Final Report
for the Period 3/15/2006 − 3/14/2010

Project Title: HYDRA: A New Paradigm for Astrophysical Modeling, Simulation, and Analysis

AISRP Grant Number: NNG06GE58G

PI: John C. Houck

Co-Is: John E. Davis, Dan Dewey, Michael S. Noble, Michael A. Nowak, Michael Wise

Web page: `http://space.mit.edu/hydra`

1 Introduction

This is the final report for the 3-year AISRP project HYDRA, covering the award period from 15 March 2006 through 15 March 2009 plus a one-year no-cost extention ending 14 March 2010.

2 Progress

2.1 Source Models

One goal of HYDRA source modeling was to support constructing source models that have realistic spatial structure and to use these models directly in the analysis of observational data.

Two sets of routines, `Source-3D` and `Event-2D`, were developed to do forward folding and comparison of 3D source models with 2D event-based data sets. These routines build on the volumetric 3D (`v3d`) routines developed during our first year of funding. One can define source models that include 3D spatial structure, including Doppler shifts in the observed spectra associated with internal bulk motions within the source. The two axes of the data/model can be chosen from the available event properties. Commonly useful 2D spaces include: X-Y sky images, wavelength–cross-dispersion grating images, or radius-energy "images". These 3D modeling capabilities are currently being used in the analysis of recent Chandra observations of SN1987A (Dewey et al., 2008), and other sources (Yu et al., 2009; Zu Hone et al., 2009). Documentation for these routines and a number of detailed example applications are available on the HYDRA web pages[1]. See also Dewey & Noble (2009).

2.2 Missions and Instruments

We aimed to couple the advanced source models of the previous section with simulations of a given observatory or instrument. A key ingredient in our thinking was the concept of simulating with appropriate fidelity, allowing a flexibility in the tradeoff of speed and accuracy. Especially in initial stages of 3D source modeling the model-data (dis)agreement is often very visible and not a subtle effect. For this kind of modeling, we found that relatively crude approximations to the instrument response (*e.g.* simple PSFs, ARFs, and RMFs) are adequate to make modeling and science progress.

---

[1]`http://space.mit.edu/hydra`

2.3 Visualization

Visualization components were developed to to support 3D modeling and data analysis applications. Greatly enhanced versions of the `SLgtk` module and the VWhere, Volview and Imdisplay visualization tools were developed and released. A S-Lang interface for the `grace` plotting package was also developed.

2.4 Fitting and Statistics

The fitting process involves adjusting model parameters and constructs to optimize agreement between model and data. A goal for HYDRA was to allow users nearly complete control over the analysis process and to let the user choose which characteristics of the data will drive the fitting engine. We first defined a scriptable software interface that supports general user-defined optimization methods. This capability is further enhanced by a new *fork-select* interface that greatly simplifies parallelization on multi-core computers.

We have explored several approaches to parallelization of model fitting and derivation of parameter confidence limits. One approach to optimization invokes parallel computation of the objective function – many slave processors contributing to a single objective function evaluation – while the actual parameter search is performed serially. Another approach to optimization invokes a parallelized search of the parameter space, with many slave processes simultaneously performing different objective function evaluations. Both are useful in practice.

The important new feature in our parallelization of the objective function computation is that this parallelization can be applied to legacy code and requires no changes in that code's implementation. This parallelization approach is implemented in a PVM-based module called *pmodel*. Essentially, *pmodel* provides a generic mechanism for transparently parallelizing serial model components. For details, see the *pmodel* web page[2].

To explore optimization methods that perform parallelized parameter space searches, we developed three different parallel optimization methods, one derived on the Levenberg-Marquardt algorithm, one derived from the simplex algorithm and a parallel version of the Differential Evolution method. The speedup gained from the parallelization is somewhat problem dependent, but for several problems of interest, the optimization time is often reduced by factors of three or more.

As yet another approach to improving parallelism in data analysis applications, we also explored the use of OpenMP-based parallelism in S-Lang scripting (Noble, 2008).

2.5 Implementation and Infrastructure

Because an existing package, ISIS, provided a robust and well-tested analysis infrastructure, we chose it as the basis for the development of HYDRA capabilities. However, because it was developed for X-ray spectroscopy, important parts of the code base are somewhat application specific. With HYDRA funding we were able to generalize important parts of this system. In part, this involved separating application specific components into modules that can be accessed through more generic interfaces. A number of generic components developed for HYDRA have simply been merged into ISIS proper, while more specialized capabilities of HYDRA are provided by the collection of modules mentioned here.

We used grant funds to purchase a 13-node Beowulf cluster (26 dual-core CPUs). Before making this equipment purchase, we did a considerable amount of research to educate ourselves about computer clusters available on the market and various options for cluster management software.

---

[2]`http://space.mit.edu/cxc/pmodel/`

This research included extensive discussions with cluster engineers from different vendors. Because the power and cooling infrastructure in our office building was inadequate for hosting a computer cluster, we arranged to have the cluster installed in MIT's Co-Location facility. Since that time, the power and cooling infrastructure of our building's computer room have been upgraded and we have moved the cluster to a much more convenient location within our building. We used free software to configure the cluster ourselves and, we continue using it to explore various methods for applying distributed computation to data analysis.

The cluster has become a focal point for nurturing multiple collaborations in high-performance astrophysical computing. One result of these efforts was a parallelizing wrapper for XSTAR, which made it feasible for us to probe thousands of photoionized gas physical scenarios in the time it has previously taken to compute only a handful of such models (Ji et al., 2009). We have also developed scripts to simplify the parallel computation of statistical confidence limits, and applied them to a number of models such as jet phenomena in active galactic nuclei (Maitra et al., 2009).

2.6   Hydra Products and Applications

The HYDRA software was designed to highly modular and we have chosen to distribute the more specialized component in modular form. Generic components associated with parallel optimization (e.g. *pmodel*) and enhanced scripting capabilities (e.g. *fork-select*) have been incorporated into the ISIS package. Separate modules are available to provide volume visualization (Volview), n-tuple filtering (VWhere), an interface to the HDF5 file format and library, the `v3d` geometric modeling routines, and the `Source-3D` and `Event-2D` routines for the analysis of event-based data sets. Some example applications are described in the papers and presentations listed on our site.

A number of new scientific modules were created using the SLIRP automatic code generator. These provide access, directly in HYDRA memory at runtime, to internal functions and data structures (such as common blocks) from the XSTAR and XSPEC astrophysical modeling tools. Through these modules collaborators may reuse important legacy functionality in ways that are fundamentally distinct from their original design, and in more rapid, scriptable fashion (Ji et al., 2009).

In the course of the project, we came to recognize a need to raise awareness in the X-ray community of the limitations inherent in traditional software approaches to spectral analysis, and how they can be mitigated in modular systems like HYDRA. Towards that end we undertook a comprehensive quantitative analysis of the defacto standard spectral modeling tool, XSPEC; our findings were summarized in (Noble & Nowak, 2008).

References

Dewey, D., & Noble, M. S., 2009, in Astronomical Society of the Pacific Conference Series, ed. D. A. Bohlender, D. Durand, & P. Dowler, Vol. 411, 234

Dewey, D., Zhekov, S. A., McCray, R., & Canizares, C. R., 2008, ApJ Lett., 676, L131

Ji, L., Noble, M., Houck, J., Schulz, N. S., Nowak, M., & Marshall, H. L., 2009, in Chandra's First Decade of Discovery, Proceedings of the conference held 22-25 September, 2009 in Boston, MA. Edited by Scott Wolk, Antonella Fruscione, and Douglas Swartz, abstract #123, ed. S. Wolk, A. Fruscione, & D. Swartz

Maitra, D., Markoff, S., Brocksopp, C., Noble, M., Nowak, M., & Wilms, J., 2009, MNRAS, 398, 1638

Noble, M. S., 2008, Concurrency and Computation: Practice and Experience, 20, 1877

Noble, M. S., & Nowak, M. A., 2008, PASP, 120, 821

Yu, Y. S., Kastner, J. H., Dewey, D., Nordon, R., Behar, E., Behar, E., & Soker, N., 2009, in Bulletin of the American Astronomical Society, Vol. 41, 461

Zu Hone, J. A., Ricker, P. M., Lamb, D. Q., & Karen Yang, H., 2009, ApJ, 699, 1004