CAL-1957

National Aeronautics and Space Administration

MARSHALL GRANT

FINAL TECHNICAL REPORT

IN-89-CR

FOR NASA GRANT NAG8-183

62499

(NASA-CR-189768) A COMPLETE DATABASE FOR THE EINSTEIN IMAGING PROPORTIONAL COUNTER Final Technical Report, 31 Jan. 1991 - 30 Jan. 1992 (Columbia Univ.) 9 p CSCL 03A

N92-16960

P9

Unclas
G3/89 0062499

Submitted to: National Aeronautics and Space Administration
Technical Officer, Code SZ
NASA Headquarters
Washington, DC 20546

Submitted by: The Trustees of Columbia University
in the City of New York
Box 20, Low Memorial Library
New York, New York 10027

Prepared by: Columbia Astrophysics Laboratory
Departments of Astronomy and Physics
Columbia University
538 West 120th Street
New York, New York 10027

Title of Research: **"A Complete Database for the Einstein Imaging Proportional Counter"**

Principal Investigator: David J. Helfand
Professor of Physics

Period Covered by Report: 31 January 1991 – 30 January 1992

January 1992

1. Introduction

Under our proposal entitled "A Complete Database for the *Einstein* Imaging Proportional Counter" we proposed to:

1) complete integrity checks on the database
2) complete a manual on the database and analysis tools we have developed for its use
3) enhance the available software to provide easy interface to off-line analysis tools
4) develop a complete set of flat field images as a function of gain, viewing geometry, background level, etc.
5) explore further the optimal editing strategies for source detection and studies of diffuse emission
6) pursue scientific problems such as determining the fraction of the X-ray background which arises from Galactic emission, searches for extended sources (e.g., galaxy clusters and synchrotron nebulae), and searches for subthreshold sources.

This proposal was funded as NAG8-183 and was in effect from January 31, 1991 to January 30, 1992. As we describe below, all of the principal goals of this proposal have been achieved. In addition, we have provided access to the database, software and data products to over a dozen outside users who found it uniquely capable of meeting their needs for IPC data. We describe below the original data which make up the archive (§2 – cf, goal 1 above), the structure of the database itself (§3 – cf, goal 2), the Op-Ed analysis system (§4 cf, goal 3), the technical advances we have achieved relevant to the analysis of IPC data (§5 – cf, goal 4), the data products we have produced (§6 – cf, goals 5 and 6), and some of the uses to which the database has been put by scientists outside Columbia over the past year (§7).

2. The Data

The imaging proportional counter (IPC) at the focus of the *Einstein* Observatory X-ray telescope collected data in the 0.1 – 4.5 keV X-ray band from 4082 $1^\circ \times 1^\circ$ fields scattered over the celestial sphere. The angular resolution of the instrument was $\sim 1'$ and the spectral resolution scaled as $R \sim 0.5(E/1\ \mathrm{keV})^{-1}$; effective exposure times ranged from $\sim 10^2$ to $\sim 10^5$ seconds (see Giacconi *et al.* 1979 and Gorenstein, Harnden and Fabricant 1981 for further details concerning the Observatory and the IPC). Over the course of the *Einstein* mission, the IPC recorded, while in the pointed mode, the position (in $8''$ cells), energy (in 32 pulse-height bins), and time-of-arrival of 230 million "events" arising from the deposition of energy in the counter gas. These events resulted from a number of distinct stimuli:

- cosmic ray interactions with the detector – low energy electrons and gamma-rays may be directly detected, while higher energy cosmic rays produce spallation in the walls of the counter as well as neutron activation of the detector and spacecraft, leading to secondary events in the instrument;
- a low-level leak of the Cm/Al fluorescence calibration source;
- detector and/or electronic malfunctions (e.g., breakdown or "sparking" in the counter gas);
- solar X-rays scattered from the residual atmosphere above the satellite and collected by the mirror;
- discrete cosmic X-ray sources such as stars, supernova remnants, galaxy clusters and active galactic nuclei (AGN);

- diffuse X-rays of Galactic origin – the hot bubble of gas $\sim$ 100 pc across which surrounds the Sun (McCammon *et al.* 1983), the ridge of emission along the Galactic plane (Iwan *et al.* 1982; Koyama *et al.* 1986), and the putative halo of hot gas surrounding the Galaxy; and
- the cosmic X-ray background.

In addition, the instrument and other satellite subsystems provided housekeeping data such as high voltage settings, temperature, gain, particle background level, etc. as well as information necessary to calculate the spacecraft's position and orientation with respect to the Earth, the Sun, and the celestial sphere.

3. The Database

From this telemetry stream, SAO constructed a number of files essential for the analysis of the data:

- .XPR is the time-ordered event file representing an orbit's worth of science data (in the jargon, a HUT). Each event is assigned 20 bytes which include its recorded x, y detector position and its position rotated into celestial coordinates, its pulse height bin and a gain-corrected energy bin assignment, and its arrival time. The file header includes such information as the start and stop time for the file, the target coordinates, the detector gain, identifying information for the observer and target, and summary statistics for the file. There are 11,230 HUT.XPR files in the total pointed-mode database.
- .TGR is the file containing information which characterizes instrumental and environmental factors affecting data quality. This includes particle background levels, high voltage settings, South Atlantic Anomaly status, aspect quality, telemetry problems, calibration source status, Earth-Sun-satellite geometry, etc. The .TGR file consists of "timing gap records" which are each 160 bits long. A new record is written each time any one of the 33 status parameters changes. There is one such .TGR file for each .XPR file.
- ASP.MAG is the file containing the satellite's position and orientation information. This includes for each telemetry major frame (40.96 seconds) a three-vector representing the satellite's position in ECI50 (Earth-centered inertial) coordinates, a translation of these coordinates to longitude, latitude, and height above the Earth, the Sun's position, the Earth-Sun angle, and the orientation of the Earth's magnetic field with respect to the spacecraft. There is one such file for each .XPR file.

Additional files were created by SAO to include such information as the star-tracker data necessary for deriving the aspect solution, for archiving ground-based calibration data, for monitoring the gain from in-orbit calibrations, etc. Our database, however, includes only the above three files which are sufficient for nearly all scientific investigations.

In limiting the database to these files we have implicitly adopted the aspect solution, the gain vs. time record, and the translation of certain telemetry records (such as guard counter rate to background level flags) produced by SAO. We retain the ability, however, to allow the user to produce new pulse-height to energy conversions, to calculate new flags for such parameters as solar contamination level (known as "viewing geometry" flags) and, most importantly, to select which data he or she decides is "good" for the scientific question at hand (see below). The full set of 33,690 files currently resides on three large-format (WORM) optical disks and comprises $\sim$ 5 Gigabytes of data. The disk reader is connected to a Convex C1-XP. The database is accessed through an extremely flexible data retrieval, editing, and analysis system Op-Ed (Optical-disk Processing for *Einstein* Data) which we describe below.

4. The Op-Ed Analysis System

The entire analysis system is written in documented standard FORTRAN. At present, it performs the following functions:

> SELECT: Allows the user to choose data on which to work by orbit number (HUT), sequence number, celestial coordinates or optical disk side (e.g., for use in running a particular analysis algorithm on the entire database), or any ranges thereof. In addition, there is an emend option to exclude some portion of a selected subset of the data or include additional datasets outside the originally selected parameter range. This program operates on an archive catalog which resides on magnetic disk and includes HUT number, sequence number, 1950 celestial coordinates, satellite roll angle, total image count rate, size of the file (typically a few hundred kilobytes), and disk location.

> LOAD: Retrieves from the optical disk all selected data. Retrieval times are a few seconds if the data reside on the currently mounted disk and ~ 45 sec if a disk must be flipped or changed. Various utilities for magnetic disk management are also available.

> SCAN: This program allows the user to create images from the time-ordered photon files edited to optimize the image for the project at hand and then to analyze the image for point or extended sources. There are 13 flags for each photon, many with multiple values, so that a total of up to 101,250 different images can be created for each HUT, sequence number, or set of overlapping sequence numbers. A few examples of editing choices and their utility include:
>
> `Viewing geometry` – to create images free of solar scattered flux.
>
> `Energy bin` – to create images in different energy bands, to search for ultrasoft or heavily obscured sources, to minimize various background contamination levels, etc.
>
> `Aspect quality` – to recover the $\sim 20\%$ of the database (100% in some fields) never before examined which was taken when only the gyros were controlling satellite attitude (typical smearing $< 2'$).
>
> `Masked out data` – to recover the full field of view of the instrument, not just the $60'$ fields of standard images. This is a $\sim 35\%$ increase in area.
>
> `Background level` – to provide the cleanest possible data for diffuse studies or to extend timing or spectral analysis on bright sources where background is not critical.
>
> Once the user has chosen appropriate editing criteria, a flat field (see Fig. 1) is produced from the entire source-subtracted database using precisely the same criteria, and an option to use this customized flat field or the default flat field is presented (e.g., if the editing criteria select only a small fraction of the total database and the statistical quality of the customized version is insufficient, the default option would be chosen). This process takes < 5 sec (see below). The flat field is then applied to the exposure map. Backgrounds are never subtracted from images to avoid compromising the statistical properties of the data.
>
> The image is then examined with a sliding circular aperture of selectable radius, and local backgrounds from concentric annuli (also selectable radii) are computed to determine significance. All features exceeding a specified threshold are tabulated.

> EXAM: Using the source list from SCAN or from a user-specified file (e.g., a radio catalog or a list of stars), this program fits a two-dimensional Gaussian to each location and reports source parameters such as position, flux, hardness ratio and spectrum.

> UTILITIES: Programs are also included in the system to allow the user to generate ASCII, AIPS, or FITS files for images, spectra, time series, etc. for use in his or her favorite image display or analysis system. One example is EXTRACT-EXTRA which creates a flat field from a 6 Mbyte hypercube of the entire source-free data set using any combination of energy channels, viewing geometries, and aspect codes, reports the image's statistics (rms, minimum, maximum, and number of photons per bin) and outputs an ASCII or FITS file. This has been of value in studying flat fielding problems with the instrument and in analyzing the diffuse X-ray background.

The system is flexible, powerful and efficient, and in many ways complements the SAO system which has more extensive image display and manipulation capability but leaves the user one step removed from the raw data. Our approach has been to provide all users with the capability to access and control the creation of images and the source detection process, and then to employ their preferred systems (IRAF-PROS, AIPS, IDL, MIDAS, etc.) for further analysis. A run through the entire database using any desired editing and search techniques (e.g., to produce the catalogs described below), takes < 40 hours. In the context of such a system, the concept of "reprocessing" loses its traditional meaning. Indeed, each user is allowed to "reprocess" and then analyze any or all of the data in a manner designed to optimize the results for the scientific problem of interest.

5. Technical Advances Achieved

In order to characterize most completely the environmental and instrumental artifacts which corrupt a dataset it is clearly desirable to have the maximum amount of information available. This situation only obtains after the mission has ended. In the course of developing the archive and analysis system described above, and in applying it to various scientific problems, we have developed considerable novel information on the performance of the IPC during the mission and have constructed various innovative approaches for the removal of instrumental signatures and contaminating signals from the data. Many of these have been described in recently published papers; we simply enumerate them here with the appropriate references.

Flat fielding: The standard IPC image contains flat fielding artifacts with $\pm 30\%$ amplitude and structure on all scales from $1 - 30$ arcminutes (Fig. 1). Ignoring these effects substantially comprises studies of faint point sources, bright diffuse sources, and the diffuse background. We have provided an analysis of these effects along with examples of the problems arising therefrom in Wu *et al.* (1991) and Hamilton *et al.* (1991). The effects on the Coma Cluster IPC image are reproduced below in Figure 2. As described above, the Op-Ed system automatically corrects all flat-fielding problems.

Solar Contamination: Up to 30% of the counts in a standard IPC image are solar photons scattered off the residual atmosphere above the satellite. This contamination can be avoided through Op-Ed's editing options or can be estimated through the use of techniques we describe in Wang *et al.* (1991).

Particle contamination: Up to 40% of the total counts (and 70% in some energy channels) in a standard image arise from cosmic rays interacting in the detector. We have established the mean intensity and its dispersion as well as the spectrum of these events (Wu *et al.* 1991). Editing can reduce the particle background intensity, and carefully selected energy bands can be

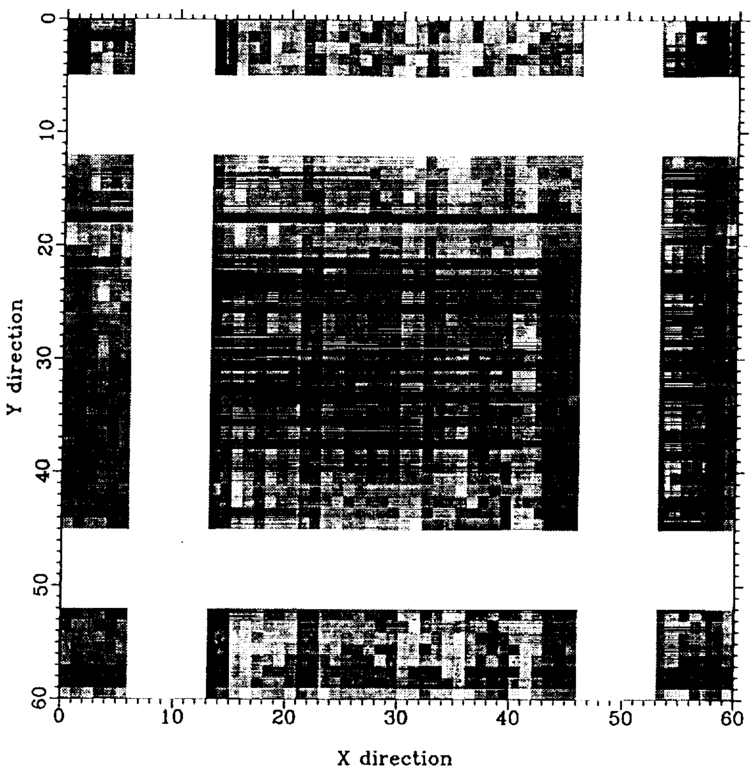

Fig. 1.—Gray-scale representation of a flat-field image for the IPC in the "broad band" (0.16–3.5 keV) derived by summing 414 source-subtracted images rotated into machine coordinates. The pixels are 1′ × 1′, and the peak-to-peak fluctuation in the image is 0.75–1.32 about a mean value of 1.0 with an rms of 0.09. Regions within ±3′.5 of the window support ribs have been deleted.

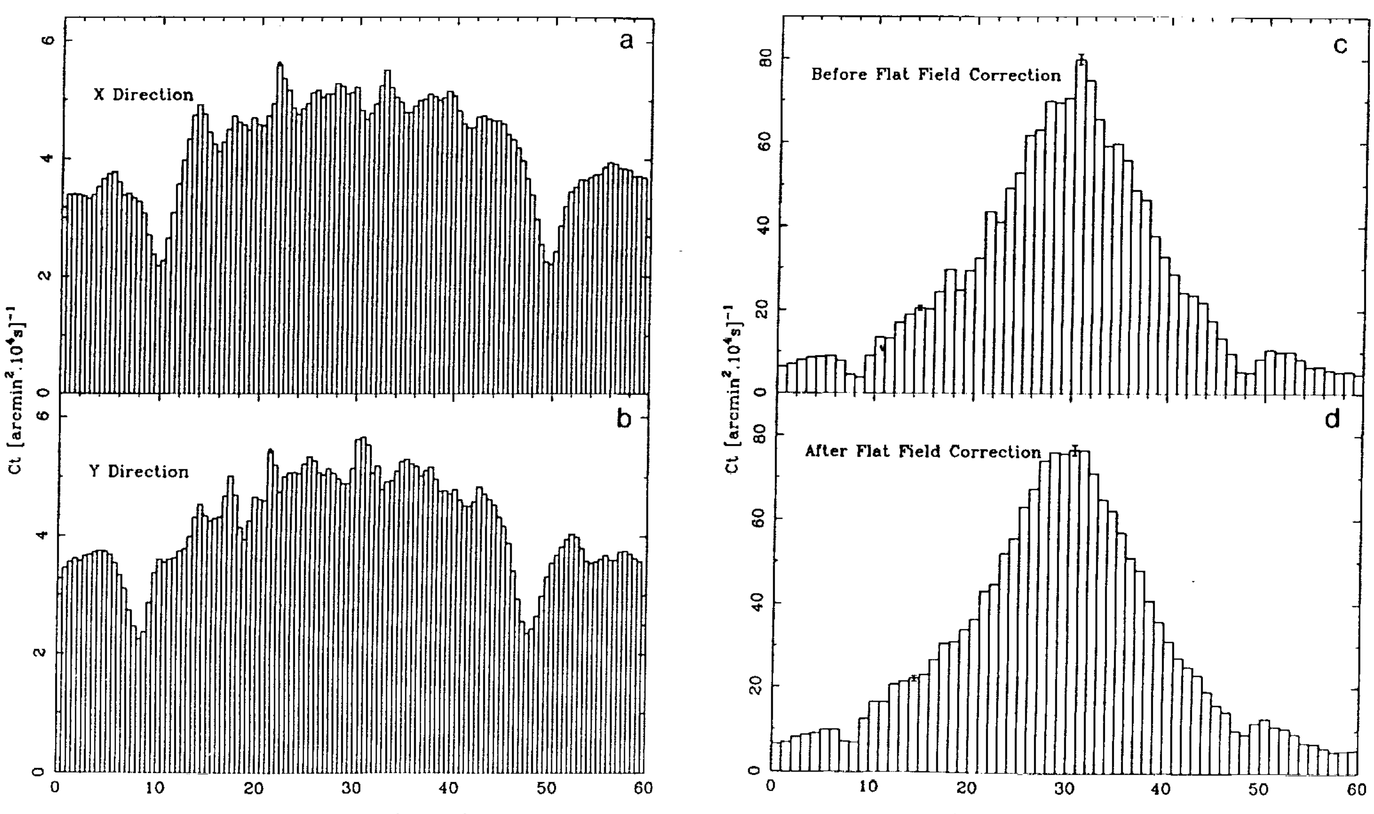

Fig. 2.—Projections along the (*a*) *x*-axis and (*b*) *y*-axis of the IPC detector with 30″ resolution. Large deviations corresponding to gain differences between adjacent detector anode wires are apparent. Error bars representing $\pm 1\ \sigma$ uncertainty resulting from counting statistics alone are indicated. (*c*) Projection along the detector *y*-axis of an observation of the Coma cluster of galaxies (IPC sequence number 1793). Apparent structure of the X-ray emission near the cluster center clearly matches the flat-field fluctuations seen in (*b*) and is purely an instrumental artifact. (*d*) Actual smooth distribution of Coma X-rays (modified only by the mirror vignetting function) is clear after the application of the flat-fielding algorithm.

chosen to maximize signal-to-noise values for a particular project. In regions with overlapping fields, the particle rate can be solved for directly (Wu *et al.* 1991).

`Calibration source leak:` The spectrum of this source of contamination has also been solved for.

`Mirror scattering:` A technique for removing the scattering wings surrounding bright point sources has been developed by Wang and Helfand (1991) and is on the system.

`Source search optimization:` Using Monte Carlo simulations, we have determined and implemented an optimum strategy for discrete source detection (Hamilton *et al.* 1991) which can be tailored to fit different scientific objectives.

Other technical issues which are addressable with our system and which could improve the quality of extracted data include the energy dependence of rib shadowing, the exact characterization of no-aspect data, and other components of the position- and time-dependent PHA-to-energy translation.

6. Data Products

In pursuit of goals 5 and 6 in our original proposal, we have developed a number of catalogs from the dataset which in themselves are of general use for the IPC researcher. These include:

The Two-Sigma Catalog: A compilation of all fluctuations exceeding a locally determined background by $\geq 2\sigma$ in each image using default editing criteria. Cross-correlation of this 10^5 "source" list with catalogs from other wavelength regimes provides a powerful way to compile large source samples for further study. For example, comparison with the Greenbank 6 cm survey has yielded a list of $> 10^3$ X-ray selected radio sources. Optical followup of the first half dozen of these objects has turned up a $z = 3.89$ quasar, the highest redshift object known at X-ray wavelengths (Helfand *et al.* 1991a), as well as several moderate redshift radio-bright galaxies which may be undergoing starbursts. Chance coincidence rates in these catalog comparisons are dependent on the number of entries and the positional uncertainty of the catalog; the rate is calculated by offsetting the input catalog positions by $10'$ in each of the cardinal directions and determining the mean number of matches.

The Hard Source Catalog: A 2σ catalog version compiled from images containing only counts above 1.0 keV; sources redundant with the main catalog are excluded.

The Soft Source Catalog: A 2σ catalog version compiled from images edited to exclude solar-scattered flux and including pulse-height bins $1-4$ ($0.1-0.5$ keV) only; sources redundant with the main 2σ catalog are excluded.

The Extended Source Catalog: Derived from analysis of the whole database using default editing, this catalog includes all objects whch exceed a local background by $\geq 3.0\sigma$ using apertures of $2\overset{'}{.}35$, $4\overset{'}{.}1$, and $6\overset{'}{.}3$.

The Burst Catalog: Results from a complete search of the database in sliding 10 sec time bins in each independent source detection cell for short-timescale transient phenomena (Helfand *et al.* 1991b).

The Diffuse Sky Map: A map derived from data without solar contamination and from which all point sources were excised and all fields containing discrete diffuse X-ray sources were eliminated. The result is an estimate of the diffuse flux in each square degree region observed by the IPC. A first approximation to this map using a protion of the dataset has recently been published by Micela *et al.* (1991) and reveals several interesting features related to the diffuse Galactic emission.

Other Catalogs: From other research programs with which we are involved, we have produced other catalogs and data sets which are now on-line as well. Examples include single-dish radio surveys of the sky at 6 cm and 20 cm (containing over 75,000 sources), deep VLA imaging surveys of the Galactic Plane at 6 cm, 20 cm, 90 cm, etc.

7. Providing Access to Op-Ed

We had originally hoped to transfer our database and analysis system to the HEASARC or some other generally accessible archive so that it could be made available to the general user community. Unfortunately, to date, technical and manpower limitations have prevented this from happening. Nonetheless, over the past eighteen months nearly a dozen researchers from various institutions have requested access to the database and, despite having no funding for this effort, we have tried to accommodate them as best we can. Examples of these projects include:

- Xavier Barcons and F. Carrera (Catabria, Spain): For a detailed study of the fluctuations and angular correlations of the diffuse background, specially prepared flat fields were required.

- Richard Mushotzky (GSFC): For another diffuse background project, flat fields were requested and supplied.

- Gillian Knapp (Princeton): To obtain fluxes and upper limits for a large sample of late-type stars, the entire database was searched with the researcher visiting Columbia for the processing.

- Shri Kulkarni (Caltech): For a project to search for both point-like and extended emission around the more than 100 radio pulsars observed as targets and serendipitously, the researcher first visited Columbia to extract the data and become familiar with the system, and then had his student work on the analysis remotely from Caltech.

- France Cordova (Penn State): The demands of two projects, one involving timing of sources in which non-standard (e.g., masked out and no-aspect) data were required, and a second in which images of diffuse sources in several different energy bands without solar contamination were needed, could be readily met by our system but would otherwise have required special lengthy reprocessing requests run at the SAO data center.

- Andrea Cimatti (Firenze): Requested us to find fluxes and/or upper limits for a sample of several hundred high redshift radio galaxies.

- Bruce Margon and Scott Anderson (U. Washington): Undertook several projects involving optically selected quasar samples from which they obtained upper limits to the QSO contribution to the diffuse X-ray background. In fact these workers provided a significant demonstration of network access to the database by copying *THE ENTIRE ARCHIVE* over the network in the space of a week or two and have now installed the system at Washington.

In addition, other faculty and students at Columbia have found the system invaluable for problems ranging from a synoptic study of cataclysmic variable periodicities, to a comparison of deep HI images of galaxy clusters with the cluster X-ray gas, to a search for short-timescale X-ray transients.

8. Summary

As the above demonstrates, we have achieved our goal of creating "A Complete Database for the *Einstein* Imaging Proportional Counter." This document serves as the Final Technical Report on this effort funded under NASA grant NAG8-183.

9. References

Elvis, M., Plummer, D., Schachter, J., & Fabbiano, G. 1992, *Ap.J. (Supplements)*, in press.

Giacconi, R., *et al.* 1979, *Ap.J.*, **234**, L1.

Gorenstein, P., Harnden, F.R., & Fabricant, D.G. 1981, *IEEE Trans. Nucl. Sci.*, **NS-28**, 869.

Hamilton, T.T., Helfand, D.J., & Wu, X.-Y 1991, *Ap.J.*, **379**, 576.

Helfand, D.J., Becker, R.H., & White, R.L. 1991a, in preparation.

Helfand, D.J., Gotthelf, E.V., & Hamilton, T.T. 1991b, *GRO Conference Proceedings*, Annapolis, Oct. 1991.

Koyama, K., Makishima, K., Tanaka, Y., & Tsunemi, H. 1986, *PASJ*, **38**, 121.

McCammon, D., Burrows, D.N., Sanders, W.T., & Kraushaar, W.L. 1983, *Ap.J.*, **269**, 107.

Micela, G., Harden, F.R., Rosner, R., Sciortino, S., & Vaiana, G.S. 1991, *Ap.J.*, **380**, 495.

Wang, Q., Hamilton, T.T., Helfand, D.J., & Wu, X.-Y. 1991, *Ap.J.*, **374**, 475.

Wang, Q., & Helfand, D.J. 1991, *Ap.J.*, **370**, 541.