Walt. Stafford

Revised Version

Manual for the Use of the High-Speed Equilibrium Ultracentrifugation Computer Program

D. E. Roark

D. A. Yphantis

Manual for Use of the High-Speed Equilibrium Computer

Program of Roark and Yphantis

Brief Description of Program

The program starts with "raw" data from the comparator readings of the photographic plate. With each set of data, a "blank" run is normally submitted. The first step in the program is an interpolated subtraction of the blank from the data.

The next step is the choosing of the zero-concentration level, or, equivalently, the concentration at the meniscus. This choice can be imposed externally, or will be made automatically by the program. Here, we will be concerned with the automatic mode. The method comes from noting that

$$\sigma_w = \frac{1}{c}\frac{dc}{d\xi} \qquad \xi = r^2/2$$

so, $$c = \frac{dc/d\xi}{\sigma_w}$$

Thus, if we know $dc/d\xi$ at some point and σ_w at that point, we can find the corresponding concentration. Although $dc/d\xi$ may be directly calculated, σ_w cannot be. To determine σ_w and c, three approximations are used. In the first two, σ_w is set equal to σ_z in some initial region. In the third, both σ_z and σ_{z+1} are found and σ_w is then set equal to a function of σ_z and σ_{z+1}

$$\sigma_w = \frac{3}{2}\sigma_z - \frac{1}{2}\sigma_{z+1}$$

By use of virial expansions, this function can be shown to be a first approximation to σ_w. By using this technique, it has been possible to choose the concentration of the first data point to within 15% in all test systems examined, and to 5% in many. Thus, even at concentrations at the meniscus as

high as 1,000μ, σ_w at the meniscus can be found to an accuracy of 15%; and the accuracy will increase steadily as one moves down the cell. For instance, if the meniscus concentration is 50μ, then at a concentration in the cell of 200μ, the accuracy in σ_w will be 4%. This technique relieves the need to have a long region of negligible concentration. In many experiments, there will be a very slight gradient at the meniscus, corresponding to a meniscus concentration of only 5 to 10μ. In the standard way of setting the zero-level for a high-speed equilibrium experiment, this would have led to perhaps a 10μ error in the choice of a base-line. With the present program, however, this error can be reduced to about 2μ. Thus, it is possible to find σ_w at low concentrations with considerably more accuracy than could be done previously.

Once the zero level has been chosen and the absolute concentrations generated from the data, the program enters a section that examines the amount of data present. Depending on how much data is available (both the number of points and the concentration span), decisions will be made as to how many different moments of the molecular weight distribution are to be calculated. If the concentration span is small, for instance, the more difficult moments, such as σ_{z+1}, will not be calculated. The number of points and concentration span to be included in the various least-square's fits will also be determined.

The next step is one of data smoothing. Before beginning this, it is best to mention the least-square fitting. The fits are quite varied: linear, quadratic, and cubic are done; some of these include weighting factors for the various points. In general, 15 points and 600μ are fitted at a time; although this quantity will be smaller if only a small amount of data is provided. If an unusually low sigma case is computed, the fit will automatically span 22

points simultaneously. In the case of an abnormally high sigma, there will be an automatic transition to ten points. As far as it is possible, the slopes at each point are found by fitting a set of points which has the point of interest at or near the center of this set. Near the meniscus or near the base, this is, of course, not possible. Here, the number of points fitted will be decreased slightly (but not below 10 points) in order to maintain the point of interest near the center of the fit. Near the two ends, the type of fit will be changed from a cubic to a quadratic.

Data smoothing is accomplished by individual cubic fits for each point (as described above) of ln c vs. $r^2/_2$. The first set of fits is performed for all points and the distances between the points and the fitted curves are calculated. On the basis of these distances, weighting factors are generated. Points nearer the curve are given greater weights than points further away. With these new weighting factors that reflect the confidence in individual points, a second set of individual fits is done for each point. It is the values from these second weighted fits that are used to replace the concentration values of the experimental points.

This smoothing technique has been tested with computer simulated data perturbed by various amounts of random noise. It has been found to result in a 3 to 5-fold reduction in random error with no systematic alteration of the data, (an essential condition for its trustworthiness). In various cases simulating typical and atypical centrifuge experiments, the systematic alteration of data due to smoothing is always less than 3μ from meniscus to base.

Following this smoothing, the various moments are calculated. First the weight-averages are calculated:

$$\sigma_w = \frac{d \ln c}{d \xi}$$

by individual fits at each point.

Then σ_{y2} is calculated:

$$\frac{1}{\sigma_{y2}} = \frac{d\left(\frac{1}{c\sigma_w}\right)}{d\left(\frac{1}{c}\right)} = \frac{d\left(\frac{dz}{dc}\right)}{d\,1/c}$$

by individual fits.

The z-average is generated from σ_{y2}:

$$\sigma_z = \frac{\sigma_w^2}{\sigma_{y2}}$$

Following this, σ_{z+1} and σ_{y3} are calculated:

$$\sigma_{z+1} = \sigma_z + \sigma_w \frac{d(\ln \sigma_z)}{d \ln c}$$

$$\frac{1}{\sigma_{y3}} = \frac{1}{\sigma_{y2}} + \frac{d(1/\sigma_{y2})}{d \ln c}$$

The next step is the calculation of the number-average molecular weight.

The expression used is

$$\sigma_n (I) = \frac{c (I)}{\int_{c(a)}^{c(I)} \frac{dc}{\sigma_w} + T}$$

The integration constant T is

$$T = \frac{c (a)}{\sigma_n (a)}$$

Thus, determining σ_n is straightforward, once one chooses a value for $\sigma_n(a)$ at some early point "a". This point "a" is generally below 100μ.

A standard way to choose $\sigma_n(a)$ is to set it equal to $\sigma_w(a)$. A better way involves the use of σ_w and σ_z, combining them so that their linear combination is equal to σ_n to first order in concentration:

$$\sigma_n(a) = \frac{3}{2}\sigma_w(a) - \frac{1}{2}\sigma_z(a)$$

This has proven quite satisfactory in all test systems examined.

Following this step, the remainder of the charge-independent moments (i.e., those moments not influenced by the 2nd virial coefficient, wherein lie most charge effects) are calculated by combinations of known moments.

With each point-average moment is an expected error estimate. The calculation of these is both complex and reliable. In a selection of over 1,800 test points from many different types of computer simulated data perturbed with random noise, the true value was within the error bars of the computed value for 91% of the points. On this basis, it is reasonable to assume that the error bars represent the expected error to a confidence level of about 90%

It is possible to use a monomer - N-mer fit of the data both in the choice of the zero level, and in the choice of the integration constant for the number-average molecular weight. To do this, both the monomer sigma and the degree of polymerization (e.g., 2 for a dimer) are imputed. However, unless the monomer sigma is known to better than 10% and unless the region near the meniscus (i.e., below 250μ concentration change) is relatively free from contaminants other than monomer and N-mer, the normal mode of operation is probably preferable to the imposed monomer-N-mer fit.

Once the sigmas have been obtained, they may be converted to (apparent) molecular weight averages:

$$M_{app} = \frac{RT}{(1-\bar{v}\rho)\omega^2}\,\sigma$$

All y-moments are independent of the 2nd virial coefficient. Thus, they are independent of the non-ideality due to charge effects (Donnan equilibria) contained in the second virial coefficient. σ_{y4}, σ_{y5} and σ_{y8} are also independent of the fourth virial coefficient. (Charge effects due to Donnan equilibria appear in even virial coefficients, with decreasing effect in higher coefficients. For mild charge effects, the only significant effect is in the second virial coefficient; for stronger effects, the fourth virial coefficient becomes involved). σ_{yo} is useful in that it is independent of the choice of zero-level in concentration. Thus, near the meniscus, particularly if there is not an extensive zero concentration region, σ_{yo} will have much smaller error bars than the other moments. However, at higher concentrations in the cell, the error bars of σ_{yo} will frequently be greater than those of σ_{y2}, for instance.

So called "Midpoint Moments" for all 11 moments are also determined. These are generated by fitting the entire set of data simultaneously. At the center of the fit, the slope can be found quite accurately. The center of the fits will not occur at the center of the cell, nor at exactly the same point for the various moments. The midpoint moments are often useful for examining rather noisy data in which the higher point-average moments become excessively noisy.

Use of the program

Input of Data

1. A heading card is the first card in a set of data. It may contain any desired information (usually starting with an identification number for the experiment).

2. The second card contains the positions of the inner and outer counter weight reference lines and the inverse camera lens magnification factor. The inner and outer reference positions are in millimeters and are taken directly

from the comparator readings of the photographic plate. These positions, and the x-coordinates of the data points need not be transformed into true distances from the center of rotation. This transformation is done internally in the program by assuming that midway between the two reference lines is 6.5 cm from the axis of rotation. A rotor stretch parameter may be used, in which case the midway point is displaced from 6.5 by an amount T. The inverse magnification factor converts photographic plate distances in millimeters into true distances in centimeters. For this reason, the decimal point is so placed that a typical factor is 0.0452. The inner and outer reference line positions, "a" and "b", are placed so that "a" begins at column 5, and "b" begins at column 13. The inverse magnification factor begins at column 21. Thus, a typical card might be:

I4 F8.3 F8.3 F9.6

Col. #	1 2 3 4	5 6 7 8 9 10 11 12	13 14 15 16 17 18 19 20	21 22 23 24 25 26 27 28
		8 2 . 4 5 9	1 1 6 . 1 6 0	. 0 4 5 4 2

If the rotor stretch parameter T is used, it is placed beginning in column 30.

If previously computed values of r are to be used, rather than computer determined values from the reference lines, "a" and "b" may be both set equal to 6.5 and the magnification factor set equal to unity.

3. Next come the set of data points in order of increasing radii. The x coordinate is in millimeters and is taken directly from the comparator, as were the positions of the reference lines. The y-coordinate is in millimeters and is also taken directly from comparator readings of the photographic plate. No subtraction of a zero-level from the y-coordinate is needed, as the program will determine what quantity should be subtracted, automatically. Thus, the "raw" comparator values constitute the input data. The format of the input is to place the x reading in the first eight columns and the y-reading in the second 8 columns. Thus, a typical data card might be

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

The punching of data cards is greatly facilitated by use of a drum card on the IBM key punches. This card will automatically skip spaces not used for data and will copy the decimal points from card to card. It becomes unnecessary to properly space each card individually. Data cards should be very carefully checked for mistakes before being submitted. Failure of the program to work is almost always attributable to a typing error in the input. Partial checking of input data is done in the program. For instance, if the x-values of the data are not monotonically increasing, a "cards out of order" warning will appear in the output and the program will go on with the next set of data. Large errors in the input of the y-values may be found by examining the output from the "Smooth Data" section. Here is listed the new concentrations and the amount of smoothing that was done for each point.

4. On (not after) the last data card, place a "1" in col. 21. This signals the end of the data.

5. After this place any "special purpose cards" that will be described later.

6. After all special purpose cards, or immediately after the last data card if no special purpose cards are used, place a totally blank card. This signals the end of special purpose cards.

7. The next card is the heading card (identification) for the blank corresponding to the above set of data.

I C(2), [illegible] now (5-Nov-86) used to input calculation level & density of 1st point e.g. to set level to 50μ input 2nn1nn0.050

8. Card (2) for the blank, containing "a", "b", and the inverse magnification factor (same format observed as in card (2) above).

9. Data cards for the blank (same format as in step 3 above).

10. Place a "1" in column 21 of last data card.

11. The next card after the last blank data card is a heading card for the next experiment. No empty card appears.

I C(2) not used

Special Purpose Cards

Normal use of the program is without any special purpose cards. The normal mode is the mode almost always used for the analysis of experimental data. However, it infrequently may be desirable to use certain special purpose cards which are detailed here. ← See pp. after 14

I C(1)

1. Monomer-N-mer fit, for choice of zero level and choice of integration constant in the calculation of the number average molecular weight:

N (the degree of polymerization) and σ_1 must be inputed. To do this, place a 1 in column 1, place N in col. 4 (unless N has two digits, in which case it is begun in col. 3), and place σ_1, beginning in col. 6. Thus, if the system is a monomer-dimer system of σ_1 of 4.3, then the special purpose card is

Col. # 1 2 3 4 5 6 7 8 9 10
       1   2   4 . 3

← section removed for PDF 11/03

1.1 Three species plot

1	2	3	4	5	6	7	8	9	10
1	2	∅	∅						

← calls for 3-species output.

2. Input of pre-determined zero-level:

I C(3)

This can be done in two ways. In the first, what is inputed is the desired concentration of the first point. A "3" goes in column 1, a "1" in col. 2 and the concentration one wishes to give to the first point in col. 6 and onward.

Thus, if c(1) is to equal 18μ:

```
Col. # 1 2 3 4 5 6 7 8 9 10
       3 1       0 . 0 1 8
```

The other method is used if the concentrations on the cards are already the desired concentrations and no zero level is to be subtrated from them. This occurs, for instance, when the scanner is used and the O.D. from the scanner is inputed. To accomplish this, place a "3" in col. 1 and a "2" in col. 4:

```
Col. # 1 2 3 4 5 6
       3     2
```

3. To have output punched onto cards:

If the output of this program is to be fed into other programs, it is useful to have the output in the form of punched cards. To do this, a special purpose card is used with a 4 in col. 1 and a 1 in col. 2:

```
Col. # 1 2 3 4 5
       4 1
```

The point number, $r^2/2$, concentration, and all 11 moments plus their error bars will be punched. Owing to the fact that not all moments are punched for each point; and that if the data is insufficient the higher moments are not calculated, the format of the punched output is unfortunately complicated. The first card that comes out is a heading card. The next card is a list of integers that code for just what moments will be outputted for what points.

The easiest way to detail this information is to list below the punch section. In this, "σ_1, $\pm$" will be short notation for $\sigma_1(I)$, $\pm$ error in $\sigma_1(I)$.

```
Write Heading
Write (7, 3 9 0 1) IC3, NI, NM, NM4, NM5, NM2, NM7
If IC3 equals 0, go to 5000
If IC3 equals 1, go to 5001
If IC3 equals 3, go to 5002
```

5000 Write (7,5004) (I, Y, $R^2/2$, σ_n, $\pm$, σ_w, $\pm$, σ_{y1}, $\pm$, σ_z, $\pm$, σ_{y4}, $\pm$, σ_{y6}, $\pm$, σ_{y2}, $\pm$, σ_{z+1}, σ_{y5}, $\pm$, σ_{y7}, $\pm$, σ_{yo}, $\pm$, σ_{y8}, $\pm$, σ_{y3}, $\pm$, I = NI, NN)

(Goes to 5001)

5001 Write (7,5005) (I, Y, $R^2/2$, σ_n, $\pm$, σ_w, $\pm$, σ_{y1}, $\pm$, σ_z, $\pm$, σ_{y4}, $\pm$, σ_{y6}, $\pm$, σ_{y2}, $\pm$, I = NN4, NN5)

(Goes to 5002)

5002 Write (7,5006) (I, Y, $R^2/2$, σ_n, $\pm$, σ_w, $\pm$, σ_{y1}, $\pm$, I = NN3, NN7)

3901 FORMAT (9 1 4)

5004 FORMAT (13, F6.3, F9.3, 6F 7.2 /, 10F 7.2 /, 10F 7.2)

5005 FORMAT (13, F6.3, F9.3, 6F 7.2 /, 8F 7.2)

5006 FORMAT (13, F6.3, F9.3, 6F 7.2)

4. Selection of minimum number of least-square points:

This special purpose card is not recommended; the program has elaborate decision sections which select how many points should be used automatically. These automatic point selections have been tested and are rather crucial for successful operation of the program. If it is wished to override this, however, a "5" is placed in col. #1 and the number of points is placed so that the last digit is in col. #4. Thus, if 19 points are desired: I C (5)

Col. #1 1 2 3 4 5

5 ⌐ 1 9

5. Selection of minimum concentration span for least-square fits.

The same caution against the use of this card applies as against the use of the above card. If, however, it is wished to override the choice of the program; place a "7" in column 1, a "1" in column 2 and the desired concentration span (in millimeters) in column 6 and onward. Thus, if 700μ fits are I C (7)

desired:

```
Col. # 1 2 3 4 5 6 7 8 9 10
       7 1       0 . 7
```

6. No blank

If for any reason no blank is to be submitted with a set of data, a special purpose card should be used. Place a 6 in column 1 and a 1 in column 2.

```
Col. # 1 2 3 4
       6 1
```

In this case, after the totally empty card signifying the end of the special purpose cards, begin with the heading card for the next run.

7. Error Estimate

The program assumes that there will be a random error of about 5% in the input concentrations. If the error is suspected to be larger than this, due to poor plate reading, for instance, a card estimating the standard deviation of the input points can be used. Place an "8" in column 1, a "1" in column 2, and the error estimate (in millimeters) in column 6 and onward. Thus, if the scanner was used and the concentration is inputted in O.D. units and the expected deviation of each point is about 0.01 O.D., then input:

```
Col. # 1 2 3 4 5 6 7 8 9 10
       8 1       0 . 0 1
```

8. Excessively Noisy Data

If data is found to be so noisy that analysis of σ_z, for instance, leads to quite noisy results, a special purpose card can be used which will automatically cause some of the fits to be done with quadratic rather than cubic polynomials. Needless to say, that although the noise is reduced, the resolution will suffer. Thus, one should use this option only in the case of noisy data which cannot be analyzed well by the normal mode. Place a 9 in col. 1 and a 1 in col. 2.

This concludes the list of input features.

We lastly will mention a few comments on the reading of the photographic plates.

In general, it has been found highly advisable to read the plates more densely than has usually been done in the past. The program, by freeing technician time from analysis of the data, makes available more time for careful reading of the plates. For the run, the plate should be read (starting slightly beyond the meniscus so that no irregularities due to the meniscus are picked up) every 200μ on the comparator x-coordinate, until the y-coordinate readings change by more than 10u between successive points. Following this, the plate should be read up to the base every 100μ on the comparator x-scale. At each x-point read, there should be five y-positions read and averaged. These five y-positions should be near the center of the fringe envelope and commonly consist of 3 white fringes and the two black fringes that are contained within the 3 white ones. A check of the quality of the readings can be done by insisting that the value of the center reading be within 10μ of the average. For this purpose, it is convenient to have a calculator next to the comparator so that the averages may be computed point by point. If the average is not within 10μ of the middle reading, the five readings should be repeated. As one moves down the plate with a particular set of 3 white and 2 black fringes, one will gradually move away from the center of the envelope. When this happens, one should jump a fringe to recenter the reading, and add the y-distance corresponding to one fringe to all subsequent readings. This jumping may occur many times during the reading of the plate. The distance corresponding to one fringe may be found by averaging twenty to fifty of such distances on a blank run.

A blank run (at the same speed as the experimental run) should be done both before and after the experiment. The blanks should be read on the comparator every 200μ of comparator-x-scale. Each point should consist of an average of 5 readings (three white fringes and two black ones as described above).

The program may be easily adapted to use with the scanner, rather than photographic plates. The y-input is in terms of O.D. units. The x-input are arbitrary scale values from the scanner output. The values of "a" and "b" on card (2) are in terms of these same arbitrary scale values. The inverse magnification factor, when multiplied by (b-a), should give the true distance in centimeters between the reference lines. This distance may be measured on a comparator. Several special purpose cards are used.

Col. # 1	2	3	4
3			2
6	1		
8			2
9	1		

No blank is inputed if the "6 1" card is used. Each set of data will be analyzed twice: once with the scanners choice of zero-concentration level and once with the automatic choice.

Addendum to the "Manual for the Use of the High-Speed Equilibrium Ultracentrifugation Computer Program" of Roark and Yphantis

The following refers to article #1 under the heading of "Special Purpose Cards" on page 9 of the manual.

Two new methods of using the #1 special purpose card have been added. These new methods are not meant to replace the existing practice of a) running without the #1 special purpose card, or b) using the card to cause a monomer-$\frac{1}{4}$-mer fit to help choose the zero concentration level and the integration constant for the calculation of σ_n; but these new uses offer additional flexibility.

The two new methods of using this special purpose card both rely on virial expansion approximations:

a) Monomer Fit

The monomer sigma is inputed:

Col. #	1	2 3 4	5	6 7 8 9 10
	1			4 . 3

Then, in the section which chooses the zero-level of concentration, σ_w is approximated by $\sigma_w \cong \frac{2}{1/\sigma_1 + 1/\sigma_2}$

In the section which chooses the integration constant for the number average molecular weight

$$\sigma_n \cong \frac{2}{1/\sigma_1 + 1/\sigma_w}$$

Both of these expressions are correct to first power in concentration. For nonideal systems, in which the nonideality is primarily in the linear term in concentration, these approximations appear quite useful.

b) N-mer Fit

Here the degree of polymerization is inputed. In the ordinary use of the program (without special purpose cards) in the determination of the integration constant for the calculation of σ_n, σ_w and σ_z are combined in such a manner that the result is σ_n to first order in concentration. However, in an ideal monomer trimer situation, for example, there will be no second virial coefficient, but an important third virial coefficient. In general, for an ideal monomer-N-mer system the first power of concentration to appear in the virial expansion will be the (N-1) power. Thus, for these cases, it is advisable to correct σ_n not to the linear term in c, but instead to the (N-1) term. This special purpose card accomplishes the correction. For a monomer-tetramer system for instance, we would use:

Col. #	1	2	3	4	5	6
	1			4		

This use of the #1 special purpose card is useful if the system is thought to follow a monomer-N-mer scheme, but the monomer sigma is uncertain.

To sum up, one can do four things, 1) not use the #1 special purpose card, and this will be the most usual choice; 2) input both σ_1 and N, causing a monomer-N-mer fit; 3) input only σ_1; or 4) input only N.