

ARCHIVING CCD/ELECTRONIC ASTRONOMICAL DATA

David Huenemoerder¹
Jet Propulsion Laboratory and California Institute of Technology
U.S.A.

The availability and advances in two-dimensional electronic detectors, in particular the charge-coupled-devices (CCDs), are a great asset to astronomical imaging and spectroscopy because of their sensitivity, dynamic range, and linearity. In some cases photographic plates still offer an advantage to imaging of large size, but the advent of large format CCDs may make a figure of merit, the area per exposure time, much more favorable for CCDs.

Photographic plates have one advantage that CCDs do not. Once the plate is exposed and properly developed, it may be stored in a vault for posterity. The primary archive is not then a question of medium, but of organization. With CCDs or other electronic data, both the storage media and organization of data are open issues.

What are the requirements of CCD data? Let's imagine that 2048 pixel square chips are standard, and that they have dynamic ranges in excess of 100,000. We then will likely use 24 bytes per pixel, and have 12 megabytes per data frame. If, as for imaging applications, chips are arranged in a two-by-two array, we will then have 48 megabytes per data frame. We can easily scale this to typical observing programs and come up with a tremendous volume of data, and conclude that we need a dense medium (such as optical disks or DAT) and also sophisticated real-time compression techniques.

What data should be saved? Everything, of course, for many reasons. (I'm assuming that "flat-fields", "bias-frames", and other relevant calibration data are averaged on-site.) Blank regions of the CCD are either non-existent or inconsequential in volume (the greedy echelle or multi-object spectra have interorder scattered light; an image has sky background), so there is no "data" to discard. Advances in computer hardware and software result in the practicality of more powerful techniques which we can only take advantage of if the original raw data is available. One astronomer's noise is another's signal, and echelle spectra produce much "noise" to be mined by alternate techniques.

We should make an effort to provide a "standard reduction" which will meet most needs, but generic reduction schemes cannot satisfy everyone. There are many levels of data

¹ Address: NASA Headquarters, Astrophysics Division, Code SZ, Washington, DC 20546

reduction and calibration, and many ways to do it. It is a personal process requiring many scientific judgements which can only be made by the investigator as directed to specific scientific goals. A standard product should be archived along with the reduction history so that judgements may be made on the suitability of the reduced data for a given purpose. The organization of an archive is as important as the data in it. A "write-only" archive is of limited utility. There must be an index to the data, probably at some central location. There must be a method to request and receive data, and the data must be provided in some portable form (such as FITS format).

Policy decisions will determine whether electronic data are ever archived. We have to decide on data ownership and proprietary periods. It is INCENTIVE to tell someone that the community will be eternally grateful if the effort is made to archive data. It is MOTIVATION to tell them that they won't get funding or observing time if they don't archive the data. These are issues for the funding agencies and observatory users' committees. (Motivation works, incentive doesn't.)

A minimalist approach is desirable. Much energy can be (and is) expended arguing about which software or hardware is "better." We can let people use what they are comfortable and familiar with, as long as we have a readily accessible index and can retrieve data in some portable format and can properly interpret it. We can relegate special needs to individual investigators; producers of generic facilities must remember that "better is the enemy of sufficient." A diversity of approaches will assure longevity and survival; homogeneity must be balanced with utility. Most of the issues in archiving deal with policy and are not specific to CCD data, but CCDs will produce a large volume of high-quality data which is in all of our interests to preserve and make accessible.