TEACHING AND EDUCATION IN CRYSTALLOGRAPHY

Contributions intended for this section should be submitted to any of the Co-editors of the Journal of Applied Crystallography. They may describe any aspect of the teaching of crystallography and may take the form of short communications, teaching notes or full research publications.

J. Appl. Cryst. (1996). **29**, 291–295

Optimized Schedule for Large Crystallography Meetings

Y. LE PAGE

National Research Council of Canada, Ottawa, Canada K1A 0R6

(*Received 24 August 1995; accepted 15 January 1996*)

Abstract

International crystallography meetings like the International Union of Crystallography, European Crystallographic Meetings and Americal Crystallographic Association ones currently offer very comprehensive programs with four parallel sessions or more over a week or so. Proper scheduling of these sessions is nearly as important for the scientific success of these meetings as the selection of topics. In order to tackle this problem objectively, it is proposed here to poll crystallographers about the sessions they most wish to attend, and to accept that schedule which minimizes the sum of residual conflicts calculated from the results of the poll and also respects a number of constraints. A novel semi-automated optimization in four steps, based on the existence of major interest profiles among participants, was applied to the scheduling of the American Crystallographic Association 1995 Annual Meeting: a simple rearrangement of a 'conflict matrix' to produce clusters of sessions with large conflicts, followed by application of constraints within clusters in step 2, then a small number of permutations within clusters only, each of them lowering the sum of conflicts while respecting constraints, and finally assignment of sessions to rooms. Without necessarily producing the absolute best solution, solutions very superior to those from intuition and respecting constraints can be built in this way with limited computing power. The experimental data were taken from 102 replies to an e-mail poll of crystallographers at large. Only 20 permutations were performed in step 3. The solution respected all constraints but one, which was relaxed in order to avoid a bad conflict. At the meeting, the room assignments turned out to be mostly verified, with just a few surprises, all having to do with outstanding contributed talks at sessions with no invited speakers. The total attendance at sessions remained high for the whole duration of the meeting. Calculations were programmed in Fortran and performed on a 386 PC clone.

1. Introduction

International crystallography meetings like those of the International Union of Crystallography (IUCr), European Crystallographic Meetings and American Crystallographic Association (ACA) have become quite large. Owing to time, physical and cost limitations, the bulk of the scientific program is held as parallel sessions in rooms with different sizes. The number of parallel sessions and the topics are decided about one year before the meeting. Meetings are usually scheduled by the Program Committee. The goal is to produce a rational layout of the pre-decided elements of the program in the available rooms while avoiding simultaneous sessions appealing to the same audience. This intuitive operation is not often given the same level of attention as other decisions about the scientific program. No matter how well a meeting is scheduled, some participants will wish to attend two or more sessions at the same time. This is unavoidable, even desirable to a degree, because it shows interest for the topics discussed. Proper scheduling is about spreading this feeling as evenly as possible over the duration of the meeting. On the other hand, a feeling by a significant number of participants that there is little for them in the scientific program on a given half-day is a clear indication of a deficiency in the schedule or the program.

2. Problem

A crystallography meeting with rd pre-decided session topics is to be held for d days in r rooms with different sizes, as r parallel day-long sessions. All p participants to the meeting are queried about the d sessions they most wish to attend at the meeting. Assign each session to a day and a room while minimizing the sum of conflicts indicated by the replies to the query and the selected schedule. In addition, some topics may span several sessions, which are then called part I, part II *etc.* of the same topic. Those parts should be scheduled in the same

room on consecutive days as much as possible. In the same vein, some topics should logically come after other ones, but not necessarily on the next day.

Solution methods for scheduling problems are classified into constructive and enumerative. Most scheduling problems are NP-hard (see *e.g.* French, 1982): their solution is dominated by an enumerative part where permutations are generated and ranked. The assignment-to-days problem of a school timetable has been extensively studied (Lawrie, 1969; White, 1975; Papoulias, 1980), but has limited resemblance to the scheduling of parallel sessions at meetings, for which no literature was found. An entirely enumerative solution would be impractical for scheduling large meetings because of the astronomical numbers of permutations involved. A simplification that will limit very much the enumerations required is therefore required.

A novel approach to the solution, based on the assumption of the existence of a number of major interest profiles among participants, is proposed below. It is accordingly expected to work well for meetings of multidisciplinary sciences like crystallography. This fairly simple solution, with no pretense to complete generality or mathematical rigour, has constructive parts (steps 1, 2 and 4) and an enumerative part (step 3), but on very small numbers of permutations. The key remark with regard to all reasonings below is that, owing to the existence of strong interest profiles within our science, *the best solution for a meeting held for d days in r rooms of the same size can be obtained by transposing days and rooms in the worst solution in r days and d rooms, and then* ***permuting on days only the sessions held in the same room***. This paradoxical remark, justified below, has at least three major consequences.

(*a*) It reduces the highly complex problem of finding an initial good solution to the much simpler problem of finding a very bad one. With random data, the complexity of the two problems would be comparable. With profiles, the data are expected to contain a small structured population of huge conflicts and a large number of small ones. The best solution has to do with the many small terms and is extremely complex to devise, especially in the presence of constraints. The worst solution has to do with the few large terms and can be obtained in a fairly straightforward way.

(*b*) It reduces from astronomical numbers to very tractable ones the number of trial permutations to be performed in order to obtain an excellent solution from the initial one.

(*c*) It simplifies the introduction of constraints between sessions, *e.g.* parts I and II of the same topic to be preferably held in the same room on consecutive days.

In the presence of strong interest profiles, the worst meeting is obtained by scheduling the sessions with the strongest conflicts at the same time, *e.g.* all macromolecular sessions on the same day *etc.* All rooms being of the same size, permutations of these sessions among rooms only make this solution no better and no worse. After transposition of days and rooms, one macromolecular session is scheduled on each day in the same room, a very reasonable way to schedule them. Can this be improved by swapping one of the macromolecular sessions with another session held in a different room? The answer is no: if the swap is with a session held on the same day in a different room, conflicts before and after the swap are identical; if the swap is with a session held on a different day, the new conflict with the other macromolecular session scheduled on that day would be so large it would be most unlikely that the solution with the lowest sum of conflicts could involve a conflict of this magnitude. Permuting on days among macromolecular sessions only, among small-molecule sessions only *etc* is then the only way to expect improvements for the sum of conflicts. Hence this apparent paradox above: under the assumption of interest profiles, the best solution can be obtained by transposition of one among the equally worst dual meetings and then permutations on days with no change of rooms.

3. Solution and algorithm

Sessions are arbitrarily numbered from 1 to *rd*. The replies from the p participants are stored as 0 or 1 values in the elements of a rectangular 'reply matrix' $\mathbf{R}$ with number of rows equal to the number of sessions and number of columns equal to the number of participants. A square and symmetrical $rd \times rd$ 'conflict matrix' $\mathbf{C} = \mathbf{R}\mathbf{R}^T$ is built from it, where $\mathbf{R}^T$ is the transpose of $\mathbf{R}$. The value of element $\mathbf{C}_{ij}$ of the conflict matrix is then the number of participants who wish to attend both session i and session j.

Without constraints, good solutions for the above problem, but not necessarily the best one, would be obtainable from singular-value-decomposition methods (C. K. Johnson, 1994, private communication). However, with the additional constraints above, the problem becomes complex for matrix algebra alone and other types of solution methods have to be used. The meetings in question having over 30 scientific sessions, a brute-force approach based on the calculation of the sum of conflicts for all possible permutations would involve astronomical numbers of permutations and then exceed the computing capabilities of existing equipment. A semi-automated algorithm based on successive improvements brought one at a time to an initial decent solution is presented here. The amount of computing required at each step is seconds on a PC and the operator can decide at each step whether or not to accept the suggestion of the computer.

3.1. *Initial solution*

We first assume all rooms to be of equal size. The problem of scheduling dr sessions for d days in r rooms

of equal size, and that of scheduling them for r days in d rooms, again of equal size, are dual problems. If interests of participants were purely along r determined profiles and d sessions were devoted to each profile, the dual of the worst meeting in r days and d rooms would take care of all conflicts. That dual meeting could be held as r separate meetings just as well. In practice, things are more complicated for two reasons:

(*a*) Macromolecules, small molecules and materials are three familiar interest profiles. Defining the five or more profiles required to schedule the parallel sessions must rely on numerical analysis rather than intuition.

(*b*) In spite of their general perception as disjoint interest profiles, there is a lot more cross-interest between macromolecular, small molecule and materials profiles than is generally thought. See below.

Taking into account points (*a*) and (*b*) above is what makes the difference between an intuitive schedule and a good one, which successfully avoids a number of hard-to-guess, but quite real, lesser conflicts as well. As an example, it came as a surprise to the author that a very significant fraction of the attendance at the session on '*Modulated Structures and Quasicrystals*' at ACA '95, a topic at the heart of the materials profile, also wished to attend '*Difficult Macromolecular Structures*', a central macromolecular topic. A closer look at the data showed that the same people also wanted to attend '*Off-the-Wall: Theory*', thus pointing to a latent 'Theory' subprofile cutting right across the traditional barriers of interest groups. These three sessions and others were scheduled at different times by the algorithm below, allowing theorists among others to remain busy all week long cross-fertilizing these difficult topics.

3.2. *Algorithm*

The algorithm below takes into account the major profiles, the constraints, the latent subprofiles and then the attendance in four successive steps, all based on magnitudes of elements of **C** rather than on a rationalization of the demography of profiles and subprofiles in terms of major and minor crystallographic disciplines.

3.2.1. *Step 1: derivation of the worst program in r days and d rooms.* The first day: Find the most attended session from the largest value among diagonal terms $\mathbf{C}_{ii}$. Permute rows 1 and i in **R**, and recalculate **C**. Find the largest $\mathbf{C}_{1j}$ term, with $j > 1$. Permute rows 2 and j in **R** and recalculate **C**. Find the index k ($k > 2$) for which the sum $\mathbf{C}_{1k} + \mathbf{C}_{2k}$ is maximum. Permute rows 3 and k and recalculate **C**. Row 4 is then permuted with row m ($m > 3$) with $\mathbf{C}_{1m} + \mathbf{C}_{2m} + \mathbf{C}_{3m}$ maximum *etc.* until row d is permuted.

This simple algorithm will find a cluster of d sessions very highly correlated with the most attended one. When applied to the data in the experimental part below, this cluster is immediately recognized to be the macromolecular core profile, but this information is not fed into the analysis. In the subsequent analysis, this cluster will be permuted, but not split up, ensuring that macromolecular crystallographers will be kept busy for the whole of the meeting, at all subsequent steps in the analysis.

The following days: Find the largest diagonal term $\mathbf{C}_{ii}$, $i > d$. Permute rows $d+1$ and i in **R**, and recalculate **C**. Find the largest $\mathbf{C}_{d+1,j}$ term, with $j > d+1$. Permute rows $d+2$ and j, *etc.*

When applied sequentially to all r days, this simple algorithm reorders matrix **C** in such a way that the largest cross terms of **C** are now elements of r $d \times d$ submatrices about the diagonal. If we read the labels of the successive rows of **R** to be the d parallel sessions on the first day, followed by the d parallel sessions on the second day *etc.*, this is the worst solution of the problem in d rooms and r days we were looking for, or very close to it. If we now read the same row labels to correspond to the sessions for the successive d days in the first room followed by those for the d days in the second room *etc.*, this is the decent solution of the dual problem discussed above. The following steps and discussion will be in terms of this initial solution in d days and the successive improvements brought to it.

3.2.2 *Step 2: introduction of constraints.* Constraints between pairs of sessions exist because the two sessions appeal to the same subset of participants. Their conflicts with other sessions are therefore expected to be similar. The assignment to days of constrained sessions is performed by hand within their cluster, based on their logical relative placement and mutual conflicts only. When constrained sessions clash with one another, one can grasp why this is so because of their small number, and what the cost of a given constraint is in terms of residual conflicts. The decision to relax constraints or not, and which ones to relax, is then taken by a human operator weighing the relative importance of these factors, rather than being left to a computer. Conflicts with nonconstrained sessions are disregarded at this stage: they will be tackled in step 3. The assignment to days of constrained sessions will not be altered in the remainder of the analysis.

3.2.3. *Step 3: minimization of the sum of residual conflicts.* In order to improve on this solution, conflicts between sessions scheduled at the same time in different rooms will be tackled in the following by permutations on days within the same room only, according to the main remark above. This reduces drastically the number of trial permutations to be considered from astronomical numbers to $2d - 2$ at each step. This number is 12 for ACA '95 and 28 for IUCr XVII. This method prevents the breakup of the above clusters, for the same reasons as previously. The residual conflicts are those $\mathbf{C}_{ij}$ terms ($j > 1$) where i and j differ by a multiple of d: the corresponding sessions are accordingly scheduled on the same day in different rooms. They lie on r lines parallel to the diagonal of **C**. The sum of conflicts for a given permutation is the sum of these terms.

The largest residual conflict is identified and sums of conflicts are calculated for each of the 2 $(d-1)$ permutations of the two corresponding sessions with the other $d-1$ sessions held in the same room on different days. The permutation respecting the constraints and with the lowest sum of conflicts is accepted as a new starting point. The process is repeated until the largest conflict cannot be resolved. The second largest conflict is now tackled *etc.*, until no further improvement can be obtained.

3.2.4. *Step 4: adapting the solution to rooms of unequal sizes.* The number of participants wishing to attend session i is the value of the corresponding diagonal term $\mathbf{C}_{ii}$. Rooms are labelled *A*, *B*, *C etc.* in decreasing order of seating capacity. Sessions are ranked in decreasing order of diagonal terms and attributed to rooms on this basis. Minor corrections to this scheme are brought in at this point in order to hold all parts of the same topic in the same room.

4. Application to ACA '95

After the duration in half-days (1, 2 or 3) of each of the 26 scientific topics dealt with at ACA '95 for 35 half-day sessions in five rooms and seven half-days was finalized mid-October 1994, these sessions were arbitrarily numbered from 1 to 35. An e-mail message with these numbered session topics was sent around on October 23 to about 2 500 addresses of crystallographers, requesting in reply a list of seven numbers corresponding to the seven sessions the recipient would most wish to attend. The message also explained how to access the ACA '95 World Wide Web (WWW) server, where additional information about the scopes of the sessions, the names and e-mail addresses of their organizers, the invited speakers in these sessions and the titles of their talks were available. Intense activity was observed right away on the server, and a total of 130 replies (12% of registrations at the meeting) was received. For reasons of available time, only the first 102 of them, received over the four days following the mailing of the message, were taken into account in the analysis. A quick demographic analysis of e-mail addresses for replies (universities/government/companies/countries) and of the replies themselves for major topics did not result in any reason to doubt that the replies came from a representative sample of participants to ACA meetings.

Only one utility was programmed, which was sufficient to take most of the tedium away from the solution. It produced a printable file with the elements of **C** and the sum of conflicts from the input of a permutation of sessions in about 1 s on a 386 PC. It was used to help build the worst solution mostly as explained in step 1 above, in about 2 h. Its result was reordered by hand within clusters to take the constraints into account as explained for step 2. At this point, one of the constraints had to be relaxed because a bad conflict would necessarily be created if parts I and II of '*Structure-Based Drug Design*' were held on successive half-days. This condition was relaxed and the two sessions were treated as independent in the remainder of the analysis. The corresponding decent solution had 178 conflicts, as compared with 252 for a random solution. Two clashes involving the two sessions of '*Data Collection Hardware, Software and Methods*' were tackled first. Two permutations minimizing these two large conflicts reduced the sum of conflicts to 148. From having participated in Program Committee scheduling exercises before, it is the opinion of the author that meetings scheduled from intuition correspond to about this kind of layout and this level of residual conflicts, and may easily miss the many minor difficult-to-guess conflicts that were tackled numerically afterwards.

This number of conflicts for the trial solution was then progressively reduced to 104 in about 20 permutations mostly as explained in step 3. Promising trial permutations were obtained by eye examination of cross-terms in printouts of **C**. Retrospectively, it would have been much faster to let the computer generate and rank the 12 trial permutations at each step: this would have taken under a minute per accepted permutation on a PC. After three days of this exercise, no further improvement could be found. Correspondence with Carroll Johnson indicated that solutions from singular value decomposition on the same data, but not taking constraints into account, produced a similar sum of residual conflicts. The solution derived as explained above and respecting the constraints was applied to theatres *A* to *E* at the *Palais des Congrès* in Montréal, based on decreasing magnitudes of corresponding diagonal terms $\mathbf{C}_{ii}$. The schedule was accepted by the ACA Council on November 5 and posted on the WWW shortly afterwards at URL http://www.cisti.nrc.ca/ACA95/timtab.html. It also appears on the back cover of the 1995 ACA *Program and Abstracts book*. A simplified version was printed on the back of the *Call for Papers*, mailed in January 1995. The time invested in the solution was recovered very many times at the meeting, with respect to a schedule from intuition.

At the meeting, predictions for relative sizes of sessions were mostly verified. Glaring exceptions were the '*Off-the-Wall*' sessions on '*Theory*' and '*Results*', new this year and inviting contributions only, which became for short times the second largest sessions on their respective half-days while they had been respectively predicted to be third and fourth largest by the analysis above. At other times, they were third and third, but with an attendance much larger than predicted. We attribute this discrepancy both to the novelty of these topics and to the fact that no topic, title or speaker was known for these sessions at the time of the poll. Otherwise, the ranking was pretty much according to forecasts, and experimental differences could be attributed to statistics on small

numbers. A number of participants wanted to be at several places all the time, but no comment about lack of interest in the program on a given parallel half-day was heard at the meeting. Total attendance numbers remained high until the close of the meeting.

5. Conclusion

The novel method proposed here for scheduling crystallography meetings from polls of crystallographers at large seems to have definite advantages over traditional scheduling from intuition. First, the wishes are obtained directly from the crystallographic community on actual topics, rather than estimated subjectively. Second, the above theoretical framework allows a very effective minimization of conflicts to be performed based on the replies to the poll. The practical feasibility and soundness of the approach have been demonstrated for ACA '95, a full-week meeting with over 1 100 participants. The schedule was known over eight months before the meeting, allowing people to make early travel plans. This method will hopefully be considered as an alternative for scheduling future meetings, thus optimizing the usefulness of these huge crystallographic forums by maximizing attendance at the various sessions through minimization of anticipated conflicts between sessions.

The warmest thanks are due to Dr Carroll Johnson from the Oak Ridge National Laboratory for his sustained encouragement, wise advice and support over the 12 months preceding the Montréal meeting, in particular with this problem. Sincere thanks are also due to Dr Martha Teeter from the Boston College for help with the huge e-mailings required in the experimental part of this study.

References

American Crystallographic Association (1995). Proc. Annual Meeting, Palais des Congrès, Montréal, Québec, July 23–28, Ser. 2, Vol. 23.

French, S. (1982). *Sequencing and Scheduling: an Introduction to the Mathematics of the Job-Shop*. Chichester: Ellis Horwood/Wiley.

Lawrie, N. L. (1975). *Comput. J.* **12**, 307–316.

Papoulias, D. B. (1980). *Eur. J. Operational Res.* **4**, 31–41.

White, D. J. (1975). *Operational Res. Q.* **26**, 875–878.