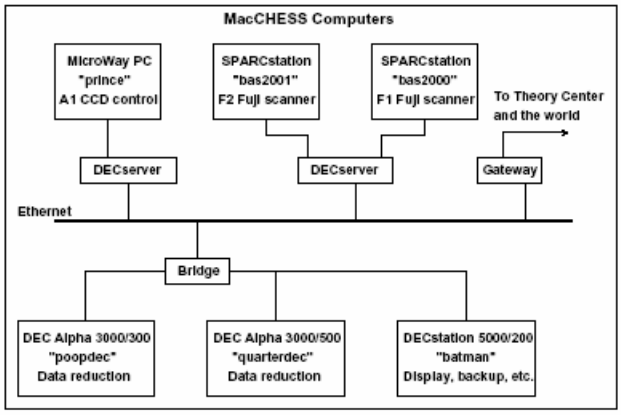
MacCHESS data reduction capability

Marian Szebenyi

Early crystallographic users of CHESS collected precession, oscillation, or Laue data on film and took the films home to process; there were no facilities here for data reduction. The 20 minute time required to digitize a film made it completely impractical to process data as fast as they were collected. With the installation of a Kodak storage phosphor data collection system in 1988, the time to digitize an image was reduced to 5 minutes, and then MacCHESS director Keith Moffat set the goal of having users reduce data on-site and go home with h, k, l, F, sigma(F)'s instead of reams of raw images. Today we are well on the way to achieving this goal.

The first figure shows the components of the MacCHESS data collection and processing system. Most data are now collected on Fuji image plates and scanned on one of the two Fuji BAS-2000 scanners. These scanners have disk storage space for about 100 images each. Using the Network File System facility, the processing computers have direct access to the raw data files. Image files from the recently installed CCD detector are currently transferred for processing via ftp or 8-mm tape, but the planned installation of NFS software on the CCD control computer will allow images to be written directly to a remote disk.

Data processing is done on two DEC Alpha workstations, one Model 3000/500 (called quarterdec) and one Model 3000/300 (poopdec). These workstations were obtained under a grant from Digital Equipment Corporation to Steve Ealick. In addition, a DECstation 2000/500 (batman) is available for operations not requiring the great speed of the Alphas. For display and processing of oscillation images, the principal software is the HKL system, created by Zbyszek Otwinowski and Wladek Minor. It incorporates the display program XDISPLAYF (also called IPVIEW)



and the data processing programs DENZO and SCALEPACK.

The single command "denview" starts up IPVIEW and DENZO. The user enters the name of the first image to process, and it appears on the display. The user can also enter the wavelength and crystal-detector distance and indicate the position on the image of the direct beam, after which IPVIEW will provide a running display of the resolution at the cursor position. Any region of the image can be enlarged in a "zoom window", to evaluate spot shape and spacing. Overloaded pixels can be marked. The display can be in false color or gray-scale; the scale is determined automatically but can be easily altered by the user.

To begin processing, the IPVIEW peak search routine is used to obtain a list of spot positions on the image; the list is written to a file which is then read by DENZO. Commands to DENZO are typed into a window which is created by "denview". Usually files of commands are set up ahead of time and read in by typing "@filename". Given the list of peaks from IPVIEW and a few basic parameters such as wavelength and crystal-detector distance, DENZO automatically determines the crystal orientation and provides a list of cell parameters in all possible crystal systems. Cell parameters and space group may be input but are not required. If they are provided, DENZO will accept the space group but redetermine the cell parameters. After indexing, DENZO refines crystal, camera, and detector parameters. Almost all parameters can be refined at once. Sample input files are available giving a reasonable sequence of refinement steps. During refinement, the predicted spot positions can be displayed on the image (see figure), so that agreement with the observed positions can be directly evaluated. The second figure shows a DENZO run midway in the refinement process. After refinement, DENZO integrates the reflections on the image, writing the results to an ASCII file, along with the refined parameters. The program is then ready to process the next image in the series. Autoindexing need be performed only once per series, but some refinement is done for each image before it is integrated.

The program SCALEPACK is run after a series of images have been processed. It determines relative scale factors between the images, can

post-refine cell parameters, mosaicity, and crystal orientation, computes Rsym's, and merges the data into a single set of h, k, l, F^2, and sigma.

In addition to HKL, programs available on the MacCHESS computers include:

* The alternative image display program M.DISPLAY, which is less convenient but more flexible than IP-VIEW.

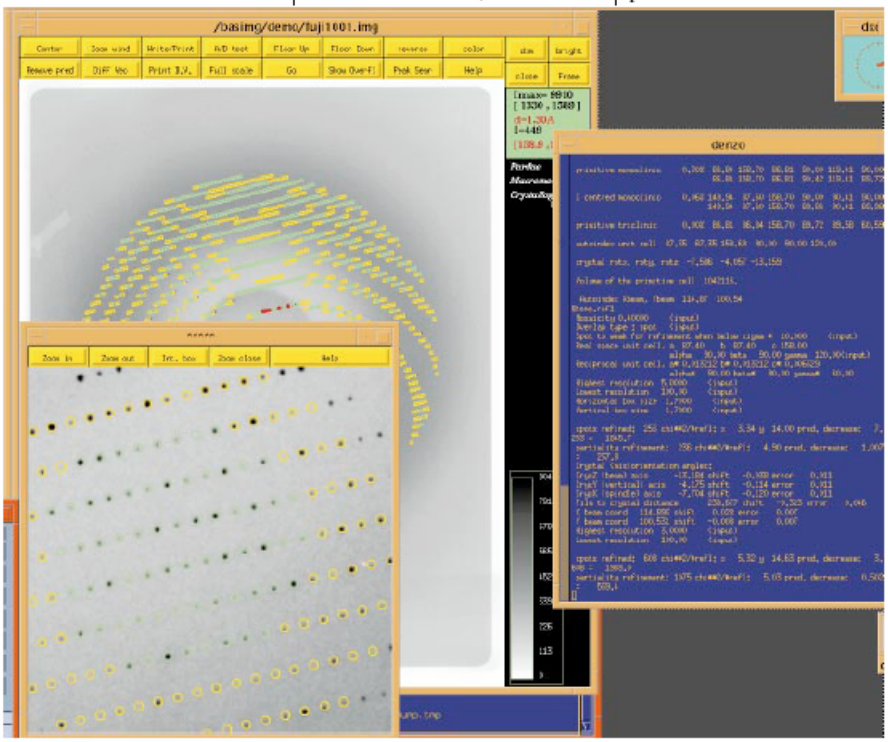
*The alternative oscillation processing suite MOSCO, a part of the CCP4 package.

* The MacCHESS Laue processing package, consisting of the indexing program M.INDX, the refinement and integration program M.LAUE, and the scaling program M.SCALE.

* Miscellaneous ancillary routines.

Users are also welcome to temporarily install their favorite programs. The operating system is OSF/1 on the Alphas and Ultrix on the DECstation. Fortran, c, and X windows are available. The window management system is Motif on the Alphas and DECwindows on the DECstation.

One of MacCHESS's long-range goals is to have a true on-line data reduction system, in which data are processed as fast as they are collected. This is commonly done on laboratory area detector systems, but conditions at CHESS are more demanding. The Fuji scanners have a readout time of about 90 seconds. The CCD detector betters this time by a factor of four. With typical exposure times of a few seconds, an on-line system needs to be able to process an image in less than 30 seconds in order to keep up. The time to process an image on quarterdec is currently about two minutes, not including the time to index one image out of each series. Improvement in speed may be possible (aside from waiting for the next generation of computers) by adaptation of DENZO to the parallel-processing environment of the supercomputers at the Cornell Theory Center. Development of foolproof on-line data processing software will also require innovative programming to make the system run as automatically as possible and yet be able to detect when something has gone wrong and human help is needed. Efforts along these lines are being pursued.



Partial computer screen showing the predicted diffraction peak positions (yellow and green circles) overlaying the image data.