Development of a Grid for Field Processing of Large Radar Data Sets on Ice Sheets

This paper describes the Cyberinfrastructure deployment designed to solve data management and computing issues for both field expeditions and lower-48 activities within the NSF Science and Technology Center for Remote Sensing of Ice Sheets (CReSIS). Polar Grid is a National Science Foundation (NSF) Major Research Instrumentation (MRI) funded partnership of CReSIS, Indiana University and [Elizabeth City State University](http://www.ecsu.edu/). The partnership goal is to acquire and deploy the computing infrastructure needed to investigate the urgent problems in glacial melting. Led by the University of Kansas, CReSIS now relies on Polar Grid hardware and expertise for post-processing data from recurring field experiments on both Greenland and Antarctica. Over the past year, Polar Grid equipment has accompanied CReSIS field teams to both Greenland and Antarctica on three separate field activities.

Polar Grid enables near real-time assessment of radar performance through daily data processing using algorithms that previously exceeded underpowered field computers or laptops. PolarGrid architecture links the dynamic components at field and base camps to the traditional computing infrastructure at ECSU, IU and KU. This architecture is designed to address the difficult technical challenges of doing research in the field: data processing, data backups, networking/bandwidth, power, physical space, and weight limitations. A newly developed field system is based on IBM BladeCenter S that delivers almost half a TeraFLOP of computing power for less than 1800W power consumption. Utilizing low-power processors, efficient multi-core parallel data processing, and efficient blade/chassis architecture, this system has a very aggressive FLOPS per watt power profile and takes up 50% less space than a traditional rack-able server comparable in computing power. This system provides very efficient data processing and online storage, and is a key for the creation of offline secondary and tertiary backups.

Polar Grid field and base camp equipment has been used during the following Polar deployments:

|  |  |
| --- | --- |
| Deployment | Polar Grid equipment |
| Greenland 2009: 3/26/2009 - 4/25/2009  Deployment at base camps in Kangerlussuaq, Illulissat, and Kulusuk, Greenland. The expedition collected 6TB of radar data | 1 IBM x3550 Server  70 Western Digital 1TB drives  1 Dell D830 laptop  1 Inspiron 9300 laptop |
| Antarctica 2008-2009 expedition was curtailed due to fuel costs, but the following equipment was sent to the Antarctica field camp, with 3.8TB of data collected. | 1 IBM x3650, dual dual-core processors 8GB of RAM, 4.5TB internal storage  20 1 TB drives  18 750GB drives  1 Dell Inspiron 9300 laptop  1 Dell Latitude D830 laptop  2 Panasonic CF-30 toughbooks |

This processing has proven responsible for identifying electronic noise sources, uncovering inaccurate hardware/radar settings, and highlighting unforeseen environmental conditions that, were Polar Grid not available, would have resulted in poor data quality across an entire field season. Specifically, two important consequences of Polar Grid field equipment have been:

1. Identified mistuned radar channel during GAMBIT deployment to Antarctica that, when corrected, resulted in immediate data improvement.

2. Field team could immediately see the impact of sub-tuned radar. As environmental conditions continue to change in ways and at speeds previously unseen, these 'on the fly' radar tunings are critical to developing good science.

Polar Grid field equipment is used for immediate, small-scale image processing, but the entire data set must be processed on more conventional clusters after it is shipped back to the United States. With assistance from IU, Kansas University processed the 2008 Ground and the 2008 Airborne MCRDS data sets. The 2008 Airborne data set were processed twice with different settings. KU also processed the 2008 Ground data in two different ways with CSARP: once with small chunks the other with large chunks. The amount of data processed was approximately 25 TB input and 15 TB output. IU's Lustre-based DC-WAN file system was mounted at KU and used to transport files between KU and IU's cluster. Computing time was an estimated 150 hours for the ground and 300 for the airborne.

Indiana University has acquired a compute resource for processing Polar Grid radar data and assembled it at the IUPUI datacenter. The system consists of 32 HP DL580 G5 nodes with six cores per node, for a total of 768 cores in the cluster, with 4 GB of memory per core, linked by an Infiniband interconnect. Benchmarks are ongoing and are not yet completed on the system to determine performance. The cluster runs both

Windows HPC and Linux cluster software.

Current challenges relate to developing TeraGrid support for Matlab-based CSARP code used for processing of data. Efforts are in progress to train researchers and students in distributed and parallel computing techniques needed to take full advantage of Tera Grid resources when processing the data and on converting the existing Matlab code to C.

Future efforts are focused on the following areas: a) continuing support for CReSIS expeditions to Greenland and Antarctica; b) acquiring a new cluster for 'Lower 48' processing; c) processing both new and historical CReSIS field data; and d) developing the Polar Grid science gateway infrastructure.

With respect to the gateway design efforts, the project will develop Polar Grid Web tools. These included the development of several filter applications in Matlab that can be used to create higher level (and more individualistic) data products from processed Polar Grid data. These filter applications are wrapped as Web services and composed into workflows using tools from the OGCE project (GFAC and XBaya, respectively). PHP client programs will be integrated with non-Java portal environments (such as Drupal) and social networks (such as Facebook).

Beyond Science and Engineering contributions, this project supports the SAR-based mapping of the underlying bedrock beneath Greenland and Antarctic glaciers, providing the level of detail needed to understand and model the behavior of these systems as they melt. The impact of glacial melting and global sea level rise on the environment is potentially devastating, so it is crucial to provide legislators and policy makers with accurate information.