J. Appl. Cryst. (1987). **20**, 235–242

A System for Collection and On-Line Integration of X-ray Diffraction Data from a Multiwire Area Detector

BY M. BLUM, P. METCALF,* S. C. HARRISON AND D. C. WILEY

Committee on Higher Degrees in Biophysics, Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA, USA

(*Received* 10 *September* 1986; *accepted* 3 *February* 1987)

Abstract

A system for collecting and measuring X-ray diffraction data from protein crystals has been developed for a multiwire area detector. Computer programs run concurrently on two microcomputers, which collect and reduce detector data to integrated intensities. The self-contained system consists of an X-ray area detector, a rotation/oscillation camera, and two microcomputers connected by a high-speed Ethernet network. One microcomputer is dedicated to operation of the detector, control of the camera, and storage of the raw data. The second microcomputer automatically integrates the data as they are collected and allows the user to monitor the quality of data as they are processed. The integration programs are written in Fortran 77 and have been designed to be portable. Additional programs for crystal alignment, detector and camera control, and graphics are written in the C programming language. A description of the system, some characteristics of the detector, and the results of data collection are presented.

1. Introduction

X-ray area detectors and their application to protein crystallographic problems have been described by a number of groups (Arndt, 1985; Bolon, Crawford, Deutsch & Quigley, 1981; Hamlin, 1985; Sobottka, Cornick, Kretsinger, Rains, Stephens & Weissman, 1984). Most systems accumulate images of the X-ray diffraction pattern in a local memory and transfer them to a dedicated computer to be integrated as they are collected (Howard, Nielsen & Xuong, 1985). We have previously described the development of a system for collecting macromolecular crystallographic data and for evaluating the stored data on a multiuser computer (VAX 11/780) (Durbin *et al.*, 1986). We have now developed a dual microprocessor system using a microcomputer to process data on-line as they are collected. The algorithms and programs are applicable to any detector producing frame data, and the processing programs can be adapted to any computer running Fortran 77 that is capable of receiving raw data frames from the detector. Although running concurrently with the data collection, the data integration does not require synchronization with the data collection. The second processor may also be used during data integration for other tasks such as graphical inspection of the data as they accumulate. Additional programs for aligning the crystal, controlling the camera and detector during data collection, and for graphically monitoring the results are written in the C programming language.

2. Data collection and processing apparatus

The system (Fig. 1) consists of a multiwire proportional X-ray counter (Nicolet/Xentronics, Madison, WI), a modified rotation/oscillation camera (Supper, Natick, MA), and two 68010 microcomputers (Cadmus, Lowell, MA). Each computer is equipped with a video terminal, printer, 1·5 Mbyte of random-access memory, a 65 Mbyte disk, and a high-speed color graphics display system. The two computers are linked together by an Ethernet cable for high-speed transfer of 512×512 pixel detector images (data frames). The computer that collects data, termed *C*1,

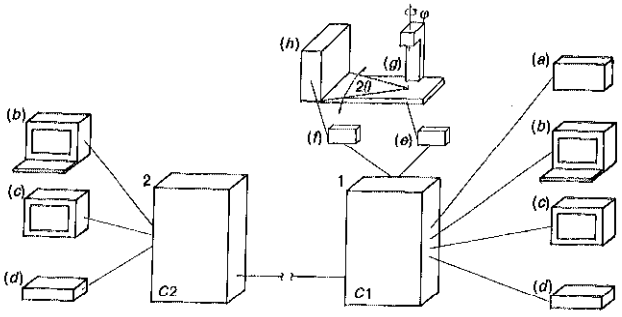


Fig. 1. Schematic diagram of the detector, computers and peripherals. (1) *C*1, data-collection microcomputer. (*a*) Tape drive, (*b*) console, (*c*) color monitor, (*d*) printer, (*e*) camera controller, (*f*) position decoding circuit, (*g*) oscillation camera with arm for swinging detector, (*h*) detector. (2) *C*2, data-processing microcomputer with peripherals labeled as in 1.

*Present address: European Molecular Biology Laboratory Outstation, c/o ILL, 38042 Grenoble Cedex, France.

0021-8898/87/030235-08$01.50

has, in addition, a 9-track magnetic tape drive and is connected to the detector position decoding circuit and the camera controller. Half a Mbyte of memory on *C*1 is dedicated to accumulation of data directly from the detector. The computer that handles the reduction of raw data frames to indexed integrated intensities, termed *C*2, has a floating-point processor to speed computation. Both computers utilize the UNIX (AT&T) operating system, which occupies approximately ½ Mybte of memory.

The camera is similar in design to a standard rotation/oscillation camera, but it has a vertical crystal rotation axis φ and an arm supporting the detector, which can be rotated around the crystal in the horizontal plane 2θ. The two camera motors and an electric shutter are controlled by a small separate microprocessor, which is connected to and controlled by *C*1.

3. Computer programs

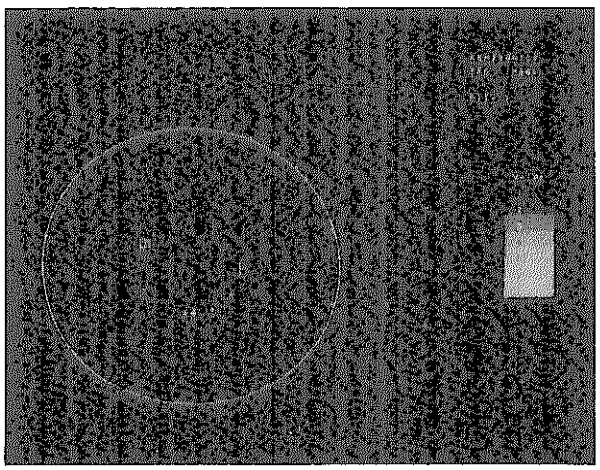
3.1. *Crystal alignment*

Crystal alignment is carried out using an interactive-graphics computer program (*ALIGN*). A number of short exposure still and oscillation images are made to ascertain the approximate orientation of the crystal. Usually 1- to 10-second exposures are sufficient to produce images showing enough spots to make the still rings recognizable. Detector images can be displayed on a video monitor within about 5 s using a color code to represent individual pixel counts. The monitor displays, in addition to the detector image, a number of variable-sized cursors, which can be moved about without disturbing the image. Crystal alignment is carried out by matching one of the still rings with a circular cursor (Fig. 2*a*). The program calculates the correction angles required to align the crystal axis corresponding to the still rings. The orientation of a crystal can be confirmed by making oscillation images with a large oscillation angle (typically 5°), in order to measure reciprocal-axis spacing and angles, and also by directing the program to superimpose on a detector image the still rings corresponding to the spacing of a particular set of lattice planes.

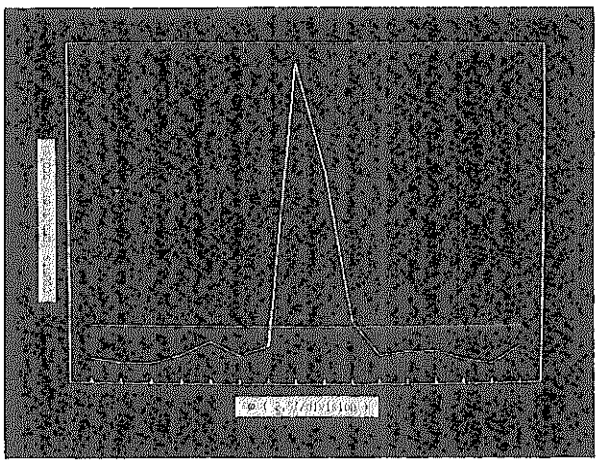
The alignment program can be used in a number of ways. Individual diffraction spots can be selected with a rectangular cursor and magnified so that the spot image covers the whole screen. The pixel counts for the spot can be displayed on the computer terminal along with statistics giving the integrated intensity, background, centroid and dimensions of the spot. A graph representing the φ profile can be generated by allowing the program to make a number of exposures while stepping through the reflection by rotating the vertical φ axis of the camera (Fig. 2*b*).

3.2. *Data collection*

Data are collected while oscillating the crystal about the rotation axis φ during each exposure. Data frames are small oscillation images whose oscillation ranges are contiguous rather than sums of still images (Xuong, Freer, Hamlin, Nielsen & Vernon, 1978; Howard *et al.*, 1985). The oscillation averages out fluctuations in beam intensity and/or small uncertainties in spindle position, ensuring a uniform and complete sampling of reciprocal space. Data from



(*a*)



(*b*)

Fig. 2. Crystal alignment and initial evaluation. (*a*) Alignment still (5 s exposure) with cursors. The + cursor marks the position of the direct beam visible through a partially transmitting Cu beam stop. The circular cursor is superimposed on the still ring so that *ALIGN* can calculate adjustments to be made to goniometer arcs. (*b*) Rocking-curve profile collected and displayed by *ALIGN*. Points on the curve represent integrated intensity inside the small box cursor (middle left of Fig. 2*a*). The crystal is moved in 2·5′ steps (marked on abscissa) about the rotation axis. The exposure time for each step is 5 s in this example. *ALIGN* automatically searches out the peak of a reflection selected with the box cursor.

more mosaic crystals or data collected using more divergent optics can be collected efficiently by increasing the oscillation range per frame without altering the manner in which the data are treated. The oscillation angle is chosen so that the majority of reflections will be fully recorded on three to five frames each. This allows accurate determination of the angular position and shape of each reflection and results in typical ranges of 5–15′. A single frame is a 512×512 pixel image of the diffraction pattern on the detector face. Each pixel is stored as a 2 byte word in memory, allowing the accumulation of, at most, 65 532 counts in the corresponding ~ 200 μm $\times \sim 200$ μm square area on the detector face. For details see Durbin *et al.* (1986).

Data frames are collected with the program *COLLECT*, which runs on *C*1. *COLLECT* serves as a controller and notebook, reminding the user to record relevant data that will be necessary for further processing, controlling and recording motions of the crystal and of the X-ray shutter, and recording statistics about each frame. Input to *COLLECT* is through a menu editing routine, which guides the user through the necessary steps to start the data collection. The program can be used to control the collection of data for on-line integration and/or to store data frames on tape for later processing.

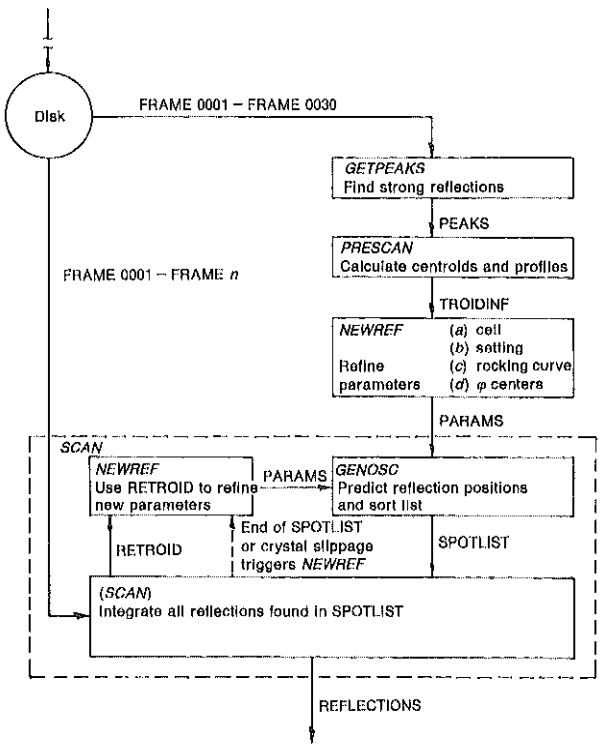


Fig. 3. Flow diagram of data integration programs. Data frames are written to disk on *C*2 across Ethernet by *COLLECT* on *C*1 (not shown). Words in italic capitals are program names (*GETPEAKS*). Solid arrows represent files (file names are capitalised: TROIDINF). The dotted box indicates that *SCAN* incorporates the functionality of three programs (*NEWREF*, *GENOSC*, *SCAN*) from the previous off-line system (Durbin *et al.*, 1986).

In the dual-processor configuration, frames are stored on the *C*2 disk so that frames can be retrieved by programs on *C*2 without disturbing the data collection on *C*1. Data frames are generated by *COLLECT* as long as room is available to store them. If no space is available on disk (or on tape, if raw frames are to be saved) because of a delay in starting the integration or because data are being collected too fast to integrate immediately (see § 4.1.3), *COLLECT* waits, with closed shutter, until it can save the most recent frame before continuing. On *C*2, reflections are integrated whenever all the frames contributing to the measurement of spot and background become available. With this scheme (Fig. 3), it is unnecessary for the integration program to exert control on the data collection.

3.3. *Preparation for integration*

In order to integrate the reflection data, an accurate knowledge of the crystal setting and camera parameters must be obtained. The most accurate orientation data could be obtained by collecting data from the crystal at several orientations to obtain a sampling of reflections well dispersed in reciprocal space. When properly indexed, these reflections would yield a well determined set of orientation parameters. This method is somewhat inconvenient, since an algorithm that relied on a well dispersed sample would require the collection of another well dispersed sample should the crystal slip during data collection. We have found that a much simpler method which takes advantage of the area detector's ability to collect many simultaneous reflections provides satisfactory orientation data.

We begin the data collection with no further knowledge of the crystal setting parameters than that obtained by the approximate alignment described earlier. After a number of frames (usually 30) have been collected and stored, they are searched for strong reflections with the program *GETPEAKS* (Fig. 3). For each of these reflections, the extent of the reflection in φ (the rocking curve) is determined and the reflection is integrated on each frame on which it is recorded using a rough integration mask. A record is written to the TROIDINF (cen*troid in*formation) file for each reflection. The record contains a reflection's calculated centroid position and the fraction of its intensity found on each frame.

The TROIDINF records are used by the refinement program (*NEWREF*) to refine accurate parameters describing the unit cell, crystal orientation and diffraction geometry. The refinement algorithm is based on methods for determining crystal orientation

parameters from oscillation films (Winkler, Schutt & Harrison, 1979) which have been modified for use with detector data (Durbin *et al.*, 1986). *NEWREF* is an interactive program. It is at this point that the user verifies that the crystal is of reasonable quality and that no gross errors have been made in alignment. If unable to proceed, a minimum amount of time has been lost, and corrections can easily be made, or a new crystal mounted, and another 20–30 frames collected. If all proceeds well to this point, an accurate set of crystal and setting parameters (PARAMS) has been obtained and the integration can begin. Data collection continues uninterrupted on *C*1 during the refinement of crystal and camera parameters on *C*2. At no time has it been necessary to interfere with the continuing collection of data.

3.4. *Integration*

The integration program *SCAN* uses the orientation information file, PARAMS, to predict a list of positions for reflections that will appear on the first n frames ($n = 50$, $\sim 4°$ of data). This file, SPOTLIST, is sorted so that reflections are listed in the order in which they will appear on the frames. During the time that the search, refinement and prediction have been occurring, additional frames have been accumulating on the disk as data collection continues on the first computer. Integration now begins and progresses faster than the data collection. Eventually *SCAN* attempts to read from disk a frame that has not yet been collected. At that time, it waits for the presence of the required frame before proceeding. As *SCAN* progresses, it deletes from the disk the frame files that are no longer necessary for retrieving spot or background data.

Two files are written by *SCAN*. The REFLECTIONS file contains a record for each reflection, recording h, k, l, x, y, φ, spot and background counts, Lorentz and polarization factors,* and prediction errors in x, y, φ. The RETROID file contains observed (see below) centroid positions and $I(\varphi)$ profiles for the stronger reflections. This file is identical in format to that used for the initial interactive refinement (TROIDINF). It is used for subsequent automatic refinements. After a number of frames have been processed, the parameters are automatically rerefined using the RETROID file. Rerefinement may occur at a predetermined point, or it can be triggered by large errors in predicting the position of strong reflections (see below).

The first rerefinement is usually programmed to occur after approximately 4° of data collection to increase the accuracy of the initial parameters, which were obtained from only 2–2·5° of data. Subsequent refinements typically are programmed to occur at 8–12° intervals. Frequent rerefinements are not unduly time consuming and compensate for crystal

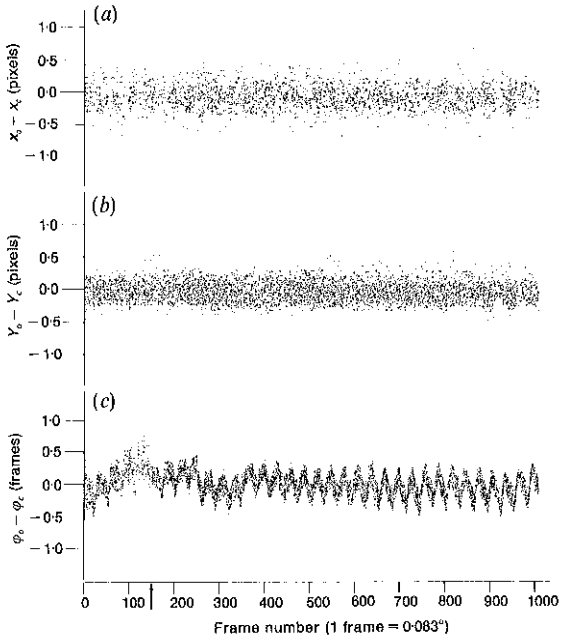


Fig. 4. The difference between observed and predicted position of a reflection during the course of data collection. Data for 16 951 reflections with I greater than $5{\cdot}0\sigma$ (σ = Poisson statistics standard deviation) from an ILTat 1.24 crystal soaked in a lead lactate solution (Table 1) showing that the positions of the strong reflections are found to be very close to the predicted positions. The prediction errors for X and Y are less than 0·5 pixel and the prediction errors in φ are less than 0·5 frames. These errors indicate that the position of the integration mask was properly calculated and that weak reflections were properly integrated. The principal contribution to errors in φ has a periodic component (period = 2·5° = half turn of the φ-axis motor), probably associated with the mechanical components of the φ-axis drive. The amplitude of this oscillation is less than 0·5 frames and therefore does not affect the placement of integration masks. The first programmed rerefinement occurred after 150 frames (arrow). A significant drift in the predicted positions is apparent here. The present system detects this drift and rerefines automatically well before it becomes so pronounced. Subsequent refinements occurred every 100 frames. (*a*) $X_{obs} - X_{calc}$ *vs* φ. Error in predicting the position of reflections on the detector face. The accuracy of the prediction depends on the knowledge of crystal and camera parameters and on the residual spatial distortion of the detector after correction. X direction is perpendicular to anode wires, so it would be expected to show the greatest distortions due to 'anode modulation'. Standard deviation is 0·37 pixels (~ 65 μm). (*b*) $Y_{obs} - Y_{calc}$ shows similar behavior but the mean deviation from predicted position is smaller. Standard deviation is 0·23 pixels (~ 45 μm). (*c*) $\varphi_{obs} - \varphi_{calc}$ *vs* φ. This quantity is independent of spatial distortions of the detector, and it is therefore a sensitive measure of the accuracy of parameters used for reflection prediction. The initial region of greater scatter is due to less-accurate parameters, based only on the first 30 frames of data. Subsequent scatter arises primarily from the periodic inaccuracy of the φ drive. Standard deviation is 0·21 frames (0·017°).

*Lorentz and polarization factors are calculated at the time reflections are predicted. *SCAN* simply passes these on as a combined Lp factor. The factor is applied any time that an intensity is calculated from the integrated counts.

slippage and possible camera misalignment. Relative insensitivity to camera alignment is a consideration with many users of varying skill and patience operating a rotating-anode X-ray generator, where filament replacement and consequent readjustment of optics and camera are a frequent occurrence. Rerefinement and prediction cycles for the 4·2 Å data set shown in Fig. 4 took less than 5 min each.

Reflections are integrated at their observed positions. Before each reflection is integrated, its position is locally refined by iterative calculation of a three-dimensional centroid of intensity inside an integration mask. For the purposes of centroid calculation, the background is intentionally underestimated, preventing extreme movement of weak reflection centroids. Integration is accomplished by summing counts within the integration mask. Background is integrated from the same detector pixels, on frames immediately 'before' and 'after' the frames on which the reflection falls.

The difference between the observed centroid position and the predicted position is stored (Fig. 4) and the r.m.s. prediction error over the strong reflections ($I > 5\sigma$) on each frame is calculated. The program monitors an averaged r.m.s. prediction error accumulated over the last ten frames, and if the value passes a preset tolerance, a rerefinement is triggered. Rerefinements (programmed or error directed) are calculated using the RETROID records for reflections falling on the most recent 50 frames. These data contain observed information about the most recent condition and alignment of the crystal, regardless of any reasonably small errors in predicting the spot positions.

Data integration and refinement cycles occur automatically with no user intervention. During the time spent refining parameters and predicting reflections, frame data collected on *C*1 are written to the *C*2 disk just as they are during integration. When integration resumes, it quickly catches up again with data collection.

3.5. *Monitoring progress*

The course of data collection can be followed in several ways. The data collection and integration programs both write relevant information to the screen. More detailed information is written by each program to a log file, which can be inspected at any time. Errors encountered are recorded both on the screens and in the log files.

At any time during the course of data collection and integration, the previously measured data may be inspected and analyzed by one of two programs, *ANALYZE* and *LOOKFILE*, without interrupting data collection or integration. *ANALYZE* operates on the *SCAN* output file REFLECTIONS. *ANALYZE* calculates symmetry *R* factors (R_{sym}) and mean and r.m.s. prediction errors for all reflections integrated, and displays these calculations as a function of several variables. These include intensity, resolution, position on the detector face, φ and redundancy. Running *ANALYZE* early in the data collection can help the user to establish that the exposure times are reasonable and give assurance that the integration is working properly. By running *ANALYZE* later, or on the completed data set, the user can determine the quality and resolution limits of the data, pinpoint any problem areas with the collection and integration system, and produce a permanent record of the data quality.

LOOKFILE provides a quick graphical method of inspecting the data. *LOOKFILE* can open any of the files associated with the integration programs including the original data frames. The program allows the user to make graphs and scatter plots on a color screen of any elements in a file and to superimpose those plots onto data-frame images. One simple scatter plot that we have found most informative is that of the prediction error in φ *vs* frame number (Fig. 4*c*). This quantity is a sensitive measure of how well the lattice used for spot prediction matches the actual reciprocal lattice. This plot quickly indicates how accurate the prediction has been during the collection run and helps identify serious problems such as severe slippage or changes in crystal lattice parameters. *ANALYZE* and *LOOKFILE* have been useful for debugging the software and the hardware during system development, as well as for monitoring actual data collection.

4. Results

4.1. *Detector, computer and software performance*

Data have been collected from several protein crystals with this system. Routine collection of diffraction data has allowed a good opportunity to evaluate the performance of the detector and the processing system. Symmetry *R* factors are not enough, however, to demonstrate that a detector is producing accurate data. Variation in both detection efficiency and positional accuracy may affect the quality of intensity data or cell-parameter determination. Durbin *et al.* (1986) described spatial corrections applied to the data from the Nicolet/Xentronics detector. The on-line processing and graphics display system gave us the ability to check the validity of those corrections, as well as evaluate the performance of the refinement and integration algorithms. We have also investigated the spatial and temporal uniformity of detection efficiency in order to evaluate the need for treatment in software of sensitivity variations.

4.1.1. *Spatial distortion of the detector face.* We have found that the detector spatial distortion, due to the

non-planar window and anode modulation, are reasonably well corrected by the treatment described by Durbin *et al.* (1986). Figs. 4(*a*) and (*b*) show that the r.m.s. error in predicting the position of a reflection on the detector face is small and that a very small number of reflections fall farther than one pixel from their predicted position. Images made with this detector show a horizontal modulation corresponding to the positions of anode wires. Anode modulation appears to increase the prediction error somewhat in the x direction. On our detector (a prototype) the magnitude of the anode modulation varies over the detector face. The extent to which this affects data quality has not been ascertained.

Unit-cell constants calculated from detector data agree well with those obtained from precession photographs. When determined from sufficient data, these parameters are stable ($\sigma_a/a < 0{\cdot}002$) from one crystal to another and are a reliable method of identifying cell-constant changes in heavy-atom derivatives.

4.1.2. *Sensitivity distortion over the detector face.* We have made a crude estimate of the spatial variation in sensitivity (quantum efficiency) over the detector face. This measurement is complicated by the method by which spatial distortions are corrected (Durbin *et al.*, 1986). After the initial correction is applied, each pixel no longer represents a unit area on the detector face. Indeed, the pixels are often not of uniform size before the correction (Hamlin *et al.*, 1981). It is therefore not possible to calculate a spatial sensitivity variation based on irradiation of the detector with a uniform field and analysis of the variation in counts from pixel to pixel. Rather, a uniform unit of area must be defined as a function of position on the detector face and the variation in counts per unit area examined. Alternatively, a collimated or focused beam may be integrated at many positions as it is moved over the detector face (Hamlin *et al.*, 1981). Integrating a beam in its entirety has the advantage of eliminating the need to define a unit area. We have attempted to implement such a measurement with a rotating-anode source.

By taking a large number of successive exposures at small angular intervals of the 2θ arm, the attenuated beam was measured at many positions along the detector equator. By turning the detector on its side and repeating the experiment, it was possible to sample along the vertical axis of the detector. These data (not shown) indicated a variation of, at most, 7% from the edge of the detector to its center, and 2–3% in the central region. This variation was mostly smooth, with the sensitivity dropping off towards the edges in the horizontal plane, but varying less systematically or extensively in the vertical direction. In several places there were small sharp changes in the sensitivity of approximately 1%. These appear to correspond to boundaries between preamplifiers on the detector wires (R. Burns). No sensitivity variation was observed that correlated with anode modulation. Accurate measurement of the variation was difficult owing to the instability of the rotating-anode X-ray source. That this temporal variation was not due to the detector was established by counting the emission from a 100 μCi ^{55}Fe source, collimated to give a 5 mm diameter image on the detector. (1 Ci = $3{\cdot}7 \times 10^{10}$ Bq.) The standard deviation of repeated measurements of the ^{55}Fe image was less than 1%, and it was accounted for by Poisson statistics.

A full characterization of the spatial variation in counting efficiency over the entire detector face would be too time consuming and cumbersome to be performed regularly. It may be necessary to develop a means for making such a measurement if data of very high accuracy are to be gathered. At present, however, the distortion measured is below the accuracy expected from the statistical limitations on the data. We therefore make no correction for sensitivity variations.

4.1.3. *Computation vs data-collection speed.* The speed with which the raw data are reduced to integrated intensities must be high enough that it does not limit the rate at which data can be collected. In processing frame data, there are two rates that will determine the overall data-processing rate: the rate at which data frames can be transferred from the data-collection device (freeing the device to collect another frame), and the rate at which the data on the frame can be integrated (freeing the data-processing computer to receive another frame). For the crystallographic problems that we have encountered, the speed of this dual microprocessor system has been adequate, although there are certainly problems which will require faster processing.

The microcomputers, configured as described above, and using Cadmus Ethernet communications software (*UNISON*) require approximately 14 s to transfer a frame from the data memory of *C*1 to the disk on *C*2. If the data frames are to be saved on tape as well, that operation requires approximately 36 s. In addition to the transfer time, the collection software and camera controller use 12 s of overhead per frame. Therefore the minimum overhead per frame when using the dual processor system and saving frames on *C*2 is 26 s.

On the integration computer, *SCAN* requires from 14 to 20 s to read a data frame directly from the local disk. The speed of this operation is limited because reading in a frame actually involves a read operation (from the file on the disk) and a write operation (to virtual memory on the disk). In addition to the read time, *SCAN* requires an average of 2·5 s to integrate each reflection. This time is also mostly due to virtual-memory access, although the program has been de-

Table 1. *Data collected with the on-line integration system*

All data were collected with rotating-anode X-ray generator (Elliot GX-13), operation at 40 kV, 40 mA, fitted with a 100 μm focusing cup and Franks double mirrors. Crystal-to-detector distance was 12 cm; oscillation angle/frame was 5′. The 4·2 Å data were collected with the detector face centered on the beam (resolution at edge of detector = 4·2 Å). The 2·7 Å data were collected with the detector rotated about the crystal by 15° (resolution at edge of detector = 2·2 Å).
Data were collected from crystals of the variant surface glycoprotein (VSG) ILTat 1.24 from *Trypanosoma brucei* (Metcalf *et al.*, 1987), and from a human histocompatibility antigen (HLA-A2) (Bjorkman, Strominger & Wiley, 1985). Each VSG data set was collected from a single 0·6 × 0·4 × 0·4 mm crystal. The HLA data were collected from three 0·2 × 0·2 × 0·1 mm crystals.

Crystal	Max. res. (Å)	Exposure (s/degree)	Number of reflections* (overall)	R_{sym}† (overall)	Number of reflections* (best batch)‡	R_{sym}† (best batch)‡
ILTat 1.24 ($P2_12_12_1$, $a = 55{\cdot}1$, $b = 98{\cdot}8$, $c = 172{\cdot}9$ Å)						
Native	4·2	1080	32213/31799/6470/221	0·050	1107/892/446/0	0·030
K_2HgI_4	4·2	1080	33108/32460/6632/578	0·072	1213/1164/582/0	0·039
Pb lactate	4·2	1080	22746/22460/6591/102	0·063		
Phenyl–Hg acetate	4·2	1080	25613/25446/6599/70	0·081	1184/1092/546/0	0·040
Native	2·7	2880	38490/35983/15086/733	0·065	505/ / /4	0·034
HLA-A2 ($P2_12_12_1$, $a = 60{\cdot}2$, $b = 80{\cdot}4$, $c = 112{\cdot}2$ Å)						
Native	2·7	3600	12640/12085/3987/227	0·066	449/350/175/0	0·039

*Number of reflections = Number collected/Number included in R_{sym}/Number of unique sets in R_{sym}/Number of rejected reflections.
†$R_{sym} = \sum(|I_i - \langle I\rangle|)/\sum(I_i)$.
‡Data were divided into batches, each containing reflections from N contiguous frames, where N was 50 frames for the 4·2 A data and 20 for the 2·7 Å data. Batches were scaled to each other by the method of Fox & Holmes (1966).

signed to reduce virtual-memory access by copying data around each reflection into a non-virtual workspace before computations are performed.

The above times limit the possible data-collection speed on the dual microprocessor system. For example, given a crystal with an orthogonal unit cell of 55 × 99 × 173 Å, the minimum exposure per frame can be estimated when collecting data to 4·2 Å (the detector edge at a crystal-to-detector distance of 12 cm). For 5′ oscillation frames, approximately 20 reflections must be integrated per frame. This will require 50 s of computation time plus an average of 17 s to read each frame; 67 s on *C*2. For *C*2 to keep up, *C*1 must not produce a frame faster than every 67 s. *C*1 requires 26 s of overhead per frame so the minimum exposure time per frame is 41 s. For similar crystals in our laboratory (Table 1), we have used 90 s exposures, but for more strongly diffracting crystals, the computing hardware may limit the rate of data collection. More sophisticated integration algorithms, incorporating dynamically learned spot shapes for more accurate masking, will require significantly faster computation. We have designed the software to be as easily transportable as possible so that advantage may be taken of newer faster microcomputers as they become available. In particular, we will soon begin using a MicroVAX II (Digital Equipment, Maynard, MA) with 8 Mbytes of memory to replace *C*2, transferring the raw frames over an Ethernet network using the TCP/IP protocol. We estimate that configuration to be at least ten times faster than the present system, owing partly to the microprocessor (approximately five times faster than the current Motorola 68010), but mostly to the ability to fit the entire data-frame stack into main memory, thereby eliminating virtual-memory operations.

Another factor limiting data-collection rates with intense X-ray sources is the saturation of the detector. Typically, the detector operates in the range of 20 kHz measured photon flux when collecting data with an Elliot GX-13 rotating anode operating at 40 kV and 40 mA with 100 μm focusing cup and Franks double mirrors at 12 cm distance from crystal to detector. A measured data rate of 20 kHz produces an estimated 8% dead-time loss.*

4.2. *Crystallography*

Using the system described here, we have obtained data from several protein crystals (Table 1). Intensities were obtained from integrated counts by applying Lorentz and polarization factors calculated by the lattice prediction routine. Neither decay nor absorption corrections were calculated. 'Batches' of reflections from contiguous frames were scaled to each other using a linear scale and exponential temperature factor. The 4·2 Å data in Table 1 were all collected at

*It takes some time for the computer to store the position of a particular event in memory. During that time, another event, or several events may occur. If n events have occurred then $n - 1$ events are lost, and only the most recent is available for storage. When such a condition is encountered, the 'late count' is incremented by one. The number of these 'late counts' accumulated during an exposure provides a means to estimate the fraction of the total number of events that were lost. If the quantity (late counts/total counts) is much less than 1, then the quantity (late counts/total counts)^2 is a valid first-order estimate of the fraction lost (R. Burns, personal communication).

room temperature with continuous exposure of the crystals to the beam. The high-resolution data from VSG (variant surface glycoprotein) were collected from a crystal cooled to 277 K and an electric shutter was installed to interrupt the X-ray beam whenever data were not being accumulated. The 'best-batch' R factors are far better than the overall R factors owing to radiation damage suffered by the crystals.

The VSG native and mercury-iodide-derivative data have been used to produce an electron density map of a VSG from *Trypanosoma brucei*, ILTat 1.24 (Metcalf, Blum, Freymann, Turner & Wiley, 1987). The mercury-atom position of the derivative was determined from the Harker sections of a 6 Å difference Patterson map. A Patterson map based solely on the mercury-derivative anomalous-scattering data was also interpretable.

Recently we have begun collecting data from crystals of the human histocompatibility antigen (HLA, Table 1). Until now, useful data from HLA crystals have only been collected with synchrotron radiation.

We have found the detector and computer programs well suited for the collection of protein diffraction data, and shown the system to be capable of accurately locating and measuring all reflections incident on the detector. We anticipate its application to a wider range of crystallographic problems as the system is developed further.

We thank R. Burns for his invaluable help with the detector and computers, R. Durbin and J. Moulai for providing an off-line version of the software, and D. Freymann, M. Silver and P. Bjorkman for collecting data with this system while it was still under development. We acknowledge support from NIH (grant CA-13202 to SCH, grants AI-21324-02 and AI-13654-10 to DCW).

References

ARNDT, U. W. (1985). *Methods Enzymol.* **114**, 472–485.

BJORKMAN, P. J., STROMINGER, J. L. & WILEY, D. C. (1985). *J. Mol. Biol.* **186**, 205–210.

BOLON, C., CRAWFORD, J., DEUTSCH, M. & QUIGLEY, G. (1981). *IEEE Trans. Nucl. Sci.* NS-**28**, 816–820.

DURBIN, R. M., BURNS, R., MOULAI, J., METCALF, P., FREYMANN, D., BLUM, M., ANDERSON, J. E., HARRISON, S. C. & WILEY, D. C. (1986). *Science*, **232**, 1127–1132.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

HAMLIN, R. (1985). *Methods Enzymol.* **114**, 416–452.

HAMLIN, R., CORK, C., HOWARD, A., NIELSEN, C., VERNON, W., MATTHEWS, D., XUONG, NG. H. & PEREZ-MENDEZ, V. (1981). *J. Appl. Cryst.* **14**, 85–93.

HOWARD, A. J., NIELSEN, C. & XUONG, NG.H. (1985). *Methods Enzymol.* **114**, 452–472.

METCALF, P., BLUM, M., FREYMANN, D., TURNER, M. J. & WILEY, D. C. (1987). *Nature (London)*, **325**, 84–86.

SOBOTTKA, S. E., CORNICK, G. G., KRETSINGER, R. H., RAINS, R. G., STEPHENS, W. A. & WEISSMAN, L. J. (1984). *Nucl. Instrum. Methods*, **220**, 575–581.

WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.

XUONG, NG. H., FREER, S., HAMLIN, R., NIELSEN, C. & VERNON, W. (1978). *Acta Cryst.* A**34**, 289–296.