

This document did not have its cover page. Here is what I know about it.
The software program is called MBLD. I believe that it was written by
M. S. Gordon and J. A. Pople in the 1970's.

Frances C. Bernstein

June 11, 2020

ABSTRACT

This program, from a minimal amount of input, builds a standardized geometric model of a molecule, and, using standard values of bondlengths, angles, and dihedral angles implicit in the program, calculates the cartesian coordinates of all atoms in the molecule. In addition, options are available which allow the user to choose his own geometric parameters.

TABLE OF CONTENTS

Abstract	i
Table of Contents	ii
Section A: Introduction	1
Section B: Input	3
Section C: MAIN Program	6
Section D: MBLD2	7
Section E: MBLD3	18
Section F: Limitations of the Standard Model	28
Section G: Timing and Accuracy	30
Appendix A: Sample Input	A1
Appendix B: Sample Output	B1
 Input Options: ATOMGEOM	 12
BONEROT	16
BONDLNTH	18
Z MATRIX	23
ELIM	27

A. Introduction

If one is to calculate properties of large molecules it is desirable to have some means of constructing standard geometric models of these molecules from their chemical formulae and a set of standard bondlengths and angles. While using such a standard model precludes the use of exact geometric parameters for any specific molecule, there is much to be said for the ability to calculate the cartesian coordinates of any molecule based on a well-defined set of rules. In the first place we may wish to discuss the electronic structure of a molecule for which there is little or no experimental data. Secondly, even for molecules whose geometries are known experimentally, a case can be made for using standard bondlengths and angles since individual geometric features are themselves caused by the electronic structure and more insight can be obtained by starting from a standard model and then discovering the driving force causing the deviation from such a model.

The "Model-Builder" computer program was written to meet the aims of the preceeding paragraph. It consists of two subroutines, MBLD2 and MBLD3. MBLD2 builds a qualitative model of the molecule, assigning bond types (single, double,...), local atom geometries (tetrahedral, trigonal, ...) and bond rotational geometries (trans, gauche,...). MBLD3 then assigns bondlengths, bond angles and dihedral angles commensurate with the qualitative model and calculates the cartesian coordinates of all atoms in the molecule. These coordinates may be obtained at two different levels of standardization. The first (Model A) distinguishes the various bond types and assigns different bondlengths accordingly. However, for some molecules one cannot uniquely define a single valence structure (for

example, many radicals and ions), and one then needs a less discriminatory model (Model B) which depends only on the types of atoms involved in a bond. Model B thus assigns the same bondlength for all bonds between any particular pair of atoms.

The program was written in FORTRAN IV for the IEM 360 Model 65. In addition, there are copies available which are compatible with the CDC 1604 (FORTRAN 63) and the Bendix G-21 (ALGOL 20). The program was originally written for the latter machine at Carnegie-Mellon University.

Since the coordinates, atomic numbers, charge, multiplicity, and numbers of atoms are all in a common block (COMMON/INFO), the program is easily meshed with other programs such as SCF routines. Alternatively, there is an option to punch this information on cards for future use. Finally, there are a number of options available (discussed in detail below) which allow the user to supercede the standard geometries. These options range from alteration of one bondlength to feeding in the entire set of bondlengths and angles. All important parameters and their specific functions are listed in Table 1, approximately in the order of their appearance in the program.

B. Input

The mandatory input data is read into the program in the following order:

Input to MAIN:

INOP,OUTOP,OP2,TIMES(4I3): Model-Builder options

Input to MBLD2:

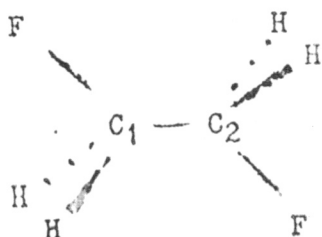
1. The name of the molecule plus any pertinent information (columns 2-72 of one card).
2. CHARGE,MULTIP(2I2): Charge and multiplicity of the molecule.
3. N(I2): The number of rows in SHORTF.
4. SHORTF(N cards, each in 7A4 format): SHORTF contains all bonding information necessary for the definition of a molecule. One card of input is required for each atom to which the user chooses to assign an explicit number. For each such atom, I, the corresponding row in SHORTF contains all atoms or groups of atoms bonded to I. If these latter atoms have themselves been numbered explicitly, then their numbers appear in the proper columns; otherwise, the input is their chemical symbols. All atoms may be numbered explicitly if one so chooses; however, any atom which is bonded to only one other atom (e.g., H, F, carboxyl oxygen) may be entered as its chemical symbol. This, of course, decreases the amount of input. In addition, the following groups may be entered by the symbols in parentheses: methyl (ME), ethyl (ET), n-propyl (NPR), isopropyl (IPR), n-Butyl (NBU), isobutyl (IBU) tertiary

butyl (TBU), and hydroxyl (OH). It should be noted that the explicit numbering is completely up to the user. There are no restrictive conventions. There are a number of sample sets of input listed in Appendix A.

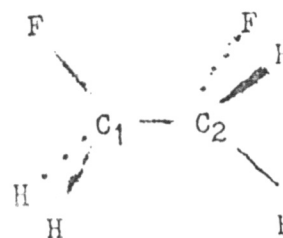
Each of the N rows of SHORTF may contain up to 7 entries. The first entry contains the chemical symbol of the corresponding atom, I, while the other entries contain the atoms or groups of atoms to which I is bonded. All entries are right justified.

5. Blank Card or geometry options followed by a blank card.

In addition to the bonding information described above, SHORTF also determines the bond rotational geometry of the molecule by the order in which the elements of each row appear in SHORTF. The convention is as follows: Consider a bond between two atoms X and Y, where X and Y are both bonded to more than one atom. Then X will appear somewhere in the row Y of SHORTF and Y will be one of the elements in row X. Unless it is directed to do otherwise, the program will then take that atom directly to the right of X in row Y, call it I, and that atom directly to the right of Y in row X, call it J, as the atoms which are defined by the bond rotational geometry of the bond X-Y. That is, if X-Y is trans then I is trans to J and all other atoms bonded to X and Y are then fixed by their bond lengths and bond angles. As an example, consider trans- and gauche-1,2-difluoroethane:



TRANS



GAUCHE

Note that only the two carbons have been numbered explicitly. SHORTF for trans would then be

COLUMN: 4 8 12 16 20

C 2 F H H

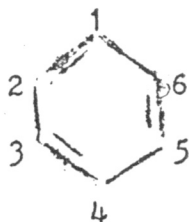
C 1 F H H

where 1 and 2 correspond to X and Y in the preceeding discussion and the two fluorines correspond to I and J. Since the standard model assigns a bond rotational geometry of trans to all such bonds, the two fluorines will be trans. Similarly, SHORTF for the gauche rotamer is

C 2 F H H

C 1 H F H

Finally, it should be noted that if X is the last entry in row Y or vice versa, the program goes back to the beginning of the row to find that atom directly to the right of X (or Y). Consider benzene as an example:



The corresponding SHORTF may be written

C	H	6	2
C	3	H	1
C	2	4	H
C	H	3	5
C	6	H	4
C	5	1	H

Here the hydrogen on Carbon 1 is trans to carbon 3 and so on. When setting up SHORTF for a molecule, one should be quite careful so that the desired rotamer is obtained. Some more complicated examples are contained in Appendix A.

C. MAIN Program

The MAIN program functions entirely as a calling program. All information pertinent to an SCF-MO routine is contained in one block of labeled common (COMMON/INFO). Initially the four control options are read: INOP, OUTOP, OP2, and TIMES. If INOP=0, the standard model is used and both MBLD2 and MBLD3 are called. If INOP=2, the experimental or some other set of alternate geometries are to be fed in and only MBLD3 need be called. If OUTOP>0 the number of atoms, atomic numbers, charge, multiplicity and coordinates are punched on cards. OP2=0 invokes Model A. OP2=1 invokes Model B. TIMES is simply the number of runs to be made using the present values of INOP, OUTOP, and OP2. If a serious error is detected in MBLD2, NALARM is set to 1 and the program is terminated.

D. MBLD2

This section of the program builds a qualitative model of the molecule. Once SHORTF has been read in, it is converted to a new array F. F is simply an expanded version of SHORTF in which all atoms not previously numbered are assigned numbers and atomic numbers. Thus all atoms appear explicitly in F. The atomic numbers are defined with the use of the matrix EL (which contains the chemical symbols of the elements) and are fed into AN. It should be noted here that it is possible for several of the matrices used in the program to obtain a zero subscript. To avoid this difficulty, the elements of all such arrays are incremented by one wherever they appear. Thus AN(I+1) is the atomic number of atom I, CN(I+1) the connectivity of atom I, and so on. In the case of AN this situation is rectified at the end of MBLD3 so that the program will be compatible with other programs to which it may be attached.

As an example of the conversion of SHORTF to F, consider SHORTF for ethane:

<u>COLUMN:</u>	4	8	12	16	20
	C	ME	H	H	H

Here only carbon, atom 1, has been numbered explicitly. The remaining atoms are always numbered in the order in which they appear in SHORTF; thus the atoms in the methyl group are numbered first and then the three hydrogens. Within the methyl group itself, the heavy atom (carbon) is first assigned a number and then the three hydrogens. The F matrix for this molecule would then be

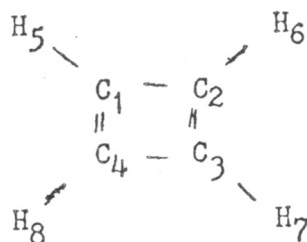
Atom		Array F		
1 (carbon)	2	6	7	8
2 (carbon)	1	3	4	5
3	2			
4	2			
5	2			
6	1			
7	1			
8	1			

Note that the three hydrogens in the methyl group have been numbered before those attached to carbon 1. If a group of atoms containing more than one heavy atom, such as ET, is encountered in SHORTF, the heavy atoms are numbered first and then their respective hydrogens. If the element of SHORTF has already been numbered it is simply carried over unchanged to F. Of course, if one or more elements of SHORTF is a group of atoms rather than a single atom, the rotational geometry of the molecule is not completely specified until SHORTF is converted to F.

Immediately following the conversion of SHORTF to F, the array NXCARD is read in a 20A4 format. If the standard model is being used NXCARD is simply a blank card and the final input; otherwise it contains the key word for one of the geometry alteration options to be discussed below. The connectivity of an atom, CN, is simply the number of atoms bonded to that atom. ATBD(J,I) is the Jth atom in bond I and defines all bonds in the molecule. Bond 1 is a fictitious bond with ATBD(1,1)=0 and ATBD(2,1)=1; thus the number of bonds in a molecule other than ring-closure bonds is just the number of atoms in the molecule. G is an array identical to F and is used to prevent bonds from being identified twice. A bond between I and J can be found both in row I and row J of F (since I is bonded to J and J is bonded to I); thus if G(K,I) is the position of J in row I and G(L,J) is the position of I in row J, both G(K,I) and G(L,J) are zeroed after the bond I-J is identified. By using G instead of F, the latter array is kept intact for later use. Once the first M bonds have

been identified, any remaining bonds must be those involved in closure of rings. These are easily identified since only those atoms participating in such bonds are left in G. By subtracting G from F, G will then contain all elements of F except those participating in ring-closure bonds. Hence G may now be used to construct all ring systems in the molecule.

ATRG(J,I) is the Jth atom in ring I. As an example of how rings are constructed by MBLD2, consider cyclobutadiene:



The first four rows of G will be

5	4	2
0	6	1
0	4	7
8	3	1

The ring closure bond in this molecule will be bond 2-3 (this is determined by the order in which ATBD is filled). MBLD2 then proceeds to build a tree structure of the ring beginning with ATBD(1, KK) where KK refers to the ring-closure bond, in this case bond 2-3. Thus the first row of TREE consists of atom 2. To find the second row the program picks out each atom bonded to 2. As each atom I is added to the tree structure it is checked off by setting CH(I+1)=1 so that I will not appear more than once in any given tree. The first two rows of TREE for cyclobutadiene will then be

```

      2
     / \
    6   1

```

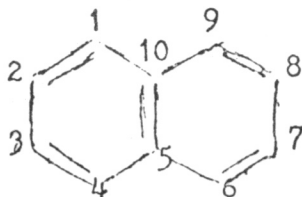
where atoms 6 and 1 are TREE(2,1) and TREE(2,2), respectively. The construction of TREE continues until ATBD(2, KK) is found (atom 3 in this case) at which point the ring has been completed:



Here the ring is 2-1-4-3. Thus by zeroing the occurrence of atom 3 in row 2 and atom 2 in row 3, we have forced the program to search backwards through the ring, identifying each atom in the ring as it proceeds.

TESTA is an integer which is initially zero. When a ring has been closed by the procedure outlined above, TESTA is given the value 1 so that the program knows to jump out of the loop controlling the tree construction. The ring is then numbered as are the atoms (using ATRG) in the ring. To allow for the possibility that two rings are defined by the same ring-closure bond, CH for ATBD(2, KK) is reset to zero as is TESTA (CH for all other atoms in the tree structure remain = 1) and the tree construction is allowed to continue to attempt to find another ring by an independent path. Finally, those elements of G pertaining to the ring-closure bond of the ring just identified are refilled. FROMA(I) indicates the origin of atom I; thus for cyclobutadiene, FROMA(2)=2 and FROMA(8)=4.

The procedure described thus far will find all rings in the molecule with one exception. If one considers naphthalene, it is obvious that there are two rings present:



12345 10 and 56789 10. The periphery of the molecule, 123456789 10, however, is not a ring and is prevented from being one by the check-off system outlined above. Since there are only two ring-closure bonds in naphthalene (3-4 and 7-8 for the numbering system above), the periphery "ring" has to be identified using the ring-closure bond of the "legitimate" rings. This, however, is prevented by the check-off system unless there are two

independent paths by which a ring-closure bond can be joined. This is not the case in naphthalene. This restriction, however, prevents the following type of ring structure from being specified properly:



I

There are three rings in the above structure, but with the above procedure only two will be identified. To get around this the program checks each pair of rings to see if they have three or more atoms in common. If they do a third, unidentified ring is present and the program identifies it. COMMAT(I) is the Ith atom the two rings have in common.

BDRG, RGAT, RGBD, and RGRG are filled quite straightforwardly once ATRG has been specified. Since rings can be attached to each other either by one atom or by a bond, RGRG is broken up into two parts: RGARG and RGBRG, corresponding to the two possibilities. For a molecule such as I, each ring is said to be bonded to the other two by a single atom at two different positions for each ring. The excess valence (XSVAL) of an atom is defined as the "normal valence" of the atom, VL, less the connectivity of that atom, CN, in the molecule. Thus if the "normal valence" of carbon is 4, then the excess valence of carbon in ethane, ethylene, and acetylene will be 0, 1, 2, respectively. In other words, the excess valence of the atom is qualitatively the number of electrons which that atom contributes to multiple bonds. Of course, this is an oversimplification with regard to true molecular structure, but for building qualitative models of molecules, it serves quite well.

Each atom in a molecule with connectivity greater than 1 is assigned a local atom geometry by the standard model; that is, all angles having atom I as the central atom are assigned the same standard value. These

standard geometries are determined by the following rules and filled into

ATGECM:

1. Any N or C atom with connectivity = 4 is tetrahedral; e.g. CH_4 .
2. A nitrogen or oxygen with connectivity 3 which is not bonded to any atom having a nonzero excess valence is pyramidal.

Examples are NH_3 , OH_3 , CH_3NH_2 .

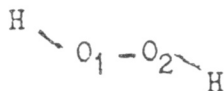
- 3a. A carbon with connectivity 3 is trigonal, e.g. CH_3 , C_2H_4 , ...
- 3b. A nitrogen or oxygen with connectivity 3 which is bonded to some atom having a nonzero excess valence is trigonal. Examples are CH_3NO_2 , vinyl amine, protonated vinyl alcohol.
- 4a. A carbon with connectivity 2 which is bonded to atoms the sum of whose excess valence is less than two is bent, e.g. CH_2 .
- 4b. A nitrogen with connectivity 2 which is bonded to atoms the sum of whose excess valence is less than three is bent, e.g. NH_2 , $\text{CH}_2 = \text{NH}$.
- 4c. An oxygen with connectivity 2 is bent.
- 5a. A carbon with connectivity 2 which is bonded to atoms the sum of whose excess valence is greater than 1 is linear. Example is acetylene.
- 5b. A nitrogen with connectivity 2 which is bonded to atoms the sum of whose excess valence is greater than 2 is linear, e.g. $\text{HC} \equiv \text{C} - \text{N} = \text{CH}_2$.

If the user wishes to change the geometry of one or more atoms, he may do so by setting the array NXCARD equal to the word ATCMGECM (left-justified) rather than a blank card. In other words, instead of having a blank card immediately following SHORTE in the input, the word ATCMGECM is placed in columns 1-80 so that $\text{NXCARD}(1) = \text{'ATCM'}$ and $\text{NXCARD}(2) = \text{'GECM'}$. The next card contains the following information:

1. The word 'TETR', 'PYRA', 'TRIG', 'BENT', 'LINE', or 'ASYM' in columns 1-4. The word chosen indicates the desired geometry.
2. The number of the atom I whose geometry is to be changed in columns 11-12, right-justified. ATGECM(I) is then set equal to the proper type of geometry.
3. If the geometry indicated is ASYM, the user has decided to assign a particular numerical value to the local geometry of atom I rather than one of the standard values. In this case the value of the angle is placed in columns 15-26 in F 12.8 format. This angle is then contained in NEWANG(I).

A similar card is inserted for each atom whose geometry is to be changed. The last card is again either a blank or the key word for the next geometry option to be invoked, again read as NXCARD.

An example using the ATOMGECM option is hydrogen peroxide, say the trans configuration:



The standard input for this molecule is

```

COLUMN:  2  4  8  12 16
          HYDROGEN PEROXIDE
          0  1
          2
           0  2  H
           0  1  H
          BLANK CARD

```

This will give both oxygens an atom geometry of 'BENT'. If we wish to change this to trigonal, the following is inserted before the blank card:

```

COLUMN:  1      8  12
          ATOMGECM
          TRIG      1
          TRIG      2

```

Now, the experimentally observed bondangle in hydrogen peroxide is 94.5° . To assign this value to both bondangles, the following cards are inserted:

```

COLUMN: 1      8  12 15
          ATOMGEOM
          ASYM      1  94.5
          ASYM      2  94.5

```

It should be reiterated here that when a particular atom geometry is assigned to an atom, all angles about that atom are assigned the same value. This restriction may be circumvented by reading in all bondlengths and angles in a manner to be discussed below.

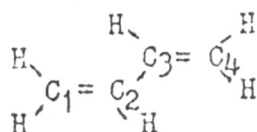
Once the atom geometries have been specified, MBLD2 proceeds to determine the bond types of all bonds in the molecule. If MODEL B is being used, these bond types are not needed and this section of the program is skipped if CP2 = 1.

EV is a dummy array identical to XSVAL and is used in the bond specification in a manner analagous to the use of G earlier. EVC(I+1) is the number of atoms bonded to I which have an excess valence greater than zero.

BT(I) is a number identifying the bond between ATBD(1,I) and ATBD(2,I). For BT(I)=1,2,3,4,5,6 the bond I is single, double, triple, aromatic, triple aromatic, dative, respectively. The only two types of dative bonds searched for are $\text{N} \begin{smallmatrix} \text{O} \\ \diagup \end{smallmatrix}$ and =N-N . Since these are special cases, all dative bonds are identified first. For NO_2 groups both oxygens will initially have EV=1 and the nitrogen will have EV \geq 2. Thus each time such a group is found, EV of both oxygens is deleted by 1 (so the bonds will only be identified once) and EVC of nitrogen is deleted by 2. Similarly, each time an N_2 group is found, EV of the end nitrogen is deleted by 2 and EVC of the other nitrogen is deleted by 1.

Once all dative bonds have been found there should be no atom having EV $>$ 0 and EVC=0. If such a case arises, SHORTF has probably been inputted incorrectly and the program prints "ATOM I HAS EV= - and EVC = 0" and then exits. Otherwise the program begins to search for all non-single bonds in the molecule in the following manner: The program steps along each atom

until it finds one with $EV > 0$ and $EVC = 1$. Such an atom corresponds to an end atom in a terminal double bond, such as C_1 or C_4 in butadiene:



NB1 and NB2 are integer variables initially set to zero. Each time a terminal multiple bond is found, BT of that bond is incremented by 1, NB2 is incremented by 1 and EV of both atoms is decreased by 1. After EV for both atoms has been reduced, if either atom has $EV = 0$, EVC for each atom attached to that atom is reduced by 1. As the end bonds are identified and their respective EV's reduced, new terminal bonds may appear. When all atoms have been scanned once, NB1 is set equal to NB2 and the cycle is repeated until, at the end of a cycle but before NB1 is replaced by NB2, $NB1 = NB2$: in other words, until a cycle produces no new multiple bonds. Thus two cycles are required for a triple bond. At the end of this procedure, all multiple bonds will have been identified except for those in completely conjugated cyclic systems since these latter have no terminal multiple bonds. TOTEV is the total excess valence of the molecule. If, at the end of the above procedure, $TOTEV = 0$, all bonds have been defined and the program exists to the end of the block. Otherwise, it begins to identify bonds in cyclic systems.

If a ring K is found for which each atom in the ring has $EV > 0$, it is identified as an aromatic ring if the number of atoms is $4N+2$ ($RGTYPE(K) = 'AROM'$); otherwise the ring is conjugated ($RGTYPE(K) = 'CONJ'$). Each bond in an aromatic ring has BT set at 3 and EV for each atom in the ring is decreased by 1. After all aromatic bonds in the system have been identified, MBLD2 proceeds to identify bonds in conjugated rings. For the latter, each bond in turn is checked for the excess valence of each of its atoms. If both atoms have $EV > 0$, BT for the bond is incremented by 1 and EV for each of its

atoms is decreased by 1. Since the identification of bonds in conjugated or aromatic rings may uncover new terminal bonds, such as in benzyne, the program calculates TOTEV again. If it is zero, the bond specification is ended; otherwise the cycle is repeated.

If both atoms in a bond are attached to more than one atom, then there exists for that bond some rotational geometry (cis, trans, gauche,...). The standard model assumes the bond rotational geometry of all such bonds to be trans. The two atoms defined by the rotational geometry of a bond are determined by the input as described earlier. However, the user may alter either the atoms defined by a bond, the actual bond rotational geometry or both by invoking the second input option. It will be remembered that an array NXCARD has been read by the program, either immediately following SHORTF or after the last atom geometry was altered. If this array contains the word BONDROT in columns 1-7, then the next input card contains the following information:

1. The word 'CIS', 'TRAN', 'GAUP', 'GAUM', or 'SKEW' in columns 1-4, right-justified. These words identify the type of geometry required, 'GAUP' and 'GAUM' referring to dihedral angles of $+60^\circ$ and -60° , respectively. 'SKEW' indicates that a numerical value for the dihedral angle is to be read.
2. Four integers, I, J, K, L, in columns 11-12, 15-16, 19-20, 23-24 (right-justified), respectively. J and K are the two atoms in the bond whose rotational geometry is to be changed. I and L (bonded to J and K, respectively) are those atoms chosen to be defined by bond J-K.
3. If the rotational geometry is specified to be 'SKEW', the value of the dihedral angle is specified in columns 30-41 in F 12.7 format.

For each such bond where rotational geometry is altered, the geometry is contained in BDRROT(X) while for a SKEW geometry the dihedral angle is placed in SKRTNG(X) where X is the bond between atoms J and K.

DIH is a 2XNBD integer array which contains the two atoms, I and L, defined by bond X. For example if X is the bond J-K then $DIH(1,X)=I$ and $DIH(2,X)=L$. If X is the bond K-J, then $DIH(1,X)=L$ and $DIH(2,X)=I$ since I is bonded to J and K to L.

As in the ATCMGECM option, once the word BONDROT has been given as input, as many bond rotational geometries may be altered as is desired. At the end of the option NXCARD again will contain a blank card or the key word for the next option to be invoked.

Consider hydrogen peroxide as an example of the BONDROT option. The SHORTF input for the trans configuration has already been given; however, the dihedral angle of H_2O_2 is in fact 111.5° . The corresponding input would then be

```
COLUMN:  2  4   8  12  16  20  24   30
HYDROGEN PEROXIDE
0 1
2
  0  2  H
  0  1  H
BONDROT
SKEW      3   1   2   4   111.5
Blank Card.
```

If we wish to specify the experimental bond angles as well, the necessary input is

```
HYDROGEN PEROXIDE
0 1
2
  0  2  H
  0  1  H
ATCMGECM
ASYM      1   94.5
ASYM      2   94.5
BONDROT
SKEW      3   1   2   4   111.5
Blank Card.
```


E. MBLD3

This part of the program is concerned with assigning numerical values to all bondlengths and angles. These values are then used to calculate the cartesian coordinates of all atoms in the molecule.

The standard model determines the bondlengths for all bonds in the molecule based on the type of atom, bond type, and connectivity (Model A) or from just the types of atoms (Model B), depending on the value of OP2. The standard values used for bondlengths in Model A are contained in Table 1. XJ represents an atom X with connectivity J, for example, C4 means a carbon with connectivity 4. These bondlengths are those in typical molecules containing the various types of bonds and were obtained from Interatomic Distances¹. Since the single bond between C3 and N3 in amides are quite different from other types of C3-N3 single bonds, the former are taken as a special case. In addition, the NO bond in NO₂ groups, NN in N₂ groups and triple aromatic bonds such as CC in benzyne are taken to be 1.24, 1.12, and 1.30Å, respectively. Table 2 contains the bondlengths used in Model B. These are simply averages of all Model A bondlengths between a given pair of atoms.

In addition to the standard bondlengths, the user has the option of choosing alternate values for any bonds in the molecule. If the array NXCARD has the value BONDLENTH (columns 1-8) then the next input card has the following information:

1. The numbers of the atoms in the bond in columns 1-2 and 5-6, right-justified.
2. The value of the bondlength in columns 10-18 in F9.6 format.

1. Interatomic Distances, L. E. Sutton, ed., The Chemical Society (London) Burlington House N. I., 1958.

3. The word 'TRUE' in columns 20-23 if another new bondlength is to be read on the next card.

When all bondlengths have been changed, the array NXCARD is again read in a 20A4 format. As before, this card should either be blank or contain the key word for the next option used.

As an example using this option, consider hydrogen peroxide once again. Suppose we would like to assign the value of 0.9\AA to both O-H bondlengths. The necessary input would then be

```

COLUMN: 12 4   8 12      20
          HYDROGEN PEROXIDE
          0 1
          2           1
            0   2   H
            0   1   H
          BONDLENGTH
          1   3   0.900   TRUE
          2   4   0.9
Blank Card

```

Once all such bondlengths have been read in, MBLD3 assigns standard bondlengths to the remaining bonds. After these have been assigned, all atoms other than those with local atom geometry of ASYM are assigned standard angles according to the following rules: 'TETR', 'PYRA', and 'BENT' geometries are assigned the tetrahedral angle. 'TRIG' is 120° and 'LINE' is 180° . For 'ASYM' the angle has been read in on cards. Finally, ROTANG(I) is filled with the proper angle for the bond rotational geometry of bond I: 180° for 'TRAN', 0° for 'CIS', $+60^\circ$ for 'GAUP' and -60° for 'GAUM'.

Aside from the first three atoms to be defined in a molecule (the coordinates of these three define an initial plane), the coordinates of any atom are, in general, found with respect to three other atoms whose coordinates are already known. In other words, three parameters are necessary to define the position of an atom in three-dimensional space. These three parameters are taken to be a bondlength and two angles, all with respect to atoms whose coordinates are known. The two angles may be two bond angles

or a bondangle and a dihedral angle or a bondangle and an angle with the plane formed by the three reference atoms. Consider the following two examples in which the coordinates of atom I are to be found and those of atoms J,K,L, are known:



In a. the bondlength I-J and the bondangles IJK and IJL may be used to find the coordinates of I. Alternatively, one may use the angle which bond I-J makes with the plane of KJL in place of one of the two valence angles. In b. one would use the dihedral angle IJKL rather than bondangle IJL. In general, one may group these atoms J,K,L into a two dimensional matrix in the following manner:

$\bar{Z}(I,1)$ is some atom to which atom I is bonded and whose coordinates are known.

$Z(I,2)$ is some atom bonded to $Z(I,1)$ forming a bondangle with I and whose coordinates are known.

$Z(I,3)$ is some atom which

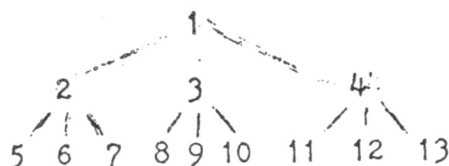
- a. is bonded to $Z(I,1)$ forming a bondangle with I and whose coordinates are known, or
- b. is bonded to $Z(I,2)$ forming a dihedral angle with $Z(I,1)$ and I and whose coordinates are known.

$Z(I,4)$ is a parameter whose value determines whether the coordinates are to be calculated using a dihedral angle and a bondangle, two bondangles or a bondangle and an angle with a plane. $Z(I,4)=0$ for the first case, ± 1 for the second and ± 2 for the third. In the latter two cases the sign of $Z(I,4)$ determines whether I is above (+) or below (-) the plane of $Z(I,1)$, $Z(I,2)$, and $Z(I,3)$.

The standard model sets up this Z matrix using a tree structure similar to that used in the determination of rings. Unless it is impossible,

each atom is defined by two bondangles rather than a bondangle and a dihedral angle. In other words, each time a new branch is added to the tree structure only one of the atoms need be defined using a dihedral angle. All other atoms in that branch can be determined using two bondangles. The third possibility, that of a bondangle and an angle with a plane, is not needed in this standard model; however, it may be implemented by using an option to be discussed below.

Consider the following hypothetical tree structure as an example.



In general, atoms 1, 2, and 3 will define a plane. Atom 4 may then be defined by the bondlength 1-4 and the bondangle 2-1-4 and 3-1-4. Of the remaining atoms in the tree, only 3 must be defined using a dihedral angle: one in each of the three branches. Each of the remaining two atoms in each branch may then be defined by two bond angles. For example, if atom 5 is defined by the bondlength 2-5, the bondangle 1-2-5 and the dihedral angle 4-1-2-5, then atom 6 may be defined by the angles 1-2-6 and 5-2-6 and the bondlength 2-6.

In order to decide which atom of each branch is to be defined by the corresponding dihedral angle, MBLD3 implements a new temporary numbering system for forming the Z matrix. With one restriction, all atoms are numbered in the order ATBD(2,I); that is, ATBD(2,1) is atom 1, ATBD(2,2) is atom 2 and so on. FROM(I) is then easily defined to be ATBD(1,I). The one restriction to this renumbering rule is as follows: One would like the first atom in each new branch of the tree to be that atom which has in fact been chosen by the user (via SHORTF or F) to be defined by a dihedral angle. Therefore, each time the program comes to a new branch in the tree all

numbers of that branch are scanned to see which atom in the branch is a member of the array DIH for the bond to which it is attached. The program knows when a new branch has been found by checking to see if $ATBD(1,I) = ATBD(1,I-1)$. If this equality is true, atoms I and I-1 are in the same branch; otherwise I has started a new branch. Once the atom belonging to DIH has been found it is given the next available number and the rest of the atoms in the branch are numbered in sequence. In our example above, if 6, 9, and 13 are members of DIH for their respective bonds, the new numbering system is



Once the entire molecule has been renumbered in this manner, it is a relatively easy matter to set up the Z matrix. If $Z(I,1)$ is bonded to only one atom which has already been defined, I must be defined by a dihedral angle; that is, $Z(I,4)=0$. Otherwise $Z(I,4)=\pm 1$. If atom $Z(I,1)$ is trigonal (in the latter case), $Z(I,4)$ is arbitrarily set equal to +1; otherwise the sign of $Z(I,4)$ is chosen so as to be consistent with a right-handed coordinate system.

$BL(I)$ is the bondlength of the bond between I and $Z(I,1)$. $ALPHA(I)$ is the angle between I, $Z(I,1)$, and $Z(I,2)$. If $Z(I,4)=0$, $BETA(I)$ is the dihedral angle for atoms I, $Z(I,1)$, $Z(I,2)$, and $Z(I,3)$; otherwise $BETA(I)$ is the bondangle I, $Z(I,1)$, $Z(I,3)$.

Although the Z matrix is filled automatically by MBLD3, one may for various reasons wish to construct his own Z-matrix. Since, using the standard model, all angles about a given atom are the same, if one wishes to assign a non-symmetric geometry, the values of ALPHA, BETA must be read in. The easiest way to do this is to read in the Z-matrix as well. A

further use of an option to specify the Z matrix as input arises in the case of a strained ring, where a strained ring is defined as one which is not completely symmetric. The standard model in general cannot handle such a ring (such as pyridine) since the ring will not close properly. In such a case being able to feed in the entire geometry would be most helpful.

It will be recalled that the first input card (in MAIN) contains a value for a parameter called INOP. Normally INOP is given the value zero in which case the standard model and/or one of the aforementioned options is used. However, if INOP=2, MBLD2 is skipped entirely and the remaining input is as follows:

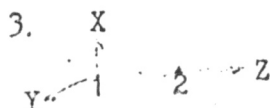
1. Name of molecule, as before
2. CHARGE, MULTIP(2I2)
3. M(I2) : The total number of atoms in the molecule.
4. For each atom I:

AN(I),	Z(I,1),	BJ(I),	Z(I,2),	ALPHA(I),	Z(I,3),	BETA(I),	Z(I,4)
I3	I4	F7.4	I4	F11.6	I4	F11.6	I4

Note that no blank card is placed in the input stream in this case.

In all cases, the angles are initially in degrees, bondlengths in angstroms. The angles are converted to radians internally.

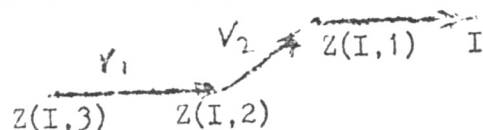
C(I,1), C(I,2), C(I,3) are the X,Y,Z coordinates of atom I, respectively. Atom 1 is taken as the origin, atom 2 along the z-axis and atom 3 in the XZ plane so that the coordinate system is as follows:



If atoms 1,2,3 are colinear, no plane has been defined. In this case each succeeding atom is defined explicitly in the same manner as was atom 3 until an atom with a nonzero X coordinate is defined. The coordinates of all remaining atoms are then found by one of the three methods to be outlined

below, depending on the value of $Z(I,4)$. It should be noted here that the temporary numbering system is still in effect outside of the loop in which Z is defined; however, when the coordinates are printed out or written on disc, all values refer to the original numbering system. Of course, if the Z -matrix is read in directly, no new numbering system has been implemented.

I. Coordinates of Atom I for $Z(I,4) = 0$:



1. Define \vec{V}_1 as the vector from $Z(I,3)$ to $Z(I,2)$ and \hat{L}_1 as the corresponding unit vector.
2. Define \vec{V}_2 as the vector from $Z(I,2)$ to $Z(I,1)$ and \hat{L}_2 as the corresponding unit vector.
3. Define \vec{V}_P as the cross product $\hat{L}_1 \times \hat{L}_2$ and \hat{L}_3 as the unit vector of \vec{V}_P .
4. Define \hat{L}_4 as the cross product $\hat{L}_3 \times \hat{L}_2$. Then \hat{L}_2 , \hat{L}_3 and \hat{L}_4 may be taken as unit vectors of a set of local coordinate axes a, b, c , respectively, centered on atom $Z(I,1)$. One may define the coordinates of atom I in terms of a, b, c , as:

$$C(I,a) = -BL(I) \cos(\text{ALPHA}(I)) \quad (1)$$

$$C(I,b) = BL(I) \sin(\text{ALPHA}(I)) \sin(\text{BETA}(I)) \quad (2)$$

$$C(I,c) = BL(I) \sin(\text{ALPHA}(I)) \cos(\text{BETA}(I)) \quad (3)$$

Transforming to a set of axes centered on $Z(I,1)$ but parallel to the x, y, z axes, one has for the coordinates of I:

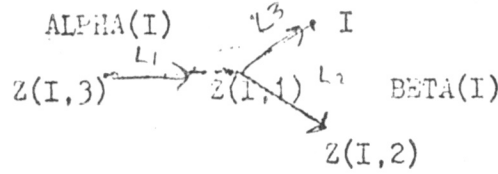
$$VJ(J) = L2(J) * C(I,a) + L3(J) * C(I,b) + L4(J) * C(I,c) \quad (4)$$

where $J = 1, 2, 3$ represents x, y, z .

Finally, the cartesian coordinates of I with respect to the origin are given by

$$C(I,J) = VJ(J) + C(Z(I,1), J) \quad (5)$$

II. Coordinates of Atom I for $Z(I,4) = \pm 1$



1. Define $\vec{V1}$ as the vector from $Z(I,3)$ to $Z(I,1)$ and $\hat{L1}$ as the corresponding unit vector.
2. Define $\vec{V2}$ as the vector from $Z(I,1)$ to $Z(I,2)$ and $\hat{L2}$ as the corresponding unit vector.
3. Define $\vec{V3}$ as the vector product $\hat{L1} \times \hat{L2}$.
4. Define \vec{VJ} as the vectore from $Z(I,1)$ to I and $\hat{L3}$ as the corresponding unit vector.

Then one may write

$$\hat{L3} = A * \hat{L1} + B * \hat{L2} + D * \vec{V3} \quad (6)$$

where A, B, D are arbitrary parameters to be determined. Taking the dot product of $\hat{L1}$ and equation (6) we obtain

$$\hat{L1} \cdot \hat{L3} = A + B * (\hat{L1} \cdot \hat{L2}) \quad (7)$$

Similarly, taking the dot product of $\hat{L2}$ and equation (6) gives

$$\hat{L2} \cdot \hat{L3} = A * (\hat{L1} \cdot \hat{L2}) + B \quad (8)$$

Combining equations (7) and (8)

$$\hat{L1} \cdot \hat{L3} = A + \{ \hat{L1} \cdot \hat{L2} \} \{ \hat{L2} \cdot \hat{L3} \} - A(\hat{L1} \cdot \hat{L2}) \quad (9)$$

Reanranging (9) yields

$$A = (\hat{L1} \cdot \hat{L3} - (\hat{L1} \cdot \hat{L2})(\hat{L2} \cdot \hat{L3})) / (1 - (\hat{L1} \cdot \hat{L2})^2) = \frac{-\cos(\text{BETA}(I)) + \text{ZETA} * \cos(\text{ALPHA}(I))}{1 - \text{ZETA}^2} \quad (10)$$

where $\text{ZETA} = -\hat{L1} \cdot \hat{L2}$,

Similarly, one may obtain the following expression for B:

$$B = (\hat{L2} \cdot \hat{L3} - (\hat{L1} \cdot \hat{L2})(\hat{L1} \cdot \hat{L3})) / (1 - (\hat{L1} \cdot \hat{L2})^2) = \frac{\cos(\text{ALPHA}(I)) - \text{ZETA} * \cos(\text{BETA}(I))}{1 - \text{ZETA}^2} \quad (11)$$

If we take the dot product of $\vec{V3}$ with (6) we obtain

$$\vec{V3} \cdot \hat{L3} = D(\vec{V3})^2 \quad (12)$$

Since

$$(\hat{L}_1 \times \hat{L}_2)^2 = 1 - (\hat{L}_1 \cdot \hat{L}_2)^2,$$

we have (multiplying both sides of the equation by D)

$$D\vec{V}_3 \cdot \hat{L}_3 = D^2(1 - (\hat{L}_1 \cdot \hat{L}_2)^2) \quad (13)$$

Taking the dot product of \hat{L}_3 and equation (6),

$$1 = A(\hat{L}_1 \cdot \hat{L}_3) + B(\hat{L}_2 \cdot \hat{L}_3) + D\vec{V}_3 \cdot \hat{L}_3 \quad (14)$$

Combining equations (13) and (14) we obtain

$$D = \pm \text{SQRT} \left\{ (1 - A(\hat{L}_1 \cdot \hat{L}_3) - B(\hat{L}_2 \cdot \hat{L}_3)) / (1 - (\hat{L}_1 \cdot \hat{L}_2)^2) \right\} \quad (15)$$

or

$$D = Z(I,4) * \text{SQRT} \left\{ (1 + A \cos(\text{ALPHA}(I)) - B \cos(\text{BETA}(I))) / (1 - \text{ZETA}^2) \right\} \quad (16)$$

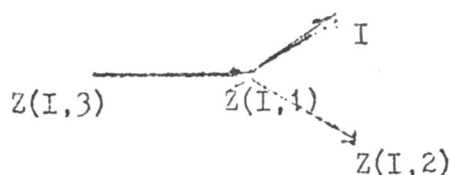
Finally then, the coordinates of atom I with respect to atom Z(I,1) are

$$VJ(J) = BL(I) \{ A(I) * L1(J) + B(I) * L2(J) + D(I) * V3(J) \} \quad (17)$$

The coordinates of I with respect to the origin are

$$C(I,J) = VJ(J) + C(Z(I,1),J) \quad (18)$$

III. Coordinates of Atom I for $Z(I,4) = \pm 2$



If we define the vectors \vec{V}_1 , \hat{L}_1 , \vec{V}_2 , \hat{L}_2 , \vec{V}_3 , \vec{V}_j , and \hat{L}_3 as in Part II, and again we start with equation (6) and following a similar procedure, it can be shown that A, B, and D are given by the following expressions where BETA(I) is now 90° - that angle which the bond Z(I,1)-I makes with the plane defined by Z(I,1), Z(I,2) and Z(I,3):

$$A(I) = \frac{1}{2} Z(I,4) * \left\{ (1 - \cos^2(\text{ALPHA}(I)) - D(I)(|V_3|) \cos(\text{BETA}(I))) / (1 - \text{ZETA}^2) \right\}^{\frac{1}{2}} \quad (19)$$

$$B(I) = A(I) * \text{ZETA} + \cos(\text{ALPHA}(I)) \quad (20)$$

$$D(I) = \cos(\text{BETA}(I)) * (|V_3|) / (1 - \text{ZETA}^2) \quad (21)$$

Then the coordinates of atom I with respect to Z(I,1) are given by equation

(17) and (18) respectively.

There is one final input option to be discussed. If one wishes to study a free radical and its parent molecule without having to type up two sets of input cards, one may simply eliminate one atom from the molecule after the coordinates have been calculated. This is done by setting NXCARD(1) equal to the word 'ELIM' in columns 1-4 and placing the number of the atom to be deleted in columns 5-8 (right-justified). For example, if we wanted to calculate methyl radical, the input would be

```
METHYL RADICAL
0 1
1
  C  H  H  H  H
ELIM  3
Blank Card
```

This would instruct the program to calculate the coordinates of methane, delete hydrogen 3, and decrease the numbers, atomic number elements, and coordinate elements of atoms 4 and 5 by 1. Unlike the previous options, only one atom may be eliminated.

If more than one of the four options discussed (ATOMGEOM, BONDROT, BONDLNTH, ELIM) are used for the same molecule, they must be used in the order discussed. For example, if both ATOMGEOM and ELIM are to be used, ATOMGEOM is invoked first, then ELIM. Unless the entire geometry is fed in by means of the Z matrix, the final card must always be blank for each molecule.

Samples of typical output from MBLD are contained in Appendix B. In addition, if OUTOP has been given the value of 2, the following information is punched out on cards:

```
M(I2 format)
CHARGE,MULTIP(2I2 format)
AM(I),I=1,M(25I3 format)
C(I,J),J=1,3 (5x,F10.5x,F10.5,5x,F10.5)=H cards.
```

F. Limitations to the Standard Model

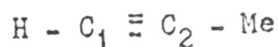
At the present time, the standard model is limited to molecules containing H, Li, C, N, O, F. In the case of Li, only standard Li-H and Li-C are available. Eventually this will be extended to the entire first and second row.

A more severe limitation arises in the case of strained rings. A

strained ring is one which is not completely symmetric with an all trans geometry. This excludes all rings systems from being defined properly by the standard model except benzene and chair cyclohexane (see Appendix A for input for the latter). All other rings will be built correctly by MBLD2; however, a nonsymmetric ring, such as pyridine will not be closed properly when the actual coordinates are calculated. In pyridine, all bondangles will be taken as 120° by the standard model, but the C-N bondlengths are different from C-C. On the other hand if one changes the atom geometry of the nitrogen, the ring will close properly. However, for more complicated ring systems, especially nonaromatic ones, even this is not possible. Of course any molecule may be run using MBLD if the entire geometry is fed in via the Z-matrix, including strained rings as well as molecules containing any atoms in the periodic table. To correct for the problem of strained rings in a general way seems infeasible since to do so would require a separate section of the program for each type of strained ring. It is possible, however, to have standard models for the most common strained rings in an analogous manner to ET, TBU,.... This is being investigated at the present time.

One final limitation concerns molecules containing 3 or more colinear atoms which are not at the end of the molecule. As mentioned earlier, if the first three atoms are colinear, the program will keep defining atoms until it finds one with a nonzero X coordinate; however, if a plane has been defined and then 3 or more colinear atoms are encountered, the molecule

cannot be handled properly. In the case of colinearity at one end of the molecule, as long as all of the colinear atoms are numbered first, the molecule will be handled properly. For example, the input for methyl acetylene would be



METHYL ACETYLENE

0 1

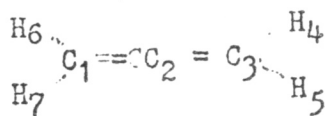
2

C 2 H

C 1 ME

Blank card.

Unfortunately, allene and similar molecules cannot be handled due to this deficiency; however, once again this problem may be circumvented by feeding in the Z matrix. Consider allene:



Atoms 1-5 are defined properly by the standard model, but atoms 6 and 7 are not since 1,2,3 do not define a plane. Ordinarily these atoms would be taken as Z(6,1), Z(6,2) and Z(6,3), respectively. However one may set

$$Z(6,1) = 1$$

$$Z(6,2) = 3$$

$$Z(6,3) = 4 \text{ or } 5$$

to get around this problem and similarly for atom 7. Hopefully, this restriction will be alleviated in the near future.

G. Timing and Accuracy

The IBM 360 single precision word is significant only to $6\frac{1}{2}$ figures. As a result, when running in single precision, the coordinates may be trusted, in general, only to 5 decimal places. However, the time required to calculate the coordinates is negligible (less than 1 second for benzene) so that the coordinates may easily be obtained in double precision if greater accuracy is required.

Table 1. Model-Builder Parameters and their Functions

<u>Parameter</u>	<u>Function</u>
SHORTF	7XN array containing all bonding information where N is the number of atoms which have been numbered explicitly. SHORTF(1,I) is the chemical symbol of atom I. SHORTF(2,I)-SHORTF(7,I) contain all atoms or groups of atoms bonded to I.
EL	A 1X10 ⁴ array containing the chemical symbols of all atoms in the periodic table. EL(1)=H, EL(2)=HE, and so on.
VL	A 1XN array containing the most common valence of each element in the first two rows of the periodic table; thus VL(6)=4 for carbon.
CHARGE	The total charge on the molecule.
MULTIP	The total spin multiplicity of the molecule.
N	The number of rows in SHORTF
F	A 6XM array which is an expanded version of SHORTF. All atoms have been explicitly numbered in F, and M is the number of atoms in the molecule.
AN	A 1X(M+1) array containing the atomic numbers of all atoms in the molecule. The atomic numbers of atom I is contained in AN(I+1) until the end of the program at which time AN(I) contains the atomic number of I (see text).
NXCARD	An array in which the key words for the four model-builder options are read.
CN	A 1X(M+1) array containing the coordination number (or connectivity) of each atom.

TABLE 1 (Cont.)

<u>Parameter</u>	<u>Function</u>
MXCN	The largest connectivity occurring in the molecule. MXCN=4 in ethane, 3 in benzene or ammonia, 2 in hydrogen peroxide.
NBD	The total number of bonds in the molecule.
ATBD	A 2XNBD array. ATBD(J,I) is the J th atom participating in bond I.
G	A 6XM array identical to F.
BDAT(J,I)	The J th bond in which atom I participates.
BDBD(J,I)	The J th bond to which bond I is attached.
ATRG(J,I)	The J th atom in ring I.
TREE	A 10X10 array used to construct ring systems (see text).
FROMA(I)	Defines that atom from which atom I originates in the tree structure.
NRG	The total number of rings in the molecule.
RGSIZE(I)	The number of atoms in ring I.
COMMAT(I)	The I th atom which two rings have in common.
MXRGSZ	The number of atoms in the largest ring in the molecule.
BDRG(J,I)	The J th bond in ring I.
RGAT(J,I)	The J th ring in which atom I participates.
RGBD(J,I)	The J th ring in which bond I participates.
NRGBD(I)	The number of rings in which bond I participates.
RGBRG(J,I)	The J th ring attached to ring I by a bond.
NRGAT(I)	The number of rings in which atom I participates.
RGARG(J,I)	The J th ring attached to ring I by just one atom.
XSVAL(I)	The excess valence of atom I; that is, the natural valence, VL(I), less the connectivity of I in the present molecule, CN(I).

TABLE 1 (Cont.)

<u>Parameter</u>	<u>Function</u>
ATGEOM(I)	The local atom geometry about atom I; i.e., trigonal, bent,...
NEWANG(I)	Contains the numerical value of the angles about atom I if the option to alter the local atom geometry of I has been invoked.
EV(I)	Identical to XSVAL(I).
EVC(I)	The number of atoms bonded to I which have EV>0. Used to define standard atom geometries and bond types.
BT(I)	The bond type (single, double,...) of bond I.
TOTEV	The total excess valence in the molecule.
RGTYPE(I)	The ring type of ring I (e.g. aromatic).
BDROT(I)	The bond rotational geometry of bond I (e.g., trans).
DIH	A 2XNBD array. Given a bond I with bond rotational geometry, DIH(1,I) and DIH(2,I) are those atoms which are defined by that geometry. For each bond I, DIH(1,I) is attached to one end of the bond and DIH(2,I) to the other.
BDTYPE(I)	The bond type of bond I.
SKRTNG(I)	Contains the dihedral angle associated with bond I if the option to alter this quantity has been invoked.
C(J,I)	The I th coordinate of atom J where I=1,2,3 corresponds to X,Y,Z respectively.
BNDLTH(I)	Contains the bondlength of atom I.
BDANG(I)	Contains the angles about atom I if the standard model has been used.
ROTANG(I)	Contains the dihedral angle about bond I if the standard model has been used.

TABLE 1 (Cont.)

<u>Parameter</u>	<u>Function</u>
NEWNO(I)	The number of atom I in the reordered numbering system (see text).
BDDEF(I)	The bond which dihedrally defines atom I.
OLDNO(I)	The number of atom I in the original numbering system.
Z	An MX4 array containing those atoms used to define the coordinates of each atom in the molecule (see text).
BL(I), ALPHA(I), BETA(I)	The three geometric parameters necessary to define the coordinates of atom I (see text).
RELVEC(R,J,K)	A subroutine which finds the vector, R, from atom K to atom J.
UNIVVEC(L,R)	A subroutine which finds the unit vector L of some vector R.
VECPRD(VP,X,Y)	A subroutine which finds the vector product, VP, of X crossed Y.

Table 2 Standard Model A Bond Lengths

Bond	Length, Å	Bond	Length, Å
Single Bonds			
H-H	0.74	C3-N2	1.40
Li-H	1.595	C3-O2	1.36
C4-H	1.09	C3-F1	1.33
C3-H	1.08	C2-C2	1.38
C2-H	1.06	C2-N3	1.33
N3-H	1.01	C2-N2	1.33
N2-H	0.99	C2-O2	1.36
O2-H	0.96	C2-F1	1.30
F1-H	0.92	N3-N3	1.45
C4-Li	2.10	N3-N2	1.45
C4-C4	1.54	N3-O2	1.36
C4-C3	1.52	N3-F1	1.36
C4-C2	1.46	N2-N2	1.45
C4-N3	1.47	N2-O2	1.41
C4-N2	1.47	N2-F1	1.36
C4-O2	1.43	O2-O2	1.48
C4-F1	1.36	O2-F1	1.42
C3-C3	1.46	F1-F1	1.42
C3-C2	1.45		
C3-N3	1.40 ^a		
Double Bonds			
C3-C3	1.34	C2-O1	1.16
C3-C2	1.31	N3-O1	1.24 ^b
C3-N2	1.32	N2-N2	1.25
C3-O1	1.22	N2-O1	1.22
C2-C2	1.28	O1-O1	1.21
C2-N2	1.32		
Triple Bonds		Aromatic Bonds	
C2-C2	1.20	C3-C3	1.40
C2-N1	1.16	C3-N2	1.34
N1-N1	1.10	N2-N2	1.35

^a 1.32 used in N-C=O group.

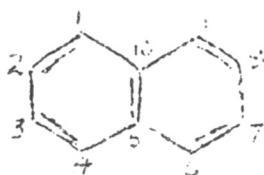
^b Partial double bonds in NO₂ and NO₃ groups.

Table 3 Standard Model B Bond Lengths

Bond	Length (\AA)
H-H	0.74
Li-H	1.60
C-H	1.08
N-H	1.00
O-H	0.96
F-H	0.92
C-Li	2.10
C-C	1.40
N-C	1.37
O-C	1.36
F-C	1.35
N-N	1.35
O-N	1.30
F-N	1.36
O-O	1.48
F-O	1.42
F-F	1.42

Appendix A: Sample Input

1. Naphthalene



COLUMN: 12 4 8 12 16

NAPHTHALENE

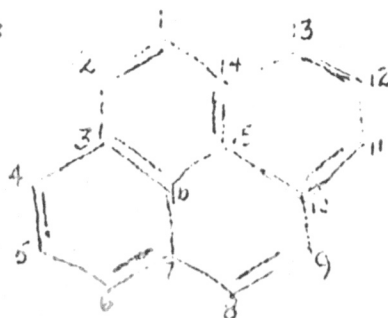
0 1

10

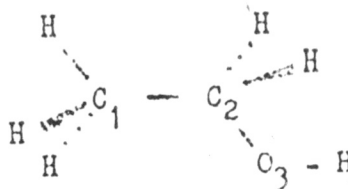
C	H	10	2
C	3	H	1
C	2	4	H
C	H	3	5
C	10	6	4
C	H	5	7
C	8	H	6
C	7	9	H
C	H	8	10
C	5	1	9

Blank Card

Note that SHORTF has been set up so that the H on carbon 1 is trans to carbon 3, the H on carbon 2 is trans to carbon 4, and so on by conforming to the rules given in Section B. Thus the atom directly to the right of C2 in row 1 is the H on C1 and similarly the atom directly to the right of C1 in row 2 is C3. Therefore H is trans to C3 according to the standard model. It should be mentioned that SHORTF for multi-ring systems is most easily set up if the molecule is numbered about the periphery of the rings, as in naphthalene above, although this is not necessary. A more complex example is pyrene:



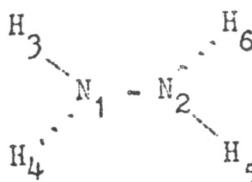
2. Ethyl Alcohol



COLUMN: 12 4 8 12 16 20 24
 ETHYL ALCCOL
 0 1
 3
 C 2 H H H
 C 1 3 H H
 O 2 H2
 BONDROT
 TRAN 1 2 3 9
 Blank Card

Note that the BONDROT option has been used to obtain a configuration in which the hydroxyl hydrogen (atom9) is in the plane of the heavy atoms. If the standard model had been used, this hydrogen would have been trans to one of the hydrogens on carbon 2 and thus it would have been out of the plane. This example serves to illustrate the necessity of being quite careful when setting up SHORTF for a molecule which has different rotational configurations.

3. Hydrazine



HYDRAZINE
 0 1
 2
 N 2 H H
 N 1 H H
 Blank Card

Note that the standard input places H_3 and H_5 trans, but H_4 and H_6 gauche. The following input will give an all staggered configuration:

HYDRAZINE
 0 1
 2
 N 2 H H
 N 1 H H
 BONDROT
 TRAN 3 1 2 6
 Blank Card

4. BF_3

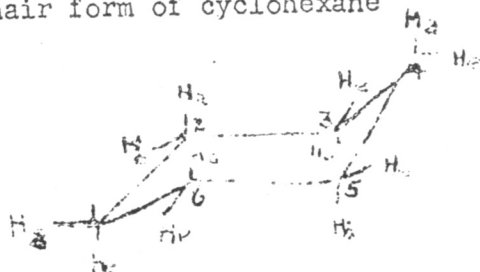
The only atoms for which standard bondlengths and angles are contained in model-builder are Li, H, C, N, O, F; however, any molecule containing any atom may be run by reading in the Z-matrix (see Section E). In addition, the standard model may be used by utilizing the ATOMGEOM and BONDLENTH options for all local atom geometries and bondlength in the molecule. For BF_3 , for example, the input would be

```

BORON TRIFLUORIDE
0 1
1
  B  F  F  F
ATOMGEOM
TRIG      1
BONDLENTH
1  2  1.3      TRUE
1  3  1.3      TRUE
Blank Card.

```

5. Chair form of cyclohexane



```

CHAIR CYCLOHEXANE
0 1
6
C  H_a  6  2  H_e
C  H_e  H_a  1  3
C  4  H_e  H_a  2
C  3  5  H_e  H_a
C  H_a  4  6  H_e
C  H_e  H_a  5  1
Blank Card.

```

The e and a subscripts on the hydrogens refer to equatorial and axial respectively. They are shown to indicate that an equatorial hydrogen is always trans to a carbon and would not appear in the actual input.

Appendix B: Sample Output

The output generated for benzene is shown below. The atom information, bond information, and ring information are all printed at the end of MBLD2. The coordinates are printed at the end of MBLD3.

BENZENE

CHARGE= 0 MULTIP= 1

C	H	6	2
C	3	H	1
C	2	4	H
C	H	3	5
C	6	H	4
C	5	1	H

ATOM INFORMATION

ATOM	SYMBOL	GEOM	ARRAY F		
1	C	TRIG	7	6	2
2	C	TRIG	3	8	1
3	C	TRIG	2	4	9
4	C	TRIG	10	3	5
5	C	TRIG	6	11	4
6	C	TRIG	5	1	12
7	H		1	0	0
8	H		2	0	0
9	H		3	0	0
10	H		4	0	0
11	H		5	0	0
12	H		6	0	0

BOND INFORMATION

BOND	ATOM,1	ATOM,2	BOND DEFINITION		BOND TYPE	BOND ROT GEOM
2	C	H	1	7	SING	
3	C	C	1	6	AROM	TRAN
4	C	C	1	2	AROM	TRAN
5	C	C	6	5	AROM	TRAN
6	C	H	6	12	SING	
7	C	C	2	3	AROM	TRAN
8	C	H	2	8	SING	
9	C	H	5	11	SING	
10	C	C	5	4	AROM	TRAN
11	C	H	3	9	SING	
12	C	H	4	10	SING	
13	C	C	3	4	AROM	TRAN

RING INFORMATION

RING	RING TYPE	ATRQ						
1	AROM	6-RING	4	5	6	1	2	3

COORDINATES

NO OF ATOM	SYMBOL	X	Y	Z
1	C	0.0	0.0	0.0
2	C	-1.21243	0.0	-0.7
3	C	-1.21243	0.0	-2.1
4	C	0.0	0.0	-2.8
5	C	1.21243	0.0	-2.1
6	C	1.21243	0.0	-0.7
7	H	0.0	0.0	1.08
8	H	-2.14774	0.0	-0.16
9	H	-2.14774	0.0	-2.64
10	H	0.0	0.0	-3.88
11	H	2.14774	0.0	-2.64
12	H	2.14774	0.0	-0.16

DICTIONARY PROGRAM

The Dictionary Program accepts as input a chemical name and generates, in SHORTF, the required input card images for Model Builder. Dictionary inputs the following set of cards.

1. Title Card (80A1)
2. Charge and Multiplicity (212)
3. Up to Three Cards Containing the Chemical Name (80A1)
4. One Blank Card

Chemical names may be constructed in the following manner.

I. The name of a simple molecule.

ex. METHANE
BENZENE
NAPHTHALENE

II. A simple molecule with one or more substituents. Locants and the number of each substituent should be written explicitly. If the locants are omitted they are assumed to be 1 through the number of that substituent present. If the number of a substituent present is omitted, it is assumed to be one (mono).

ex. METHYL BENZENE
or 1 MONO METHYL BENZENE
TRI FLUORO BENZENE
or 1 2 3 TRI FLUORO BENZENE
1 2 DI FLUORO 4 HYDROXY BENZENE

III. A simple or complex construction, I. or II., with one or more parenthetical groups. A parenthetical group consists of one or more locants and a left and right parenthesis enclosing a Type II group whose base is a substituent name instead of a simple molecule name.

ex. 1 (1 2 DI FLUORO BUTYL) BENZENE
1 ETHYL 2 5 (1 3 DI HYDROXY 2 FLUORO BUTYL) BENZENE

Spaces are used to delimit the different parts of a name and any number of them may be used. If a name is continued on a second or third card, the first column of each continuation card must contain a non-blank character which is part of the name.

The following symbols may be used in the construction of names.