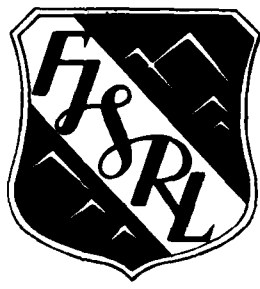


MICROCOPY RESOLUTION TEST CHART  
NATIONAL BUREAU OF STANDARDS-1963-A

1

DTIC FILE COPY S DTIC ELECTE D  
MAR 14 1988  
AD



FRANK J. SEILER RESEARCH LABORATORY  
FJSRL-TR-88-0001  
FEBRUARY 1988

AD-A191 797

# MOPAC WORKSHOP MANUAL

C. J. DYMEK, JR.  
J. J. P. STEWART  
D. M. STORCH



APPROVED FOR PUBLIC RELEASE;  
DISTRIBUTION UNLIMITED.

88 3 11 011

PROJECT 2303

AIR FORCE SYSTEMS COMMAND  
UNITED STATES AIR FORCE

FJSRL-TR-88-0001


This document was prepared by the Electrochemistry Division, Directorate of Chemical Sciences, Frank J. Seiler Research Laboratory, United States Air Force Academy, CO. The research was conducted under Project Work Unit number 2303-F3-07. Lt Col Chester J. Dymek, Jr. was the project scientist.

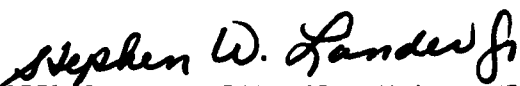
When U.S. Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the government may have formulated, furnished or in any way supplied the said drawings, specifications or other data is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

Inquiries concerning the technical content of this document should be addressed to the Frank J. Seiler Research Laboratory (AFSC), FJSRL/NC, USAF Academy, Colorado Springs, CO 80840. Phone AC 303-472-2655.

This report has been reviewed by the Commander and is releasable to the National Technical Information Service (NTIS). At NTIS it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

  
CHESTER J. DYMEK, JR., Lt Col, USAF  
Project Scientist

  
STEPHEN W. LANDER, JR., Major, USAF  
Director, Chemical Sciences

KENNETH E. SIEGENTHALER, Lt Col, USAF  
Chief Scientist

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

Printed in the United States of America. Qualified requestors may obtain additional copies from the Defense Documentation Center. All others should apply to:

National Technical Information Service  
6285 Port Royal Road  
Springfield VA 22161

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			4. PERFORMING ORGANIZATION REPORT NUMBER(S) FJSRL-TR-80-0001		
6a. NAME OF PERFORMING ORGANIZATION Frank J. Seiler Research Lab			5b. OFFICE SYMBOL (If applicable) FJSRL/NC	5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6c. ADDRESS (City, State and ZIP Code) USAF Academy Colorado Springs CO 80840-6528			7b. ADDRESS (City, State and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AF Office of Scientific Research		8b. OFFICE SYMBOL (If applicable) AFOSR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER		
8c. ADDRESS (City, State and ZIP Code) Bldg. 410 Bolling AFB DC 20332			10. SOURCE OF FUNDING NOS.		
11. TITLE (Include Security Classification) MOPAC Workshop Manual			PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2303	TASK NO. F3
12. PERSONAL AUTHOR(S) Chester J. Dymek, Jr., James J. P. Stewart and Donn M. Storch			WORK UNIT NO. 07		
13a. TYPE OF REPORT Technical Report		13b. TIME COVERED FROM Aug 87 TO Jan 88	14. DATE OF REPORT (Yr., Mo., Day) 88 Feb 01		15. PAGE COUNT 50
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD 0704	GROUP 1901	SUB. GR. 2102	MOPAC, MNDO, AM1, Semi-Empirical Calculations, Workshop		
19. ABSTRACT (Continue on reverse if necessary and identify by block number) MOPAC is a general-purpose, semi-empirical molecular orbital program for the study of chemical reactions involving molecules, ions, and linear polymers. It implements the semi-empirical Hamiltonians MNDO, AM1, and MINDO/3 and combines the calculations of vibrational spectra, thermodynamic quantities, isotopic substitution effects, and force constants in a fully integrated program. This manuscript is designed to be used by participants in a 2-day workshop on the MOPAC computational chemistry program. It is aimed at beginning users, assuming some fundamental knowledge of quantum mechanics, but not requiring familiarity with the higher mathematics involved in molecular orbital calculations.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT <input type="checkbox"/> DTIC USERS <input type="checkbox"/>			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL Chester J. Dymek, Jr.			22b. TELEPHONE NUMBER (Include Area Code) (303) 472-2655	22c. OFFICE SYMBOL FJSRL/NC	

## TABLE OF CONTENTS

Foreword . . . . .	ii
List of Figures . . . . .	iii
1. Overview of MOPAC Method . . . . .	1
2. Description of theoretical/computational process . . . . .	5
3. How to use MOPAC . . . . .	11
a. Creating the data file, CH3OH.DAT . . . . .	11
b. Reading the output file, CH3OH.OUT . . . . .	16
4. How to use DRAW . . . . .	19
a. Drawing the initial geometry . . . . .	20
b. Drawing the optimized geometry . . . . .	21
5. Examples of applications . . . . .	23
a. Naphthalene and the keyword, SYMMETRY . . . . .	23
b. Vibrational modes and IR spectrum of Al <sub>2</sub> Cl <sub>7</sub> <sup>-</sup> . . . . .	26
c. Reaction path calculation . . . . .	32
6. Other considerations . . . . .	38
a. Restart files . . . . .	38
b. Dummy atoms . . . . .	39
c. Sparkles . . . . .	39
d. Condensed phase systems . . . . .	40
e. A final word . . . . .	41
Notes . . . . .	42



Accession For	<input checked="" type="checkbox"/> DTIC <input type="checkbox"/> NTIS <input type="checkbox"/> CRA&I <input type="checkbox"/> TAB <input type="checkbox"/> Unannounced <input type="checkbox"/> Justification	By	Distribution/	Availability Codes	Avail and/or Special
					A-1

## FOREWORD

This manual is designed to be used by participants in a 2-day workshop on the MOPAC computational chemistry program. MOPAC implements the semi-empirical molecular orbital methods pioneered by Professor Michael J. S. Dewar, University of Texas at Austin. This manual is aimed at beginning users, assuming some fundamental knowledge of quantum mechanics, but not requiring familiarity with the higher mathematics involved in molecular orbital calculations. The workshop addresses one objective in a joint U.S. Air Force Academy Department of Chemistry -- Frank J. Seiler Research Laboratory project to establish a center-of-excellence in computational chemistry at the Air Force Academy. That objective is to promote use of MOPAC in DOD and associated research groups while demonstrating its applicability to real problems of interest to the Air Force. The project is funded by the Air Force Office of Scientific Research.

## LIST OF FIGURES

	Page
Fig. 1. MOPAC flow diagram . . . . .	6
Fig. 2. CH <sub>3</sub> OH geometry . . . . .	14
Fig. 3. Naphthalene geometry definition with symmetry . . . . .	24
Fig. 4. Al <sub>2</sub> Cl <sub>7</sub> <sup>-</sup> vibrations . . . . .	30
Fig. 5. Al <sub>2</sub> Cl <sub>7</sub> <sup>-</sup> calculated vs experimental IR spectra . . . . .	31
Fig. 6. Br <sup>-</sup> + CH <sub>3</sub> Cl reaction profiles . . . . .	35
Fig. 7. Br <sup>-</sup> + CH <sub>3</sub> Cl transition state . . . . .	37



## MOPAC WORKSHOP

### Outline of workshop:

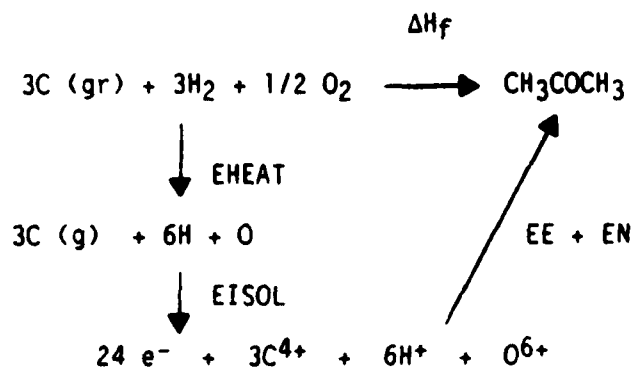
1. Overview of the MOPAC method.
2. Description of theoretical/computational process.
3. How to use MOPAC.
4. How to use DRAW.
5. Examples of applications.
6. Other considerations.

1. Overview of the MOPAC method.

MOPAC is a semi-empirical molecular orbital program for calculating various properties of molecular and ionic systems. The geometries, charge distributions, dipoles, the make-up of the molecular orbitals, and other similar information on systems of interest can be calculated. Thermodynamic information can be obtained from the calculated heats of formations ( $\Delta H_f$ ) of

systems. Kinetic information can be obtained from calculations which map the change in  $\Delta H_f$  with chosen reaction coordinates and identify the transition states for reactions. The infrared spectra of molecules can be calculated for comparison with experimental spectra. Other more specialized capabilities, such as the calculation of polymer properties, are also available. The MOPAC Manual is the document which shows how the full range of MOPAC's capabilities can be used. This workshop is intended to get you started as a MOPAC user and to give you a sense of its powers and limitations.

As the term, "semi-empirical," implies, a MOPAC calculation is based partly on quantum mechanical theory and partly on experimental data. Theoretical considerations led to the form of the Hamiltonian energy operator used in the secular equations. The secular equations are solved using matrix algebra to obtain the system eigenfunctions and the corresponding eigenvalues. These eigenvalues are then used to calculate the system's electronic energy (EE). EE is combined with the calculated nuclear repulsion (EN), the energies of the isolated atoms (EISOL), and the heats of atomization of the atoms in the system (EHEAT) to give the heat of formation of the system. The diagram below illustrates this process for the example of acetone.



There are three different Hamiltonians which were developed by the research group of Michael J. S. Dewar at the University of Texas at Austin and are available in the MOPAC program: MNDO, AM1, and MINDO/3.

Experimental data enter into the method because the secular equations set up from these Hamiltonians contain functions (describing the behavior of each atom in the system) which include adjustable parameters. These parameters are determined not from first principles (i.e., purely theoretical considerations), but rather by a fitting process called "parameterization." Parameterization uses experimental data from a set of molecules to determine the values of these parameters. There are seven parameters for each element in the MNDO method. The AM1 method has the same seven and a few more per element while the MINDO/3 method has seven per pair of atoms. In the fourth edition of MOPAC, the atoms which have been parameterized are: H, Li, B, C, N, O, F, Al, Si, P, S, Cl, Ge, Br, Sn, I, Hg, Pb. (Currently only MNDO is parameterized for all these atoms.)

MOPAC can be run on normal Control Data Corporation (CDC), Data General, Gould, and Digital Equipment Corporation (VAX) computers, as well as on the CDC 205 and Cray-XMP "supercomputers." Versions are also available for several microcomputers, including IBM PC-AT and XT and Zenith.

The process in which MOPAC is a valuable tool begins with a clear statement of the chemical problem being dealt with: Is a particular molecule stable? What is the most stable geometry of a particular ion? What is the electronic structure of a molecule? What is the enthalpy of reaction for a

particular reaction? How large is the energy of activation for the reaction when it proceeds in a certain way? What are the frequencies of the normal modes of a particular molecule? Which compound requires the least amount of activation to react with a specific substrate? And so on.

Chemists answer these kinds of questions every day without MOPAC. They get their answers from thermodynamic tables, the literature, or experiments they perform themselves. MOPAC is useful when these sources cannot easily or reliably provide the desired information. But its greatest value emerges when it is used to predict, interpret, and confirm experimental results.

Once the system and the information desired about it are determined, the MOPAC user constructs a data file which models the system and describes the desired information in a form which the MOPAC program can use to carry out the needed calculations. The essential parts of the data file are the KEYWORDS, which tell the program how the calculation is to be carried out and what output is desired, and the GEOMETRY DEFINITION, which gives the coordinates of the atoms in the system. The user then enters a simple command to start the calculation and the MOPAC program takes over. It performs all the requested calculations and creates a number of files containing the results. Most of the desired information will be in an output file. In the third section of this manual, we will show how to create a data file (labeled FILENAME.DAT) and how to interpret an output file (labeled FILENAME.OUT). But first we will give in the following section a brief outline of what goes on when, as we said above, "the MOPAC program takes over."

## 2. Description of theoretical/computational process.

First MOPAC checks the KEYWORDS for compatibility. For example, if both AM1 and MINDO-3 were entered, the calculation would stop since only one Hamiltonian can be used. It also checks that the geometry definition is not fatally flawed, e.g., two atoms positioned at the same point in space. If the data file passes this inspection, the subroutine MOLDAT is called.

MOLDAT retrieves (from a set of reference data) the parameters and other atomic data for the atoms comprising the system. It sets up the many arrays (or multi-dimensional matrices) to be used in the calculations based on the total number of electrons in the system and other factors, such as whether a UHF (unrestricted Hartree-Fock) or an RHF (restricted Hartree-Fock, this being the default selection) calculation has been requested.

MOPAC then proceeds to one of the seven subroutines shown in the flow diagram in Fig 1 depending on which KEYWORD has been selected. We will follow what happens when none of them have been selected so that the default subroutine, FLEPO, is called. (This is a geometry optimization procedure still named after two of its creators, Fletcher and Powell, although with the fourth edition, another procedure, the BFGS method, is used). The GEOMETRY package (of which FLEPO is a part) communicates with the .ELECTRONICS package through subroutine, COMPG. These two packages comprise about 90% of the program. Following the basic workings of FLEPO requires some understanding of the terms listed below. In this list and in the following discussion, a singly underlined symbol indicates a vector and a doubly underlined symbol indicates a matrix.

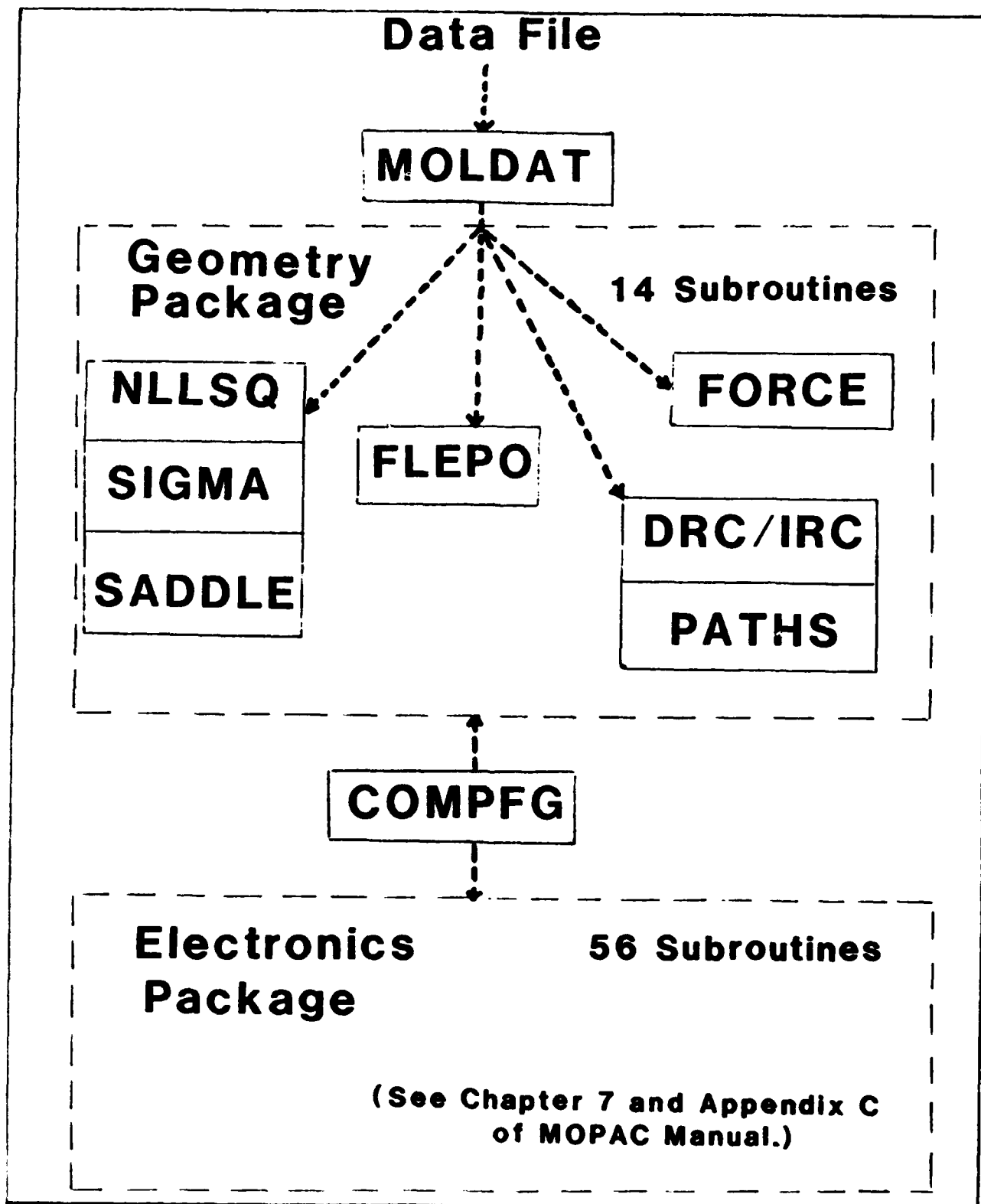


Fig. 1. MOPAC flow diagram.

$x_i^0$  - the  $i$ th coordinate of the  $3N-6$  coordinates (where  $N$  is the number of atoms in the system) initially entered in the geometry definition of the data file.

$x_i^k$  - the  $i$ th coordinate after completion of the  $k$ th cycle of the calculation. What a cycle consists of will become clear as we proceed.

$\underline{x}^k$  - the vector in  $(3N-6)$ -space with the internal coordinates as components  $(x_1^k, x_2^k, x_3^k, \dots, x_{(3N-6)}^k)$

$g_i^k$  - the derivative of  $\Delta H_f$  with respect to  $x_i$  in the  $k$ th cycle of the calculation.

$\underline{g}^k$  - the vector with the  $g_i^k$  as components, i.e.,  $(g_1^k, g_2^k, g_3^k, \dots, g_{(3N-6)}^k)$ .

$\underline{p}^k$  - the change in  $\underline{x}$  from one cycle to the next, i.e.,  $\underline{p}^k = \underline{x}^{k+1} - \underline{x}^k$ .

$\underline{y}^k$  - the change in  $\underline{g}$  from one cycle to the next, i.e.,  $\underline{y}^k = \underline{g}^{k+1} - \underline{g}^k$ .

$\underline{G}^k$  - the Hessian matrix, this is the matrix of the second derivatives of  $\Delta H_f$  with respect to the coordinates,  $d^2\Delta H_f/dx_i dx_j$ , on the  $k$ th cycle. Thus it is a square matrix with  $3N-6$  rows.

$\underline{H}^k$  - the inverse Hessian, i.e., the inverse of  $\underline{G}^k$ .

$\underline{d}^k$  - the "search direction vector" used to calculate  $\underline{p}^k$ . It is given by  $\underline{d}^k = \underline{H}^k \underline{g}^k$ .

$\alpha$  - a scaling factor for  $\underline{d}^k$  which accounts for the fact that the approach to minimum energy on the 3N-6 dimensional surface will not be along a pure parabolic surface.

With these definitions in hand, let us follow what FLEPO does. Using the coordinates,  $x_i^0$ , which you entered in the data file, it performs a self-consistent field molecular orbital (SCF MO) calculation in the ELECTRONICS package using the MNDO, AM1, or MINDO/3 methods. We will not go into the details of how these calculations are done since they are described in their essential form in any introductory quantum mechanics text.\* The result is the initial  $\Delta H_f^0$  for the system. FLEPO also calculates the values for  $g_i^0$  in the GEOMETRY package. The coordinates,  $x_i^1$ , to be used in the next cycle are then calculated from

$$x_i^1 = x_i^0 - 0.01 * \text{SIGN}(g_i^0),$$

where  $\text{SIGN}(g_i^0) = 1$  if  $g_i^0 > 0$  and  $= -1$  otherwise. The second term in this equation corresponds to the  $i$ th component of  $\underline{p}^k$ , which, in subsequent calculations, will be calculated differently, as we shall see. With the new values,  $x_i^1$ , the next cycle, which also consists of only a single SCF calculation (subsequent cycles typically consist of four SCF's), is completed and the results,  $\Delta H_f^1$  and  $\underline{g}^1$ , are obtained. The vectors,  $(\underline{y}^0 = \underline{g}^1 - \underline{g}^0)$  and

---

\*also see "MOPAC: A Semi-Empirical Molecular Orbital Program," James J. P. Stewart, Handbook of Computer Aided Modeling and Design, ACS Monograph Series, 1988.



( $\underline{p}^0 = \underline{x}^1 - \underline{x}^0$ ), are then used to calculate the diagonal elements,  $G_{ij} = y_i^0/p_j^0$ , of the Hessian matrix,  $\underline{G}^k$ , so that  $\underline{H}^k$  and  $\underline{d}^k$  can be obtained (for  $k=1$ ) as defined above. FLEPO next calls the subroutine, LINMIN, to calculate a value for  $\alpha$  using a parabolic fit. With  $\alpha$  and  $\underline{d}^k$  in hand, LINMIN calculates  $\underline{p}^k$  from

$$\underline{p}^k = \alpha \underline{d}^k .$$

Now FLEPO can calculate  $\underline{x}$  for the next cycle ( $k+1=2$ ) using

$$\underline{x}^{k+1} = \underline{x}^k + \underline{p}^k .$$

The next cycle results in new values for  $\Delta H_f$ ,  $\underline{g}$ ,  $\underline{G}$ ,  $\underline{d}$ , and  $\alpha$ , as well as the coordinates for the following cycle using the same equations. The equation used to calculate the inverse of the Hessian in a following cycle is given below:

$$\underline{H}^{k+1} = \underline{H}^k - (\underline{H}^k \underline{y}^k \underline{p}^{kt} + \underline{p}^k \underline{y}^{kt} \underline{H}^k) / S + Q (\underline{p}^k \underline{p}^{kt}) / S ,$$

$$\text{where } Q = 1 + \underline{y}^{kt} \underline{H}^k \underline{y}^k / \underline{p}^{kt} \underline{y}^k$$

$$\text{and } S = \underline{p}^{kt} \underline{y}^k .$$

(Note: the superscript, t, indicates the transpose of the vector)

What must be kept in mind at this point is that the inverse Hessian and all the other quantities described above are calculated to obtain progressively better coordinates,  $\underline{x}_j$ , for the system. Coordinates get

"better" when the heat of formation calculated from them is lower than that obtained in the previous cycle.

After each cycle FLEPO calculates an indicator of how close the calculated value of  $\Delta H_f$  is to a minimum. This indicator, GNORM, is given by

$$\text{GNORM} = \sqrt{\sum g_i^2} = |g|$$

which has the units kcal/mol/Å and can be considered a measure of how sensitive  $\Delta H_f$  is to small changes in the system geometry.

FLEPO continues to run through these cycles until two tests are passed:

(1) the  $|g_i|$  values (i.e., the individual components of GNORM) must be below a certain limit which is a calculated function of the system, and

(2) any one of the following:

(a) "test-on-x-satisfied", i.e., the magnitude of the vector,  $\underline{d}^k$ , is very small (less than 0.0001 Å),

(b) Herbert's test, i.e., the projected decrease in  $\Delta H_f$  is less than 0.001 kcal/mol,

(c) "gradient satisfied", i.e., GNORM is less than KEYWORD-specified value (default is 1 kcal/mol/Å).

(d) " $\Delta H_f$  test satisfied", i.e., the change in  $\Delta H_f$  from cycle to cycle is less than 0.002 kcal/mol.

While the process we have described is controlled by only one of the subroutines shown in the flow diagram, it is a very important process which occurs in most of the MOPAC calculations which are done. Brief descriptions of what is done when KEYWORDS select the other subroutines are given in the MOPAC Manual. The description above was given not only to satisfy the curiosity of users who prefer to have some understanding of what happens when "the MOPAC program takes over," but also to introduce some terms which must be understood to properly interpret MOPAC results.

### 3. How to use MOPAC.

We'll start by calculating the heat of formation of a simple, well known molecule, methanol (so a good filename would be CH3OH). We need a computer on which MOPAC has been loaded and access to a directory in which we can create a data file and initiate the MOPAC calculation. We then enter a command to edit the data file, e.g., EDIT/EDT CH3OH.DAT for a VAX or vi ch3oh.dat for a UNIX operating system. Since that file presumably does not exist, we get a blank screen and begin creating!

#### a. Creating the data file, CH3OH.DAT.

The first line will be a series of KEYWORDS, which tell the program which properties we want calculated for the system we will enter, how much CPU time it will be allowed, the precision we want, what outputs we want, etc. The

full list of KEYWORDS and their effects on the calculation is given in Chapter 2 of the Manual. In every calculation a  $\Delta H_f$ , geometry, and charge distribution for the system are output whether you want them or not. So for our simple calculation, this first line will look like this:

```
T=3600 AM1 PRECISE
```

T=3600 gives the computer 3600 seconds of CPU time to complete the calculation, AM1 tells the program to use the AM1 method rather than the MNDO method, which is the default choice, and PRECISE puts into effect criteria which will give us a more precise result than the default criteria, but at the expense of more computer time.

The second and third lines are comments which we may use to describe for ourselves what the calculation is, for example:

```
AM1 CALCULATION OF HEAT OF FORMATION OF METHANOL WITH PRECISE
```

The first of these three lines must always be used for the purpose described above, and the second and third lines must be left blank if no comments are desired.

Now we get to the starting geometry definition. The geometry is defined in internal coordinates with a connectivity list, which is the list of numbers in three columns under the heading "Connectivity" in the table below. How it works should become clear as we discuss this table. The numbers in the

columns, "Distance," "Angle," and "Dihedral" are the 3N-6 coordinates,  $x_i^\circ$ , which were discussed in Section 2. The same geometry can be defined in many ways, and what is shown below is just one example:

Atom	Distance	Angle	Dihedral	Connectivity		
I	I-A	I-A-B	I-A-B-C	A	B	C
1   C						
2   O	1.43	1	0	0	0	1
3   H	0.96	1	108.0	1	0	0
4   H	1.10	1	109.0	1	180.0	1
5   H	1.10	1	109.0	1	-60.0	1
6   H	1.10	1	109.0	1	60.0	1
	.....BLANK LINE.....					

The system described above is shown in Fig 2. The drawing was created from the CH3OH.DAT file using the DRAW program, which will be discussed later. The geometry definition can be understood by following the definition for one atom, say atom number 6, a hydrogen designated H(6).

Distance I-A: H(6) is 1.10 Å from C(1) since in the connectivity list, A is defined to be atom number 1. This is the distance 6-1.

Angle I-A-B: H(6) forms an angle of 109.0 degrees with C(1) and O(2) since in the connectivity list, A and B are defined to be atoms 1 and 2, respectively. This is the angle 6-1-2.

Dihedral I-A-B-C: The H(6)-C(1) bond is rotated 60.0 degrees clockwise from the O(2)-H(3) bond when one looks down the O(2)-C(1) bond, with the O(2) closer to the eye. See Fig 2. This is the dihedral 6-1-2-3.

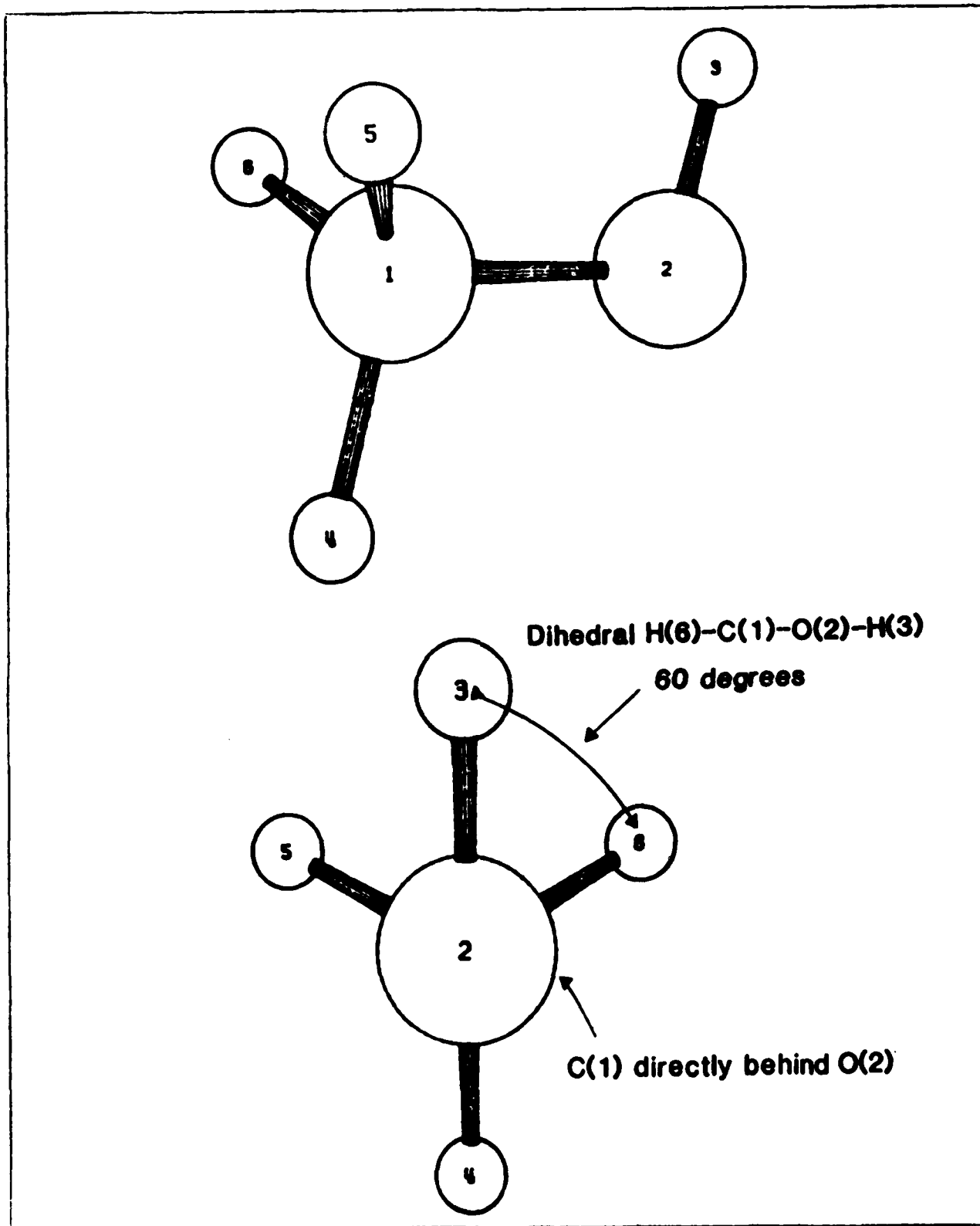


Fig. 2.  $\text{CH}_3\text{OH}$  geometry.

The number "1" following a coordinate tells MOPAC to optimize that coordinate. A "0" in this position tells MOPAC to keep that coordinate fixed. There must be at least one space between each entry in a row, but we normally put in more for ease in reading the data file. The 0's in the rows for O(2) and H(3) are used to indicate that the entries can not be defined and are there only to maintain format/spacing requirements up to the connectivity list.

The atom numbers and the column headings shown in the geometry description above cannot be put in the data file. Only the part within the dashed line is entered. Thus the completed CH3OH.DAT file will look like this:

T=3600 AM1 PRECISE  
 AM1 CALCULATION OF HEAT OF FORMATION OF METHANOL WITH PRECISE

C									
O	1.43	1	0	0	0	0	1		
H	0.96	1	108.0	1	0	0	2	1	
H	1.10	1	109.0	1	180.0	1	1	2	3
H	1.10	1	109.0	1	-60.0	1	1	2	3
H	1.10	1	109.0	1	60.0	1	1	2	3

(NOT FORGETTING THAT THE KEYWORD LINE IS THE FIRST LINE IN THE FILE) Now a command is entered to start the calculation or at least submit it to a queue where it will wait its turn (e.g., MOPAC CH3OH 60 on a VAX, where the 60 is an indicator of the length of time estimated for the job and thus the queue to which the job should be sent).

Next you go for a cup of coffee, lunch, or a holiday, depending on how many atoms are in the system, the kind of computer you're using, and how busy

it is. With CH3OH using a not-too-busy VAX, you'll probably have just enough time to pour yourself a cup of coffee before the job is done. Assuming it has been successfully completed you should now have some additional files in your directory, e.g., CH3OH.OUT and CH3OH.ARC on the VAX. Next let's take a look at how the results of the calculation are presented and what they mean.

b. Reading the output file, CH3OH.OUT.

It must be kept in mind that while the kind of output and the order in which it is presented in this file is fairly typical, changes in keywords can make an output file look quite different. Understanding the parts of the file we will discuss below will be very helpful in reading any MOPAC output file.

(1) After some initial labeling, the KEYWORDS are listed. These should be checked to insure that this is the result of the calculation you think it is!

(2) Next the essential ingredients of the data file used are presented. This listing, especially the geometry definition, lets you know on what system this calculation was done.

(3) Following the listing of the CARTESIAN COORDINATES is a statement of how many singly and doubly occupied energy levels (or MO's) were calculated for the system. This can be revealing if you expected the system to be a singlet in the ground state and find that there is a singly occupied MO.



(4) Next come the interatomic distances for your initial geometry followed by some information on how many cycles were required before the tests were passed which signalled the calculation to end. The GNORM (GRAD) and  $\Delta H_f$  (HEAT) at the end of each cycle are also given.

(5) Then come the main results of the calculation, the FINAL HEAT OF FORMATION and IONIZATION POTENTIAL, both of which can be readily compared with experimental results. There is also some other self-explanatory information given on the system and the calculation. The calculated value of -57.0 kcal/mol does not agree very well with the experimental value of -48.2. The error of -8.8 kcal/mol is larger than the average absolute error in  $\Delta H_f$  obtained using the methods included in MOPAC. A brief table giving some examples of these errors is given below.

Examples of Average Errors

Species Containing:	AM1			MNDO		
	$\Delta H_f$ (kcal/mol)	bond length, Å	bond angle	$\Delta H_f$ (kcal/mol)	bond length, Å	bond angle
CHNO*	5.5	0.014	3°	6.3	0.014	2.8°
CH	5.1			6.0		
F				9.8		
Cl				5.6		
Br				7.2		
I				5.8		
S				7.7		

\*138 compound test set.

(6) Next is the listing of the internal coordinates for the final optimized geometry. Rather than studying this listing, however, we normally use the DRAW program discussed in Section 4 to examine the geometry. Again it is important to know how well MOPAC methods predict geometries. The table above indicates the kinds of errors in geometries you can expect using MOPAC methods.

(7) The next listing, INTERATOMIC DISTANCES, gives calculated distances between the atoms in the final optimized geometry. Here again DRAW is usually a more desirable vehicle for studying this aspect of the system because of the visual displays it presents.

(8) The energies (in eV) of the MO energy levels are then printed under the heading, EIGENVALUES. Each value corresponds to an MO which contains at most two electrons. Thus for degenerate orbitals, two values which are virtually identical will appear in succession. Since the number of filled levels is given (See (3) above), the gap between the highest occupied MO (HOMO) and lowest unoccupied MO (LUMO) can be determined and related to reactivity or other properties of the system. Note that these are the MO's formed only from the valence electrons of the atoms in the system.

(9) The ATOM ELECTRON DENSITY given in the next table is obtained for each atom as follows. The table of ATOMIC ORBITAL ELECTRON POPULATIONS (two tables further down in the CH3OH.OUT file), gives the electron population of each atomic orbital in the basis set used to construct the LCAO MO's. They are given in the same order in which the atoms in the system are numbered.

Thus the first 4 values give the electron population for the 2s, 2px, 2py, and 2pz orbitals of C(1). (Only the valence atomic orbitals are included.) These values are the diagonal elements of the DENSITY MATRIX, which is essentially the square of the MO's obtained from the SCF calculation. The sum of these four electron populations gives the ATOM ELECTRON DENSITY for C(1), 4.0733 in the CH3OH file. Since these four orbitals would carry a charge of 4.0000 in the unbonded carbon, the CHARGE column shows a net charge on C(1) of -0.0733. The same analysis for the other atoms in the molecule gives the overall charge distribution of the system.

(10) While the charge distribution cannot be directly compared with experimental data, the total dipole moment of the molecule can be calculated from these results and compared with the experimental result. It is given in the lower right corner of the table below the one on NET ATOMIC CHARGES AND DIPOLE CONTRIBUTIONS. For this system it is 1.621 debyes.

(11) Finally there is the table showing the optimized geometry in cartesian coordinates.

The CH3OH.ARC file is an abbreviated version of the .OUT file which is convenient for long term storage of results. It is also the file which is normally used by the DRAW program to display the results of a MOPAC calculation in graphic form. We now discuss how DRAW can be used.

#### 4. How to use DRAW.

DRAW is a program designed to give graphical displays of systems whose

geometry definitions are contained in output and data files in MOPAC. The DRAW Manual gives a complete description of the many capabilities of this program. Within the software itself there is a HELP routine which can be called to assist you while you are using it. In this workshop we will demonstrate some of the more commonly used functions of DRAW.

a. Drawing the initial geometry.

Drawing the initial geometry in a data file is almost always done to insure that the GEOMETRY DEFINITION is what you intended. After exiting the data file, you issue the command, DRAW, to get into the program. In response to a prompt, you enter the filename, CH3OH.DAT, to bring that file into the program. Next, the letter "T" followed by the name for the graphics terminal you are using is entered, e.g., T 4107 for a Tektronix 4107 terminal. This produces a picture of the system in the data file using a stick model with atoms represented by the numbers they were assigned in the file. Now there are a number of ways you can manipulate this picture to get a better look at the geometry. Some commonly used examples are given below.

(1) You can rotate the molecule by issuing a command such as ROTATE CARTESIAN 2 0 90 45, which will rotate the molecule around the number 2 atom in cartesian coordinates 0, 90, and 45 degrees about the x, y, and z axes, respectively. Or you can issue the command, ROTATE PAIR 1 2 90, which will rotate the system 90 degrees about the axis connecting the pair of atoms, 1-2.

(2) You can enter the DISPLAY mode (entering DISPLAY) and change the style of the picture. For example, a popular ball and stick picture

designated ORTEP can be produced by entering DISPLAY ORTEP QUIT, where QUIT tells the program you want it to quit the display mode and execute the change (to ORTEP). The HELP routine gives other styles which are available.

(3) You can change the way the atoms are labelled by entering DISPLAY LABEL, which results in a prompt asking you whether you want atom labels to be their chemical symbols (enter SYMBOL), removed altogether (enter MASK), to be user-defined (enter USER), etc. Again, once you have specified your choice, enter QUIT to put the change into effect.

(4) Bonds are drawn between atoms which are close enough to be within the sum of their Van der Waals radii from each other. This criterion can be altered by entering DISPLAY BONDS, which results in a prompt for you to select a choice of different criteria, including simply removing bonds which you do not wish to appear. Again QUIT puts the changes into effect.

(5) With the system drawn on the terminal you can get information about the geometry by entering INFORMATION to get into the INFORMATION mode. Then, for example, you can find the distance between any two atoms. Entering 2 6 immediately gives the distance between atoms 2 and 6. Entering 2 3 6 gives the angle between 2, 3, and 6. Entering 1 2 3 6 gives the dihedral angle.

b. Drawing the optimized geometry.

Drawing the optimized geometry following a successful calculation is obviously the best way to see how it turned out. You generate the picture in

the way described above for the data file, here using the .ARC or .OUT files. It can then be manipulated in the same way as described above for the data file picture. But there are other things you can do with an .OUT or .ARC file, some examples of which are given below.

(1) Entering the INFORMATION mode with the command I I Q gives a summary of the information contained in an ARC file. Or you can ask for other specific information about the system as outlined in the HELP routine.

(2) If the ARC file contains several optimized geometries because a reaction path calculation was done (more on this in the next section), then the various points along the reaction path can be viewed in sequence by entering NEXT after each one is shown.

(3) If the KEYWORD, "FORCE," was used to get the vibrational modes of the molecule (more on this later), then the .OUT file must be brought into DRAW. Now the various modes can be viewed by entering a command such as DISPLAY VIBRATION 14 QUIT, which results in the molecule being displayed in stick style with arrows indicating the direction of motion of the atoms for the 14th normal mode.

(4) The geometry of the system can be changed in the DRAW program using the EDIT command. In the EDIT mode atoms can be removed or changed into atoms of other elements. Bond lengths, angles, and dihedrals can be changed, and atoms can be added. The new geometry is drawn immediately upon QUITting the EDIT mode. A data file corresponding to the new geometry is generated by issuing the OUTPUT command.

We have now seen how the DRAW program can be used to help interpret and present MOPAC results. In the next section we will work through some examples of MOPAC calculations that will help to improve your understanding of the capabilities we have discussed so far. We will also introduce some new functions and approaches that should help you begin to appreciate the wide-ranging powers of this program.

## 5. Examples of applications.

### a. Naphthalene and the keyword, SYMMETRY.

The data file and the DRAW-generated geometry of this file for naphthalene are shown in Fig. 3. You should satisfy yourself that the geometry shown is consistent with the geometry definition in the data file. The keyword, SYMMETRY, and the additional set of numbers below the geometry-ending "blank line" are the only new features beyond what has already been introduced.

Our knowledge of molecular structure tells us that many of the geometric features of this molecule are equivalent due to its symmetry. For example, we know that the bonds 3-2, 7-2, 5-1, and 8-1 must all be the same length in the equilibrium geometry. We also know that the angles 3-2-1, 5-1-2, 7-2-1, and 8-1-2 must be equal. For reasons which we will discuss later, it is advantageous to use this knowledge by optimizing one of the coordinates in each of these equivalent sets and fixing the others in the set at the optimized value.

In this example the bond length 3-2 is flagged with a "1", indicating that this coordinate is to be optimized. The distances 7-2, 5-1, and 8-1 are

t=3600 SYMMETRY  
naphthalene

C	0.0	0	0.0	0	0.0	0	0	0	0
C	1.4	1	0.0	0	0.0	0	1	0	0
C	1.4	1	120.0	1	0.0	0	2	1	0
C	1.4	1	120.0	1	0.0	0	3	2	1
C	1.4	0	120.0	0	0.0	0	1	2	3
C	1.4	0	120.0	0	0.0	0	5	1	2
C	1.4	0	120.0	0	180.0	0	2	1	5
C	1.4	0	120.0	0	180.0	0	1	2	3
C	1.4	0	120.0	0	0.0	0	8	1	2
C	1.4	0	120.0	0	0.0	0	7	2	1
H	1.1	1	120.0	1	180.0	0	3	2	1
H	1.1	1	120.0	1	180.0	0	4	3	2
H	1.1	0	120.0	0	180.0	0	5	1	2
H	1.1	0	120.0	0	180.0	0	6	5	1
H	1.1	0	120.0	0	180.0	0	7	2	1
H	1.1	0	120.0	0	180.0	0	8	1	2
H	1.1	0	120.0	0	180.0	0	9	8	1
H	1.1	0	120.0	0	180.0	0	10	7	2
O	0.0	0	0.0	0	0.0	0	0	0	0

3	1	5
3	1	7
3	1	8
4	1	6
4	1	9
4	1	10
3	2	5
3	2	7
3	2	8
4	2	6
4	2	9
4	2	10
11	1	13
11	1	15
11	1	16
12	1	14
12	1	17
12	1	18
11	2	13
11	2	15
11	2	16
12	2	14
12	2	17
12	2	18

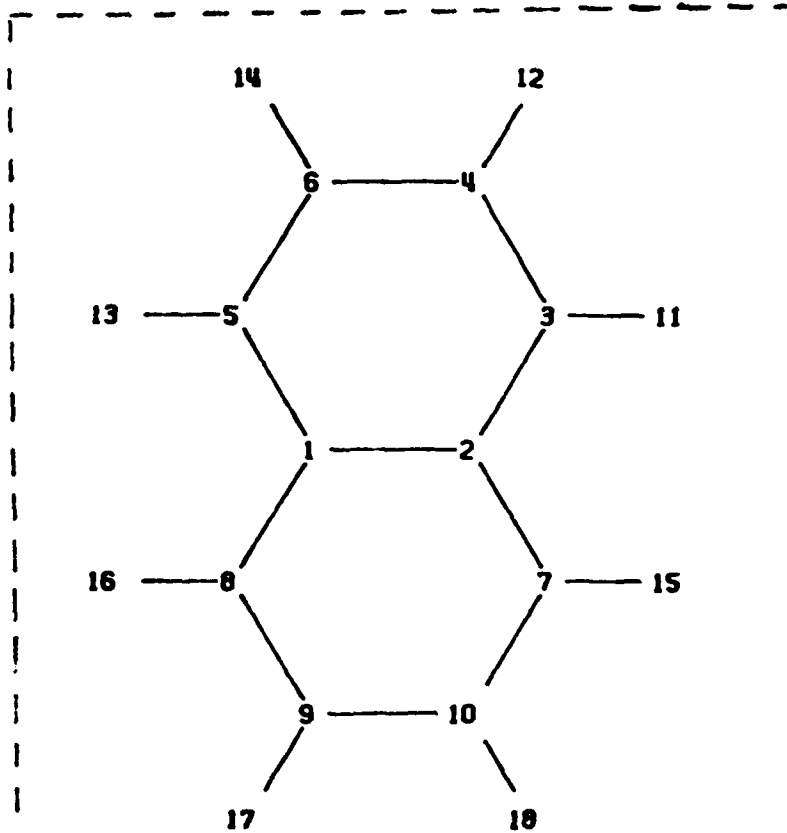


Fig. 3. Naphthalene geometry definition with symmetry.



flagged with "0" to indicate that they are not optimized by the normal procedure. Instead, we will include symmetry definitions in the data file to assign them the same value as the 3-2 optimized value. How is this done? Note the first three lines of the symmetry numbers below the "blank line." The first, "3 1 5", tells MOPAC that the distance coordinate for atom 5 (in this example, 5-1, as seen from the connectivity list) must be assigned the same value as the optimized value for the distance coordinate for atom 3 (in this example, 3-2, as seen from the connectivity list). The number "1" in between the "3" and the "5" on the line is a code that identifies the equivalent coordinates as a distance. The next two lines, "3 1 7" and "3 1 8", say the same rule applies to the 7-2 and 8-1 distances. The seventh through ninth lines in the symmetry numbers, "3 2 5", "3 2 7", and "3 2 8", have the code number "2" for "angle" in the middle. Thus the angle 3-2-1 will be optimized (as indicated by the "1" flag in the geometry definition) and the angles 5-1-2, 7-2-1, and 8-1-2 (all of which are flagged with "0") will be assigned the optimized 3-2-1 value.

To summarize what we did in this specific example of applying the SYMMETRY function: the middle number of each line indicates which type of coordinate is being "symmetrized," the first number tells which atom will have that coordinate optimized, and the third number tells which atom will have that coordinate set equal to the equivalent coordinate of the optimized atom. In other applications, the atom corresponding to the first number (called the "reference atom") may not be optimized, and the middle number may represent a more complex relation than simple equality. A complete description of the ways the SYMMETRY functions can be used is given in Chapter 2 of the MOPAC Manual.

Note that the numbering of the atoms in this example is not the conventional IUPAC numbering which might seem more desirable. The numbering scheme was chosen to allow the maximum number of equivalent sets of coordinates to be "symmetrized." An ability to find the best numbering scheme will require some practice with the SYMMETRY function.

But why use symmetry? First it can save significantly on the time required for a calculation, a factor which obviously becomes more important with larger systems. A not so obvious advantage is that the appearance of exactly equal coordinates in the tabulated output geometries makes the geometries easier to read. It also results in marginally better geometries and heats of formation. When IR spectra are calculated from "symmetrized" geometries, degenerate frequencies come out to be almost exactly equal, whereas differences of up to 20 wavenumbers may result when "unsymmetrized" geometries are used. More advanced uses of SYMMETRY involve finding transition states which are known or suspected to be symmetrical and in defining reaction paths.

b. Vibrational modes and IR spectrum of  $\text{Al}_2\text{Cl}_7^-$ .

MOPAC generates many results which can be compared with experimental data. One of the most interesting of these is the vibrational frequencies of a system calculated by using the keyword, FORCE. In this example we will demonstrate how this function is used to generate the IR spectrum of the  $\text{Al}_2\text{Cl}_7^-$  ion.

We begin by calculating an optimized geometry for  $\text{Al}_2\text{Cl}_7^-$  in the same way we did the  $\text{CH}_3\text{OH}$  molecule. We can then create a data file to do the FORCE calculation by editing the resulting .ARC file. We simply delete all of the file except the keywords and geometry definition which appears in the .ARC file under the "FINAL GEOMETRY." We add the keyword, FORCE, to produce the data file shown below:

```

FORCE CHARGE=-1 T=1000
CALC OF AL2CL7- - FORCE CALC
Al      0.000000  0      0.000000  0      0.000000  0      0      0      0      0.9483
Cl      2.109516  1      0.000000  0      0.000000  0      1      0      0      -0.4071
Cl      2.110764  1      113.297184  1      0.000000  0      1      2      0      -0.4220
Cl      2.110761  1      113.200224  1      129.047021  1      1      2      3      -0.4210
Cl      2.295359  1      109.799824  1      -115.040056  1      1      2      3      -0.3013
Al      2.297374  1      125.471304  1      -1.177332  1      5      1      2      0.9419
Cl      2.121224  1      101.504516  1      -179.671620  1      6      5      1      -0.4249
Cl      2.112476  1      107.241245  1      -61.477832  1      6      5      1      -0.4125
Cl      2.112740  1      107.168291  1      62.123575  1      6      5      1      -0.4126
0      0.000000  0      0.000000  0      0.000000  0      0      0      0

```

(NOTE: The charges on the atoms printed in the extreme right column, which come with the .ARC file, do not need to be deleted for the data file to work.)

Performing a MOPAC calculation on this file results in an output file which initially presents the same kind of information we discussed for the  $\text{CH}_3\text{OH}$  output file. After the INTERATOMIC DISTANCES, we see a new listing, the INTERNAL COORDINATE DERIVATIVES, which shows how the energy varies with changes in the internal coordinates of the optimized geometry. These should (and almost certainly will) be less than 0.5 kcal/mol/Å. Closely following this are two listings of the "PRINCIPAL MOMENTS OF INERTIA" and the "ORIENTATION OF MOLECULE IN FORCE CALCULATION." The latter gives the cartesian coordinates of the atoms with the system center of gravity at the

zero of the coordinate system. Next comes a list of the steps used to set up the Hessian which is diagonalized to get the force constants. A condensed form of the Hessian follows in the table designated "FORCE MATRIX IN MILLIDYNES/ANGSTROMS." The numbers in this table indicate the strength or "stiffness" of the bonds connecting the atoms in the molecule. Next is a listing of the results of the "NORMAL COORDINATE ANALYSIS." This is of limited value for most users.

Finally we get to the key result, the "DESCRIPTION OF VIBRATIONS." The first few vibrations may not be listed if their frequencies are less than 50  $\text{cm}^{-1}$  since these tend to have large uncertainties. Those modes with higher frequencies are listed in order of increasing frequency. We will discuss only the main features of this listing; others are discussed in Chapter 5 of the MOPAC Manual.

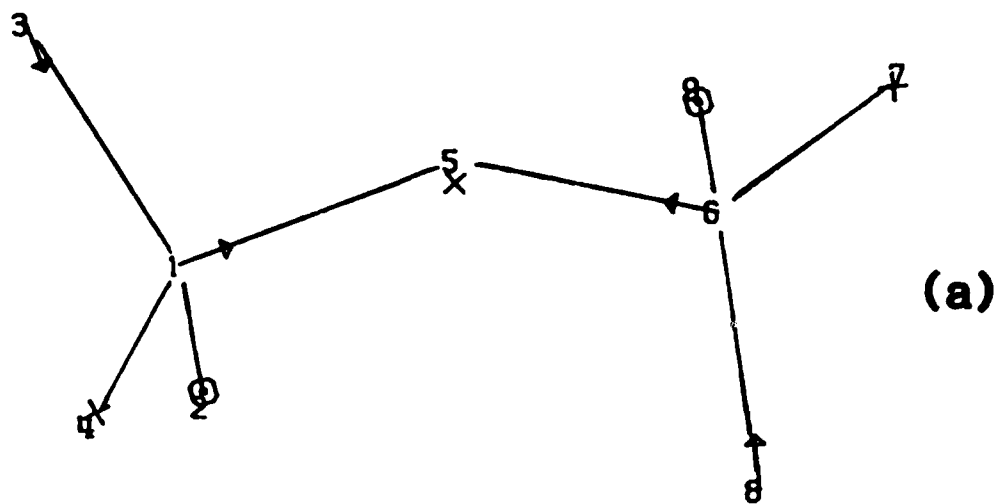
VIBRATION		ATOM PAIR	ENERGY CONTRIBUTION	RADIAL
14				
FREQ.	313.89	Cl 5 -- Al 6	29.1% ( 99.6%)	64.2%
T-DIPOLE	1.0483	Al 1 -- Cl 5	19.5%	61.9%
TRAVEL	0.0800	Al 1 -- Cl 2	11.0%	35.9%
RED. MASS	16.7786	Al 6 -- Cl 8	10.7%	28.6%
		Al 6 -- Cl 9	10.7%	28.5%
15				
FREQ.	348.16	Al 1 -- Cl 5	35.9% ( 78.9%)	86.6%
T-DIPOLE	4.9923	Cl 5 -- Al 6	35.7%	86.3%
TRAVEL	0.0753	Al 1 -- Cl 2	4.9%	17.0%
RED. MASS	17.0540	Al 1 -- Cl 4	4.0%	20.0%

The "FREQ" in  $\text{cm}^{-1}$  and "T-DIPOLE" (transition dipole in Debyes) are given in the left column as seen for the examples shown above for the vibrations, 14 and 15, for  $\text{Al}_2\text{Cl}_7^-$ . The next column lists all pairs of atoms that contribute significantly to the energy of the mode. In the next column is an indication of how much each pair contributes to the energy of the mode. A discussion of

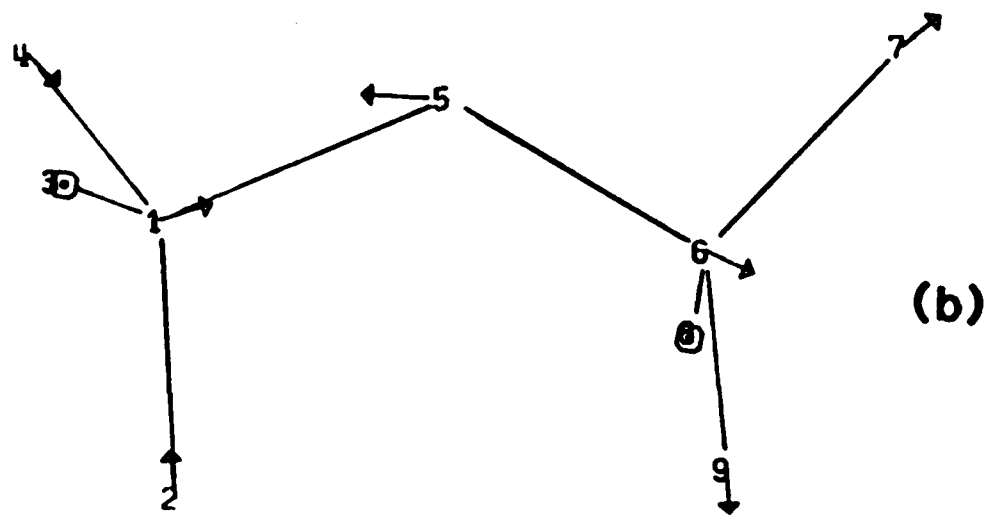
how these numbers are calculated and what they strictly mean is given in the Manual. For the beginning user, the following may suffice. The energy of a photon stimulating a given mode is distributed throughout the molecule. Pairs of atoms with the highest "ENERGY CONTRIBUTION" are seen as absorbing most of that photon's energy. The number in parentheses following the top pair is the energy of that pair divided by the total energy of the mode. If the 78.9% given for the Al1-C15 pair in vibration 15 seems high (especially since the second pair, C15-Al6, has about the same value for the ENERGY DISTRIBUTION), it need only be recognized that both pairs share a common atom, C15, so that in effect there is double counting in this method for describing the way the energy is distributed. (It is probably small consolation that other methods of description are even less satisfying!)

The last column gives a measure of how much the motion of the atoms in the pair is radial (as occurs in pure stretching) for the mode. The smaller this number is, the more the motion between the atoms can be characterized as bending or wagging. This information is useful in describing the mode in traditional terms.

Vibrations 14 and 15 can be reasonably assumed to be the symmetric and asymmetric stretches in the Al-Cl-Al bridge. Diagrams of these modes can be obtained by using DRAW on the output (not .ARC) file. The command, DISPLAY VIBRATION 14 QUIT, in DRAW gives the picture of the mode shown in Fig. 4.a. Vibration 15 is shown in Fig. 4.b. Fig. 5 shows the calculated frequencies and their intensities (obtained as T-DIPOLE squared multiplied by the frequency) for  $\text{Al}_2\text{Cl}_7^-$  compared to the IR spectrum obtained experimentally.

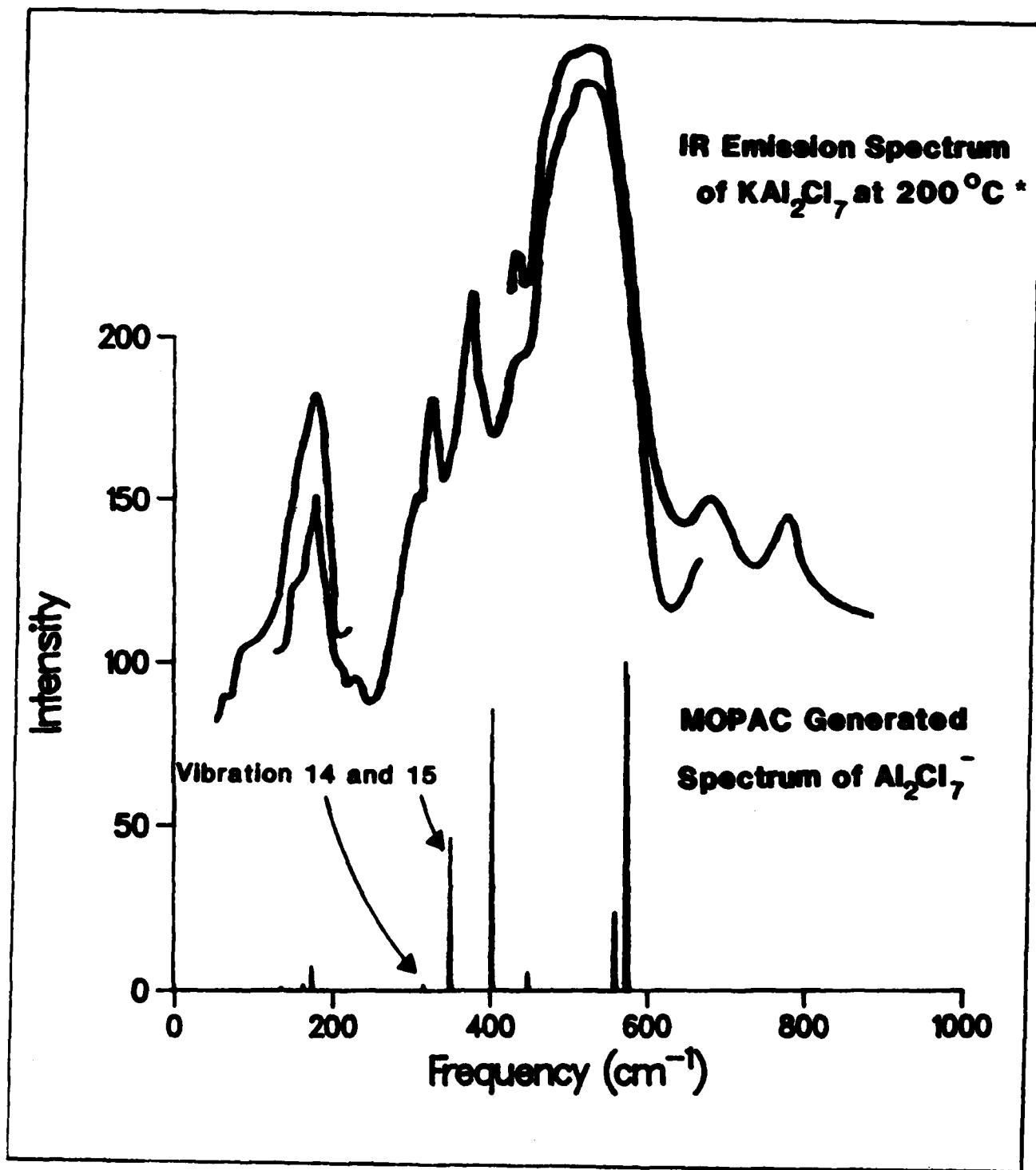


**Vibration 14: Al-Cl-Al symmetric stretch**  
 $313.8 \text{ cm}^{-1}$



**Vibration 15: Al-Cl-Al asymmetric stretch**  
 $348.16 \text{ cm}^{-1}$

Fig. 4.  $\text{Al}_2\text{Cl}_7^-$  vibrations

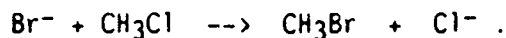


\* Hviistendahl, J.; Klæboe, P.; Rytter, E.; Øye, H. A. *Inorg. Chem.* 1984, 23, p. 706.

Fig. 5. Al<sub>2</sub>Cl<sub>7</sub><sup>-</sup> calculated vs experimental IR spectra.

c. Reaction path calculation.

In this example we consider the problem of finding an energy of activation for the reaction,



The key to determining an activation energy is identifying and calculating the  $\Delta H_f^\ddagger$  for the transition state for the reaction. Usually in MOPAC calculations, this involves a three step process. First, we get an approximate geometry for the transition state by finding the maximum in a reaction profile (i.e., a plot of  $\Delta H_f^\ddagger$  for the reacting system vs. the selected reaction coordinate). Next we refine that geometry by locating a stationary point at that maximum (i.e., a point at which the first derivative of the energy with respect to the coordinates approaches zero). Finally we characterize that stationary point to determine if it is indeed a transition state. We will now briefly describe how each of these steps can be carried out for our example.

(1) Finding maximum in reaction profile.

We begin by creating a data file for the initial geometry of the reacting system as shown below.

T=7200 CHARGE=-1  
 REAX OF BR- WITH CH3CL

C	0	0	0	0	0	0	0	0	0
CL	1.8	1	0	0	0	0	1	0	0
H	1.2	1	109.0	1	0	0	1	2	0
H	1.2	1	109.0	1	120.0	1	1	2	3
H	1.2	1	109.0	1	-120.0	1	1	2	3
BR	20.0	-1	180.0	1	0.0	1	1	2	3

10.0 5.0 4.0 3.5 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



The keyword, CHARGE=-1, gives the total charge of the system, which initially (with the Br<sup>-</sup> well removed from the CH<sub>3</sub>Cl) resides entirely on the Br<sup>-</sup>. All the coordinates in the geometry description are flagged with "1", indicating they are to be optimized, except for one. This one coordinate, the Br<sup>-</sup>-C distance, has been chosen to be the "reaction coordinate." It is flagged with a "-1" to indicate that it will be initially fixed at the 20.0 Å specified in the geometry definition. ΔH<sub>f</sub> for this initial configuration will be calculated for the geometry in which all the other coordinates have been optimized. Then the Br<sup>-</sup> will be fixed at shorter distances from the C and the process will be repeated. The values at which the Br<sup>-</sup>-C distance will be fixed are specified in the line immediately below the "blank line" at the end of the initial geometry definition. The Br<sup>-</sup>-C distances should be decreased to beyond the value it would be expected to have when the reaction has occurred. The intent here is to generate a reaction profile (that is, a plot of the system heat of formation as a function of the reaction coordinate) which passes through a maximum (see Fig. 6.a.). The transition state is then assumed to be near that maximum and further efforts can be made to locate it and characterize it as a transition state.

The output from this calculation is essentially a sequence of sets of results similar to the output from the calculation we did previously on CH<sub>3</sub>OH. At the very beginning of the output, the "POINTS ON REACTION COORDINATE" are listed. These are the values we selected to have the Br<sup>-</sup>-C distance fixed at. There will be a set of results for each fixed Br<sup>-</sup>-C distance, and these can be interpreted in the same way as previously discussed for the CH<sub>3</sub>OH output.

If the  $\Delta H_f$  values are plotted as a function of these fixed distances along the "reaction coordinate," we obtain the curve shown in Fig. 6(a). Note that there is a minimum in the curve before the maximum is reached. This corresponds to the Van der Waals complex formed by these two species. The three darkened points in the vicinity of the maximum indicate where we must look to find the transition state. Thus we next perform another reaction path calculation (with smaller steps) starting with the first of these points, at 2.4 Å, and ending with the last point at 2.2 Å. We can create the data file by editing the .ARC file from the first reaction path calculation. We simply delete from this file everything except the keywords and geometry definition of the "FINAL GEOMETRY" obtained for the 2.4 Å calculation. Next we add, after the "blank line," the points on the reaction path going through the maximum as shown in the data file below:

```

T=7200 CHARGE=-1
REAX OF BR- WITH CH3CL

C      0.000000  0      0.000000  0      0.000000  0      0      0      0      0.3632
Cl     1.972184  1      0.000000  0      0.000000  0      1      0      0     -0.5739
H      1.094551  1      97.129845  1      0.000000  0      1      2      0      0.0235
H      1.094488  1      97.227117  1     120.024835  1      1      2      3      0.0234
H      1.094510  1      97.215600  1    -120.033132  1      1      2      3      0.0234
Br     2.400000 -1     179.917573  1     -5.614908  1      1      2      3     -0.8595

2.38 2.36 2.34 2.32 2.30 2.28 2.26 2.24 2.22 2.20

```

The results of this calculation are plotted in Fig. 6(b). We are now in a position to make an attempt to confirm what appears to be a valid transition state.

(2) Locating a stationary point at the maximum.

We choose the point just before the maximum in the curve, at 2.26 Å, as the geometry which we will now allow to fully optimize. This means we will

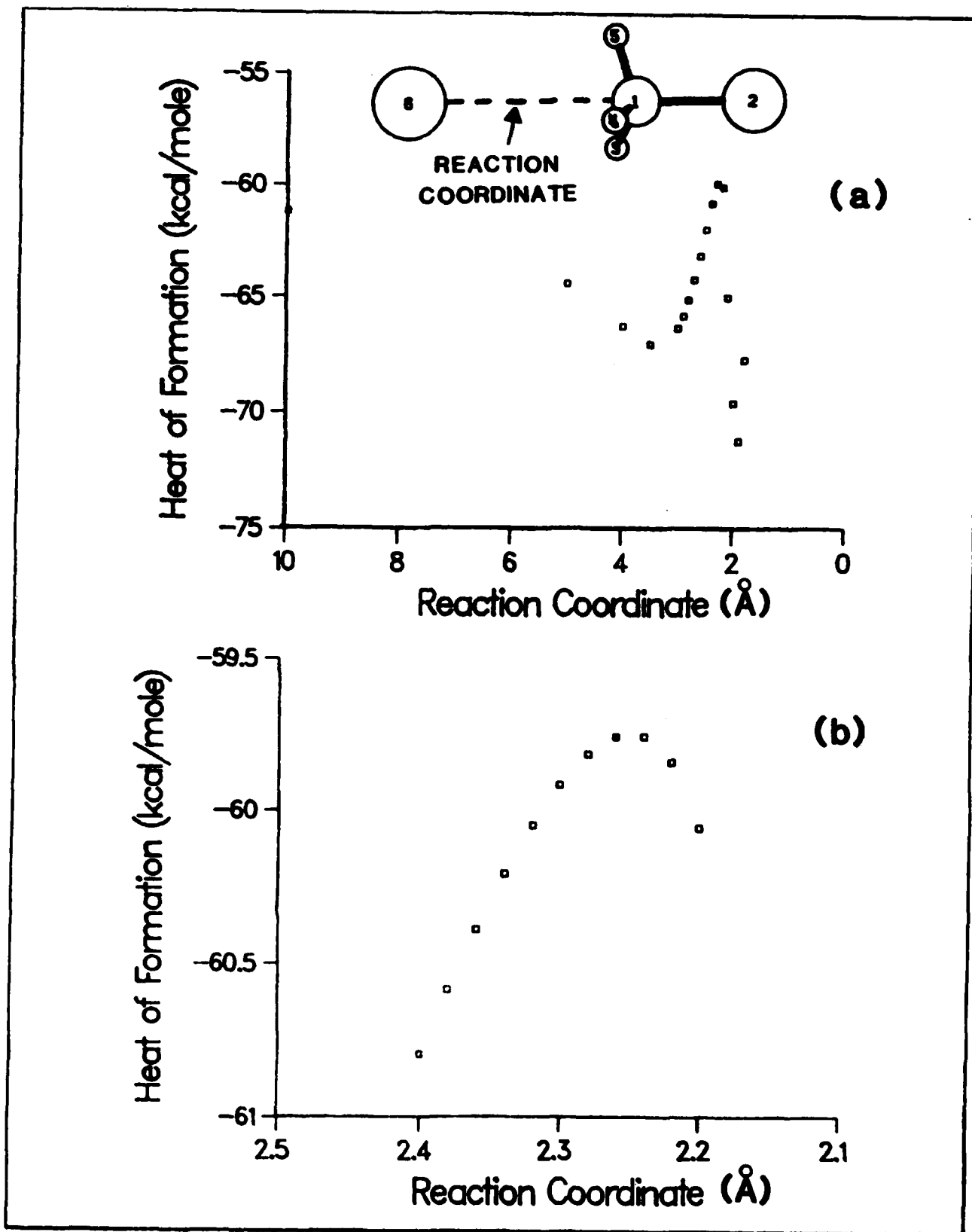


Fig. 6.  $\text{Br}^- + \text{CH}_3\text{Cl}$  reaction profiles.

not fix the Br-C distance. However we must also tell MOPAC that we do not want to find the minimum in  $\Delta H_f$  from this starting geometry, which would take us back down the reaction profile. What we want is the maximum on the reaction profile, i.e., the transition state. To do this, we must use a keyword (we will use NLLSQ) which tells MOPAC to minimize the GNORM, the derivative of  $\Delta H_f$  with respect to the coordinates. (Other keywords which can be used to locate the transition state are SIGMA and SADDLE, and these are discussed in the MOPAC Manual.) Again we edit the .ARC file from this last calculation to create a data file which has the keywords and geometry definition of interest; in this case, that for the 2.26 Å point. We change the "-1" flag in the .ARC file to a "1" for the Br-C distance to indicate that it will be optimized along with the other coordinates. Then we add the keyword, NLLSQ. If the calculation is successfully concluded, the resulting GNORM-optimized geometry and  $\Delta H_f$  correspond almost certainly to the transition state.

### (3) Characterization of the stationary point.

We now test whether the stationary point obtained in Step (2) is indeed a transition state. The final geometry is extracted out of the .ARC file and another data file created. The NLLSQ keyword should be removed and the keyword, FORCE, used. This gives the frequencies of the vibrations for the stationary state. If the transition state is valid, there will be one, and only one, vibrational mode with a negative frequency. In our example, vibration 1 is seen to be the only mode with a negative frequency. Fig. 7 shows a DRAW-generated picture of this mode, which effectively describes the actual reaction coordinate as well as the geometry of the transition state.

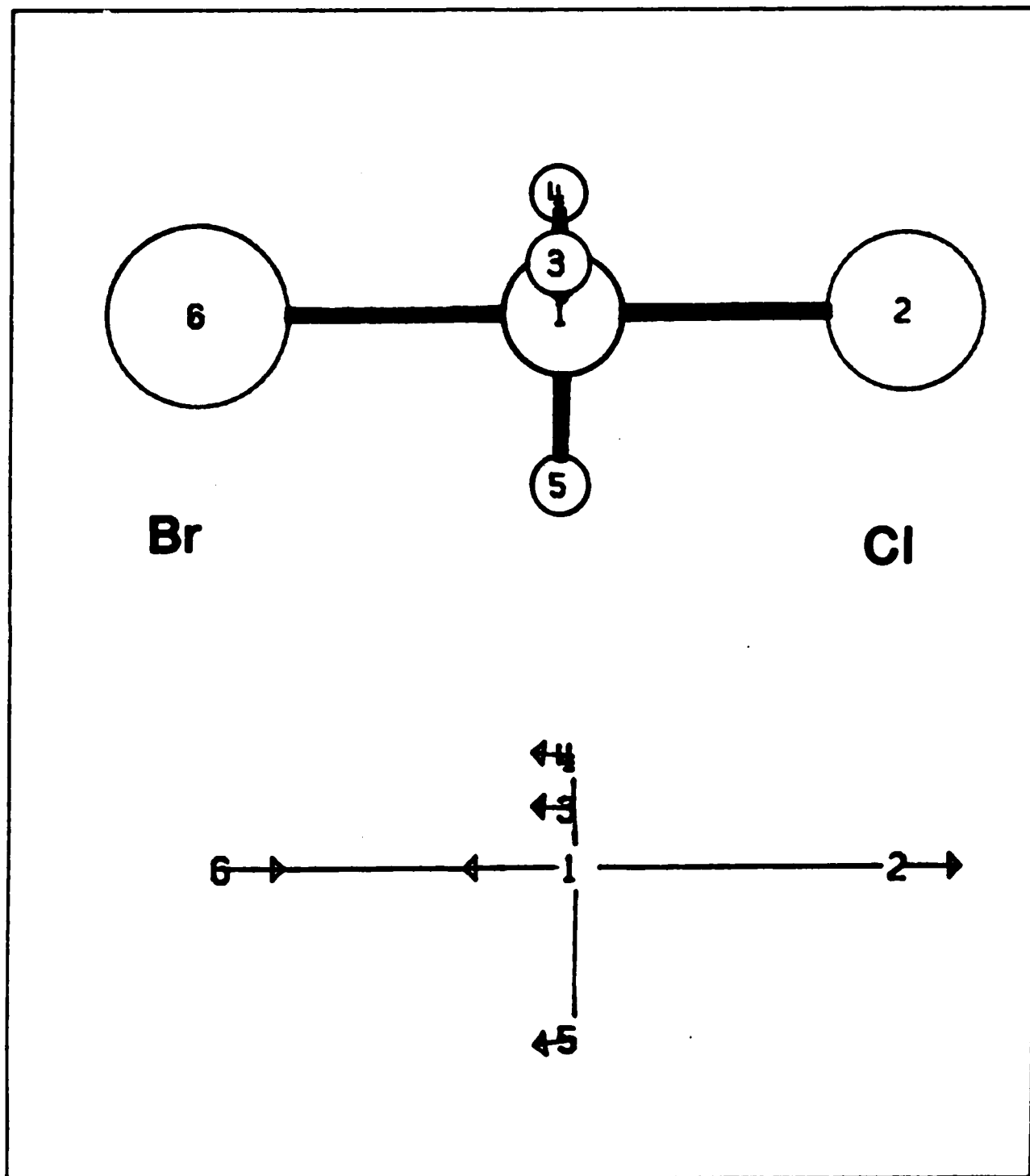
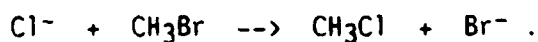


Fig. 7.  $\text{Br}^- + \text{CH}_3\text{Cl}$  transition state.

Note that the vibration corresponds to the dissociation of the transition state into CH<sub>3</sub>Br and Cl<sup>-</sup>.

Once the transition state has been fully confirmed, the energy of activation can be obtained as the difference in ΔH<sub>f</sub> between the fully separated species and the transition state. It is left as a useful exercise to find the transition state for the reverse reaction,



#### 6. Other considerations.

We now discuss some other aspects of the MOPAC program which are not essential to operation at the introductory level but are very useful even to the beginning user.

##### a. Restart files.

If a job uses more than one hour of CPU time, a restart file, called FILENAME.RES, is created which saves the results which have been calculated in that hour. A file containing the density matrix, FILENAME.DEN, is also created. After each additional hour, a new set of these files is created to replace the previous one. If for any reason the calculation is terminated (e.g., a computer shutdown, failure to achieve a self-consistent field), the calculation can be picked up from the point at which the last restart file was created. This can be accomplished by adding the keyword, RESTART, to the data file for the calculation and resubmitting it. The geometry definition must

not be changed in this resubmitted file. If the calculation was stopped because of some problem with the calculation itself, the problem would normally be described at the end of the output file. In this case the keywords, RESTART and ISCF (which tells MOPAC to perform only one SCF and then stop), should be added to the data file. The .ARC file output from this calculation gives the geometry before the problem with the calculation forced it to end. This file probably can be edited to create a new data file which avoids the calculational problem.

b. Dummy atoms.

A dummy atom is a pure mathematical point which can be entered into a geometry definition in the same way that other atoms are entered. Coordinates for other atoms can then be specified by reference to the dummy atom. The symbol for a dummy atom is XX. Dummy atoms are used when their presence as a point of reference simplifies a geometry definition. This may be especially useful when there are more than one species in the system being calculated. It can be almost indispensable in the calculation of specific reaction paths. For example, suppose it is desired to follow the reaction of  $\text{Cl}_2$  and  $\text{CH}_2=\text{CH}_2$  with the C=C and  $\text{Cl}_2$  bonds approaching each other while remaining parallel to each other. Dummy atoms centered in the C=C and  $\text{Cl}_2$  bonds greatly simplify this reaction path calculation.

c. Sparkles.

Sparkles are computational constructs made to simulate unpolarizable ions of diameter 1.4 Å. The charge of the sparkle is set by the symbol used to enter it into the geometry definition: +, ++, -, or --. They have no heats

of atomization, no ionization potentials, and no orbitals. They are computationally designed so that other atoms or sparkles can approach them to within about 2 Å quite easily but can come closer than 1.4 Å only with great difficulty. The most obvious use of sparkles is to simulate counterions. Other uses are discussed in Chapter 6 of the Manual.

d. Condensed phase systems.

The species in a MOPAC model are surrounded by vacuum so that calculated results most closely approximate the behavior of the "real" species in the gas phase. MOPAC results can be applied to liquid phase system, but only with great caution. In aqueous systems, solvated species will in most cases have significantly different properties than what they have in the gas phase. This can be attributed to the high dielectric constant of water which significantly modifies electrostatic interactions. Furthermore, solute-solvent interactions change certain properties (or even the chemical nature) of both solute and solvent species. As a fairly obvious example, MOPAC predicts the system,  $\text{NH}_3 + \text{HCOOH}$ , to be stable in the gas phase. However, in aqueous solution, we know that solvated  $\text{NH}_4^+$  and  $\text{HCOO}^-$  are the stable species. In some cases, addition of solvent molecules to the model can be tried in order to predict behavior in solution, but this must be done very carefully. For species in low dielectric constant solvents, the effect on accuracy of not including solvent species in the model is less. Some properties, such as certain vibrational frequencies, can be fairly insensitive to phase, as experimental data have shown.



e. A final word.

The best way to develop your ability to use MOPAC effectively is to use it on systems with which you are familiar. When problems with interpretation or apparent bugs in the program arise which you cannot solve using the MOPAC Manual, you are encouraged to contact Dr. Jimmy Stewart at FJSRL.

Write to:           Dr. J. J. P. Stewart  
                      FJSRL/NC  
                      US Air Force Academy  
                      Colorado Springs CO 80840-6528  
                      USA

or call at:         (303) 472-2655 or AV 259-2655.

or:                   stewart@usafa.arpa on the arpa network.

MOPAC (the Manual and program) is revised on a nearly annual basis and the current version is always available from the Quantum Chemistry Program Exchange (QCPE); Department of Chemistry; Indiana University; Bloomington, Indiana 47405. Contact the Editor, Richard Counts at (812) 335-4784 for details.

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

END

DATE

FILMED

6-88

DTIC