GBT Software Development Areas
from the GBT Future Instrumentation Workshop
held September 7-9, 2006

This is a summary and exploration of the major software ideas and issues raised during the GBT Future Instrumentation Workshop on September 7-9, 2006. Four major topics are addressed: (1) cluster computing and processing parallelization, (2) data storage, access, and transport, (3) data pipelining, and (4) FPGAs. It is worth noting that these topics are not discussed in any particular order. However, the first three topics are becoming an increasing concern across all fields of astronomy, not just for the GBT. Data production rates and data volumes are dramatically increasing and our current data handling paradigms will be increasingly ineffective when buffeted by the deluge of data from future instrumentation.

Cluster Computing & Processing Parallelization:

There was some interest expressed in using clusters in both instrument development and in data reduction for future instrumentation. In fact, there are already such clusters in use at the Green Bank Telescope specifically as well as the other facilities of the NRAO. For example, the GASP pulsar backend used at the GBT is a cluster of computers. Scott Ransom has a cluster in Charlottesville dedicated to pulsar data reduction. The ALMA correlator utilizes a Beowulf computer cluster. Also, Jon Romney is heading a group currently exploring cluster computing options for the VLBA.

Indeed, many astronomical institutions are already employing high performance parallel computing clusters for data reduction and analysis as well as instrumentation. One example is the Astronomical Wide-field Imaging System for Europe (Astro-WISE). Astro-WISE is employing such a system which depends on a distributed database and federated data server to efficiently process enormous amounts of data. If one were to do a survey of computing trends in astronomy today, one would find anything from small Beowulf clusters to large massively parallel platforms to collections of MPPs distributed across grids, such as the NSF TeraGrid facility. Additionally, there is a growing focus on MPPs and grids. This is especially apparent in the large development projects such as the Large Synoptic Survey Telescope (LSST).

Cluster computing and parallel processing bring a different mindset to software design and development. One must build parallelization into the software in order to make full use of a cluster of computers. There are frameworks available to aid software engineers. MPI is a library specification for message-passing. It is a proposed standard put forth by a broadly-based committee of vendors, implementers, and users. MPI was designed for high performance on both massively parallel machines and on workstation clusters. A number of MPI implementations exist, some of which are freely available.

We view cluster computing and processing parallelization as an important trend in astronomical software development – both in instrumentation development and in data analysis and reduction. Our primary concerns at the GBT in this area center are twofold,

(1) cultivating parallel software processing expertise and (2) ensuring that applications are developed with an eye toward parallelization.

Data Storage, Access & Transport:

As with cluster computing and process parallelization, the topics of data storage, transfer and access are important to all astronomical facilities. The volume of data being produced by telescope facilities is rising dramatically. At ADASS XVI, Robert Seaman et al asserted that "observatories' data storage costs will threaten their budgets. Data transport latency and bandwidth will threaten not just budgets, but available technology and human patience."[1] The problem of data storage will not be solved purely through the usage of physical media; data handling techniques are also extremely important. Seaman et al offered up data compression as a key technique required to produce both acceptable data throughput and affordable data handling. Many astronomical facilities are actively investigating additional data handling techniques as well as pushing the state of the art in data storage.

The GBT is certainly no different in this regard from other telescopes and this was discussed at some length during the future GBT instrumentation workshop. Currently, the raw data acquired by the GBT is stored to hard disk and archived to hard disk. Observers transport their data from the GBT to their home institutions via network protocols, tapes, or portable hard drives. As the data volume grows, we will need to employ alternative approaches to data handling, storage, and transport. Therefore, GBT staff needs to remain conversant in various data handling techniques (e.g. data compression) and remain competitive by utilizing the state of the art in physical storage.

Data Pipelining:

Several times during the workshop, the topic of data reduction pipelines arose. One motivation for the discussion of data pipelining was that belief that future instrumentation will most certainly generate larger volumes of data that we are presently handling at the Green Bank Telescope. Several participants pointed out that when handling large data volumes, an observer can easily get overwhelmed and lose the high-level interpretation of the data in the details. Currently at the Green Bank Telescope, we do not provide an automatic data reduction pipeline. Automating such a data pipeline for GBT data is tricky because of the telescope's high degree of configurability and application to a wide variety of problem domains. However, we do possess the basic building blocks from which data reduction pipelines could be assembled. Given a good specification of a pipeline decision tree and scientific guidance in determining at which pipeline segments the observer will want to make decisions, an automated pipeline is possible – at least for standard observing modes and configurations.

If we were to construct and make automated data pipelines available, it is our belief that we would need to provide a framework which is flexible and accommodates both the novice observer as well as the experienced observer. We see three basic requirements for such a flexible system: (1) the observer needs to be able to interrupt the pipeline at any

[1] Seaman, Robert, Pence, William, Dickinson, Mark, Valdes, Frank, White, Rick, & Zarate, Nelson. Astronomical Tiled Image Compression: How & Why. ADASS XVI. Tucson, Arizona. October 2006.

point, (2) the observer needs to be able to swap components of the pipeline easily and quickly, and (3) the observer needs to be able to insert his/her own custom components.

FPGA Development:

FPGAs were a hot topic at the workshop and we feel that they will feature prominently in future Green Bank Telescope instrumentation endeavors. From a software perspective, we expect that proper training and the employment of good software engineering practices will be sufficient preparation where FPGA development is concerned. As with cluster computing and processing parallelization, we believe that we need to cultivate expertise in this area. A recent article in EDN[2] states that "design software is becoming a critical part of the FPGA's value." One can then extrapolate and assert that as telescope instrumentation becomes more and more complex, FPGA-design software needs to address challenges that in the past only existed in the ASIC space, such as timing closure, power closure, and system-level design." Therefore, the hardware engineer who employs FPGAs also needs to become conversant in good software engineering practices and software design in order to make best use of the technology.

Summary:

The GBT Software Development Division is committed to current projects through September 2007, and so it will be some time before we can address these issues. As effort becomes available, we will have to be pragmatic but visionary in selecting the appropriate areas in which to focus our attention. As with some of the hardware developments discussed at the workshop, many of these topics will be amenable to university collaborations. All of these considerations will be taken in to account as we continue to refine the GBT development program.

---

[2] Cabel, Denis. Trends Shaping the FPGA Industry. EDN. November, 9, 2006. http://www.edn.com/index.asp?layout=article&articleid=CA6389417&industryid=23439