also applicable to the evaluation of diffraction patterns of the matrix and of precipitates and is layed out for arbitrary crystal systems.

References


**MITHRIL – an integrated direct-methods computer program.** By C. J. GILMORE, Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland

(Received 28 June 1983; accepted 11 October 1983)

**Abstract**

A new direct-methods computer program, **MITHRIL**, is described. Written in a neutral subset of Fortran IV, it is built around a heavily modified **MULTAN80** system. It incorporates many recent theoretical developments in direct methods including the use of quartet and quintet invariants, a new method for estimating triplets, **YZARC** and **MAGEX**, and random-phase tangent refinement. It can be run as a menu-driven interactive real-time package or in the more conventional batch mode. Several levels of user–program interaction are provided.

**Introduction**

The need for a new direct-methods computer program arises in part from recent theoretical developments in the subject. **MULTAN80** (Main, Fiske, Germain, Hull, Declercq, Lessinger & Woolfson, 1980) incorporates magic integer theory, tangent formula weighting and early figures of merit. It can be readily interfaced to the **MAGEX** (Hull, Viterbo, Woolfson & Shao-Hui, 1981) and **YZARC** (Baggio, Woolfson, Declercq & Germain, 1978) programs. **SIMPEL** (Overbeek & Schenk, 1978) uses quartets and a further development in tangent refinement techniques in a symbolic addition environment. **SHELXTL** (Sheldrick, 1981) also uses quartets as an important figure of merit and provides features such as trio relationships and two-phase seminvariants. **RANTAN** (Yao Jia-Xing, 1981) is a **MULTAN80** package in which phase permutation is replaced by the assignment of random phases to unphased reflections. **QTAN** (Langs & Hauptman, 1975) also uses quartets as figures of merit and the **MDKS** formula (Hauptman, 1972) as a method of filtering out possible aberrant triplets. One purpose of **MITHRIL** is to take the best features of these packages and unite them in a single program, at the same time adding new options such as quintet invariants, the third-neighbourhood quartet formula and the active use of higher invariants.

There is an additional need for a new program linked to new developments in computer hardware. The 32-bit supermini computer is now widely used in the crystallographic laboratory, and as its price falls this use is likely to grow. Such a machine can provide the user with multitasking capability, but in a suitable laboratory environment it allows real-time interactive use, so that it is possible for the crystallographer to influence the flow of analysis as it proceeds. These facilities are not readily available in current programs. At the same time, of course, it should be possible to run the package in the more usual batch environment. Cheap low-resolution graphics facilities are also widely available, and a comprehensive program should provide simple graphics facilities, particularly for the interpretation of **E** maps.

Finally, it must be remembered that there are still many structures which present difficulties to direct methods, so that it is essential to have a program at one’s disposal that offers as many options as possible.

**The modules of the MITHRIL program**

The name **MITHRIL** is an acronym for Multan with Interactive facilities, Triplet checking, Higher invariants, Random phasing, Intelligent control of flow and options and Linear equations. This describes the system quite succinctly. A flowchart of the program modules is shown in Fig. 1. In this section each module and its facilities is

0021-8898/84/010042-05$01.50 © 1984 International Union of Crystallography
described in detail, but first there is a general outline of the user-program interface.

Four levels of user interaction are provided, ranging from the batch mode to a full dialogue requiring user decisions at key points. This level is set dynamically and can be changed at any time, so that some modules may be run interactively and others under complete program control. The interactive modes are menu driven with a separate menu for each module outlining available commands. Each instruction consists of a four-character key word and a free-format list of input fields. Null fields generate default options. Each command has a sensible default value so that the user need only enter non-default options. Furthermore, there are four levels of default complexity, designed to cover all situations between the simplest and most difficult problems. These default levels are also dynamic and may be altered from module to module. Finally, the user need only specify those modules for which non-default options are to be used, the program will automatically run any modules that are needed but have not been explicitly called by the user.

NORMALIZATION: This is an extensive modification of the MULTAN80 program. The normalization procedure is often neglected by crystallographers since it is so automatic, but decisions made at this point can have drastic implications for subsequent steps in the analysis, and full control over the process may be needed. Accordingly, the package offers the following features.

1) Equivalent reflections and systematic absences are removed. The former distort the phasing procedure, while the latter can render relationships which rely on small E magnitudes incorrect. In this latter category are the \( \varphi_0 \) triplets and the negative quartets and quintets.

2) Editing facilities are provided to remove or modify structure factors before normalization, or to remove or modify the subsequent E magnitudes. The former facility alters the normalization process, whereas the latter modifies only some of the normalized structure factors. Allied to these provisions is a set of optional \( \delta \) limits and a maximum permitted E magnitude. Experience has shown that E magnitudes \( \geq 3 \) prevent the weighting schemes in a weighted multi-solution phasing environment from working with optimum efficiency – they tend to drive most weights to unity very rapidly. The relatively crude device of setting an upper limit can often remove this problem. The \( \delta \) limits can be useful to exclude low-angle data which are subject to large systematic error, or the high-angle reflections which can be very sensitive to small changes in overall temperature factor.

3) Allowance is made for a lack of knowledge of the detailed contents of the unit cell.

4) Should the traditional K-curve-Wilson plot techniques prove inadequate, direct input of E magnitudes is permitted and the normalization procedure is bypassed. This allows interfacing to other normalization procedures.

TRIPLETS: The Cochran distribution (Cochran, 1955) is used with the addition of two formulae for independently checking the triplet cosine. These are the MDKS formula (Hauptman, 1972) and a related technique (LE) in which a quintet extension of a triplet is used to derive a joint conditional probability distribution involving six E magnitudes (Hauptman, unpublished results). This is manipulated to give a system of ten simultaneous linear equations in which the triple-phase invariant is one undetermined variable and can be calculated in the usual way. Both these methods give only approximate estimates for the cosines, but they can be used to alter the relative weights of the triplets and for indicating which relationships may be troublesome. This facility provides one link with the QTAN system.

The triplets are searched for those which give indications of the phases of one-phase seminvariants. If MDKS or LE has been used, an analysis of these triplets in terms of the estimated cosine is also given. This can be a useful adjunct in the decisions concerning the reliabilities of the \( \varphi \) relationships.

QUARTETS: Quartet invariants of the form

\[ \varphi_h + \varphi_a + \varphi_t + \varphi_{-h-k-t} = \varphi_4 \]

are a very important component of MITHRIL. Three types of quartet must be distinguished – that for which \( \cos \varphi_4 \) is estimated to be zero (a positive quartet), one in which \( \cos \varphi_4 \) is estimated to be \( \pi \) (a negative quartet), and the enantiomorph-sensitive relationships between these two extremes. The negative quartets in particular are very useful. They can be used both as a figure of merit NQEST (De Titta, Edmonds, Langs & Hauptman, 1975; Gilmore, 1977) and in an active mode to generate new phases (Freer & Gilmore, 1980). Two formulae are provided: the seven-magnitude second-neighbourhood formula, and the more powerful third-neighbourhood, 13-magnitude formula (Hauptman, 1977a, b). Missing members of these neighbourhoods are permitted. The missing magnitudes are assigned values of unity. The use of the second-neighbourhood formula is now widespread, but the third-neighbourhood formula is still somewhat neglected. It introduces a floating vector \( p \) and a related vector \( q \) such that

\[ p + q + h + k = 0 \]

whilst \( |E_p| \) and \( |E_q| \) are 'large'. This gives rise to a multiplicity of neighbourhoods and allows a series of semi-independent estimates to be made for each quartet; any relationship exhibiting discrepancies is discarded. This involves an increase in computer time, but it can be partly reduced by limiting the vectors \( p \) and \( q \) to span only the top 50–100 \( E \) magnitudes. The larger number of magnitudes involved in the third neighbourhood often produces invariants of much lower variance than the second-neighbourhood formula, and it can also alleviate the \( 1/N \) dependence of quartet reliability (\( N \) is the number of atoms – assumed equal – in the unit cell). This can be valuable for poor-quality data sets and especially large problems (Gilmore, Hardy, MacNicol & Wilson, 1977). The use of the third neighbourhood is related to the generation of trios although it puts such relationships on a quantitative basis.

Positive quartets may also be generated if requested. They can sometimes be useful in situations where there is a deficiency in the number of triplets available, but the use here is somewhat experimental. Positive quartets are correlated with triplets (Giacovazzo, 1980) and this correlation is dependent on the \( E \)’s involved. It is handled in the way described by Gilmore (Freer & Gilmore, 1980). Higher invariants are put on the same scale as triplets using the concept of an equivalent \( k \) which is obtained from the
variance of the appropriate probability distribution (Freer & Gilmore, 1980).

The negative quartet module is always called in situations where the space group is symmorphic. Negative quartets are also generated in cases where difficulty in solving the structure is indicated by the user.

**QUINTETS**: There are no completely reliable quintet distributions in the current literature, and quintet theory needs to be further developed. However, quintets can be valuable in situations where negative invariants are needed but where quartets are difficult to generate, since the second neighbourhood of a quintet involves ten cross terms, not all of which need to be ‘small’, in contrast to the equivalent quartet formula which requires three small cross terms. However, quintet reliability has a $1/N^{3/2}$ dependence so there are inherent difficulties in deriving accurate relationships. Only negative quintets are generated. The distributions are calculated in determinant form using both the Giacovazzo formalism (Giacovazzo, 1977) and that proposed by Hauptman & Fortier (1977). For acceptance, both distributions must indicate a reliably negative invariant. The maximum number of missing neighbours permitted is under user control.

**CONVERGENCE MAPPING**: The convergence mapping module follows invariant generation. It performs two quite distinct functions.

1. The collection of invariants: all the invariants generated by previous modules are loaded for active use in the phasing procedures which follow, although it is possible to exclude higher invariants. Triplets are optionally weighted via their MDKS or LE cosine estimate, if it is available. A rather simpler method than that used by Busetta & Comber (1974) is employed. The invariants $\varphi_3$ are split into four classes:
   
   - (i) $\langle \cos \varphi_3 \rangle \geq 0.7$
   - (ii) $0.7 > \langle \cos \varphi_3 \rangle \geq 0.0$
   - (iii) $0.0 > \langle \cos \varphi_3 \rangle \geq -0.7$
   - (iv) $\langle \cos \varphi_3 \rangle < -0.7$

   Each class (i)–(iv) is assigned a weight $\geq 0.0$ by the user, and this is used to multiply the $\kappa$ value for the relationship, where
   
   $$\kappa = 2 |E_k E_k E_{k-1}| / N^{1/2}$$

   ($E_k$, $E_k$, $E_{k-1}$ are the $E$'s involved in the triplet). By giving relationships in class (i) weights $> 10$ they can be upweighted and play a larger role in the phasing; in a similar way, those in class (iv) can be downweighted or removed completely. The MDKS and LE formulae are unreliable in the two remaining classes and unit weights are usually used here. It is worth emphasising that this weighting scheme radically alters the phasing path as decided by the convergence method, and even if the MDKS estimates are unreliable, the resulting convergence map may be sufficiently different from the original that previous problems may disappear. In particular, unreliable triplets which appear early in the convergence map, even if not detected by the MDKS/LE tests, may now appear in a much less critical area of the phasing.

   Specific relationships may be deleted from or added to the list. Any two-, three-, four- or five-phase invariant or seminvariant may be added. This can be particularly useful as an adjunct to symbolic addition (Karle & Karle, 1966) where indications of relationships between symbolic phases or of potentially unreliable invariants can emerge. This information can now be supplied to a multisolution process in a simple way. The user must estimate the reliability of any input relationship by supplying a $\kappa$ value.

2. Convergence mapping: once the relationships have been suitably ordered, the convergence map chooses suitable reflections which define the origin, the enantiomorph (if relevant), the known phases and the permuted phases. The user is provided with the usual options to define or partially define origin and enantiomorph, the permuted reflections, the maximum and minimum number of phase sets and the values of any known phases. Three criteria are applied in the acceptance of $\sum$, results: (i) A minimum probability; (ii) a minimum number of indications $\sum \pm \leq N +$, $N -$ (where $N +$ is the number of zero indications and $N -$ the number of $\varphi$ indications). This is usually $3$. (iii) A consistency ratio (c) defined as

   $$c = \frac{\max(N +, N -)}{[N + + (N -)]}; \text{ usually } > 0.9.$$  

   Caution in accepting a $\sum$ determined phase cannot be overemphasised.

   Care is also needed with the active employment of quartets. It is obvious that, with a triplet, two known phases can derive a third, whereas, for a quartet, three known phases are required to give a fourth. If the phase angles are very approximate, as they are in the early stages of direct methods, quartets will tend to propagate errors more than triplets. The situation is even worse with quintets. Accordingly, the number of higher invariants should be a maximum of ca 20% of the total number of relationships used.

   At this point the negative invariants are assembled for use in NQEST and its quintet related form NQINT. It is important to use reliable invariants even though they are being used only for figures of merit. Thus only relationships with an equivalent $\kappa \geq 1.0$ are accepted, and a minimum of 25 invariants is needed for either NQEST or NQINT.

   **YZARC**: Usually one proceeds directly to tangent refinement from convergence mapping, but MITHRIL offers two additional modules which can be run between them. The first of these is YZARC. It uses sets of random phases as a starting point and refines them via least squares or steepest descents. Usually only a subset of reflections is phased in this way and these phases are passed to tangent refinement. Normally some 50–100 different sets are processed in this way. Some of the facilities offered in MITHRIL differ from the standard YZARC procedure.

   (1) All the relationships collected during convergence (except quintets) are used. The inclusion of quartets alters the refinement; sometimes it improves the radius of convergence, but sometimes it does not.

   (2) The problem of when to stop refinement has always been difficult in YZARC. An alternative method based on NQEST combined with NQINT is offered as an option. After an initial round of $n$ cycles, NQEST and NQINT (if available) are calculated. If at this point a combined NQEST, NQINT figure of merit is greater than a specified cut-off (e) refinement is terminated, otherwise it continues. It checks this figure of merit after each cycle, and refinement continues whilst it continues to fall until it hits a minimum at which point it stops. Experience dictates that suitable
values for $n$ and $e$ above are 7 and 0-0, respectively, but they are user options. This method has the effect of reducing the number of refinement cycles - usually by a factor of six - with a commensurate fall in the computer time required. As with the traditional techniques of stopping $YZARC$ refinement, it is not always successful.

(3) $YZARC$ normally phases the bottom 100 or so reflections from the convergence map. An option is provided whereby the convergence map is bypassed except for origin definition, and the top reflections ordered on $E$ magnitude alone are phased. This usually creates a singular matrix if least-squares is employed so that steepest descent is normally used. This can be useful in circumstances where the convergence map selects a subset of highly linked reflections in which certain parity groups are not represented. Bypassing the map can pass a more representative set of reflections.

(4) The final figures of merit are augmented by the inclusion of $NQEST$ and $NQINT$.

$MAGEX$: $MAGEX$ is another module run between convergence merging and tangent refinement. It has been inserted in the $MITHRIL$ package without any major modifications. Higher invariants are not currently employed here except that an $NQEST$ and/or $NQINT$ value is assigned to each possible solution just as in $YZARC$. This can be used in ranking solutions.

$TANGENT$ $REFINEMENT$: The tangent formula carries out expansion and refinement of phase angles. The following options are provided.

(1) All relationships accepted by $CONVERGE$ are used actively to generate new phase angles. Enantiomorph-sensitivity quartets are manipulated using the procedure described by van der Putten & Schenck (1979).

(2) Three early figures of merit - the two $MULTAN80$ options of $PSIZERO$ combined with $R_{Karle}$ and $PSIZERO$ alone to which is added $NQEST$ and/or $NQINT$. Any combination of these three can be specified. The early figures of merit are useful when it is proposed to generate large numbers of phase sets. There is a link here with $SHELXTL$ and $QTAN$, both of which rely heavily on $NQEST$ to promote refinement efficiency. These figures of merit can also be used in $MITHRIL$ to filter out unacceptable phase sets from $YZARC$ and $MAGEX$ after only a few cycles of tangent refinement.

(3) Two weighting schemes as in $MULTAN80$ - the traditional $MULTAN$ scheme and Hull-Irwin statistical weights (Hull & Irwin, 1978).

(4) Random phases can be used instead of phase permutation.

(5) In its interactive mode, the program will attempt to identify any solution with exceptional figures of merit. It is possible to stop at this point, compute an $E$ map then continue with tangent refinement if the map seems unpromising.

Five figures of merit: $ABSFOM$, $PSIZERO$, $R_{Karle}$, $NQEST$ and $NQINT$, combined together with user controlled weights give a single combined figure of merit and this is used to rank the solutions from tangent refinement.

$MAPS$: The final step is the calculation of one or more $E$ maps. It is in map interpretation that the high level of user interaction which can be provided becomes important, particularly when there are many maps to search. It is possible to generate and search all the appropriate maps quickly, and reject unsuitable solutions on the basis of peak heights, peak positions or fragmentation patterns before attempting a full interpretation. Simple graphics facilities are also provided for those users with graphics terminals. The interfacing is carried out using the $SIMPLEPLOT$ routines (Butland, undated). Even simple graphics can quite transform the chore of searching $E$ maps for recognizable fragments.

$RECYCLING$: Four methods of recycling are provided.

(1) Weighted Fourier syntheses (Sim, 1959, 1960).

(2) Karle recycling (Karle, 1968).

(3) Karle recycling with random phases for the unphased reflections (Yao Jia-Xing, 1983).

(4) Groups of correct orientation but known or unknown position (Main, 1976).

Provision is made for the calculation of an $E^2 - 1$ vector map. User control of the sharpening procedure is possible. Examination of such maps can be useful in difficult cases (Nixon, 1978).

Conclusions

$MITHRIL$ has been in use for about one year, and has proved very successful. Structures with more than 90 atoms in the asymmetric unit have been easily solved. It can tackle all the problems accessible to $MULTAN80$, $QTAN$, $SIMPEL$ etc., as well as structures which are resistant to these programs. It is particularly powerful in symmorphic space groups because of the introduction of negative invariants. Even in non-symmorphic situations, higher invariants often introduce worth while improvements. The resulting $E$ maps tend to be cleaner and more easily interpreted, especially by an automatic peak interpretation routine. In terms of its demands on computer time, $MITHRIL$ uses 10% more processor time than $MULTAN80$ for a similar calculation. A similar figure is applicable for runs of $YZARC$ and $MAGEX$. When higher invariants are invoked, this can increase by 50–100% depending on the circumstances.

This project was the brainchild of Professor M. M. Woolfson who did much of the initial organization, and has provided continual support. Authors of the direct-methods packages cited in the text also provided assistance. The research has been supported by a NATO Research Grant, the SERC and the University of Glasgow. Copies of the program are available from the author.

References


