLECULAR WEIGHT = 30.03

ROTATIONAL CONSTANTS IN CM(-1)

A = 9.658776 B = 1.286026 C = 1.134917

PRINCIPAL MOMENTS OF INERTIA IN UNITS OF 10**(-40)*GRAM-CM**2

A = 2.898170 B = 21.766874 C = 24.665044

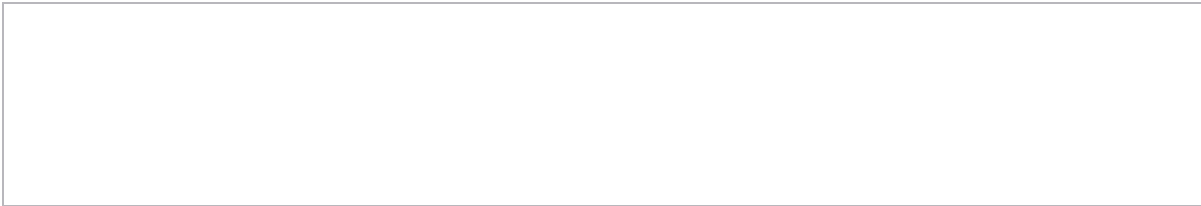
ORIENTATION OF MOLECULE IN FORCE CALCULATION

NO.	ATOM	X	Y	Z
1	O	-0.6047	0.0000	0.0000
2	C	0.6061	0.0000	0.0000
3	H	1.1881	0.9305	0.0000
4	H	1.1881	-0.9305	0.0000

FIRST DERIVATIVES WILL BE USED IN THE CALCULATION OF SECOND DERIVATIVES

FOR POINT-GROUP C2v THERE ARE 4 UNIQUE SYMMETRY FUNCTIONS.

STEP:	1	TIME =	0.00	SECS,	INTEGRAL =	0.00	TIME LEFT:	172800.00
STEP:	2	TIME =	0.00	SECS,	INTEGRAL =	0.00	TIME LEFT:	172800.00
STEP:	3	TIME =	0.00	SECS,	INTEGRAL =	0.00	TIME LEFT:	172800.00
STEP:	4	TIME =	0.02	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	5	TIME =	0.00	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	6	TIME =	0.00	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	7	TIME =	0.00	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	8	TIME =	0.00	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	9	TIME =	0.00	SECS,	INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	10				INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	11				INTEGRAL =	0.02	TIME LEFT:	172799.98
STEP:	12				INTEGRAL =	0.02	TIME LEFT:	172799.98



FORCE MATRIX IN MILLIDYNES/ANGSTROM

	O	1		C	2		H	3		H	4

O	1	13.833631									

C	2	12.333946	15.892804		
H	3	1.028611	3.252661	3.992890	
H	4	1.028662	3.252790	0.476883	3.992890

HEAT OF FORMATION = -20.697747 KCALS/MOLE

ZERO POINT ENERGY 15.272 KCAL/MOL

NORMAL COORDINATE ANALYSIS

Root No.	1	2	3	4	5	6
	1 B2	1 B1	1 A1	2 A1	2 B2	3 A1
	1100.7	1155.9	1351.1	1792.9	2615.9	2666.5
1	0.0000	0.0000	0.0285	0.2910	0.0000	-0.0030
2	-0.0486	0.0000	0.0000	0.0000	0.0039	0.0000
3	0.0000	0.0262	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0275	-0.3732	0.0000	-0.0392
5	0.0971	0.0000	0.0000	0.0000	0.0584	0.0000
6	0.0000	-0.1076	0.0000	0.0000	0.0000	0.0000
7	0.3810	0.0000	-0.3900	-0.0859	-0.2766	0.2572
8	-0.1931	0.0000	0.2659	-0.1442	-0.3786	0.4040
9	0.0000	0.4331	0.0000	0.0000	0.0000	0.0000
10	-0.3810	0.0000	-0.3900	-0.0859	0.2766	0.2572
11	-0.1931	0.0000	-0.2659	0.1442	-0.3786	-0.4040
12	0.0000	0.4331	0.0000	0.0000	0.0000	0.0000

MASS-WEIGHTED COORDINATE ANALYSIS

Root No.	1	2	3	4	5	6
	1 B2	1 B1	1 A1	2 A1	2 B2	3 A1
	1100.7	1155.9	1351.1	1792.9	2615.9	2666.5
1	0.0000	0.0000	0.1662	-0.6627	0.0000	0.0171
2	-0.2698	0.0000	0.0000	0.0000	0.0222	0.0000
3	0.0000	0.1443	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.1386	0.7365	0.0000	0.1960
5	0.4673	0.0000	0.0000	0.0000	0.2908	0.0000
6	0.0000	-0.5132	0.0000	0.0000	0.0000	0.0000
7	0.5310	0.0000	-0.5704	0.0491	-0.3991	-0.3724
8	-0.2691	0.0000	0.3889	0.0824	-0.5461	-0.5848
9	0.0000	0.5982	0.0000	0.0000	0.0000	0.0000
10	-0.5310	0.0000	-0.5704	0.0491	0.3991	-0.3724
11	-0.2691	0.0000	-0.3889	-0.0824	-0.5461	0.5848
12	0.0000	0.5982	0.0000	0.0000	0.0000	0.0000

DESCRIPTION OF VIBRATIONS

--	--	--	--	--	--

VIBRATION	1	1B2	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	1100.67		C 2 -- H 3	+43.6% (78.5%)	14.3%
T-DIPOLE	0.2926		C 2 -- H 4	+43.6% (78.5%)	14.3%
TRAVEL	0.1293		O 1 -- C 2	+12.7% (42.4%)	0.0%
RED. MASS	0.9107				
VIBRATION	2	1B1	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	1155.91		C 2 -- H 4	+45.3% (62.6%)	0.0%
T-DIPOLE	0.1461		C 2 -- H 3	+45.3% (62.6%)	0.0%
TRAVEL	0.1409		O 1 -- C 2	+9.3% (28.4%)	0.0%
RED. MASS	1.0983				
VIBRATION	3	1A1	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	1351.10		C 2 -- H 3	+49.6% (61.4%)	0.3%
T-DIPOLE	0.1257		C 2 -- H 4	+49.6% (61.4%)	0.3%
TRAVEL	0.1945		O 1 -- C 2	+0.8% (7.6%)	100.0%
RED. MASS	0.4745				

VIBRATION	4	2A1	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	1792.88		O 1 -- C 2	+60.1% (100.1%)	100.0%
T-DIPOLE	2.4960		C 2 -- H 3	+20.0% (57.7%)	18.1%
TRAVEL	0.0526		C 2 -- H 4	+20.0% (57.7%)	18.1%
RED. MASS	6.6199				
VIBRATION	5	2B2	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	2615.87		C 2 -- H 3	+49.5% (72.4%)	98.7%
T-DIPOLE	0.1198		C 2 -- H 4	+49.5% (72.4%)	98.7%
TRAVEL	0.1315		O 1 -- C 2	+1.0% (10.1%)	0.0%
RED. MASS	0.5078				
VIBRATION	6	3A1	ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
FREQ.	2666.50		C 2 -- H 4	+49.5% (69.5%)	95.6%
T-DIPOLE	0.0707		C 2 -- H 3	+49.5% (69.5%)	95.6%
TRAVEL	0.1419		O 1 -- C 2	+1.0% (9.9%)	100.0%
RED. MASS	0.4834				



FORCE CONSTANT IN INTERNAL COORDINATES (Millidynes/A)

ATOM NUMBER	CHEMICAL SYMBOL	BOND LENGTH FORCE CONSTANT	BOND ANGLE FORCE CONSTANT	TWIST ANGLE FORCE CONSTANT
1	O	0.000000	0.000000	0.000000
2	C	13.804069	0.000000	0.000000
3	H	3.899044	0.937825	0.000000
4	H	3.899044	0.937825	0.257944

TOTAL CPU TIME: 0.03 SECONDS

== MOPAC DONE ==

Use of symmetry to simplify the geometry optimization of benzene

SYMMETRY
Benzene

C	0.0	+0	0.0	+0	0.0	+0			
C	1.4	+1	0.0	+0	0.0	+0	1		
C	1.4	+0	120.0	+0	0.0	+0	2	1	
C	1.4	+0	120.0	+0	0.0	+0	3	2	1
C	1.4	+0	120.0	+0	0.0	+0	4	3	2
C	1.4	+0	120.0	+0	0.0	+0	5	4	3
H	1.1	+1	120.0	+0	180.0	+0	1	6	5
H	1.1	+0	120.0	+0	180.0	+0	5	6	1
H	1.1	+0	120.0	+0	180.0	+0	2	1	6
H	1.1	+0	120.0	+0	180.0	+0	6	1	2
H	1.1	+0	120.0	+0	180.0	+0	3	2	1
H	1.1	+0	120.0	+0	180.0	+0	4	3	2
2	1	3	4	5	6				
7	1	8	9	10	11	12			

How to set up a data-set for a Solid-State Calculation

For examples of data-sets for solids, please see: [SOLIDS](#)

A solid-state data set can be identified by the presence of three "Tv" atoms at the end of the geometry. A working definition would be as follows:

Unit cell

The geometry used represents a [unit cell](#) that can be translated in three dimensions to form an infinite crystal. This means that the unit cell is a volume defined by three sets of two planes. No space-group symmetry can be used. The unit cell used must be large enough to hold a sphere of radius 4 Angstroms. This is not a problem when the [primitive unit cell](#) is large, but if it is small then two or more primitive unit cells must be put together to make a unit cell that is large enough to hold the sphere. Thus in rock salt, halite, the primitive unit cell is a cube of size 5.64 Angstroms. In order to hold a sphere of diameter 8 Angstroms, eight primitive unit cells would need to be used. This would give rise to a unit cell of edge 11.28 Angstroms.

The unit cell should be defined in Cartesian coordinates, with atom 1 at the origin, that is, at coordinates (0.0, 0.0, 0.0).

Translation vectors

The translation vectors are represented in a unique way. At first, it might appear unnecessarily complicated, but in practice it is quite easy to use.

An atom in the unit cell is selected. The choice of atom is not important computationally, but can help in visualizing the structure later on. For example, if 1-Chloronaphthalene were being calculated, the chlorine atom would be an obvious choice, as it is easily located in the solid. This atom is put at the start of the geometry, that is, it is defined as atom 1. Its coordinates are

Let the unit cell used be defined as lattice point (0,0,0), then the three translation vectors are defined by lattice points (1,0,0), (0,1,0), and (0,0,1). Each translation vector will move atom 1 to the corresponding position in the adjacent unit cell. These positions are defined using the symbols "Tv". That is, the position of each Tv is defined as being where atom 1 would be if it were translated through one unit cell.

How to build a solid-state data set

A good starting point is to begin with a [cif](#) file and use [MERCURY](#) to generate a starting geometry. Open the cif file with MERCURY, and set up a 2 by 2 by 2 set of lattice points by using "Calculate" => "Packing/Slicing" => "Pack" then change the "1.0" to "2.0" for "a", "b", and "c". Save the resulting structure as an "ent" file.

Open the "ent" file in a graphics package, e.g., CAChe, and identify the atom to be used as atom 1. Re-arrange the atoms so that the atom selected is atom 1.

Identify the three atoms that atom 1 would become if it were translated through one lattice point in each of the three directions. Rename these atoms "Tv".

Delete all atoms that are not in lattice point (0,0,0). This might be tedious.

Save the file as Cartesian coordinates.

Edit the Cartesian coordinates to add optimization flags.

The lattice parameters a, b, and c, a, b, and g, are not used directly, but are used by MERCURY in generating the

lattice points.

Add keywords, title and comment, and run the data set using MOPAC.

Data Set for Polarizability Calculation

polar									
p-Nitroaniline									
C	0.000000000	+0	0.00000000	+0	0.00000000	+0			
C	1.42257500	+1	0.00000000	+0	0.00000000	+0	1	0	0
C	1.38320734	+1	119.5675708	+1	0.00000000	+0	2	1	0
C	1.41092128	+1	120.0832340	+1	-0.0144890	+1	3	2	1
C	1.41090330	+1	120.5179086	+1	-0.0088243	+1	4	3	2
C	1.38303031	+1	120.0641715	+1	-0.0054105	+1	5	4	3
N	1.38208898	+1	119.9019759	+1	-179.9846365	+1	1	2	3
N	1.45925645	+1	119.7481108	+1	179.9969816	+1	4	3	5
O	1.22348751	+1	117.9219301	+1	-179.9955699	+1	8	4	3
O	1.22345766	+1	117.9407076	+1	179.9982323	+1	8	4	9
H	1.08679394	+1	120.3019464	+1	179.9942765	+1	2	1	3
H	1.09706779	+1	120.8513014	+1	-179.9981944	+1	3	2	4
H	1.00365802	+1	121.5053119	+1	-0.0401016	+1	7	1	2
H	1.08684815	+1	120.1183662	+1	179.9986059	+1	6	5	4
H	1.09702234	+1	119.0764578	+1	179.9989202	+1	5	4	6
H	1.00365598	+1	121.5013061	+1	-179.9251444	+1	7	1	13

Description of Polarizability calculation output file

(Move the cursor over the file to see what the parts are. The first part is common to all calculations, so scroll down to "TDHF POLARIZABILITIES" to get to the start of the polarizability calculation.)

```
*****
** Site#:      0                For non-commercial use only          Version 7.263W **
*****
** Cite this work as: MOPAC2007, James J. P. Stewart, Stewart Computational **
** Chemistry, Version 7.263W web: HTTP://OpenMOPAC.net   Days remaining: 162 **
*****
**
**                                     MOPAC2007                        **
**                                     **
**                                     **
*****
```

PM6 CALCULATION RESULTS

```
*****
*  CALCULATION DONE:                               Fri Oct  5 09:12:43 2007  *
*  POLAR      -  CALCULATE FIRST, SECOND AND THIRD-ORDER POLARIZABILITIES
*  T=         -  A TIME OF 172800.0 SECONDS REQUESTED
*  DUMP=N     -  RESTART FILE WRITTEN EVERY  7200.000 SECONDS
*****
polar
p-Nitroaniline
```

ATOM NUMBER (I)	CHEMICAL SYMBOL	BOND LENGTH (ANGSTROMS) NA:I	BOND ANGLE (DEGREES) NB:NA:I	TWIST ANGLE (DEGREES) NC:NB:NA:I	NA	NB	NC
1	C	0.0000000	0.000000	0.000000			
2	C	1.4233530 *	0.000000	0.000000	1	0	0
3	C	1.3839170 *	119.551256 *	0.000000	2	1	0

[Lots of lines deleted]

ATOMIC ORBITAL ELECTRON POPULATIONS

```
1.07648  0.90708  0.83411  0.78810  1.10256  1.00684  1.08304  1.22790
1.10660  1.05136  0.91975  0.80854  1.10893  1.00696  0.95491  1.30825
1.10657  0.94436  1.02677  0.80858  1.10252  1.09864  0.99128  1.22785
1.44161  1.17064  1.13374  1.75959  1.29060  0.94808  0.97388  0.92398
1.91999  1.66947  1.33007  1.57361  1.91998  1.09224  1.90726  1.57359
0.81993  0.83645  0.73047  0.81994  0.83641  0.73050
1***** TDHF POLARIZABILITIES *****
```

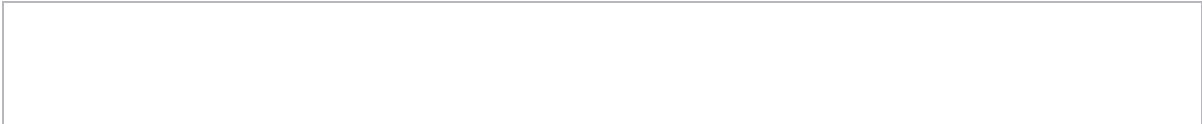
MOLECULAR WEIGHT = 138.13

ROTATIONAL CONSTANTS IN CM(-1)

A = 0.129251 B = 0.027297 C = 0.022537

PRINCIPAL MOMENTS OF INERTIA IN UNITS OF 10**(-40)*GRAM-CM**2

A = 216.577222 B = 1025.488877 C = 1242.066059



ROTATION MATRIX FOR ORIENTATION OF MOLECULE:

```
0.498516  -0.866881  -0.000309
0.866881   0.498516   0.000159
0.000016  -0.000347   1.000000
```

CARTESIAN COORDINATES

NO.	ATOM	X	Y	Z
-----	------	---	---	---

1	C	-2.1523	0.0001	0.0002
2	C	-1.4432	-1.2331	-0.0002
3	C	-0.0600	-1.2250	-0.0003
4	C	0.6401	-0.0001	-0.0002
5	C	-0.0596	1.2251	-0.0002
6	C	-1.4426	1.2331	-0.0005
7	N	-3.5344	0.0001	0.0005
8	N	2.0994	-0.0001	0.0001
9	O	2.6723	1.0810	0.0003
10	O	2.6726	-1.0809	0.0001
11	H	-1.9832	-2.1762	-0.0005
12	H	0.5081	-2.1636	-0.0004
13	H	-4.0589	-0.8556	-0.0001
14	H	-1.9826	2.1763	-0.0005
15	H	0.5085	2.1635	0.0000
16	H	-4.0589	0.8558	0.0003

ENERGY OF "REORIENTED" SYSTEM WITHOUT FIELD: 12.0699965485

NFREQ= 3 IWFLB= 0 IBET= 1 IGAM= 1
ATOL= 0.10000D-02 BTOL= 0.10000D-02 MAXITU= 500 MAXITA= 150

CALCULATION OF STATIC FIELD QUANTITIES

+++++ ALPHA AT 0.00000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	149.2600	-0.0014	-0.0108
*=Y:	-0.0014	106.4164	0.0009
*=Z:	-0.0111	0.0009	47.1637

ISOTROPIC AVERAGE ALPHA = 100.94667 A.U. = 14.95875 ANG.**3

+++++ BETA (SECOND HARMONIC GENERATION) AT 0.00000 EV.

COMPONENTS OF BETA

	*XX	*XY	*YY	*XZ	*YZ	*ZZ
*=X	-1763.70174	-0.00871	334.60500	-0.05335	0.00392	-1.46052
*=Y	-0.00887	334.48768	-0.03253	0.00374	-0.07887	-0.00208
*=Z	-0.05716	0.00382	-0.08294	-1.47796	-0.00208	0.02071

AVERAGE BETAX(SHG) VALUE AT 0.00000 EV = -858.3883
AVERAGE BETAY(SHG) VALUE AT 0.00000 EV = -0.0260
AVERAGE BETAZ(SHG) VALUE AT 0.00000 EV = -0.0685

AVERAGE BETA (SHG) VALUE AT 0.00000 EV = 858.3883

+++++ ALPHA AT 0.00000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	149.2600	-0.0014	-0.0108
*=Y:	-0.0014	106.4164	0.0009
*=Z:	-0.0111	0.0009	47.1637

ISOTROPIC AVERAGE ALPHA = 100.94667 A.U. = 14.95875 ANG.**3

GAMMA (THIRD HARMONIC GENERATION) AT 0.00000 EV.

GAMMA(X,X,X,X) = 119727.34761
GAMMA(Y,Y,Y,Y) = 6631.42257
GAMMA(Z,Z,Z,Z) = 23.04466
GAMMA(X,X,Y,Y) = -9486.72474
GAMMA(X,X,Z,Z) = 1248.88853
GAMMA(Y,Y,X,X) = -9486.75791
GAMMA(Y,Y,Z,Z) = 547.64178
GAMMA(Z,Z,X,X) = 1248.80255
GAMMA(Z,Z,Y,Y) = 547.52506

AVERAGE GAMMA VALUE AT 0.00000 = 22200.23802

CALCULATION FOR A FREQUENCY OF 0.25000 EV = 0.00919 A.U.
WAVELENGTH OF 4959.37 NM = 2016.38525 CM(-1)

+++++ ALPHA AT 0.25000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	149.6008	-0.0014	-0.0110
*=Y:	-0.0014	106.5021	0.0009
*=Z:	-0.0112	0.0009	47.1718

ISOTROPIC AVERAGE ALPHA = 101.09158 A.U. = 14.98023 ANG.**3

+++++ BETA (SECOND HARMONIC GENERATION) AT 0.25000 EV.

COMPONENTS OF BETA

	*XX	*XY	*YY	*XZ	*YZ	*ZZ
*=X	-1804.54238	-0.00962	343.11338	-0.06014	0.00359	-1.69468
*=Y	-0.01047	337.88196	-0.03275	0.00429	-0.08017	-0.00208
*=Z	-0.05411	0.00386	-0.08639	-1.51257	-0.00212	0.02106

AVERAGE BETAX(SHG) VALUE AT 0.25000 EV = -879.8939
AVERAGE BETAY(SHG) VALUE AT 0.25000 EV = -0.0269
AVERAGE BETAZ(SHG) VALUE AT 0.25000 EV = -0.0716

AVERAGE BETA (SHG) VALUE AT 0.25000 EV = 879.8939

+++++ ALPHA AT 0.75000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	152.4417	-0.0014	-0.0118
*=Y:	-0.0014	107.2041	0.0009
*=Z:	-0.0120	0.0008	47.2317

ISOTROPIC AVERAGE ALPHA = 102.29249 A.U. = 15.15818 ANG.**3

GAMMA (THIRD HARMONIC GENERATION) AT 0.25000 EV.

GAMMA(X,X,X,X) = 127264.84063

GAMMA(Y,Y,Y,Y) = 7010.42806
GAMMA(Z,Z,Z,Z) = 23.50275
GAMMA(X,X,Y,Y) = -10230.82001
GAMMA(X,X,Z,Z) = 1334.24709
GAMMA(Y,Y,X,X) = -9957.55914
GAMMA(Y,Y,Z,Z) = 560.83959
GAMMA(Z,Z,X,X) = 1269.82865
GAMMA(Z,Z,Y,Y) = 549.78398

AVERAGE GAMMA VALUE AT 0.25000 = 23565.01832

CALCULATION FOR A FREQUENCY OF 0.50000 EV = 0.01837 A.U.
WAVELENGTH OF 2479.68 NM = 4032.77050 CM(-1)

+++++ ALPHA AT 0.50000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	150.6406	-0.0014	-0.0113
*=Y:	-0.0014	106.7625	0.0009
*=Z:	-0.0115	0.0009	47.1939

ISOTROPIC AVERAGE ALPHA = 101.53230 A.U. = 15.04553 ANG.**3

+++++ BETA (SECOND HARMONIC GENERATION) AT 0.50000 EV.

COMPONENTS OF BETA

	*XX	*XY	*YY	*XZ	*YZ	*ZZ
*=X	-1937.93395	-0.01274	371.62769	-0.08052	0.00255	-2.46010
*=Y	-0.01606	348.40042	-0.03346	0.00609	-0.08111	-0.00206
*=Z	-0.04241	0.00403	-0.09817	-1.59848	-0.00222	0.02202

AVERAGE BETAX(SHG) VALUE AT 0.50000 EV = -950.2061
AVERAGE BETAY(SHG) VALUE AT 0.50000 EV = -0.0297
AVERAGE BETAZ(SHG) VALUE AT 0.50000 EV = -0.0796

AVERAGE BETA (SHG) VALUE AT 0.50000 EV = 950.2061

+++++ ALPHA AT 1.50000 EV.

COMPONENTS OF ALPHA

	*X	*Y	*Z
*=X:	163.9962	-0.0017	-0.0155
*=Y:	-0.0017	109.7957	0.0006
*=Z:	-0.0160	0.0006	47.4438

ISOTROPIC AVERAGE ALPHA = 107.07857 A.U. = 15.86741 ANG.**3

GAMMA (THIRD HARMONIC GENERATION) AT 0.50000 EV.

GAMMA(X,X,X,X) = 155208.90571
GAMMA(Y,Y,Y,Y) = 8369.84767
GAMMA(Z,Z,Z,Z) = 25.12361
GAMMA(X,X,Y,Y) = -13111.41695
GAMMA(X,X,Z,Z) = 1664.44729
GAMMA(Y,Y,X,X) = -11620.43412
GAMMA(Y,Y,Z,Z) = 605.53234
GAMMA(Z,Z,X,X) = 1339.63919
GAMMA(Z,Z,Y,Y) = 555.97263

AVERAGE GAMMA VALUE AT 0.50000 = 28607.52347

TOTAL CPU TIME: 1.48 SECONDS

== MOPAC DONE ==

Approximations used in MNDO, AM1, PM3, RM1, PM6, and MNDO-d

All the 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 that is not obvious from their names.

MNDO stands for Modified Neglect of Diatomic Overlap.

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

RM1 is Recife Model 1, developed by the group of A. M. Simas, Recife, Brazil

MNDO-*d*, is MNDO with *d* orbitals.

CNDO (Complete Neglect of Differential Overlap) ignored all terms involving two different atomic orbitals on one atom.

CNDO is an example of a ZDO (Zero Differential Overlap) 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, etc 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.

Parameters used in Semiempirical Methods

All the semiempirical methods contain sets of parameters. Atomic and diatomic parameters exist in PM6, while MNDO, AM1, PM3, and MNDO-*d* use only single-atom parameters. Not all parameters are optimized for all methods; for example, in MNDO and AM1 the two electron one center integrals are normally taken from atomic spectra. In the following Table, 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.

Table of Parameters

Parameter	Description	Units	MNDO	AM1	PM3	PM6
U_{ss}, U_{pp}, U_{dd}	$s, p,$ and d atomic orbital one-electron one-center integrals	eV	*	*	*	*
$b_s, b_p,$ and b_d	$s, p,$ and d atomic orbital one-electron two-center resonance integral terms	eV	*	*	*	*
$x_s, x_p,$ and x_d	$s, p,$ and d Slater atomic orbital exponent	bohr ⁻¹	*	*	*	*
$x_{sn}, x_{pn},$ and x_{dn}	$s, p,$ and d Slater atomic orbital internal exponent	bohr ⁻¹	*	*	*	*
a_A	Atom A core-core repulsion term	Å ⁻¹	*	*	*	
a_{AB}	Atoms A and B core-core repulsion term	Å ⁻¹				*
G_{ss}	s - s atomic orbital one center two electron repulsion integral	eV	+	+	*	*
G_{sp}	s - p atomic orbital one center two electron repulsion integral	eV	+	+	*	*
G_{pp}	p - p atomic orbital one center two electron repulsion integral	eV	+	+	*	*
G_{p2}	p - p' atomic orbital one center two electron repulsion integral	eV	+	+	*	*
H_{sp}	s - p atomic orbital one-center two-electron exchange integral	eV	+	+	*	*
K_{nA} or a_{nA}	A Gaussian multiplier for n th Gaussian of atom A	none		*	*	*
L_{nA} or b_{nA}	A Gaussian exponent multiplier Gaussian of atom A	Å ⁻²		*	*	*
M_{nA} or c_{nA}	A Radius of center of n th Gaussian of atom A	Å		*	*	*

Basic Roothaan-Hall Equations

Secular Equation

$$c_i/F-e_iS/c_i = 0$$

c_i : eigenvector; F : Fock matrix; e_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

$$F_{mn}^a = H_{mn} + S_l S_s [P_{mn}^{a+b} \langle mn|ls \rangle - P_{mn}^a \langle ml|ns \rangle].$$

Or, spin-free

$$F_{mn} = H_{mn} + S_l S_s [P_{mn} \langle mn|ls \rangle - (1/2) P_{mn} \langle ml|ns \rangle].$$

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:

$$c_i/F - e_i S/c_i = 0.$$

In semiempirical theory the Coulson density matrix is used, e.g.:

$$P_{ls}^a = S_i^{occ}(c_{li}^a c_{si}^a)$$

where the sum is over all occupied spin molecular orbitals. In RHF calculations, only the total density matrix is calculated:

$$P_{ls} = 2S_i^{occ}(c_{li} c_{si})$$

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:

$$P_{ls}^a = 1 - S_{i=occ+1}^N(c_{li}^a c_{si}^a)$$

and

$$P_{ls} = 2 - S_{i=occ+1}^N(c_{li} c_{si})$$

An important exception to this rule is the calculation of the one-electron two-center integral H_{mn} , which is approximated by:

$$H_{mn} = S_{mn}(U_{mm} + U_{nn})/2.$$

where S_{mn} is the overlap integral between atomic orbital f_m on an atom, and f_n 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 [DH_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 f_m and f_n are on different centers the NDDO matrix element F_{mn}^a reduces to

$$F_{mn}^a = H_{mn} + S_l^A S_s^B P_{ls}^a \langle ml|ns \rangle.$$

Equivalent expressions exist for F_{mn}^b and P_{mn}^b . Thus no Coulombic terms are present in the two-center Fock matrix elements.

If f_m and f_n are different but on the same center, then, since a minimal basis set is being used, all integrals of the type $\langle mn|ls \rangle$ are zero by the orthogonality of the atomic orbitals unless $m = n$ and $l = s$, or $m = l$ and $m = s$. The off-diagonal one-center NDDO Fock matrix elements become:

$$F_{mn}^a = H_{mn} + 2P_{mn}^{a+b} \langle mn|mn \rangle - P_{mn}^a (\langle mn|mn \rangle + \langle mm|nn \rangle).$$

If f_m is the same as f_n , then, because of the symmetry of the two-electron integrals, the diagonal NDDO Fock matrix elements reduce to:

$$F_{mm}^a = H_{mm} + S_n^A (P_{mm}^{a+b} \langle mm|nn \rangle - P_{nn}^a \langle mn|mn \rangle) + S_B S_l^B S_s^B P_{ls}^{a+b} \langle mm|ls \rangle.$$

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:

$$\langle pp|p'p' \rangle = 1/2(\langle pp|pp \rangle - \langle pp|p'p' \rangle)$$

Proof: If the molecular frame is rotated by 45° about the z axis the atomic bases mix thus:

$$R(45^\circ)p_x = 0.7071(p_x + p_y)$$

$$R(45^\circ)p_y = 0.7071(p_y - p_x)$$

$$\begin{aligned} R(45^\circ)\langle p_x p_y / p_x p_y \rangle &= 1/4 \langle (p_x + p_y)(p_y - p_x) | (p_x + p_y)(p_y - p_x) \rangle \\ &= 1/4 (\langle p_x p_x | p_x p_x \rangle + \langle p_y p_y | p_y p_y \rangle - \langle p_x p_x | p_y p_y \rangle - \langle p_y p_y | p_x p_x \rangle) \end{aligned}$$

or

$$R(45^\circ)\langle p_x p_y / p_x p_y \rangle = 1/2 (\langle p_x p_x | p_x p_x \rangle - \langle p_y p_y | p_x p_x \rangle)$$

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:

$$\begin{aligned} F_{ss}^\alpha &: P_{ss}^\beta G_{ss} + (P_{p_x}^{\alpha+\beta} + P_{p_y}^{\alpha+\beta} + P_{p_z}^{\alpha+\beta}) G_{sp} - (P_{p_x}^\alpha + P_{p_y}^\alpha + P_{p_z}^\alpha) H_{sp} \\ F_{sp}^\alpha &: 2P_{sp}^{\alpha+\beta} H_{sp} - P_{sp}^\alpha (H_{sp} + G_{sp}) \\ F_{pp}^\alpha &: P_{ss}^{\alpha+\beta} G_{sp} - P_{ss}^\alpha H_{sp} + P_{pp}^\beta G_{pp} + (P_{p'p'}^{\alpha+\beta} + P_{p''p''}^{\alpha+\beta}) G_{p2} - \frac{1}{2} (P_{p'p'}^\alpha + P_{p''p''}^\alpha) (G_{pp} - G_{p2}) \end{aligned}$$

$$F_{pp'}^\alpha = P_{pp'}^{\alpha+\beta}(G_{pp} - G_{p2}) - \frac{1}{2}P_{pp'}^\alpha(G_{pp} + G_{p2})$$

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 1.

Table 1: Two-Electron Two-Center Integrals (Local Frame)

1	$\langle ss ss\rangle$	12	$\langle s\ p_s/p_p\ p_p\rangle$
2	$\langle ss p_p\ p_p\rangle$	13	$\langle s\ p_s/p_s\ p_s\rangle$
3	$\langle ss p_s\ p_s\rangle$	14	$\langle ss s\ p_s\rangle$
4	$\langle p_p\ p_p/ss\rangle$	15	$\langle p_p\ p_p/s\ p_s\rangle$
5	$\langle p_s\ p_s/ss\rangle$	16	$\langle p_s\ p_s/s\ p_s\rangle$
6	$\langle p_p\ p_p/p_p\ p_p\rangle$	17	$\langle s\ p_p/s\ p_p\rangle$
7	$\langle p_p\ p_p/p_p'\ p_p'\rangle$	18	$\langle s\ p_s/s\ p_s\rangle$
8	$\langle p_p\ p_p/p_s\ p_s\rangle$	19	$\langle s\ p_p/p_p\ p_s\rangle$
9	$\langle p_s\ p_s/p_p\ p_p\rangle$	20	$\langle p_p\ p_s/s\ p_p\rangle$
10	$\langle p_s\ p_s/p_s\ p_s\rangle$	21	$\langle p_p\ p_s/p_p\ p_s\rangle$
11	$\langle s\ p_s/ss\rangle$	22	$\langle p_p\ p_p'/p_p\ p_p'\rangle$

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 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:

$$\langle p_p\ p_p'/p_p\ p_p'\rangle = 1/2(\langle p_p\ p_p/p_p\ p_p\rangle - \langle p_p\ p_p/p_p'\ p_p'\rangle).$$

The electron density distributions are approximated by a series of point charges. There are four possible types of distribution. These are given in Table 2.

Table 2: Types of Electron Density Distribution

Monopole	Unit negative charge centered on the
(1 charge)	nucleus
Dipole	+1/2 charge located at position (x,y,z),

(2 charges)	-1/2 charge located at position (-x,-y,-z)
Linear Quadrupole	+1/2 charge located at the nucleus, -1/4 charge
(3 charges)	at positions (x,y,z) and at (-x,-y,-z)
Square Quadrupole	Four charges of magnitude +1/4, -1/4, +1/4
(4 charges)	and -1/4 forming a square centered on the nucleus.

These are used to represent the four types of atomic orbital products (Table 3).

Table 3: 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

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 $\langle p_p p_p / p_p ' p_p' \rangle$, 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 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 $2^{1/2}D_2$ 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 4.

Table 4: Additive Terms for Two-Electron Integrals

Multipole	Monocentric Equivalent	Name
Monopole	G_{ss}	AM
Dipole	H_{sp}	AD
Quadrupole	$H_{pp} = 1/2(G_{pp} - G_p^2)$	AQ

In practice, $1/2(G_{pp} - G_p^2)$ 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

$$a_n = 1/2AD_n - 1/2(4D_1^2 + AD_n^{-2})^{-1/2}.$$

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,

$$a_n = 1/4AQ_n - 1/2(4D_2^2 + AQ_n^{-2})^{-1/2} + 1/4(8D_2^2 + AQ_n^{-2})^{-1/2},$$

H2H2 1 integral
H3C 10 integrals
H3H1 1 integral
H3H2 1 integral
H3H3 1 integral
H4C 10 integrals
H4H1 1 integral
H4H2 1 integral
H4H3 1 integral
H4H4 1 integral

for a total of 150 integrals. To see how these are related, run methane with HCORE and compare the two-electron matrix with the following integrals:

TWO-ELECTRON MATRIX IN HCORE (Methane)										
12.2300	0.0000	11.4700	0.0000	0.0000	11.4700	0.0000	0.0000	0.0000	11.4700	CC
0.0000	2.4300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
11.4700	0.0000	11.0800	0.0000	0.0000	9.8400	0.0000	0.0000	0.0000	9.8400	
0.0000	0.0000	0.0000	2.4300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.6200	0.0000	0.0000	0.0000	0.0000	0.0000	
11.4700	0.0000	9.8400	0.0000	0.0000	11.0800	0.0000	0.0000	0.0000	9.8400	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	2.4300	0.0000	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6200	0.0000	0.0000	
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6200	0.0000	
11.4700	0.0000	9.8400	0.0000	0.0000	9.8400	0.0000	0.0000	0.0000	11.0800	
9.0917	-0.6191	8.6000	-1.9992	0.2460	9.3181	0.0000	0.0000	0.0000	8.5239	H1C
12.8480										H1H1
9.0917	-0.7361	8.6315	0.9582	-0.1402	8.7063	1.7088	-0.2500	0.3254	9.1042	H2C
6.8458										H2H1
12.8480										H2H2
9.0917	-0.7361	8.6315	0.9582	-0.1402	8.7063	-1.7088	0.2500	-0.3254	9.1042	H3C
6.8458										H3H1
6.8458										H3H2
12.8480										H3H3
9.0917	2.0912	9.3930	0.0827	0.0344	8.5252	0.0000	0.0000	0.0000	8.5239	H4C
6.8458										H4H1
6.8458										H4H2
6.8458										H4H3
12.8480										H4H4

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:

$$\langle ss|ss \rangle = \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AM_B}]^2 0.529177^2)^{1/2}}.$$

$\langle ss|p_p p_p \rangle$: $p_p p_p$ 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:

$$\begin{aligned} \langle ss|p_p p_p \rangle = & \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AM_B}]^2 0.529177^2)^{1/2}} \\ & + \frac{\frac{1}{2}14.399}{(R_{AB}^2 + (2D_2^B)^2 1/2[\frac{1}{AM_A} + \frac{1}{AQ_B}]^2 0.529177^2)^{1/2}} \\ & - \frac{14.399}{(R_{AB}^2 + 1/2[\frac{1}{AM_A} + \frac{1}{AQ_B}]^2 0.529177^2)^{1/2}}. \end{aligned}$$

The one-center one-electron integral H_{mm}

This represents the energy an electron in atomic orbital f_m 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_{mm} . H_{mm} is derived from the fundamental equation

$$H_{mm} = U_{mm} - S_{B/A} Z_B \langle mm|BB \rangle$$

by equating the core-electron integral $\langle mm|BB \rangle$ to the corresponding two-electron integral, thus:

$$\langle mm|BB \rangle = Z_B \langle mm|ss \rangle$$

The two-center one-electron integral H_{mn}

Sometimes called the resonance integral, H_{mn} is approximated using the overlap integral, S_{mn} . 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

$$f_m = N r^{n-1} e^{-\alpha r}$$

the overlap integral is given by:

$$S_{mn} = \langle f_m f_n \rangle.$$

H_{mn} is approximated by:

$$H_{mn} = S_{mn} 1/2 (\alpha_m + \alpha_n);$$

Core-core repulsion integrals

From simple electrostatics the core-core repulsion integral in eV is:

$$E_N(A,B) = 14.399 \, Z_A Z_B / R_{AB}$$

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:

$$E_N(A,B) = Z_A Z_B \langle AA | BB \rangle.$$

MNDO modification to the core-core term

The MNDO approximation to the screening effect is :

$$E_N(A,B) = Z_A Z_B \langle s_A s_A | s_B s_B \rangle (1 + e^{-a_A R_{AB}} + e^{-a_B R_{AB}}).$$

Again, O-H and N-H interactions are treated differently. For these interactions, use

$$E_N(A,B) = Z_A Z_B \langle s_A s_A | s_B s_B \rangle (1 + R_{AB} e^{-a_B R_{AB}} + e^{-a_H R_{AB}}).$$

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`).

Calculation of DH_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 formation for a gaseous atom from its standard state, $E_{atom}(A)$, are added to the electronic plus nuclear energy. This yields:

$$\text{DH}_f = (E_{elect} + E_{nuc}) - S_A E_{isol}(A) + S_A E_{atom}(A),$$

or

$$\text{DH}_f = E_{tot} - S_A E_{isol}(A) + S_A E_{atom}(A).$$

This is the quantity which MOPAC calls the "Heat of Formation". An alternative but equivalent definition of DH_f , more suited for comparison with experimental DH_f , is:

" DH_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 DH_f at 298K. This follows from the way in which semiempirical methods are parameterized: the reference DH_f are conventionally given at 298K. This means that semiempirical methods will reproduce DH_f for 298K.

Secondly, note that DH_f are for gas-phase systems. To calculate DH_f in the liquid or solid phases, additional terms are necessary.

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:

$$\begin{aligned} G_{ss} &= \langle f_s f_s | 1/r | f_s f_s \rangle = 12.848 \text{ eV} \\ U_{ss} &= \langle f_s | H | f_s \rangle = -11.906276 \text{ eV} \\ x_s &= 1.331967 \text{ Bohr} \\ b_s &= -6.989064 \text{ eV} \\ E_{atom} &= 52.102 \text{ kcal/mol} \end{aligned}$$

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:

$$\langle \phi | \phi \rangle = (e^{-a}) \left(\frac{a^2}{3} + a + 1 \right)$$

where $a = xR/a_o$.

At the optimum H-H distance of 0.9831571Å, this yields an overlap integral of 0.4643. The nearest-neighbor one-

electron integral is thus

$$H(1,2) = S_{1,2}(\beta_s + \beta_s)/2 = -3.2457eV.$$

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

$$E_{e,n} = -Z_n < \varphi_s \varphi_s | \varphi_s \varphi_s > .$$

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 $<ii|ss>$, $<ii|sp_x>$, $<ii|p_xp_x>$, $<ii|sp_y>$, $<ii|p_xp_y>$, $<ii|p_y p_y>$, $<ii|sp_z>$, $<ii|p_xp_z>$, $<ii|p_y p_z>$, and $<ii|p_zp_z>$.

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

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.

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 f_a . Consider $F(1,1)$. The total initial population of f_I is 1.0, composed of equal amounts of a and b electron density. An electron in f_I 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:

$$F(1,1) = -51.7124 + 1/2(12.848) + 2(9.6585 + 7.0635) + 6.3622$$

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

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

$$\langle \psi_i | \psi_j \rangle = \sum_{\lambda} c_{\lambda i} c_{\lambda j} = \delta(i, j).$$

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.

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.

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.

Calculation of heat of formation

The heat of formation is defined as:

$$DH_f = E_{elect} + E_{nuc} - E_{isol} + E_{atom}$$

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.

(For a single atom, there is no nuclear term, therefore:

$$DH_f = E_{elect} - E_{isol} + E_{atom}$$

But $E_{elect} = E_{isol}$ for an isolated atom, therefore

$$DH_f = E_{atom}$$

where E_{atom} is the experimental heat of formation of the isolated atom from the element in its standard state. For example, in nitrogen this is half the heat required to break a N_2 molecule into its separated atoms. Isolated nitrogen atoms have the configuration $1s^2 2s^2 2p^3$, and the state [4S_u](#), i.e. quartet S. Experimentally, this energy is 113 kcal/mol. This means that all semiempirical methods predict the heats of formation of isolated atoms with a zero error, except in those few instances when the predicted ground state is incorrect.)

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:

$$\begin{aligned} E_{elect} = & 3(+1.0000)(-51.7124 + -5.4823) \\ & + 6(+0.6667)(-3.2457 + - \\ & 6.4652) \\ & + 3(-0.3333)(-0.6992 + -0.3611) \end{aligned}$$

or

$$E_{elect} = -210.0898 \text{ eV}.$$

E_{nuc} is a relatively straightforward calculation, and is equal to 130.2902 eV. The total energy of the system is thus -79.7996 eV.

We are now ready to calculate the DH_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

$$DH_f = 119.780 \text{ kcal/mol.}$$

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 DH_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 the Table may prove useful.

Table: 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		

These numbers may be used in conjunction with the semiempirical electronic and nuclear energies to calculate the

heat of formation.

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 G 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 G point.

At this point in the calculation, conventionally, the density matrix for each unit cell is constructed. Instead, the G-point density and one-electron density matrices are combined with "G-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 G 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 Ångstrom is sufficient. For one rare group of compounds a larger translation vector is needed. Solids with delocalized p-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.

The Madelung Problem

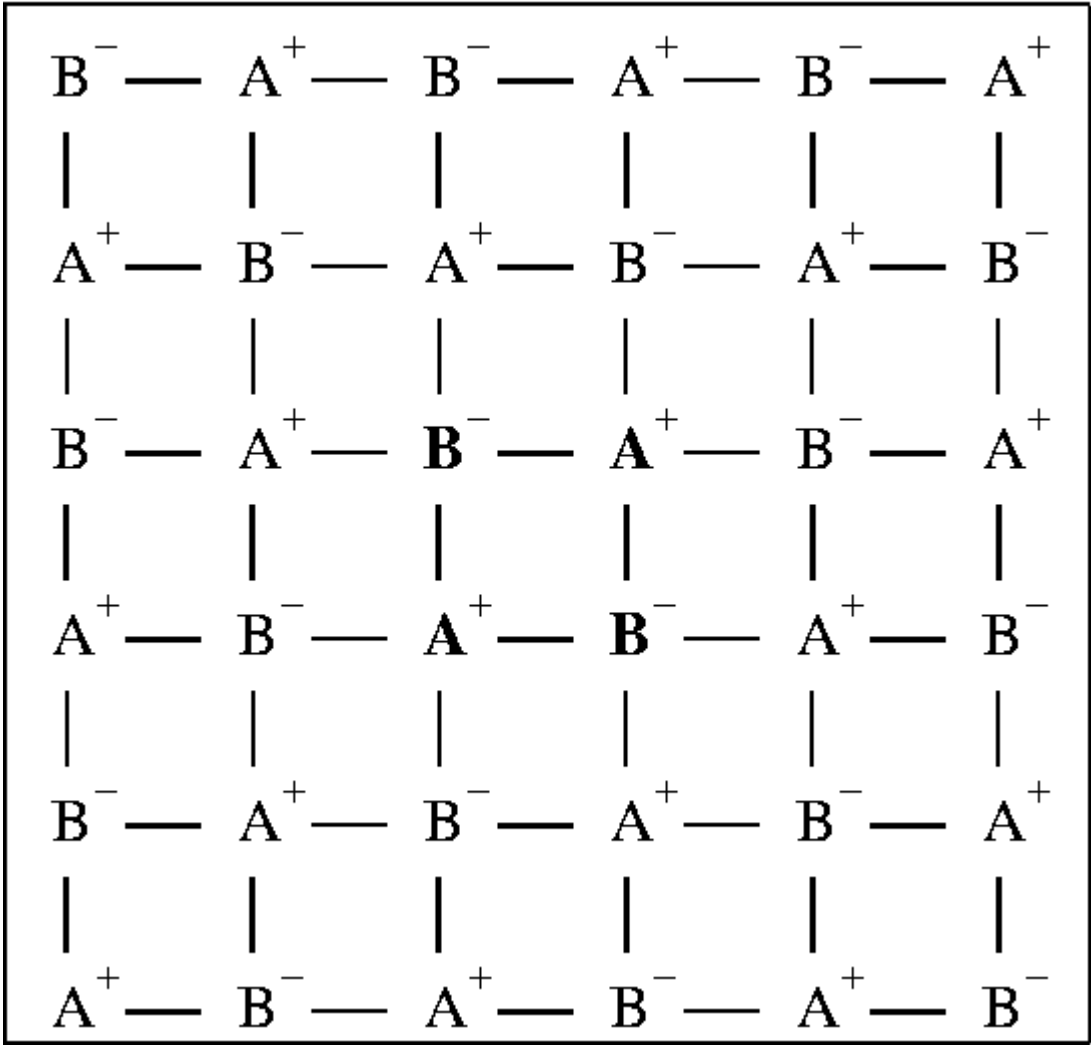
The number of diatomic interactions in molecules is finite, being $(N(N+1))/2$ in number, where N is the number of atoms. Solids, on the other hand, have potentially an almost infinite number of diatomic interactions. In order to reduce this enormous number of interactions to a manageable amount, approximations are necessary.

Quantum effects arising from the overlaps of atomic orbitals can extend to a distance of a few tens of Ångstroms. However, the overlaps themselves become negligible at distances of only a few Ångstroms. Overlap integrals decrease in value at large distances in proportion to the inverse exponential of their orbital exponents. Because of this fact, all overlap integrals involving atoms separated by more than a specified distance can be ignored. In practice, a limit of 5-7 Å should be used. One advantage of limiting overlap integrals is that the calculation runs much faster. A diamond calculation involving 64 carbon atoms took 31 minutes to do one SCF calculation. When overlaps involving atoms separated by more than 7 Ångstroms were ignored, the calculation took less than 3 minutes.

Electrostatic terms are more difficult to approximate. Unlike quantum effects, electrostatic terms remain significant even at quite large distances, say several tens of Ångstroms. Simply ignoring all terms beyond a specific distance results in large errors. Consider a simple two-dimensional ionic lattice consisting of alternating positive and negative charges in a square array (Figure). Each positive ion is surrounded at a distance r by four negative ions, and at a distance $2^{1/2}r$ by four positive ions. If all electrostatic interactions at distances greater than a preset value are ignored, then each positive ion will experience an electrostatic field due to 4 or 8 more or less positive ions than negative ions. Consider first the effect of 4 excess positive ions. Because of electrostatic repulsion, these ions would raise the energy of the ion by $1328/R$ kcal/mol, where R was the cutoff-distance for electrostatic interactions. Similarly, if there were 4 excess negative ions, each positive ion would be stabilized by $1328/R$ kcal/mol. Even for large R , say 20 Ångstroms, the effect would amount to many tens of kcal/mol.

One option would be to include all interactions in a finite set of fundamental unit cells. Since each unit cell is, by definition, electrically neutral, this would result in each ion experiencing a field arising from all other ions. The sum of the charges on all the ions each ion would experience would be simply equal to the negative of the ion's own charge. Consider a cell consisting of 2A(+).2B(-), in which the four ions form a square with alternating charges, e.g. A(+)-B(-)-A(+)-B(-). If 9 cells, forming a 3 by 3 grid, were used to represent the solid, then ion A(+) would experience the environment of 17 A(+) ions and 18 B(-) ions. This is similar to the situation which occurs in a molecule.

Figure: Two-dimensional ionic lattice



This approximation is acceptable for polymer systems, but it can induce a small error in the symmetry of the electric field in a two-dimensional system. Consider the electric field on the ions A(+) and B(-) in our system. Each ion would experience an electrostatic environment which did not have square symmetry. To see this, consider the number of B(-) ions the A(+) ion would interact with, 18 in all. It is clearly impossible to position 18 B(-) ions in such a way as to yield a potential having square symmetry at ion A(+).

For a three dimensional system, the problem is much more serious. For a unit cell consisting of A(+)-B(-), every "A" ion would experience a highly asymmetric potential. Increasing the number of unit cells involved does not solve the problem--with increasing numbers of unit cells, the effects arising from distant ions becomes less, but the number of distant ions increases. Put another way, the influence of an ion falls off as r^{-2} , but the number of such ions increases as r^2 . This phenomenon is peculiar to 3-D systems.

To correct this error in the symmetry of the potential, a new approximation is needed, in which the environment of each ion has the correct symmetry. In addition to having the correct symmetry, the function must have a continuous first derivative, in order for the derivatives to be calculable.

A cutoff distance is selected, this is called [CUTOFP](#); by default this is set at 30 Ångstroms. The electrostatic terms arising from all atoms up to $2/3CUTOFP$ are treated as normal. Between $2/3CUTOFP$ and $4/3CUTOFP$, the apparent distance of the atom is reduced, and for atoms separated by more than $CUTOFP$, the apparent distance is set to $CUTOFP$. These three domains can be summarized as follows:

Domain	Apparent Distance*
$R_{AB} < (2/3)C$	R_{AB}

$(2/3)C < R_{AB} < (4/3)C$	$2R_{AB} - C/3 - \frac{3}{4}R^2/C$
$R_{AB} > (4/3)C$	C

∗: C=CUTOFP

Inspection of this function shows that in the region of $R=(2/3)CUTOFP$, the function value is $(2/3)CUTOFP$ and the derivative is 1.0. In the region $R=(4/3)CUTOFP$ the function has the value $CUTOFP$, and has a derivative of zero. In between these two limits, the value of the derivative decreases smoothly from 1.0 to 0.0.

Piezoelectric Effect

The piezoelectric effect is a phenomenon exhibited by certain crystals of expansion in one direction (axis) and contraction in another direction, as the result of an applied electric field. Also, an applied mechanical strain in one direction causes electric charges of opposite signs to appear on different faces of the crystal. This phenomenon can be modeled by simply altering the different translation vectors of a solid. That is, altering the Tv bond-length from the optimized value would simulate stress in that direction. After geometry optimization, with the Tv bond length not optimized, the change in electric charges from the values for the relaxed crystal would be a measure of the piezoelectric effect..

This type of phenomenon might be easier to model using Cartesian coordinates.

Refractive Index

Introduction

For a vacuum, the refractive index is 1.000. This means that light travels through a vacuum at the speed c . When light goes through a medium such as glass, the time needed to travel a given distance increases. The conventional way of looking at this is to say that light is slowed down as it passes through the glass. An alternative, and equivalent, explanation would be to say that the apparent distance the light has to travel has increased. The extra distance can be expressed as a ratio, thus a refractive index of 1.4 would mean that the light was slowed by a factor of 1.4 or that the distance was increased by the same factor. In three dimensions, the refractive index represents the extra volume that appears to exist when light goes through a medium. Thus if a glass had a refractive index of 1.4, it would mean that a 10cc block of it would appear to have a volume of 14 cc, if the volume was determined using light.

Refractive index is related to polarizability, in that it represents the extra volume that appears to exist when light goes through a medium.

The refractive index of a solid cannot be calculated directly. Nor can the polarizability of a solid be calculated, if translation vectors are used.

How to Calculate Refractive Index

The polarizability at various energies (wavelengths) can be calculated for the cluster, using [POLAR](#), if translation vectors are removed and the "dangling bonds" are satisfied. The best atom for this is the capped bond ([Cb](#)). Obviously, the calculated polarizability includes the contributions from the surface, so the surface effects must be removed. This is easily done by doing calculations on different clusters. Consider a cluster of size 10x10x10 Å. If this was increased to 15x15x15, the surface area (surface effect) would increase from 6x10x10 to 6x15x15 square Angstroms, while the volume (what we want) would increase from 10x10x10 to 15x15x15 cubic Angstroms.

Given the polarizability, a , and the volume, V , of the cluster, the average refractive index, n , is readily given by $n = (1+a/v)$. If [LET](#) is used, the system will not be re-oriented, and the individual polarizabilities (in x , y , and z) can be calculated. From these, the refractive indices in different directions can be calculated.

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.

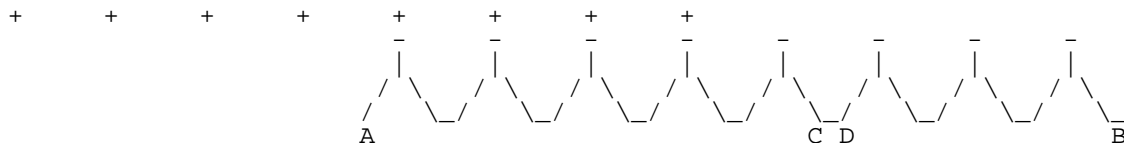
The Cartesian derivatives of the surrounding unit cells represent the forces acting on the atoms in those cells arising from the atoms of the CUC. Again, this is an unbalanced set of forces, and those atoms near to the cell boundaries are likely to have large resultant forces.

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.

Charge Balanced Unit Cells

Obviously, the net charge in a unit cell is zero. A less obvious need is for the dipole of the unit cell to be as small as possible. Consider the highly artificial system:



This system represents a simple hydrocarbon backbone with anionic sites covalently bound. The anionic charge is balanced by cation counterions. However, although there is one cation for each anion, and the unit cell is uncharged, the unit cell does have a large dipole.

Why is this undesirable? For the purpose of this discussion, let [CUTOFF](#) be infinite. Then electrons in atom *A* would be stabilized by every cationic charge and destabilized by every anionic charge. With increasing distance the effect of the cations and anions would more-or-less cancel out. However, at large distances an asymmetry in the charge distribution is introduced. To the left of *A* there would be 8 anions and 12 cations, while to the right of *A* there would be 16 anions and 12 cations. The potential effect of this asymmetry would be a net positive potential at *A*. Similarly, at *B* there would be a net negative potential arising from the charges.

As a result of the Born-von Kármán periodic boundary conditions, atom *A* is bonded to atom *B*. Because of the symmetry of this system, *A* and *B* should have the same charge (roughly zero), but because of the large unit cell dipole, there will be an induced dipole in the bond *A-B*. There will be induced dipoles in all bonds of the *A-B* type, for example *C-D*, but these will be very small compared to that in the *A-B* bond.

The overall effect of this induced bond dipole is that the molecular geometry in the region of *A* and *B* is deformed. In a correctly optimized system, the effect of having a large dipole in the unit cell is to induce gradient effects at the ends of the unit cell.

To avoid these effects, the unit cell should be chosen in such a way as to make the dipole of the unit cell as small as possible. For most systems, the default value of `CUTOFF` used by the program will prevent this problem from occurring.

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 for polytetrahydrofuran. When band structures are to be calculated, [the order of atoms is important](#).

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.

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, $1/m$, $2/m$, ...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.

Starting electronic configuration

As MECI requires the space parts of the a and b 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 the [Table](#).

Table: 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

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 p-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.

Microstates

Microstates, which are normally represented by a Slater determinant, are normally written as an antisymmetrized product of p a- and q b-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 $[(p+q)!]^{1/2}$ 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:

$$\Psi_j = \frac{1}{\sqrt{N!}} \sum_{P=1}^{N!} (-1)^P P(\prod_{k=1}^N \psi_k^j)$$

where Y_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 Y_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 a 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 a and b M.O.s of the same index occur, then a precedes b, 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 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 the Table.

Table: Microstates for 5 electrons in 5 M.O.s

	Electron Configuration				Electron Configuration		
	Alpha	Beta	M _S		Alpha	Beta	M _S
M.O.	1 2 3 4 5	1 2 3 4 5		M.O.	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

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 symbol . 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, $\sqrt{1/10}$ (

$j_1 = 3/2, m_1 = 3/2, j_2 = 1, m_2 = -1, j = 5/2$).

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.

Table 1: 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)x(0,1)	=	1	(1,1)x(1,1)	=	1
2	(1,2)x(0,2)	=	2	(1,2)x(1,2)	=	4
3	(2,3)x(1,3)	=	9	(2,3)x(2,3)	=	9
4	(2,4)x(1,4)	=	24	(2,4)x(2,4)	=	36
5	(3,5)x(2,5)	=	100	(3,5)x(3,5)	=	100

(*n,m*) means *n* electrons in *m* M.O.s.

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 1), 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 2.

Table 2: Number of Microstates for CIS, CISD, and CISDT

C.I.= k	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)

Footnotes

3-J symbol
The symbol is of form

$$\langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle = \left\{ \frac{(j+m)!(j-m)!(j_1-m_1)!(j_2-m_2)!(j_1+j_2-j)!(2j+1)}{(j_1+m_1)!(j_2+m_2)!(j_1-j_2+j)!(j_2-j_1+j)!(j_1+j_2+j+1)!} \right\}^{\frac{1}{2}}$$
$$\delta(m, m_1+m_2) \sum_r (-1)^{j_1+r-m_1} \frac{(j_1+m_1+r)!(j_2+j-r-m_1)!}{r!(j-m-r)!(j_1-m_1-r)!(j_2-j+m_1+r)!}$$

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 $\langle j_1 j_2 m_1 m_2 | j_1 j_2 j m \rangle$

Energy of microstates

The electronic energy, E_r , of any microstate Y_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. y_i , J_{ii} is the two-electron Coulomb integral $\langle y_i y_i | y_j y_j \rangle$, and K_{ii} is the two-electron exchange integral $\langle y_i y_j | y_i y_j \rangle$. In this section it is more convenient to express it in terms of molecular orbital occupancies:

$$E_r = \sum_{i=B}^A H_{ii} (O_i^{\alpha r} + O_i^{\beta r}) + \sum_{ij=B}^A \left(\frac{1}{2} (J_{ij} - K_{ij}) (O_i^{\alpha r} O_j^{\alpha r} + O_i^{\beta r} O_j^{\beta r}) + J_{ij} O_i^{\alpha r} O_j^{\beta r} \right)$$

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 the Multi Electron Configuration Interaction

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^+ .

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

$$\langle Y_a | H | Y_b \rangle$$

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.

$Y_a = Y_b$. 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 y_i in Y_a and y_j in Y_b ; $Y_a = Y_b$. Assuming y_i and y_j to be a-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 y_i and y_j in Y_a and y_k and y_l in Y_b ; $Y_a = Y_b$. 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 a 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 reconstructed.

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 the [Table](#) will be obtained. These numbers of spin states will be obtained irrespective of the chemical nature of the system.

Table: States arising from $C.I.=n$

No. of M.O.s	States Arising from			States Arising from		
in MECI	Odd Electron Systems			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

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:

$$\begin{aligned} S_x a &= \frac{1}{2}b & S_x b &= \frac{1}{2}a \\ S_y a &= i\frac{1}{2}b & S_y b &= -i\frac{1}{2}a \\ S_z a &= \frac{1}{2}a & S_z b &= -\frac{1}{2}b \end{aligned}$$

Using these expressions, various useful identities can be established:

$$\begin{aligned} S^2 &= S_x^2 + S_y^2 + S_z^2 \\ I^+ &= (S_x + iS_y) & I^+ b &= a & I^+ a &= 0 \\ I^- &= (S_x - iS_y) & I^- a &= b & I^- b &= 0 \end{aligned}$$

$$\begin{aligned} S_x^2 + S_y^2 &= (I^+ I^-) + i(S_x S_y - S_y S_x) \\ &= (I^- I^+) + i(S_y S_x - S_x S_y) \\ &= \frac{1}{2}(I^+ I^- + I^- I^+) \end{aligned}$$

$$\text{and finally } i(S_y S_x - S_x S_y) = S_z.$$

For any microstate Y, the expectation value of the S^2 operator is given by

$$\langle S^2 \rangle = \langle Y | S_z^2 + S_y^2 + S_x^2 | Y \rangle$$

The first part of this expression is obvious, *vis*:

$$\langle Y | S_z^2 | Y \rangle = \frac{1}{4}(N^a + N^b)$$

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 a electrons, and q , the number of b 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 F , 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.

$Y_a = Y_b$: Since the two wavefunctions are the same, this corresponds to the expectation value of a microstate, and has already been derived.

2.

Except for y_i in Y_a and y_j in Y_b ; $Y_a = Y_b$: Assuming y_i and y_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:

$$\begin{aligned}\langle y_i y_j \rangle &= \langle m_i m_j \rangle = d(i, j) \\ \langle y_i y_k \rangle &= d(i, k) \\ \langle y_i y_l \rangle &= d(i, l)\end{aligned}$$

Therefore, $\langle Y_a / S^2 / Y_b \rangle = 0$.

3.

Except for y_i and y_j in Y_a and y_k and y_l in Y_b ; $Y_a = Y_b$. 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 a spin and two are of b 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 y_k and y_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 y' has the opposite spin to that of y .

Thus, only if y_i and y_j are spatially identical with y_k and y_l will $\langle Y_a / S^2 / Y_b \rangle$ 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 $\langle Y_a / S^2 / Y_b \rangle = 1$ if

$$(I^+ + I^-)(\psi_i + \psi_j) = (\psi_k + \psi_l), \text{ otherwise } \langle Y_a / S^2 / Y_b \rangle = 0.$$

4.

If more than two differences exist, $\langle Y_a / S^2 / Y_b \rangle = 0$.

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 not calculated in MOPAC.

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 p 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 semiemp

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 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. $(1/2,1/2,1/2)$, 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 the Table.

Table: Space-group operations for O_h^7 (diamond)

$\{E 000\}$	$\{C_2(1,1,0) 1/2,1/2,1/2\}$	$\{I 1/2,1/2,1/2\}$	$\{s_d(0,1,-1) 000\}$
$\{C_3(1,1,1) 000\}$	$\{C_2(1,0,1) 1/2,1/2,1/2\}$	$\{S_4(0,0,1) 000\}$	$\{S_6(1,1,1) 1/2,1/2,1/2\}$
$\{C_3(1,-1,-1) 000\}$	$\{C_2(0,1,1) 1/2,1/2,1/2\}$	$\{S_4(0,1,0) 000\}$	$\{S_6(1,-1,-1) 1/2,1/2,1/2\}$
$\{C_3(-1,1,-1) 000\}$	$\{C_2(0,1,-1) 1/2,1/2,1/2\}$	$\{S_4(1,0,0) 000\}$	$\{S_6(-1,1,-1) 1/2,1/2,1/2\}$
$\{C_3(-1,-1,1) 000\}$	$\{C_2(1,0,-1) 1/2,1/2,1/2\}$	$\{S_4^2(0,0,1) 000\}$	$\{S_6(-1,-1,1) 1/2,1/2,1/2\}$
$\{C_3^2(1,1,1) 000\}$	$\{C_2(1,-1,0) 1/2,1/2,1/2\}$	$\{S_4^2(0,1,0) 000\}$	$\{S_6^5(1,1,1) 1/2,1/2,1/2\}$
$\{C_3^2(1,-1,-1) 000\}$	$\{C_4(0,0,1) 1/2,1/2,1/2\}$	$\{S_4^2(1,0,0) 000\}$	$\{S_6^5(1,-1,-1) 1/2,1/2,1/2\}$
$\{C_3^2(-1,1,-1) 000\}$	$\{C_4(0,1,0) 1/2,1/2,1/2\}$	$\{s_d(1,1,0) 000\}$	$\{S_6^5(-1,1,-1) 1/2,1/2,1/2\}$
$\{C_3^2(-1,-1,1) 000\}$	$\{C_4(1,0,0) 1/2,1/2,1/2\}$	$\{s_d(1,0,1) 000\}$	$\{S_6^5(-1,-1,1) 1/2,1/2,1/2\}$
$\{C_2(0,0,1) 000\}$	$\{C_4^3(0,0,1) 1/2,1/2,1/2\}$	$\{s_d(0,1,1) 000\}$	$\{s_h(0,0,1) 1/2,1/2,1/2\}$
$\{C_2(0,1,0) 000\}$	$\{C_4^3(0,1,0) 1/2,1/2,1/2\}$	$\{s_d(1,-1,0) 000\}$	$\{s_h(0,1,0) 1/2,1/2,1/2\}$
$\{C_2(1,0,0) 000\}$	$\{C_4^3(1,0,0) 1/2,1/2,1/2\}$	$\{s_d(1,0,-1) 000\}$	$\{s_h(1,0,0) 1/2,1/2,1/2\}$

Calculation of spin-states

In order to calculate the spin-state, the expectation value of S^2 is calculated.

$$\begin{aligned} \langle \Phi_k | S^2 | \Phi_k \rangle &= S(S+1) = S_z^2 + 2I^+ I^- \\ &= \frac{1}{2}(p+q) - \\ &\quad \sum_i \left\{ C_{ik} C_{ik} \left((1/4)(N_i^\alpha - N_i^\beta)^2 + \sum_l O_{li}^\alpha O_{li}^\beta \right) + \sum_j 2 \left[C_{ik} C_{jk} [\delta(\Psi_i, (I^+ I^-) \Psi_j)] \right] \right\} \end{aligned}$$

where C_{ik} is the coefficient of microstate Y_i in State Φ_k , N_i^a is the number of alpha electrons in microstate Y_i , N_i^b is the number of beta electrons in microstate Y_i , O_{lk}^α is the occupancy of alpha M.O. l in microstate Y_k , O_{lk}^β is the occupancy of beta M.O. l in microstate Y_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.

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 the Figure.

Figure: 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

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 A₁ symmetry. The second state is a triplet, with energy 2.7109eV above the SCF energy, or 2.7158eV above the ground state, and has A₂ 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 if [ROOT=2](#) is used. See also [Oscillator Strength](#)

Franck-Condon considerations

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. `MECI C.I.=3 GNORM=0.01`.
- 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 DH_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 DH_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.

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](#). In MOPAC 6, 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), in which all calculations use the ground-state geometry and 1SCF.

Table : 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

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 CH₂O, and that increasing the C.I. lowers the energy steadily.

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 $\text{state}(1) = (1/2)^{1/2}(\text{Microstate}(a) + \text{Microstate}(b))$ is different from $\text{state}(2) = (1/2)^{1/2}(\text{Microstate}(a) - \text{Microstate}(b))$. 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.

EigenFollowing

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 [NONE](#). 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 [RMIN](#) and [RMAX](#) 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 [OMIN](#), 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 [OMIN](#) 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 [RMIN](#), [RMAX](#) and [OMIN](#) 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. [RMIN=-100](#) [RMAX=100](#) effectively inhibits step rejection. Similarly setting [OMIN=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 [RMIN](#), which for the default value of [RMIN=0](#) is equivalent to a lower energy. If the ratio is below [RMIN](#), the step is rejected, the trust radius reduced by a factor of two and a new step is predicted. The [RMIN](#), [RMAX](#) and [OMIN](#) features has been introduced in the current version of EF to improve the stability of TS optimizations. Setting [RMIN](#) and [RMAX](#) close to one will give a very stable, but also very slow, optimization. Wide limits on [RMIN](#) and [RMAX](#) 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 DDMIN, and the optimization terminates after only a few cycles with the "TRUST RADIUS BELOW DDMIN" warning long before the stationary point is reached. In such cases the user can add `DDMIN=0.0` and `RMIN` set to some negative value, say -10, thereby allowing steps which allow the energy to increase. An alternative is to use `LET DDMIN=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. `OMIN` 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 `OMIN`. 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 an 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 becomes 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 indicates 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 these cases there are no problems following high lying modes.

Unfortunately setting `RECALC=1` is very expensive in terms of computer time, but used in conjunction with `OMIN=0.90` (or possibly higher), and maybe also tighter limits on `RMIN` and `RMAX`, it represents an option of locating transition 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 refine 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.

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 $(a/4)d$ 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 $-(a/4)d$ is made. The energy is then re-evaluated. Given the three energies, a prediction is made as to the value of a 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 a 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 DH_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

$$x_1 = x_0 + 0.01 * \text{sign}(g_0)$$

from which a trial inverse Hessian can be constructed:

$$H_1(i,i) = 0.01 * \text{sign}(g_0(i)) / y_1(i).$$

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 the sign of $(g_0(i)) / y_1(i)$ 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.

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 DH_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.

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 DH_f decreases, the cycle is completed. If it does not drop, the step-size is reduced, and the DH_f recalculated. Only when the DH_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.

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., $n = 1/cr$. 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 n 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 m 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.

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

$$H_{i,j} = \frac{1}{2} \left(\frac{g_j^{+\delta_i} - g_j^{-\delta_i}}{2\delta} + \frac{g_i^{+\delta_j} - g_i^{-\delta_j}}{2\delta} \right).$$

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:

$$H_{i,j} = \frac{\left(\frac{\delta E}{\delta x_i}\right)_{+0.5\Delta x_j} - \left(\frac{\delta E}{\delta x_i}\right)_{-0.5\Delta x_j}}{\Delta x_j}.$$

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 $(1/2)^{1/2}$) by re-defining the Hessian as:

$$H_{i,j} = \frac{1}{2} \left(\frac{\left(\frac{\delta E}{\delta x_i} \right)_{+0.5\Delta x_j} - \left(\frac{\delta E}{\delta x_i} \right)_{-0.5\Delta x_j}}{\Delta x_j} + \frac{\left(\frac{\delta E}{\delta x_j} \right)_{+0.5\Delta x_i} - \left(\frac{\delta E}{\delta x_j} \right)_{-0.5\Delta x_i}}{\Delta x_i} \right).$$

A call to the energy - gradient function `COMPFG` will generate the gradients in kcal/mol/Ångstrom at a given geometry. These can then be converted into millidynes/Ångstrom (or 10⁵ dynes/cm) as follows:

$$H_{i,j}(\text{millidynes}/\text{\AA}) = 10^5 \frac{(\text{Kcal to ergs})}{(\text{\AA to cm})^2(\text{Mole to molecule})} H_{i,j}(\text{kcal/mol}/\text{\AA}^2)$$

or

$$H_{i,j}(\text{millidynes}/\text{\AA}) = 10^5 \frac{4.184 * 10^3 * 10^7}{(10^{-8*2})(6.022 * 10^{23})} H_{i,j}(\text{kcal/mol}/\text{\AA}^2).$$

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 10⁵.

Diagonalization of this matrix yields eigenvalues, e, which represent the quantities (k/m)^{1/2}, from which the vibrational frequencies can be calculated:

$$\bar{\nu}_i = \frac{1}{2\pi c} \sqrt{e_i}.$$

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 T_x , T_y , T_z , R_x , R_y and R_z (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.

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.

Results for the analysis of vibrations in the range $i50$ to $+50\text{ cm}^{-1}$, although printed, are not described accurately in the vibrational analysis, due to numerical difficulties.

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 2^{1/2} 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:

$$(\frac{M_B}{M_A + M_B})M_A;$$

likewise, for B :

$$(\frac{M_A}{M_A + M_B})M_B.$$

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:

$$\mu = \sum_A <\psi_A M_A \psi_A> \times <\psi_A | \psi_A> .$$

When written in this way, the quantum nature of the expression is obvious. However, for computational convenience, the effective mass is rewritten as:

$$\mu = \sum_A (\psi_A^2)^2 \times M_A$$

or

$$\mu = \sum_A (\sum_{i=1}^3 c_{A_i}^2)^2 M_A.$$

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.

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.

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:

$$E = \frac{1}{2}kx^2.$$

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.

Force Constants

Internal coordinate force constants, f_{ic} , can be derived from the Cartesian coordinates and the Cartesian force constant matrix by use of:

$$f_{ic}(l) = \sum_j \sum_k \frac{d(\mathbf{c.c.}_j)}{d(\mathbf{i.c.}_l)} \frac{d^2 E}{d(\mathbf{c.c.}_j) d(\mathbf{c.c.}_k)} \frac{d(\mathbf{c.c.}_k)}{d(\mathbf{i.c.}_l)}.$$

As with the gradients, the calculation of $\frac{d(\mathbf{c.c.}_j)}{d(\mathbf{i.c.}_l)}$ 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 [operating on the Hessian matrix](#).

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.

Transition dipole

In [FORCE](#) calculations, the transition dipole moment printed (T-DIPOLE) represents the electric dipole moment associated with the transition between the ground state and the vibrationally excited state. The units are Debye. To understand how it is calculated, consider a diatomic A-B, with $R_{A-B} = 2.0$ Ångstroms, and the partial charge on $Q_A = -Q_B = 0.5$ electrons. The ground-state dipole is $4.803 \cdot (R_{A-B}) \cdot (Q_A)$ or 4.803 Debye. If the atoms A or B have p-orbitals, then adjustments would need to be made for hybridization effects, these would be a few percent of the dipole terms arising from the partial charges.

During the calculation of the Hessian matrix, each atom is displaced in the X , Y , and Z directions. From these displacements, the derivative of the dipole with respect to motion in these directions, dD/dX_A , dD/dY_A , and dD/dZ_A , can readily be calculated. Later on, the normal modes are generated. The associated normalized eigenvectors, Y , represent the motion of the atoms in the normal modes. This allows the components of the transition dipole, d_{1n} , to be calculated using:

$$d_{1n}(X) = \sum_i S_i dD/dX_i Y(X)_i$$

and thus the transition dipole can be calculated:

$$d_{1n}(X) = (d_{1n}(X)^2 + d_{1n}(Y)^2 + d_{1n}(Z)^2)^{1/2}$$

Description of DRC/IRC

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.

IRC Reaction Coordinate Definition

The IRC is a purely geometric structure, that is, it is time-independent. A plot of the IRC has two axes: potential energy and reaction coordinate. While potential energy is easy to understand - it's just the heat of formation - the meaning of the reaction coordinate is not so obvious. To understand the meaning used here, consider the first half of a reaction, going from reactants to the transition state. In this discussion, it is a requirement that the Potential Energy Surface (PES) increases monotonically in going from reactant(s) to the transition state. If this requirement is not true, then the PES can be divided into two or more sections in such a manner that in each section the requirement is true. The discussion that follows would then apply to each section.

Both reactants and transition state have well-defined geometries. For the transition state, it is the stationary point on the PES which is the highest point on the minimum energy path from reactants to products. If the reactant is a single entity, e.g. the keto-enol tautomerization reaction, then it is that entity in its lowest energy configuration. If the reactant consists of two entities, e.g. the Diels Alder reaction or the $\text{Sn2 CH}_3\text{Cl} + \text{F}(-)$ reaction, then it is the associated complex. If no associated complex is formed, then it is the two isolated species.

The "length" of the reaction coordinate can be defined as follows: In the definition of the Cartesian geometry, let both the transition state and the reactants share the same center of mass. No rotation is permitted. Given that the coordinate of atom i in the reactant(s) is $R_i(x,y,z)$, and the corresponding coordinate in the products is $P_i(x,y,z)$, then the length, L , of the reaction coordinate is:

$$L = (S_i(R_i(x,y,z) - P_i(x,y,z))^2)^{1/2}$$

The reaction coordinate is represented by a set of points. The difference in geometry between each adjacent point on the reaction coordinate can be defined in a similar manner. By default, the [step-size](#) is 0.05 Angstroms, so the difference, Dl between the geometries of two points, $X(j)$ and $X(j+1)$, on a reaction coordinate is:

$$Dl = 0.05 = (S_i(X(j)_i(x,y,z) - X(j+1)_i(x,y,z))^2)^{1/2}$$

Note that the addition of all the step-sizes is *not* equal to the length of the reaction coordinate. It would only be equal if the reaction coordinate involved all atoms moving in straight lines; as this only occurs in diatomic reactions, this condition is normally not satisfied. In the output, the difference between each pair of points is a constant, when [X-PRIORITY](#) is in force, and the distance from the transition state is also printed.

DRC Reaction Coordinate Definition

Unlike the IRC, the default coordinate for the DRC is time, with a default [step size](#) of 0.1 fs. For long trajectories, particularly those of several picoseconds, a time step on 1 or more femtoseconds is suggested.

Equations used in the DRC and IRC

DRC

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:

$$V(i) = V(i) + V'(i)t + V''(i)t^2 + V'''(i)t^3$$

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 size of Δt , the time interval, starts off at 0.1fs ($1 \cdot 10^{-16}$ s), and changes depending on 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. The time interval is not printed, however typical values are between 0.1 and 1fs.

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 calculation of the IRC is that of a fully damped DRC. Within the calculation, a fictional time interval is used to determine the degree of motion, but at the end of each step, all velocity components are removed. The time interval, Δt , starts off at 0.1fs ($1 \cdot 10^{-16}$ s), and changes depending on 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. Typical values of Δt are between 0.1 and 1fs.

Specifying a temperature in a DRC

The behavior of a system at a specified temperature can be modeled by use of [KINETIC](#). This keyword allows extra kinetic energy to be added to the system. In order to determine how much extra energy to add, an understanding of the issues involved is essential.

Energy issues involved in a dynamic system

The total energy of a system can be expressed as three sums:

- (A) The heat of formation of the system, DH_{f0} . This is an irreducible minimum, and represents the energy of the system at equilibrium.
- (B) The potential energy of the system. This is the heat of formation of the system with its current geometry, DH_f , minus DH_{f0} . It represents the energy of distortion from the equilibrium geometry.
- (C) The kinetic energy of the system. This is the sum of the vibrational energies of motion of all the atoms in the system.

Energy term (A) is a constant, regardless of temperature. Energy terms (B) plus (C) represent the internal energy (enthalpy), U , of the system. At zero Kelvin, U is zero. At any other temperature, T , the enthalpy represents the integral of the heat capacity from zero to T . In a DRC calculation, any desired temperature can be specified by defining the associated internal energy. The internal energy can be calculated using [THERMO](#), and specifying the temperature to be used. In the output of a [THERMO](#) calculation, the enthalpy needed is given at the intersection of [ENTHALPY](#) and [VIB](#). In the following example, this would be 578.5 cal/mol, or 0.578 kcal/mol.

TEMP. (K)	PARTITION FUNCTION	H.O.F. KCAL/MOL	ENTHALPY CAL/MOLE	HEAT CAPACITY CAL/K/MOL	ENTROPY CAL/K/MOL
298.00	VIB.	0.2342D+01	578.4823	4.2357	3.6320
	ROT.	0.3897D+04	888.2813	2.9808	19.4109
	INT.	0.9125D+04	1466.7636	7.2165	23.0429
	TRA.	0.1515D+27	1480.4688	4.9680	36.0322
	TOT.		10.536	2947.2324	59.0751

To set up a run at a given temperature, the DH_f of the optimized system is needed. Calculate this first, then do a [THERMO](#) calculation to get the enthalpy at the desired temperature. The next step might be unexpected. Distort the geometry of the system slightly, and re-calculate the heat of formation at the distorted geometry. Make sure that it has increased by at least 0.2 kcal/mol, and preferably by a large fraction of the enthalpy. A non-equilibrium starting geometry is needed in a DRC calculation because otherwise the atoms would not be moving, and adding in excess kinetic energy would not be possible (You can't make the atoms move faster if they were not moving originally).

Work out how much extra kinetic energy would need to be added to equal the vibrational enthalpy at the desired temperature.

- Let the calculated equilibrium heat of formation be -100.000 kcal/mol.
- Let the desired internal vibrational energy be 0.578 kcal/mol.
- Let the heat of formation of the distorted geometry be -99.500 kcal/mol.

Then the extra kinetic energy needed would be 0.078 kcal/mol.

Set up the DRC calculation using the distorted geometry and `KINETIC=0.078`. Run this system. In the output, the starting heat of formation will be that of the distorted system, and the kinetic energy will be zero or almost zero. Over the next few femtoseconds, the heat of formation will become more negative, and the kinetic energy will rise. When it exceeds 0.2 kcal/mol, the velocity vector for the system will be well defined, and the extra kinetic energy will then be added. This will temporarily confuse the predictor-corrector error reduction function, so errors in the calculation will rise for a while, but once they become small the total energy of the system (heat of formation of

the current geometry plus kinetic energy) will be equal to DH_{f0} plus the desired enthalpy.

A legitimate question is, why do it this way, why not add the velocity vector "by hand"? The problem is, the kinetic energy must not include rotational or translational terms. Translational contributions to the enthalpy are irrelevant for a single molecule, and the rotational contributions are conserved, so they, too, can be ignored (or can they?), all that matters is the dynamics of the non-translating, non-rotating system. Defining a velocity vector that achieves this is definitely non-trivial.

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 calculation of the IRC is that of a fully damped DRC. Within the calculation, a fictional time interval is used to determine the degree of motion. This time interval, Dt , starts off at 0.1 fs (1×10^{-16} s), and changes depending on the factor needed to re-normalize the velocity vector. If it is significantly different from unity, Dt will be reduced; if it is very close to unity, Dt will be increased. Typical values of Dt are between 0.1 and 1 fs.

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 [IRC=*n*](#) and [RESTART](#), run the IRC calculation. If [RESTART](#) is specified with [IRC=*n*](#) then the restart is assumed to be from the FORCE calculation. If, in an [IRC](#) calculation, [RESTART](#) is specified, and [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 [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 [DRC=*n.nnn*](#), where *n.nnn* is the half-life in femtoseconds. If *n.nn* is 0.0 this corresponds to infinite damping simulating the [irc0](#). 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 bypassed.

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.

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>

For the IRC, the default is movement, i.e. `X-PRIO`, for the DRC, the default is `T-PRIO`.

By default, only the energies involved are printed (one line per point). To allow the geometry to be printed, [LARGE](#) is provided. 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 `irc0`.

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](#).
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.

To run a specific number of steps, use [CYCLES=*nnn*](#). This is the easiest way to limit the run.

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 1](#) 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.

Figure 1: Example of DRC calculation

```

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
  
```

Example 2: [Figure 2](#) 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 irc0. IRC on its own (without the "=*n*") implies an IRC calculation from the starting position--here the RESTART position--without initial displacement.

Figure 2: Example of IRC 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
  
```

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 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.

[IRC=1](#) [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.

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 1.103816Å, 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 the Table. A nitrogen molecule is thus a good approximation to a harmonic oscillator.

Table: Stretching Curve for Nitrogen Molecule

N-N DIST	ΔH_f	GRADIENT
(Ångstroms)	(kcal/mol)	(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

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, m , (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:

$$k = 2 \times 0.31755 / (0.0118)^2 \text{ (per mole)}$$

$$k = 4561.2 \text{ kcal/mol/\AA}^2 \text{ (per mole)}$$

$$k = 1.9084 \times 10^{30} \text{ ergs/cm}^2 \text{ (per mole)}$$

$$k = 31.69 \times 10^5 \text{ dynes/cm (per molecule)}$$

$$\text{(Experimentally, for N}_2\text{, } k = 23 \times 10^5 \text{ dynes/cm)}$$

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/Å² and a period of 12.185 fs.

3.

From the vibrational frequency. Given a "frequency" (wavenumber) of vibration of N₂ of $\tilde{\nu} = 2738.8$ cm⁻¹, the period of oscillation, in seconds, is given directly by:

$$r = \frac{1}{c\tilde{\nu}} = \frac{1}{2738.8 \times 2.998 \times 10^{10}},$$

or as 12.179 fs.

Summarizing, by three different methods the period of oscillation of N₂ is calculated to be 12.1851, 12.185 and 12.179 fs, average 12.183 fs.

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 $R(\text{N-N}) = 1.094$ 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 1.

Table 1: Velocities in DRC for N₂ Molecule

Time	Calculated	Linear	Diff. in
	Velocity	Velocity	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

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 2).

Table 2: Modified Velocities in DRC for N₂ Molecule

	Calculated	Simple Harmonic	Diff.
Time	Velocity	25325.Sin(0.5296t)	
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

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.

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 ([a qualification appears elsewhere](#)) .

Limitations on Point-Groups

- 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 SE_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.

Representation of Point Groups

The 57 groups recognized in MOPAC are given in Table 1.

Table 1: 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	

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 2. 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 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.

Table 2: Subset of Group D_{2h}

G	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

Each point-group is assumed to contain the totally symmetric representation, here A_{1g}. Operations are represented as 3x3 Euler matrices, thus C_{2x}, C_{2y} and C_{2z} would be represented as in the Figure All operations not given can be generated as products of operations already known, thus C_{2x} = C_{2y} × C_{2z}.

Figure: Representation of Symmetry Operations

$$\begin{array}{c} C_{2x}: \\ \begin{array}{ccc} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{array} \end{array}$$

$$\begin{array}{c} C_{2y}: \\ \begin{array}{ccc} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{array} \end{array}$$

$$\begin{array}{c} C_{2z}: \\ \begin{array}{ccc} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{array} \end{array}$$

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 D_2 to use the same character table.

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 R_3 . If two moments are zero, the system is either $C_{\infty v}$ or $D_{\infty h}$; the presence of a horizontal plane of symmetry distinguishes between them.

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 p orbitals point along the x axis.

Tolerance for Identification of Point-Groups

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 Å, 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 Å. Rotating the ring by 45 degrees (a C_8 operation) would place the atoms at a distance of only 0.18 Å 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.

Orientation of the Abelian Point 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 o

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.

Irreducible Representations of 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 c'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.

Irreducible Representations of 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.

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 a or b 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:

$$\langle \Psi_i | R | \Psi_j \rangle = \frac{1}{N!} \sum_{P=1}^{N!} (-1)^P P \left(\langle \prod_{k=1}^N \psi_k^i | R | \prod_{Q=1}^{N!} (-1)^Q Q \left(\prod_{l=1}^N \psi_l^j \right) \right) .$$

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. y_k in microstate Y_i with the result of operator R acting on M.O. y_k in microstate Y_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 a and b 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 $N=3$:

$$\langle \Psi_i | R | \Psi_j \rangle = \frac{1}{3!} \sum_{P=1}^{3!} \sum_{Q=1}^{3!} (-1)^P (-1)^Q$$

$$\begin{aligned} \langle \Psi_i | R | \Psi_j \rangle = & \langle \psi_1^i | R | \psi_1^j \rangle \langle \psi_2^i | R | \psi_2^j \rangle \langle \psi_3^i | R | \psi_3^j \rangle \\ & - \langle \psi_1^i | R | \psi_1^j \rangle \langle \psi_2^i | R | \psi_3^j \rangle \langle \psi_3^i | R | \psi_2^j \rangle \\ & - \langle \psi_1^i | R | \psi_2^j \rangle \langle \psi_2^i | R | \psi_1^j \rangle \langle \psi_3^i | R | \psi_3^j \rangle \\ & + \langle \psi_1^i | R | \psi_2^j \rangle \langle \psi_2^i | R | \psi_3^j \rangle \langle \psi_3^i | R | \psi_1^j \rangle \\ & + \langle \psi_1^i | R | \psi_3^j \rangle \langle \psi_2^i | R | \psi_1^j \rangle \langle \psi_3^i | R | \psi_2^j \rangle \\ & - \langle \psi_1^i | R | \psi_3^j \rangle \langle \psi_2^i | R | \psi_2^j \rangle \langle \psi_3^i | R | \psi_1^j \rangle \end{aligned}$$

or

$$\begin{aligned} \langle \Psi_i | R | \Psi_j \rangle = & \chi_{11}^{ij} \chi_{22}^{ij} \chi_{33}^{ij} + \chi_{12}^{ij} \chi_{23}^{ij} \chi_{31}^{ij} + \chi_{13}^{ij} \chi_{21}^{ij} \chi_{32}^{ij} \\ & - \chi_{11}^{ij} \chi_{23}^{ij} \chi_{32}^{ij} - \chi_{12}^{ij} \chi_{21}^{ij} \chi_{33}^{ij} - \chi_{13}^{ij} \chi_{22}^{ij} \chi_{31}^{ij} \end{aligned}$$

For higher numbers of electrons, the associated determinant is solved using standard methods.

The total character, $\langle \Psi_a | R | \Psi_a \rangle$, is obtained by multiplying the characters for the α and β parts together:

$$\langle \Psi_a | R | \Psi_a \rangle = \langle \Psi_a^\alpha | R | \Psi_a^\alpha \rangle \langle \Psi_a^\beta | R | \Psi_a^\beta \rangle.$$

If the positron equivalent is taken for only one set of electrons, e.g. either the a or the b 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: S, P, D, Q, and G. Even quite simple systems can achieve quite high angular momentum, thus acetylene, with a Σ_g^+ (the HOMO p and LUMO p*) will contain a 1G_g state, i.e., the angular momentum will be 4.

At present J-J coupling is not supported.

Symmetry of Atomic States

Atomic States have symbols such as $(^{2S+1})L_{M_S}$. Thus a vanadium atom, $3d^34s^2$ would give rise to the states: 4P , 4D , 4F , 4G , 4H , 6S , and 6D . Some of these States occur more than once, for example the 4P State occurs four times. All components of a single State that have the same spin, S , and angular momentum, L , are degenerate. Following Griffith, "The Theory of Transition Metal Ions", the L orbital angular momenta values are represented in MOPAC by the letters S, P, D, F, G, H, I, K, L, M corresponding to the angular momenta $L = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$. Note that symbol "J" is missing, this is because it is used extensively in other aspects of atomic theory.

All atomic States that can exist for a $s-p-d$ basis set with any allowed number of electrons are automatically identified. By specifying the component of spin, using $M_S=n.n$, individual magnetic components within a manifold such as 6S can be identified.

Atomic theory is both beautiful and complicated. Simple spin-spin and orbit-orbit coupling, i.e., Russell-Saunders or L-S coupling, can be modeled, but relativistic coupling ($j-j$ coupling) cannot be modeled, due the the absence of the spin-orbit interaction terms. Despite the lack of spin-orbit interactions, a lot of elegant symmetry work can still be done. One of the most beautiful sections of atomic symmetry theory involves the coupling of two space angular momenta. Coupling coefficients involve the Wigner or Clebsch-Gordon $3j$ symbol $\langle L_1 L_2 m_1 m_2 | L_1 L_2 L m \rangle$ which equals:

$$\frac{\{(L+m)!(L-m)!(L_1-m_1)!(L_2-m_2)!(L_1+L_2-L)!(2L+1)/((L_1+m_1)!(L_2+m_2)!(L_1-L_2+L)!(L_2-L_1+L)!(L_1+L_2+L+1)!\}^{0.5}}{\times d(m,m_1+m_2)S_r(-1)^{(L_1+r-m_1)}\{((L_1+m_1+r)!(L_2+L-r-m_1)!)/(r!(L-m-r)!(L_1-m_1-r)!(L_2-L+m_1+r)!)\}}$$

with $0! = 1$. All aspects of this can be modeled. For users of MOPAC who are brave enough to explore this topic, two tools are provided:

A [set of Wigner coefficients](#) for all values of l_1 and l_2 up to 6. This is a large file, over 1Mb. Each coefficient is expressed four ways:

- (A) As a decimal.
- (B) As a rational fraction.
- (C) As a rational fraction with a common denominator
- (D) As a rational fraction expressed as powers of prime numbers, thus the coefficient for the most complicated term, $\langle 6,4,5,-3 | 6,5,9,1 \rangle$, is $(367^2)/(2^2.3.5.7.11.13.17)$

A FORTRAN 90 [source code](#) to generate these coefficients.

An illustrative exercise in atomic theory is to use the shift [step-down operators](#), I_- , to move through the associated Hilbert spin-space, e.g., to go from $6S$ with $M_S = 5/2$ to $M_S = -5/2$, then show that the State is annihilated on application of the step-down operator once more. Moving through the orbital angular momentum space is tedious, and definitely not worth doing.

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 the Table. 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 DHf 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, 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.

Table: Fundamental Physical Constants

Physical Constant	Symbol	Value	Units
Speed of Light	c	299 792 458	m sec ⁻¹ (Definition)
Planck constant	<i>h</i>	6.626 075 5(40) x10 ⁻³⁴	J sec
		6.626 075 5(40) x10 ⁻²⁷	erg s
Avogadro constant	<i>N</i>	6.022 136 7(36)x10 ²³	mol ⁻¹
Molar gas constant	<i>R</i>	1.987 215 6	cal/mol/degree
		8.314 510(70)	J/mol/K
Volume of 1 mol of gas	<i>V</i> ₀	22.414 10(19)	l/mol (at 1 atm, 25 C)
Electron volt	eV	1.602 177 33(49)x10 ⁻¹⁹	J
Electron charge	e	1.602 177 33(49)x10 ⁻¹⁹	C
Hartree	<i>E_h</i>	27.211 396 1(81)	eV
Electrostatic energy	<i>E_ha_o</i>	14.399 651 782 565	eV
Bohr radius	<i>a_o</i>	0.529 177 249(24)x10 ⁻¹⁰	m
Boltzmann constant	<i>k = R/N</i>	1.380 658(12)x10 ⁻²³	J/K
		1.380 658(12)x10 ⁻¹⁶	erg/K
pi	p	3.141 592 653 589 79	
Joule	J/cal	4.184	J/cal (Definition)
cm ⁻¹	<i>hc/eV</i>	1.239 842 4x10 ⁻⁴	eV
cm ⁻¹	<i>hcN/(1000J/cal)</i>	2.859 144x10 ⁻³	kcal/mol
cm ⁻¹	<i>hcN/(J/cal)</i>	2.859 144	cal/mol
cm ⁻¹		1.196 266x10 ⁸	erg = dyne Å ⁻¹
Atomic unit (a.u.)		8.657 10x10 ⁻³³	e.s.u.
a.u.		2.541 747 8x10 ⁻⁴⁰	Debye
a.u.		51.422 082	V m ⁻¹
kcal/mol		6.947 700x10 ⁻¹³	erg
1 J		1.x10 ⁷	erg
1 eV		23.060 542 301 389	kcal/mol
1a.u.		627.509 6	kcal/mol
1 atm		1.013 25x10 ⁵	Pa
		1.013 25x10 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.

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*.

The only alternative way to calculate the I.P. is to calculate the DH_f of the parent species, then, without allowing the geometry to relax, calculate the DH_f of the ionized system. The difference in DH_f , in kcal.mol^{-1} , divided by 23.06, is the predicted I.P., in eV.

Dipole moments.

For neutral systems, the dipole moment is calculated from the atomic charges and the lone-pairs as

$$\mu_x = cC \sum_A Q_A x_A + cCa_o2 \sum_A P(s - p_x)_A D_1(A)$$

$$\mu_y = cC \sum_A Q_A y_A + cCa_o2 \sum_A P(s - p_y)_A D_1(A)$$

$$\mu_z = cC \sum_A Q_A z_A + cCa_o2 \sum_A P(s - p_z)_A D_1(A)$$

$$\mu = \mu_x + \mu_y + \mu_z$$

Where c = speed of light, C = charge on the electron, and a_o = Bohr radius, or $cC = 2.99792458*1.60217733 = 4.8032066$, and $cCa_o2 = 2.99792458*1.60217733 = 4.8032066*0.529177249*2.0 = 5.0834948$.. [D1\(A\)](#) is defined elsewhere.

Conversion factors between Dipole units

Quantity	Dipole as	Factor (units)	Factor(value)	System
1 electron*1 Angstrom	Debye (D)	cC	4.80320	esu
1 Debye (D)	Dipole Length (m)	1/(cC)	2.0819*10 ⁻¹¹	SI

$$1 \text{ statvolt} = 299.79 \text{ volts} = c*10^{-6} \text{ volts (c in m/s)}$$

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:

$$\mu_x = cC \sum_A Q_A (x_A - x_{cog}) + cCa_o2 \sum_A P(s - p_x)_A D(A)$$

$$\mu_y = cC \sum_A Q_A (y_A - y_{cog}) + cCa_o2 \sum_A P(s - p_y)_A D(A)$$

$$\mu_z = cC \sum_A Q_A (z_A - z_{cog}) + cCa_o2 \sum_A P(s - p_z)_A D(A)$$

$$\mu = \mu_x + \mu_y + \mu_z,$$

where x_{cog} is the x -coordinate of the center of gravity of the system

$$x_{cog} = \sum_A M_A x_A,$$

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.

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.

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. As with the Coulson population, the unit of the Mulliken population analysis is the electron. For the hydrogen atom, the P_{ss} population would be 1.00.

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 the Table.

Table: Comparison of Observed and Calculated Pyramidalization of Nitrogen

Compound	MNDO	AM1	PM3	PM6	Exp
Ammonia	Py	Py	Py		Py
Aniline	Py	Py	Py		Py
Formamide	Py	Flat	Py		Flat
Acetamide	Py	Flat	Py		Flat
N-methyl formamide	Py	Flat	Py		Flat
N-methyl acetamide	Flat	Flat	Py		Flat
<i>k</i>	6.17	3.32	7.19		

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. Values of k in kcal/mol are also given in the Table. 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.

"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".

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.

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.

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.

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 1.

Figure 1: Example of an S_N2 reaction path calculation

```

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
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

```

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 2. 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.

Figure 2: Example of a rotation barrier 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
3 1 4 5 6 7 8
3 2 4 5 6 7 8
6 7 7
6 11 8

```

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 3](#).

Figure 3: Data set to stretch a polymer

```

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

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.

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_N2 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,

$$S_i<y_i^4>$$

is a maximum. Since

$$S_iS_j<y_i^2><y_i^2>$$

is a constant,

$$S_iS_{j<i}<y_i^2><y_i^2>$$

must be a minimum.

The operation to localize M.O. consists of a series of binary unitary transforms of the type:

$$|\psi_i>=a|\psi_k>+b|\psi_l>$$

$$|\psi_j>=-b|\psi_k>+a|\psi_l>$$

where $|y_k>$ and $|y_l>$ are normal M.O.s, and $|y_i>$ and $|y_j>$ are the LMOs.

The ratio a/b is given by

$$a/b=\frac{1}{4}\arctan\left(\frac{4(<\psi_k\psi_l^3>-<\psi_k^3\psi_l>)}{<\psi_k^4>+<\psi_l^4>-6<\psi_k^2\psi_l^2>}\right)$$

Note that in normal semiempirical work: $<\phi_\lambda|\phi_\sigma>=\delta(\lambda,\sigma)$.

From this it follows that, given $\psi_k=\sum_\lambda C_{\lambda k}\phi_\lambda$,

$$<\psi_k\psi_l^3>=\sum_\lambda C_{\lambda k}C_{\lambda l}^3$$

In order to preserve rotational invariance, all contributions on each atom must be added together. This gives:

$$<\psi_k^4>=\sum_A(\sum_{\lambda\in A}C_{\lambda k}^2)^2$$

$$\langle \psi_k^3 \psi_l \rangle = \sum_A \left(\sum_{\lambda \in A} C_{\lambda k}^2 \right) \sum_{\lambda \in A} C_{\lambda k} C_{\lambda l}$$

$$\langle \psi_k^2 \psi_l^2 \rangle = \sum_A \left(\sum_{\lambda \in A} C_{\lambda k}^2 \right) \left(\sum_{\lambda \in A} C_{\lambda l}^2 \right)$$

Oscillator Strength

The intensity of a UV-Visible absorption band is a function of the oscillator strength and of the energy of the absorption band. Both quantities are printed when a C.I. calculation is done, and MECI is used. In the output, the oscillator strength, in electron-Angstroms is printed under the heading "POLARIZATION".

Absorption intensity is proportional to the energy of the transition and to the square of the oscillator strength, i.e., the absorption intensity is proportional to: $n[(m_x)^2 + (m_y)^2 + (m_z)^2]$.

Theory

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, k , is given (Wilson Decius and Cross, "Molecular Vibrations", p 163, McGraw-Hill (1955)) in terms of the oscillator integral:

$$\langle \Phi_{ground} | \vec{r} | \Phi_{excited} \rangle, \quad 1$$

by

$$\kappa = \frac{8\pi^3}{3ch} \nu_{n'n''} (N_{n'} - N_{n''}) \langle \Phi_{ground} | \vec{r} | \Phi_{excited} \rangle^2.$$

For electronic photoexcitations, Φ are state functions:

$$\Phi = \sum_i c_i \Psi_i,$$

and the Ψ are [microstates](#).

In order to evaluate 1, 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

$$\langle \psi_j | \vec{r} | \psi_i \rangle = - \langle \psi_i | \vec{r} | \psi_j \rangle .$$

The starting point for evaluating 1 is to calculate the molecular orbital oscillator:

$$\langle \psi_i | \vec{r} | \psi_j \rangle .$$

In order to solve this integral, the following approximations will be made:

$$\langle \phi_\lambda | \vec{r} | \phi_\sigma \rangle = 0$$

and

$$\langle \phi_\lambda | \vec{r} | \phi_\lambda \rangle = r.$$

The operator \vec{r} can be expanded into three terms:

$$\vec{r} = \vec{x} + \vec{y} + \vec{z},$$

where \vec{x} , \vec{y} and \vec{z} are Cartesian position operators. For each atomic orbital, the position would be that of the nucleus.

The zero or origin of the position operators is not immediately obvious. In general it would *not* be the origin of the Cartesian coordinate system used. The origin of the position operator would depend on the individual M.O. and would be defined so that:

$$\langle \psi_i | \vec{x} | \psi_i \rangle = \langle \psi_i | \vec{y} | \psi_i \rangle = \langle \psi_i | \vec{z} | \psi_i \rangle = 0.$$

To determine the origin, the position operators are replaced by $\vec{x} + \vec{x}_0$, where \vec{x}_0 is the distance from the origin of the Cartesian coordinate system used to the origin of the position operator. Then:

$$\langle \psi_i | \vec{x} + \vec{x}_0 | \psi_i \rangle = 0,$$

from which it follows that:

$$\langle \psi_i | \vec{x}_0 | \psi_i \rangle = - \langle \psi_i | \vec{x} | \psi_i \rangle .$$

The integrals over molecular orbitals are evaluated by first calculating $- \langle \psi_i | \vec{x} | \psi_i \rangle$

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. Not used by [MOZYME](#)

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. Not used by [MOZYME](#)

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. Not used by [MOZYME](#)

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`) Not used by [MOZYME](#)

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. Not used by [MOZYME](#)

7. In a [MOZYME](#) calculation, damp SCF oscillations using `DAMP=n.nn`, where $n.nn$ should be in the range $1.0 < n.nn < 0.0$. Values in the region of 0.5 are usually successful. For solids, the default factor is 0.5, so $n.nn$ should be less than 0.5.

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.

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.

To a degree, configuration interaction can correct this picture. When $\text{C.I.}=2$ is used, Figure 1 and Figure 2, 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.

Figure 1: Effect of C.I. on ΔH_f in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)

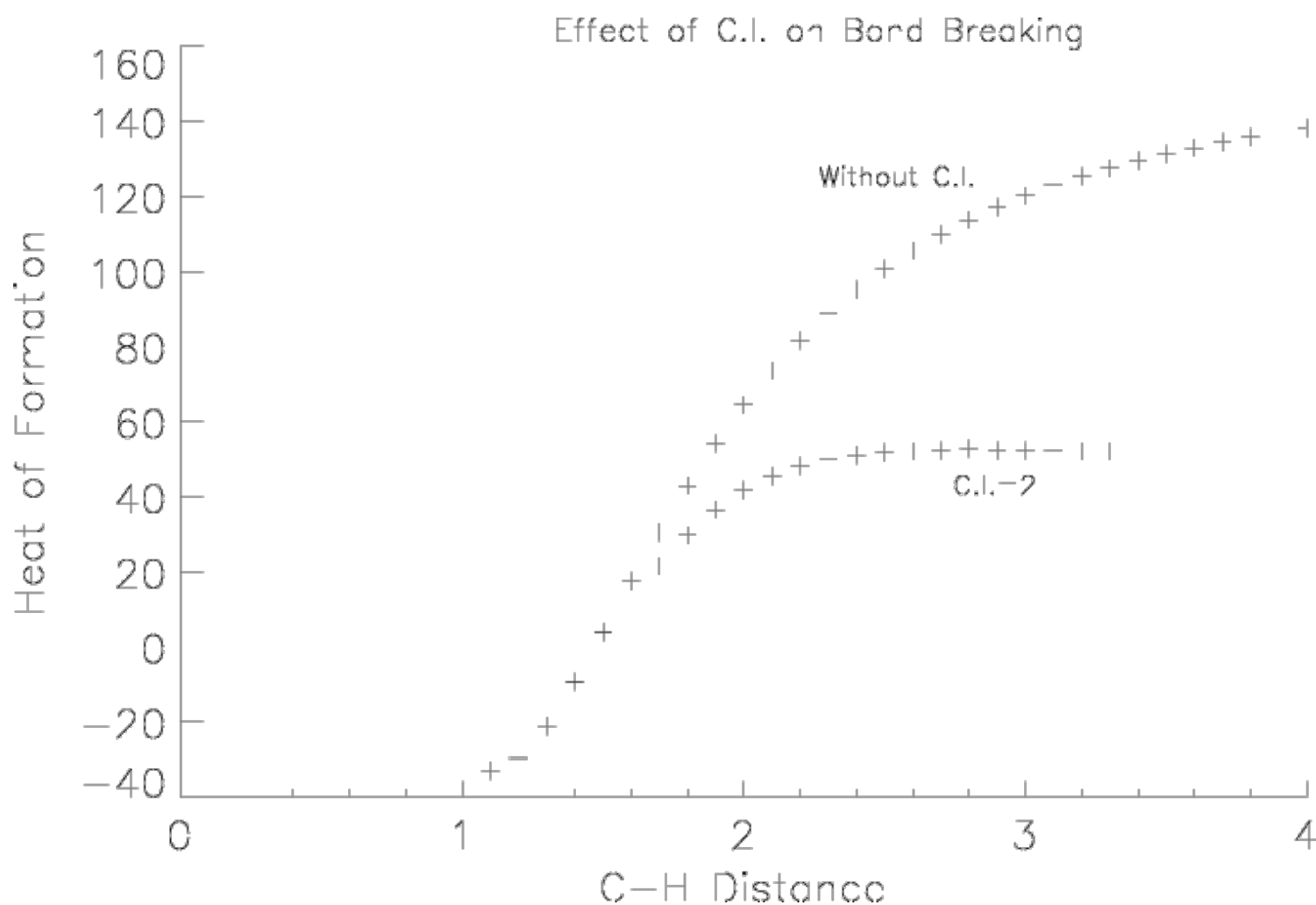
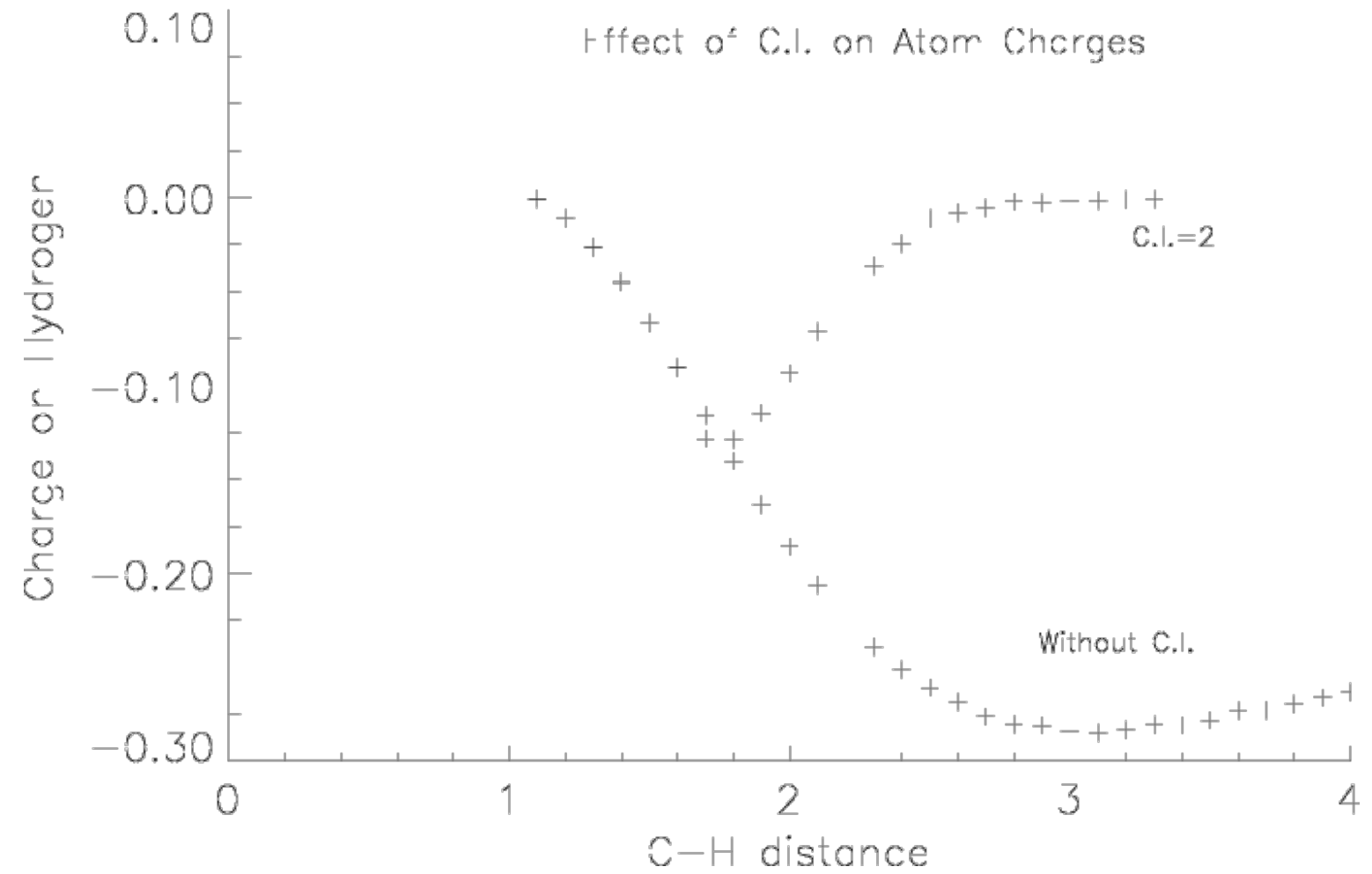


Figure 2: Effect of C.I. on the charge in a bond-breaking reaction (for the reaction $\text{CH}_2\text{O} \rightarrow \text{CHO} + \text{H}$)



Sparkles

Eight extra "elements" have been put into MOPAC. These represent pure ionic charges, roughly equivalent to the following chemical entities:

Chemical Symbol	Equivalent to
Fr	Half plus cation, for use with transition metal complexes. Example: To have $[\text{Ti}^{\text{III}}\text{F}_6]^{3-}$ in a more realistic environment, while still having it octahedral, use: $[\text{Ti}^{\text{III}}\text{F}_6\text{Fr}_8]^+$.
+	Tetramethyl ammonium, Potassium or Cesium cation + electron
++	Barium di-cation + 2 electrons
+3	Al^{3+} , a tri-cation, a generic transition metal ion, M^{3+} .
At	Half minus anion, for use with transition metal complexes. Example: To have $[\text{Ti}^{\text{III}}(\text{H}_2\text{O})_6]^{3+}$ in a more realistic environment, while still having it octahedral, use: $[\text{Ti}^{\text{III}}(\text{H}_2\text{O})_6\text{At}_8]^-$.
-	Borohydride Halogen, or Nitrate anion minus electron
--	Sulfate, oxalate di-anion minus 2 electrons
-3	PO_4^{3-} , a tri-anionic ion, X^{3-} .

For the purposes of discussion these entities are called 'sparkles': the name arises from consideration of their behavior.

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: $\text{CH}_2\text{-CH}_2$, $\text{CH}_2\text{-NH}$, 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 b-value. The b value is used in the calculation of the one-electron two-center integral. Because the b 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 $(1/2)^{1/2}(\text{Cb}+\text{A})$. 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 $(1/2)^{1/2}(\text{Cb}-\text{A})$, 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 b-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](#).

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

$$\begin{aligned} \varphi = & \frac{(2\xi)^{n+1/2}}{(2n)!^{1/2}} \left[\frac{(2l+1)(l-m)!}{2(l+m)!} \right]^{1/2} \left[\frac{\sin^m \theta}{2^l} r^{n-1} e^{-\xi r} \frac{e^{im\phi}}{(2\pi)^{1/2}} \right] \\ & \sum_{j=0}^{1/2(l-m)} \frac{(2(l-j))! (-1)^j}{j! (l-j)! (l-m-2j)!} (\cos \theta)^{l-m-2j}. \end{aligned}$$

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

$$\begin{aligned} < \varphi_a \varphi_b > = \frac{(2\xi_a)^{na+1/2}(2\xi_b)^{nb+1/2}}{((2n_a)!(2n_b)!)^{1/2}} \frac{[(2l_a+1)(2l_b+1)]^{1/2}}{2} \\ &\int_0^\infty sin^m\theta_a sin^m\theta_b r_a^{na-1} r_b^{nb-1} e^{-r_a\xi_a} e^{-r_b\xi_b} \frac{e^{im\phi} e^{im\phi*}}{2\pi} \\ &\sum_{j_a=0}^{1/2(l_a-m)} C_{i_a m j_a} (cos\theta_a)^{l_a-m-2j_a} \sum_{j_b=0}^{1/2(l_b-m)} C_{i_b m j_b} (cos\theta_b)^{l_b-m-2j_b} dv \end{aligned}$$

It is impractical to solve this integral using polar coordinates. Instead, a prolate spheroidal coordinate system is used. Using the identities:

$$\begin{aligned} r_a &= \frac{R(\mu+v)}{2}; \quad cos\theta_a = \frac{(1+\mu v)}{(\mu+v)}; \quad sin\theta_a = \frac{((\mu^2-1)(1-v^2))^{1/2}}{(\mu+v)} \\ r_b &= \frac{R(\mu-v)}{2}; \quad cos\theta_b = \frac{(1-\mu v)}{(\mu-v)}; \quad sin\theta_b = \frac{((\mu^2-1)(1-v^2))^{1/2}}{(\mu-v)}, \end{aligned}$$

this gives $dv = \frac{R^3}{8}(\mu + v)(\mu - v)d\mu dv d\phi.$

Substituting these identities into the previous expression we get:

$$\begin{aligned} < \varphi_a \varphi_b > &= \sum_{j=0}^{1/2(l-m)} \frac{(2(l-j))! (-1)^j}{j! (l-j)! (l-m-2j)!} (cos\theta)^{l-m-2j}. \\ &\frac{((\mu^2-1)(1-v^2))^m R^{na-1}}{(\mu+v)^m (\mu-v)^m} \frac{R^{nb-1}}{2^{na-1}} (\mu+v)^{na-1} \frac{R^{nb-1}}{2^{nb-1}} (\mu-v)^{nb-1} \end{aligned}$$

$$\frac{e^{-R\xi_a(\mu+\nu)/2}e^{-R\xi_b(\mu-\nu)/2}}{2\pi}$$

$$\sum_{j_a=0}^{(l_a-m)/2}\sum_{j_b=0}^{(l_b-m)/2}C_{l_a m j_a}C_{l_b m j_b}\frac{(1+\mu\nu)^{l_a-m-2j_a}(1-\mu\nu)^{l_b-m-2j_b}}{(\mu+\nu)^{l_a-m-2j_a}(\mu-\nu)^{l_b-m-2j_b}}$$

$$\frac{R^3}{8}(\mu+\nu)(\mu-\nu)d\mu d\nu d\phi,$$

which, on integrating over and rearranging, gives:

$$<\varphi_a\varphi_b>=\int_{-1}^1\int_1^\infty\frac{\xi_a^{n_a+1/2}\xi_b^{n_b+1/2}}{2}\left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!}\right]^{1/2}R^{n_a+n_b+1}$$

$$\sum_{j_a=0}^{(l_a-m)/2}\sum_{j_b=0}^{(l_b-m)/2}C_{l_a m j_a}C_{l_b m j_b}(\mu^2-1)^m(1-\nu^2)^m(\mu+\nu)^{n_a-l_a+2j_a}$$

$$(\mu-\nu)^{n_b-l_b+2j_b}(1+\mu\nu)^{l_a-m-2j_a}(1-\mu\nu)^{l_b-m-2j_b}e^{-R\xi_a(\mu+\nu)/2}e^{-R\xi_b(\mu-\nu)/2}d\mu d\nu.$$

This is a product of six simple expressions of type $(a+b)^n$. Expanding each term as a binomial generates six summations:

$$<\varphi_a\varphi_b>=\int_{-1}^1\int_1^\infty\frac{\xi_a^{n_a+1/2}\xi_b^{n_b+1/2}}{2}\left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!}\right]^{1/2}R^{n_a+n_b+1}$$

$$\sum_{j_a=0}^{(l_a-m)/2}\sum_{j_b=0}^{(l_b-m)/2}C_{l_a m j_a}C_{l_b m j_b}\sum_{k_a=0}^m\sum_{k_b=0}^m\sum_{P_a}^{n_a-l_a+2j_a}\sum_{P_b}^{n_b-l_b+2j_b}\sum_{q_a}^{l_a-m-2j_a}\sum_{q_b}^{l_b-m-2j_b}$$

$$\frac{(l_b-m-2j_b)!}{(l_b-m-2j_b-q_b)!q_b!}\frac{(l_a-m-2j_a)!}{(l_a-m-2j_a-q_a)!q_a!}\frac{(n_b-l_b+2j_b)!}{(n_b-l_b+2j_b-P_b)!P_b!}$$

$$\frac{(n_a-l_a+2j_a)!}{(n_a-l_a+2j_a-P_a)!P_a!}\frac{m!^2}{(m-k_b)!k_b!(m-k_a)!k_a!}$$

Using integration by parts, and making use of the following integrals (The "A" and "B" overlap integrals):

$$\int_1^\infty x^n e^{-ax}dx=e^{-a}\sum_{\mu=1}^{n+1}\frac{n!}{a^\mu(n-\mu+1)}=A_n(a)$$

$$\int_{-1}^1 x^n e^{-ax}dx=-e^{-a}\sum_{\mu=1}^{n+1}\frac{n!}{a^\mu(n-\mu+1)}-e^a\sum_{\mu=1}^{n+1}\frac{n(-1)^{n-\mu}}{a^\mu(n-\mu+1)!}=B_n(a),$$

the overlap integral becomes

$$\begin{aligned}
 \langle \varphi_a \varphi_b \rangle &= \frac{\xi^{n_a+1/2} \xi^{n_b+1/2}}{2} \left[\frac{(2l_a+1)(2l_b+1)}{(2n_a)!(2n_b)!} \right]^{1/2} R^{n_a+n_b+1} \\
 &\sum_{j_a=0}^{(l_a-m)/2} C_{l_a m j_a} \sum_{j_b=0}^{(l_b-m)/2} C_{l_b m j_b} \sum_{k_a=0}^m \sum_{k_b=0}^m \sum_{P_a=0}^{n_a-l_a+2j_a} \sum_{P_b=0}^{n_b-l_b+2j_b} \sum_{q_a=0}^{l_a-m-2j_a} \sum_{q_b=0}^{l_b-m-2j_b} \\
 &\frac{(l_b-m-2j_b)!}{(l_b-m-2j_b-q_b)!q_b!} \frac{(l_a-m-2j_a)!}{(l_a-m-2j_a-q_a)!q_a!} \frac{(n_b-l_b+2j_b)!}{(n_b-l_b+2j_b-P_b)!P_b!} \\
 &\frac{(n_a-l_a+2j_a)!}{(n_a-l_a+2j_a-P_a)!P_a!} \frac{m!^2}{(m-k_b)!k_b!(m-k_a)!k_a!} (-1)^{k_a+k_b+m+P_b+q_b} \\
 &B_{2k_a+P_a+P_b+q_a+q_b} \left(\frac{R(\xi_a - \xi_b)}{2} \right) A_{2k_b+n_a-l_a+2j_a+n_b-l_b+2j_b-P_a-P_b+q_a+q_b} \\
 &\left(\frac{R(\xi_a + \xi_b)}{2} \right) d\mu \, dv,
 \end{aligned}$$

in which the coefficients C_{lmj} have the numerical values given in the Table.

Table: 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. This is the most convenient form for algorithmic use. In this form, the overlap integral can be found in function SS.

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:

$$E_{isol}(A) = E_{\text{neutral atom}}(A) - E_{\text{nucleus}}(A) - E_{\text{valence electrons}}(A)$$

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 $E_{\text{neutral atom}}$.

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:

$$E_{isol} = aU_{ss} + bU_{pp} + (a-1)G_{ss} + a.bG_{sp} + (b(b-1))/2 G_p^2 - bH_{sp} - cH_{pp}$$

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 $1/2(G_{pp} - G_p^2)$, the expression for systems with 2 to 4 p electrons is recast as:

$$E_{isol} = aU_{ss} + bU_{pp} + (a-1)G_{ss} + a.bG_{sp} + ((b(b-1))/2 + c/2)G_p^2 - (a-1)bH_{sp} - c/2G_{pp},$$

or

$$E_{isol} = 2U_{ss} + bU_{pp} + G_{ss} + 2.bG_{sp} + ((b(b-1))/2 + c/2)G_p^2 - (a-1)bH_{sp} - c/2G_{pp}.$$

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 $s^m d^n$, in which there are m_a s -electrons and m_b b s -electrons, and n_a d -electrons and n_b b d -electrons, and the total angular quantum number is L , is:

$$E_{isol} = mU_{ss} + nU_{dd} + (m(m-1))/2 G_{ss} + m.nG_{sd} - (m_a n_a + m_b n_b)H_{sd} \\ + (n(n-1))/2 \frac{G_{dd}^0}{5} + (-4(n_a^2 + n_b^2) + 13n - 3/2(L(L+1))) \frac{G_{dd}^2}{49}$$

$$+(-(n_a^2+n_b^2)/2-9/2n+5/6(L(L+1)))\frac{G_{dd}^4}{49}.$$

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:

$$<^{n+1}L|H|^{n+1}L>= \frac{1}{2}n(n-1)(A-8B)+\frac{3}{2}[6n-L(L+1)]B.$$

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:

$$\begin{aligned} G_{dd}^0 &= F_0 = F^0 \\ G_{dd}^2 &= F_2 = F^2/49 \\ G_{dd}^4 &= F_4 = F^4/441 \end{aligned}$$

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 the [Table](#). 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

Note also that there are no elements with both p and d valence electrons, therefore terms of the type G_{pd} are not necessary.

Table: Two Electron Energy Contributions to EISOL for Atoms in their Ground States

Element		State		G_{ss}	G_{sp}	H_{sp}	G_{pp}	G_{p2}	G_{sd}	H_{sd}	G_{dd}^0	G_{dd}^2	G_{dd}^4
		Config.	Mult.:	1	1	-1	-1/2	1/2	1	-1/5	1	-1/49	-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	$2s^22p^1$	$2P$		1	2	1							
6	C	$2s^22p^2$	$3P$		1	4	2	1	3					
7	N	$2s^22p^3$	$4S$		1	6	3	3	9					
8	O	$2s^22p^4$	$3P$		1	8	4	1	13					
9	F	$2s^22p^5$	$2P$		1	10	5		20					
10	Ne	$2s^22p^6$	$1S$		1	12	6		30					
11	Na	$3s^1$	$2S$											
12	Mg	$3s^2$	$1S$		1									
13	Al	$3s^23p^1$	$2P$		1	2	1							
14	Si	$3s^23p^2$	$3P$		1	4	2	1	3					
15	P	$3s^23p^3$	$4S$		1	6	3	3	9					
16	S	$3s^23p^4$	$3P$		1	8	4	1	13					
17	Cl	$3s^23p^5$	$2P$		1	10	5		20					
18	Ar	$3s^23p^6$	$1S$		1	12	6		30					
19	K	$4s^1$	$2S$											
20	Ca	$4s^2$	$1S$		1									
21	Sc	$4s^23d^1$	$2D$		1					2	1			
22	Ti	$4s^23d^2$	$3F$		1					4	2	1	8	1
23	V	$4s^23d^3$	$4F$		1					6	3	3	15	8
24	Cr	$4s^13d^5$	$7S$							5	5	10	35	35
25	Mn	$4s^23d^5$	$6S$		1					10	5	10	35	35
26	Fe	$4s^23d^6$	$5D$		1					12	6	15	35	35
27	Co	$4s^23d^7$	$4F$		1					14	7	21	43	36
28	Ni	$4s^23d^8$	$3F$		1					16	8	28	50	43
29	Cu	$4s^13d^{10}$	$2S$							10	5	45	70	70
30	Zn	$4s^2$	$1S$		1									
31	Ga	$4s^24p^1$	$2P$		1	2	1							
32	Ge	$4s^24p^2$	$3P$		1	4	2	1	3					
33	As	$4s^24p^3$	$4S$		1	6	3	3	9					
34	Se	$4s^24p^4$	$3P$		1	8	4	1	13					
35	Br	$4s^24p^5$	$2P$		1	10	5		20					
		2 6												

36	Kr	$4s\ 4p$	$1S$		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
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									
57	La	$6s^2 5d^1 *$	$^2D^*$		1					2	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
74	W	$6s^1 5d^5$	$7S^{**}$							5	5	10	35	35
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	$2\ 2$	$3P$		1	4	2	1	3					

		6s 6p												
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					

* The correct ground state of lanthanum is $6s^2 4f^1, ^2F(u)$, but for the methods in MOPAC, methods that do not use "f" orbitals, the low-lying excited state $6s^2 5d^1, ^2D(g)$, is used instead.

** The lowest state of tungsten is 5D_0 , but the lowest state ignoring spin-orbit coupling is 7S_3 .
MOPAC does not use relativistic terms, so spin-orbit splitting is not considered.
Because of this, the ground state used in MOPAC for W is 7S .

Polarizability and Hyperpolarizability Calculation

Units for the series expansion "charge - dipole - alpha - beta - gamma" are as follows:

Charge: positrons (Charge on a fluoride ion is "-1", on a sodium cation is "+1")

Dipole: Debye

Polarizability (alpha): Cubic atomic units and cubic Angstroms. $1 \text{ au}^3 = (0.529)^3 \text{ Angstroms}^3$

First hyperpolarizability (beta): ESU or atomic units (atomic units)⁵ and (Angstroms)⁵. The description given by D. Kanis, M. Ratner and T. Marks, Chem. Rev. vol 94, pg 239 (1994), is as follows:

"The units for molecular hyperpolarizabilities can be somewhat complicated. Most contributions quote β in either atomic or esu units. Dimensional analysis of the quadratic hyperpolarizability indicates that it possesses the units of dipole moment³/energy², or in atomic units:

$$\beta = e^3 a_0^3 / E_h^2$$

where a_0 is the Bohr radius and E_h is the ionization energy of hydrogen (a Hartree). Substituting E_h (Hartree) $= e^2 / a_0$

$$\beta = e^3 a_0^3 / (e^2 / a_0)^2 = a_0^5 / e$$

in esu units; the second-order response is given as $\text{cm}^5 \text{statcoulomb}^{-1}$ which is identical to $\text{cm}^5 \text{esu}^{-1}$. Sometimes these units are simply referred to as esu. The conversion factor between atomic units and esu is as follows:

$$1 \text{ atomic unit} = (a_0^5 / e) (0.529 \times 10^{-8} \text{cm} / a_0)^5 (e / 4.803 \times 10^{-10} \text{esu}) \text{ or}$$

$$1 \text{ atomic unit} = 8.641 \times 10^{-33} \text{ cm}^5 \cdot \text{esu}^{-1}$$

in SI units, $1 \text{ atomic unit} = 3.206 \times 10^{-53} \text{ C}^3 \text{m}^3 \text{J}^{-2}$ and $1 = 2.693 \times 10^{20} \text{ cm}^5 \text{esu}^{-1}$."

(In earlier MOPACs, the quantity "4.803" was not used, therefore until December 2003 the value in $\text{cm}^5 \text{esu}^{-1}$ was too large by a factor of 4.803).

Second hyperpolarizability (gamma): Esu.

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 the [Table](#). See also [units](#).

Table: 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	$\gamma(-\omega; \omega, -\omega, \omega)$

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

IWFLB=3 optical rectification

E=(n₁, n₂, n₃, ...) 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.

COSMO (Conductor-like Screening Model)

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,

$$f(\epsilon) = \frac{(\epsilon - 1)}{(\epsilon + \frac{1}{2})},$$

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). 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 the Table 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.

Table: 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
N(Me)_4^+	AM1	157.1	101.1	56.0	59.9
N(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).

COSMO Keywords

EPS=*n.nn* Defines the dielectric constant of the solvent. This keyword triggers the whole COSMO.

NSPA=*nn* Controls the number of segments, default = 42.

DISEX=*n.nn* 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 energies*, (e.g. ions) larger values may be preferable. The calculation time may increase as $DISEX^2$ (until all interactions are calculated accurately).

RSOLV=*n.nn* Effective VDW radius of the solvent molecule. Default = 1.0Å.

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 the Table.

Table: Van der Waals radii (Angstroms) 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

- 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.
- 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 segmentatation.
- 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.

Footnotes

* : The 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.

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 Wang* 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.

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 `ESPLOT`. `ESPLOT` is very simple to use: the command is `esplot <filename>`, where `<filename>` is the name of the data-set. `ESPLOT` generates an on-line picture of the PMEP, and a HPGL file suitable for use in generating hard-copy. Because `ESPLOT` is so simple, it will not be discussed further. Instead, the rest of this discussion applies to the MOPAC calculation.

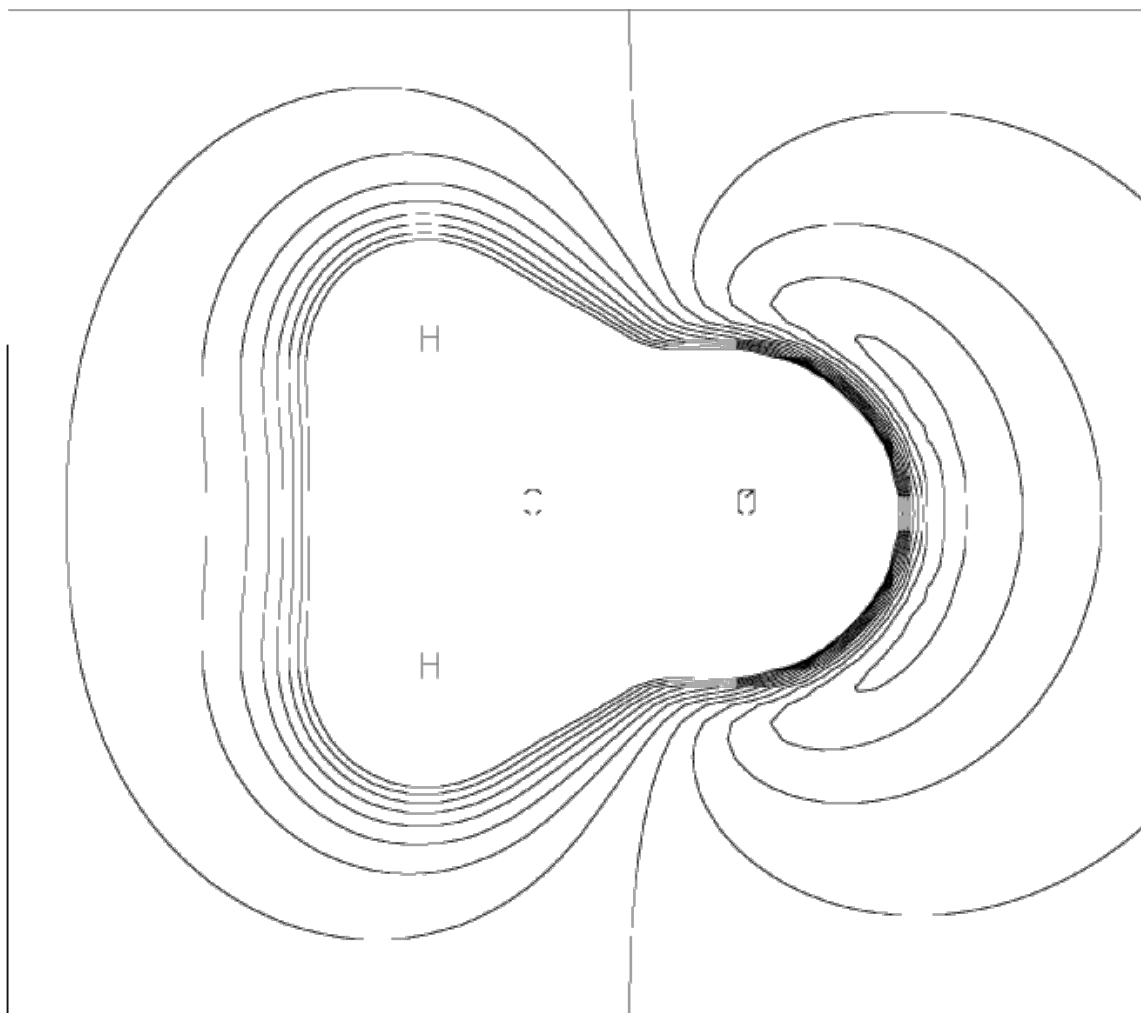
The grid generated by MOPAC consists of a 2-D array of points representing a cross-section through the system. The distance between points is a constant Ångstroms. The size represented by the grid is roughly 4 Ångstroms plus twice the size of the system. For example, N_2 has a N-N distance of Å, and 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 [PMEP](#) and `PRTMEP`. An example of a data-set for the PMEP procedure is shown in Figure 1. The PMEP plot for this data set is shown in Figure 2. This plot can be compared with the [MEP plot](#).

Figure 1: Data Set for PMEP Calculation of Formaldehyde

```
1scf AM1 PMEP MINMEP PRTMEP
Formaldehyde (Cross-section in plane
of molecule)
Generate a 2-D grid of PMEP 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 2: Parametric Molecular Electrostatic Potential around Formaldehyde
(contrast this with the [MEP plot](#))



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.

Atomic Charges

By use of [OPMEP](#), 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.

Footnotes

*: Current address: IRBM, Via Pontina Km.30.600, 00040 Pomezia (Roma), Italy

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 the Figure. Note that the keywords [STEP1=*n.nn*](#) and [STEP2=*m.mm*](#) are essential.

Figure: Example of a GRID Calculation

```
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
```

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.

Keywords used with the GRID option are [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. *n* x *m* should be less than 400000

A Note on Thermochemistry

Thermodynamic quantities, such as heat capacity, entropy, and internal energy, can be calculated using the vibrational frequencies (energies), moments of inertia of the molecule, its symmetry number, and a knowledge of the temperature. In this section, the relationships of these quantities is described.

Some derived quantities, which will be used in this section only are: Moment of inertia: I ($1 \text{ amu } \text{\AA}^2 = 1.660540 \times 10^{-40} \text{ g cm}^2$).

Rotational constants: A , B , and C (e.g.) $A = h/(8\pi^2 I)$

With I in $\text{amu } \text{\AA}^2$ then: $A \text{ (in MHz)} = 5.053791 \times 10^5 / I$

$A \text{ (in cm}^{-1}\text{)} = 5.053791 \times 10^5 / cI = 16.85763 / I$

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.

$$E_{eq} = E_{el} + E_{\text{nuclear-nuclear}}$$

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 $\text{cal.degree}^{-1}.\text{mol}^{-1}$) is:

Vibrational terms

The vibrational contribution to the internal energy arises from population of the vibrational energy levels. The vibrational partition coefficient, Q_{vib} , is given by:

$$Q_{vib} = \prod_i \frac{1}{[1 - \exp(-h\nu_i/kT)]}$$

E_{vib} , for a molecule at the temperature T as:

$$E_{vib} = \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\}$$

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:

$$E_{vib} = N_a \sum_i \left\{ \frac{h\nu_i}{2} + \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} \right\} \quad \text{Equation 1}$$

Note that the first term in the above equation is the zero-point vibration energy, E_{zpe} . Hence, the second term is the additional vibrational contribution due to the temperature increase from 0 K to T K. Namely,

$$E_{vib} = E_{zpe} + E_{vib}(T)$$
$$E_{zpe} = N_a \sum_i \frac{h\nu_i}{2}$$
$$E_{vib}(T) = N_a \sum_i \frac{h\nu_i \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]}$$

Equation 2

The value of E_{vib} from GAUSSIAN 82 and 86 includes E_{zpe} as defined by Equation 1 and Equation 2.

$$S_{vib} = R \sum_i \left\{ \frac{(h\nu_i/kT) \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]} - \ln[1 - \exp(-h\nu_i/kT)] \right\}$$
$$C_{vib} = R \sum_i \left\{ \frac{(h\nu_i/kT)^2 \exp(-h\nu_i/kT)}{[1 - \exp(-h\nu_i/kT)]^2} \right\}$$

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

s is the symmetry number (Examples of symmetry numbers are shown in the Table). I is moment of inertia. I_A , I_B , and I_C are moments of inertia about A, B, and C axes.

Table: 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 :	60
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

Linear molecule

Values for Q_{rot} , E_{rot} , and S_{rot} for a linear molecule are defined below:

$$Q_{ro} = \frac{8\pi^2 I k T}{\sigma h^2}$$
$$E_{ro} = 2/2)RT$$
$$S_{rot} =$$

$$\begin{aligned}
 & R \ln \left[\frac{8\pi^2 I k T}{\sigma h^2} \right] + R \\
 &= R \ln I + R \ln T - R \ln \sigma - 4.349203
 \end{aligned}$$

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

Values for Q_{rot} , E_{rot} , and S_{rot} for a non-linear molecule are defined below:

$$\begin{aligned}
 Q_{\text{rot}} &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\frac{8\pi^2 k T}{h^2} \right]^{3/2} \sqrt{I_A I_B I_C} \\
 &= \left(\frac{\sqrt{\pi}}{\sigma} \right) \left[\left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \right]^{1/2} \left(\frac{k T}{hc} \right)^{3/2} \\
 E_{\text{rot}} &= (3/2)RT \\
 S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \left(\frac{\pi}{\sqrt{\sigma}} \right) \left(\frac{8\pi^2 c I_A}{h} \right) \left(\frac{8\pi^2 c I_B}{h} \right) \left(\frac{8\pi^2 c I_C}{h} \right) \left(\frac{k T}{hc} \right)^3 \right\} + (3/2)R \\
 &= (R/2) \ln (I_A I_B I_C) + (3/2)R \ln T - R \ln \sigma - 5.3863921
 \end{aligned}$$

Here, -5.386 3921 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

Given that M is the molecular weight, then the values for Q_{tra} , E_{tra} , S_{tra} , and C_{tra} for a molecule are as defined below:

$$\begin{aligned}
 Q_{\text{tra}} &= \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 \\
 E_{\text{tra}} &= (3/2)RT \\
 S_{\text{tra}} &= R \left\{ \frac{5}{2} + \frac{3}{2} \ln \left(\frac{2\pi k}{h^2} \right) + \ln k + \frac{3}{2} \ln \left(\frac{M}{N_a} \right) + \frac{5}{2} \ln T - \ln p \right\} \\
 &= (5/2)R \ln T + (3/2)R \ln M - R \ln p - 2.31482
 \end{aligned}$$

$$C_{tra} = (5/2)R$$

or $H_{tra} = (5/2)RT$ due to the PV term (cf. $H = U + PV$). The internal energy U at T is:

$$U = E_{eq} + [E_{vib} + E_{rot} + E_{tra}]$$

or, in terms of the zero point energy, E_{zpe} , and real vibrations, $E_{vib}(T)$,

$$U = E_{eq} + [(E_{zpe} + E_{vib}(T)) + E_{rot} + E_{tra}].$$

Enthalpy H for one mole of gas is defined as

$$H = U + PV$$

Assumption of an ideal gas (i.e., $PV = RT$) leads to

$$H = U + PV = U + RT$$

Thus, Gibbs free energy G can be calculated as:

$$G = H - TS$$

Thermochemistry in MOPAC

It should be noted that M.O. parameters for MNDO, AM1, etc., 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, DH_f), force constants, normal vibration frequencies, etc. are all related to the values at 25°C, not to 0 K. Therefore, the E_0 calculated in FORCE is not the true E_0 . Its use as E_0 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):

$$E_{SCF} = E_{eq} + E_{zpe} + E_{vib}(298.15) + E_{rot} + E_{tra} + PV + S[-E_{elec}(\text{atom}) + DH_f(\text{atom})].$$

To avoid the complication arising from the definition of E_{SCF} , within the thermodynamics calculation the Standard Enthalpy of Formation, DH , is calculated by

$$DH = E_{SCF} + (H_T - H_{298}).$$

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 DH for the elements in their standard state at T , and use these DH 's to get the DH for the compound you're working with (or use tables from the literature).

This problem is, however, not normally important, because the most common use of H_T is for calculating the thermodynamics of reactions at temperatures other than 298K. For all reactions, the types and number of atoms

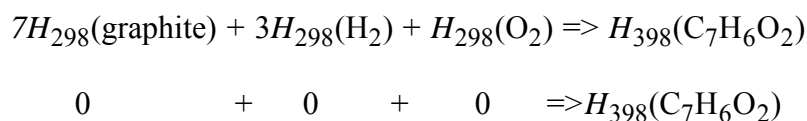
must be the same in reactants and products, therefore the fact that the H_T are relative to the elements in their standard state at 298K is irrelevant. Consider the simple Diels-Alder reaction:



The heat of this reaction at 298K is $H_{298}(\text{C}_6\text{H}_{10}) - H_{298}(\text{C}_2\text{H}_4) - H_{298}(\text{C}_4\text{H}_6)$. At any other temperature, the heat of reaction would be:

$$H_R = H_T(\text{C}_6\text{H}_{10}) - H_T(\text{C}_2\text{H}_4) - H_T(\text{C}_4\text{H}_6).$$

Care must, however, be taken to account for changes in volume - if any of the reactants or products are gaseous, then appropriate corrections must be made to H_R . Complications arise only if absolute heats of formation are needed. Thus, if the heat of formation of benzoic acid ($\text{C}_7\text{H}_6\text{O}_2$) at 398K (100C) is needed, the $H_{398}(\text{C}_7\text{H}_6\text{O}_2)$ generated by MOPAC would be for the reaction:



Note that on the left side, the temperatures are 298K. For H_2 and O_2 , the heats of formation at 398K can readily be calculated, but for graphite the calculation is more complicated. The easiest way to generate a balanced equation would be to use tables of heats of formation of the elements at non-standard temperatures.

Finally, as mentioned above, changes in volume must also be taken into account: if the reaction volume changes, then $R\text{DN}(T-298)$ must be added or subtracted, where R is the gas constant ($\sim 2\text{cals/degree/mol}$), and DN is the change in volume. Thus for the formation of methane from graphite and hydrogen, 2 volumes of reactant (H_2 + graphite) yield 1 volume of methane, therefore $\text{DN} = 1$.

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} :

$$v_i = \omega_i c$$

$$E_{\omega_i} = \exp(-hv_i/kT) = \exp(-\omega_i hc/kT) = \exp(-\omega_i C_1)$$

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, the earlier Equations can be written as follows:

Vibration

$$Q_{\text{vib}} = \prod_i \frac{1}{(1 - E_{\omega_i})}$$

$$\begin{aligned}
 E_{\text{zpe}} &= \frac{0.5N_a h c}{4.184 \times 10^7} \sum_i \omega_i \\
 &= 1.429572 \sum_i \omega_i \\
 E_{\text{vib}}(T) &= N_a h c \sum_i \frac{\omega_i E_{\omega_i}}{1 - E_{\omega_i}} = (R/k) h c \sum_i \frac{\omega_i E_{\omega_i}}{1 - E_{\omega_i}} \\
 S_{\text{vib}} &= R(hc/kT) \sum_i \left\{ \frac{\omega_i E_{\omega_i}}{(1 - E_{\omega_i})} \right\} - R \sum_i \ln(1 - E_{\omega_i}) \\
 &= RC_1 \sum_i \left\{ \frac{\omega_i E_{\omega_i}}{(1 - E_{\omega_i})} \right\} - R \sum_i \ln(1 - E_{\omega_i}) \\
 C_{\text{vib}} &= R(hc/kT)^2 \sum_i \left\{ \frac{\omega_i^2 E_{\omega_i}}{(1 - E_{\omega_i})^2} \right\} \\
 &= RC_1^2 \sum_i \left\{ \frac{\omega_i^2 E_{\omega_i}}{(1 - E_{\omega_i})^2} \right\}
 \end{aligned}$$

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

$$\begin{aligned}
 Q_{\text{rot}} &= \frac{1}{\sigma} \left[\frac{\pi}{(ABCC_1^3)} \right]^{1/2} \\
 E_{\text{rot}} &= (3/2)RT \\
 S_{\text{rot}} &= \frac{R}{2} \ln \left\{ \frac{\pi}{\sigma^2 ABC} \left(\frac{kT}{hc} \right)^3 \right\} + (3/2)R \\
 &= 0.5R3 \ln(kT/hc) - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3 \\
 &= 0.5R - 3 \ln C_1 - 2 \ln \sigma + \ln \left(\frac{\pi}{ABC} \right) + 3 \\
 C_{\text{rot}} &= (3/2)R
 \end{aligned}$$

Translation

$$Q_{\text{tra}} = \left(\frac{\sqrt{2\pi M k T / N_a}}{h} \right)^3 = \left(\frac{\sqrt{1.660540 \times 10^{-24} \times 2\pi M k T}}{h} \right)^3$$

$$E_{\text{tra}} = (3/2)RT$$

$$H_{\text{tra}} = (3/2)RT + pV = (5/2)RT \text{ cf. } pV = RT$$

$$\begin{aligned} S_{\text{tra}} &= (R/2)[5 \ln T + 3 \ln M] - 2.31482 \text{ cf. } p = 1 \text{ atm} \\ &= 0.993608[5 \ln T + 3 \ln M] - 2.31482 \end{aligned}$$

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:

$$(E_{\text{vib}} - E_{\text{zpe}}) + E_{\text{rot}} + (E_{\text{tra}} + pV)$$

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 using U and H and Escf, thus:

$$\Delta H = E_{\text{SCF}} + (H_T - H_{298})$$

Note that E_{zero} is already counted in E_{scf} .

By using

$$H = U + pV = U + RT$$

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.

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 the [Figure](#).

Figure: Example of data for SADDLE calculation

```

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:

```

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:

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.

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 $\text{DRC}+\text{TS}$ sequence do the $\text{DRC}+\text{TS}$ 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.

Gibbs Free Energy

Definitions

The following definitions are useful in considering free energy relationships:

- The heat of formation, DHf, of an element in its standard state is zero.
- The free energy of formation, DFf, of an element in its standard state is zero.
- $F = H - TS$.
- $DFf = DHf - TDS$.

Application to some Simple Reactions

Consider the reaction: $H_2 + Cl_2 \rightarrow 2HCl$

From Table 1, the DHf change can be calculated using:

$$\begin{array}{rclcl} HCl & - & (1/2)H_2 & - & (1/2)Cl_2 \\ -22.06 & - & 0.00 & - & 0.00 \end{array}$$

to be -22.06 kcal/mol.

Table 1: Thermodynamic Properties of some Compounds

Compound	Formula	DHf (kcal/mol)	DFf (kcal/mol)	S (cal/mol/K)
Hydrogen Chloride (g)	HCl	-22.06	-22.77	44.62
Water (g)	H2O	-57.80	-54.64	45.11
Water (l)	H2O	-68.32	-56.69	16.72
Methane (g)	CH4	-17.89	-12.14	44.50
Oxalic acid (s)	C2H2O4	-197.6	-166.8	28.70

The corresponding entropy change, DS, requires using the entropies for the elements in their standard state, Table 2, and can be calculated using:

$$\begin{array}{rclcl} HCl & - & (1/2)H_2 & - & (1/2)Cl_2 \\ 44.62 & - & (1/2)31.21 & - & (1/2)53.29 \end{array}$$

to be +2.37 cal/mol/K.

Using these results, the free energy change, DFf, can be calculated:

$$\begin{array}{rclcl} DFf & = & DHf & - & TDS \\ -22.77 & = & -22.06 & - & 298 \times 2.37 / 1000 \end{array}$$

which agrees with the value given in the CRC handbook (Table 1, here)

In like manner, the free energy of water, as the vapor, can be calculated:

$$DFf(H_2O) = DHf(H_2O) - TDS(H_2O - H_2 - (1/2)O_2)$$

as

-54.64 = -57.80 - 298x(45.11 - 31.21 - 0.5x49.00)/1000

Similarly, the free energy of water, as the liquid, can be calculated:

DFf(H₂O) = DHf(H₂O) - TDS(H₂O-H₂-(1/2)O₂)

as

-56.70 = -68.32 - 298x(16.72 - 31.21 - 0.5x49.00)/1000

which is close to the reported DFf of -56.69 kcal/mol.

For oxalic acid, the equivalent calculation is

DFf(C₂H₂O₄) = DHf(C₂H₂O₄) - TDS(C₂H₂O₄-H₂-2O₂-2C)

or

-166.8 = -197.6 - 298x(28.70 - 31.21 - 2x49.00 - 2x1.36)/1000

Table 2: Entropies for the Elements at 298K (cal/K/mol)

I	II	Transition Metals									III	IV	V	VI	VII
H 31.21															
Li 6.7	Be 32.53										B 1.56	C 1.36	N 45.77	O 49.00	F 48.6
Na 12.2	Mg 7.77										Al 39.30	Si 4.47	P 10.6	S 7.62	Cl 53.29
K 15.2	Ca 9.95	Ti 7.24	V 7.05	Cr 5.68	Mn 7.59	Fe 6.49	Co 6.8	Ni 7.20	Cu 7.96	Zn 9.95	Ga	Ge 10.14	As 8.4	Se	Br 36.4
Rb 16.6	Sr 13.0	Zr 9.18		Mo 6.83				Pd 8.9	Ag 10.21	Cd 12.3	In	Sn 12.3	Sb 43.06	Te 11.88	I 27.9
Cs 19.8	Ba 40.70							Pt 10.0		Hg 18.5	Tl 15.4	Pb 15.51	Bi 13.6		

("CRC Handbook of Chemistry and Physics," 60th Edition, R. C. Weast, (Ed.), CRC Press, Boca Raton, FL, 1980.)

In MOPAC, the quantities DHf and S are calculable. DHf is the calculated heat of formation coming out of the SCF calculations, S, the entropy, is printed near the end of a THERMO calculation. In the following output from a methane calculation, the entropy terms is in **bold** and *italicized*:

CALCULATED THERMODYNAMIC PROPERTIES						
*						
TEMP. (K)	PARTITION FUNCTION	H.O.F. KCAL/MOL	ENTHALPY CAL/MOLE	HEAT CAPACITY CAL/K/MOL	ENTROPY CAL/K/MOL	
298	VIB.	0.1006D+01	24.0439	0.5429	0.0927	
	ROT.	0.3850D+02	888.2854	2.9808	10.2356	
	INT.	0.3874D+02	912.3293	3.5238	10.3283	
	TRA.	0.6212D+26	1480.4756	4.9680	34.2611	
	TOT.	-8.790	2392.8049	8.4918	44.5894	

The entropies of the pure elements should be taken from Table 2. A comparison of the three gaseous systems in Table 1 is presented in Table 3:

Table 3: Comparison of Observed and Calculated Thermodynamic Quantities

Compound	Formula	Observed (Table 1)			Calculated (AM1)		
		DHf (kcal/mol)	DFf (kcal/mol)	S (cal/mol/K)	DHf (kcal/mol)	DFf (kcal/mol)	S (cal/mol/K)
Hydrogen Chloride (g)	HCl	-22.06	-22.77	44.62	-24.61	-25.32	44.63
Water (g)	H2O	-57.80	-54.64	45.11	-59.25	-56.09	45.09
Methane (g)	CH ₄	-17.89	-12.14	44.50	-8.79	-3.04	44.59

General

For MOPAC2007 [[147](#)]

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 [[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.

General Description of 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](#)], AM1 [[3](#)], PM3 [[4](#)], PM6, [RM1](#), 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 unique, 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, MOPAC was an obvious contender.

DIPOLE

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.

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.

- [Eigenfollowing](#) (EF)
- [The BFGS function optimizer](#)
- [Optimization of one unknown](#)
- [Over-riding the default options](#)
- [Locating transition states](#)

PM3

The PM3 method [4] is to be used.

The PM3(tm) method is not in MOPAC, but it is so similar to PM3 that PM3(tm) calculations can be run using the EXTERNAL option. To run PM3(tm) calculations, use keywords PM3 and EXTERNAL=<filename>. The contents of the file <filename> would be the normal PM3(tm) parameters. An example of such a parameter set is as follows:

Table: Example of PM3(tm) parameter set (for Ti)

atnum	22
n	4. 0
uss	- 26. 45829779
upp	- 21. 17197024
udd	- 36. 34653108
betas	- 23. 30450933
betap	- 1. 07045796
betad	- 2. 69874461
zs	1. 04638304
zp	1. 25955299
zd	1. 31534641
alpha	1. 57689752
gss	12. 64676054
gpp	6. 92904760
gdd	13. 01302063
gsd	4. 28866787
hsd	1. 35838275
a1	- 0. 09046372
b1	4. 25904322
c1	1. 25471803
a2	- 0. 01221641
b2	4. 07610129
c2	2. 79140274
hform	112. 30000000
zval	4. 00000000

SAFE

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`.

UNSAFE

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.