Graphical Analysis and Visualization of Three-Dimensional Properties of Molecules and Solids

MARK D. JACKSON

Department of Chemistry, Florida Atlantic University, Boca Raton, Florida 33431

Received August 21, 1990

A simple approach is presented for visualizing three-dimensional properties of molecular and crystalline systems by using PC-based molecular drawing software. This allows a scientist with lower-cost PC-graphics capability to carry out analysis of data, such as wave functions and electron density, which would otherwise require a dedicated graphics workstation and expensive specialized software. In this approach, a data set $\{x,y,z,f_1(x,y,z)\}$ is transformed to a new set of data $\{x,y,z\}$ for values of f_1 meeting specified requirements. The function f_1, which varies throughout a chemical system is difficult to visualize. However, a more manageable data set is obtained if the set of Cartesian coordinates corresponding to locations in the system where f_1 has a specified value are used. This set of points can then be visualized and plotted with standard graphics software or PC-based molecular modeling programs. In addition, the use of PC-based molecular graphics permits one to superimpose the data describing molecular properties on drawings of the molecular geometry. The entire object can then be rotated to facilitate a clear analysis of the property of the molecule.

INTRODUCTION

One of the practical limitations to the use of quantum chemical methods, such as molecular orbital or electronic structure theory, is the difficulty encountered in analyzing three-dimensional properties of a system. The electron density, wave functions, and electrostatic potential all vary with the spatial location in the molecule or crystalline solid. Each property can be described by a function, $f_1(x,y,z)$. In order to visualize this 3-D function, $f_1(x,y,z)$, a four-variable data set must be handled. Three-variable data sets, $z = f(x,y)$, are easy to visualize using standard contouring approaches. However, for 3-D properties the entire data set cannot be clearly visualized simultaneously. One approach involves specifying a value of the function and contouring the portion of $f_1(x,y,z)$ with these values. However, this usually requires specialized, expensive software and a dedicated graphics workstation. For example, the CHEM-X molecular modeling software package[1] allows the input of such data describing 3-D molecular properties. The program then contours a user-specified value and displays the result on the terminal screen. The only limitation to this approach is a practical one, an investment of at least $40 000 is required for academic users and $150 000 for commercial customers (total cost for workstation and software).

Therefore, routine use requires a procedure which utilizes a PC or graphics terminal equipped with cheaper molecular modeling/drawing and graphics software. In this study, an approach is presented which allows a user to analyze $f(x,y,z)$ and obtain output suitable for visualization using PC-based software costing about $300. As an illustration of the method, the localized molecular orbitals (LMO's) of a Zintl anion[2] are studied.

METHODOLOGY

The goal of this procedure is to obtain from a three-dimensional $f_1(x,y,z)$ a simpler $f_2(x,y)$ which be can more easily visualized and plotted. The 3-D function may correspond to charge density, wave functions, or electrostatic potential in a chemical system. The first step requires the determination of f_1 for the property of interest at a sufficient number of points to accurately describe the property. Normally, a uniform 3-D grid of 3375 points ($15 \times 15 \times 15$) gives a reasonable treatment. This tabulation of x,y,z Cartesian coordinates and $f_1(x,y,z)$ must be tailored to the particular property and computer program.

The value of the function to be plotted is then chosen by the user. By using a simple Fortran program, such as the one given in the Appendix, each $f_1(x,y,z)$ entry in the data set is compared to the prescribed function value. If the $f_1(x,y,z)$ is within a certain percentage or tolerance of the desired value (20% for example) the x,y,z coordinates are assigned to a new data set $\{x',y',z'\}$ which can be manipulated more easily than the original $f_1(x,y,z)$. The $\{x',y',z'\}$, or $\{x',y',f_2(x',y')\}$, can be visualized by using standard graphics software or any PC-based molecular drawing program. With either method the data can be viewed as a set of points $\{x',y',z'\}$ or as a function $z' = f_2(x',y')$.

Two approaches have been investigated for carrying out this analysis. First, general graphing software is used to plot the set of points. For this study SAS/Graph[3] is used, but any software package capable of surface plots, i.e., plotting a function of two variables, would also work. SAS was chosen because it is available at many computer facilities. In the second approach, the data were analyzed by the ChemCad+ molecular modeling/drawing package.[4] This software sells for about $300, which makes it accessible to almost all budgets, academic or industrial. Other drawing packages could also be used with the approach discussed in this work. However, the program must allow the input of a Cartesian description of atomic locations. ChemCad+ also serves as a graphical interface for several computational chemistry programs used in our laboratory. The ChemCad+ approach is found to be superior to SAS/Graph for chemical problems.

SAS/Graph is a general purpose, widely available graphics package designed to work with a large variety of graphics terminals and plotters. For the purposes of this study the "G3D" subroutine is used. This SAS routine is designed to plot a set of $\{x,y,z\}$ values.

There are several drawbacks to this method. First, SAS/Graph is designed to plot "surfaces", $z = f(x,y)$, and it therefore ignores some of the data if two data points have equivalent z values. For almost any tabulation of a symmetric molecule using a uniform grid, this feature of SAS creates difficulties. However, it can be overcome by plotting unconnected data points, instead of contoured surfaces, and by shifting the values of the offending data by small, random amounts. An alternative is to use a nonuniform grid of points to generate the original function values. This produces accurate plots but the connectivity of the data cannot be clearly visualized. A second drawback limits the use of SAS/Graph for chemical problems. There is no convenient way to su-

0095-2338/91/1631-0127$02.50/0

Chart I

```
      IMPLICIT REAL*8(A-H,O-Z)
      IMPLICIT INTEGER*4(I-N)
      DIMENSION TITLE(20),X(10000),Y(10000),Z(10000),F(10000)
      DIMENSION XMOL(30),YMOL(30),ZMOL(30)
      DATA TITLE/20*'    '/
C     ROUTINE TO CONVERT F(X,Y,Z,) TO F(X',Y')= Z IF A GIVEN
C     CONDITION IS OBEYED -  I.E. IF F=CONTOUR (.03 EG.)
C       USED TO CONVERT 3D GRID OF FUNCTION VALUES TO
C       A FORM COMPATIBLE WITH SAS/GRAPH G3D ROUTINE
C     UNIT 20 - FILE FOR SAS/GRAPH
C     UNIT 10 - FILE FOR CHEMCAD PROGRAM
C   ALSO CALCULATES "average" POSITION AND SIZE OF LMO
      READ(5,2)TITLE
      WRITE(6,2)TITLE
      WRITE(6,60)
C   READ IN COORDINATES OF MOLECULE - WILL ADD TO OUTPUT FILE AS
C   WELL AS INCLUDE IN CHEMCAD FILE.  FOR THIS FILE ALL ATOMS TYPES
C   WILL BE DENOTED WITH "C"
C    NMOL IS THE NUMBER OF ATOMS IN THE MOLECULE
      READ(5,10)NMOL
      DO 24 IMOL=1,NMOL
      READ(5,20)XMOL(IMOL),YMOL(IMOL),ZMOL(IMOL)
      WRITE(6,20)XMOL(IMOL),YMOL(IMOL),ZMOL(IMOL)
      WRITE(10,90)XMOL(IMOL),YMOL(IMOL),ZMOL(IMOL)
   24 CONTINUE
C
C    NPTS IS THE NUMBER OF POINTS IN 3-D GRID OF FUNCTION
C    CONTOUR IS THE VALUE OF THE FUNCTION DESIRED FOR PLOT
C
      READ(5,10)NPTS,CONTOUR
C
C    ASSUME 20% TOLERANCE FOR PLOTTING DATA
C
      IF(CONTOUR.GT.0.0D00)FMIN=0.80D00*CONTOUR
      IF(CONTOUR.GT.0.0D00)FMAX=1.20D00*CONTOUR
C  FOR NEGATIVE CONTOURS
      IF(CONTOUR.LT.0.0D00)FMAX=0.80D00*CONTOUR
      IF(CONTOUR.LT.0.0D00)FMIN=1.20D00*CONTOUR
C  FOR ZERO CONTOUR
      IF(CONTOUR.EQ.0.0D00)FMIN=-0.0001D00
      IF(CONTOUR.EQ.0.0D00)FMAX=0.0001D00
C
      WRITE(6,10)NPTS,CONTOUR,FMIN,FMAX
      NEWPTS=0
C      WRITE(6,50)
C   TEST FUNCTION VALUES
C
      DO 25 I=1,NPTS
      READ(5,30)X(I),Y(I),Z(I),F(I)
      IF(F(I).GT.FMAX) GO TO 25
      IF(F(I).LT.FMIN) GO TO 25
      FNEW=Z(I)
      NEWPTS=NEWPTS+1
      WRITE(6,20)X(I),Y(I),FNEW
      WRITE(10,80)X(I),Y(I),FNEW
      WRITE(20,20)X(I),Y(I),FNEW
   25 CONTINUE
C
C  CONNECTIVITY DATA FOR MOLECULAR CAD/DRAFT OUTPUT
C
      WRITE(10,100)
      DO 26 I=1,NMOL
      K=I+1
      DO 26 J=K,NMOL
      WRITE(10,110)I,J
```

Chart I (*Continued*)

```
   26 CONTINUE
      WRITE(10,100)
C  RESET ARRAYS TO ZERO
      DO 34 II=1,NPTS
      X(II)=0.0D00
      Y(II)=0.0D00
      Z(II)=0.0D00
   34 CONTINUE
      XTOTAL=0.0D00
      YTOTAL=0.0D00
      ZTOTAL=0.0D00
C
      REWIND 20
C   CALCULATE POSITION OF LMO
      DO 35 J=1,NEWPTS
      READ(20,20)X(J),Y(J),Z(J)
      XTOTAL=XTOTAL+X(J)
      YTOTAL=YTOTAL+Y(J)
      ZTOTAL=ZTOTAL+Z(J)
   35 CONTINUE
      DFRACT=DFLOAT(NEWPTS)
      XAVER=XTOTAL/DFRACT
      YAVER=YTOTAL/DFRACT
      ZAVER=ZTOTAL/DFRACT
      WRITE(6,70)XAVER,YAVER,ZAVER
C  CALC. "SIZE" OF LMO
      XDEL2=0.0D00
      YDEL2=0.0D00
      ZDEL2=0.0D00
      DO 37 J=1,NEWPTS
      XDEL2=XDEL2+(X(J)-XAVER)**2
      YDEL2=YDEL2+(Y(J)-YAVER)**2
      ZDEL2=ZDEL2+(Z(J)-ZAVER)**2
   37 CONTINUE
      XDEL2=XDEL2/DFRACT
      YDEL2=YDEL2/DFRACT
      ZDEL2=ZDEL2/DFRACT
      DELR2=(XDEL2+YDEL2+ZDEL2)
      DELR=DSQRT(DELR2)
      DELRX=DSQRT(XDEL2)
      DELRY=DSQRT(YDEL2)
      DELRZ=DSQRT(ZDEL2)
      WRITE(6,75)DELR,DELRX,DELRY,DELRZ
C  CALC. DISTANCE OF LMO FROM EACH ATOM IN THE MOLECULE
      DO 38 K=1,NMOL
      XDELA2=(XMOL(K)-XAVER)**2
      YDELA2=(YMOL(K)-YAVER)**2
      ZDELA2=(ZMOL(K)-ZAVER)**2
      DELAR2=(XDELA2+YDELA2+ZDELA2)
      DELAR=DSQRT(DELAR2)
      DELAX=DSQRT(XDELA2)
      DELAY=DSQRT(YDELA2)
      DELAZ=DSQRT(ZDELA2)
      WRITE(6,76)K,DELAR,DELAX,DELAY,DELAZ
   38 CONTINUE
C
    2 FORMAT(20A4)
   10 FORMAT(I5,3F10.6)
   20 FORMAT(3F10.6)
   30 FORMAT(4F10.6)
   40 FORMAT(4X,'REFINED DATA SET')
   50 FORMAT(4X,'ORIGINAL DATA SET')
   60 FORMAT(4X,' ATOMIC COORDINATES')
   70 FORMAT(4X,'CENTER OF LMO CHARGE CLOUD (X,Y,Z)'3F10.4)
   75 FORMAT(4X,'SIZE OF LMO CHARGE CLOUD (R,X,Y,Z);RMS'4F10.4)
```

Chart I (*Continued*)

```
     76 FORMAT(4X,'DISTANCE FROM ATOM ',I3,' =',F8.3,':: DX,DY,DZ ',3F8.3)
     80 FORMAT(1X,'* ',3F10.5)
     90 FORMAT(1X,'C ',3F10.5)
    100 FORMAT(2X,'0  0  0')
    110 FORMAT(2I3,'  1')
        STOP
        END
```

Table I. ChemCad+ Cartesian Input for H_2O

H	−0.901400	−0.055516	0.000000	
O	0.059733	−0.055516	0.000000	(atom no. and coordinates)
H	0.284701	0.878914	0.000000	
0 0 0				(bonding data to follow)
1 2 1				(atom 1 − atom 2), single bond
2 3 1				(atom 2 − atom 3), single bond
0 0 0				(end of file)

Table II. Insertion of Molecular Properties Data into the H_2O ChemCad+ File

H	−0.901400	−0.055516	0.000000	(atomic coordinate data)
O	0.059733	−0.055516	0.000000	
H	0.284701	0.878914	0.000000	
*	x_1	y_1	z_1	(molecular property data)
*	x_2	y_2	z_2	
....				
*	x_n	y_n	z_n	
0 0 0				(bond connectivity)
1 2 1				
2 3 1				
0 0 0				

perimpose or overlay the function, corresponding to a property of the system, on a screen-plot of the molecular structure. For molecular properties it is essential to correlate the function with the locations of atoms in the system. One way to overcome this is to include the coordinates of the atomic positions as part of the data to be plotted. By clustering several points on the atomic sites, the positions of atoms relative to charge density or other molecular property are distinguishable. The final drawback is the most crucial. The viewing orientation of a plot is difficult to optimize with SAS/Graph. Ideally, one needs to "interactively" orient the graph to determine the best view of the molecule and property. Though a specific limitation for SAS/Graph, this problem is shared by many general 3-D graphing routines. Therefore, the second option, use of molecular graphics software, should be adopted to analyze the molecular property if the orientation of the molecule is required or desirable.

In order to use PC-based software, such as ChemCad+, the data describing the molecular property must be modified to fit the format of the drawing/graphic software. In the current study, this is required since the graphics program cannot be modified.

The ChemCad+ program permits the input of the molecular description as a set of atomic labels, Cartesian coordinates, and bond identification data. An example of an input file for H_2O is given in Table I. The first three lines give the Cartesian coordinates of the H and O atoms. Line 4, "O O O", denotes the start of bonding information. Lines 5 and 6 contain the bond definitions. Three integers, i,j,k, define each bond. The first two, i,j, identify the numbers of the bonded atoms. The integer k refers to the strength of the bond; k = 1 defines a single bond. For water, single bonds are defined between "O" and each of the "H" atoms.

In order to study molecular properties the table of $\{x',y',z'\}$, values generated from an analysis of the $f_1(x,y,z)$ is added to such a molecular data file. Each of the data points corresponding to the molecular property is assigned a dummy atom symbol (* or X). This is illustrated in Table II. The set of $\{x'_i,y'_i,z'_i\}$ represents the data extracted from the $f_1(x,y,z)$ describing the molecular property. This file is then called by the ChemCad+ molecular graphics program. The atoms

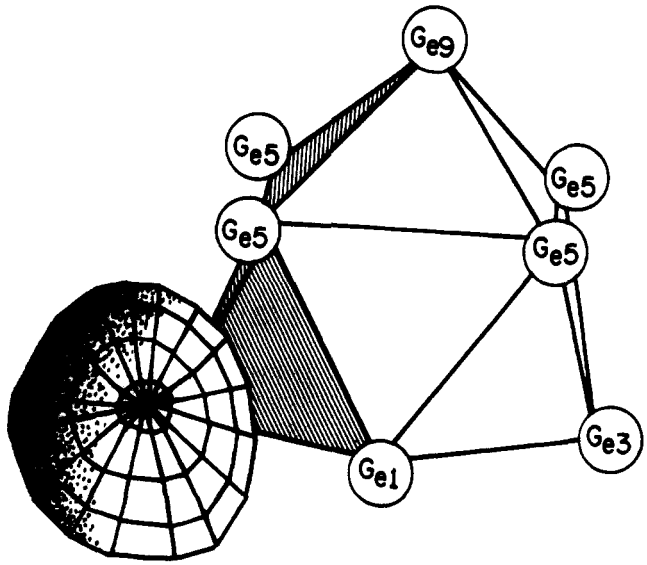

Figure 1. Ge_9^{2-} LMO corresponding to a lone pair on one of the Ge(3) atoms in the base of the polyhedron.

appear as normal bonded components, but the molecular property data shows up as a "cloud" of points. The entire system can then be rotated to obtain the optimum view. If desired, the entire drawing can be printed or plotted.

This method takes advantage of the molecular drawing and rotating capabilities of the ChemCad+ package. For other choices of software, the only requirement is that the program allows for enough "atoms", i.e., $\{x',y',z'\}$ values, to give a realistic picture of the molecular function.

The Fortran program used to carry out the data manipulation is listed in the Appendix. In addition to extracting $\{x,y,z\}$ from (x,y,z,f_1), the program also characterizes the "size", "location", and "shape" of the molecular property. For localized molecular orbitals, LMO's, the location, size, and shape of the charge cloud assist in understanding the bonding in the molecule. The use of the PC-based ChemCad+ program to study LMO's of a Zintl anion is illustrated in the following section.

SAMPLE CALCULATION USING CHEMCAD+

The procedure outlined in the previous section is now used to visualize and analyze localized molecular orbitals (LMO's) calculated for the Ge_9^{2-} anion using scattered-wave-Xα theory.[5] This nine-atom Zintl anion, crystallized in an organic matrix, has been characterized by using X-ray diffraction.[2] It exhibits C_{2v} symmetry with four symmetry-distinct atoms, denoted Ge(1), Ge(3), Ge(5), and Ge(9). The anion can be viewed as a capped-twisted-antiprism.

The SW-Xα method is a common technique for studying the electronic structure of molecular systems.[6] Chemical properties of polyhedral molecules, such as molecular orbitals, localized molecular orbitals (LMO's), and electron density, exceed the scope of traditional 2-D contouring methods. Determining the plane for the contouring of LMO data is especially difficult. In addition, for multicentered bonding the 2-D contouring may overlook important features of the electronic structure.

The first step in the process is the calculation of the 3-D

Table III. Abbreviated ChemCad+ File for Ge_9^{2-} Localized Molecular Orbital Shown in Figure 1

```
H    -2.98700    0.00000   -2.70900
B     0.00000    3.78100   -2.48000
C     2.68400   -2.60900    1.46900
N     0.00000    0.00000    4.50100
H     2.98700    0.00000   -2.70900
B     0.00000   -3.78100   -2.48000
C     2.68400    2.60900    1.46900
C    -2.68400   -2.60900    1.46900
C    -2.68400    2.60900    1.46900
*    -1.25000    1.25000    0.62500
*    -0.62500    1.25000    0.62500
......
*     0.62500    3.75000    3.12500
*     0.00000   -0.62500    4.37500
*    -1.87500    0.62500    4.37500
*     1.87500    0.62500    4.37500
*    -1.87500    2.50000    4.37500
*     1.87500    2.50000    4.37500
*     0.00000    3.12500    4.37500
*    -0.62500   -1.87500    5.00000
*     0.00000   -1.87500    5.00000
*     0.62500   -1.87500    5.00000
*     0.00000   -1.25000    5.62500
*     0.62500   -1.25000    5.62500
 0   0   0
 1   2   1
 1   3   1
 1   4   1
 1   5   1
 1   6   1
 1   7   1
 1   8   1
 1   9   1
 2   3   1
.......
 4   7   1
 4   8   1
 4   9   1
 5   6   1
 5   7   1
 5   8   1
 5   9   1
 6   7   1
 6   8   1
 6   9   1
 7   8   1
 7   9   1
 8   9   1
 0   0   0
```

$f_1(x,y,z)$. This entails calculation of the value of the LMO at each point of a 15 × 15 × 15 three-dimensional grid. Only points with a function value of greater than 0.0001 are retained. For the Ge_9^{2-} tabulations, about one-third of the points can be discarded with this cutoff value. The output of this calculation is a set of Cartesian coordinates, $\{x,y,z\}$, and the value of the LMO at each point. Using the program in the Appendix, this data set is transformed to a set of $\{x',y',z'\}$ corresponding to a specific value of the LMO. In addition the ChemCad+ bonding data is also written out by the program. The points which describe the LMO and the connectivity for the atoms in the molecule are appended to a file listing the molecular coordinates. This entire data set (molecular data, LMO coordinates, and bond data) is called into ChemCad+. An abbreviated input file for Ge_9^{2-} is shown in Table III. Once in ChemCad+, the molecule (with associated LMO data), can be interactively rotated, scaled, or duplicated as needed to provide a clear view of the LMO location and shape.

The valence electrons in Ge_9^{2-} are described with 19 LMO's. In Figure 1 one of these is presented. The three-dimensional aspect of the drawing has been enhanced by shading. For the LMO in Figure 1, the drawing indicates that this pair of electrons constitutes a lone pair, i.e., a nonbonding contribution, associated with one of the Ge(3) atoms.

In addition to preparing a file for the ChemCad+ program, the routine listed in the Appendix also provides an analysis of the relative position and shape of the LMO. First, the "center" of the LMO is calculated. This is useful for locating lone pairs and determining associations with particular atoms. In addition, the average x, y, and z distance of each point of the LMO from the "center" is determined. This provides a measure of the "shape" of the LMO. Spherical LMO's correspond to average x, y, and z widths of roughly equal value. Finally, the distances from the LMO center to each of the Ge atoms in the molecule are calculated. This aids in locating the LMO relative to atoms in the molecule. The LMO pictured in Figure 1 is found to be roughly spherical and located near one of the Ge(3) atoms.

CONCLUSIONS

The method presented in this work gives a simple approach for visualizing 3-D properties of chemical systems using PC-based molecular drawing software and standard graphing programs. By converting the $f_1(x,y,z)$ to a set of Cartesian coordinates, $\{x',y',z'\}$, corresponding to a specified value of f_1, the data can be easily manipulated. For properties requiring reference to the molecular structure, the PC-based ChemCad+ molecular drawing program is far superior to the SAS approach. The low cost of ChemCad+, $300, makes this a very attractive alternative to more expensive graphics workstation software. In addition, any molecular drawing package which accepts Cartesian data and allows user-defined rotations will work as well as ChemCad+. If a general plot is required, any graphics program capable of surface plots, i.e., "z" as a function of (x,y), can be used.

ACKNOWLEDGMENT

I thank the FAU Chemistry department for the purchase of computer software and hardware to carry out this work. In addition, the FAU Academic Computing Center provided access to the SAS/Graph software and computational support for the research.

APPENDIX

A Fortran computer program (see Chart I) was used to convert $f_1(x,y,z)$ to a set of Cartesian coordinates $\{x,y,z\}$ corresponding to a specific function value. In addition, this program analyzes the shape and relative location of the localized molecular orbital, LMO, in a molecule.

REFERENCES AND NOTES

(1) CHEM-X is a molecular modeling/graphics program from Chemical Design, Inc., 200 Route 17 South, Suite 120, Mahwah, NJ 07430. This program requires a graphics workstation.
(2) Corbett, J. D. Polyatomic Zintl Anions of the Post-Transition Elements. *Chem. Rev.* **1984**, *85*, 383–397.
(3) SAS/Graph is a standard statistical and graphics software package from the SAS Institute Inc., P.O. Box 8000, Cary, NC 27511. This package normally runs on a mainframe, Vax, or comparable minicomputer. The system is utilized with a graphics device of the users choice (plotter or terminal).
(4) ChemCad+ is a PC-based molecular modeling/drawing program from C_Graph Software, P.O. Box 5641, Austin, TX 78763.
(5) Jackson, M. D. Localized-Bonding Patterns in Nine-Atom Germanium Zintl Anions. Manuscript in preparation.
(6) Case, D. A. Electronic Structure Calculations Using the Xα Method. *Annu. Rev. Phys. Chem.* **1982**, *33*, 151–171.