1958

An electron diffraction investigation of the structure of free hydrocarbon molecules

Russell Aubrey Bonham
Iowa State College

AN ELECTRON DIFFRACTION INVESTIGATION OF THE STRUCTURE OF FREE HYDROCARBON MOLECULES

by

Russell Aubrey Bonham

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Physical Chemistry

Approved:

Signature was redacted for privacy.
In Charge of Major Work

Signature was redacted for privacy.
Head of Major Department

Signature was redacted for privacy.
Dean of Graduate College

Iowa State College
1958

TABLE OF CONTENTS

I. INTRODUCTION

The purpose of this work was to determine accurate molecular parameters for ethylene, isobutylene, n-butane, and n-heptane. Since the results of early electron diffraction and spectroscopic work on ethylene and isobutylene (1) are in conflict with some of the current ideas about the structure of molecules, it was felt that these two molecules should be checked by more accurate methods. In the cases of n-butane and n-heptane, it was hoped that electron diffraction studies could determine the shape of the molecules since there has been some controversy (2-8) in recent years as to whether these molecules are fully extended or coiled. Also, it was hoped that the approximate barrier to hindered rotation about carbon-carbon bonds in the normal hydrocarbons and about the methyl groups in isobutylene could be determined.

The history of structure determinations of gaseous molecules dates back to 1915-1929 when P. J. W. Debye, Bewilogua, and Ehrhardt (9, 10) obtained the first results of diffraction experiments performed on gas molecules with X-rays. At about the same time Davisson and Germer (11) demonstrated experimentally the wave nature of electrons. Shortly after the wave nature had been shown, Mark and Wierl (12) obtained crude structural parameters of a gas molecule by electron scattering experiments.

As a result of the application of quantum mechanics to

the ideas that had been formulated on the scattering of X-rays and electrons, a rigorous mathematical scattering theory was developed (13). These ideas, coupled with the theory, indicated that X-rays see only the nebulous planetary electrons around the nuclei of each atom while the electrons see the nuclei of the atoms screened by the planetary electrons. Therefore, X-rays determine nuclear positions only indirectly by determining the shape of the electron cloud, whereas electrons actually give a direct measurement of nuclear positions. In the case of electrons, the ratio of elastic to inelastic scattering approaches the atomic number of the atom from which the scattering takes place as the scattering angle is increased. With X-rays this ratio goes to zero as the scattering angle is increased. This last difficulty, together with the smaller wave lengths used with electrons, makes electron diffraction a more suitable tool for disclosing the structure of gas molecules. However, electron diffraction has a disadvantage due to the rapid decrease in intensity that makes it difficult to obtain accurate intensity measurements.

The theory of electron diffraction indicates that the scattering which determines the molecular parameters appears as small sinusoidal oscillations about the smooth rapidly decreasing atomic scattering. By using low contrast photographic plates Mark and Wierl (12) were able to correlate the apparent maxima and minima on the photographic plates with

theoretical intensity curves. To do this, the relative intensities and positions of the maxima and minima were compared with a series of theoretical curves computed from various trial structures of the molecule. The model in best agreement with the experiment was then chosen as the most probable structure of the molecule. Although Mark and Wierl originated this method of structure determination, Pauling and Brockway (14) were responsible for a major part of its refinement and use in structure work. This method was termed the visual method because of the way in which the relative intensities were obtained. The visual method's great advantage was its speed. In fact by 1936 some one hundred and forty different molecules had been investigated by the method (15).

Before 1932 the theory had only been worked out for molecules that were assumed to be rigid frameworks of scattering centers. It had also been assumed that all the electron scattering that occurred from a molecule was elastic. James (16) was able to show how to modify the scattering equations so that thermal motion could be accounted for. The contribution to the total intensity from inelastic scattering was taken into account by Morse (17) and Bewilogua (18) in their work on the X-ray incoherent scattering factor, but the experiments were still too crude to measure accurately the effects of all the new contributions to the theory. In 1935

Pauling and Brockway (19) introduced the radial distribution function, which was the Fourier sine transform of the molecular scattering intensity. This too was crude since the intensities were estimated visually and only at the maxima and minima of the molecular intensity function.

It was not until Finbak (20) and P. P. Debye (21) performed the first rotating sector experiments that objective measurements of intensities were obtained rather than visual estimates. A rotating sector was used so that the recorded intensity was in the useful range of response of the photographic plate. The sector was a thin cardioid shaped plate which was spun parallel to and slightly above the photographic plate. With the photographic records of sector experiments it was possible to use with some precision recording microphotometers to measure the experimental intensities. A technique developed with unsectored plates by Degard, Pierard, and Van der Grinten (22) was used with sectored plates which were spun to average out emulsion errors while the microphotometer recordings were being made.

Almost simultaneously, improvements were being made in the radial distribution method. Degard (23) suggested that an artificial damping factor be used to minimize integral termination errors in the Fourier inversion technique. By 1940 Walter and Beach (24) had suggested an analytic method for improving radial distribution curves, and Spurr and

Schomaker (25) had employed I.B.M. machines to compute radial distribution curves from visually estimated intensities at many points along the theoretical scattering curve.

In the years that have followed, an almost completely objective procedure for the determination of molecular structure has been devised. The problems involved in making electron diffraction a precise tool for structure determination have been investigated by P. J. W. Debye (26) and Harvey, Keidel, Bauer, Coffin, and Hastings (27-30) at Cornell University, C. Finbak and Hassel and Viervoll (20, 31) at the University of Oslo, Norway, I. L. and J. Karle (32-34) at the Naval Research Laboratory, L. S. Bartell, L. O. Brockway and R. Schwendeman (1, 35-40) at the University of Michigan, Y. Morino, K. Kuchitsu, T. Shimanouchi, A. Takahashi, and K. Maeda (41, 42) at the University of Tokyo, Japan, and Schomaker and Glauber (43, 44) at the California Institute of Technology. It appears, however, that there are still many unsolved problems in this field.

The more recent work in electron diffraction, in this and other laboratories, indicates several directions which future electron diffraction investigators may take. As in the past, electron diffraction will continue to be an important means of accurate structure determination. Precise bond distances and vibrational amplitudes will be obtained (45) as well as approximate barriers to hindered rotation (46). In addition,

studies of electron distributions in atoms by means of electron diffraction are possible (40). Some recent work done in this laboratory indicates the possibility of determining qualitative differences between the form factors for spherical atoms and those for bonded atoms. It may also be possible to study vibrational amplitudes involved in hindered rotation as a function of temperature (47, 48). These advances and possibilities for the future are a direct consequence of the ability to obtain more accurate experimental intensities. Perhaps the most important advance in the interpretation of electron diffraction data has been brought about by the development of high speed digital computers. By using these computers it is now possible to use more exact theoretical expressions and more powerful methods of analysis in the interpretation of electron diffraction data.

II. THE THEORY OF ELECTRON DIFFRACTION

A. Derivation of the Intensity Function

Two main theoretical approaches have been used in explaining the observed intensity of electrons diffracted by atoms. The kinematical theory assumes that every infinitesimal element of volume of the atom scatters under the influence of the incident beam only. The total amplitude is then obtained by summing up the amplitudes scattered from all the various volume elements taking account of the difference in phase due to the different positions occupied by the volume elements. This method is valid as long as the scattered intensity is much weaker than the intensity of the incident beam. The other approach is the method of partial waves, which has been applied by Hoerni and Ibers (49, 50) to the problem of electron scattering by atoms. This method gives a more rigorous solution to the atomic scattering problem than does the Born approximation or kinematical approach. For instance, the kinematical approach does not take into account corrections due to a shift in phase of the incident wave whereas the method of partial waves does.

For molecular scattering the simplest method of attack is to assume that the total scattered amplitude from a molecule in a rarefied gas is the sum of the amplitudes scattered from the individual atoms in the molecule. A more rigorous

starting assumption would lead to a more complicated formulation of the theory. This complication does not seem warranted since the above assumption leads to a theory which seems to correlate very well with the experimental results as long as the atoms in the molecule do not differ greatly in atomic number. If the scattered amplitudes from the atoms are computed by the method of partial waves, the resultant molecular scattering theory is termed pseudokinematical by Hoerni (51). Use of the kinematical atomic amplitudes to compute the molecular amplitude is referred to as the molecular kinematical theory, and it has been shown by Hoerni (51) that there is little difference between the results of the two methods of approach to the molecular problem. Since the kinematical theory is mathematically simpler, it will be employed here.

The time independent Schrödinger equation may be put in the convenient form for scattering problems

$$(1) \qquad (\nabla^2 + k^2)\,\Psi(\vec{r}) = 2m\,V(\vec{r})\,\Psi(\vec{r})/\hbar^2 = U(r)\,,$$

where

$$k = \sqrt{2mE}\,/\hbar\,.$$

If $\Psi(\vec{r})$ and $U(\vec{r})$ are both expanded in eigenfunctions of the operator $(\nabla^2 + k^2)$, where the eigenfunctions are box normalized then,

$$(2) \qquad \Psi(\vec{r}) = \sum_{n=o}^{\infty} a_n u_n(\vec{r})$$

and $$U(r) = \sum_{n=o}^{\infty} b_n u_n(\vec{r}),$$

where the $u_n(\vec{r})$ form a complete normalized orthogonal set, the a_n are to be determined, and the b_n are defined as below:

$$(3) \qquad b_n = 2m \int d^3x' \, V(\vec{r}') \, \Psi(\vec{r}') \, u_n(\vec{r}')/\hbar^2 .$$

The notation $\int d^3x'$ or $\int d^3x$ refers to an integration over all space. The original eigenvalue equation (1) can now be written

$$(4) \qquad \sum_{n=o}^{\infty} a_n(\nabla^2 + k^2) \, u_n(\vec{r}) = \sum_{n=o}^{\infty} b_n \, u_n(\vec{r}) ,$$

but since the $u_n(\vec{r})$ are eigenfunctions of the operator, $(\nabla^2 + k^2)$, such that

$$(5) \qquad (\nabla^2 + k^2)u_n(\vec{r}) = \lambda_n u_n(\vec{r}),$$

then substitution into equation (4) gives

$$(6) \qquad \sum_{n=o}^{\infty} (a_n\lambda_n - b_n)u_n(\vec{r}) = 0.$$

Since the $u_n(\vec{r})$ are linearly independent, the coefficients $a_n\lambda_n - b_n$ must all vanish. Thus

$$a_n = b_n / \lambda_n,$$

and, by substitution, it is possible to formulate the integral

equation for scattering

$$(7) \quad \Psi(\vec{r})_{sc.} = 2m \int d^3x' V(\vec{r}') \Psi(\vec{r}') \sum_{n=o}^{\infty} u_n^*(\vec{r}') \, u_n(\vec{r}) / \lambda_n \hbar^2 ,$$

where $\sum_{n=o}^{\infty} u_n^*(\vec{r}') \, u_n(\vec{r}) / \lambda_n$ is a Green's function denoted as $G(\vec{r}', \vec{r})$. Here it should be noted that $\Psi(\vec{r})_{sc.}$ is a particular solution of the inhomogeneous equation (1) and represents the amplitude of the scattered wave if $G(\vec{r}', \vec{r})$ is chosen for an outgoing wave. A general solution of the homogeneous equation

$$(\nabla^2 + k^2) \Psi(\vec{r})_o = 0$$

is given by $$\Psi(\vec{r})_o = e^{i\vec{k}\cdot\vec{r}} .$$

The complete solution of the Schrödinger equation for the scattering problem is made up of a particular solution of the inhomogeneous equation plus a general solution of the homogeneous equation. If we choose the incident beam to be composed of plane waves in the z direction, then

$$\Psi(\vec{r})_o = e^{ikz}$$

is the plane wave solution of the homogeneous equation for an incident monochromatic wave coming from the left. The complete solution is then

$$(8)\quad \Psi(\vec{r}) = \Psi(\vec{r})_o + \Psi(\vec{r})_{sc.} = e^{ikz} + 2m\int d^3x'\, V(\vec{r}')\,\Psi(\vec{r}')\,G(\vec{r}',\vec{r})\,/\hbar^2$$

The value of the Green's function $G(\vec{r}',\vec{r})$ for a scattered wave may be obtained from Mott and Massey (52) or Morse and Feshbach (53). For the scattered wave, it is $-e^{ik|\vec{r}-\vec{r}'|}/4\pi\,|\vec{r}-\vec{r}'|$. For atomic scattering the potential $V(\vec{r}')$ falls off rapidly with distance so that the potential is essentially zero at the point at which the intensity of the scattered wave is to be measured. Because of this, it is possible to expand $|\vec{r}-\vec{r}'|$ to $r - (\vec{r}\cdot\vec{r}'/r)$ in the phase and $|\vec{r}-\vec{r}'|$ to r in the denominator. This gives

$$(9)\quad \Psi(\vec{r})_{sc.} = -me^{ikr}\int d^3x' V(\vec{r}')\Psi(\vec{r}')\,e^{-ik\vec{r}\cdot\vec{r}'/r}/2\pi\hbar^2 r\ ,$$

where $\hbar k(\vec{r}/r)$ is the momentum of the scattered wave. If

$$\Psi_o(\vec{r}) = e^{ikz} >> -me^{ikr}\int d^3x' V(r')\Psi(\vec{r}')e^{-ik\vec{r}\cdot\vec{r}'/r}/2\pi\hbar^2 r\ ,$$

then to good approximation $\Psi(\vec{r}')$ may be replaced by $\Psi(\vec{r}')_o$. This is known as the first Born approximation and is generally valid if

$$V(\vec{r}') << E_o\ ,$$

where E_o is the energy of the incident beam. The second Born approximation would be obtained by letting

$$\Psi(\vec{r}') = \Psi_o(\vec{r}') + \Psi(\vec{r}')_{sc.}\ ,$$

where $\Psi(\vec{r}')_{sc.}$ is determined by the first Born approximation.

For the first Born approximation,

$$(10)\quad \Psi(\vec{r})_{sc.} = -\,me^{ikr}\int d^3x' V(\vec{r}')e^{i(\vec{k}-k(\vec{r}/r))\cdot\vec{r}'}/2\pi\hbar^2 r$$

and $$|\vec{k} - k(\vec{r}/r)| = 4\pi \sin(\theta/2)/\lambda = s,$$

where λ is the wave length of the incident wave and θ is the scattering angle. The scalar s will be called the scattering variable, and it should be noted that $\hbar\vec{s}$ is the change in momentum of the scattered beam. So far, of course, we have only been considering perfectly elastic collisions. If $V(\vec{r}')$ is a spherically symmetric potential, equation (10) may be further reduced to

$$(11)\quad \Psi(\vec{r})_{sc.} = -(2m\, e^{ikr}/\hbar^2 r)\int_0^\infty dr' r'^2 V(r') \sin sr'/sr' \quad .$$

It is of interest to express the scattered amplitude in a form analogous to that used in X-rays. To do this, it is convenient to introduce the charge distribution $\rho(\vec{r})$ such that

$$(12)\quad V(\vec{r}') = e^2 \int d^3x\, \rho(\vec{r})/4\pi\, |\vec{r}' - \vec{r}| \quad .$$

In order to introduce the charge density into the scattering equations, it will be convenient to make use of the folding theorem of Fourier integrals (54). This theorem says

that the product of the Fourier transforms of two functions g(x) and f(x) is the Fourier transform of the integral of the products. In the case we are considering here, this results in

$$(13)\quad e^2 \int d^3x' \rho(\vec{r}') e^{i\vec{s}\cdot\vec{r}'} \int d^3x e^{i\vec{s}\cdot(\vec{r}-\vec{r}')}/4\pi|\vec{r}-\vec{r}'| = \int d^3x' V(\vec{r}')\, e^{i\vec{s}\cdot\vec{r}'} .$$

It should be noted that $\int d^3x(1/|\vec{r}-\vec{r}'|)e^{i\vec{s}\cdot(\vec{r}-\vec{r}')}$, where d^3x refers to the integration over all $\vec{r}-\vec{r}'$ space, is easily evaluated (52) as $4\pi/s^2$. Substituting this result in equation (13) gives

$$(14)\quad \int d^3x' V(\vec{r}')\, e^{i\vec{s}\cdot\vec{r}'} = e^2 \int d^3x' \rho(\vec{r}') e^{i\vec{s}\cdot\vec{r}'}/s^2 .$$

It is of interest to note that $-1/4\pi|\vec{r}-\vec{r}'|$ is the Green's function for a coulomb potential. Our scattered amplitude is now reduced to

$$(15)\quad \Psi(\vec{r})_{sc.} = -2me^2\, e^{ikr} \int d^3x' \rho(\vec{r}') e^{i\vec{s}\cdot\vec{r}'}/\hbar^2 s^2 r .$$

We will now assume an atomic charge distribution of the form

$$\rho(\vec{r}) = -Z\delta(\vec{r}) + \rho_{el.}(\vec{r}) ,$$

where $\delta(\vec{r})$ is the Dirac delta function (55) defined by the properties $\delta(\vec{r}) = 0$ for $r \neq 0$ and $\int d^3x\, \delta(\vec{r}) = 1$ while $\rho_{el.}(\vec{r})$ stands for the probability density of the planetary electrons in the atom. Substituting this into

equation (15) gives

$$(16)\quad \psi(\vec{r})_{sc.} = 2me^2 e^{ikr}/Z \int d^3x' \, \delta(\vec{r}') e^{i\vec{s}\cdot\vec{r}'} - \int d^3x' \rho_{el.}(\vec{r}') e^{i\vec{s}\cdot\vec{r}'} / \hbar^2 s^2 r \quad .$$

If one assumes that the distribution of electrons in the atom is spherically symmetric, then, using this fact with the definition of the δ-function, equation (16) reduces to

$$(17)\quad \psi(\vec{r})_{sc} = 2me^2 e^{ikr}(Z - F(s)) / \hbar^2 s^2 r \quad ,$$

where F(s) is given by the equation

$$(18)\quad F(s) = \int_0^\infty dr \; 4\pi r^2 \rho_{el.}(r) \sin sr/sr \quad .$$

Here F(s) is the X-ray atom form factor for the elastic scattering of X-rays. The scattered intensity for electrons is then given by

$$(19)\quad I \propto 4m^2 e^4 (Z - F(s))^2 / \hbar^4 s^4 r^2 \quad ,$$

whereas the X-ray elastic scattering intensity is proportional to $F(s)^2$. This points out the fact that electron scattering is nuclear scattering screened by the planetary electrons while X-ray scattering is due to the planetary electrons. In the proportionality (19) an equality sign could be introduced where the necessary proportionality constant would be the product of the intensity of the incident electron beam and the number of atoms per unit area exposed to the beam.

Next we consider the problem of scattering from a fixed framework of rigid atoms. To do this, it is assumed that the total amplitude of the scattered beam observed at some point is just the sum of the amplitudes scattered from each atom in the framework. If the charge distribution is assumed to be

$$(20)\quad \rho(\vec{r}') = \sum_{j=1}^{N} \rho_j(\vec{r}') = \sum_{j=1}^{N} [-Z_j \delta(\vec{r}' - \vec{r}_j) + \rho_{j\ el.}(\vec{r}' - \vec{r}_j)] ,$$

then equation (17) reduces to

$$(21)\quad \Psi(r)_{sc} = \sum_{j=1}^{N} 2\ me^2 e^{ikr}(Z_j - F_j(s))e^{i\vec{s}\cdot\vec{r}_j}/\hbar^2 s^2 r ,$$

where the sum is over all the atoms in the rigid framework and $F_j(s)$ is the X-ray atom form factor for the j^{th} atom. If the intensity is now computed, remembering the proportionality factor, which depends on beam intensity and sample concentration, the result is

$$(22)\quad I = (NI_o)\ 4m^2e^4 \left\{ \sum_j (Z_j - F_j(s))^2 + \sum_{jk}{}' (Z_j - F_j(s)) (Z_k - F_k(s))e^{i\vec{s}\cdot\vec{r}_{jk}} / \hbar^4 s^4 r^2 \right\}.$$

The primed summation indicates that terms are omitted for which i = j. Here N is the number of molecules per unit area exposed to the beam and I_o is the intensity of the incident beam. The summation of $(Z_j - F_j(s))^2$ is, of course, due to scattering from the individual atoms. The vector

$$\vec{r}_{jk} = \vec{r}_j - \vec{r}_k$$

in the second term is the distance between the j^{th} and k^{th} nuclei, and it is this coherent phase relation that makes it possible to determine internuclear parameters.

In equation (22) only elastic scattering has been considered. Thus, to improve the theory, it is necessary to include a correction for inelastic scattering. Morse (17) has shown that for inelastic collisions of fast electrons with an atom of nuclear charge Ze the intensity is

$$(23) \quad I = I_o N 4m^2e^4 S(s) / s^4 4r^2 \ ,$$

and

$$S(s) = ZS'(s) \ ,$$

where S'(s) is the factor tabulated by Bewilogua (18) and S(s), is referred to as the X-ray atom form factor for inelastic scattering. For a light atom such as carbon, the inelastic intensity is roughly sixteen per cent of the elastic atomic scattering intensity at large s although it is greater at small s. If equation (23) is added to equation (22), the result is

$$(24) \quad I_T = I_o N 4m^2e^4 \Big[\sum_j \{(Z_j - F_j(s))^2 + S_j(s)\} + \sum_{ij}{}' (Z_i - F_i(s))(Z_j - F_j(s)) e^{i \vec{s} \cdot \vec{r}_{ij}} \Big] / s^4 4r^2 \ .$$

In Equation (24) the molecules or rigid framework of atoms may be averaged over all possible orientations to give for the intensity contribution from the primed summation

$$(25)\quad I_m = I_o N 4m^2e^4 \left[\sum_{ij}' (Z_i - F_i(s))(Z_j - F_j(s)) \sin sr_{ij}/sr_{ij} \right] / \not{n} 4s 4r^2 .$$

A molecule is actually not a rigid framework, therefore, the intensity must also be averaged over distributions of internuclear distances. Thus

$$(26)\quad \langle I_m \rangle_{av.} = I_o N 4m^2e^4 \left[\sum_{ij}' (Z_i - F_i(s))(Z_j - F_j(s)) \int_o^\infty dr\, 4\pi r^2 \rho_{ij}(r) \sin sr/sr \right] / \not{n} 4s 4r^2$$

It should be noted that for a diatomic molecule at temperatures not too great

$$4\pi r'^2 \rho_{o_{ij}}(r') = \Psi_o^*(r') \Psi_o(r') ,$$

where $\Psi_o(r')$ is the radial part of the nuclear vibrational wave function for the lowest state. Equation (26) may be incorporated into equation (24) giving

$$(27)\quad I_T = N I_o 4m^2e^4 \left[\sum_j \left\{ (Z_j - F_j(s))^2 + S_j(s) \right\} + \sum_{ij}' (Z_i - F_i(s))(Z_j - F_j(s)) \int_o^\infty dr P_{ij}(r) \sin sr/sr \right] / \not{n} 4s 4r^2 ,$$

where

$$P_{ij}(r) = 4\pi r^2 \rho_{ij}(r) \quad .$$

As noted above, $P_{ij}(r)\,dr$ is the probability that the separation of the i^{th} and j^{th} nuclei is between r and r + dr. Equation (27) as written above is the equation that is used in the analysis of diffraction patterns in this laboratory.

B. Derivation of the Radial Distribution Function

The last term on the right of equation (27) contains the Fourier transform of the probability distribution of nuclear separations $P_{ij}(r)$. It is of interest to be able to obtain $P_{ij}(r)$ as a function of the experimentally measurable intensity I_T. If equation (27) is rearranged into the suggestive form

$$(28)\quad \left[\sum_{ij}{}' (Z_i - F_i(s))(Z_j - F_j(s)) \Big/ \sum_k \left\{ (Z_k - F_k(s))^2 + S_k(s) \right\}\right] \int_0^\infty dr P_{ij}(r) \sin(sr)/sr = \left[s^4 / h^4 r^2 I_T /\right.$$

$$\left. 4 I_e N m^2 e^4 \sum_k \left\{ (Z_k - F_k(s))^2 + S_k(s) \right\}\right] - 1$$

It is easily seen that the Fourier sine transform of the quantity on the left would give the desired probability distribution of internuclear distances if the coefficient of the integral were independent of the scattering variable s. Two

new quantities $M(s)$ and $M(s)_c$ may be defined as

$$(29)\quad M(s) = \left(\sum_{ij}{}'(Z_i-F_i(s))(Z_j-F_j(s))/\sum_k \left\{(Z_k-F_k(s))^2 + S_k(s)\right\}\right) \int_0^\infty dr P_{ij}(r)\ \sin(sr)/sr$$

and

$$(30)\quad M(s)_c = \left(\sum_{ij}{}' Z_iZ_j/\sum_k (Z_k^2+Z_k)\right) \int_0^\infty dr P_{ij}(r)\sin(sr)/sr\ .$$

Here $Z_iZ_j/\sum_k (Z_k^2 + Z_k)$ is the limit of $(Z_i-F_i(s))(Z_j-F_j(s)/\sum_k \left\{(Z_k-F_k(s))^2 + S_k(s)\right\}$ as s goes to infinity. The Fourier sine transform of equation (30) is

$$(31)\quad \int_0^\infty sM(s)_c \sin(sr)ds = \pi\left(\sum_{ij}{}' Z_iZ_j/\sum_k (Z_k^2+Z_k)\right)P_{ij}(r)/2r\ ,$$

where the quantity on the right is referred to as the unmodified radial distribution function and is conventionally denoted by $D(r)$. It can be seen from equation (31) that $D(r)$ must be nowhere negative because the probability $P_{ij}(r)$ cannot be negative. This, coupled with the fact that $M(s)$ is not significantly different from $M(s)_c$ over most of the range of the scattering variable, allows us to get an approximate $D(r)$ function using $M(s)$ rather than $M(s)_c$ in equation (31).

If a sector is used in obtaining the experimental intensity, the intensity obtained, I_r, is proportional to s^nI_T, where n is usually between two and four. In order to

obtain M(s) from this intensity, it is necessary to determine a function B such that

$$(I_r/B)-1 = M(s) \ .$$

This B function is determined by an iterative procedure in which an arbitrarily chosen B function is successively modified so that the resultant D(r) function is nowhere negative and B is smooth enough to keep false detail from being introduced into D(r). With the parameters obtained from the D(r) function, it is then possible to compute theoretically $M(s)_c$-M(s) and add this to the experimental M(s). This, in effect, reduces the experimental M(s) to an experimental $M(s)_c$ by correcting for the non-nuclear scattering. The radial distribution curve derived from the corrected M(s) function with the use of $M(s)_c$ data at small s is the curve from which final parameters may be obtained. It is still necessary, however, to specify the form of the function $P_{ij}(r)$ if theoretical intensity curves are to be computed. Two possible approximations will be considered here.

1. Gaussian distribution of internuclear distances

Assuming the nuclear potential to be that of a simple harmonic oscillator, the distribution function is

$$(32) \quad P_{ij}(r) = (1/\sqrt{2\pi}\,\ell_{ij})e^{-x_{ij}^2/2\ell_{ij}^2} \ ,$$

where

$$x_{ij} = r - r_{ij} .$$

Here ℓ_{ij} is the root mean square amplitude of vibration and r_{ij} is the most probable distance between the i^{th} and j^{th} nuclei. If this definition is used in equation (29) or (30) and the integration carried out, the results to first order corrections in the argument of the sine in M(s) and $M(s)_c$ functions are

$$(33) \quad M(s) = \left(\sum_{ij}' (Z_i - F_i(s))(Z_j - F_j(s)) / \sum_k \left\{ (Z_k - F_k(s))^2 + S_k(s) \right\} \right) e^{\frac{-\ell_{ij}^2 s^2}{2}} \sin\left[s(r_{ij} - \ell_{ij}^2 / r_{ij}) \right] / s r_{ij}$$

and

$$(34) \quad M(s)_c = \left(\sum_{ij}' Z_i Z_j / \sum_k (Z_k^2 + Z_k) \right) e^{\frac{-\ell_{ij}^2 s^2}{2}} \sin\left[s(r_{ij} - \ell_{ij}^2 / r_{ij}) \right] / s r_{ij} .$$

Here corrections to the amplitude have been neglected. It is now possible to give several definitions of bond length in the Gaussian sense. The maximum and center of gravity of the $P_{ij}(r)/r$ peak are given as

$$r_m(1) = r_{ij} - \ell_{ij}^2 / r_{ij}$$

and

$$r_g(1) = r_{ij} - \ell_{ij}^2/r_{ij}$$

respectively. The maximum and center of gravity of the $P_{ij}(r)$ function by itself is

$$r_m(o) = r_{ij}$$

and

$$r_g(o) = r_{ij}$$

respectively. Here $r_m(n)$ denotes the position of the maximum value of the $P_{ij}(r)/r^n$ function and $r_g(n)$ denotes the center of gravity of the same function. It should be noted that r_{ij} is also the position of the minimum of the potential function. One other possible means of defining an internuclear parameter would be as the average value of r^{-2} defined as

$$\langle (r^{-2}) \rangle_{Av.}^{-1/2} = r_{ij} - 3\ell_{ij}^2/2r_{ij} \quad .$$

This last definition corresponds closely with that used in spectroscopy. In this laboratory bond distances are reported using the nomenclature $r_m(n)$ and $r_g(n)$. It is important that the particular definition of bond distance used be specified since discrepancies arising from using different definitions may amount to as much as one hundredth of an angstrom unit. Since actual $P_{ij}(r)/r$ peaks in an experimental D(r) function are not exactly Gaussian, there is a limit to how well the

peaks may be characterized by the Gaussian approximation. It has been shown for diatomic molecules (45) that differences as large as .008 of an angstrom may occur between the center of gravity and the maximum of the radial distribution peak. To characterize $P_{ij}(r)/r$ peaks more accurately, it is necessary to consider a more accurate description of $P_{ij}(r)$.

2. Distribution of internuclear distances based on the Morse potential function (45)

The Morse potential function is given as

$$V(r) = De^{-2a(r-r_e)} - 2De^{-a(r-r_e)} , \tag{35}$$

where D is the potential energy at the minimum, r_e is the distance from the origin to the potential minimum, and a is an asymmetry constant. The ground state wave function for this potential is

$$\psi_o(r) = Ke^{-\left[(a/2)(1/\ell^2a^2-1)(r-r_e)+(1/2\ell^2a^2)e^{-a(r-r_e)}\right]} , \tag{36}$$

where K is a constant and ℓ is the root mean square amplitude of vibration. If the substitution $x = r - r_e$ is made in equation (36), the probability distribution $P_o(r)$ is computed as

$$P_o(r) = K^2 e^{-\left[(1/\ell^2a^2)-1)ax + (1/\ell^2a^2)e^{-ax}\right]} . \tag{37}$$

Next, expanding the exponentials and neglecting terms in x^2

and higher powers of x than x^3 gives the approximate normalized result

$$(38) \quad P_o(r) = (1/\sqrt{2\pi}\ell)(1+ax+ax^3/6\ell^2)e^{-x^2/2\ell^2} .$$

Here the neglected terms would correspond to a correction in the maximum or center of gravity of $P_o(r)$ of about 10^{-4} angstroms.

The position of the maximum of the function $P_o(r)/r^n$ designated as $r_m(n)$ is given by

$$(39) \quad r_m(n) = r_e + a\ell^2 - (n\ell^2/r_e).$$

The center of gravity of the function $P_o(r)/r^n$ is

$$(40) \quad r_g(n) = r_e + (3a\ell^2/2) - (n\ell^2/r_e)$$

and the average value is

$$(41) \quad \langle (r^n) \rangle_{Av.}^{1/n} = r_e + (3a\ell^2/2) + ((n-1)\ell^2/2r_e)$$

These results have been derived for diatomic molecules, but to fairly good approximation they may be used for terminal bonded distances in polyatomic molecules. For non-bonded distances, however, this approximation is not usually applicable since the potential well has not yet been well characterized and since the bending modes are frequently not in zero point vibrational states. Actually, if the asymmetry constant a is allowed to become an arbitrary parameter characterizing the asymmetry of the $P_{ij}(r)$ peaks, then

equation (38) may be generalized for polyatomic molecules.

If the $P_c(r)$ for a diatomic molecule is substituted in equation (30) and the integral evaluated keeping only terms important to the first order correction in the argument of the sine, the result is

$$(42)\quad M(s)_c = (Z_1Z_2/\sum_{k=1}^{2}(Z_k^2+Z_k))e^{-\ell^2 s^2/2}\sin(s[r_o+\rho(s)])/sr_o \ ,$$

where

$$(43)\quad \rho(s) = a\ell^2 - (\ell^2/r_o) + (a\ell^2/2)(1 - s^2\ell^2/3)$$

for $s^2 < 3/\ell^2$. This result may be generalized to

$$(44)\quad M(s)_c = (\sum_{ij}{}' Z_iZ_j/\sum_{k}(Z_k^2+Z_k))e^{-\ell_{ij}^2 s^2/2}\sin[s(r_{ij})+\rho_{ij}(s)]/$$

$$sr_{ij} \ ,$$

where

$$(45)\quad \rho_{ij}(s) = a_{ij}\ell_{ij}^2 - \ell_{ij}^2/r_{ij} + (a_{ij}\ell_{ij}^2/2)(1-s^2\ell_{ij}^2/3) \ .$$

This expression may be used for terminal bonded distances assuming that the asymmetry constants a_{ij} are the same as those in diatomic molecules, which generally have a value of two for a . There is also the further possibility of using a least squares technique to fit the above theoretical expressions to experimental data where the a_{ij}, ℓ_{ij}, and r_{ij} are all

parameters to be determined.

It is convenient to define the radial distribution function f(r) as

$$(46) \quad f(r) = \int_0^{s_{max}} s\, M(s)_c\, e^{-bs^2} \sin(sr)\, ds ,$$

where the factor e^{-bs^2} is a combination damping factor to prevent integral termination errors and weighting function. The weighting function is important since there is a larger error in the experimental $sM(s)$ at large values of s than at intermediate values. The constant b is often chosen so that $e^{-bs^2} = 0.1$, where $s = 30$.

Substituting equation (30) into equation (46) and replacing the variable r in equation (30) with ρ gives

$$(47) \quad f(r) = \left(\sum_{ij}{}' Z_i Z_j / \sum_k (Z_k^2 + Z_k)\right) \int_0^\infty d\rho \int_0^{s_{max}} ds\, \rho^{-1} P_{ij}(\rho) \sin(sr) \sin(s\rho) e^{-bs^2} .$$

Integrating over s and remembering that the use of an adequate e^{-bs^2} function makes the upper limit s_{max} effectively infinity gives

$$(48) \quad f(r) = \left(\sum_{ij}{}' Z_i Z_j / \sum_k (Z_k^2 + Z_k)\right) \sqrt{\pi/6b} \int_{-\infty}^{\infty} \rho^{-1} P_{ij}(\rho) e^{\frac{-(r-\rho)^2}{4b}} d\rho .$$

Substituting $P_o(r)$ from equation (38) for $P_{ij}(\rho)$ and

integrating gives

$$(49)\quad f(r) = (1/\sqrt{\pi(4b+2\ell_{ij}^2)})\left[\sum_{ij}{}'(Z_iZ_j/\sum_k(Z_k^2+Z_k)\right](1/r_{ij})$$

$$e^{-(r-r_{ij})^2/4b+2\ell_{ij}^2}\left\{1+(a_{ij}-r_{ij}^{-1})\ell_{ij}^2(r-r_{ij})\right.$$

$$\cdot(2b+\ell_{ij}^2)^{-1} + a_{ij}b(r-r_{ij})(2b+\ell_{ij}^2)^{-1}$$

$$\left. + a_{ij}\ell_{ij}^4(r-r_{ij})^3(2b+\ell_{ij}^2)^{-3}/6\right\}.$$

Equation (49) is useful in computing anharmonicity corrections. If the small contributions to f(r) due to the terms containing a_{ij} are computed separately for terminal bond distances and subtracted from the experimental f(r) function, the result is a sum of pure Gaussian peaks and is simpler to evaluate by a least squares technique.

III. COMPUTATIONAL PROCEDURES

In order to analyze diffraction data rigorously in the determination of molecular structure, it is necessary to have rapid methods of computation available to compute intensity and radial distribution functions. In this laboratory the calculation of the functions was programmed for the I.B.M. 650.

A. Theoretical Intensity Function

A program was written and coded for the functions

$$(50)\quad M(q) = \Big[\sum_{ij}{}'(Z_i - F_i(q))(Z_j - F_j(q))\Big/\sum_k \{(Z_k - F_k(q))^2 + S_k(q)\}\Big] e^{-\ell_{ij}^2 \pi^2 q^2/200} \Big[\sin(\pi q r_{ij}/10)\Big]/\pi q r_{ij}/10$$

and

$$(51)\quad M(q)_c = \Big(\sum_{ij}{}' Z_i Z_j \Big/ \sum_k (Z_k^2 + Z_k)\Big) e^{-\ell_{ij}^2 \pi^2 q^2/200} \Big[\sin(\pi q r_{ij}/10)\Big]/\pi q r_{ij}/10 \quad .$$

The details of the program are given in Appendix A. The purpose of the program is fourfold. First, M(q) is computed at integral q values over the entire experimental range, where

$$q = 10\ s/\pi$$

and

$$s = (4\pi/\lambda\ ) \sin (\theta/2) .$$

This is useful as an aid in drawing background functions through the experiment with various theoretical models. It also aids in the computation of radial distribution functions by providing $M(q)_c$ data to graft onto experimental data at small angle scattering, where experimental data are not available. The function $M(q)_c$ can be computed to any desired value of q so that it may be used to correct for integral termination effects. The difference $M(q)_c - M(q)$ is also tabulated so that it may be added to the experiment to correct the experimental intensity to pure nuclear scattering. As an added guide in determining the background function, the program makes available the function $\sum_k \{(Z_k - F_k(q))^2 + S_k(q)\}/q$ computed at integral values of q. It takes the I.B.M. 650 about sixty-eight seconds per term in the primed summation to compute one hundred values of each of the four quantities $M(q)$, $M(q)_c$, $M(q)_c - M(q)$, and $\sum_k \{(Z_k - F_k(q))^2 + S_k(q)\}/q$.

B. Radial Distribution Function

The expression

$$(52)\quad f(r) = (\pi^2/100)\int_0^{q_{max}} qM(q)_c e^{-aq^2} \sin (2\pi qr/20)dq$$

is programmed on the I.B.M. 650 to compute the radial distribution function f(r) at .05 angstrom or .025 angstrom intervals. Two programs have been coded with different aims of operation. The first evaluates f(r) at .05 angstrom intervals and then evaluates the function

$$(53)\quad B_N(q) = B_o(q)/1+(\pi^3/200)\int_o^{r_1} f(r)\ \sin\ (2\pi qr/20)dr$$

at integral values of q, where r_1 represents the distance from the origin to the beginning of the first molecular feature and $B_o(q)$ is the old background function. The $B_N(q)$ function is an improved background function for which the corresponding f(r) function has the absolute value of its spurious features reduced in the region zero to r_1. In practice, a smooth curve is drawn through the $B_N(q)$ points to obtain a new background. This procedure is often observed to improve other regions of the f(r) function as well as the region from zero to r_1. Since f(r) is a probability function, a background function must be chosen so that f(r) is nowhere negative. At the same time, however, the background function must be smooth so that false detail is not introduced into the f(r) function. The evaluation of the function $B_N(q)$ in this program is thus an aid in determining a final radial distribution function. Also, in this program experimental values of M(q) and $M(q)_c$ may be obtained if desired. Details of this program are contained in Appendix B.

The second radial distribution program, which evaluates f(r) at .025 angstrom intervals, is used to obtain an accurate representation of f(r) once the smoothness and positive area requirements are met. This f(r) function is used as input data for least squares analysis of individual radial distribution peaks. The program is discussed in Appendix C, and techniques for interpreting radial distribution functions are discussed in the next section.

C. Analysis of the Radial Distribution Function

Two basic problems in the analysis of the radial distribution function are the deduction of the molecular parameters and the estimation of the reliability of these parameters. The problem of obtaining molecular parameters can perhaps best be approached by application of a least squares method for fitting the part of the radial distribution function which can be characterized by a summation of Gaussian functions. The parameters obtained for the calculated function giving the best fit are approximately the $r_g(1)$ and l_{ij} parameters defined previously. If it is possible to make complete asymmetry corrections to a radial distribution peak then the parameters obtained from a least squares fit with a Gaussian peak can be directly related to the $r_g(1)$, $r_m(1)$ and l_{ij} parameters. In the case of internuclear distances involved in internal rotation, however, the radial distribution peaks re-

flecting the effect of this motion cannot be meaningfully characterized by Gaussian functions. Therefore, for a complete analysis of a radial distribution function with internal rotation special methods must be employed. When final molecular parameters are obtained some estimate of their reliability must be made. A detailed discussion of these problems is presented below.

1. Methods for obtaining parameters from the radial distribution function

There are several methods for the dissection of any part of a radial distribution function not involving internal rotation. Trial and error curve fitting, the method of moments, and various methods of least squares have been used (56). In this laboratory the method of steepest ascents for the solution of the least squares problem has been used successfully.

The basic problem is that of finding the maximum of the function (57)

$$Y(\theta_i) = (-1/2)\sum_r (f(r)_{exp} - f(r,\theta_i)_{calc})^2 .$$

Here $f(r)_{exp}$ is the value of the experimental radial distribution function and $f(r,\theta_i)_{calc}$ is the calculated value of the radial distribution function where the θ_i (i = 1, ...N) parameters include a scale factor k, internuclear parameters - r_{ij}, and root mean square amplitudes of vibration - ℓ_{ij} .

The summation is made at 0.025 angstrom intervals. If the function $Y(\theta_i)(i = 1, \ldots N)$ is thought of as determining a surface in an N + 1 dimensional space, the set of estimated parameters θ_i^o $(i = 1, \ldots N)$ determine a point on the surface on the contour line $Y(\theta_i) = Y(\theta_i^o)$. Another point on a different contour line is determined by the set of parameters

$$\theta_i^1 = \theta_i^o + \lambda_i^o t \quad (i = 1, \ldots N)$$

where the λ_i^o are the components of the gradient of $Y(\theta_i)$ $(i = 1, \ldots N)$ evaluated at the point θ_i^o, $(i = 1, \ldots N)$ and t is a parameter chosen so that it maximizes $Y(\theta_i^1)$. Geometrically this corresponds to a move in the N + 1 dimensional space from the point $\theta_i^o(i = 1, \ldots N)$ on the line $Y(\theta_i) = Y(\theta_i^o)$ in the direction of steepest ascent to the point $\theta^1(i = 1, \ldots N)$ on the line $Y(\theta_i) = Y(\theta_i^1)$. A similar process may be repeated to get to a point $\theta_i{}^2(i = 1, \ldots N)$ on the line $Y(\theta_i) = Y(\theta_i{}^2)$. Continued iteration will eventually lead to the extremum. Methods for arriving at suitable values of t for the refinement of the θ_i and a program for the application of the steepest ascent method to the dissection of $f(r)_{exp}$ functions using a digital computer are given in Appendix D.

In practice the method is applied assuming that $f(r)_{exp}$ can be approximated by a sum of Gaussian functions with weighting factors proportional to the product of the atomic numbers of the atoms involved in each internuclear

distance. It would be possible to use this same method to characterize $f(r)_{exp}$ peaks using a more general function in which a parameter to characterize the asymmetry is added. The method easily resolves peaks of roughly comparable areas that are about 0.2 angstroms apart. At the extremum the average deviation per point has been found to be about of the same magnitude as the observed noise level in the $f(r)_{exp}$ function. One drawback to the method is the fact that the parameters for peaks with small weighting factors are refined only very slowly in the presence of peaks with large weighting factors. Also the method as currently programmed takes between an hour and an hour and one-half on the I.B.M. 650 digital computer to resolve three Gaussian peaks characterized by thirty-six experimental points if initial estimates are used that are within ten per cent of the final values. The main advantage is that the parameters are obtained from the $f(r)_{exp}$ function by use of a completely objective method. The above approach is capable of giving precise molecular parameters as long as the Gaussian approximation is valid but a different approach must be used in dealing with sections of the f(r) function depending on internal rotation.

Two methods for handling the problem of internal rotation in electron diffraction work on molecules have generally been used (46, 58). The first method is applied in the case of a high potential barrier. In this method it is assumed that the

effects of the skeletal vibrations may be separated from the total probability distribution leaving a probability distribution the breadth of which is characteristic of the internal rotation. It is also assumed that the librational modes associated with the internal rotation are in their ground states.

The curvatures at the bottoms of the potential wells may then be determined from the amplitudes of motion due to the internal rotation. Thus if it is assumed that the complete potential function can be approximated by a superposition of cosine functions it is then possible to determine the amplitudes of the cosine functions, from the parabolic shape at the minimum of the potential wells. An estimate of the barriers hindering rotation in the molecules may now be made.

In the case of a low barrier it is assumed that the internuclear separations involved in the internal rotation are averaged over all the angles of rotation weighted by a factor of the form $e^{-V(\theta n)/kT}$. Here $V(\theta_n)$ is the potential function for the internal rotation. The same assumption is again made as to the separability of the vibrational and rotational contributions to the $f(r)_{exp}$ curve. The potential barrier to the internal rotation is determined by a comparison procedure. To do this a cosine potential approximation is made. Then by varying the parameters involved in the expression for $V(\theta_n)$ and computing the contributions to the M(s) function or the radial distribution function the most probable

parameters are determined by a comparison with the corresponding experimental functions.

Of the molecules considered in this study only in n-butane is it possible to determine the barrier to hindered rotation in an unambiguous manner. Thermodynamic information indicates that the energy difference between the trans and gauche forms is in the region where the low energy barrier approximation is valid. The best value reported for the energy difference is about 800 cal. mole^{-1} (4, 6). The carbon carbon internuclear distance in n-butane depending on the barrier to internal rotation can thus be expressed as a function of the angle of rotation weighted by a distribution function depending on the barrier shape and absolute temperature. This distribution function can be expressed as

$$P(\theta_n) = e^{-V(\theta_n)/kT} \Big/ \sum_{\theta_n=0^\circ}^{2\pi} e^{-V(\theta_n)kT} \Delta\theta_n$$

where $V(\theta_n)$ is the potential function approximated by the relation

$$V(\theta_n) = V_1 + V_2 - V_1 \cos \theta_n - V_2 \cos 3\theta_n ,$$

where k is the Boltzman constant and T is the absolute temperature. The quantities $2(V_1 + V_2)$ and $3V_1/2$ are the barrier to free rotation and the approximate energy difference between the trans and gauche forms respectively. The contribution to the M(s) curve from the terminal carbon carbon

internuclear distance depending on the internal rotation can be expressed as

$$M(s) = (2Z_c^2/\sum_R (Z_R^2+Z_R))e^{-\ell_{cc}^2 s^2/2} \sum_{\theta_n=0}^{2\pi} (\sin sr_{cc}(\theta_n)) \; P(\theta_n) \Delta \theta_n$$

where $r_{cc}(\theta_n) = \left[2r_1^2(1 + \cos \theta_n) + r_2^2\right]^{1/2}$

in which r_1 and r_2 are constants determined from the geometry of the molecule. By varying the parameters V_1 and V_2 of the potential function and using a reasonable value of kT it is possible to compute the contribution to the scattering curve from the atom pairs involved in internal rotation. Once this has been done a Fourier sine inversion of that part of the function involved in internal rotation will give the $f(r)_{exp}$ function due to internal rotation. Then by varying the parameters V_1 and V_2, the most probable values are determined by comparing the calculated M(s) functions and f(r) functions for the internal rotation with the corresponding experimental functions. Once the problem of obtaining molecular parameters has been solved it is of interest to determine the reliability of these parameters.

2. Treatment of errors

There are two basic sources of error in electron diffrac-

tion work. One source lies in errors in the theoretical approximations while the second source stems from errors in the experimental work. The main errors in the theoretical approximations are due to failures in the Born approximation (52). These are of two types, one affecting the results at large angle scattering and the other affecting the results at small angle scattering. The effect at large angle scattering is due to the differing atomic fields of atom pairs of differing atomic number. This effect enters into the expression for M(s) as a phase factor (51). Fortunately, this effect is negligible (43, 44) for atomic number differences of five, which are the largest differences encountered in the molecules treated here.

At small angles difficulties arise because of possible polarization of the atomic electrons by the incident beam and the use of X-ray atom form factors for spherical atoms. Of this first effect little is known (52) although agreement between theory and experiment would seem to indicate that such an effect must be small. It appears from experimental work performed in this laboratory that there are smooth deviations of the experimental intensities at small scattering angles from the intensities calculated with the use of X-ray atomic form factors for spherical atoms. If this effect is real, one possible interpretation might be that the observed deviations are due to the distortion of the electron charge cloud

due to chemical binding. It would be interesting to examine diatomic molecules such as nitrogen or oxygen to try to determine effective molecular form factors for bonded atoms by electron diffraction. The form factor effect, while influencing the shape of the atomic background, should cause appreciable error in the molecular scattering function only at small scattering angles.

The experimental errors fall into three classifications (35). First, there are the systematic errors in measurement of the scattering angle and errors in the settings and calibration of the electron diffraction unit. These errors mainly affect the determination of the internuclear distance parameters and not so much the vibrational amplitudes.

The second class of errors to be considered is the random errors in the determination of the intensity as a function of angle. These errors affect both internuclear distance parameters and vibrational amplitudes and show up as noise in the $f(r)_{exp}$ function. The errors in the parameters are roughly given by the relation (35)

$$\delta r_{ij} \sim \delta l_{ij} \sim (y/A)(2b + l_{ij}^2)^{1/2}$$

where y is the maximum amplitude of the noise in the neighborhood of an $f(r)_{exp}$ peak whose height is A.

The third classification contains the errors which appear systematically in the intensity measurements, including differ-

ing indices of resolution. One possible source of systematic error in the determination of the intensity would lie in the uncertainty of the determination of the factors relating the optical density measured from the photographic plates to the intensity. The index of resolution is the ratio of the experimental molecular scattering intensity to the calculated theoretical molecular scattering intensity.

The errors in the third classification do not affect the positions of the nodes of the M(s) function appreciably but do affect the amplitude of the M(s) function. The parameter r_{ij}, since it may be determined from the positions of the nodes, is not affected while the ℓ_{ij} values are since the envelope of the M(s) function determines the value of the ℓ_{ij} parameter. Another possible source of error of this same general type arises from extraneous scattering. The error introduced by this effect is small since the extraneous scattering generally runs from zero at small scattering angles to perhaps 6 per cent of the total scattering intensity at large angles in a smooth fashion.

For the molecules studied here the errors in r_{ij} were generally in the range $\delta r_{ij} = 0.004$ angstroms to $\delta r_{ij} = 0.015$ angstroms. Similar uncertainties were computed for the ℓ_{ij} values also. Although this discussion of error is mainly slanted toward the determination of parameter errors from the $f(r)_{exp}$ function independent estimates of the error may be made from the M(s) function.

IV. EXPERIMENTAL METHODS

A. Equipment and Procedure

The experimental work discussed here was done at the University of Michigan using the sector electron diffraction unit constructed by Brockway and Bartell (38). A discussion of the apparatus and the experimental techniques involved in obtaining scattering intensities is given below.

In this unit the incident electron beam obtained from an electron gun with an accelerating potential of forty thousand volts is allowed to intersect a narrow beam of molecules injected into the main chamber of the diffraction unit. The accelerating potential is regulated to $\pm$ 4 volts and measured by using a potentiometer in conjunction with a voltage divider. A beam of molecules is formed by allowing the gaseous specimen in a sample bulb to expand into the diffraction chamber through a platinum nozzle the throat diameter of which is about 0.4 millimeters. Pressures of the gas in the sample bulb ranged from 45 to 500 millimeters depending on the molecular weight of the gas species being studied. In order to obtain intensities over a large range of the scattering variable exposures were taken at two nozzle to photographic plate distances. Long distance exposures were made at a plate distance of 25 centimeters and covered a range in s from approximately s = 3 to s = 14. Short distance ex-

posures were made at a plate distance of 10 centimeters and covered a range in s from approximately s = 7 to s = 33. The distance between the nozzle and plate was measured with a high precision cathetometer.

Exposure time was governed by an electronic timer coupled with an electrostatic shutter. A cam on the stopcock connecting the nozzle to the sample bulb operated the shutter. In this work exposure times of 0.1 to 0.4 seconds and a beam current of 0.3 of a micro ampere were used.

To record the scattered intensity Kodak medium lantern slides 3-1/4" x 4" were used. In order to compensate for the rapid fall off of the total intensity a rotating sector was used. The sector is a cardioid-shaped thin metal plate spun at 1500 rpm parallel to and 8 millimeters above the photographic plate. By use of the rotating sector the exposure time at s = 30 was made approximately 1000 times larger than the exposure time at s = 3. It was then possible to obtain accurate microphotometer recordings of the intensity. To measure the intensity the photographic plates were mounted on a special stage (59) and rotated (22) at 600 rpm while the plate was being scanned by a Leeds and Northrup recording microphotometer. In this way possible errors in the photographic emulsion were averaged out while the record of the intensity was made on special graph paper which was graduated directly in optical density units. The intensity was deter-

mined by measuring at close radial intervals the optical densities of four photographic plates for each distance. Since both the left and right hand sides of the microphotometer traces scanning the full diameter of each plate were measured, eight optical density readings in all were averaged at each particular value of s.

In order to convert the optical density readings to intensities a procedure developed by Bartell and Brockway was used. This procedure is based on two assumptions. The reciprocity relation

$$\text{Exposure} = \text{Intensity} \times \text{Time}$$

was assumed to be valid for the experimental conditions encountered in this work. It was also assumed that the exposures from two sets of electron diffraction patterns taken under identical conditions with the same gas and differing only in total exposure time must be the same function of the plate radius except for a multiplicative constant. Then according to the procedure of Bartell and Brockway a plot of $D_1(r)/D_2(r)$ vs. $D_1^{(r)}$, where $D_1(r)$ and $D_2(r)$ were the optical densities of the two exposures, usually gives a linear curve over the optical density range from 0.3 to 0.9. The slope of this linear relation then determines an emulsion calibration constant α such that

$$I = D_{Av} + \alpha D_{Av}^2 .$$

The α 's found in this work were in the range 0.13 to 0.16. It was found that the intensities determined according to the procedure outlined above were smooth to 3 to 4 parts per ten thousand of the measured total intensity.

B. Experimental Intensity Functions

As we have shown previously in the section on the theory of electron diffraction the total observed intensity is given by

$$I_T = (I_m + I_a)$$

where I_m is the intensity due to the molecular scattering and I_a is the intensity due to the total atomic scattering. Actually in practice there exists a small amount of extraneous scattering so that the measured intensity is

$$I_\gamma = \gamma(I_m + I_a + I_e)$$

where I_e is the intensity due to the extraneous scattering and γ is a modification function due to the sector. Since the sector is not perfect a sector correction or function which corrects I_γ for the errors in the sector must be applied. It is also convenient to multiply the corrected intensity by a plotting function so that the resultant intensity function may be more easily graphed. For short distance data the plotting function

$$P_F = (a + r - br^2)$$

was used where r is the radial distance on the photographic plate and a and b are arbitrary constants. The long distance data for n-butane and n-heptane were modified by dividing through by the function

$$(s-\pi/2)^{-1}\sum_k \left\{(Z_k-F_k(s-\pi/2))^2 + S_k(s-\pi/2)\right\}$$

while the long distance data for ethylene and isobutylene were not modified.

The sector correction used for isobutylene was determined by reading with a traveling microscope the radial distance from the center of the sector as a function of angular opening. The function tabulated was kr^3/ϕ_r as a function of r where ϕ_r was the measured angle and r was the radial distance. For the molecules ethylene, n-butane and n-heptane an improved sector correction was used. This sector correction was obtained by dividing experimental intensity data for argon by theoretical intensity data obtained from Hartree Fock calculations. This procedure had the advantage that imperfections in the sector were indicated in greater detail. However, because the Hartree Fock results emphasize the electronic shell structure in argon to a greater degree than does the experiment certain smooth features were introduced into the sector correction that were characteristic of the differences between the theoretical and experimental results. To compensate for this effect the background curves had to be

made to reflect this difference. Comparison of the background curve of isobutylene with those of ethylene, n-butane and n-heptane illustrates this point.

In general the final plotted intensity through which the background function was drawn was

$$I_r = I_\gamma \times (\text{Sector correction}) \times (\text{Plotting function}) .$$

It should be noted that I_r is the same intensity function discussed under Appendix B.

V. THE STRUCTURES OF ETHYLENE, ISOBUTYLENE, n-BUTANE AND n-HEPTANE

A. Introduction

The structure of ethylene has previously been examined by both infrared (59, 61) and electron diffraction methods (1). The electron diffraction results were obtained twenty years ago and were in agreement with the theoretical ideas of the time with regard to bond angles in hydrocarbons (1). According to these ideas the angle between two hydrogen atoms both bonded to a carbon atom should be approximately the tetrahedral angle independent of the other functional groups on the carbon atom. Since then a more recent theory of directed valence has been developed (62) which suggests that the angles between bonds in ethylene should be 120°. The most recent spectroscopic work (59) supports the 120° picture. According to these ideas, the analogous bond angles in isobutylene should also be 120° except for small distortions due to steric repulsion. However, early work (63) on isobutylene indicated that the angle between the two methyl groups was nearly the tetrahedral angle. Since spectroscopic work on this molecule has not been done it was felt that a modern electron diffraction determination of the structure should be attempted to check the earlier results in light of their disagreement with present ideas.

In the course of investigating isobutylene it was found that the experimental C-C double bond distance was in disagreement with the corresponding distance reported in the spectroscopic study of ethylene (59, 63). Also, the angle between the methyl groups appeared to be closer to the tetrahedral angle (63) than the 120° angle predicted by the current directed valence picture. At this point it was decided that a reinvestigation of ethylene should be attempted to see whether the differences between the two molecules were due to chemical effects or to inaccuracies in the structure determination for ethylene.

It is important that the structure of ethylene be known accurately since ethylene is taken as the prototype of non-aromatic organic compounds with double bonds, and the 120° bond angle has generally been considered to be a consequence of "pure" sp^2 sigma bonding. Since, in chemistry, graphs of bond orders against internuclear distances are used to correlate structural data it is important that accurate parameter measurements be available for the compounds used to determine points on these graphs.

Another reason for interest in the structure of isobutylene is the study of the effect on a C-C single bond of an adjacent double bond. Many investigators (62) have established that a C-C single bond adjacent to a triple bond is substantially shorter than the normal single bond distance of

1.54 angstroms. On this basis it has seemed anomalous that the reported value for the single bond distance in isobutylene and several other olefins is identical to the "normal" single bond distance (63).

The structures of the normal hydrocarbons in the gaseous state were first investigated by Wierl (65, 66) who examined the series from n-propane to n-hexane. Wierl reported that very little difference existed in the diffraction patterns and was unable to determine the overall shapes of the molecules although he reported values for the C-C single bond distance for all the molecules studied. These hydrocarbons have been studied spectroscopically by both the infrared and raman techniques (67) in the solid and liquid states. The longer chain normal hydrocarbons have also been studied in the solid state by X-rays (68, 69). Spectroscopic results indicate that the normal hydrocarbons are quite rigidly held in the extended trans configuration in the solid state but that there exist some gauche forms in the liquid state (67). The main structural problem of the normal hydrocarbons in the gaseous state besides determining the parameters for the bonded distances is the determination of the distributions between internal rotational configurations of the molecule. A great deal of work has already been done on this last problem.

Gaseous viscosity measurements on n-heptane and n-butane

by Malaven and Mack (70) and later by McCoulrey, et al. (2) were interpreted as indicating that the molecules are largely in a "highly coiled" or "crumpled" state. However, theoretical work by Stein (3) on molecular polarizabilities using experimental values determined from light scattering experiments indicated that these molecules were mainly in extended trans states.

Two statistical thermodynamic approaches have been made (4, 7) to this problem using heat capacity data and different starting assumptions. In the earlier work by Aston (7) a potential function of the form $V_o \cos \theta$ was assumed to represent the rotation about the central C-C axis in n-butane. A more rigorous approach was used by Pitzer (4-6) who assumed a two parameter potential function and computed not only the total barrier to rotation about the central C-C axis in n-butane but also the energy difference between the trans and gauche forms. Since current electron diffraction methods are one of the most direct methods of structure determination it was felt that an investigation of n-butane and n-heptane might resolve the controversy about the shape of free hydrocarbon molecules.

B. Ethylene

A sample of ethylene was obtained from the Matheson Company with less than 1/2 per cent impurities. Diffraction patterns for a large number of plates were obtained at two

different plate distances and three long distance plates and four short distance plates were selected for microphotometering. The resultant I_r function was computed from the experimental data by use of the I.B.M. 604 digital computer. The I.B.M. 650 computer was used to determine the radial distribution function and the background function.

The peaks in the resultant radial distribution curve were reasonably well resolved and the C-H and C-C bonded distances were both determined by graphical and steepest ascents techniques. It is interesting to note that the H-H peaks at 1.8 angstroms, 2.5 angstroms and 3.1 angstroms show up on the radial distribution curve and plainly indicate the planar model for ethylene. Once a planar structure is assumed the ethylene structure is determined by three parameters and these are obtained from the three main features of the f(r) function. The first two peaks may be corrected for asymmetry effects but not enough is known about the C-H non-bonded peak to characterize it in this way. The internuclear distance determined by a Gaussian function which best fits an f(r) peak seems to be closely related to $r_g(1)$ since $r_g(1)$ is also the center of gravity of the radial distribution peak. Thus it appears from the standpoint of electron diffraction that if asymmetries are unknown it is best to define internuclear distances in terms of $r_g(1)$ or $r_g(0)$. The position of the maximum of a radial distribution peak where the cubic asymme-

Figure 1. A plot of the experimental I_r and B(s) functions for ethylene

(The heavier lines indicate the B(s) functions)

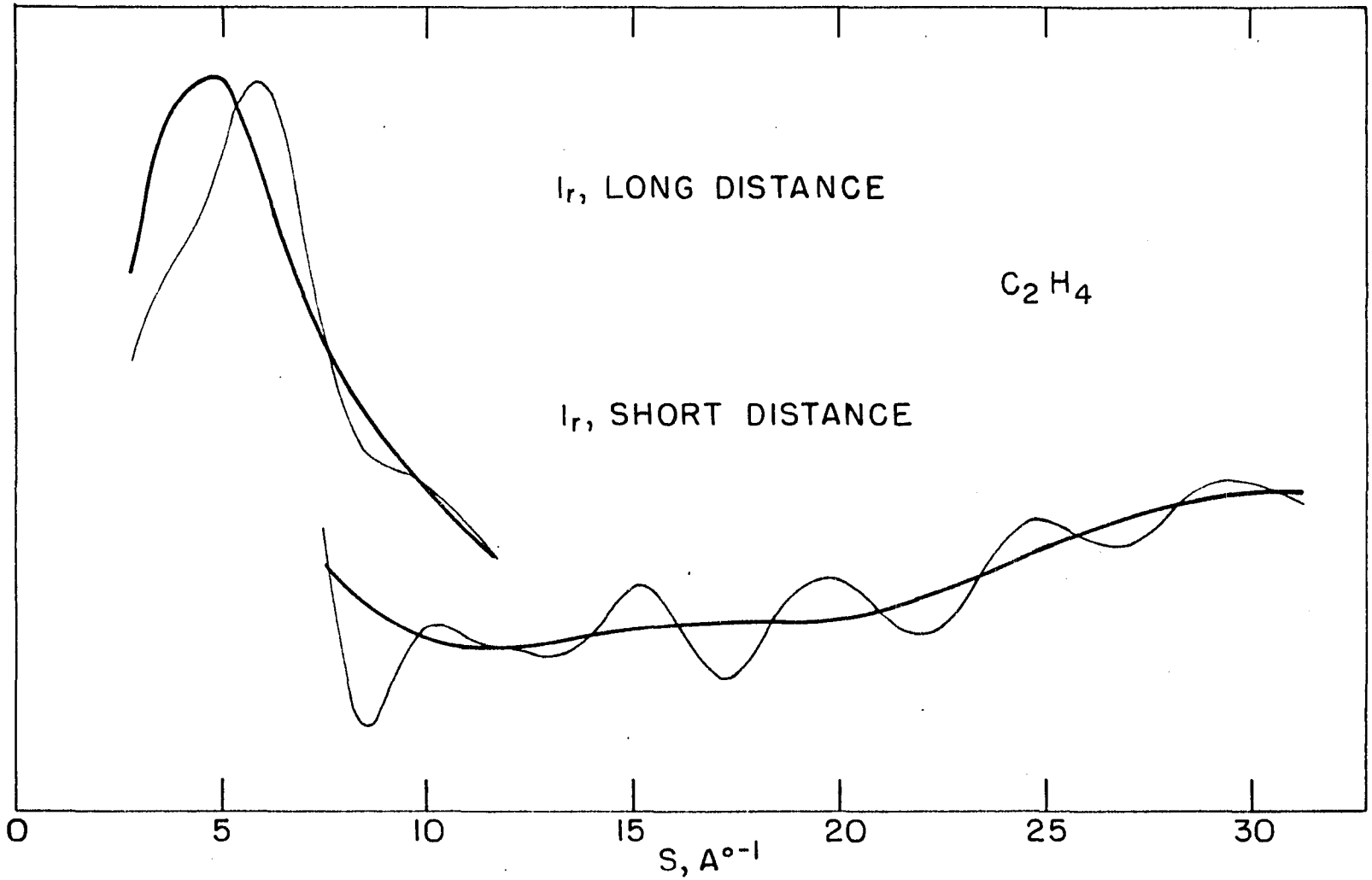

I_r, LONG DISTANCE
C_2H_4
I_r, SHORT DISTANCE
0
5
10
15
20
25
30
S, A°-1

Figure 2. Plot of the experimental M(s) function and calculated M(s) functions

(The H-C-H angle is the only parameter that has been varied and the distance parameters used were $r_{C=C}$ = 1.334 and r_{C-H} = 1.085.)

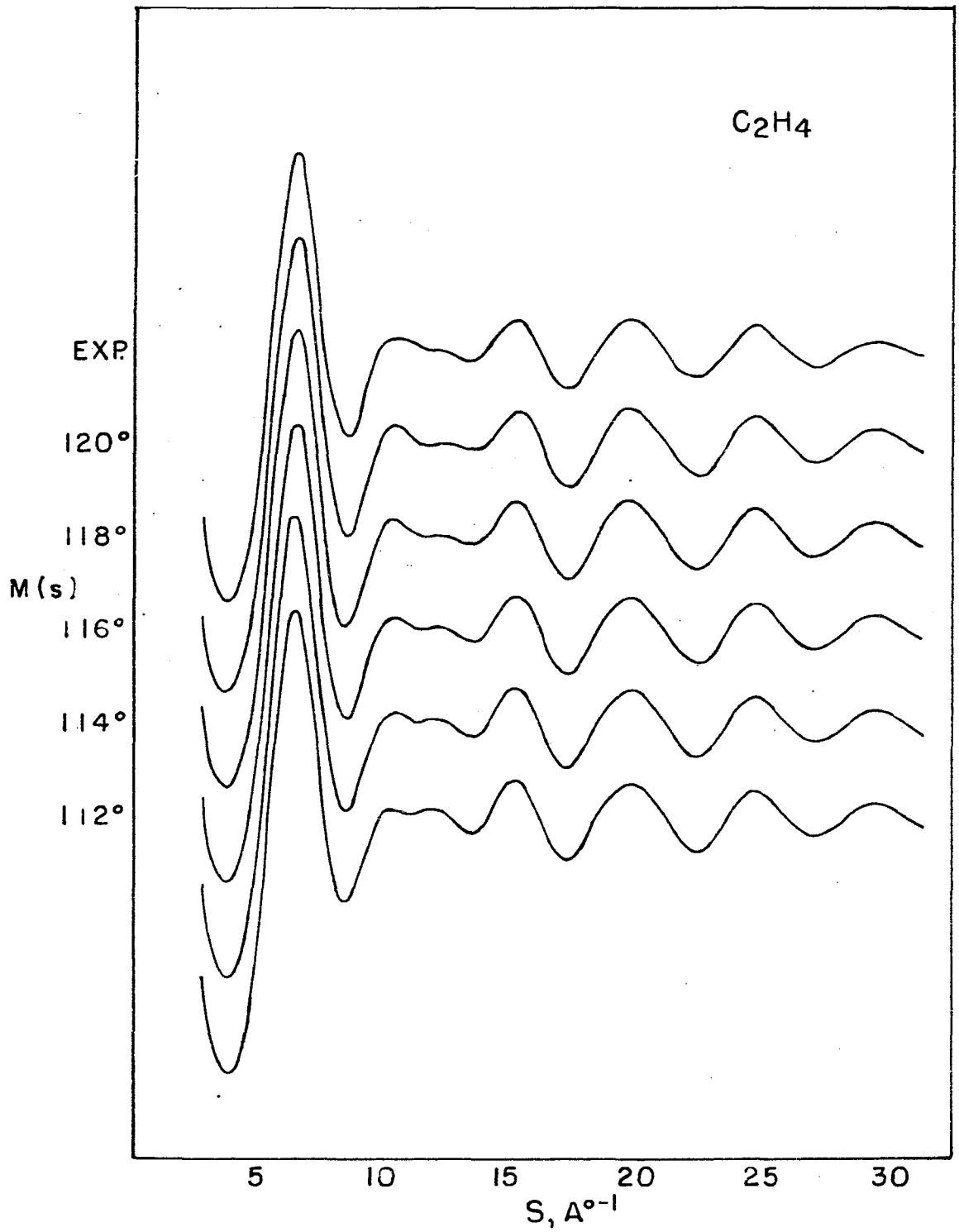

C2H4
EXP.
120°
118°
M(s)
116°
114°
112°
5
10
15
20
25
30
S, A°-1

Figure 3. Plot of the experimental M(s) function and the best calculated M(s) and $M(s)_c$ functions

($M(s)_c$ function differs from M(s) by the correction for non-nuclear scattering.)

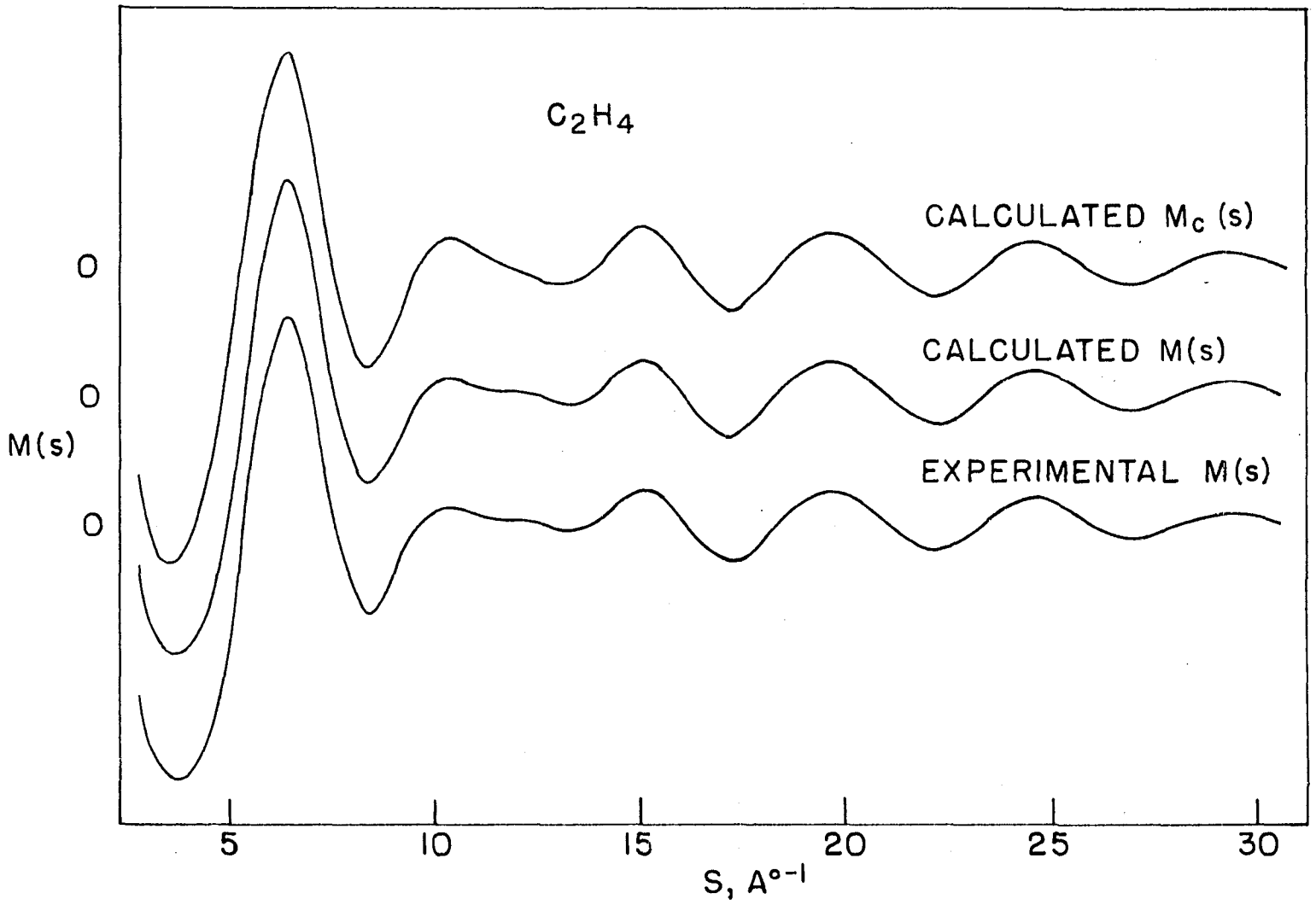
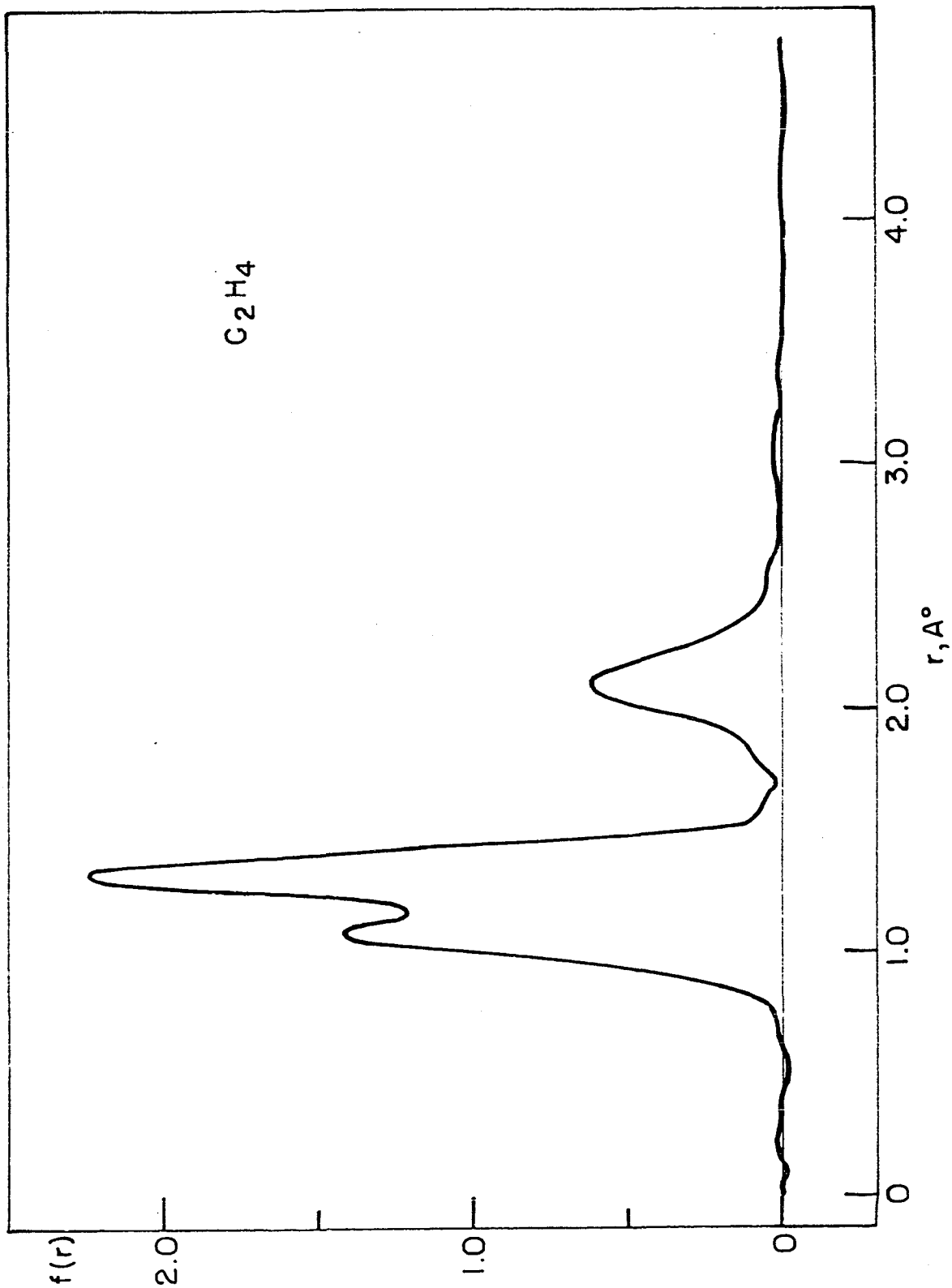

C_2H_4
CALCULATED M_c (s)
CALCULATED M(s)
EXPERIMENTAL M(s)
M(s)
0
0
0
5
10
15
20
25
30
S, A°-1

Figure 4. The radial distribution function for ethylene


C_2H_4
f(r)
2.0
1.0
0
0
1.0
2.0
3.0
4.0
r, A°

try corrections have been subtracted out is related to $r_g(1)$ approximately by the expression

$$r_{RD\ max} = r_g(1) - a\ell_{ij}^4/(4b + 2\ell_{ij}^2)$$

where a is the asymmetry constant, b is the damping factor and ℓ_{ij} is the root mean square amplitude of vibration. Table 1 shows the various values obtained for the ethylene parameters.

Table 1. Internuclear distance parameters for ethylene

Peak	$r_{RD\ max}$[a]	$r_g(1)$	$r_m(1)$	$r_g(0)$	$r_m(0)$
C-H	1.0763	1.0795	1.0740	1.0849	1.0794
C=C	1.3319	1.3325	1.3304	1.3339	1.3318
C---H	2.1059	2.1140		2.1190	

[a]Values were determined by the graphical method

It is of interest to note the spread in internuclear distances due to differing operational definitions. This spread would make it seem desirable in molecular structure work to standardize the nomenclature so that possible ambiguities in reporting internuclear distances would not arise.

The angle in ethylene was also obtained by a correlation method using the M(s) function where values of the ratio of $s_{expt.}$ to $s_{calc.}$ for the extrema of all the molecular features

indicate that a good fit was obtained. This is of course not an independent determination of the angle but merely a different way of looking at the same data. Table 2 shows the results of this correlation procedure.

Table 2. Correlation of maximum and minimum positions between the experimental and calculated M(s) curves

Feature	s_{exp}	s_{calc}	s_{exp}/s_{calc}
1	3.63	3.64	0.997
2	6.24	6.25	0.998
3	8.22	8.19	1.003
4	10.06	10.07	0.999
5	12.94	12.95	0.999
6	14.68	14.64	1.002
7	16.69	16.65	1.002
8	19.04	19.04	1.000
9	21.52	21.53	0.999
10	23.90	23.82	1.003
11	26.16	26.15	1.000
12	28.65	28.48	1.005
Average			1.0005
Average deviation			0.0020
Standard deviation			0.0023

Morino and Hirota have introduced the so-called reliability factor (46)

$$R = \sum_s W(s)\ s\ \left|\ M(s)_{exp}/J - M(s)_{calc}\ \right|\ /\sum_s W(s)\ s\ \left|\ M(s)_{calc}\ \right|$$

where $W(s)$ is a suitable weighting function and J is the index

of resolution. The factor R is a measure of the reliability of the structure determination and Morino reports values of 30 to 40 per cent for R with uncertainties in his reported parameters of about 0.01 angstroms. For ethylene values of better than 5 per cent were obtained for R.

A table summarizing the final molecular parameters for ethylene along with a comparison of experimental and calculated root mean square amplitudes of vibration and the estimated uncertainties in the experimental parameters is presented below.

Table 3. Molecular parameters and their uncertainties for ethylene

Peak	$r_g(1)$[a]	$r_g(0)$	δr	$l_{ij\ calc}$[b]	$l_{ij\ exp}$	δl_{ij}
C-H	1.080	1.085	$\pm$0.005	0.077	0.080	$\pm$0.006
C=C	1.332	1.334	$\pm$0.003	0.041	0.046	$\pm$0.005
C-..H	2.114	2.119	$\pm$0.008	0.091	0.102	$\pm$0.010
	$\angle$ H-C-H			115.9 $\pm$ 2°		
	Index of resolution			.76[c]		

[a]All values obtained by the graphical procedure

[b]Calculated from infrared spectroscopic results. See references (41) and (42)

[c]Index of resolution is the ratio of the amplitudes of the experimental M(s) and the calculated M(s) for the best theoretical model.

Although the parameters presented in the above table were obtained by a graphical method using the upper two-thirds of each peak they were also checked by the method of steepest ascents using the entire radial distribution function. The two methods agreed for the internuclear separations to better than 0.001 angstroms and agreed to about 0.002 angstroms for the root mean square amplitudes of vibration. It was found on dissection of the peaks that the C-H and C=C peaks could be characterized quite well by Gaussian functions after suitable asymmetry corrections were made. An integral termination correction was also made which affected the l_{ij} value for the C=C distance and the internuclear distance for the C-H peak. The observed noise level appeared to be about one per cent of the maximum of the experimental f(r) function. The peak for the C---H non-bonded distance, however, was appreciably skewed. Although an empirical asymmetry constant could be obtained for this peak, no theoretical framework exists which would permit corrections for this asymmetry to be made in the calculation of the H-C-H angle.

C. Isobutylene

A sample of isobutylene with less than 1/2 per cent impurities was obtained from the Matheson Company. A radial distribution curve and background function were obtained by the methods outlined earlier. It should be pointed out that

Figure 5. A plot of the experimental I_r and B(s) functions for isobutylene

(The heavier lines indicate the B(s) functions.)

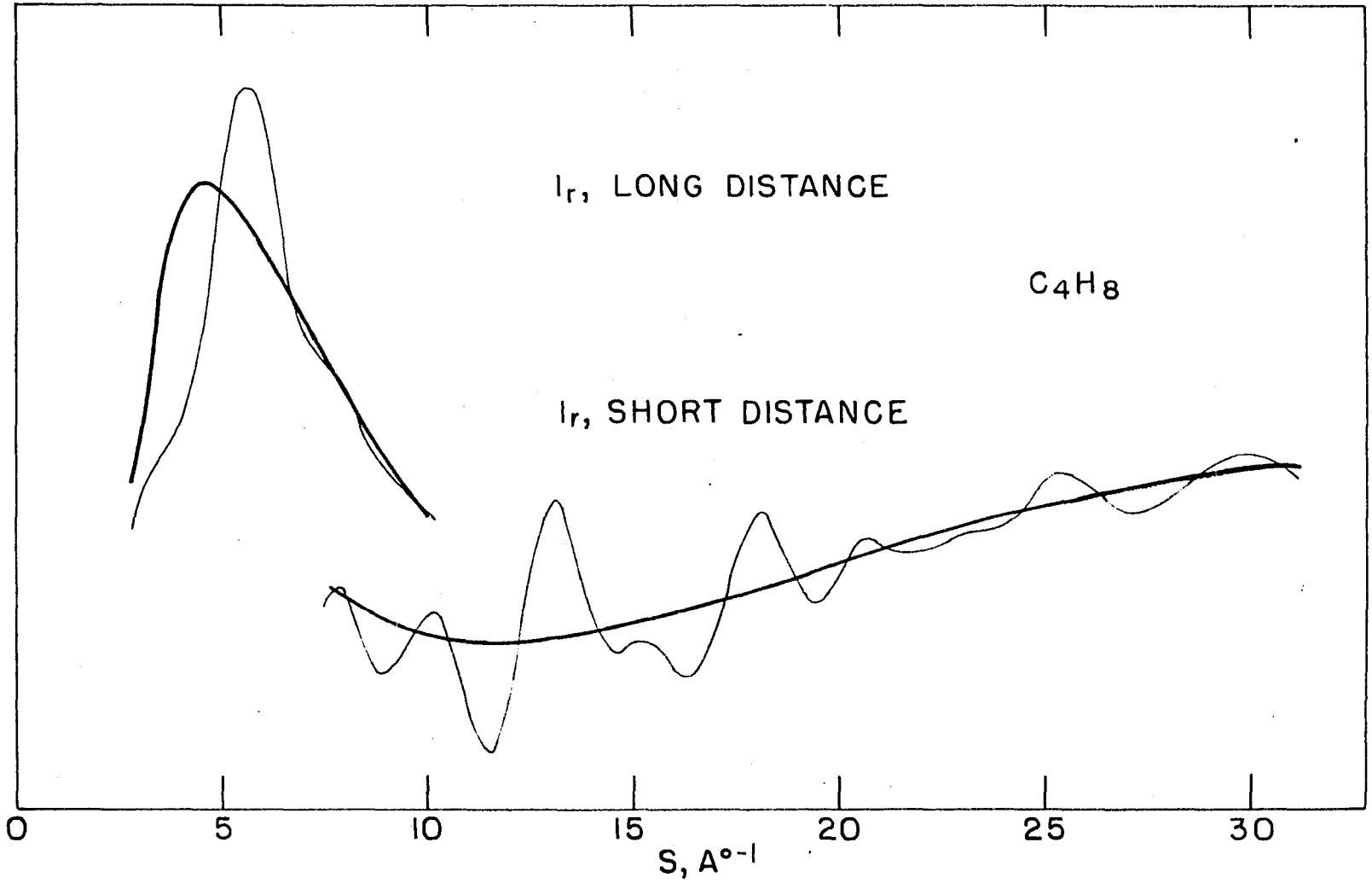
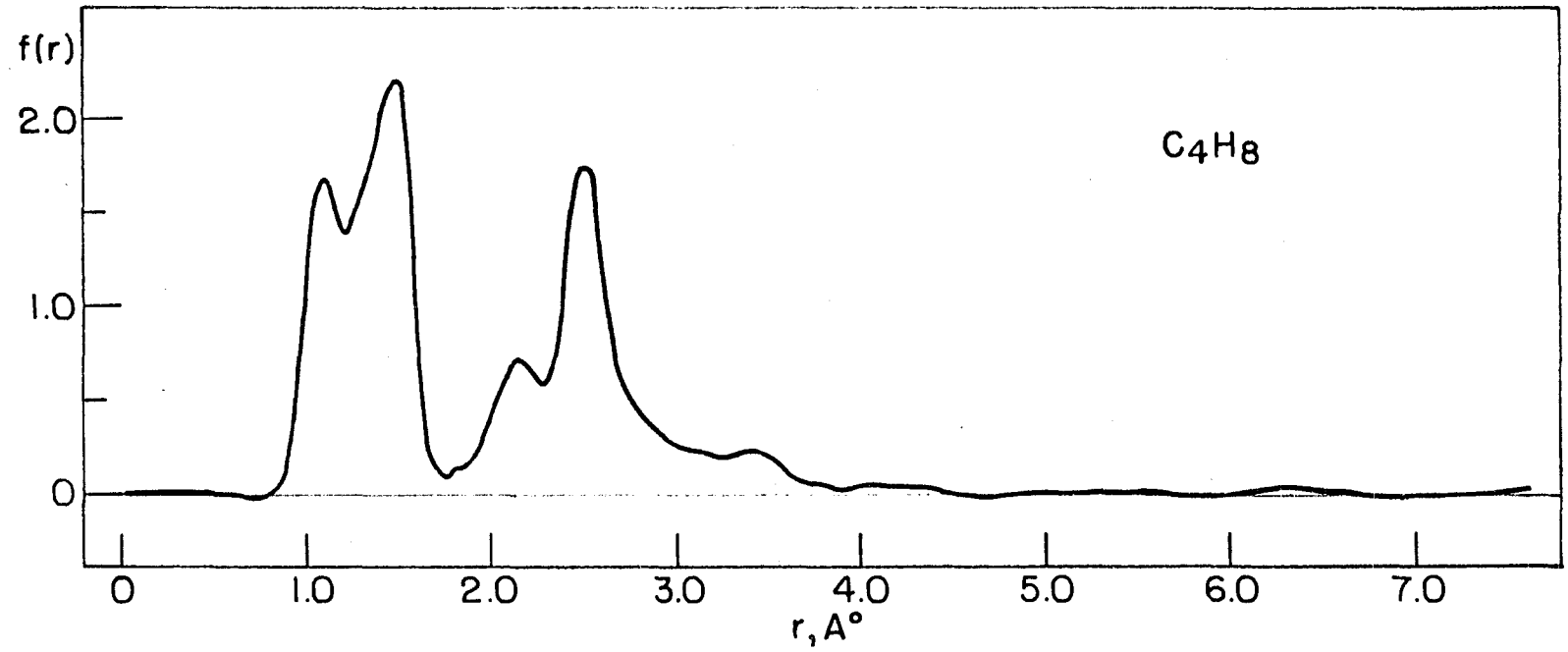
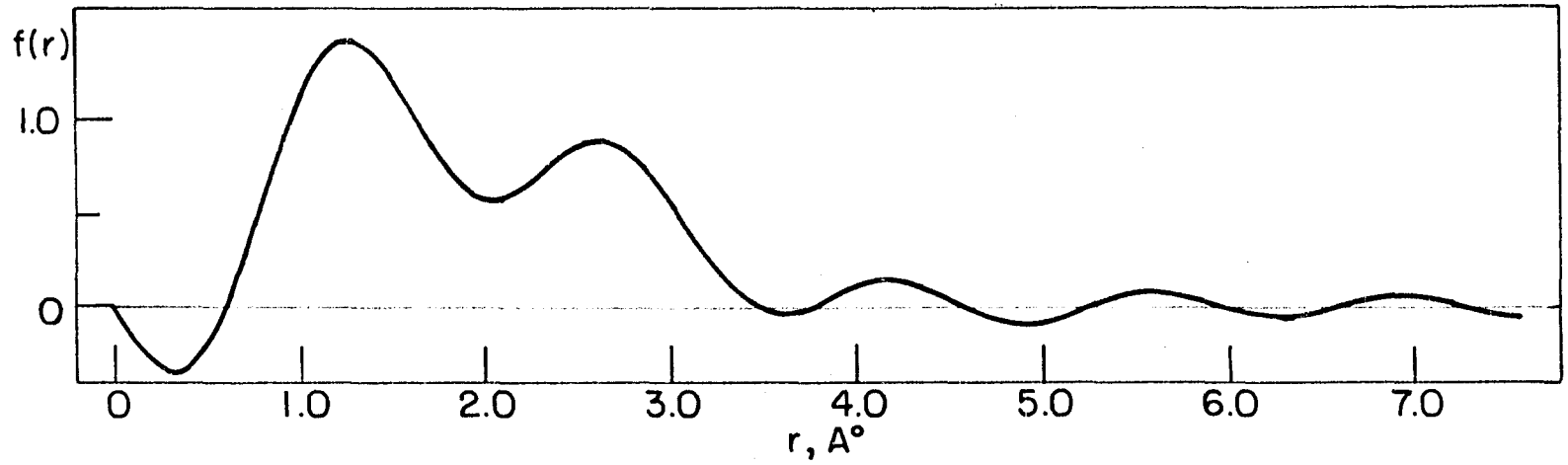

I_r, LONG DISTANCE
C_4H_8
I_r, SHORT DISTANCE
0
5
10
15
20
25
30
S, A°-1

Figure 6. Radial distribution functions for isobutylene

(a) The radial distribution function for isobutylene

(b) The radial distribution function using $M(s)_c$ data from $s = 0$ to $s = 5$


f(r)
2.0
1.0
0
C_4H_8
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°


f(r)
1.0
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°

the old sector correction was used for this molecule since the new sector correction had not been developed until the analysis of isobutylene was almost finished.

At least nine internuclear distance parameters are needed to determine the structure of isobutylene if the C-C skeleton is assumed to have the point group symmetry, C_{2v}, both methyl groups are assumed to be equivalent with equal C-H distances and H-H distances, and the ethylenic C-H distances are assumed to be equivalent. Two of these parameters determine the configurations of the methyl groups. The remaining seven parameters may be reduced to five if the CH_2 group is assumed to be the same as the CH_2 group in ethylene. Unfortunately a careful investigation of possible theoretical models for isobutylene indicates that within the limits of experimental error it would be impossible to determine the difference between various configurations of the methyl groups. It has been reported (64) that the barrier to rotation of the methyl groups is 2700 cal. $mole^{-1}$ (7). This would seem to favor a fairly rigid configuration of the methyl groups. There are only two sterically compatible configurations; in the first the molecule has the point group symmetry C_{2v} and in the second C_s. The configuration of each methyl group with respect to the double bond in the case possessing C_{2v} symmetry is in the same configuration as the one methyl group in acetaldehyde (58). For these two most probable configurations

M(s) curves were calculated but it was impossible on the basis of correlation with the experiment to choose between them.

The remaining five parameters were obtained from analysis of the f(r) function although the angle between the methyl groups could not be determined with the normal precision because of the unfavorable geometry of the molecule. The non-bonded C-C distances form an equilateral triangle approximately so that the radial distribution function exhibits a composite peak formed from the two different C-C non-bonded distances. As long as the separation between two Gaussian f(r) component peaks is less than the vibrational amplitudes of the peaks the sum of the two is again almost exactly a Gaussian peak. Because of this, small changes in the C-C non-bonded distances in either direction from the equilateral configuration do not appreciably affect the shape of the f(r) or M(s) functions. Therefore, the precision with which the Me-C-Me angle may be determined is severely limited.

In Table 4 a list of the parameters which are independent of the internal rotation for isobutylene is presented with a comparison of calculated and experimental amplitudes of vibration and estimated uncertainties for the parameters.

The uncertainties in the determination of the C-C non-bonded distances and their corresponding amplitudes of vibration were estimated from the sensitivity of the correlation procedure. The other parameters were all obtained from an

Table 4. Molecular parameters for isobutylene

Peak	$r_g(1)$	$r_g(0)$	δr	$l_{ij\ calc}$ [a]	$l_{ij\ exp}$	δl_{ij}
C-H_{ave}	1.102	1.107	±0.007	0.077	0.075	±0.009
C-H_{Et}		(1.085)[b]				
C-H_{me}		1.115				
C=C	1.324	1.326	±0.006	0.041	0.045	±0.006
C-C	1.499	1.501	±0.006	0.055	0.058	±0.007
C---H_{ave}	2.149	2.154	±0.020		0.100	±0.015
C---H_{Et}		(2.119)[b]				
H---H_{ave}	1.844	1.849	±0.045	0.123	0.100	±0.055
H_{me}---H_{me}		1.813[c]				
M_e---M_e		2.494[d]	±0.040	0.079	0.080	±0.020
CH_2---M_e		2.494[d]	±0.020			
	∠ Me-C-Me			112.5 ± 4.0°		
	∠ C-C-H[e]			109.5 ± 1.4°		
	∠ H-C-H[e]			109.0 ± 2.8°		
	Index of resolution			.87		

[a] See reference (41) and (42)

[b] Values assumed from ethylene structure

[c] Geometrically most compatible value

[d] Determined by correlation procedure with M(s) function

[e] Angles refer to the methyl hydrogens

analysis of the radial distribution function by the method of steepest ascents. The estimates of error were based on the procedure outlined previously but additional allowances were made for the effect of integral termination errors. A radial distribution curve of the somewhat arbitrary part of the $M(s)_c$ function from s = 0 to s = 5 was obtained so that the sensitivity of the $f(r)_{exp}$ function to this data could be determined.

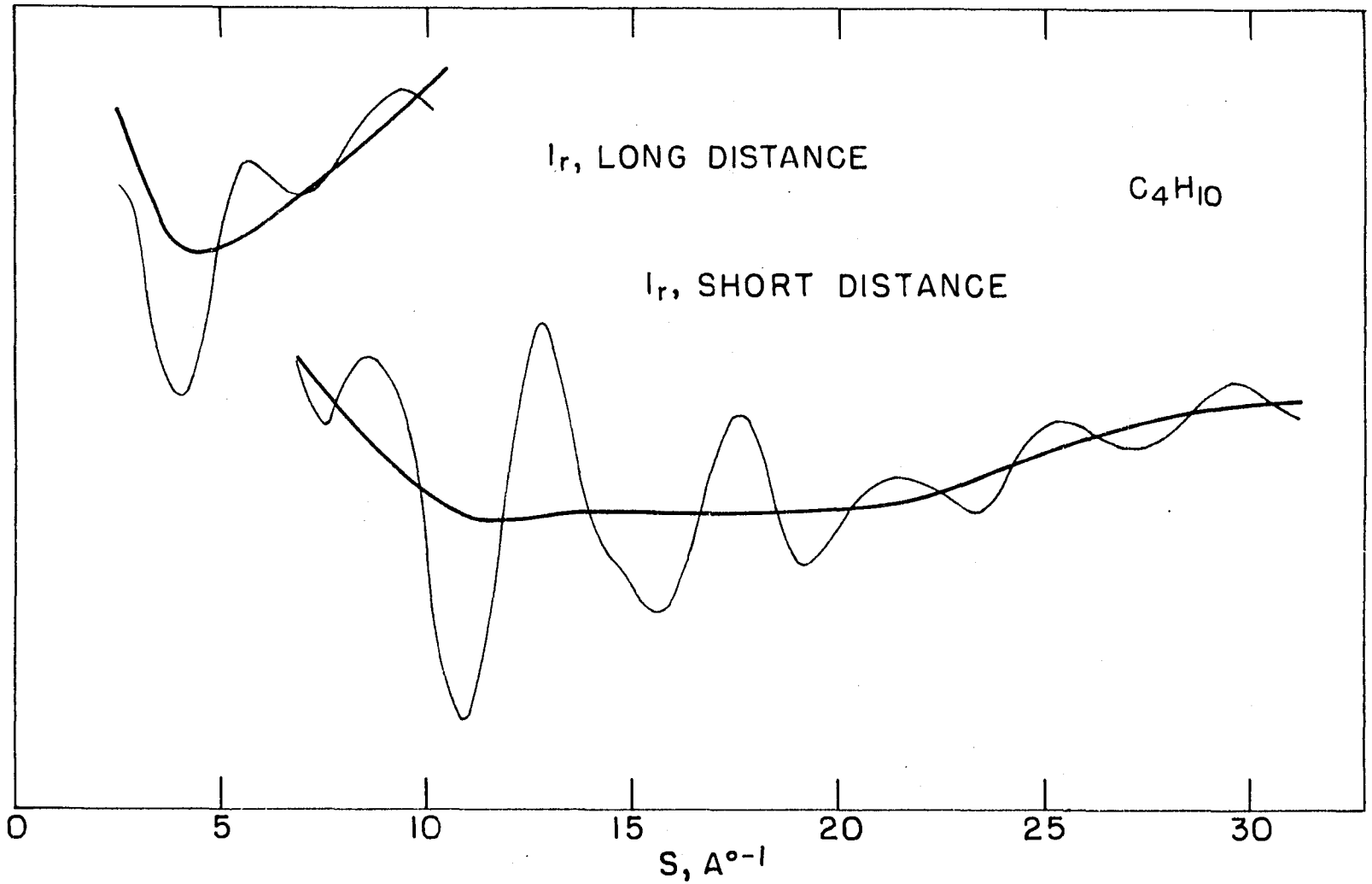
D. n-Butane

A sample of n-butane with less than 1/2 per cent impurities was obtained from the matheson Company and a radial distribution curve was computed by following the procedure outlined previously. It was assumed that all methyl and all methylene groups were equivalent and that the hydrogen atoms were related by three-fold and mirror symmetry operations respectively. Once these assumptions are made along with the assumption that all C-C bonded distances and C-C-C angles are equal it is possible to determine the structure of n-butane by determining nine parameters. Of these nine parameters, six determine the geometry of the framework independent of the internal motion and three determine the internal motions of the methyl groups and the carbon framework.

To simplify the problem further, it is assumed that the C-H bonded distances are the same throughout the molecule and

Figure 7. A plot of the experimental I_r and B(s) functions for n-butane

(The heavier lines indicate the B(s) functions.)


I_r, LONG DISTANCE
C_4H_{10}
I_r, SHORT DISTANCE
0
5
10
15
20
25
30
S, A°-1

that all the C-C-H angles are the same. Actually, the distances obtained from n-butane and n-heptane for the average carbon hydrogen bonded distances differ by 0.01 angstroms. This difference is so close to the estimated errors in the measurements of the bond distances that no interpretation of this shift in terms of differing carbon hydrogen distances for CH_3 and CH_2 groups is made. If proposed studies of n-pentane and n-hexane show a consistent trend, more significance might be attached to the indicated difference. The parameters determining the geometry of the framework independent of the internal motion are presented in a table below with estimated errors. It should be noted that it was not possible to determine the smallest carbon carbon nonbonded internuclear distance directly. Since the neighborhood of the feature determining this distance is confused by peaks depending on internal rotation, the position of the nonbonded carbon carbon peak was obtained from geometrical considerations using the position of the peak determining the trans configuration.

The parameters determining the internal motion of the methyl groups appear to be impossible to determine since their effect on the radial distribution curve is obscured by the effect of the internal motion of the carbon skeleton. The peak reflecting the internal rotation of the carbon framework about the trans position is fairly well resolved. If it is

Table 5. Molecular parameters and uncertainties for n-butane

Peak	$r_g(1)$	$r_g(0)$	δr	$l_{ij\ calc}$[a]	$l_{ij\ exp}$	δl_{ij}
$C\text{-}H^{b}_{ave}$	1.096	1.103	$\pm$0.007	0.077	0.087	$\pm$0.012
$C\text{-}C^{b}$	1.529	1.531	$\pm$0.006	0.055	0.057	$\pm$0.006
$C\text{---}H^{b}$	2.179	2.187	$\pm$0.015		0.150	$\pm$0.020
$C\text{---}C^{c}$	2.540	2.546	$\pm$0.020			
$C\text{---}C^{d}_{Trans}$	3.901	3.905	$\pm$0.015		0.120	$\pm$0.025
	$\angle$ C-C-C			112.5 $\pm$ 1.4°		
	$\angle$ C-C-H			111.2 $\pm$ 1.4°		
	Index of resolution			.85		

[a]See references (41) and (42)

[b]Calculated by method of steepest ascents

[c]Calculated from trans C---C distance. Smallest C---C non-bonded distance in molecule

[d]Estimated from maximum of f(r) peak

assumed that the stable forms for the carbon skeleton are the trans form and two gauche forms (67) then all the areas in the observed radial distribution function may be accounted for. The potential barrier to rotation for the trans gauche model may be approximated by the function

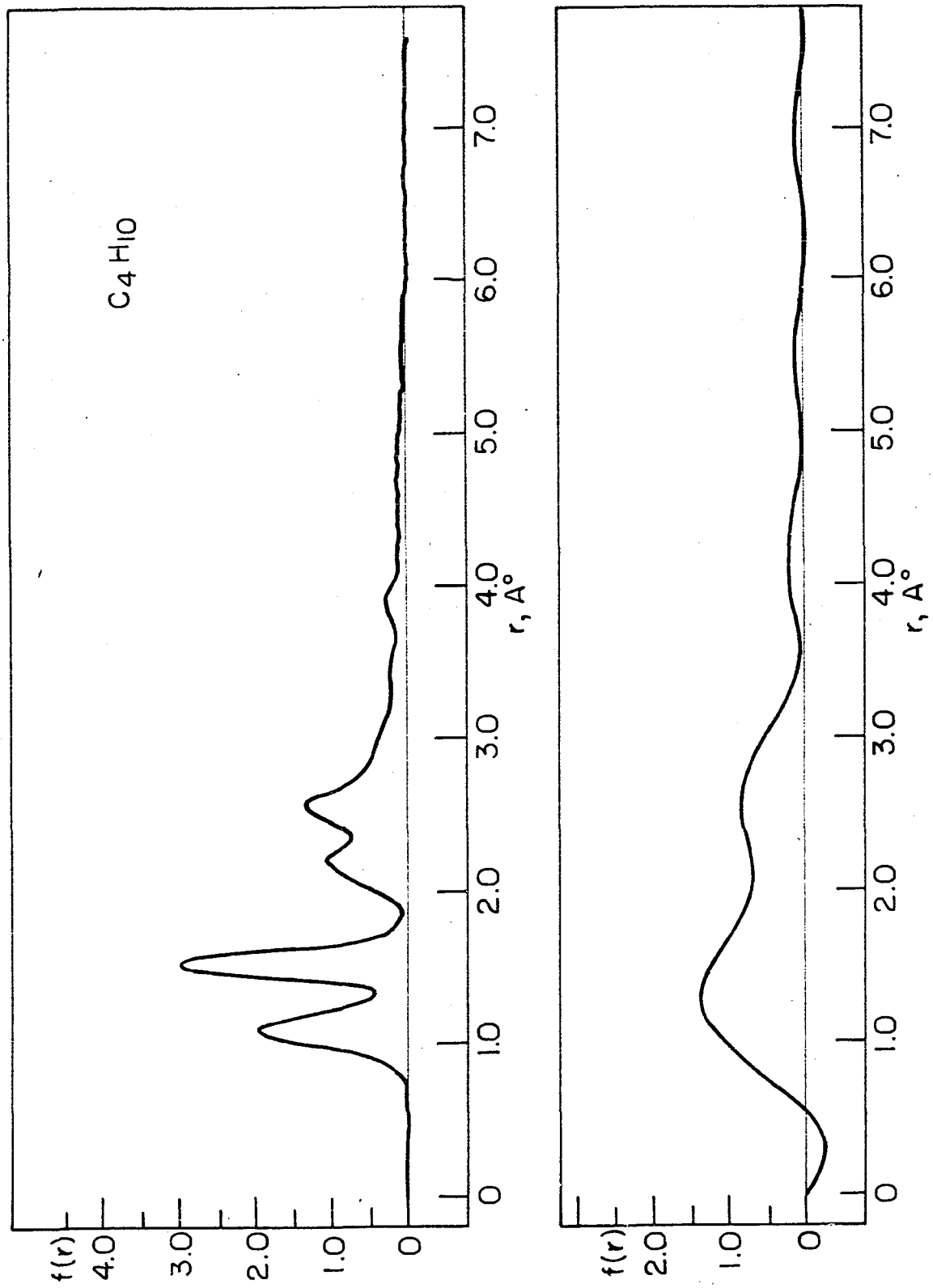
$$V(\theta) = V_1 + V_2 - V_1 \cos \theta - V_2 \cos 3\theta$$

where θ is the angle of rotation measured from the trans

Figure 8. Some radial distribution functions for n-butane

(a) The radial distribution function for n-butane

(b) The radial distribution function for n-butane using $M(s)_c$ data from s = 0 to s = 5


C4H10
f(r)
4.0
3.0
2.0
1.0
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°
f(r)
2.0
1.0
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°

position, $2(V_1 + V_2)$ is the barrier to free rotation and $(3/2)$ V_1 is the approximate energy difference between the trans and gauche forms. It was found that the trans C-C peak calculated using this potential and Pitzer's values (6) for the barriers were in agreement with the experimental trans peak. Experimentally it appeared that there were 65 per cent of the molecules in the trans configuration and 35 per cent of the molecules in the gauche configurations. The values used for the barriers were 3600 cal. $mole^{-1}$ for the barrier to free rotation and a value of 800 cal. $mole^{-1}$ for the energy difference between the trans and gauche forms. In the present study it appears that the uncertainty in the determination of the barrier heights is rather large, although the experiment is more sensitive to determination of the trans gauche energy difference than the trans cis energy difference.

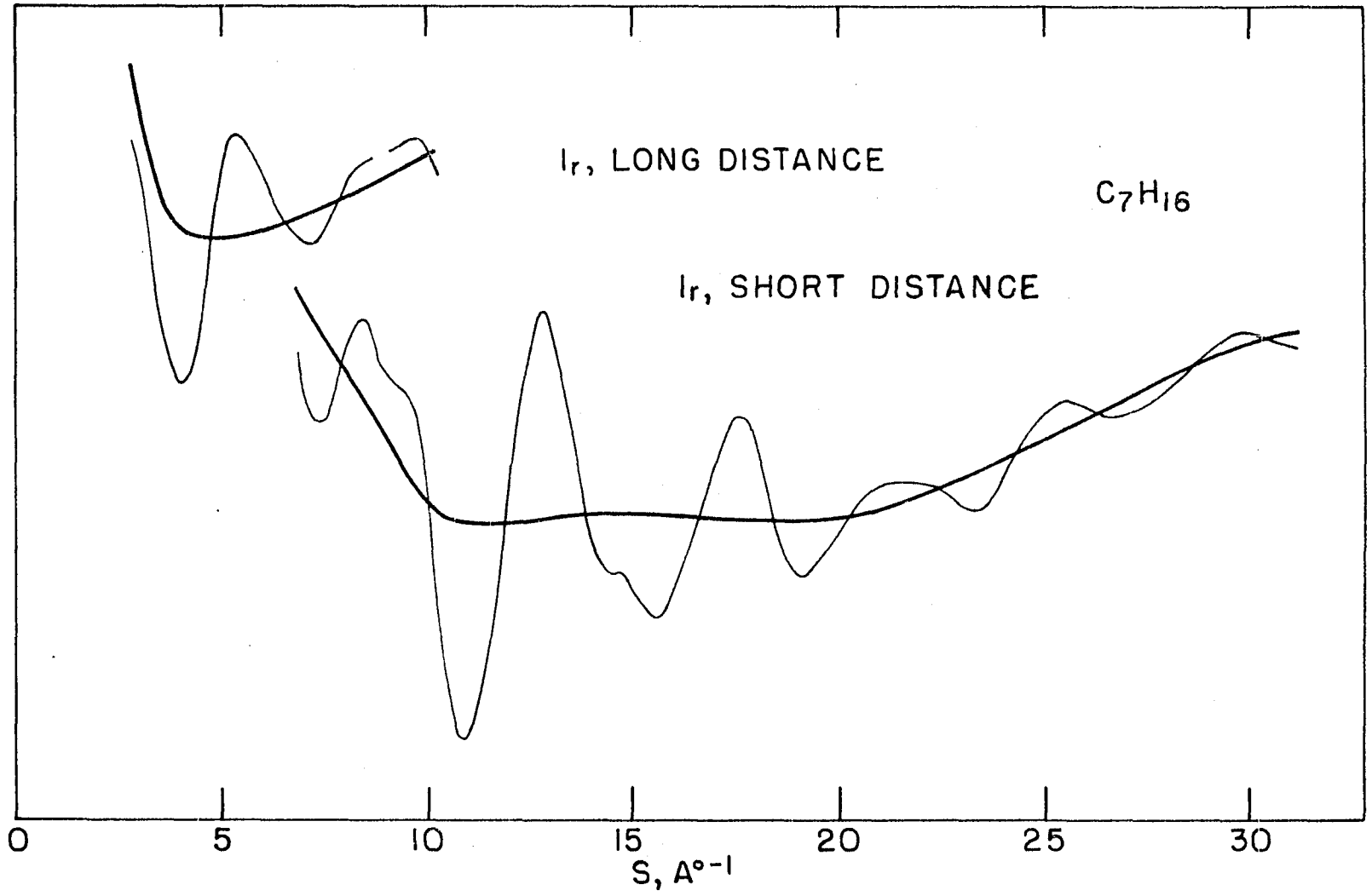
A radial distribution function was calculated using the $M(s)_c$ data from s = 0 to s = 5. These $M(s)_c$ values are somewhat arbitrary, and by performing the Fourier inversion it is possible to determine the sensitivity of the trans C-C peak to these values.

E. n-Heptane

A sample of n-heptane of better than 99.5 per cent purity was obtained from Professor R. S. Hansen of Iowa State College. A radial distribution curve for n-heptane and a radial

Figure 9. A plot of the experimental I_r and B(s) functions for n-heptane

(The heavier lines indicate the B(s) functions.)


I_r, LONG DISTANCE
C_7H_{16}
I_r, SHORT DISTANCE
0
5
10
15
20
25
30
S, A°-1

distribution curve using the $M(s)_c$ data for n-heptane over the range s = 0 to s = 5 were obtained. The same assumptions that were made in the case of n-butane about the equivalence and symmetry of all methylene and all methyl groups were made also for n-heptane. Since no shift was observed in the C-C single bond distance going from n-butane to n-heptane it was also assumed that all C-C single bond distances were equivalent.

These assumptions reduced the number of parameters required to determine the geometry of the framework independent of internal rotation to six which can be subsequently reduced to four by assuming that all the C-H bond distances, C-C-H angles, and C-C-C angles are the same. As in n-butane it is not possible to determine the parameters for the internal rotation of the methyl groups, but an analysis of the possible configurations of the carbon framework may be made. The notation T for trans and G for gauche will be adopted for describing the possible configurations of the molecule. It takes four symbols to describe an internal rotational configuration of the carbon skeleton since there are four C-C single bonds about which rotation may take place. The nomenclature used below follows that proposed by Mizushima (67) except that the differentiation between gauche forms is included in a weighting factor rather than by enumerating the gauche forms separately. For instance, the internal rotation-

al configurations denoted in Mizushima's notation as TTTG and TTTG' each with a weight of two will be referred to here as TTTG with a weight of four. It is impossible to determine from the electron diffraction data the percentages of the molecules in each internal rotational configuration without making assumptions about the relative barriers to internal rotation about each bond. It is possible nevertheless to estimate directly the percentages of molecules in groups of internal rotational configurations.

In the analysis of the parameters for the rigid framework it was found that the smallest C--C non-bonded distance was masked by radial distribution features depending on the internal motion of the carbon framework. However, it was noted that three of the peaks determining the geometry of the molecule in the completely trans state were well resolved. From these peaks three separate calculations of the smallest C--C distance were made with an internal consistency of 0.005 angstroms. The three peaks used in the calculation are labeled A, B, and C in order of increasing internuclear separation and are presented in a table below along with the rest of the parameters determined for n-heptane.

In order to obtain a description of the shape of n-heptane molecules the areas of the three peaks, A, B, and C were determined. The assignment of areas was subject to rather large uncertainty as indicated in Table 7 due to over-

Table 6. Molecular parameters and uncertainties for n-heptane

Peaks		$r_g(1)$[a]	$r_g(0)$	δr	l_{ij} calc[b]	l_{ij} exp	δl_{ij}
	C-H	1.106	1.113	±0.007	0.077	0.084	±0.012
	C-C	1.530	1.532	±0.006	0.055	0.058	±0.006
	C---H	2.167	2.175	±0.015		0.105	±0.020
	C---C	2.549[c]	2.553	±0.020			
A	C---C	3.920	3.924	±0.015		0.100	±0.025
B	C---C	5.086	5.089	±0.020		0.115	±0.030
C	C---C	6.431	6.433	±0.020		0.120	±0.035
		∠ C-C-H			109.6 ± 1.4°		
		∠ C-C-C			112.8 ± 1.4°		
		Index of resolution			1.00		

[a]Determined by the method of steepest ascents

[b]See references (41) and (42)

[c]Calculated from $r_g(1)$ values of peaks A, B, and C

lapping of peaks A, B, and C by broad shallow peaks associated with the more highly coiled molecules. Then by comparing the experimental area with the areas calculated assuming that all the molecules were in the TTTT configuration it was possible to estimate the percentages of the molecules in various groups of rotational configurations in a sample of n-heptane at about 285° Kelvin.

The percentages were also calculated for a simple model

Figure 10. Some radial distribution functions for n-heptane

(a) The radial distribution function for n-heptane

(b) The radial distribution function for n-heptane using $M(s)_c$ data from s = 0 to s = 5

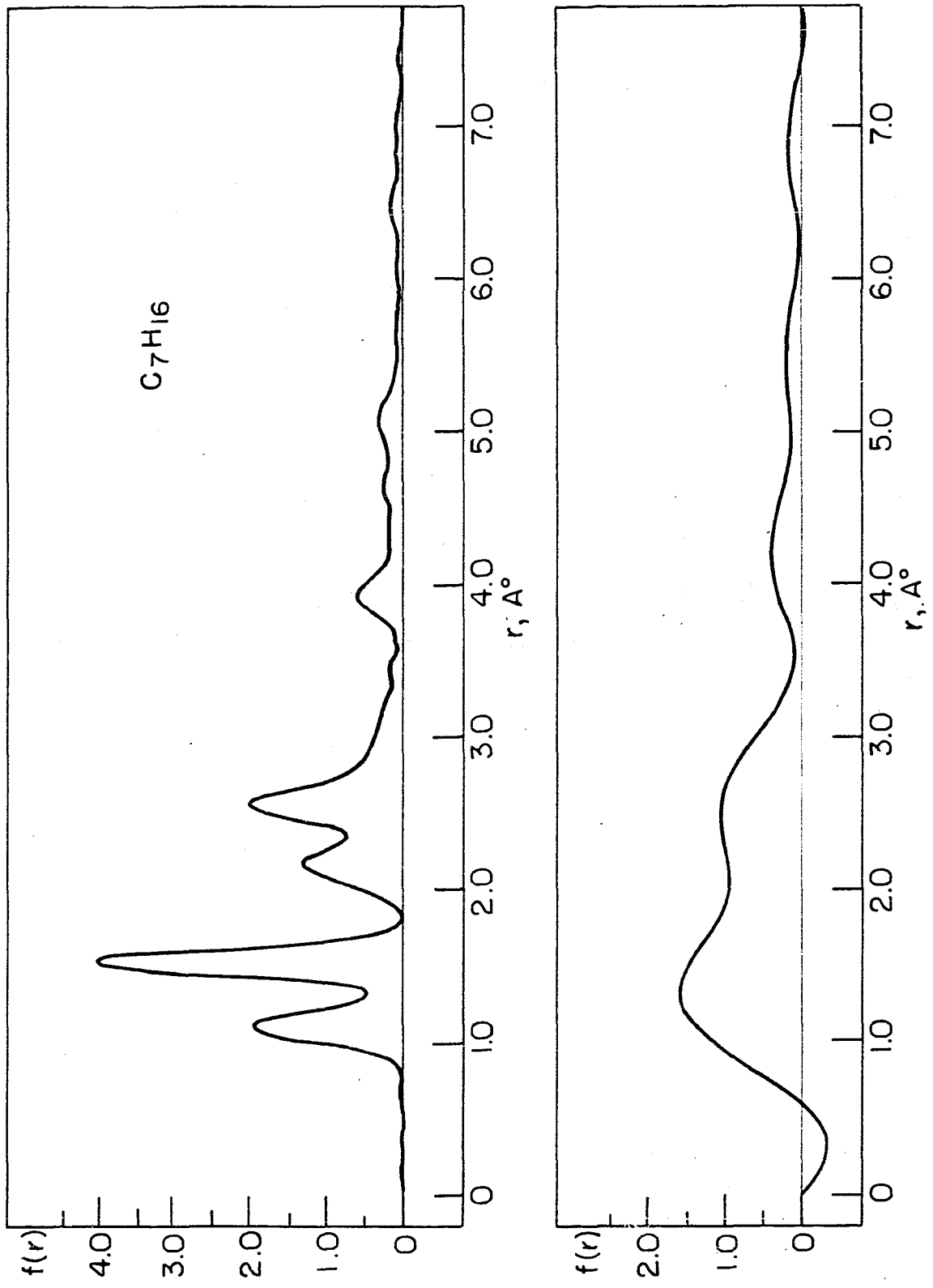

C7H16
f(r)
4.0
3.0
2.0
1.0
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°
f(r)
2.0
1.0
0
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
r, A°

Table 7. Analysis of the internal rotational configurations in n-heptane

Peak	Per cent of trans area[a]	Contributing configurations	Per cent of trans area per form	No. of configurations	Calculated per cent of area[b]	Calculated per cent of trans area
A	67 ± 20	TTTT	100	1	17.9	65.1
		TTTG	75	4	14.4	
		TTGT	75	4	14.4	
		TTGG	50	8	5.2	
		GTTG	50	4	2.6	
		TGTG	50	8	5.2	
		TGGT	50	4	2.6	
		TGGG	25	16	1.4	
		GTGG	25	16	1.4	
B	47 ± 20	TTTT	100	1	17.9	42.2
		TTTG	66.7	4	12.8	
		TTGT	33.3	4	6.4	
		TTGG	33.3	8	3.5	
		GTTG	33.3	4	1.7	
C	30 ± 25	TTTT	100	1	17.9	27.5
		TTTG	50	4	9.6	

[a]Estimated from the experimental radial distribution function

[b]Computed using the potential barriers for n-butane

in which it was assumed that the a priori probabilities of all configurations were equal, the barriers to rotation about all C-C bonds were the same and that the values for the potential barriers to rotation were the same as those in n-butane. Actually some of the configurations contributing to peak A may be sterically improbable (67) but the exclusion of these configurations would not affect the theoretical percentage by more than 3 per cent. A complete list of the internal rotational configurations contributing to each of the three peaks as well as the area of each, the area percentage compared with that calculated for an all trans sample, and the area percentage calculated for the above model is presented in Table 7. The diffraction results definitely establish the existence of a substantial proportion of extended molecules, but show also that the barrier to folding of the internal rotational isomers is not greatly in excess of thermal energy at room temperature.

F. Discussion of the Structures

A comparison of the results obtained from various structural determinations of ethylene, isobutylene, propylene and acetaldehyde are made in Table 8. The angle denoted by C=C-C will be chosen for comparison purposes. It should be noted that the corresponding angle generally associated with pure sp^2 hybridization is about 120° while that associated with

Table 8. Structural parameters for ethylene and methyl substituted ethylene

Molecule	$r_{C=C}$, A°	r_{C-C}, A°	r_{C-H}, A°	∠ C=C-C	Method	Date	Reference
CH_2CH_2	1.30 ± 0.10				MED[a]	1932	65
CH_2CH_2	1.34 ± 0.02		1.06 ± 0.03	125.0 ± 2.5°	ED[b]	1937	1
CH_2CH_2	1.331		1.085	121.0°	IR[c]	1939	71
CH_2CH_2	1.353 ± 0.010		1.071 ± 0.010	120.0 ± 0.3°	IR	1942	59
CH_2CH_2	1.334 ± 0.003		1.085 ± 0.005	122.1 ± 1.0°	SMED[d]	1957	present study
Me $CHCH_2$	1.33	1.49		125.0°	SVED[e]	1956	63
Me $CHCH_2$	1.353	1.488	1.09[f]	124.8°	MW[g]	1957	72
Me CH O		1.504 ± 0.010	1.124 ± 0.020[h]	123.6 ± 1.5°[i]	SMED	1955	58
Me CH O		1.501 ± 0.005	1.086 ± 0.005[h]	124.0 ± 0.1°[i]	MW	1957	73
Me_2CCH_2	1.33	1.50		123.5°	SVED	1956	63
Me_2CCH_2	1.326 ± 0.006	1.501 ± 0.006	1.107[j] ± 0.007	123.8 ± 2.0°	SMED	1957	present study

[a]Microphotometer electron diffraction method
[b]Electron diffraction visual method
[c]Infrared method
[d]Sector microphotometer electron diffraction method
[e]Sector visual electron diffraction method
[f]Distance assumed for methyl group
[g]Micro wave method
[h]Methyl group C-H distance
[i]The O=C-C angle
[j]Average C-H distance in molecule

Table 9. Comparison of some molecular parameters for selected aliphatic hydrocarbons

Molecule	r_{C-C}, A°	r_{C-H}, A°	$r_{C\cdots C}$ A°	∠ C-C-C	∠ C-C-H	Method	Date	Reference
n-Propane	1.52 ±0.05					ED[a]	1932	65
n-Butane	1.51 ±0.05					ED	1932	65
n-Butane	1.531±0.006[b]	1.103±0.007	2.546±0.020	112.5±1.4°	111.2±1.4°	SMED[c]	1957	present study
n-Pentane	1.53 ±0.05					ED	1932	65
n-Hexane	1.54 ±0.05					ED	1932	65
n-Heptane	1.532±0.006[b]	1.113±0.007	2.553±0.020	112.8±1.4°	109.6±1.4°	SMED	1957	present study
Sebacic acid	1.50 -1.54[d]			114°		XR[e]		68
h-C_{23} H_{48}			2.549±0.004			XR	1953	69
n-C_{36} H_{74}	1.534±0.006		2.546±0.004	112.2±0.3°		XR	1956	75

[a]Electron diffraction microphotometer method

[b]Bond distances reported as $r_g(0)$

[c]Sector microphotometer method of electron diffraction

[d]Carbon carbon single bond distances alternate between these values

[e]X-ray crystallographic structure determination in solid state

pure sp^3 hybridization is about 125°.

A number of interesting conclusions may be drawn from the more recent structural results in Table 8. First, the carbon carbon double bond distance appears to be nearly 1.333 angstroms as determined by the electron diffraction method in the present study on ethylene. This value falls in line with the smooth curve obtained by plotting the bond order as determined from the molecular orbital approach against the bond distance for acetylene, benzene, graphite, and ethane (62). The value of 1.353 reported by Gallaway and Barker (59) does not fall on the above curve. The double bond distance for isobutylene is in agreement with the value found in ethylene by the present study but the microwave results on propylene are not. It should be noted, however, that the ethylenic C-H bonds and angles in propylene were assumed (72) to be the same as those reported by Gallaway and Barker for ethylene and these assumptions appear to be appreciably in error. Also the distances for the hydrogens in the methyl group were assumed, and the values chosen appear rather low.

A second point is the shortening of the C-C single bonds adjacent to double bonds. This shortening has been attributed by Goldish, Hedberg, and Schomaker (74) to hyperconjugation of the methyl groups with the double bond. It has also been observed that the ionization of ethylene decreases with methyl substitution and Coulson (62) has explained this effect in

terms of hyperconjugation with a slight inductive effect. Although the experimentally determined shortening of the C-C bonds in these compounds had not been determined at the time the shortening of the C-C bond in methylacetylene was observed and Coulson explained this by attributing one-half of the 0.080 angstrom shortening to hybridization and one-half to hyperconjugation. In view of the bond angles observed in isobutylene and propene it would appear that the effect of hybridization in these compounds would be much less than in methyl acetylene.

The third aspect is the value of the bond angle C=C-C for methyl substituted ethylenes. The experimental values appear to be closer to the tetrahedral value of 125.3° for the angle in the case of the methyl substituted ethylenes. However, this effect is more subtle than the shortening of the bond distance and no simple explanation properly correlating bond angles in similar compounds has been proposed. If the double bond distance from the present study is substituted into the moments of inertia as determined by Gallaway and Barker (59) the resultant C=C-C angle is larger by several degrees than 120°. There also appear to be discrepancies between the electron diffraction results and the microwave results for C-H bonds. A more careful study of the asymmetry problem as it relates to both methods may help resolve this difference.

In the present study it appears that C-C single bonds in aliphatic compounds have the constant value 1.532 angstroms rather than the value of 1.544 angstroms usually quoted as the normal C-C distance. The apparent shift in the average bonded C-H distance between n-butane and n-heptane is too small to be interpreted reliably in terms of different distances for CH_3 and CH_2 groups. A shift in the C-C-C angle from the tetrahedral value is noted and such a shift is observed from the X-ray work on larger hydrocarbons (68, 69, 75).

The molecular configurations of n-butane and n-heptane were also determined. It was found that the experimental results determined in this study were in agreement with the results of Pitzer's (6) statistical thermodynamical calculation. The energy differences are given by Pitzer (6) as 800 cal. $mole^{-1}$ between the trans and gauche forms and 3600 cal. $mole^{-1}$ between the trans and cis forms. In n-heptane it was found that about 38 per cent of the molecules in a given sample at 285° Kelvin were in TTTT and TTTG configurations. About 33 per cent of the molecules were in GTTG, TGTT, and TTGG configurations and 28 per cent were in TGTG, TGGG, GTGG and TGGT configurations. Although the uncertainties in the percentages quoted above were rather large it appeared from the analysis of the radial distribution function that the proposed distribution was more probable than the completely extended picture (3) and definitely more probable than the "crumpled" or "highly coiled" picture (2).

VI. SUMMARY

The purpose of this work was to characterize the structures of ethylene, isobutylene, n-butane and n-heptane. The studies of ethylene and isobutylene were concerned with double bond lengths, bond distances adjacent to double bonds and bond angles associated with double bonds. In the case of n-butane and n-heptane the internal rotational configurations of the molecules were investigated as well as the bonded internuclear distances. The results of these investigations are tabulated in Tables 1 through 7. In Table 8 and 9, the principle results are summarized and comparisons are made with related molecules.

In order to obtain the best possible analysis of the accurate intensity data obtained by the sector microphotometer technique of electron diffraction, high speed digital computers were used. A survey of the theory of electron diffraction was made. Computation of the theoretical expressions useful in the interpretation of the diffraction data were programmed for the I.B.M. 650 digital computer along with operations for numerical analysis of the data. Because of the high speed of the computer it was possible to use more powerful numerical methods and more rigorous theoretical expressions for interpretation of the diffraction data than had previously been used. The discussions of these programs are presented in the appendices.

The structural parameters were obtained in this study mainly from analysis of the radial distribution function. Analytical methods for analysis of the radial distribution function are discussed for the cases with and without internal rotation. The reliability of the molecular parameters is considered and estimates are made of the various errors involved.

It was found that the length of the C=C double bond in ethylene and isobutylene is significantly shorter than that reported for the most definitive spectroscopic study of ethylene. The C=C-C and C=C-H angles in isobutylene and ethylene respectively were found to be larger than 120°. It was also found that the C-C single bond distance in isobutylene is shortened .03 angstroms by the influence of the adjacent double bond, as judged by a comparison with the single bond distance in saturated compounds. The root mean square amplitudes of vibration and internuclear distances were reported for all bonded distances along with important bond angles for all molecules studied. In the case of n-butane the electron diffraction structure was found to be in agreement with that implied by the barrier to internal rotation calculated by Pitzer. For n-heptane the shape of the molecule was described in terms of percentages of molecules in various internal rotational configurations. The results for n-heptane indicated conclusively that the molecules in the gaseous state

at 285° Kelvin are not highly coiled in contrast to the interpretation by McCoulrey, et al., of vapor phase viscosity measurements (2), but rather are more likely to be extended in predominantly trans configurations. The diffraction results suggest, however, as in the case of n-butane, that the excess energy of gauche configurations over trans is not large compared with thermal energy at room temperature.

VII. REFERENCES

1. Pauling, L. and Brockway, L. O., J. Am. Chem. Soc. 59, 1223 (1937).

2. McCoulrey, J. C., McCrea, J. N., and Ubhelohde, A. R., J. Chem. Soc. 1951, 1961.

3. Stein, R. S., J. Chem. Phys. 21, 1193 (1953).

4. Pitzer, K. S., J. Chem. Phys. 8, 711 (1940).

5. Pitzer, K. S., Ind. Eng. Chem. 36, 829 (1944).

6. Pitzer, K. S., Disc. Faraday Soc. 1951 No. 10, 66.

7. Aston, J. G., Isserow, S., Szasz, G. J., and Kennedy, R. M., J. Chem. Phys. 12, 336 (1944).

8. Stuart, H. A. and Schiessl, S. V., Ann. Physik 62, 321 (1948).

9. Debye, P. J. W., J. Math. and Phys. 4, 133 (1925).

10. Debye, P. J. W., Bewilogua, L., and Ehrhardt, F., Phys. Zeit. 30, 84 (1929).

11. Davisson, C. J. and Germer, L. H., Phys. Rev. 30, 705 (1927).

12. Mark, H. and Wierl, R., Naturwissenschaften 18, 205 (1930).

13. Mott, N. F., Proc. Roy. Soc. (London) A 127, 658 (1930).

14. Pauling, L. and Brockway, L. O., J. Chem. Phys. 2, 867 (1934).

15. Brockway, L. O., Revs. Mod. Phys. 8, 231 (1936).

16. James, R. W., Phys. Zeit. 33, 737 (1932).

17. Morse, P. M., Phys. Zeit. 33, 443 (1932).

18. Bewilogua, L., Phys. Zeit. 32, 740 (1931).

19. Pauling, L. and Brockway, L. O., J. Am. Chem. Soc. 57, 2684 (1935).

20. Finbak, C., Avhandl Norske Videnskaps - Akad. Oslo. I. Mat. - Nuturv. Kl. No. 13 (1937).

21. Debye, P. P., Phys. Zeit. 40, 66 (1939).

22. Degard, C., Pierard, J., and Van der Grinten, W., Nature 136, 142 (1935).

23. Degard, C., Bull. Soc. Roy. Sci. Liege 12, 383 (1937).

24. Walter, J. and Beach, J. W., J. Chem. Phys. 8, 601 (1940).

25. Spurr, R. and Schomaker, V., J. Am. Chem. Soc. 64, 2693 (1942).

26. Debye, P. J. W., J. Chem. Phys. 9, 55 (1941).

27. Harvey, R. B., Keidel, F. A., and Bauer, S. H., J. App. Phys. 21, 860 (1950).

28. Coffin, K. P. and Bauer, S. H., Rev. Sci. Instr. 33, 115 (1952).

29. Bauer, S. H., J. Chem. Phys. 18, 27 (1950).

30. Hastings, J. M., and Bauer, S. H., J. Chem. Phys. 18, 13 (1950).

31. Hassel, O. and Viervoll, H., Acta Chem. Scand. 1, 149 (1947).

32. Karle, I. L. and Karle, J., J. Chem. Phys. 17, 1052 (1949).

33. Karle, J. and Karle, I. L., J. Chem. Phys. 18, 957 (1950).

34. Karle, J. and Karle, I. L., J. Chem. Phys. 18, 963 (1950).

35. Bartell, L. S., Brockway, L. O., and Schwendeman, J. Chem. Phys. 23, 1854 (1955).

36. Bartell, L. S., and Brockway, L. O., Nature 171, 978 (1953).

37. Bartell, L. S., and Brockway, L. O., J. Chem. Phys. 23, 1860 (1955).

38. Brockway, L. O. and Bartell, L. S., Rev. Sci. Instr. 25, 569 (1954).

39. Bartell, L. S., and Brockway, L. O., J. Appl. Phys. 24, 656 (1953).

40. Bartell, L. S. and Brockway, L. O., Phys. Rev. 90, 833 (1953).

41. Morino, Y., Kuchitsu, K., and Shimanouchi, T., J. Chem. Phys. 20, 726 (1952).

42. Morino, Y., Kuchitsu, K., Takahashi, A., and Maeda, K., J. Chem. Phys. 21, 1927 (1953).

43. Schomaker, V. and Glauber, R., Nature 170, 291 (1952).

44. Schomaker, V. and Glauber, R., Phys. Rev. 89, 667 (1953).

45. Bartell, L. S., J. Chem. Phys. 23, 1219 (1955).

46. Morino, Y. and Hirota, E., Tokyo, Japan. On hindered rotation in molecules by electron diffraction. Private communication. 1957.

47. Swick, D. A., Karle, I. L., and Karle, J., J. Chem. Phys. 22, 1242 (1954).

48. Swick, D. A. and Karle, I. L., J. Chem. Phys. 23, 1499 (1955).

49. Ibers, J. A. and Hoerni, J. A., Acta Cryst. 7, 405 (1954).

50. Hoerni, J. A. and Ibers, J. A., Phys. Rev. 91, 1182 (1953).

51. Hoerni, J. A., Phys. Rev. 102, 1530 (1956).

52. Mott, N. F. and Massey, H. S. W., "The Theory of Atomic Collisions", Oxford University Press, London, England, 1949.

53. Morse, P. M. and Feshbach, H., "Methods of Theoretical Physics", Vol. 2, McGraw-Hill Book Co., Inc., New York, New York, 1953.

54. Titchmarsh, E. L., "Introduction to the Theory of Fourier Integrals", Oxford University Press, London, England, 1937.

55. Dirac, P. A. M., "Quantum Mechanics", Oxford University Press, London, England, 1947.

56. Anderson, Richard Eugene. The molecular structures of trifluoromethyl bromide, trifluoromethyl iodide, trifluoromethyl cyanide, and trifluoromethyl sulfurpentafluoride. Unpublished Ph.D. Thesis, University of Michigan Library (Ann Arbor, Michigan), 1956.

57. Booth, A. D., "Numerical Methods", Butterworth Scientific Publications, London, England, 1955.

58. Schwendeman, Richard Henry. A critical evaluation and improvement of the procedures for electron diffraction by gases and the determination of the molecular structures of carbon tetrachloride, trifluoroethane, methyltrifluorosilane, acetaldehyde, and trifluoroacetaldehyde. Unpublished Ph.D. Thesis. University of Michigan Library (Ann Arbor, Michigan), 1955.

59. Gallaway, W. S. and Barker, E. F., J. Chem. Phys. 10, 88 (1942).

60. Badger, R. M., Phys. Rev. 45, 648 (1934).

61. Scheib, W. and Lueg, P., Z. Physik 81, 764 (1933).

62. Coulson, C. A., "Valence", Clarendon Press, Oxford, England, 1952.

63. Schomaker, V. and McHugh, J., Pasadena, Calif.: On the structures of propylene and isobutylene. Prevate communication. 1956.

64. Aston, J. G., Disc. Faraday Soc. 1951 No. 10, 73.

65. Wierl, R., Ann. Physik 13, 453 (1932).

66. Wierl, R., Ann. Physik 8, 521 (1931).

67. Mizushima, San-Ichiro, "Structure of Molecules and Internal Rotation", Academic Press, Inc., New York, New York, 1954.

68. Wyckoff, Ralph W. G., "Crystal Structures", Vol. 2, Interscience Publishers, Inc., New York, New York, 1957.

69. Smith, A. E., J. Chem. Phys. 21, 2229 (1953).

70. Melaven, R. M. and Mack, E., J. Am. Chem. Co. 54, 888 (1932).

71. Thompson, H. W., Trans. Faraday Soc. 35, 697 (1939).

72. Lide, D. R. and Mann, D. E., J. Chem. Phys. 27, 868 (1957).

73. Kilb, R. W., Chun, C. L., and Wilson, E. B., J. Chem. Phys. 26, 1695 (1957).

74. Goldish, E., Hedberg, K., Schomaker, V., J. Am. Chem. Soc. 78, 2714 (1956).

75. Shearer, H. M. M., and Vand, V., Acta Cryst. 9, 379 (1956).

VIII. ACKNOWLEDGMENTS

I wish to thank Professor L. S. Bartell for his valuable instruction, help and advice given during the course of this research.

For the use of the electron diffraction unit I would like to thank Professor L. O. Brockway of the University of Michigan.

For their helpful suggestions, I would like to thank Professor B. C. Carlson, Mr. Robert Fitzwater and Mr. Robert White.

I would like to thank the donors to the Petroleum Research Fund administered by the American Chemical Society for supporting this research.

For their help with some of the calculations I would like to thank Miss N. Carlson, Mr. G. Ferguson, Miss E. Bortle and Mr. D. Kohl.

I wish to thank my wife for her help with the calculations and for her moral support throughout the study.

And finally I want to thank Mrs. H. F. Hollenbeck for typing the rough draft of the thesis.

IX. APPENDIX A: THEORETICAL INTENSITY PROGRAM FOR ELECTRON DIFFRACTION OF GASES

A. Mathematical Method

This program evaluates the functions

$$M(q) = \sum_{ij}' (Z_i - F_i(q))(Z_j - F_j(q)) e^{-\ell_{ij}^2 \pi^2 q^2/200} \sin \frac{2\pi q r_{ij}}{20} \Big/ \sum_k \left\{ (Z_k - F_k(q))^2 + S_k(q) \right\} \pi q r_{ij}/10$$

and

$$M(q)_c = \sum_{ij}' Z_i Z_j e^{-\ell_{ij}^2 \pi^2 q^2/200} \sin \frac{2\pi q r_{ij}}{20} \Big/ \sum_k (Z_k^2 + Z_k) \pi q r_{ij}/10$$

Here the F(q) terms are the X-ray atom form factors for coherent scattering, and the S(q) terms are the X-ray atom form factors for incoherent scattering. The Z's are the atomic numbers of the atoms involved, and the prime on the summation indicates that terms for which i=j are not included. It should be noted that ℓ_{ij} is the root-mean-square amplitude of vibration and r_{ij} is the internuclear separation for the ijth atom pair. Also, q is the scattering variable defined as $40 \sin(\theta/2)/\lambda$ where λ is the wave length of the incident beam and θ is the scattering angle.

The form factors are obtained from a stored table, and the functions are computed from the Hasting's approximations

$$e^{-\ell_{ij}^2 \pi^2 q^2/200} = /2\pi/\lfloor b_o + x^2(b_2 + x^2(b_4 + x^2(b_6 + x^2 b_8)))\rfloor$$

where $x = \ell_{ij}\, \pi\, q/10$, $b_o = 2.511261$, $b_2 = 1.172801$, $b_4 = .494618$, $b_6 = -.063417$, and $b_8 = .029461$; and

$$\sin(\pi/2)(qr_{ij}/5) = x\lfloor c_1 + x^2(c_3 + c_5 x^2)\rfloor$$

where $x = qr_{ij}/5$, $c_1 = 1.5706268$, $c_3 = -.6432292$, and $c_5 = .0727102$.

As output data, both $M(q)$ and $M(q)_c$ are obtained as well as $M(q)_c - M(q)$ and $\sum_k \lfloor (Z_k - F_k(q))^2 + S_k(q)\rfloor/q$. This last quantity is proportional to the theoretical background after the intensity has been modified by a q^3 sector and is useful as a guide in drawing the theoretical background. The quantity $M(q)_c - M(q)$ may be added to the experimental $M(q)$ curve to correct the data to pure nuclear scattering.

B. Range and Accuracy

This program used fixed decimal arithmetic and allows for five significant figures in all calculations. The function $M(q)$ may be computed to $q=100$ while $M(q)_c$ may be computed to $q=200$ or larger if desired. It should be noted that there are no restrictions on the size of the r_{ij} terms, but ℓ_{ij} must be less than $.472A^{\circ}$.

C. Storage

There are five hundred words of storage for the F(q) values and five hundred words of storage for the S(q) values. The F(q) and S(q) values are stored in order of increasing atomic number: $F_1(q)$ in 0-99, $S_1(q)$ in 500-599, $F_2(q)$ in 100-199, $S_2(q)$ in 600-699, etc. (The form factors referring to the atom with the smallest atomic number are stored in 0-99, and the code number used for the initial value of $F_1(q)$ is one. The code number for the initial value of $F_2(q)$ is two, etc.) These code numbers are used on the detail cards (see input-output) to indicate the initial storage locations of the $F_i(q)$ and $F_j(q)$ values. The initial storage locations of the $S_i(q)$ and $S_j(q)$ values are determined in the program from the value given for the location of the $F_i(q)$ and $F_j(q)$ values.

D. Speed

The program, on the average, takes sixty-nine seconds to compute one term of the primed summations for one hundred values of q. For a molecule with fifteen distances, the program takes seventeen and one-half minutes to complete the calculation for one hundred q values.

E. Equipment

The basic I.B.M. 650 digital computer with no special

attachments and two thousand words of drum storage is required.

F. Error Checks

There are no built-in error checks. A plot of the function will generally reveal any serious machine error.

G. Input-Output

The format of input and output cards is discussed in detail under "Detailed Operating Instructions." Input consists of one control card indicating the number of terms in the primed and unprimed summations plus the atomic number and multiplicity of each kind of atom present in the molecule. After this control card, there follows one detail card for each primed sum term. Each of these cards contains the ℓ_{ij} and r_{ij} values for the term as well as the locations of the initial $F_i(q)$ and $F_j(q)$ values. The detail card also has Z_i, Z_j, and the multiplicity of the term (e.g. the number of times the $_{ij}$th and $_{ji}$th atom pairs occur in the molecule). The output cards have $M(q)_c$, $M(q)$, $M(q)_c - M(q)$, q, and $\sum_k [(Z_k - F_k(q))^2 + S_k(q)]/q$ for two different values of q on each card.

H. Detailed Operating Instructions

Card formats

(1) Control Card Format

Word	Information	Decimal Form		Notes
1	No. of atoms of atomic no. Z_i; atomic no. Z_i	ooxxx	ooxxx	Should be in order of increasing atomic number.
2	No. of atoms of atomic no. Z_i; atomic no. Z_i	ooxxx	ooxxx	
3	No. of atoms of atomic no. Z_i; atomic no. Z_i	ooxxx	ooxxx	
4	No. of atoms of atomic no. Z_i; atomic no. Z_i	ooxxx	ooxxx	
5	No. of atoms of atomic no. Z_i; atomic no. Z_i	ooxxx	ooxxx	
6	No. of atoms of differing atomic numbers	ooooo	oooox	Limit of 5
7	No. of different distances in the molecule	ooooo	ooxxx	Limit of 15

(2) Detail Card Format

Word	Information	Decimal Form	Notes
1	r_{ij} value	oooooxx.xxx	No restrictions
2	ℓ_{ij} value	o.xxxoooooo	Max. limit of 0.472
3	Code no. of $F_i(s)$	oooxoooooo	1-5 (Note that F values are stored in order of increasing Z) See Storage.
4	Code no. $F_j(s)$	oooxoooooo	
5	Z_i, atomic no.	ooxxx.ooooo	No restriction
6	Z_j, atomic no.	ooxxx.ooooo	No restriction
7	M_{ij}, the multiplicity of the ijth atom pair	ooooooooooxx	No restriction

(3) Output Card Format

Word	Information	Decimal Form
1	$M(q_1)_c$	ox.xxxxxxxx
2	$M(q_1)$	ox.xxxxxxxx
3	$M(q_1)_c - M(q_1)$	ox.xxxxxxxx
4	q_1 and $\sum_k \left\{ (Z_k - F_k(q_1))^2 + S_k(q_1) \right\}$	xxx xx.xxxxx
5	$M(q_2)_c$	ox.xxxxxxxx
6	$M(q_2)$	ox.xxxxxxxx
7	$M(q_2)_c - M(q_2)$	ox.xxxxxxxx
8	q_2 and $\sum_k \left\{ (Z_k - F_k(q_2))^2 + S_k(q_2) \right\}$	xxx xx.xxxxx

(4) Form Factor Card Format

Word	Information	Decimal Form
1	Location of $F_i(q_1)$ or $S_i(q_1)$	24 oxxx 1990
2	$F_i(q_1)$ or $S_i(q_1)$	oooxx.xxxxx
3	Location of $F_i(q_2)$ or $S_i(q_2)$	24 oxxx 1991
4	$F_i(q_2)$ or $S_i(q_2)$	ooo xx.xxxxx
5	Location of $F_i(q_3)$ or $S_i(q_3)$	24 oxxx 1992
6	$F_i(q_3)$ or $S_i(q_3)$	oooxx.xxxxx
7	Location of $F_i(q_4)$ or $S_i(q_4)$	24 oxxx 1993
8	$F_i(q_4)$ or $S_i(q_4)$	oooxx.xxxxx

Note that it takes twenty-five cards for a table of coherent factors and twenty-five cards for a table of incoherent factors for each atom of different atomic number.

I. Operating Procedure

The console is set at 70 1201 1970. The overflow sense switch is set at sense, and all other switches are placed in position for normal operation. The first two cards of the program deck are drum clear cards, which set the entire contents of the drum to minus zeros. Of the last four cards of the program deck two are used and two are separated from the deck depending upon the range of q to be calculated. The cards marked q=100 and q=101 are retained at the end of the program deck if the theoretical curves are to be run out to q=100, and the cards marked q=200 and q=201 are retained if the curves are run out to q=200. The detail cards are placed after the complete program deck. The detail control card is first, followed by the detail cards for the primed summation terms. The program is written for continuous operation so that as many decks or detail cards as desired may be run. One set of molecular intensity curves is calculated for each detail control card. If a detail control card is fed in without being followed by detail cards, the program will still compute the function

$$\sum_k \left[(Z_k - F_k(q))^2 - S_k(q)\right]/q.$$

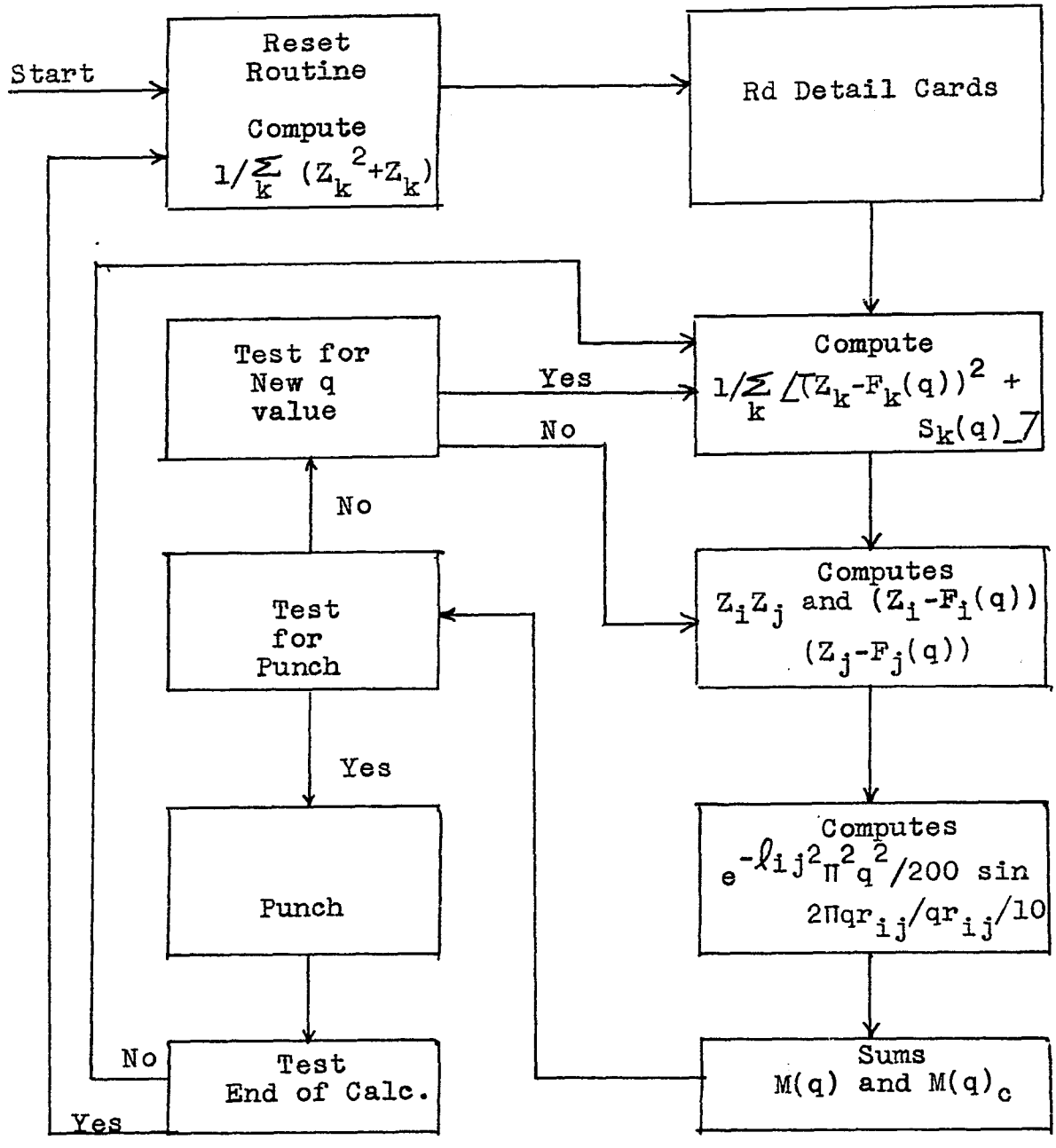
It should be noted that the deck of F(q) and S(q) values, which are part of the program deck, must be prepared from tables. The program requires one hundred values of F(q) at

integral q values and one hundred values of S(q) at integral q values for each different atom in the molecule.

The control panel wiring is "straight in-straight out" and if a board is not available the I.B.M. 650 manual should be consulted for the wiring information.

The detailed program instructions may be obtained from Dr. L. S. Bartell, Chemistry Department, Iowa State College, Ames, Iowa.

FLOW DIAGRAM OF THEORETICAL INTENSITY PROGRAM



X. APPENDIX B: RADIAL DISTRIBUTION PROGRAM FOR ELECTRON DIFFRACTION OF GASES

A. Description

In the sector microphotometer method of electron diffraction the rotating sector modifies the intensity of the scattered electrons received by the photographic plate in such a way as to level the precipitous fall off of intensity with scattering angle. This makes it possible to keep the exposure over the plate within the optimum range of response of the emulsion and thereby permits relative intensities of diffraction features to be measured objectively by microphotometry.

The molecular scattering function is defined by the relation

$$M(s) = (I_r/B) - 1 \quad ,$$

where M(s) is the molecular scattering function, I_r is the experimentally measured intensity, and B is a background function evenly cleaving the sinusoidal molecular features that must obey two restrictions: It must be smooth so that no false peaks in the radial distribution function are introduced by inversion of the B function, and it must be drawn so that the radial distribution function is nowhere negative.

A radial distribution program was coded for the I.B.M. 650 with modifications to aid in the determination of a B

function compatible with the restrictions. The program computes the Fourier integral

$$f(r) = \int_o^{s_{max}} e^{-bs^2} \, sM(s)_c \sin(sr)ds \quad ,$$

where, provided s_{max} and b are sufficiently large

$$f(r) = (2/\pi)(\pi/16b)^{1/2}(\sum_{ij}{}' Z_iZ_j/\sum_k Z_k^2+Z_k)$$

$$\int_{-\infty}^{+\infty} \rho^{-1} P_{ij}(\rho) e^{-\frac{(r-\rho)^2}{4b}} \, d\rho$$

Here $P_{ij}(\rho)$ is the probability function of the distribution of distances between the i^{th} and j^{th} atoms, the Z's are the atomic numbers of the atoms, and s is the scattering variable

$$s = (4\pi/\lambda) \sin(\theta/2)$$

where λ is the electron wavelength and θ is the scattering angle.

When experimental data are obtained only over the range s=3 to s=30, it is advantageous to use a damping function to minimize series termination errors and to use theoretical $M(s)_c$ values from s=0 to s=3. The theoretical values are calculated from the well known equation

$$M(s)_c = (\sum_{ij}{}' Z_iZ_j/\sum_k Z_k^2+Z_k) e^{-\ell_{ij}^2 s^2/2} \sin sr_{ij}/sr_{ij}$$

and a damping factor is inserted into the integral expression such that

$$f(r) = \int_0^{s_{max}} e^{-bs^2} \, sM(s)_c \sin(sr) \, ds$$

where b is a constant determined so that $e^{-bs^2} = 0.1$, when $s = 30$. In the program allowance is made for use of an arbitrarily chosen damping factor in order to weight the experimental intensities in the best possible manner. An arbitrary number of theoretical $M(s)_c$ values may also be chosen. This makes it possible to invert a complete theoretical model, which can sometimes be a help in interpretation of radial distribution functions.

The program uses I_r and B values to compute $M(s)$, which is the experimental molecular scattering function arising from screened nuclear scattering. $M(s)$ must be corrected to the pure nuclear scattering function, $M(s)_c$. This is done by adding to $M(s)$ the theoretically computed correction function

$$M(s)_c - M(s) = \Big[\Big(\sum{}' Z_i Z_j / \sum_k Z_k^2 + Z_k \Big) - \Big(\sum_{ij}{}' (Z_i - F_i(s))(Z_j - F_j(s)) / \sum_k \big\{ (Z_k - F_k(s))^2 + S_k(s) \big\} \Big) \Big] \cdot \Big(e^{-l_{ij}^2 s^2/2} \sin sr_{ij} / sr_{ij} \Big)$$

where $F_j(s)$ is the coherent X-ray atom form factor of the j^{th}

atom and $S_k(s)$ is the incoherent X-ray atom form factor of the k^{th} atom.

As an aid in choosing the best background function, the program takes f(r) values from r=0 to any arbitrarily selected distance, r_1, up to 2.5 A^o and performs a Fourier inversion on them. The Fourier integral for this case is

$$sM(s')_c = \int_o^{r_1} f(r) \sin(sr)dr \quad ,$$

where r_1 is chosen to be at smaller r than the first molecular feature appearing in the f(r) function. Inversion of f(r) values out to this feature indicates what changes in the background are necessary to smooth the base line of the initial part of the f(r) curve. If a smooth curve is drawn through the indicated changes, it is often observed that improvements occur also in other parts of the f(r) function.

A negative region at large values of r in the f(r) function can generally be removed by changes in the index of resolution of the theoretical data. In the program an assumed index of resolution is multiplied by the theoretical $M(s)_c$ that is used at small scattering values and also by the correction applied to M(s) to reduce it to pure nuclear scattering. This index of resolution is varied until an acceptable f(r) function is obtained.

If data are obtained at two different distances, there is then the additional problem of what to do about over-

lapping data. In the present program no attempt is made to weight the overlapping data. Instead, the data from one distance are used to the center of the overlap region, and from here, the data from the other distance are used. This is justifiable since the B function is essentially arbitrary over the first few s values and the last few s values for any distance. The question of differing indices of resolution for different camera distances is left until later refinement.

Since the data and background, even using current techniques, are essentially arbitrary out to s=5, the reliability of any molecular parameters at large distances must be tested. To do this, the $M(s)_c$ values out to s=5 are inverted, and the features at large distances in this inversion are compared to the molecular features to determine the sensitivity of derived features to arbitrary choices made in $M(s)_c$ at small s.

The advantage in using the radial distribution method over the correlation method is that an f(r) function with a noise level of about 1% of the maximum value of the function can be obtained in four iterations. For a molecule with a largest distance of about 6A°, this takes approximately three fourths of an hour of machine time. In comparison, the time needed to obtain one theoretical model for the correlation method for a molecule with fifteen distances is twenty minutes

of machine time.

As an added advantage, the program can be made to give the $M(s)$ and $M(s)_c$ values whenever they are desired.

B. Mathematical Method

The program approximates the Fourier integral

$$f(r) = \int_0^\infty sM(s) \sin (sr)\, ds$$

by the sum

$$f(r) = \pi^2/100 \sum_{q=1}^{N} q\, M(q) \sin (2\pi\, qr/20)\, \Delta q$$

where $s=\pi q/10$ and $\Delta q=1$. The function $f(r)$ is evaluated at $.05A^\circ$ intervals from zero to $10A^\circ$. Since $f(r)$ is periodic with a wave length of $20A^\circ$, it is not possible to go further than $10A^\circ$.

For the computation of $f(r)$ the argument of the sine is rewritten in the form $2\pi\, n(.0025)$ where n is any integer. It should be noted that r can be written as $n(.05)$ where n is an integer; and since q is also an integer, the product qn is also an integer. The sums involved in the integral approximation are written out to the first few terms below:

$f(.05)=\pi/50(M(1)\sin 2\pi\lfloor(.0025)\rfloor + 2M(2)\sin 2\pi\lfloor 2(.0025)\rfloor$
$+3M(3)\sin 2\pi\lfloor 3(.0025)\rfloor + \cdots)$

$f(.10)=\pi/50(M(1)\sin 2\pi\lfloor 2(.0025)\rfloor + 2M(2)\sin 2\pi\lfloor 4(.0025)\rfloor$
$+ 3M(3) \sin 2\pi \lfloor 6(.0025)\rfloor + \cdots)$

$$f(.15) = \pi/50(M(1)\sin 2\pi/3(.0025)/ + 2M(2)\sin 2\pi/6(.0025)/ + 3M(3)\sin 2\pi/9(.0025)/ + \cdots)$$

The complete calculation is represented in matrix notation as

$$\underline{f(r)} = A_q\underline{M(q)}$$

or

$$\begin{pmatrix} f(.05) \\ f(.10) \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ f(n) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \cdot & \cdot & \cdot \\ 2 & 4 & 6 & 8 & 10 & 12 & \cdot & \cdot & \cdot \\ 3 & 6 & 9 & 12 & 15 & 18 & \cdot & \cdot & \cdot \\ 4 & 8 & 12 & 16 & 20 & 24 & \cdot & \cdot & \cdot \\ 5 & 10 & 15 & 20 & 25 & 30 & \cdot & \cdot & \cdot \\ 6 & 12 & 18 & 24 & 30 & 36 & \cdot & \cdot & \cdot \\ \cdot & & & & & & & & \\ \cdot & & & & & & & & \\ \cdot & & & & & & & & \end{pmatrix} \begin{pmatrix} M(1) \\ 2M(2) \\ 3M(3) \\ 4M(4) \\ \cdot \\ \cdot \\ \cdot \\ nM(n) \end{pmatrix}$$

where the sine matrix is completely characterized by the value of n in the sine argument.

It is faster to compute the sums one term at a time rather than one f(r) value at a time. This is done by computing an $M(q)_c$ value and then multiplying it by the proper set of sines, adding each answer to its proper f(r) summation. To do this, the program uses a sine table evaluated at 2π (.0025) radian intervals from zero to 2π. Thus the first intensity is multiplied by successive sine values starting at the first; the second intensity is multiplied by two and then every other sine value starting at the second. This process is continued until all the sums are evaluated. When a value of the sine goes out of the zero to 2π limit, the program

senses this and reduces the sine to its proper value in the table.

C. Range and Accuracy

All of the computation is performed in fixed decimal form. Decimal points on input and output data are given in the discussion of operating instructions.

If a theoretical $M(s)_c$ function for a diatomic molecule is inverted by this program, the center of the resultant f(r) peak is accurate to better than one part per ten thousand. The sine table is exact to about six decimal places but is smooth to eight.

To test the error in the integration, a finer subdivision than $\Delta q = 1$ was chosen using experimental intensity data for ethylene. When a value of 1/2 was used for Δq the resultant values of the f(r) function deviated by no more than 0.3% from the corresponding values for the $\Delta q = 1$ integration. In this process it was found that corresponding peak maxima in the two f(r) functions differed by no more than three ten thousandths of an angstrom.

D. Storage

Storage space is available for two hundred f(r) values (10 A^o), two hundred B values, fifty f(r) values for reinversion up to 2.5A^o, and one hundred modification function

values. Some additional storage space at scattered addresses on the drum is available for modifications.

E. Speed

The program takes about thirteen seconds per M(q) value when evaluating the f(r) function to 10A°. The total times to compute f(r) functions out to various r values are eight minutes for 3.20A°, eleven minutes for 5.2A°, seventeen minutes for 8.00A°, and twenty three minutes for 10.00A°.

F. Equipment

A standard IBM 650 computer with a two thousand word magnetic drum is required. No special accessories are needed.

G. Error Checks

There are no programmed error checks, but any mistake of a serious nature would probably show up as an undamped sine function running through the f(r) function.

H. Input-Output

Input consists of one special card indicating the number of r values at which the f(r) function is to be evaluated, the number of theoretical $M(q)_c$ values to be grafted onto the experimental data, the total number of $M(q)_c$ values used, the index of resolution, the number of f(r) values between the

origin and the beginning of the first molecular feature, and the number of background values to be modified. Each remaining input card contains q - the scattering variable, B - the background value (or 1.00000 if $M(q)_c$ is a theoretical value), a modification function - usually of the form $e^{-\alpha q^2}$ (or unity if no modification function is used), $(M(q)_c - M(q))$ - the constant coefficient correction (or left blank for $M(q)_c$ theoretical values), and either the $M(q)_c$ theoretical value or the experimental $I_{(sector)}$ value.

The radial distribution output cards have eight r and eight f(r) values per card, and the new background cards have eight q and eight B values per card.

If there are two control cards in the program deck that are optional. If they are put on the back of the program deck, the $M(q)_c$ and M(q) values are punched out, one of each per answer card. The punching out of $M(q)_c$ and M(q) does not take additional time, but these values are not particularly useful until the f(r) function is reasonably good.

I. Detailed Operating Instructions

1. Input cards

The first detail card is punched as follows:

Word	Information	Decimal Form	Notes
1	No. of r values	oooxxxoooo	Multiple of 8
2	No. of M(q) values	ooooooooxxx	No restriction
3	No. of theoretical M(q) values	ooooooooxxx	No restriction
4	% resolution	oooooox.xxx	0 - 100
5	No. of f(r) values to reinvert	ooooooooxx	0 - 50
6	No. of background values	ooooooooxxx	Multiple of 8

For each M(q) value a detail card must be made as follows:

Word	Information	Decimal Form	Notes
1	q, scattering variable	ooooooooxxx	Integral values
2	B, background values	oooxx.xxxxx	5 decimal places
5	$e^{-\alpha q^2}$, modification function	x.xxxxooooo	Arbitrary
6	M(q) - M(q), constant coefficient correction	ox.xxxxxxxx	From theoret. model
8	$M(q)_T$, theoretical intensity data	x.xxxxxoooo	From theoret. model
or			
8	I, experimental intensity	oooxx.xxxxx	5 decimal places

2. Output cards

There are three types of output cards: the output of the radial distribution function, the output of the new back-

ground, and the $M(q)_c$ and $M(q)$ values. Formats for these are listed below.

(1) Radial Distribution Output

Word	Information	Decimal Form r	Decimal Form f(r)
1	r_1, $f(r_1)$	x.xx	xx.xxxxx
2	r_2,$f(r_2)$	x.xx	xx.xxxxx
3	r_3, $f(r_3)$	x.xx	xx.xxxxx
4	r_4, $f(r_4)$	x.xx	xx.xxxxx
5	r_5, $f(r_5)$	x.xx	xx.xxxxx
6	r_6, $f(r_6)$	x.xx	xx.xxxxx
7	r_7, $f(r_7)$	x.xx	xx.xxxxx
8	r_8, $f(r_8)$	x.xx	xx.xxxxx

(2) New Background Output

Word	Information	Decimal Form q	Decimal Form B(q)
1	q_1, $B(q_1)$	xxx	xx.xxxxx
2	q_2, $B(q_2)$	xxx	xx.xxxxx
3	q_3, $B(q_3)$	xxx	xx.xxxxx
4	q_4, $B(q_4)$	xxx	xx.xxxxx
5	q_5, $B(q_5)$	xxx	xx.xxxxx
6	q_6, $B(q_6)$	xxx	xx.xxxxx
7	q_7, $B(q_7)$	xxx	xx.xxxxx
8	q_8, $B(q_8)$	xxx	xx.xxxxx

(3) $M(q)_c$ and $M(q)$ Output

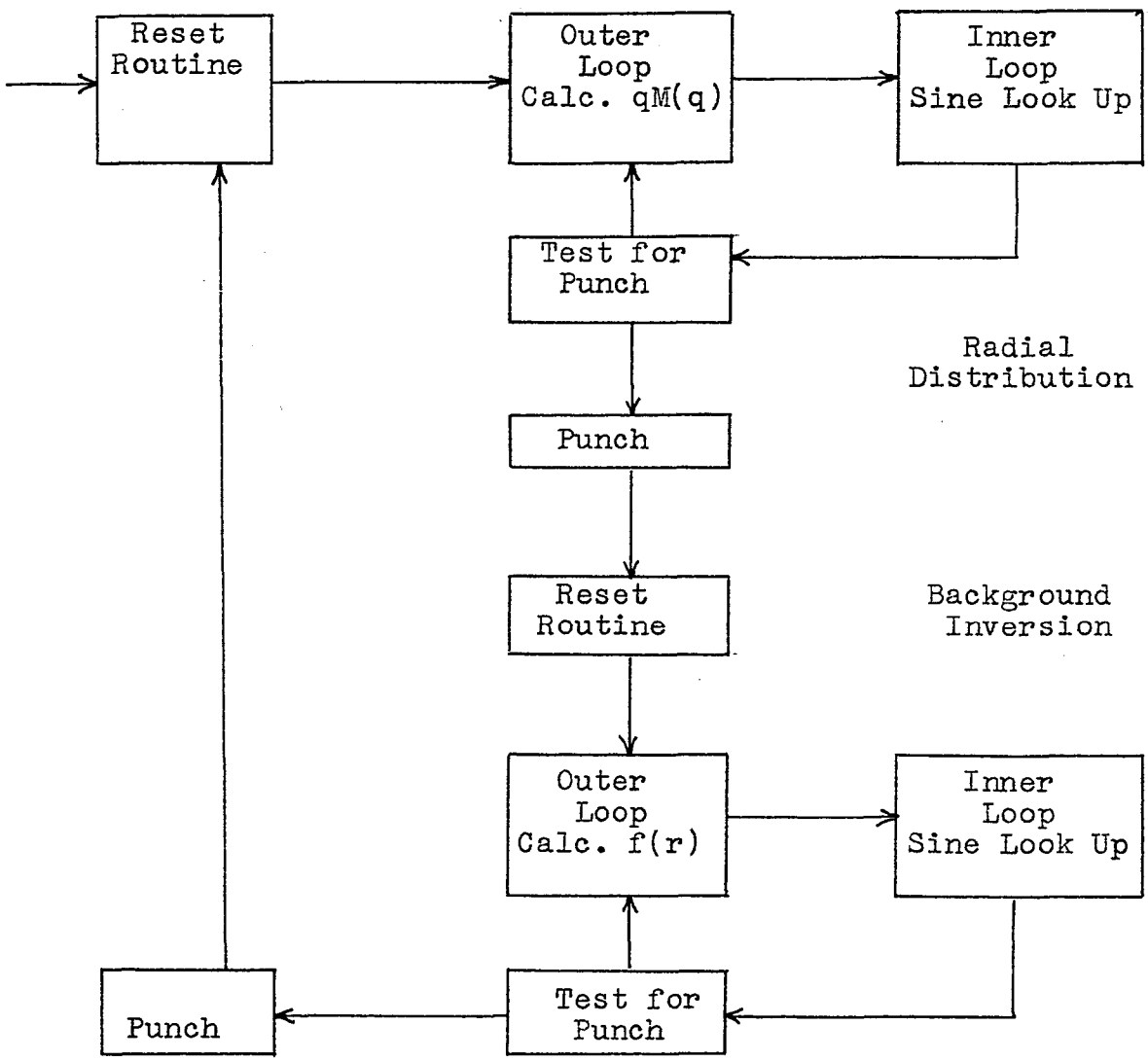
Word	Information	Decimal Form
1	$M(q)_c$	x.xxxxxxxxx
2	$M(q)$	x.xxxxxxxxx

J. Operating Procedure

The console is set at 70 1201 1951. Sense switches are set at stop, and other switches are at normal operating positions. The first two cards in the program deck are drum clear cards that clear the entire drum to minus zeros. The last two cards of the program deck determine whether or not the $M(q)$ and $M(q)_c$ values are to be punched out. Following this, is the first detail card, which is the control card. The detail cards for the intensities follow next. It should be noted that any number of decks of detail cards, each preceded by its own control card, may be placed in the card feed hopper. This is possible since the program, on finishing one calculation, resets itself; and if there are cards in the feed hopper, it will start a new calculation. When the last card has been calculated, "end of file" is depressed. The computer will stop with the input-output light on. The control panel wiring is the so called "Straight in-straight out" wiring, and if a "straight in-straight out" board is not available, the IBM 650 manual for wiring instructions should be consulted.

The detailed program instructions may be obtained from Dr. L. S. Bartell, Chemistry Department, Iowa State College, Ames, Iowa.

FLOW DIAGRAM OF RADIAL DISTRIBUTION PROGRAM



XI. APPENDIX C: REFINED ARBITRARY SECTION PROGRAM FOR THE RADIAL DISTRIBUTION FUNCTION

A. Mathematical Method

This program is essentially the same as the one described in Appendix B in the way that the calculation of the f(r) function is set up. It differs in that the f(r) function may be evaluated between arbitrary limits in the range r=o to r=10 angstroms and is evaluated at finer intervals than in the previous program. The f(r) function is normally evaluated by this program at .025 angstrom intervals. To do this, the same input data that are used in the regular radial distribution program are used as input data in this program except that a new detail control card must be made.

If it is desired to approximate the integration by a summation with Δ q smaller than one, the M(q) experimental curve may be read at half q intervals and the f(r) function started at .05 angstroms or a multiple of .05 angstroms. In this way the f(r) function is evaluated only at .05 angstrom intervals but uses twice as many values of M(q). This change in the choice of Δ q also allows the f(r) function to be evaluated out to 20 angstroms. It should be noted, however, that this f(r) function must be divided by four to be placed on an absolute scale.

The f(r) values obtained by the above scheme with experi-

mental data for ethylene agreed with f(r) values obtained from the regular program within 0.3 per cent in absolute value. Also the maxima of f(r) peaks computed by the two methods of integration agreed to three ten thousands of an angstrom unit. A more rigorous test of the approximations of the integral by the summations would involve the use of synthetic intensity data.

For a detailed discussion of the mathematical method the reader is referred to Appendix B. The arbitrary section program differs from the one cited only in that it does not have a background modification and it uses a sine table of eight hundred values rather than four hundred values.

B. Range and Accuracy

All the computation is performed in fixed decimal form. Decimal points on input and output data are given in the discussion on operating instructions.

The sine values used in this program were calculated using the Bell Lab interpretive routine and are accurate to eight places. The f(r) function calculated with Δq equal to one has the same errors as the f(r) function calculated in the normal radial distribution function. Also the function f(r) may be calculated over any region from zero to ten or zero to twenty angstroms depending on whether Δq is one or one half.

C. Storage

There is sufficient storage space to evaluate f(r) over the complete zero to ten or zero to twenty angstrom range. There is also some additional unused storage space for modifications.

D. Speed

The program takes about ten minutes to evaluate the f(r) function over a three angstrom section.

E. Equipment

A standard IBM 650 computer with a two thousand word magnetic drum is required. No special accessories are necessary.

F. Error Checks

There are no programmed error checks, but any mistake of a serious nature would probably show up as an undamped sine function running through the f(r) function.

G. Input-Output

Input consists of one special control card followed by a detail card for each M(q) value.

The output cards have eight r and eight f(r) values per card. A detailed account of input and output cards

follows.

H. Detailed Operating Instructions

The first detail card is punched as follows:

Word	Information	Decimal Form	Notes
1	Initial r value	x.xxoooooooo	Multiple of .025
2	Number of M(q) values	ooooooooxxx	No restriction
3	Number of r values	ooxxxxoooo	Multiple of 8 added to 1399
4	Number of $M(q)_T$ values	ooooooooxxx	No restriction
5	Index of resolution	oooooox.xxx	0-1.00

For each M(q) value a detail card must be made as follows:

Word	Information	Decimal Form	Notes
1	q, scattering variable	ooooooooxxx	Integral values
2	B, background values	oooxx.xxxxx	5 decimal places
5	$e^{-\alpha q^2}$, modification function	x.xxxxooooo	Arbitrary
6	$M(q)_c$ - M(q), constant coefficient correction	ox.xxxxxxxx	From theoret. model
8	$M(q)_T$, theoretical intensity data	x.xxxxxoooo	From theoret. model
or			
8	I, experimental intensity	oooxx.xxxxx	5 decimal places

I. Output Cards

The radial distribution output cards are as follows:

Word	Information	Decimal Form	
1	r_1, $f(r_1)$	r	f(r)
2	r_2, $f(r_2)$	x.xxx	x.xxxxx
3	r_3, $f(r_3)$	x.xxx	x.xxxxx
4	r_4, $f(r_4)$	x.xxx	x.xxxxx
5	r_5, $f(r_5)$	x.xxx	x.xxxxx
6	r_6, $f(r_6)$	x.xxx	x.xxxxx
7	r_7, $f(r_7)$	x.xxx	x.xxxxx
8	r_8, $f(r_8)$	x.xxx	x.xxxxx

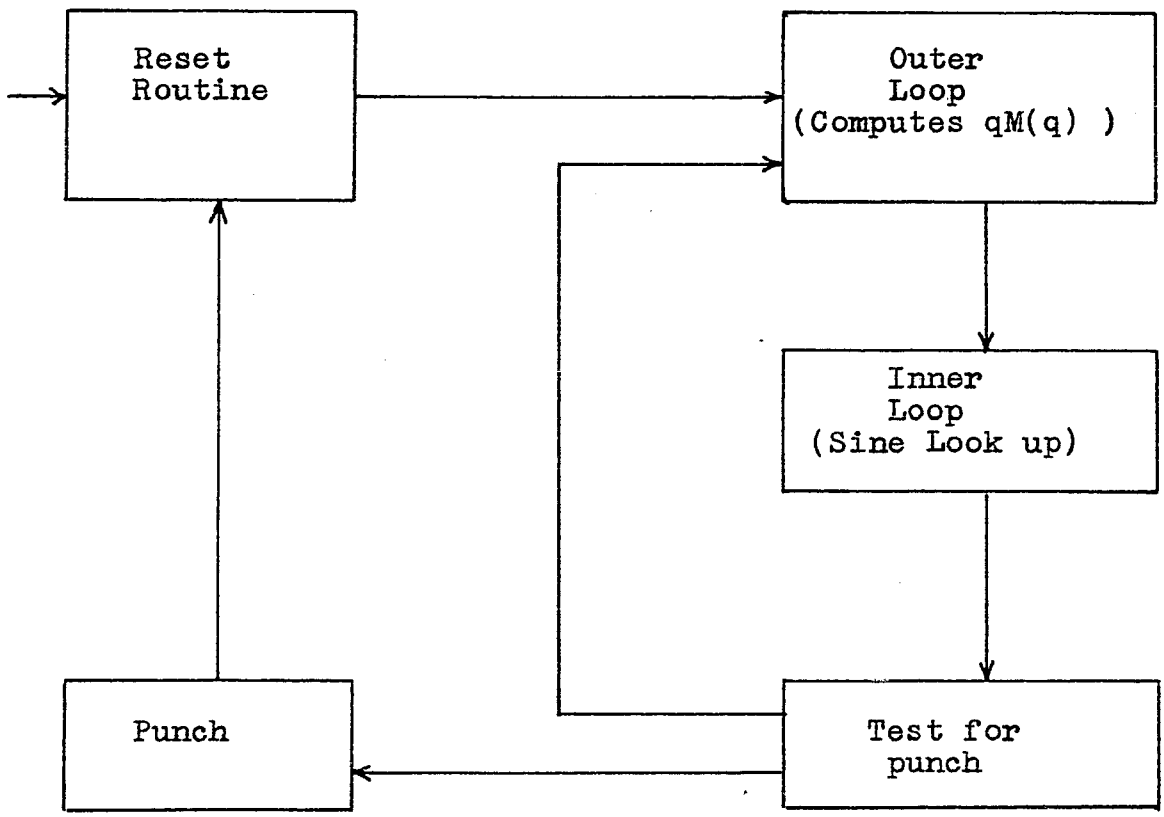
J. Operating Procedure

The console is set at 70 1201 0800. Sense switches are set at stop, and other switches are at normal operating positions. The program deck is followed by a control card and then the detail card deck. It should be noted that any number of decks of detail cards, each preceded by its own control card, may be placed in the card feed hopper. This is possible since the program, on finishing one calculation, resets itself; and if there are cards in the feed hopper, it will start a new calculation. When the last card has been calculated, "end of file" is depressed. The computer will stop with the input-output light on. The control panel wiring is the so called "straight in-straight out" wiring, and if a

"straight in-straight out" board is not available, the IBM 650 manual for wiring instructions should be consulted.

The detailed program instructions may be obtained from Dr. L. S. Bartell, Chemistry Department, Iowa State College, Ames, Iowa.

Flow Diagram of Radial Distribution Program



XII. APPENDIX D: ANALYSIS OF RADIAL DISTRIBUTION CURVES BY THE METHOD OF STEEPEST ASCENTS

A. Description

If the rough structure of a molecule is known, then the problem of analyzing the parts of the radial distribution curve which are independent of internal rotation, may be solved by finding the calculated radial distribution function which comes closest to matching the experimental one. One method of accomplishing this is to determine the parameters for a calculated function minimizing the sum of the squares of the differences between the experimental $f(r)_{exp}$ function and the calculated one at points close enough to characterize the shapes of both. The parameters, θ_i, upon which the calculated $f(r, \theta_i)_{calc}$ function depends are the ℓ_{ij} or root mean square amplitudes of vibration, the r_{ij} or average internuclear distances, and k, a scale factor relating the amplitude of the experimental $f(r)_{exp}$ function to that of the calculated $f(r, \theta_i)_{calc}$ function. It will be assumed that the experimental $f(r)_{exp}$ function can, in general, be characterized by a sum of Gaussian $P_{ij}(r)$ peaks. Often, it is possible to characterize the asymmetry of individual peaks from theoretical considerations and, accordingly, to make corrections and reduce them to Gaussian functions. It should be pointed out that the error in the Gaussian approximation is usually not

much greater than the uncertainty in the $f(r)_{exp}$ function except for molecules in which internal rotations introduce conspicuous asymmetries.

The problem may be stated mathematically as finding values of the θ_i for which the function

$$Y(\theta_i) = -1/2 \sum_r (f(r)_{exp} - f(r, \theta_i)_{calc})^2$$

has a maximum. The summation over r is taken in the present program of 0.025 angstrom intervals. Since $Y(\theta_i)$ is a negative definite quadratic form, it will have an absolute maximum. In practice, if there occur subsidiary maxima or minima, the possibility exists that the method will break down. It will be assumed for this discussion that $Y(\theta_i)$ has no subsidiary maxima or minima sufficiently close to the desired extremum to cause confusion.

If a set of estimated parameters, θ_i (i=1, ...N), are available for a given $f(r)_{exp}$ function that are within ten percent or so of the parameters giving the best fit, we may set up an iterative process for improving the initial estimates. Consider the set of estimated parameters, θ_i° (i=1, ...N). A new set of parameters can be chosen such that

$$\theta_i^1 = \theta_i^\circ + \lambda_i^\circ t \quad (i=1, \ldots N) \quad ,$$

where the λ_i° are the components of the gradient of $Y(\theta_i)$

evaluated at the guesses and t is an arbitrary parameter. It can be shown that, unless the initial guesses are correct, there exist values of t for which the function $Y(\theta_i^1)$ is larger than $Y(\theta_i^o)$ and that, as long as there is only one maximum in the vicinity of the θ_i (i=1, ...N), the values of t leading to greater $Y(\theta_i^1)$ are always positive. The value of t leading to a maximum in $Y(\theta_i^1)$ is approximately given by the equation

$$t = M_o'(o) / \sum_r \left(\sum_{i=1}^{N} \lambda_i^o \, \partial f(r, \theta_i^o) / \partial \theta_i^o \right)^2 ,$$

where

$$M_o'(o) = \sum_{i=1}^{N} \lambda_{io}^{o\,2} .$$

To show this, $f(r, \theta_i')$ is expanded in a Taylor's series expansion to first powers in the parameter t so that

$$f(r, \theta^o{}_i + \lambda_i^o t) = f(r, \theta^o{}_i) + t \sum_{i=1}^{N} \lambda^o{}_{io}$$

$$x \left(\partial f(r, \theta^o{}_i) / \partial \theta^o{}_i \right) .$$

This result is then substituted into the equation for $Y(\theta^1{}_i)$ and the value of t for which

$$M_o'(t) = d/dt \, Y(\theta_i') = \sum_{i=1}^{N} \lambda^o{}_{io} \lambda_{il} = 0$$

is computed. The substitution

$$\theta_i = \theta_i^{\circ} + \lambda^{c}_{io}\, t \qquad (i=1, \ldots N)$$

can be interpreted as a move in the direction of steepest ascent in a hyperspace of N+1 dimensions from the point θ_i° (i=1, ...N) on the line

$$Y(\theta_i) = Y(\theta_i^{\circ})$$

to the point θ_i (i=1, ...N) on the line

$$Y(\theta_i') = -1/2 \sum_r (f(r)_{exp} - f(r, \theta_i)_{calc})^2 \quad .$$

We can now set up an iterative procedure for finding the maximum by generalizing the above equations to

$$Y(\theta_i^{n+1}) = Y(\theta_i^{n}) + M_n'(o)^2/2 \sum_r \Big(\sum_{i=1}^{N} \lambda^{\circ}_{in} \partial f(r,\theta_i^{n}) / \partial \theta_i^{n}\Big)^2$$

and

$$\theta_i^{n+1} = \theta_i^{n} + \lambda^{\circ}_{in}\, t \quad ,$$

where

$$M_n'(o) = \sum_{i=1}^{N} \lambda^{\circ}_{in}{}^2 \quad .$$

This generalization corresponds to moving in the direction of steepest ascent on the surface $Y(\theta_i)$ from the point θ_i° (i=1, ...N) on the line determined by

$$Y(\theta_i) = Y(\theta_i^{\circ}) \qquad (i=1, \ldots N)$$

to the point θ_i' (i=1, ...N) on the line determined by

$$Y(\theta_i) = Y(\theta_i') \qquad (i=1, \ldots N) .$$

Then, the next step is to move from the point θ_i' (i=1, ...N) on the line determined by

$$Y(\theta_i) = Y(\theta_i') \qquad (i=1, \ldots N)$$

along the surface $Y(\theta_i)$ in the direction of steepest ascent to the point $\theta^2{}_i$ (i=1, ...N) on the line determined by

$$Y(\theta_i) = Y(\theta^2{}_i) \qquad (i=1, \ldots N)$$

and so on until the maximum is reached. At the maximum point, $Y(\theta_i)$ is not necessarily zero because of experimental and rounding errors, but the value of $M_n'(o)$ cannot be improved.

B. Computational Method

In this particular application of the method of steepest ascents, t is left as an arbitrary parameter rather than chosen so that it maximizes a particular $Y(\theta_i^n)$. In operation, a set of tests is used to determine whether or not a particular value of t leads to an improved value of $Y(\theta_i^n)$. If it does, then the point determined by this value of t is used as a starting point for another ascent toward the maximum. This process, if continued, leads to the same maximum as outlined above. It is simpler mathematically since t does not

have to be computed but requires more iterations to converge to the maximum.

The tests involved in determining whether or not a given value of t leads to an improvement in the function

$$M_{n+1}(t) = Y(\theta_i^n + \lambda^o_{in}\, t) = Y(\theta_i{}^{n+1})$$

are based on the assumption that in the neighborhood of the maximum the function $M'_{n+1}(t)$ can be approximated by a straight line. At $t = o$, $M'_{n+1}(t)$ is given by the equation

$$M'_{n+1}(o) = \sum_{i=1}^{N} \lambda^{o^2}_{in} \geqslant 0 \ .$$

Therefore, if a calculated value of $M'_{n+1}(t)$ is greater than zero, then the value of t used is likely to be less than the value of t for which $M'_{n+1}(t) = 0$. Thus, any value of t for which $M'_{n+1}(t) \geqslant 0$ leads to a new set of parameters for which the value of $Y(\theta_i^{n+1})$ is an improvement over the value of $Y(\theta^n{}_i)$. If, however, $M'_{n+1}(t)$ is less than zero and less in absolute magnitude than $M'_{n+1}(o)$, the value of t used in computing $M'_{n+1}(t)$ is assumed to lead to a poorer value of $Y(\theta^{n+1}_i)$ and a smaller t is chosen for a new ascent from the same starting point as used before.

Whenever an improved value of $Y(\theta_i)$ is found, a new ascent is started from the point $\theta_i{}^{n+1}$ $(i=1, \ldots N)$ on the line

$$Y(\theta_i^{n+1}) = Y(\theta_i) \qquad (i=1, \ldots N) \ .$$

The function to be maximized is now $M'_{n+2}(t)$, but, in general, it is helpful to make a new choice of t. To aid in finding an improved value of $Y(\theta_i)$, a procedure for varying t is set up so that if $M'_{n+1}(t)$ is greater than zero, then the t value for $M'_{n+2}(t)$ is chosen to be larger than the t value for $M'_{n+1}(t)$ and if

$$0 < M'_{n+1}(t) \leqslant | M'_{n+1}(o) |$$

then the t value of $M'_{n+2}(t)$ is chosen to have a smaller value than the t value in $M'_{n+1}(t)$.

It is possible to include these tests in a program for digital computation such that the parameters, θ_i, leading to a maximum value of $Y(\theta_i)$ can be computed automatically without further decisions by the operator once initial choices of t and the θ_i^o values are made. The initial choice of t in the present program is obtained from the approximate relation

$$t = k^2 / \sum_r (f(r, \theta_i))^2 \quad .$$

In this program the theoretical expression used for $f(r, \theta_i)$ is given by the equation

$$f(r, \theta_i) = k \sum_{j=1}^{M} (c_j / r_j \sqrt{2b + \ell_j^2}) \; e^{-x_j^2/4b + 2\ell_j^2} \quad ,$$

where k is .398944 times the index of resolution, c_j is $(m_{kl} Z_k Z_l / \sum_r (Z_r^2 + Z_r))_j$, where m_{kl} is the multiplicity of

the klth internuclear distance and the Z's are the atomic numbers, b is a damping factor, and x_j is $r-r_j$. The expressions for the λ_j's are then

$$\lambda_k = \sum_r (f(r)_{exp} - f(r, \theta_i)_{calc}) \; f^{(r, \theta_i)calc}/k \quad ,$$

$$\lambda_{r_j} = k \sum_r (f(r)_{exp} - f(r, \theta_i)_{calc})(c_j/r_j \sqrt{2b + \ell_j^2}\,)$$

$$[(x_j/2b + \ell_j^2) - 1/r_j] \; e^{-x_j^2/4b + 2\ell_j^2} \quad ,$$

and

$$\lambda_{\ell_j} = k \sum_r (f(r)_{exp} - f(r, \theta_i)_{calc})(c_j/r_j \sqrt{2b+\ell_j^2})$$

$$[(x_j^2 \ell_j/(2b+\ell_j^2)^2) - \ell_j/2b+\ell_j^2] \; e^{-x_j^2/4b+2\ell_j^2} \quad .$$

To program the method of steepest ascents for the I.B.M. 650 digital computer, the Bell Laboratories interpretive routine was used. This routine allows the programmer to code the program sequentially and in floating decimal form but takes up half of the storage space on the drum leaving only one thousand words of storage for the steepest ascent program.

C. Storage

There is room for the resolution of up to twenty-five radial distribution peaks using up to three hundred and fifty

pieces of experimental data at .025 angstrom intervals.

D. Range and Accuracy

The Bell Laboratories interpretive routine uses an eight place floating decimal subroutine so that all numbers in the range 10^{-50} to 10^{49} are represented by eight significant figures. It appears that the limiting value of $M'(o)$ attainable in most problems lies in the range 1 to .5. This appears to be due to rounding errors in the calculation and also to experimental error in the data. Generally, at this point, the sum of the squares of the deviations is such that the average deviation per experimental point is about one per cent of the maximum value of the f(r) function in the range being fitted. An additional limitation appears to be due to the fact that peaks with small c_j values are not sensitive to change in the presence of peaks with large c_j values.

E. Speed

The program takes about one and one quarter hours to resolve three radial distribution peaks. This involves the computation of about seven sets of f(r) functions at thirty-six or so experimental points. On the average, for thirty-six experimental points the program takes three minutes per cycle for each radial distribution peak.

F. Equipment

A standard I.B.M. 650 computer with a two thousand word magnetic drum is required. No special accessories are necessary.

G. Error Checks

The Bell Laboratories interpretive routine has a number of programmed error stops. For the details, "I.B.M. Technical Newsletter No. 11" should be consulted.

H. Input-Output

Input consists of the experimental f(r) values at .025 angstrom intervals, the initial value of r, and the guesses for the parameters to be fitted. The output consists of the calculated f(r) values for each iteration, the new parameters if desired, and the quantities $M'(o)$, $\sum_r (f(r)_{exp}-f(r)_{calc})^2$, and t.

I. Detailed Operating Instructions

A normal mode of operation Bell Laboratories interpretive deck is used for this program. This deck is placed in the card feed hopper and followed by a drum clear card, which clears the half of the drum not used by the interpretive routine to zeros. Next the load cards for the steepest ascent program are placed in the card feed hopper. These cards

include the input cards mentioned above. The last card of the deck is a transfer card.

For operation, the storage entry switches are set to 70 1951 1331, and the address selection switches are set to 1338. The programmed switch is put in the stop position, and the display switch is set to the upper accumulator. Next, the Bell Laboratories normal mode of operation deck, followed by drum clear card and steepest ascent program deck with transfer card, is placed in the card feed hopper. The computer reset and program start switches are depressed in that order and the card feed started. The program will stop after the first set of f(r) values have been computed, and 1131 will be displayed on the address selection lights. At this point, the number 02020 $\lfloor 50 + \delta \rfloor$ 160 should be dialed with the storage selection switches. The number δ is an exponent which is compared with the exponent of $M'(o)$. If the $M'(o)$ exponent is less than δ , then the parameters used to compute $M'(o)$ are punched out. After the exponent has been set, the program start switch is depressed, and the program runs until the test on the exponent is made. At this time, the program stops to display $M'(o)$ in the storage entry lights and 1120 in the address selection lights. A depression of the program start key causes the program to continue until an improved $M'(o)$ has been calculated. For continuous operation, the programmed switch should be set to run after 02020 $\lfloor 50 + \delta \rfloor$

160 has been entered into storage. This is conveniently done at the second stop to display $M'(o)$.

The board wiring is outlined in the "I.B.M. Technical Newsletter No. 11", and this source should be consulted to familiarize the operator with the interpretive routine.

The detailed program instructions may be obtained from Dr. L. S. Bartell, Chemistry Department, Iowa State College, Ames, Iowa.

FLOW DIAGRAM