

CAMEL JOCKEY, An Absorption Correction Program

BY HOWARD D. FLACK

*Laboratoire de Cristallographie aux Rayons X, Université de Genève, 24, Quai Ernest-Ansermet,
1211 Genève 4, Switzerland*

(Received 3 March 1975; accepted 5 May 1975)

A greatly improved version of the program *CAMEL JOCKEY* for the absorption correction method of Flack [*Acta Cryst.* (1974). A30, 569–573] is described. The program is written in FORTRAN for a minicomputer.

Introduction

In a recent paper (Flack, 1974) a practical method was described for the calculation of absorption corrections with intensity measurements from azimuthal scans. This method gave good results in the correction of intensity data from SmAu₆ (Flack, Moreau & Parthé, 1974). Since the time of writing the previous paper, the computer program *CAMEL JOCKEY* used to implement these absorption corrections, has been considerably revised and improved, and is now available for distribution. In modifying our program, we were particularly interested in investigating the possibility of running the program on a minicomputer similar to those used in an automatic four-circle diffractometer. With such a goal attained, it is possible to envisage an *in situ* absorption correction carried out directly by the diffractometer.

Method

The theory of our absorption correction method has already been described by Flack (1974). This basic approach has not been changed. However a problem of practical importance with the previous versions of *CAMEL JOCKEY* was the necessity of fixing the limits on the values of i, j, m and $i^2 + j^2 + m^2$ at the outset. These integers describe indices of coefficients in the Fourier expansion of the normalized transmission factor. With low values on these limits a good representation of the observed transmission factors is not produced, whilst with high limits strong correlations between the parameters are obtained, giving rise to severe problems of precision on a small-word-length computer.

We have overcome these difficulties by making use of a variance ratio test (*cf.* Hamilton, 1964). The procedure is as follows: We suppose that the transmission has been represented by $p-1$ parameters (Fourier coefficients) in a linear least-squares fit. This has given a weighted sum of squares R_{p-1} . Another possible parameter (the p th) is chosen and the transmission factors are fitted to all p parameters to give a new weighted sum of squares R_p . Hamilton (1964) shows that the null hypothesis

$$H_0: x_p = 0$$

where x_p is the p th parameter, may be tested by computing the variance ratio

$$(n-p) \frac{R_{p-1} - R_p}{R_p}$$

and comparing it with $F_{1, n-p, \alpha}$ where n is the number of observations. Depending on whether this ratio is significant or not, the parameter is accepted or rejected and the process is continued by considering another trial parameter. We thus start with no parameters and build up our Fourier representation with those parameters which are significant. The parameters are tested in an order which we expect *a priori* to represent their significance – possibly most significant being first.

Strictly speaking our statistical test is wrong. We are using *multiple* monivariate tests as a *multivariate* test. The situation is saved by choosing a suitably low value of the significance level α and by retesting those parameters which were indicated not significant the first time. This retesting is carried out once all the original parameters have been tested and in fact we cycle over the parameters three times as do Wind (1972) and O'Connell (1974).

For each parameter considered, the value of $1 - R_1^2 = 1/b_{11}s_1^2$ in the notation of Hamilton (1964) is also calculated. R_1 is the multiple-correlation coefficient of the trial parameter with those parameters previously accepted. If this value (R_1) is larger than an amount fixed in the program the parameter is 'killed' – no longer to be considered as a possible parameter. Such a test reduces the size and number of large terms in the correlation matrix and prevents difficulties of precision due to rounding errors.

The variance ratio and the multiple correlation tests are sufficiently powerful that it is now possible to consider a_{i00} ($i=0, 1, 2, \dots$) as parameters to be tested. Previously it was necessary to fix $a_{000} = 1$ and the others to 0, in order to avoid rounding errors.

A further problem with the original *CAMEL JOCKEY* was that some interpolated values of the transmission factor which were to be used as an absorption correction turned out to be negative. We have tried unsuccessfully to build constraints into the Fourier expansion of the transmission factor in such a way that it yielded only positive values. Nevertheless the solution adopted in our current *CAMEL JOCKEY* is to fix on $(A')^{1/2}$ whence all interpolated values of A' are positive. Likewise a system of unit weights is used. The system of $1/\sigma_A^2$ was found to be unsatis-

factory as forcing the observations in the medium range of θ to dominate the least-squares fit. In consequence the calculated corrections at higher θ angle were unreasonable.

Further remarks

A crude test has been carried out to see if it is possible to reduce the number of intensity measurements necessary for an absorption correction. For this purpose, the ψ -scan data of SmAu₆ (Flack, 1974) were limited to those intensity measurements with $\psi=0$. Working under identical conditions to Flack (1974) and Flack, Moreau & Parthé (1974), structure refinement produced an R value of 20.2%, using only a spherical absorption correction, 16.6% with the limited ψ -scan data compared to 12.8% with the full ψ -scan data. The unfortunate conclusion is that, even in the favourable case of SmAu₆ with 16 equivalent reflexions a certain amount of ψ scanning is necessary. However recent developments in the control software of four-circle diffractometers make these measurements automatic (*cf. User Manual PW 1100, 1974*).

Computational details

Table 1 shows information on those computers on which the program has been run. *CAMEL JOCKEY* is a collection of nine small programs (called *CAM1*, *CAM2* etc.) which have to be run successively. No overlay structure is necessary. Intermediate results are in binary on disc files. Owing to the number of files (9) used, magnetic tapes would not be suitable. All calculations are in single precision.

Table 1. Information on computers on which *CAMEL JOCKEY* has been run

Manufacturer	Machine	Store used in K	No of bits for F.P. number	Compilation + execution time
DGC	Nova 840*	16	32	~1.25 h
CDC	3800	6	48	3 min 55 s
CDC	7600	4	60	4.96 s
UNIVAC	1108	6	36	6 min

* The Nova we used has non-standard peripherals of a card reader and a line printer.

The program is written in FORTRAN following the prescriptions of the X-RAY system (1972) (Stewart, Kruger, Ammon, Dickinson & Hall, 1972) for pidgin Fortran. This makes the program easily adaptable on any machine having a Fortran compiler. There is one notable exception to the pidgin Fortran for the Nova where the binary read and write instructions are:

READ BINARY (i) L and WRITE BINARY (i) L
and not

READ (i) L and WRITE (i) L

as usual. All such instructions are clearly marked in the source deck.

The source deck consists of ca. 2300 cards with 196 test data cards. The program is fully commented in the source deck which also contains a description of data cards and a glossary of symbols.

The present program is limited to 70 trial parameters of which up to 30 can be accepted. Those parameters (Fourier coefficients) a_{ijm} or b_{ijm} with $i \leq 2$, $j \leq 7$, $m \leq 7$ and $i/2 + j/7 + m/7 \leq \text{QMX}$ with $\text{QMX} = 1.22$ are considered on trial. The program stops trying parameters if one of the following conditions is achieved:

- Three cycles of trials completed;
- 30 parameters accepted;
- Value of $(R_p/n-p)^{1/2} \leq 0.04$.

The program uses approximate values of the F statistic for $\alpha = 0.10$

$$F_1, n-p, \alpha = 104.0/(n-p) + 2.71.$$

The multiple-correlation coefficient is tested by killing those parameters with

$$1 - R^2 \leq 0.15.$$

Discussion

In order to calculate the weighted sum of squares and the multiple-correlation coefficient for each trial parameter, it is necessary for the *CAMEL JOCKEY* to invert the matrix of the normal equations. This is necessarily a time-consuming procedure. There exists a method (*cf. Wind, 1972*) which avoids the matrix inversion step by making use of the modified Gram-Schmidt transformations. We have seriously considered using this technique but have found that it is more costly in terms of memory space than the matrix inversion procedure. On the minicomputer it would not have been possible to consider a sufficient number of parameters.

The author wishes to thank Dr H. Wind for an illuminating discussion on the subject matter of this paper. The financial support of 'Le Fonds National Suisse pour la Recherche Scientifique' was welcome under research contract No 2.1730.74.

References

- FLACK, H. D. (1974). *Acta Cryst.* A **30**, 569-573.
 FLACK, H. D., MOREAU, J. M. & PARTHÉ, E. (1974). *Acta Cryst.* B **30**, 820-821.
 HAMILTON, W. C. (1964). *Statistics in Physical Science*. New York: Ronald Press.
 O'CONNELL, M. J. (1974). *Comput. Phys. Commun.* **8**, 56-69.
 STEWART, J. M., KRUGER, G. J., AMMON, H. L., DICKINSON, C. & HALL, S. R. (1972). *The X-ray System-Version of June 1972*. Technical report TR-192. Computer Science Center Univ. of Maryland.
User Manual PW 1100, (1974). Revision Sept. 1974. Eindhoven: Philips.
 WIND, H. (1972). *Proc. 1972 CERN Computing and Data Processing School*, Report CERN 72-21, pp. 53-106.