Chapter 4.2.2

CAMEL JOCKEY

An experimental absorptionextinction-misalignment correction or the quintessence of empiricism

H.D. Flack, Laboratoire de Cristallographie, University of Geneva. Switzerland. email: <u>howard.flack@cryst.unige.ch</u>

4.2.2.1 Introduction

The program **CAMEL JOCKEY** was written by H.D. Flack of the Laboratoire Disciplinaire de Cristallographie aux Rayons X of the University of Geneva. This work was supported by the Fonds National Suisse de la Recherche Scientifique, project no. 2.173-0.74. The program was debugged and tested on the UNIVAC 1108 of the Centre Universitaire d'Informatique of the University of Geneva.

The object of the program is to make an absorption, extinction and misalignment correction from intensity measurements on the crystal under study. No detailed knowledge of the shape or mosaic distribution of the crystal is required. The method is thus applicable to crystals of irregular shape or crystals contained within a support (such as a glass capillary with mother liquor, opposed anvils of a high pressure cell, housing of a liquid helium cryostat, one crystal included inside another). The program may also be used to correct for miscentring of the crystal or misalignment of the diffractometer. This feature could be useful in cases where the centring is difficult to carry out (e.g. liquid helium cryostat or crystal inclusions). The present code has no limitations or assumptions with regard to space group, number of measurements or the strength of the absorption-extinction-misalignment. However it assumes bisecting geometry for the main data file, and hence is only of use for data collected with point detectors.

4.2.2.2 Required Intensity Measurements

Two sets of intensity measurements from the crystal are required:

- The standard data collection which is generally the sequential measurement of all reflections in the asymmetric unit of reciprocal space. These intensities will be corrected for absorption-extinction-misalignment from the results of the calculations made with the second set of measurements.
- The CAMEL data collection of sets of intensity measurements which in the absence of absorption-extinction-misalignment would have exactly the same intensity value but which in their presence show differences of intensity. The requirements of a CAMEL data collection are fulfilled by measurements of symmetry equivalent reflections and of azimuthal scans (ψ) of these reflections. Azimuthal scans are obtained by rotation of the crystal about the normal to its reflecting plane which on a 4-circle diffractometer results from a concurrent movement of the circles ω , χ , and ϕ .

A number of *basic* or parent reflections are chosen. For each of these we measure all of the intensity equivalent reflections at all the ψ angles ψ_{min} , $\psi_{min}+\Delta\psi$, $\psi_{min}+2\Delta\psi$, ..., ψ_{max} which are accessible to the diffractometer. See Section **1.12** below for some help on how to choose the parent reflections, ψ_{min} , $\Delta\psi$, and ψ_{max} .

4.2.2.3 The method of finding the CAMEL correction

It is a statistical curve fitting method devoid of a physical model. Let h(ij) be the intensity of the *i* th. measurement before CAMEL correction. *j* indicates that it belongs to the *j* th. set of equivalent reflections (*j* th. parent).

g(ij) be the intensity of the *i* th. measurement after correction for absorptionextinction-misalignment. We can write...

g(ij) = h(ij) * camcor(i)

where **camcor(i)** is the CAMEL absorption-extinction-misalignment correction for the *i th*. measurement. We write

 $camcor(i) = c(0)*f(0i) + c(1)*f(1i) + c(2)*f(2i) + \dots$

where f(n) are trigonometric-power functions of the diffractometer angles ω , 2θ , χ , and ϕ and the reflection intensity, selected to obey the symmetry of the diffractometer, and c(n) are coefficients, calculated by a least-squares method, which give the best representation of the absorption-extinction-misalignment.

This least-squares method tries to make all of the measurements of an equivalent reflection as nearly equal as possible under certain restrictions or constraints which are described in detail below. The same method shows which and how many of the functions f(n) are necessary to describe the absorption-extinction-misalignment.

4.2.2.4 What functions are used for *camcor* ?

A typical function f(n) is

 $f = ((intensity)^{*h})^{*}sin(i^{*}\omega)^{*}sin(j^{*}2\theta)^{*}cos(k^{*}\chi)^{*}sin(l^{*}\phi))$

where h,i,j,k, and l are the orders of the function. The CAMEL method tests sequentially a set of candidate functions likely to be useful in representing the CAMEL correction. If the function is significant it is added to the list of functions to be used for *camcor*, otherwise it is temporarily rejected. The set of candidate functions is examined a maximum of three times and thus a function not included on the first cycle (examination) may be included on a subsequent cycle. The number of candidate functions examined is limited by several means in the program. Firstly the user has to give maximum values for h,i,j,k, and l. Secondly the user has to give a value for the maximum value of q permitted for a candidate function where

q = h/maxh + i/maxi + j/maxj + k/maxk + l/maxl.

Thirdly the user has to indicate the maximum number of examinations-trials to be made. The list of candidate functions is sorted according to the value of q in order that the low order functions are tried first. Normally the low order functions are the most important in the fit. This means that the absorption, extinction and misalignment corrections are carried out simultaneously. A further test in the

program calculates the degree of correlation of a candidate function with the previously accepted functions. If the candidate function is too strongly correlated, it is definitely rejected from *camcor*.

4.2.2.5 A limitation of the CAMEL measurements

The measurement of equivalent reflections by azimuthal scans does not give any information on the variation of *camcor* as a pure function of 2θ . In physical terms this means that these measurements tell us nothing about an absorbing spherical crystal suffering isotropic extinction. Thus the CAMEL program always multiplies the CAMEL correction by the absorption correction for a sphere of radius equivalent to the crystal under study. To correct for the purely isotropic part of the extinction, an isotropic extinction parameter should be included with the atomic parameters in the final stages of structure refinement by least-squares.

4.2.2.6 Constraints

Two types of constraint are used in the CAMEL program.

1. The first constraint is exact. It is...

Σ *camcor(i)* = number of observations

This relation implies a mean correction of 1.0 which in view of the inability of ψ scan measurements to measure variations of absorption or extinction as a function of 20 only, is a physically reasonable choice. This constraint acts as fixing a scale factor for the CAMEL correction. We could equally well have chosen a mean correction of 10.0 (for example). However without this constraint the CAMEL calculates all corrections as zero, giving a sum of squares of zero. This statistically heavenly result is, need we say it, useless physically. This first constraint being exact, it is built directly into the model by letting f(0) be equal to 1.0 and calculating c(0) so that the constraint condition is obeyed.

2. The second constraint is approximate and its 'strength' may be varied from the data cards. The constraint may be stated as follows...

The weighted mean intensity of jth. parent reflection before correction should be approximately equal to this intensity after correction, for all parents i.e. let g(j) be the weighted mean value of the corrected intensity of the j th. parent, h(j) be the weighted mean value of the uncorrected intensity of the j th parent, then

$$h_j \sim g_j$$
 (for all j).
It was found in earlier versions of this program where this second constraint was
not implemented, that often the program CAMEL reduced the sum of squares by
changing, in an unreasonable manner, the parent intensity after correction. With
reflections not in the CAMEL set an unsatisfactory result was obtained. The

constraint is implemented by introducing additional observational equations into the least squares, following Lawson and Hanson. One observational equation is introduced for each parent reflection. This equation has the form

downwt x h_j x $\sqrt{w2_j}$ = downwt x g_j x $\sqrt{w2_j}$

where h_j and g_j are defined above, $w2_j$ is the weight associated with the j th. parent reflection and *downwt* is a downweighting factor to control the strength of these extra observational equations. Clearly *downwt* = 0.0 means this constraint has no effect.

4.2.2.7 Brief description of the mathematical procedures

The function minimized by the least-squares is

 $s = \sum_{i} (w1_i \times (g_{ij} - g_j)^2) + \sum_{i} (downwt^2 \times w2_i \times (h_j - g_j)^2)$

where Σ_1 indicates the sum over the reflection observations, $w1_i$ is the weight of the *i* th. observation, Σ_2 indicates a sum over the parent reflections. This Σ_2 arises from introducing the approximate constraints as additional observational equations.

Doing the alegbra we arrive at the least-squares normal equations. However the program works by transforming the model in terms of the coefficients c and functions f into an othogonalized model in terms of coefficients a. This transformation is called the (modified) Gram-Schmidt transformation and leads to a diagonal least-squares matrix. Once function testing is completed the a coefficients are converted into c coefficients. The program stops trying to fit when the relative error has reached a value specified by the user. In this respect the relative error is defined by ...

error = $\sqrt{(residual sum of squares)/(initial sum of squares about means)}$ A function is considered to be significant in the description of *camcor* if the reduction in the weighted sum of squares it produces is greater than

(current residual sum of squares)/(number of functions wanted - number already accepted)

The various intensity R values printed by the program are defined...

R(j) before =sqrt(sum(((h(ij)-j))/sigma(ij))**2)/sum((h(ij)/sigma(ij))**2))

R(j) after =sqrt(sum(((g(ij)-g(j))/sigma(ij))**2/sum((h(ij)/sigma(ij))**2)))

An expected value of R(j) can be calculated by assuming that only statistical errors are present...

R(j) expected = sqrt(number of obs(j)/sum((h(ij)/sigma(ij))**2)) where sigma(ij) is the standard deviation of the *ij* observation.

4.2.2.8 Weighting schemes

$$w1_i = 1.0 / (h_{ij}^2)$$

 $w2_j = 1.0 / (h_j^2)$

These weighting schemes have been chosen so that strong and weak reflections contribute approximately equally to the weighted sum of squares.

4.2.2.9 Alignment correction and symmetry of diffractometer

A conventional four circle diffractometer is such that a reflection may be measured in one of the following eight angular settings.

1. 2θ, ω, χ, Ø 2. -2θ**.** ω, $\phi + \pi$ -χ, 3. -2θ, χ+π -ω, Ø 4. -ω, 2θ, $-\chi + \pi$, $\phi + \pi$ 5. $\omega + \pi$, -2 θ , ø χ, 6. $\omega + \pi$, 2 θ , -χ, $\phi + \pi$ 7. -ω+π, 2θ, $\chi + \pi$, ϕ 8. -ω+π, -2θ, -χ+π, φ+π

Measurements in these eight angularly equivalent positions should give identical intensities within experimental error if the diffractometer is aligned and the crystal centred. Thus in the *diffractometer aligned* option, the program will produce a *camcor* obeying the above symmetry relations. This is achieved by using only those functions which obey the above relations. Clearly if the crystal is miscentred or the diffractometer misaligned, differences in intensity will occur amongst the eight settings and these may be used in the CAMEL to produce a misalignment correction. In the *diffractometer misaligned* option the *camcor* is made to obey only the following symmetrically equivalent settings.

- 1. ω, 2θ, χ, φ
- 2. ω+π, -2θ, χ, φ

In theory only the first would be necessary for a totally misaligned system. However careful study of the eight positions above will show that (on a Philips PW110, Syntex P21, and Picker) only the first four are mechanically attainable due to restricted movement on the ω circle. Hence for the *misaligned* option we assume a symmetry which cannot be experimentally disproved. The reader is left to decide why this is done.

4.2.2.10 Some computational details

Refer to the program listing for complete information on the way the program is arranged. However three points are worth mentioning.

1. The intensity, parent index and the cosine and sine of the angles ω , 2θ , χ and ϕ of the CAMEL measurements are kept in core. In order to save space limited precision cosine and sine for each angle are packed into one word.

2. The trigonometric functions of higher order angles are calculated by an iterative procedure....

 $cos((n+1)^*a) = 2.0^*cos(a)^*cos(n^*a) - cos((n-1)^*a)$ $sin((n+1)^*a) = 2.0^*cos(a)^*sin(n^*a) - sin((n-1)^*a)$

This procedure saves a considerable amount of computer time.

3. CAMEL controls for itself the size of memory it needs. However if at any moment CAMEL does not have sufficient memory available it simulates the reading of a *maxsize* card with the necessary memory space indicated. At the end of the program the memory size is set back to its original value.

4.2.2.11 Operation of the program

CAMEL should be run before any averaging. The file is read putting into core, the intensity and angular data on each measurement of a parent reflection.

a. Only those measurements with intensity greater than a specified number of standard deviations are used.

b. Those parents of which there are less than a specified number of measurements are eliminated.

c. A list of candidate functions is constructed according to the maximum values of h, i, j, k, l, and q given by the user, and whether the diffractometer and crystal are assumed aligned. The list of candidate functions is sorted according to the value of q.

d. The multi-dimensional fit using the modified Gram-Schmidt transformations

is performed. The fit always includes one parameter for each parent representing the weighted-mean corrected-intensity divided by the maximum intensity in the file. Fitting stops when the desired precision is attained, or when three cycles have been completed, or when the maximum number of functions desired has been included, or when the desired number of trials has been made.

e. The coefficients a in the orthogonalized space are transformed into the coefficients c of the original model. The constant c(0) is also calculated.

f. All intensity measurements are corrected for CAMEL and spherical absorption corrections and put onto a new file. *tbar*, the absorption averaged path length for a spherical crystal is also calculated.

4.2.2.12 How to choose reflections for the CAMEL data collection ?

Two questions of vital importance for the success of CAMEL are :

• What basic or parent reflections to choose...

• What values of ψ_{min} , $\Delta \psi$, ψ_{max} to choose...

Our accumulated wisdom (prejudice) to date shows that the following guide lines should be followed. The basic objective is to cover the angular space (in ω , 2 θ , χ and ϕ) of the standard data collection as thoroughly as possible.

1. Choose parent reflections to have general Miller indices *hkl* to give the largest possible number of equivalent reflections.

2. Choose parent reflections to cover the range of 2θ of the standard data collection. Parent reflections should be in roughly equally spaced intervals of 2θ .

3. If you want to do only an absorption-alignment correction avoid choosing the very strongest reflections as parent reflections.

4. If you want an absorption-extinction correction, double the number of parent reflections, trying at each value of 2θ to choose a strong (or very strong) reflection and a medium-weak reflection.

5. Avoid using very weak reflections.

6. The number of measurements necessary depends on the strength and anisotropy of the absorption-extinction-misalignment. A weakly absorbing-extinguishing almost spherical crystal requires less measurements than a highly absorbing-extinguishing thin platelet.

7. Likewise the number of measurements depends on the improvement in the accuracy (indicated by the intensity R values) desired.

8. Crystals of low symmetry will require more parent reflections, each measured with a smaller $\Delta \psi$ than crystals of high symmetry.

9. For an aligned diffractometer and crystal, the geometry is such that $\psi = \psi + 180$ (in degrees). Hence a normal choice for ψ_{min} is 0 and for ψ_{max} is 179.99. Do not worry that a significant proportion of the desired measuring positions will be inaccessible to the diffractometer due to mechanical restrictions.

10. For a misaligned diffractometer and/or miscentered crystal, it will be necessary to make measurements in the four equivalent angular settings of an aligned diffractometer. Two of the four positions are achieved by letting ψ go from 0 to 360 degrees. The other two require some other diffractometer dependent device.

11. It may be possible in the case of crystals with very high symmetry where there are a large number of equivalent reflections that the ψ -scans would be unnecessary.

12. Try and choose at least one reflection which contains the directions of the greatest anisotropy in the absorption (and if possible, the extinction)

4.2.2.13 How to choose the values of the input parameters ?

The spherical absorption corrections should be for a sphere of radius equivalent to the crystal under study. We have tried three prescriptions to calculate this equivalent radius.

1. the radius of a sphere of volume equal to the crystal under study.

2. the semi-dimension of a cube of volume equal to the crystal under study. $r_{12}(0.125) r_{12}(1/2)$

3. the mean linear dimension of the crystal.

On the crystals we have studied prescription 2 seems to give the best results as judged by the isotropic temperature factors obtained after least-squares refinement of the atomic parameters.

Blanks on the CAMEL card will produce default setting values which can be used in a first run of CAMEL and will usually produce satisfactory results. If the results are satisfactory, in any subsequent rerun of CAMEL on the the same data, reduce the number of pages printed and limit the maximum values of the number of functions, number of trials and maximum h, i, j, k and l to those found in the satisfactory run. This will save a considerable amount of time and some memory space. In case of a preliminary fit which is judged to be unsatisfactory it will be necessary to change the value of some of the input parameters. In the following we attempt to explain the effect on the fit, the R values, machine time and memory space of changing the input parameters.

a. maximum number of functions expected in the final expression.

increasing this parameter will cause a slight increase in memory space (space for this is 3^* max number of functions), and will cause fitting to take longer if the previous fit stopped because the maximum number of functions had been reached. It will also change the condition for acceptance of a candidate function thus allowing functions which make a smaller contribution to the reduction in the sum of squares to be accepted. The latter effect, together with the possible increase in the total number of functions in the final fit will mean a better fit and lower intensity R values.

b. maximum order in the intensity function and the trigonometric functions.

An increase in some or all of these values will cause an increase in memory space which is difficult to predict. Since the number of candidate functions increases, the machine time also increases (linearly with the number of candidate functions). A better fit is not assured since the extra candidate functions may be judged not significant and hence it is easy to waste machine time unnecessarily. It is worth trying to increase a maximum order if a previous fit did not reduce the intensity R values sufficiently and functions were accepted with the previous maximum order in the particular function.

The maximum order in the intensity function controls whether an absorption or an absorption-extinction correction is carried out. Only ask for an absorption-extinction correction if you have made the intensity measure-ments as described above for absorption-extinction. Otherwise the program will almost certainly blow its head. If ψ -scan measurements have not been carried out, there is no information on the variation with ω and hence the maximum order in ω should be set to zero.

c. aligned - misaligned diffractometer and crystal

The aligned option gives a *camcor* obeying the full diffractometer symmetry by limiting the candidate functions. This is the normal situation. The misaligned option gives a camcor with no observable symmetry. To use the misaligned option requires intensity measurments in each of the four possible angular settings. The

misaligned option produces more trial functions for the fit and hence may take longer. The memory space will be increased a little.

d. maximum number of trials(functions) to be made for the fit.

An increase in this parameter has no effect on memory space but controls directly the machine time. It has an indirect effect on the fit and intensity R values by stopping the process before three cycles have been carried out or the desired precision attained.

e. relative error tolerated on the sample fit.

This parameter controls directly the reduction in the intensity R values that may be obtained, given a sufficient number of functions to try. However one should be careful not to overfit. With an overfit the CAMEL reflections will have very small intensity R values but the absorption-extinction corrections for the standard data will be in error by a larger amount. What has happened is that the fit now takes care of the statistical variations in the CAMEL data which do not correspond to an absorption-extinction effect. In order to pass more and more nearly through the experimental points, the camcor makes a tremendous number of wiggles which probably have no physical meaning. We prefer a smoother curve which passes less exactly through the data. Examination of the *expected* R values and *expected* residual sum of squares will allow this situation to be detected. The parameter has no effect on memory size but increases the machine time.

f. control parameter for function choice.

Increasing this parameter has much the same effect as increasing the maximum orders in the intensity function and trigonometric functions. The difference is that increasing the maximum orders will produce more candidate functions with a high order in one of the angles (i.e. higher order axial functions) whereas the control parameter will produce functions where the sum of the orders is higher (more mixed functions).

g. correlation selection parameter.

Increasing this parameter will make the program refuse more functions which are strongly correlated with the best linear combination of those already accepted. This has no effect on memory size, but increases the machine time because it is now more difficult to make a fit. Likewise it will tend to decrease the reduction in the sum of squares. The parameter is useful in eliminating functions which produce excellent values of the intensity R values for an utterly non-physical absorption-extinction correction. The correlation selection parameter is related to, but is not, the multiple correlation coefficient.

h. downweighting factor.

A value of zero of this factor leaves the mean parent intensities after correction free to vary as they like. increasing the value from zero will tie the mean parent corrected intensities more and more down to their values before correction. This enables a more physical result to be obtained. The value of downwt is scaled so that a value of 1.0 doubles the original sum of squares.

4.2.2.14 What about multiple reflections ?

This has as yet (July 1976) never been tried but it should be possible to use the program for this purpose since the variations caused by the multiple reflections obey the symmetry of the functions used by CAMEL. Multiple reflections cause sharp spikes on the ψ scan curves. Hence ψ -scans measured with $\Delta \psi$ very small would be necessary for the program to distinguish between the rapidly varying (due to multiple reflections) and slowly varying (due to absorption-extinction) reflection intensities. The maximum orders of the functions would have to be set rather high.

4.2.2.15 How to know whether the CAMEL correction is good ?

Study the intensity *R* values to see if there has been a suitable reduction for the parent reflections. Look as well at the indices of the accepted functions to make sure that only sensible functions have been chosen, *e.g.* functions *a* or *b* (*h0j00*) should never be present, only functions *a* or *b* (*h0jkl*) should be present if ψ -scans have not been carried out, only functions *a* or *b* (*0ijkl*) should be present if only an absorption correction is wanted. Look as well at the mean parent intensities before and after CAMEL correction. For each parent the values should be about equal. The value of the constraint constant should be approximately 1.0.

Study also the expected residual sum of squares (based only on measurement errors) and the residual sum of squares.

4.2.2.16 Using CAMEL JOCKEY in *WinGX*

As stated above, two files are required for CAMEL JOCKEY, a file containing the psi-scans and the file containing the data to be corrected. Since both of these files must contain the diffractometer angle settings for each reflection, the standard SHELX-style reflection files are of no use. In the *WinGX* environment, these two files are called PSISCANS.DATA and REFLECTIONS.DATA. They have exactly the same format, each record containing the following information :

h, k, l, 2 θ , ω_{eul} , χ_{eul} , ϕ_{eul} , ϕ_{kappa} , ψ_{eul} , ψ_{xtal} , F^2 , $\sigma(F^2)$, (s0(i),s1(i),l=1,3)

This should be read in the Fortran format (3i4,8f8.0,2f10.0,6f8.5). The direction cosines *s*0,*s*1 are as defined in SHELX (*i.e.* referred to the reciprocal axes) and are optional. They are not used at all when reading PSISCANS.DATA and if present in the REFLECTIONS.DATA file are merely copied verbatim to the corrected output SHELX style reflection file. The format for PSICANS.DATA is identical to that used for the PSIABS program, and both files may be prepared from CAD4 diffractometer

files using the program XCAD4. Since both PISABS and CAMEL JOCKEY assume the main reflection data have been collected in bisecting geometry (*i.e.* with a point-detector), they are quite useless for data collected with area-detectors. CAMEL JOCKEY uses only the Eulerian angles in the files, while PISABS uses the kappa angles. The crystal-based azimuth angle ψ_{xtal} , as defined by Schwarzenbach & Flack (1989), is not currently used by either program.

The only other information required is the linear absorption coefficient (in mm⁻¹), the mean crystal dimension in mm (for θ -dependent correction) and the Laue symmetry used for determing the equivalent reflections. These are entered in the dialog box shown below.

K CAMEL JOCKEY Control Pane	el	X
Absorption coefficient mu (mm-1) Mean crystal dimension (mm)	0.234 🗢 2/m mmm 4/m	 ▲ ■
	<u>O</u> K <u>C</u> ance	el

The corrected data are written to the file CAMEL_JOCKEY.HKL and a summary of the corrections applied is given in CAMEL_JOCKEY.LST

4.2.2.17 References

Flack, H. D. (1974) Acta Cryst. A30, 569-573
Flack, H. D., Moreau, J.M. and Parthe, E. (1974) Acta Cryst. B30, 820-821
Flack, H. D. (1975) J.Appl.Cryst. 8, 520-521
Flack, H. D. (1977) Acta Cryst. A33, 890-898
Lawson, C.L. and Hanson. R. J. (1974) Solving Least Squares Problems, pp 148-157
Prentice Hall Inc.
Wind, H. (1972). Proc. 1972 CERN Computing and Data Processing School, CERN 72-21, pp 53-106.
Schwarzenbach, D. and Flack, H. (1989) J. Appl. Cryst. 22, 601-605.