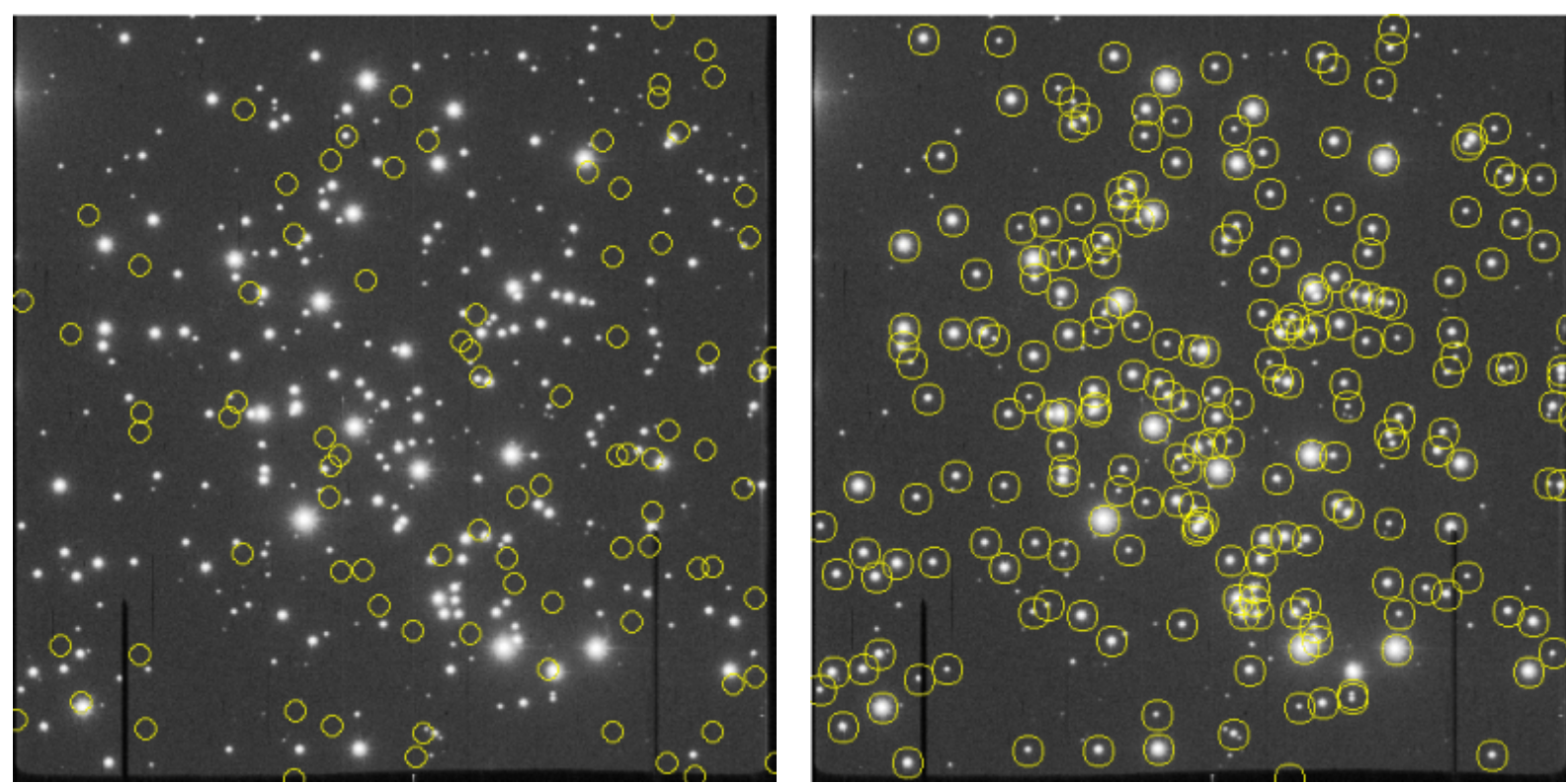


# Creating Data that Never Die: Building a Spectrograph Data Pipeline in the Virtual Observatory Era

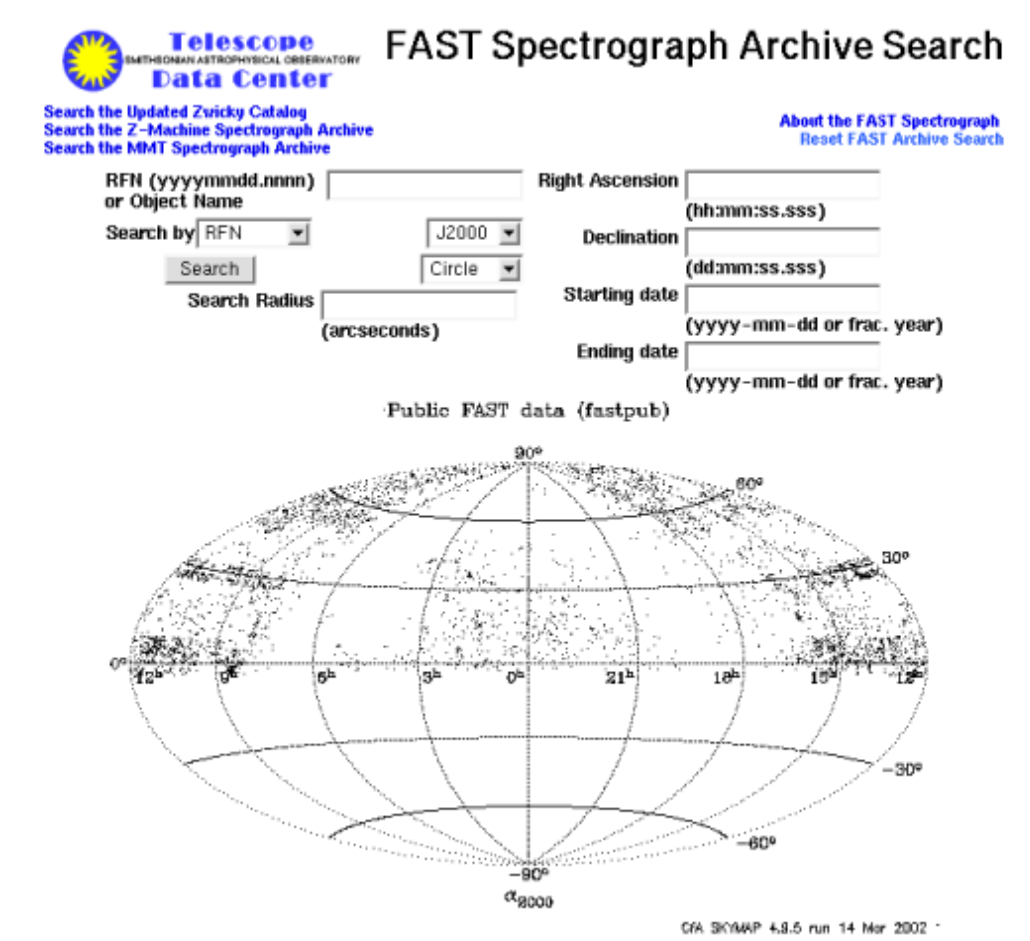
Doug Mink, William Wyatt, John Roll, Susan Tokarz, Maureen Conroy,  
Nelson Caldwell, Michael Kurtz, Margaret Geller  
*Smithsonian Astrophysical Observatory*



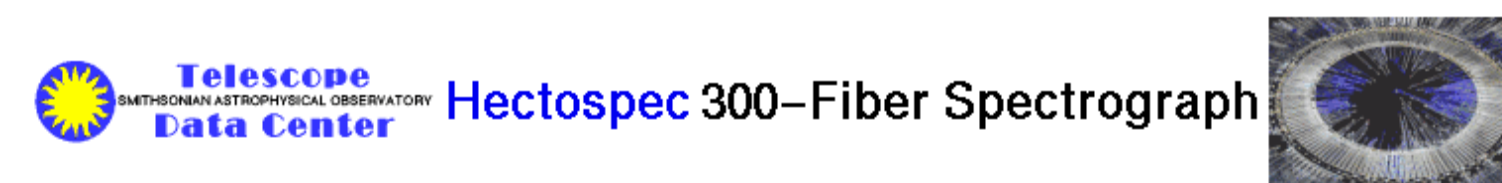
## Program Conception

- Instrument information from the Web
- Previous observations from ADS and VO
- Catalogs from VO
- Positions from images reduced using VO catalogs and WCSTools

Data pipelines for modern complex astronomical instruments do not begin when the data is taken and end when it is delivered to the user. Information must flow between the observatory and the observer from the time a project is conceived and between the observatory and the world well past the time when the original observers have extracted all the information they want from the data. For the 300-fiber Hectospec low dispersion spectrograph on the MMT, the SAO Telescope Data Center is constructing a data pipeline which provides assistance from preparing and submitting observing proposals through observation, reduction, and analysis to publication and an afterlife in the Virtual Observatory. We will describe our semi-automatic pipeline and how it has evolved over the first nine months of operation.



**Virtual Observatory**  
After a proprietary period all spectra will be made available through the VO as we already do for some of the spectra from our FAST spectrograph.



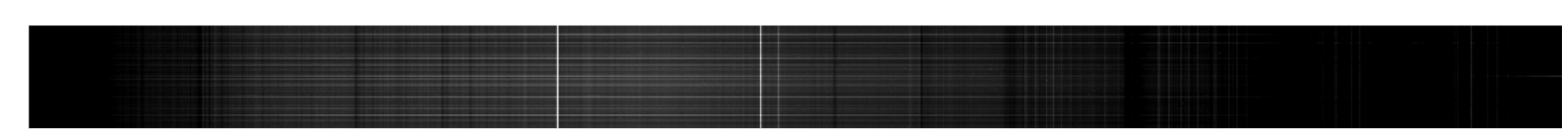
**Hardware** Hectospec is a moderate-resolution, multiobject optical spectrograph fed by 300 optical fibers. It operates at the f/5 Cassegrain focus of the 6.7-meter MMT on Mount Hopkins in Arizona. The instrument is composed of a robotic positioner and a large, bench-mounted spectrograph, joined by a bundle of 25m long optical fibers. Dual robots, dubbed Fred and Ginger, reconfigure all 300 optical fibers in just 300 seconds.

**Processing** [Preparation] [Observation] [Reduction] [Distribution] [Archiving]  
Getting Good Coordinates for Hectospec Hectospec Header Keyword Dictionary

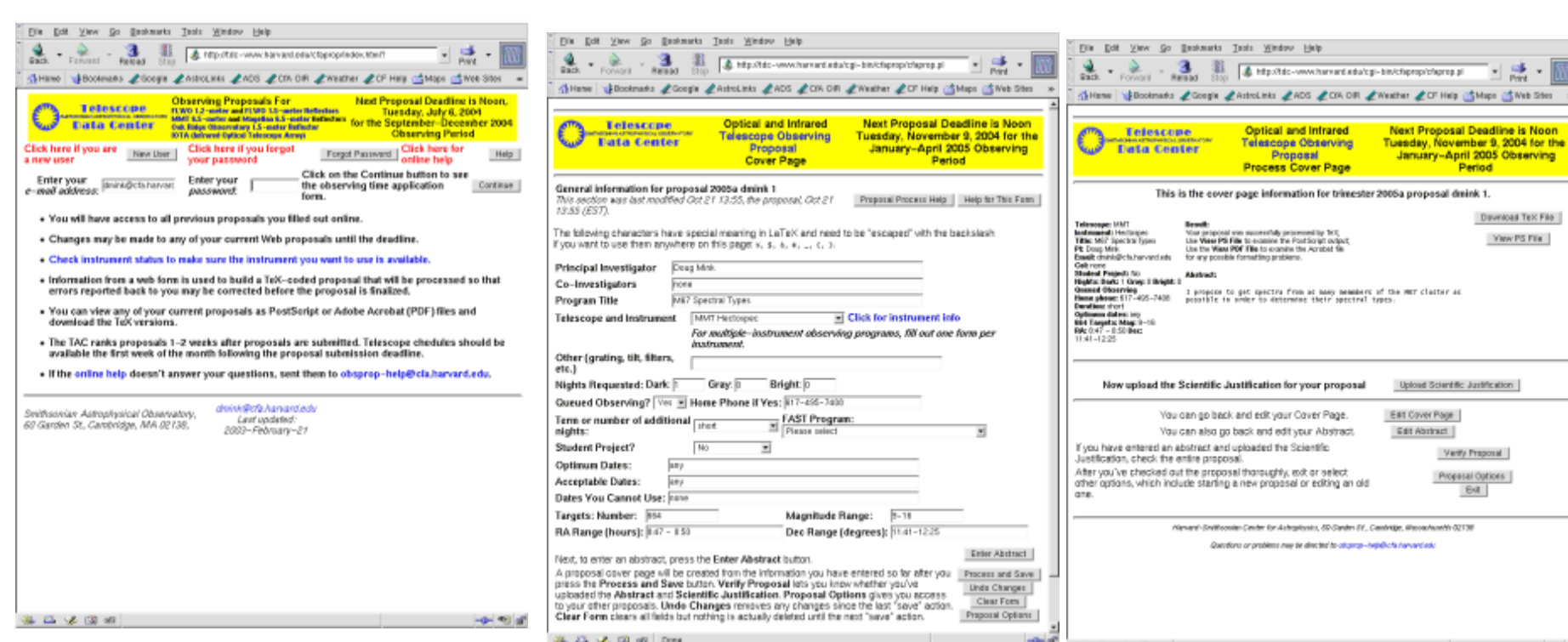
**Sample Image**  
(Click for larger image)

**Sample Extracted Spectra**

Good signal-to-noise data from a very preliminary reduction of engineering data taken in October 2003 (absorption and emission lines are marked based on cross-correlation redshifts)  
(Click for larger graphs)  
[Here is some information on sky subtraction testing.](#)



300 stacked spectra as seen by SAOImage



## Proposal Submission and Evaluation

- Web-based Proposal based on NOAO Perl scripts  
-> online archive and database of proposals
- Web-based TAC evaluation software  
-> database of accepted proposals

## Data Archiving

- Raw data
- Reduced data in single file per pointing
- Reduced data in one file per object
- PI's have access to their data

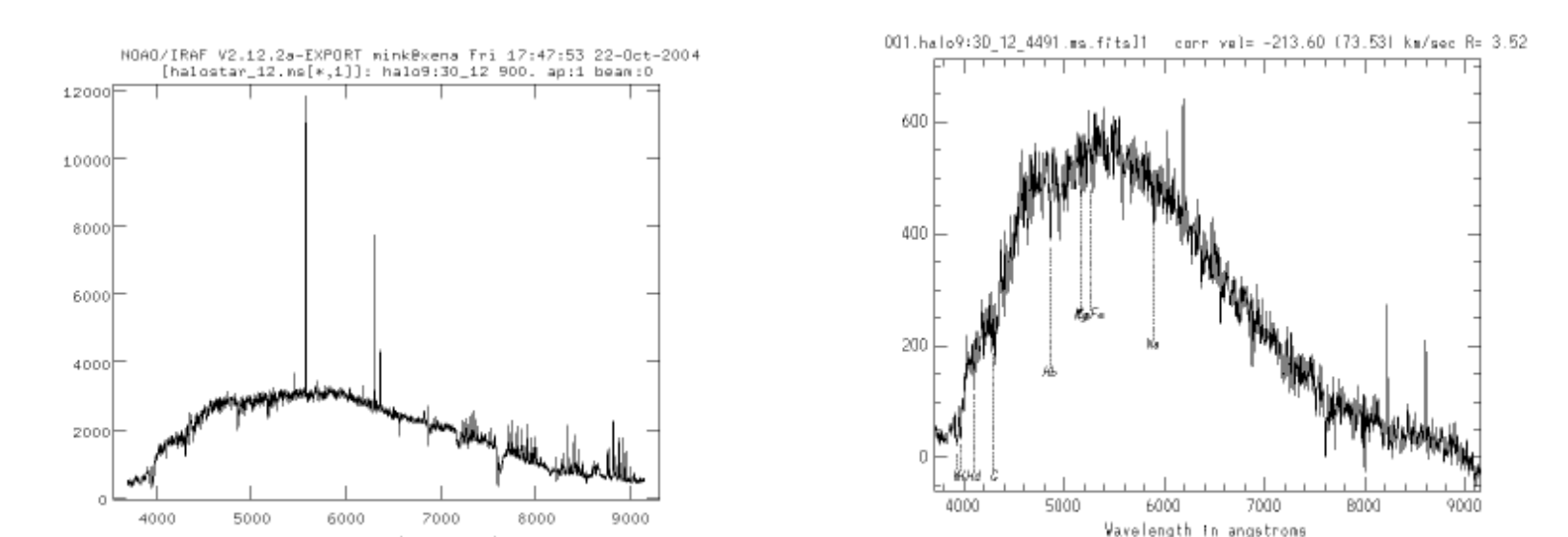
This description of Hectospec on TDC website is used by the scientist planning to observe, the person reducing the data, and the VO user who wants to understand what they are getting.



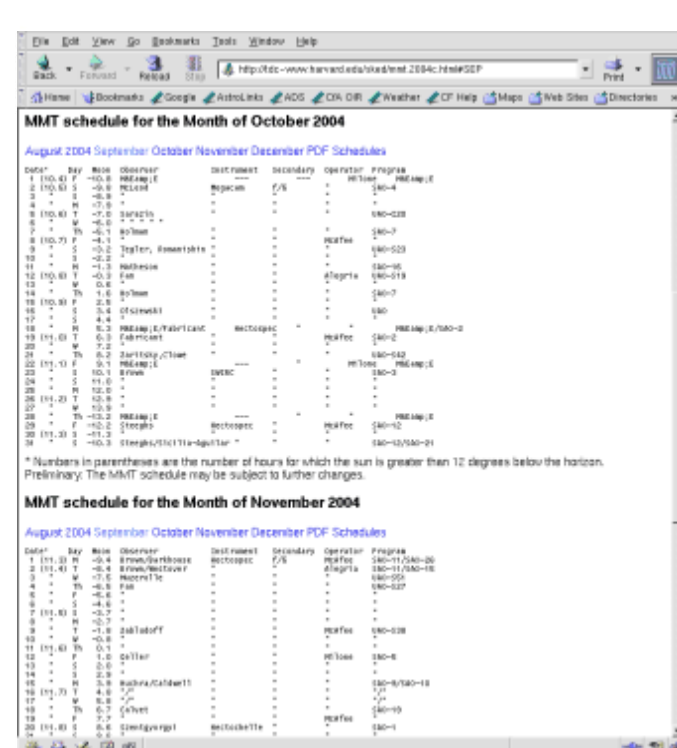
Hectospec focal plane



Hectospec on the MMT



Extracted spectrum is mostly sky Same spectrum without sky

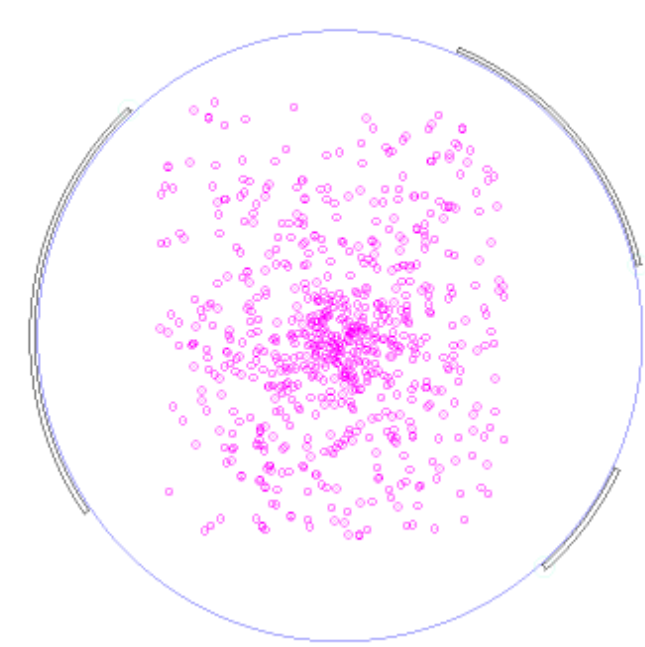


## Scheduling

- Web-accessible telescope schedules
- TAC-assigned program numbers follow data throughout its life

## Data Reduction

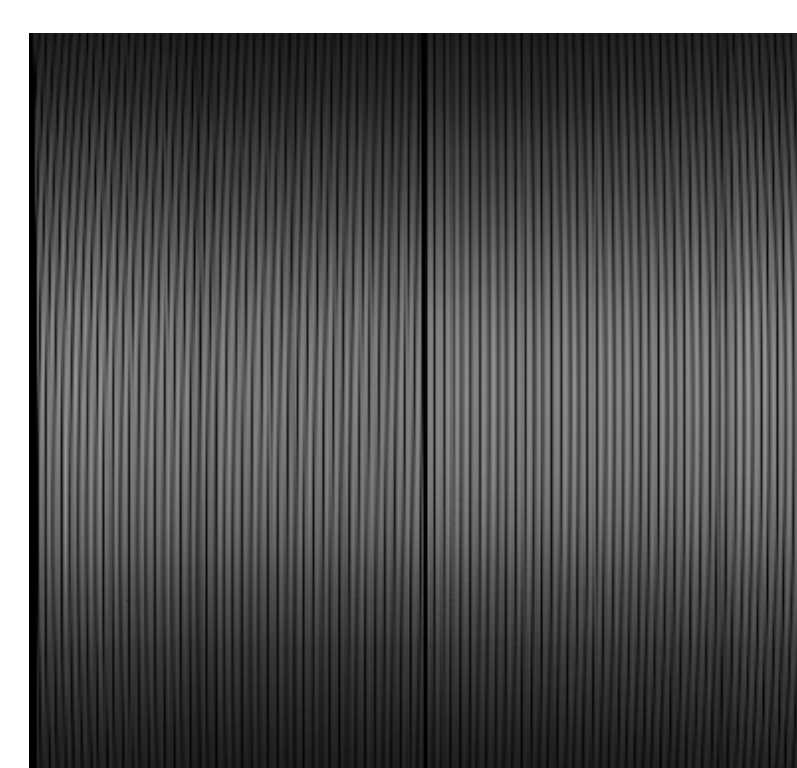
- Local IRAF CI and KSH scripts
- IRAF noao.twodspec



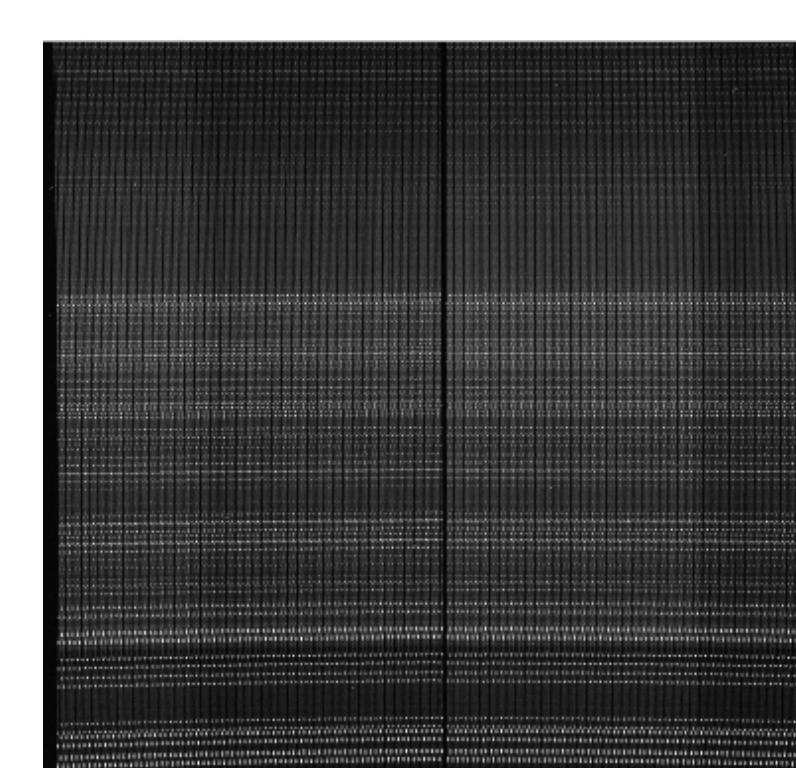
Fibers positioned for M67

## Observing Preparation

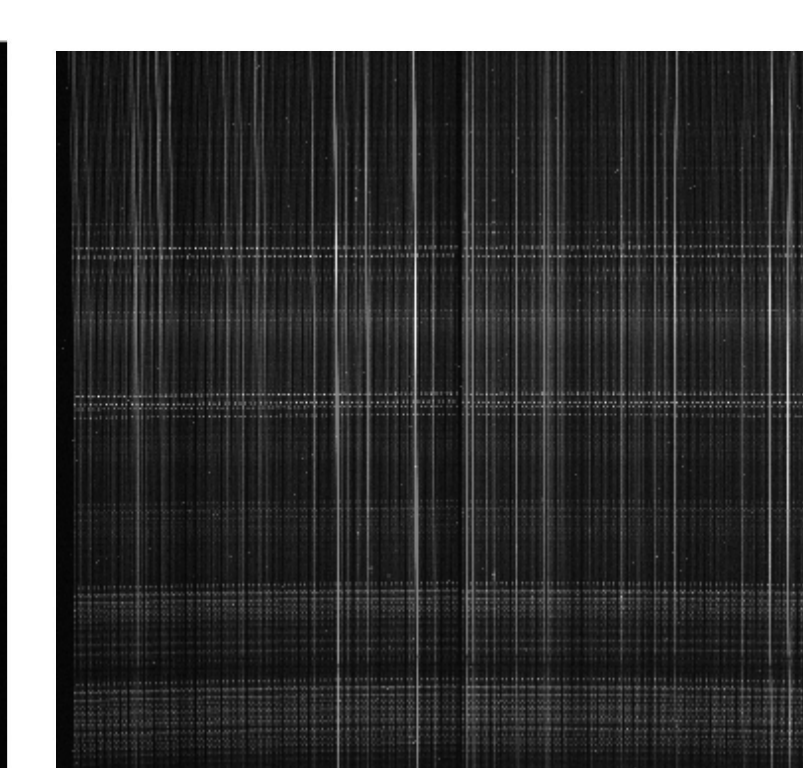
- Fiber positions from PI object catalogs
- Sky positions found using deep all-sky catalogs
- Interactively assign fibers to program objects



Dome flat spectra



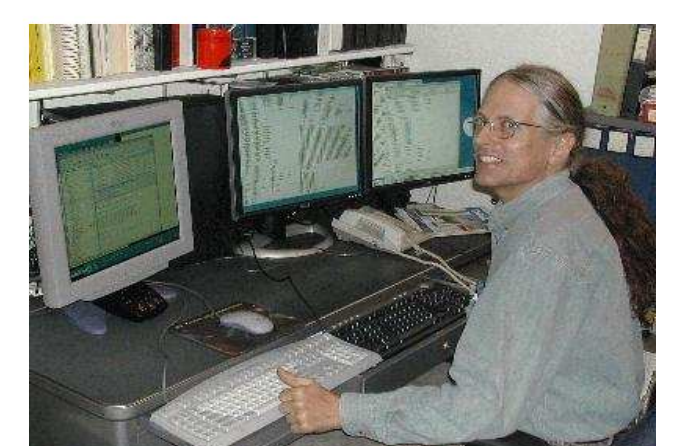
Calibration lamp spectra



Object spectra with night sky

## Data Acquisition

- 3 exposures per pointing are taken so cosmic rays can be removed
- Control software uses IRAF ICE
- Data is immediately ftp'ed to Cambridge



Doug Mink  
dmink@cfa.harvard.edu  
<http://tdc-www.harvard.edu/mink/>