THE GONG++ DATA PROCESSING PIPELINE

Frank Hill, John Bolding, Clifford Toner, Thierry Corbard, Steve Wampler,
Bret Goodrich, Jean Goodrich, Patricia Eliason, and Kerri Donaldson Hanna

National Solar Observatory,
950 N. Cherry Ave., Tucson, Arizona, USA, 85719
Tel: +1-520-318-8318 Fax: +1-520-318-8278
E-mail: fhill@noao.edu

ABSTRACT

We describe the hardware and software for the new GONG++ data processing system, and discuss our current experience in developing a pipeline for local helioseismology.

INTRODUCTION

The recent upgrade of the GONG detectors from 256 × 256 to 1024 × 1024 pixel cameras (known as GONG+, Harvey, Tucker & Britanik 1998) can only reach its full scientific potential with a corresponding major upgrade to the data processing system. This system, known as GONG++, is currently under development. It will produce new data products, including subsurface synoptic flow maps over continuous rotations, subsurface active region flow maps, farside images for space weather applications, tracked and remapped data cubes for general local helioseismology, ring diagrams, time-distance diagrams, holographic images, high-degree mode parameter estimates, merged velocity images, merged magnetic field images, and magnetic synoptic charts. In order to deal with the greatly increased volume and complexity of the data processing, as compared to the earlier GONG Classic processing task, we have adopted an automated pipeline approach. This approach was employed by the Stanford MDI/SOI Project and has been shown to be extremely effective in dealing with the varied demands and volume of helioseismic data processing and scientific analysis (Kosovichev *et al.* 1997). Here we describe the GONG++ hardware and present our current experience with the integration of a "mini-pipeline", which produces subsurface synoptic flow maps using ring diagram analysis.

HARDWARE

The GONG++ hardware system had to meet a minimum set of requirements: it had to be reliable, provide high-performance computing capability and substantial data storage resources, and allow the use of existing code without major revision. The selected configuration is comprised of a symmetric multi-processor SunFire 4800 with twelve 900-MHz CPUs and 24 GB RAM. The data storage requirement has been met with six T3 disk arrays holding a total capacity of 4.7 TB, and a StorEdge L180 Tape Library, with a capacity of 18 TB on 180 LTO (Linear Tape Open) cartridges and four LTO drives. Figure 1 shows the system, which is now installed at Tucson.

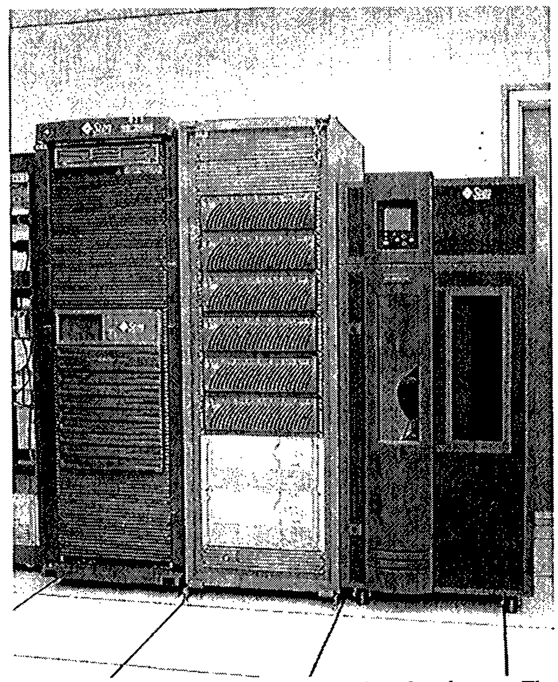

Figure 1: The GONG++ processing hardware. The rack on the left contains the SunFire 4800 with twelve 900-MHz CPUs and 24 GB of memory; the middle rack holds the six T3 disk arrays and the UPS systems, and the box on the right is the L180 StorEdge tape library.

The hardware can adequately fulfill the performance requirement of keeping cadence with the data collection. The 22.7-TB storage capacity is sufficient for about 650 days of anticipated GONG++ data products, which are expected to accumulate at a rate of about 35 GB per day. GONG's considerable experience with Sun equipment and the Solaris operating system, should expedite the task of porting usable code to the new system. The processing pipeline system and the tape library archive control system needed to be created. In order to accelerate development, commercial products will be used for these components.

Proc. of SOHO 12/GONG+ 2002 'Local and Global Helioseismology: The Present and Future', Big Bear Lake, California (U.S.A.), 27 October – 1 November 2002 (ESA SP-517, February 2003)

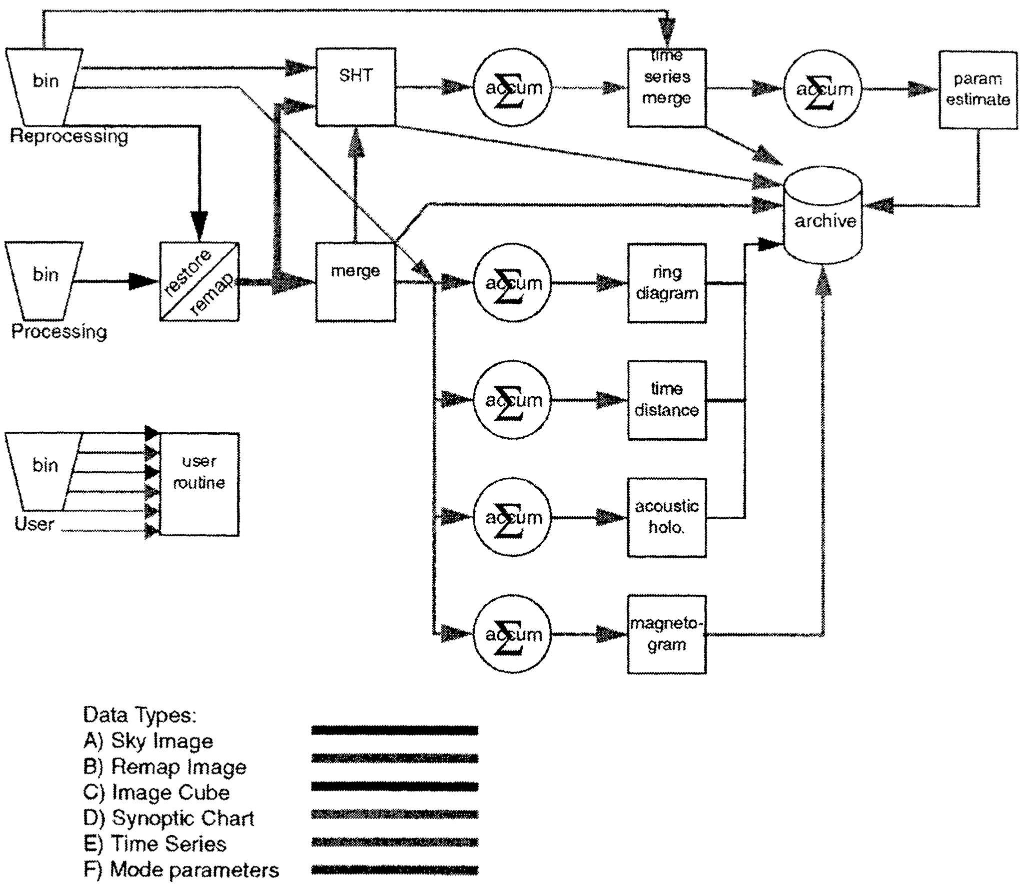

Figure 2: A flow chart/block diagram of the GONG++ data processing pipeline. The square blocks represent analysis modules. The arrows represent the flow of the data, which will be controlled by the OPUS pipeline package and the VERITAS™ hierarchical storage management system. The circles represent the accumulator stages that monitor the temporal coverage of the data products and initiate processing steps as needed. The trapezoids represent the input to the pipeline, and the cylinder represents the on-line archive.

SOFTWARE

The GONG++ system is essentially comprised of analysis modules, accumulators, a pipeline structure, and an archive. The processing data flow, as shown in Figure 2, begins with the ingestion of the data from the upstream calibration process. The data proceeds through an optional restoration stage and is then remapped to a common registered sky image. The registered images either enter a global analysis pipeline via a spherical harmonic transform (SHT) or are merged for use by local methods. As the data flows through the pipeline, the accumulator modules monitor the progress and control the processing schedule by periodically checking for sufficient time coverage before initiating the next stage. The accumulator modules are an integral part of the local helioseismology pipeline – it would be formidable task for an operator to successfully keep track of the large number of areas in a dense-pack ring diagram synoptic flow map.

Each local helioseismology module in Figure 2 actually contains several sub-modules. For example, the ring diagram box produces an image cube (i.e. a dense-pack set, Haber *et al.* 2002), computes the 3-dimensional power spectrum, fits the rings, and performs the inversion. The time-distance (Zhao, Kosovichev & Duvall 2002) and acoustic holography (Braun & Lindsey, 2001) modules will also contain several steps, as will the production of magnetograms. All valuable and computationally expensive products will be archived, among them, merged images, data cubes, fitted parameters, and inversion output.

The pipeline also provides an entry point to reprocess data starting from intermediate data products that can be extracted from the archive. Finally, users will have access to the archive and the ability to run their own processing, albeit at a lower priority than the production process.

Two essential pieces of this data handling software are the pipeline control system and the tape library and archive management system. Both of these components are being constructed using commercially available software, whereas the analysis modules and accumulators are being written in house. The use of commercial products for the core systems greatly accelerates the pipeline development timeline and reduces the overall cost. We have selected OPUS for the pipeline control system, and VERITAS™ as the tape library and file management system. The web sites for these packages are listed in the references section of this paper.

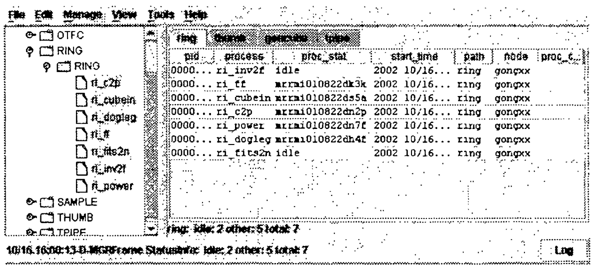

Figure 3: The OPUS Process Manager screen. This shows an operational ring-fitting pipeline composed of seven steps while it is processing the 189 areas of a dense pack. In this snapshot, two of the processes are idle, while the remaining five are working on five different files in the pack.

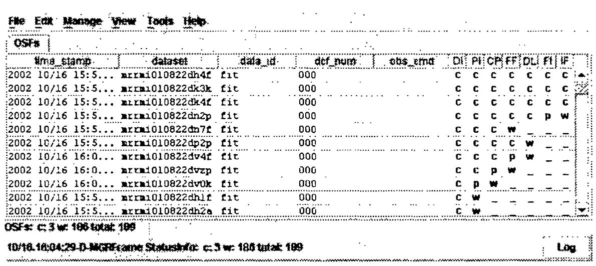

Figure 4: The OPUS Observation Manager screen. This shows a portion of the list of the 189 files/regions in the dense pack. On the right are seven columns, one for each of the seven steps in the pipeline, showing the status of the file as it moves through the steps. In these columns, the letter c indicates completion, p indicates processing, and w shows a file waiting to begin the step. Here, 3 of the 189 files have completed the entire processing chain.

OPUS was originally developed at the Space Telescope Science Institute to process HST telemetry data, but is actually a powerful, general-purpose tool that provides a fully distributed pipeline processing system for any series of applications. The OPUS environment allows multiple instances of multiple processes to run in multiple pipelines on multiple nodes. It is particularly well suited for GONG++ local helioseismology reduction, which typically utilizes hundreds of small areas on the disk. Figures 3 and 4 show a snippet of the OPUS interfaces – the Process Manager and the Observation Manager.

VERITAS™ is a commercial software package that is used primarily to operate tape library systems and provide automated backup services. It can also be used as a hierarchical storage management (HSM) system wherein the tape library behaves as a very large near-line disk partition. VERITAS™ can be called from a command line, making its integration with OPUS easily accomplished, and the distribution of files across the tape library is completely controllable, which is essential for helioseismic processing.

MINI-PIPELINE EXPERIENCE

We found that the construction and implementation of a mini-pipeline to produce flow maps from tracked and remapped data cubes was extremely easy. An example of one of the flow maps produced by the pipeline is shown in Figure 5. Because OPUS only requires that the code be executed via a Unix shell script, it was simple to link together analysis modules written in IRAF, Fortran 90, Fortran 77, and C. The pipeline makes full use of all available processors that are assigned to the task without the need to translate the code into distributed parallel processes. We have recently installed the process to merge images, which completes the core steps in the ring-diagram pipeline. The next major task will be the integration of OPUS and the VERITAS™ HSM system.

ACKNOWLEDGEMENTS

This work utilizes data obtained by the Global Oscillation Network Group (GONG) project, managed by the National Solar Observatory, which is operated by AURA, Inc. under a cooperative agreement with the National Science Foundation. The data were acquired by instruments operated by the Big Bear Solar Observatory, High Altitude Observatory, Learmonth Solar Observatory, Udaipur Solar Observatory, Instituto de Astrofísico de Canarias, and Cerro Tololo Interamerican Observatory.

REFERENCES

Braun, D.C., and Lindsey, C., 2001, *Astrophys. J.* **560**, L189.

Haber, D. A., Hindman, B. W., Toomre, J., Bogart, R. S., Larsen, R. M., and Hill, F., 2002, *Astrophys. J.* **570**, 855

Harvey, J., Tucker, R., and Britanik, L. 1998, *Structure and Dynamics of the Interior of the Sun and Sun-like Stars* (Eds. S. G. Korzennik & A. Wilson), ESA SP-

418, ESA Publications Division, Noordwijk, The Netherlands, 209.

Kosovichev, A. G., and 33 co-authors, 1997, *Sol. Phys.* **170**, 43.

OPUS web site: www.stsci.edu/resources/software_hardware/opus

VERITAS™ web site: http://www.veritas.com/

Zhao, J., Kosovichev, A. G., and Duvall, T. L. Jr., 2002, *Astrophys. J.* **557**, 384.

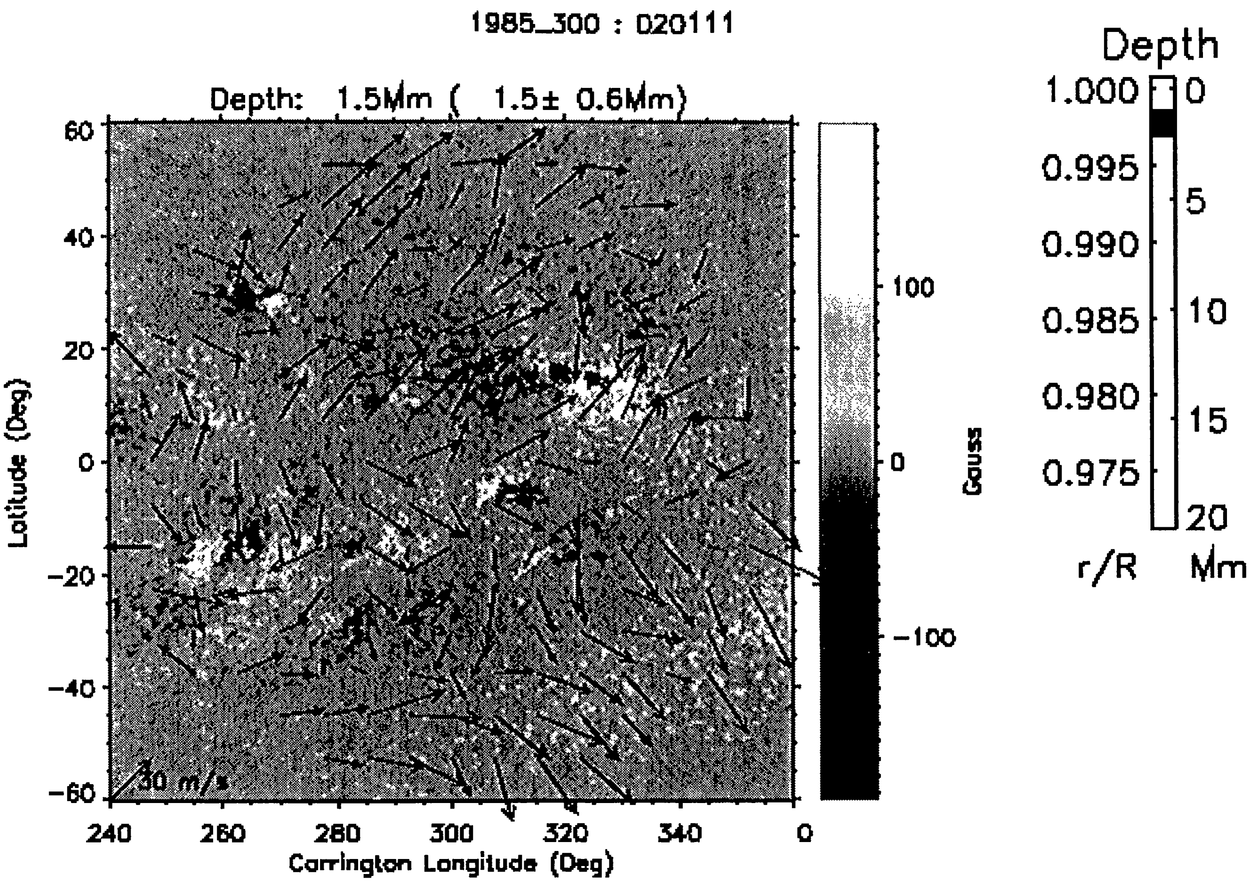

Figure 5: An example dense-pack flow map from a single day of merged GONG+ images, produced through the GONG++ ring - fitting pipeline. The background image is the simultaneous solar magnetic field, obtained from a GONG+ magnetogram.