PolarGrid: Cyberinfrastructure for Polar Science

Lead Institution: Indiana University, Bloomington, IN

PI: Geoffrey Fox
CGL: Community Grids Laboratory
Computer Science Department, School of Informatics
Indiana University, Bloomington, IN

Co-PI Craig Stewart
UITS University Information Technology Services Research Computing
Indiana University, Bloomington, IN

Co-PI Marlon Pierce
Community Grids Laboratory
Indiana University, Bloomington, IN

Co-PI: Linda Hayden, Malcolm LeCompte
CERSER: Center of Excellence in Remote Sensing Education and Research
Elizabeth City State University, Elizabeth City, NC

Collaborators:
Prasad Gogineni, David Braaten and Cornelis van der Veen
CReSIS: Center for Remote Sensing of Ice Sheets
University of Kansas, Lawrence, KS

Ken Jezek
Byrd Polar Research Center
The Ohio State University, Columbus, OH

Sridhar Anandakrishnan
Pennsylvania State University, University Park, PA

Project Summary

Intellectual Merit: Recent polar satellite observations show disintegration of ice shelves in West Antarctica and speed-up of several glaciers in southern Greenland. The great ice sheets in Antarctica and Greenland interact with the global climate in a complex manner, and the impact on global sea level of their retreat would be profound. Most of the existing ice-sheet models, including those used by the Intergovernmental Panel on Climate Change (IPCC), cannot explain the rapid changes being observed. The Center for Remote Sensing of Ice Sheets (CReSIS) is developing new technologies to perform 3-D characterization of ice sheets to understand the physics of rapid changes, and develop models to explain observed changes and predict future behavior. In particular, CReSIS has demonstrated that Synthetic Aperture Radar (SAR) can image the beds of ice-sheets. This will enable a new-generation of high resolution ice-sheet models with realistic boundary conditions, but it will require distributed Cyberinfrastructure to gather and process data and assimilate them with large simulations. We propose a sophisticated Cyberinfrastructure instrument that will both enable the crucial ice-sheet science and educate and train a diverse workforce in both Polar science and Cyberinfrastructure.

Instrument: This proposal addresses these compelling scientific challenges with a PolarGrid Cyberinfrastructure, aimed at research, education and outreach. PolarGrid consists of intermittently disconnected field and base grids feeding information to “lower 48” data and computing resources. True real-time processing at the field camp is backed up with increasing fidelity but increasing delay at the base and “lower 48” systems. The requested system includes an expedition grid consisting of ruggedized laptops in a field grid linked to a low power multi-core base camp cluster; a prototype and two production expedition grids feed into a 17 Teraflops "lower 48" system at Indiana University and Elizabeth City State (ECSU) split between research, education and training. This will give ECSU a top-ranked 5 Teraflop MSI High performance computing system, building on its distance education and undergraduate laboratory infrastructure to create tremendous outreach capabilities. We request upgrades to student laboratories, video-conferencing and key scientific software to support the ambitious education and training goals. PolarGrid will be integrated with TeraGrid for both resource utilization and curricula sharing. We will follow modern open data access standards so that raw, processed and simulated data can be archived outside PolarGrid by and for the full science community.

Management: We have formed the Cyberinfrastructure Center for Polar Science (CICPS) with experts in Polar Science, Remote Sensing and Cyberinfrastructure. This center includes the lead institutions for this proposal: Indiana University, which is internationally known for its broad expertise in research and infrastructure for eScience; and ECSU, a founding member of CReSIS with a center of excellence in remote sensing. CICPS includes CReSIS institutions as collaborators and will drive PolarGrid to meet their goals while using the best known technologies. Impressive institutional commitments include a new building and faculty lines at ECSU and resources and system support from Indiana University. We also commit the substantial initial effort needed to build the portal, workflow and Grid (Web) services that are required to make PolarGrid real.

Broader Impact: We exploit strong existing outreach activities at CReSIS and Indiana University to involve Minority Serving Institutions (MSI) directly in our work. A strong education and training activity led by ECSU (a Historically Black University) will involve undergraduates through new curricula and research experiences at all CICPS participants. Students trained and educated on PolarGrid will participate in internships and enhance the entry of a diverse workforce into exciting important science. Existing collaborations with the Association of Computer/Information Sciences and Engineering Departments at Minority Institutions (ADMI) and the NSF-funded MSI Cyberinfrastructure Empowerment Coalition (MSI-CIEC) will ensure our outreach has broad impact. The innovative architecture of PolarGrid with intermittently disconnected components has applications to other power- and bandwidth-challenged applications.

1. Introduction

Building a Polar Grid: This proposal will deploy the Cyberinfrastructure (abbreviated CI), called PolarGrid, needed to support polar science, and in particular the field research program of CReSIS (Center for Remote Sensing of Ice Sheets). The major components, including field, base and core systems, are depicted in **Figure 1**. In past years, most of the data collected in the polar regions were not processed in real time. Instead, data were shipped to computing facilities in the continental US and analyzed well after collection. Real-time processing and data analysis are urgently needed, both in the field and at supporting computing centers, to adjust collection strategies in response to new information contained in the data. The polar community must have access to a state-of-the-art computing facility to process the large volumes of data to be collected by the polar community in 2007-2008 in particular, as a part of International Polar Year (IPY) activities, and to support large-scale ice-sheet models.

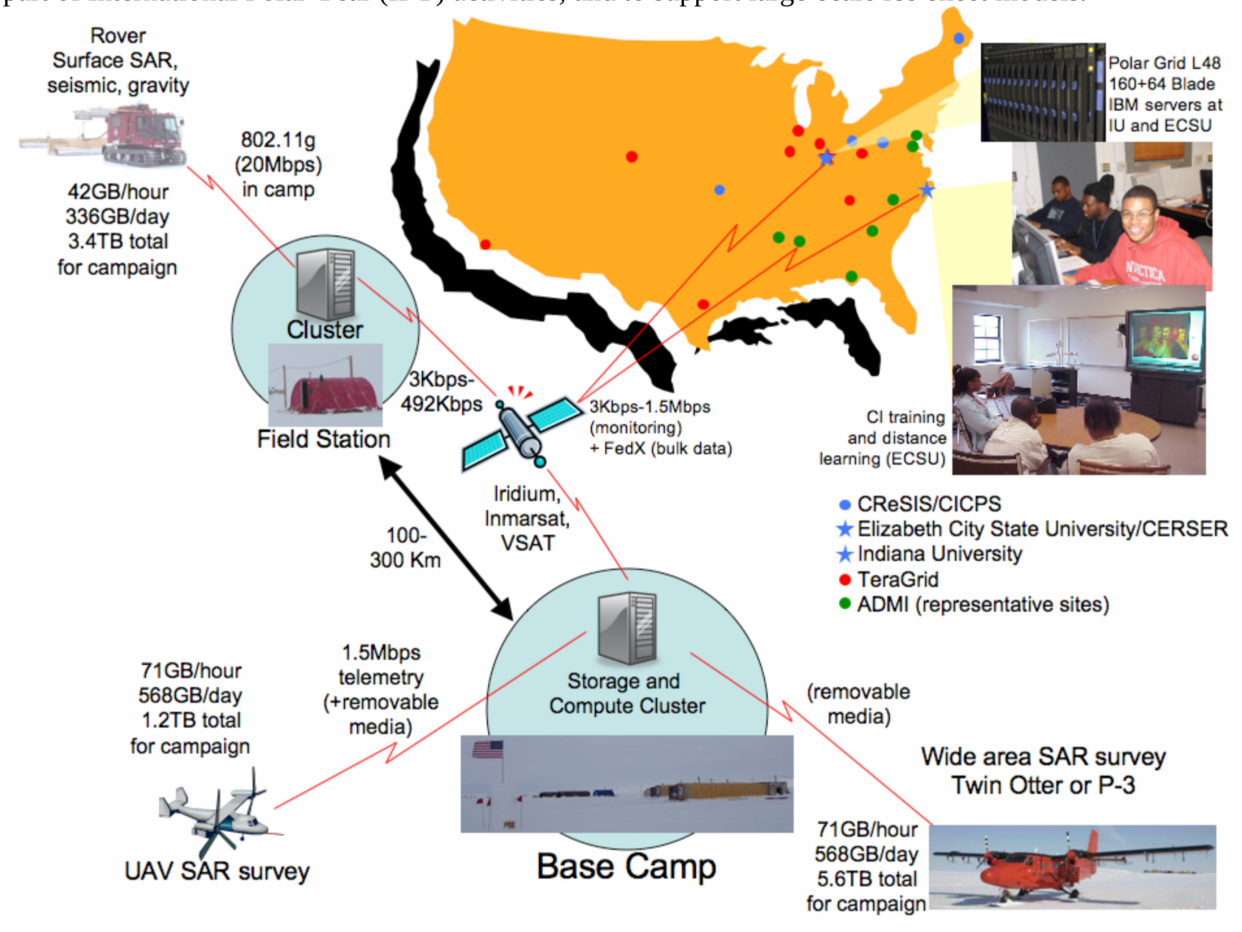


Figure 1 illustrates PolarGrid's major components and concept of operations.

The PolarGrid we propose will transform the capabilities of US polar researchers by enabling them to process and evaluate data as they are collected. This capability is achieved through a combination of laptop computers carried by researchers as they collect data in the field, a 64-core computer cluster to be added to the polar research base camp, and a larger cluster located in the continental US for final data processing and modeling work. The 64-core base camp computer cluster is a system sufficiently powerful that just half a decade ago it might have been considered a university's major supercomputer. This system must be deployed in the field: much of the data will be collected with Synthetic Aperture Radars (SAR) and will require considerable post-processing, and low network bandwidth between the polar regions and larger supercomputer facilities will preclude real-time supercomputing at centers such as the TeraGrid.

However, there will be considerable post-collection analysis and scientific investigation on the collected data that will require the resources of the TeraGrid. Video conferencing hardware to create a small educational-grid is also included as part of this proposal to aid the dissemination of new research findings to students, particularly students from traditionally underserved groups, at our partner institutions.

Urgency of Better Polar Science: Accurate prediction of the future evolution of polar ice sheets is a matter of global importance. The ice sheets in the Antarctic and Arctic interact with the Earth's climate in a complex manner, and the impact on global sea level of their retreat will be profound. Current models suggest that ice sheets respond to climate change on time scales ranging from centuries to millennia, creating the belief that there is yet considerable time to develop strategies for responding to the effects of the melting of the polar ice sheets.

This belief has been undermined by recent data. Satellite observations have shown that rapid changes in the polar ice sheets are occurring that cannot be explained by existing models. Recent rapid changes in the polar ice sheets include the disintegration of ice shelves in West Antarctica, accelerated melting of several glaciers in southern Greenland [1][2], and most recently the breaking free of the Ayles Ice Shelf from its attachment to land in northern Canada [3]. Most of the existing ice-sheet models, including those used by the Intergovernmental Panel on Climate Change (IPCC) [4], cannot explain the rapid changes being observed. The measured contribution to sea level from a few areas of the Greenland and Antarctic ice sheets is already exceeding the contribution estimated by the IPCC. This trend is especially alarming because the impact of even very modest sea level rise on humanity is substantial. A 1.5 meter increase in sea level would affect about 17 million people in Bangladesh alone [5]. Thus, models that incorporate mechanisms for rapid changes are needed to explain and predict the role of the great ice sheets in sea level rise.

CReSIS (Center for Remote Sensing of Ice Sheets) [6] is developing technologies and techniques to perform 3-D characterization of the ice sheets, and will collect data on a large scale and in great detail about areas undergoing rapid changes in both Greenland and the Antarctic during 2008 and 2009. This coincides with the International Polar Year (IPY) [7], with CReSIS as one of the key participants in IPY activities. CReSIS is led by the University of Kansas, with partner institutions including Elizabeth City State University (ECSU), Haskell Indian Nations University (Haskell), The Ohio State University, Pennsylvania State University, and the University of Maine.

CReSIS aims to involve a large part of the glaciological community to effectively design field campaigns and collect the data required for a qualitatively improved understanding of polar ice sheets. The quantity of data collected, the computations required to convert raw data into ice sheet images and measurements, and the modeling and analysis required to understand these data, create unprecedented challenges for the information technology and computer science involved in polar studies.

We propose to create a Cyberinfrastructure [8] that will enable CReSIS to manage the polar ice sheet data to be collected, analyze them, and make better predictions of future ice sheet behavior than are possible with existing models and facilities. By Cyberinfrastructure we mean advanced instruments, computing systems, data storage facilities, visualization environments, software, and people connected by advanced networks to enable scientific discoveries that would otherwise not be possible. Dr. Arden Bement, director of the National Science Foundation, has called Cyberinfrastructure the engine for change that will drive "a second revolution in information technology, one that may well usher in a new technological age that will dwarf, in sheer transformational scope and power, anything we have yet experienced in the current information age" [9].

This project will support the scientific activities in CReSIS and also develop tools that will be useful to the broader polar scientific community. The proposed PolarGrid will include satellite remote sensing (altimetry and interferometric synthetic aperture radar InSAR), measurement of a 3-D characterization of areas undergoing changes (ice-penetrating radar, synthetic aperture radar to image the ice-bed interface, seismic probes), reconstruction of climate records and glaciological modeling. In the next section we describe the PolarGrid scientific drivers. After that, we describe the Cyberinfrastructure we propose to develop and implement, followed by discussion of educational activities, broader impacts, and management of the project.

2. PolarGrid Research

2.1 Science Research Challenges for PolarGrid

The traditional view regarding the relationship of polar ice sheets and climate has been that polar ice sheets respond slowly to changes in climate. Most existing models, including those used by the Intergovernmental Panel on Climate Change (IPCC), predict only a small contribution from Greenland and Antarctica to global sea level over the next century [10]. However, evidence for active ice sheets, both present-day and in the past, has mounted, and we now understand that large ice sheets can undergo rapid changes [11]. For example, analysis of deep-sea sediment cores has revealed major oscillations in the Laurentide Ice Sheet volume, believed to be associated with surges of the Hudson Bay lobe [12]. During meltwater pulse 1A, *ca*. 14,000 years B.P., sea levels rose about 20 m over a period of 400 years, at an average rate of 5 cm/yr [13], compared to the rate of rise of 1-2 mm/yr during the 20th century [14].

The observed rapid changes cannot be explained by conventional models, and thus critically important predictions about the future responses of polar ice sheets to climate change are certain to be wrong in potentially disastrous ways. The existing models are based on a shallow-ice approximation in which driving forces are balanced by friction at the glacier bed. The models incorporate basal sliding where the bed is lubricated or for deformation of a soft sediment layer. However, they do not fully capture the controlling dynamics of ice streams and outlet glaciers, and do not allow for rapid along-flow propagation of force perturbations, including those resulting from removal of peripheral ice shelves.

Recent studies have focused attention on dynamic behavior of the Greenland Ice Sheet and on rapid changes occurring that cannot be explained by existing ice-sheet models. Peripheral outlet glaciers are thinning rapidly at double the rates that can be explained by increased ablation, pointing to ice-dynamical effects [14] [16][15][2][1]. At the same time, the speed of many outlet glaciers has increased significantly, with Jakobshavn Isbræ speeding from 6.8 km/yr in 1995 to 12 km/yr in 2005 [1]. This acceleration is likely a result of enhanced basal lubrication [18] [2] or possibly a dynamical response to weakening and subsequent disintegration of buttressing tidal ice tongues [15] [37].

In West Antarctica, some of the ice streams draining into the Ross and Ronne-Filchner ice shelves have been the subject of extensive field campaigns aimed at better understanding the controls on ice streams (19). These ice streams appear to be capable of rapid changes, including margin migration and complete shut-down. In the Antarctic Peninsula, several of the peripheral ice shelves have disintegrated or retreated, with a total area in excess of 14,000 square kilometers lost over the past two decades. Vaughan and others [41] have linked ice shelf collapse to southward migration of the -9^{O} C isotherm, presumed to correspond to the thermal limit of ice-shelf viability. While Vaughan [17] reported no significant acceleration of input glaciers following the breakup of the Wordie Ice Shelf, elsewhere in the Peninsula, ice-shelf breakup has led to flow acceleration of grounded glaciers (e.g. [39][33][1]). These observations suggest that the West Antarctic Ice Sheet may be on the verge of contributing to future sea level rise and have reinvigorated debate about the stability of the marine-based ice sheets.

While there is considerable evidence now that fast glacial changes occur and have occurred in the past, there is no consensus yet regarding how this takes place. Better understanding will only come from an ambitious campaign to collect more data and development of new and better models using such data. The National Science Foundation established the Center for Remote Sensing of Ice Sheets (CReSIS) in 2005 for just this purpose. Major advances in the quantity and quality of data about polar ice sheets is a major objective of the 2008 and 2009 field data campaigns, in which new measurement techniques will multiply the quantities of data available to the polar research community. The field program will consist of coordinated satellite, airborne and *in situ* measurements. Satellite data acquisitions are being planned as a part of Global Interagency IPY Polar Snapshot Year (GIIPSY) [36]. The field program includes plans to perform airborne measurements to collect data over an area of 200 km x 200 km with grid spacing of 5-10 km and a 20 km x 20 km area with grid spacing of 1 km or less, and surface-based observations including ice cores, seismics and fine-resolution radars [6].

2.2 *Data Analysis and Management Challenges*

The increase in measurement capabilities and data management needs is being driven by development of synthetic aperture radar (SAR) systems by CReSIS that provide new capabilities to image the ice-bed interface. In 2006, this system enabled CReSIS researchers to image the ice-bed interface through a 3-km thick ice sheet. This was the first time ever that an ice bed this deep had been imaged. Understanding the ice-bed interface through such deep ice sheets is critical to understanding rapid glacial change. Figure 2 shows a SAR image mosaic of the ice-bed interface at Summit, Greenland [38].

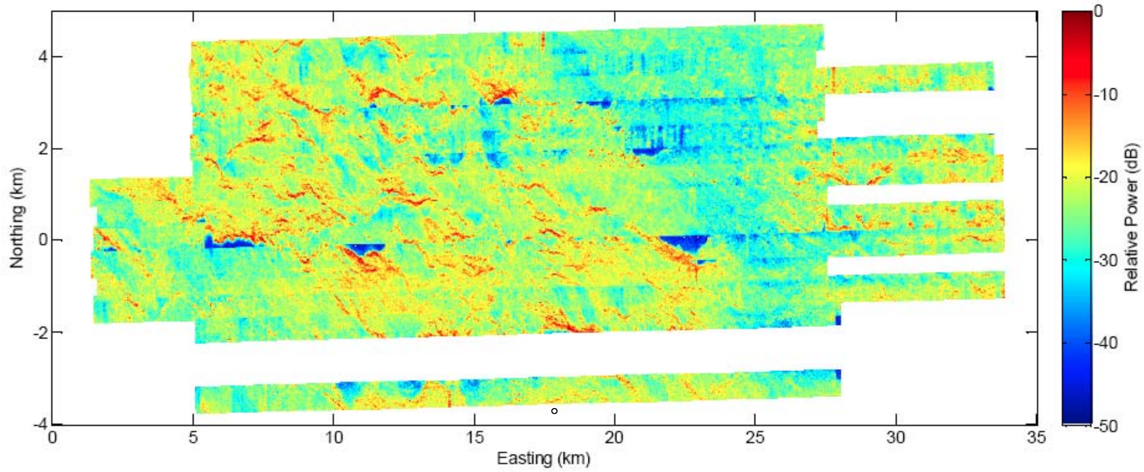
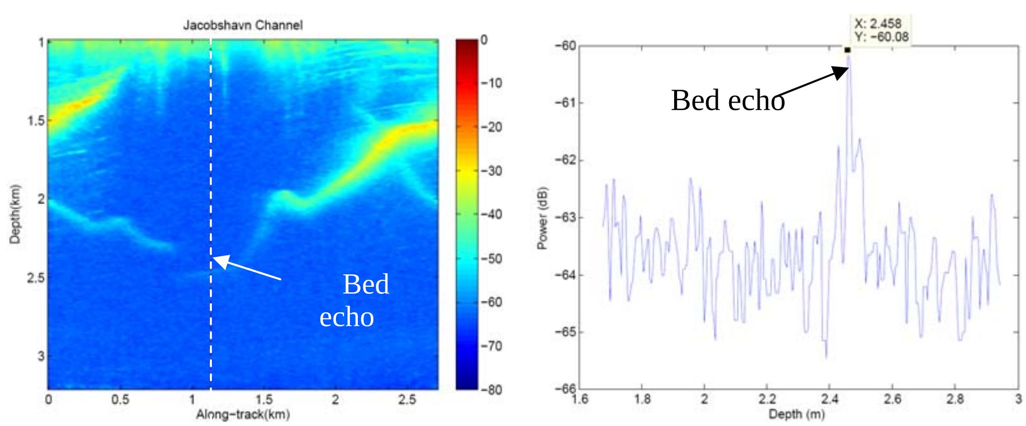


Figure 2. SAR mosaic of 120-200 MHz from the eight long data sequences. The origin is at Summit Camp, Greenland (72.5783°N and 38.4596° W). The polarization is VV and the resolution is 10-30m cross-track and 30m along track with 15 looks.

We have also used a similar radar that operates over a narrower bandwidth from a Twin Otter aircraft in 2006. We successfully obtained InSAR images of the ice bed and accomplished the first-ever radar sounding of the Jacobshavn channel [35]. Figure 3 shows sample results from airborne radar data processed in sounder mode across the Jacobshavn channel. We could obtain these results only because of SAR processing and digital beam forming to reject surface clutter that masks weak echoes from the bed.



3a 3b

Figure 3a Echogram across Jacobshavn channel (3b) A-scope along the white dotted line in 3a.

2.2.1 Computational and Data Challenges

SAR imaging of the ice-bed interface has the potential to make a significant contribution to glaciology by providing insights regarding the conditions at this critical boundary, including terrain relief and roughness as well as presence and distribution of liquid water. However, SAR image formation is computationally intensive. While the raw radar data, called phase histories, can be collected and stored for post-processing analysis, the image formation process requires a significant amount of signal processing, regardless of whether it is done in real-time or in post-processing. SAR imaging of the ice-bed interface through 3-km thick ice may require iterative focusing to compensate for the unknown variability in the electrical properties of the intervening ice. The image shown above does not include this.

To reject surface clutter, we need to generate seven complex SAR images from each antenna element to support digital beamforming for reducing surface clutter [38]. The clutter masks weak returns from the ice-bed interface, particularly over areas such as outlet glaciers and margins of the ice sheet undergoing rapid changes. The digital beamforming capability also allows us to image the ice-bed interface, map internal layers and measure ice thickness simultaneously.

Initial CReSIS SAR data collection flights may produce in excess of 6 million complex samples per second per receiver channel. A single 5-hour flight mission could collect as much as one terabyte of raw SAR data. Table 1 shows estimated data volume that CReSIS plans to collect during the 2008 and 2009 field seasons.

Table 1: Estimated data volume from field experiments in 2008 and 2009 field seasons in comparison with existing data sets.

Field Season	Data (TeraBytes)	Description of data sets
1993-2004	0.8	Airborne and surface-based radar sounders
Summer 2004	0.4	Airborne depth sounder, Surface-based radar sounder
Summer 2005	1	Airborne depth sounder, Surface-based radar at summit
Winter 2006	1.5	WAIS camp, surface-based radar
2006-2007	12	Airborne radar with SAR with digital beamforming and other instrumentation
2008 and 2009	20	Airborne and surface-based radars with SAR with digital beam-forming and other instrumentation

Effective management and analysis of 35 terabytes of data is a challenge under any circumstances. Management of large data sets when they originate from the polar regions is a major challenge.

2.3 Power Consumption and Network Bandwidth Challenges

The Polar Grid poses a number of interesting problems with respect to power and networking. Field observing networks in general tend to be in remote locations far from critical infrastructure and therefore must provide their own power for both the observing equipment (sensors, computers, and storage devices) and communications equipment. Power budgets can be reduced by using low power controllers, advanced power management policies, and minimizing communications power costs. However, the need to field more and more complex experimental packages and the desire to have continuously reported data (e.g. video and audio) increase bandwidth requirements and result in long or continuous communications times and greater power requirements. We also note that as well as remote science observations, energy exploration, emergency response and military operations have similar challenges. IU is exploring these outside this project.

The key characteristic of this project is the large amount of data that will be collected during each campaign: about 1TB per day. A primary problem is the difficulty of providing network connectivity between each node of the grid: field camps, base camp and the CICPS data center in the continental US as sketched in figure 1. With a distance of 100-300 Km between base and field camps and the remoteness of the base camp from the data center (3000-6000 Km) the only practical solution is satellite communications. Three options, Iridium (10Kbps max), Inmarsat (492Kbps max), and VSAT (2.5Mbps max) are available. Each has opportunities and costs, but none can provide sufficient bandwidth for transferring all data collected back to the CICPS data center. The solution is to bring sufficient computing capability to the field to perform necessary analyses on-site, and provide these results along with monitoring and status information over the available bandwidth. Primary data will be stored on-site and returned to the data center by air on removable media on a regular schedule. Continuous low-bandwidth connection to the continental US allows users can see what data has been acquired from the field and can potentially request that ad hoc analyses be run on data sets in the field to answer questions long before the primary data are available at the CICPS data center. Another significant problem is meeting the power requirements for storage and computing devices needed in the field for mobile data acquisition by UAV and tracked ground vehicle. These devices are expected to operate for eight hours independently of external power and will during that time collect 300-500GB per shift from SAR devices.

3.0 PolarGrid Instrument: Computer and Computational Science Resources

Polar science is an exemplar of the type of science that will benefit from an advanced Cyberinfrastructure. Cyberinfrastructure enables a dynamic Data-Information-Knowledge pipeline, accelerating new scientific discovery. This will result in better quality predictions rendered in shorter periods of time. Some of the hardware acquisitions will be quite novel, to meet the challenges of operating an IT infrastructure in the polar regions.

The impact of the "data deluge" has motivated substantial progress in the areas of data management, real-time sensors and data assimilation algorithms. Other important general Grid research areas of relevance to PolarGrid include scientific gateways [43] to manage the use of Grid-based services. A critical matter is the development of tools that enable polar scientists to make use of sophisticated supercomputers and Grid computing systems without first having to develop deep expertise in computational science themselves. The PolarGrid implementation and development will focus on the following key areas sketched in Figure 1.

- Field data collection systems to be taken with Polar Science researchers as they collect data
- A base camp 64-core cluster, allowing near real-time analysis of radar data by the polar field teams.
- A large 17 Teraflops cluster, to be integrated with the NSF-funded TeraGrid, to serve the polar research community in final processing of data and supporting simulations. This is split between Indiana and ECSU to support research and education/training respectively.
- The implementation of new improvements to the speed and capabilities of ice sheet models, and implementation of web portals to make the data, data products, and resources easily accessible to the polar research community.
- An educational videoconferencing Grid to support educational activities.

We propose two sets of field and base camp equipment, to simultaneously support operations in the Arctic and Antarctic. We are funding from IU's commitment an initial prototype system that will be used to confirm the design of the production expedition Grids. It will be used later in other applications described in Section 2.4 where we face similar power and bandwidth challenges

IU will lead the design and development of PolarGrid software based on their experience with the CIMA [20] [21], LEAD [22] and QuakeSim [28] projects, which cover key areas of remote instruments, sensors, Geographical Information Systems (GIS), simulation, and data assimilation. The entire new area of eScience, in which IU is a leader, aims to develop tools by collaboration between computer and discipline scientists so as to put into the hands of discipline scientists the tools needed to rapidly make new discoveries. We include a discussion of this software in the instrument section as the Cyberinfrastructure will only be effective if we link hardware and grid-technology based software.

3.1 Field and Base Station Descriptions

To equip **field stations** with appropriate data collection equipment, we propose 10 laptop computers and a portable server. Field operations require laptop computers which are easily portable and can function in the extreme polar environments. We propose to equip field researchers with a Panasonic Toughbook 30 laptop computer, designed to operate in temperatures as low as -20°C and tested according to military (MIL-STD-810F) specifications [49]. The laptops will be used to collect data from field sensors and perform initial analysis and data filtering. The laptops and all associated equipment will be bundled conveniently and securely within a self-contained, rugged field kit (a Pelican 1600) for researchers. A Trimble Recon GPS XC global positioning system will be connected to the laptops, to record the exact location of data collection. The data collected in polar regions are hard to obtain. Each laptop will have with it an external 100-GB USB2 drive, allowing data to be backed up in the field. Movement of data from laptops to base station equipment in an extreme environment is best done via 802.11 wireless communications. Each laptop will be outfitted with the appropriate (rugged) wireless communication equipment. Access to power is a critical issue; two small, foldable solar power panels (Brunton Solaris 26) will provide the capability of charging laptop batteries in the field.

Due to the limited storage capability of even a robust laptop computer, field stations will be equipped with a portable server capable of storing 1 terabyte (TB) of data (enclosed in a “polar-ready” case). The basic pattern of operation, then, will be that data will be downloaded from instruments to laptops in the field, and from laptops to a portable server in the field camp. The portable server will be transported from field camp to base camp, and data will be transferred from the portable server to the base camp computing system and analyzed there.

We propose to equip the **base camps** with identical or similar hardware to prevent field researchers and IT staff from having to manage multiple discrete systems while on an expedition. In the same way that two sets of field-station hardware are proposed, two sets of base-camp hardware are also necessary. While system portability is not maximized to the same extreme as the field-station hardware, the base-camp hardware also requires portability and a similar degree of ruggedness.

The foundation of each base camp is a 64-core data analysis cluster that will be used to refine, filter, and analyze data collected in the field. The data analysis cluster and the portable server for the field camps will be very similar, for ease of use by the field teams. The proposed base camp cluster takes advantage of new processor chip technology. In years past, there was just a single processor element within a CPU chip. Today, multiple processor cores are integrated within a single CPU. We propose CPUs that include four processor cores per CPU, and two CPUs per server within the cluster. We propose IBM x3550 servers as the basic server within the cluster. Using multi-core CPUs enables the creation of a cluster with what amounts to 64 processors, using just eight relatively small and light servers. This approach places a system capable of 0.5 Teraflops at the base camp, using less than 6 kW of power (well within the 100+ kW camp capacity). The cluster can store a total of 13 terabytes of data, enough for an entire field campaign. The primary data storage medium is external hard drives that can be easily shipped back to the continental US for processing. The cluster will be encased in a rack built to military specifications, and equipped with a UPS system to protect against possible variations in the power supply of the base camps. Details are in the budget justification.

3.2 Lower-48 Systems

An important component of the PolarGrid is to support on-demand data processing from the field. This requires that we establish communications between the CReSIS field camps and base camp, and from base camp to the primary computer cluster at IU. Currently, CReSIS performs on-site tests to ensure the quality of data collected. The basic filed data collection operational model is to select an area of potential interest based on available data and observations, to conduct field measurements and to process the collected data afterwards. Often, this post-processing identifies new and unsuspected features that then become targets for future surveys. For example, the sub-glacial trough underlying Jakobshavn Isbræ, West Greenland, extends much farther inland than previously believed. However, with the capability of real-time processing and data analysis, field campaigns can be adjusted or modified as dictated by the new observations in a matter of hours or days. The ability to manage data and perform near real-time analysis in the Arctic and Antarctic will transform capabilities and effectiveness of data collection in the field. The proposed grid-enabled processing cluster, to be located at IU and ECSU, will support final data processing and support ice-sheet modeling by the polar research community generally. This facility will consist of an IBM e1350 BladeCenter cluster to be installed at IU and ECSU. The cluster will have a processing capability of 17 Teraflops, with 0.9 terabytes of RAM memory and 32 terabytes of local disk storage. The IBM e1350 BladeCenter cluster will utilize Ethernet as its primary network interconnection. The proposed cluster will be based on the new, Intel 2.33 GHz quad-core Clovertown processor. The architecture of the proposed new system is based on the existing 20 Teraflops Big Red cluster at IU; the details of the architecture for this cluster are available online [50]. The proposed new cluster differs only in underlying processor technology and in the use of a single Ethernet interconnection. We will use the same operating system and management software as are currently deployed in the Big Red cluster.

The performance of the 3-level field-base-Lower 48 Grid increases by about an order of magnitude at each level, corresponding to increasing load as one moves from real-time to longer timescale analysis. The purchase of the system will be timed so that it provide the field and base grids for the 2008 and 2009 CReSIS expeditions while the Lower 48 system is back loaded with a 20%-40%-40% split across years 1, 2 and 3 to reflect the exponential growth in data analysis requirements illustrated in Table 1. The expedition grids use current experience to ensure there is enough power to analyze the data needed to monitor and guide the real-time experiments. The new CReSIS data will drive a new generation of high resolution simulations which will become increasingly important. As documented in several earth science fields [51] petascale machines will be needed for the largest simulations. PolarGrid will focus on interesting scale simulations developing new data assimilation approaches and providing the timely (from real-time in the field to a few days' delay in the Lower 48) analysis of the observational data. The link to the TeraGrid will allow us to support simulations that are too large for PolarGrid. We divide the Lower 48 systems between IU and ECSU, which have respectively 160 and 64 nodes of IBM Blade systems which we will populate with the latest multicore chips that we expect to be four-core at the start and 8- or 16-core for final purchases. There are several interesting synergies here with the same nodes supporting both research and education and a uniform architecture extending to grid components supporting the ECSU student laboratories. This will support our goal of training ECSU and other MSI students to be familiar with the very latest Cyberinfrastructure. We only propose 32 terabytes of disk on the Lower 48 system as IU has just installed its new 535-terabyte data capacitor. We are not aware of any MSI campus system in the Top 500 list [56] and our proposed system would today rank ECSU at about 150 in this list.

PolarGrid will depend on effective parallelization of key codes. As the CReSIS group gathers more data defining the ice-sheet physics, it will be possible to develop finer resolution simulations requiring substantial simulation resources and challenging parallelization issues for irregular geometries that match the non-uniform ice-sheet behavior. The IU team members have substantial experience in parallel computing [27][31]. Although parallelization of such codes is non-trivial, the IU collaborators have done extensive work in finite difference and finite element approaches in related fields (for example, earthquake, atmospheric, fluid dynamics and flow in porous media) that can be leveraged. We will also look at parallel algorithms for the antenna simulations and SAR data analysis needed by CReSIS. All such simulations, whether parallel or sequential, will be wrapped as services with the needed provenance (metadata) to define the conditions and data used in any simulation.

We propose to integrate the new major processing cluster supporting PolarGrid with the NSF-funded TeraGrid [46]. The TeraGrid is the NSF's flagship effort to create an advanced, national Cyberinfrastructure. By next year, the aggregate processing capability of the TeraGrid will be in excess of 500 Teraflops – the largest aggregation of (unclassified) supercomputers in the world. Integration of PolarGrid with the TeraGrid will lead to many benefits for the polar science community and the weather and climate communities generally. IU is well prepared to achieve this integration. IU participates in the TeraGrid in several ways. IU is a Resource Provider to the TeraGrid – meaning that computation, storage, and software resources at IU [57] are available to the national community of users authorized to use the TeraGrid. IU, particularly the Fox and Gannon groups, plays a key role in developing scientific gateways that enable discipline scientists to access the TeraGrid.

The purpose of integration of the proposed main processing cluster with the TeraGrid is to enhance the capabilities of the PolarGrid. We will leverage the authentication and accounting infrastructure of the TeraGrid, as well as existing TeraGrid projects, creating scientific gateways to provide intuitive web access to the system for polar science researchers. Most importantly, since the TeraGrid is the NSF's primary effort to create a national Cyberinfrastructure, integration of the PolarGrid facilities is the best way to make PolarGrid accessible to the broader polar and global climate research communities. Furthermore, while we have attempted to scale the proposed main computational facility to meet the specific modeling needs of the glacier modeling community, the data collected by CReSIS, and the data products created by CReSIS, will be used by a much larger and broader weather and climate modeling community. By integrating the proposed new facility with the TeraGrid, it will be possible to use the data and data products created by CReSIS and PolarGrid anywhere within the TeraGrid. This capability will maximize the utility of the data collected by CReSIS to the US research community.

IU will also enable integration of PolarGrid with the DOE- and NSF-funded Open Science Grid [45]. All development and deployment activities will be implemented on the basis of best practices available today. The Open Grid Forum (OGF) [26] is the major community activity in Cyberinfrastructure, aiming at both defining best practices and identifying requirements for new applications like PolarGrid. IU leads the OGF eScience activities (Fox), while Gannon and Plale are members of the OGF steering group.

TeraGrid and the Open Science Grid typify "resource grids" that can be used for many different application domains. In this proposal we are funding the needed dedicated resources for PolarGrid, while other existing and new projects will develop the collection of services and portals (Science Gateways [46][47]) that form the PolarGrid Cyberinfrastructure.

Table 2 PolarGrid hardware summary.

Institution	**Hardware**
Indiana University	160 e1350 Blades each with two 4-core 2.33 GHz Intel processors with 4GB memory, 24 TB Disk
ECSU Grid and Laboratory Servers	64 e1350 Blades each with two 4-core 2.33 GHz Intel processors with 4GB memory 8 TB Disk
2 Base Camp Grids	10 Panasonic Ruggedized Toughbook laptops (EACH) 8 IBM x3550 with two 4-core 1.86 Ghz Intel Xeon 5320 processors with 8GB memory (EACH)
2 Field Camp Grids	(a) 10 Panasonic Ruggedized Toughbook laptops (EACH) (b) 1 IBM x3650 with two 4-core 1.86 Ghz Intel Xeon 5320 processors with 16GB memory (EACH)

3.3 Education and Training Cyberinfrastructure

We request the resources to support the education, training and outreach goals described in section 4. The major part of this part of Cyberinfrastructure is described in the above section as the 5 Teraflop ECSU Grid and Laboratory system. This is augmented by client PC's for student laboratories and scientific analysis software IDL, Matlab and ArcGIS. We also include video conferencing hardware to support distance education.

4.0 Impact

4.1 Education and Training

We have a powerful Education, Outreach and Training (EOT) component led by Elizabeth City State University (ECSU). The EOT goals are designed to foster the integration of polar science research and education through use of the virtual classroom environment and on-site training laboratories which are equipped with Cyberinfrastructure especially configured for the training environment. The EOT goal will be to make a wide range of CI training resources available including on-site training, on-line tutorials and courses, virtual presentations from workshops and seminars, software tools, and other resources for education, outreach and training.

ECSU will assume primary responsibility for the training and outreach activities associated with the PolarGrid Infrastructure project. "Educational settings, audiences, and goals are too important to be adequately addressed as afterthoughts or add-ons to CyberInfrastructure projects, and instead must be treated as high priorities integrated in a project's overall design" [54]. To that end the PolarGrid project includes support for training activities associated with the new partnerships across academia, government agencies, the private sector and polar scientists targeted as users of the PolarGrid. The targeted audience will first be the CReSIS community of scientists and the minority serving institutions associated with the current CI-TEAM project. The audience will later be extended to include the larger polar science community.

Crucial to the effective use of CI are trained personnel and interdisciplinary education [55]. The EOT objectives are structured to maximize the training and preparation of the targeted audiences to effectively use the PolarGrid as a research and education tool.

- Objective 1: ECSU will enhance its Master's in Applied Mathematics with a concentration in remote sensing.
- Objective 2: IU will pilot new computational science and Informatics application modules, while TeraGrid will ensure that its training materials are appropriate for broader participation.
- Objective 3: CReSIS scientists and students will develop understanding of how to use the PolarGrid capabilities to address ice-sheet and other polar science problems.
- Objective 4: The Association of Computer/Information Sciences and Engineering Departments at Minority Institutions (ADMI) [19] will be provided with access to curriculum, including CyberInfrastructure, remote sensing, modeling and polar-science-specific components which are outside of their current capabilities.
- Objective 5: Code writers, users and experts on high performance computing performance and the larger polar science community will be brought together through parallelization activities.

4.2 Proposed classes

To achieve these objectives, the project proposes extension of the virtual classroom capabilities and the on-site training laboratories to include CI to be used for training purposes. ECSU has allocated total usage of the E.V. Wilkins building to house the proposed training facility associated with PolarGrid. E.V. Wilkins is the former information technology building and as such is equipped with raised floors, connectivity and electrical capabilities. The facility will support a vigorous schedule of virtual and on-site training on Grid technology, PolarGrid data access, and remote sensing data usage. Each organization will contribute its expertise; ECSU provides remote sensing, Haskell GIS, Kansas and Maine aspects of ice-sheet science, the TeraGrid team covers core CI and IU Grid workflow, data systems and sensors. The figure below shows the plan to build upon the existing expertise to aggressively design and deliver content to the targeted audiences. We will meld Virtual classrooms, on-site training, workshops, academic year and summer CI research opportunities, the Cyberinfrastructure concept from TeraGrid and MSI-CIEC, online tutorials, parallelization activities, middle school outreach and of course an excellent project portal (web site).

The EOT management team will establish a regular schedule of seminars and summer programs and talks that will be made available via the virtual classroom. This will involve interactive broadcasting to all partners and archiving the content for future viewing. The CReSIS educational program already involves synergy between the different CReSIS partners, by re-using material between institutions while each retains its own curriculum. The PolarGrid EOT program will follow a similar model. A schedule of courses will be developed and advertised, allowing all partners to directly enroll students or to embed the content in their individual home institution course offerings. A proposed schedule of course offerings is shown in Table 3.

4.3 Outreach and Broader Participation

The education, training and outreach leadership team for PolarGrid will be composed of representatives of ECSU (Linda Hayden and Malcolm LeCompte), CReSIS (Gary K. Webber, Education Coordinator), San Diego Supercomputer Center (Diane A. Baxter, Director of Education), TeraGrid (Scott Lathrop, Director of Education, Outreach and Training), ADMI (Andrea Lawrence, ADMI President and Chairperson of the Department of Computer Science at Spelman College) and MSI-CIEC (Alex Ramirez, Executive Director for Information Technology Initiatives, Hispanic Association of Colleges and Universities [HACU]).

Table 3 PolarGrid courses.

Institution/ Organization	Courses (timeline)
ECSU MS in Math with Remote Sensing concentration	RS 501: Geophysical Remote Sensing (Fall) RS 502: Geographic Information Systems and Geophysical Signal Processing (Spring) RS 503: Digital Image Processing and Analysis (Fall) RS 504: General Analytic Methods of Remote Sensing (Spring) RS 505: Geophysical Modeling (Spring) RS 506: Microwave Remote Sensing Principles and Applications (Fall)
CReSIS courses	SAR Interferometry, InSAR, and Applications (2-yr Fall cycle) Advanced Glacier Dynamics (2-yr Fall cycle) Ice and Climate (every Fall) Seismic Imaging of and beneath Glaciers (every Fall) Holistic Ice-Sheet Modeling (2-yr Fall cycle) Principles of Microwave Remote Sensing (2-yr Fall cycle) Geophysical Signal Processing (2-yr Fall cycle) RF Circuit Design (2-yr Fall cycle) Geophysics of Glaciers (2-yr Spring cycle)
TeraGrid courses	Access Grid Tutorials (-each semester) Globus Grid Security Infrastructure (-each semester) TeraGrid Overview (-each semester) TeraGrid Polar Science Gateways (-each semester) Introduction to The Open Science Grid (-each semester) Web-based Tutorials (-each semester)

We will exploit strong existing outreach activities at CReSIS and IU to involve Minority Serving Institutions (MSI) directly in our work. A strong education and training activity led by the Historically Black University Elizabeth City State will involve undergraduates through new curricula and research experiences. The ADMI association of MSI Computer Science and Engineering departments and the NSF-funded MSI Cyberinfrastructure Empowerment Coalition will ensure our outreach has broad impact.

Both the proposed virtual classroom enhancements and on-site laboratories within the educational Grid will support existing CI-TEAM program outreach activities involving ADMI institutions. The CI-TEAM project funded recently by NSF will provide each of the ADMI institutions involved with videoconferencing capabilities for use by the faculty and student teams. These teams of faculty and students are committed to not only full participation in the virtual classroom events but also to initiating a Cyberinfrastructure or polar science related research project on their individual campuses during the academic year. Support for the ADMI institution teams is included in the CI-TEAM budget. Also included in the CI-TEAM budget are funds for an extended summer research training program in Cyberinfrastructure for undergraduate students from the ADMI institutions. The summer program will be organized in a typical Research Experience for Undergraduates (REU) fashion on the campus of ECSU during the summers of 2007 and 2008.

As a result of its affiliation with CReSIS and its program in Terrestrial Remote Sensing, ECSU and its Center of Excellence in Remote Sensing Education and Research (CERSER) are engaged in a number of polar science research projects. As described below, each will benefit from the capability of a grid-computing network located at ECSU. Each of these projects, as well as future programs, will be aimed at achieving an increased understanding of the Antarctic environment, and also will have broader implications for improved understanding of physical processes occurring in both of Earth's polar regions.

ECSU, in collaboration with scientists at The Ohio State University and NASA Goddard Space Flight Center's Cryospheric Sciences Branch, is working to determine average air temperature trends at the surface of the Antarctic Ice Sheet. Satellite data (AVHRR thermal infrared Image and SSM/I brightness temperature data) accumulated over approximately 30 years of observations will be examined. The satellite data are spatially and temporally correlated with *in-situ* surface temperature data recorded by numerous Automatic Weather Stations (AWS) distributed across the southern continent. The AWS data provide ground truth for accurately interpreting the temperatures deduced from satellite remote sensing observations. Accessing and processing large volumes of high-resolution AVHRR, LANDSAT Thematic Mapper, SSM/I and ICESAT GLAS data acquired over a large continental land mass will be aided by a local grid-computing network. Grid-enabled access to analysis tools and data products will facilitate the widest possible use and exploitation of these important results beyond the faculty and students at ECSU.

Beyond its implications for ECSU research activities, the creating of a grid computing facility at ECSU would have far-reaching implications for its Computer Science (CS) academic program curriculum. This project will facilitate a transformation of the current ECSU CS curriculum by supporting principles of parallel computing and distributed computer architectures throughout the curriculum, starting at the most basic level.

In addition to modifying the content of its standard curriculum to include distributed computing topics, the ECSU computer science department will also adopt new courses devoted to more focused study in this field. A course in "Cluster and Grid Computing" will be introduced into the CS curriculum for students concentrating in scientific computing. The content of this course will include communication, application, security, and management technologies that enable grid computing. Students will write programs for grid computing and be introduced to current grid computing tools. They will also test programs on a computer grid comprised of local and remote clusters. Programming will be performed in Java and C++ languages.

A course in distributed scientific computing will be developed as an integral component of the ECSU graduate program in Applied Mathematics with a concentration in Remote Sensing. The course will be designed for graduate students in this program, who are likely to be challenged by the need to solve computationally intensive problems on parallel and distributed machines. Recent advances in parallel algorithms for scientific computing will be surveyed and their impact on improving performance of scientific applications will be studied. Students in this program typically come from a variety of academic disciplines (including for example, the geosciences, mathematics and computer science), and course content will emphasize a range of topics including algorithm design, numerical methods, computer architecture, software, and run-time environments. Students will engage in exercises using parallel algorithms for scientific problems on parallel and distributed machines.

All of the required courses in the remote sensing concentration will reap significant benefits from the grid computing techniques and methods learned in the new parallel scientific computing course described above. Opportunities to reinforce computing skills with actual application exercises drawn from real research using real data and grid-based algorithms will be especially beneficial in certain of the courses now offered in the curriculum, including Geographical Information Systems and Geophysical Signal Processing (RS-502), Geophysical Modeling (RS-505) and Digital Image Processing (RS-503).

Videoconference facilities and the on-site laboratories with the educational Grid acquired under this grant will play a major role in achieving the PolarGrid EOT objectives related to both the CReSIS minority and non-minority partners and to the greater community of polar scientists.

4.4 Broader Impact

The proposed PolarGrid configuration represents the first attempt by the Grid and glaciological communities to establish a comprehensive data analysis and distribution center and to make data available to the scientific community in a timely manner. While other data centers exist, such as the National Snow and Ice Data Center (NSIDC) and the World Data Center (WDC), these are not well-equipped to handle the large quantities of data now becoming available. Further, the proposed Cyberinfrastructure will allow data from various remote-sensing platforms and other sources to be linked and visualized in a logical and consistent manner. This will provide an invaluable service to the international polar science community.

The education, training and outreach activities in Section 4 illustrate our major commitment to bring science to undergraduates and to Minority Serving Institutions. As well as the significant CReSIS and ECSU outreach, we have the support of MSI-CIEC documented in the support letter from PI Richard Alo. MSI-CIEC allows us to tap into an alliance of over 335 MSIs and so broaden our outreach, as in our plan to use education and training to familiarize students from first ECSU and other ADMI and MSI-CIEC schools with the proposed grid. It is worth noting that the collaboration between CReSIS and Indiana stemmed from activities of the forerunner MSI Cyberinfrastructure Institute [23] to MSI-CIEC of which Fox (PI of this proposal) was PI.

5.0 Management and Project Schedule

5.1 Summary of Project Goals and Timeline

PolarGrid's ambitious proposed work will build upon previous and currently funded work. The NSF and other agencies have substantially funded LEAD, CIMA, OGCE, and the QuakeSim project, as described elsewhere, to implement Grid middleware for building science gateways to the TeraGrid. These gateways support both real-time and archived data from both instruments and applications, and integrate data and applications using workflow and related technologies. We will substantially leverage these tools to build the initial PolarGrid TeraGrid Science Gateway. This will be funded through an NSF CI-TEAM project led by ECSU, along with internal IU funding.

We divide the current proposed work into four tracks, with yearly milestones indicated below.

Hardware and Software Acquisition (Track 1): We will acquire the hardware listed in Table 2 over three years. The Lower 48 equipment will be split, with about 70% at IU and 30% at ECSU. In order to maximize the power of the multicore technology we require, the acquisition will be split 20%/40%/40% over Year 1/Year2/Year 3. In addition to hardware, ECSU will acquire Geographical Information Systems software (ArcGIS) and scientific visualization/analysis software (IDL, Matlab). These are described in detail in the proposal budget justification. IU (Stewart) and ECSU (Hadyen) will lead.

EOT (Track 2): These activities are summarized in Section 4 and Table 3. ECSU (Hayden) will lead this effort.

Scientific Software Development and Parallelization (Track 3): As described in the text, improved SAR analysis and glacier modeling are crucially important to this project. In addition, we have long-term goals for improving sensor design. The core algorithm development is a CReSIS activity and not part of this proposal, but we will address problems of parallelization as part of PolarGrid. IU (Fox) and KU (Gogineni) will lead this track. Deliverables are as follows. *Year 1:* Make current SAR and FEM codes into Grid services. *Year 2:* Investigate parallelization of analysis codes using field equipment; exploit multicore for this. Investigate parallelization of glacier modeling codes. *Year 3:* Integrate parallelized codes into production systems (Track 1). Investigate parallelization of sensor design codes.

PolarGrid System Design and Deployment (Track 4): We will build upon our team's extensive experience building Grids and Science Gateways. We have a substantial amount of software in hand for collecting data from instruments, managing real-time and archival data and metadata, managing scientific application, coupling data and codes through workflows, and building science portals to serve as user interfaces. The interesting work here will be to extend this software to support the real-time, in-the-field requirements of the field and base camps (Section 3.1), their integration with the on-demand Lower-48 systems at IU and ECSU (Section 3.2), and finally, their integration with the TeraGrid. IU will internally fund the field and base camp prototype system by April 2008.

Our general conception of PolarGrid is depicted in Figure 1. Field stations and base camps will have some capacity for on-site data analysis, but some problems will be too large for these systems, and the data must be physically transferred back to the Lower-48 system for high-priority, on-demand processing. Post-campaign analysis will take place on the TeraGrid, and data will be persistently archived in appropriate, pre-existing "open access" digital libraries for community access. The integrated system resulting from this work will form the basis of the PolarGrid Operations Center.

Deliverables are as follows. *Year 1:* Using team-derived usage scenarios, design PolarGrid Operations Center gateway as a follow-on to the CI-TEAM funded PolarGrid TeraGrid Gateway. This will include access to data analysis and simulation codes developed by CReSIS, workflows, third-party GIS software, and data management systems, and will support the intermittent subsystem interconnections. Train ECSU students and CReSIS scientists on the use of the PolarGrid gateway. *Year 2:* Deploy the system for the field campaign. Train ADMI members on PolarGrid usage. Update to use improved CReSIS analysis and modeling codes (Track 3). *Year 3:* Deploy the "final" version of the operations system. Expand MSI users to include MSI- CEIC members. Update to use parallelized codes from Track 3.

5.2 Management Strategy

The novelty of PolarGrid's CI demands collaboration between computer and ice-sheet scientists for progress. Thus as discussed in Section 5, we have set up the Cyberinfrastructure Center for Polar Science (CICPS) as an overarching framework of our collaboration to deploy and support the research, education, training and outreach for Polar Science Cyberinfrastructure. The founding members of CICPS include several members of CReSIS: Kansas, Elizabeth City State, Ohio State, and Penn State. These partners provide world-class expertise in remote sensing, glacier modeling, and education. They are joined by IU, which provides world-class expertise in Cyberinfrastructure and high performance computing. Together, the CICPS team is qualified to address the unique challenges in data management, computational steering, and high performance modeling that are required to understand and predict glacier melt.

Our team members have established a track record for collaboration through CReSIS and more recently through the NSF CI-TEAM Project. To build on these previous collaborations and coordinate our team members' individual strengths to take on the challenges of this proposal, we will adopt the following strategy. An *executive council* will be composed of the principal investigators. Overall project management will be handled by Fox. CICPS research activities will be divided into four teams covering, respectively, Models and applications; Sensors (led by Gogineni); Computer science, CI, and high performance computing (led by Fox and Gannon); and Education, Outreach and Training (led by Hayden). We will hold 1-2 "all hands" meetings per year. To ensure our work remains in contact with the larger community, we will set up an *external advisory board* that mirrors the above structure. This group will include Kelvin Droegemeier (Earth Science and CI), Scott Lathrop (EOT and CI) and ADMI representative Andrea Lawrence (Education and Broadening Participation).

CICPS is founded with the vision that Cyberinfrastructure will have a profound impact on polar science. We will realize this vision through a set of projects that are of the highest quality in all dimensions: Polar Science, Computer Science, Cyberinfrastructure, and EOT. The first of these projects is an NSF CI-TEAM project (PI: Hayden, Co-PIs: Fox and Gogineni), "Cyberinfrastructure for Remote Sensing of Ice Sheets," which will establish a CReSIS Science Gateway for TeraGrid working with IU, MSI-CIEC (Minority-Serving Institution Cyberinfrastructure Empowerment Coalition) and TeraGrid.

As shown in support letters and the budget justification, there is substantial institutional commitment associated with this proposal. ECSU pledges a new building and two faculty lines; IU $218,602 towards the Cyberinfrastructure (including the prototype expedition grid and $129,208 to core Lower 48 system) and a full-time support person for 3 years valued at $300,000. We also note the collaboration letters from Kansas, Hydrospheric and Biospheric Sciences Laboratory at NASA Goddard, New Hampshire, Penn State and Ohio State stressing the leverage of CReSIS activities. Further, the education and outreach support is documented with letters from ADMI (Spelman College and Mississippi Valley State University), TeraGrid and MSI-CIEC.

6.0 Related Work

IU is a member of TeraGrid and its Information Technology unit is highly regarded both in networking, where it runs the national Internet2 operations center, and computing, where it recently installed Big Red, a 2048 processor Supercomputer that is ranked #31 in the Top 500 list [56] generated each year by experts at the International Supercomputer Conference and is currently the most powerful campus computer in the USA. Dr. Craig Stewart leads IU's research computing activity. This includes campus computing resources such as Big Red as well as IU's involvement in TeraGrid and Open Science Grid.

The Community Grids Laboratory has substantial experience in building application Grids, including QuakeSim [28], which supports sensors, repositories and simulations in earthquake forecasting. This has many characteristics similar to the Cyberinfrastructure needed in PolarGrid. Indiana also has important outreach activities, including the NSF CI-TEAM project MSI-CIEC (Minority Serving Institution Cyberinfrastructure Empowerment Coalition), of which Fox is co-PI. Other relevant activities that will be leveraged include the NSF-funded Open Grid Computing Environment (OGCE) portal consortium [43] [48] (involving the Texas Advanced Computing Center [TACC], the National Center for Supercomputing Applications [NCSA], the San Diego State University, and the University of Michigan).

The Common Instrument Middleware Architecture (CIMA) project [20][21] is developing middleware to address a number of problems associated with integration of instruments and sensors into computing and storage grids. CIMA middleware provides a standard, Web Services-based protocol for interacting with sensors and instruments, flexibility in the underlying network transport and the ability to work across heterogeneous and hierarchical, possibly intermittent networks involving multiple field buses. CIMA also aims at efficient and high throughput data transport including compression, and abstraction of instrument capabilities and functions to allow graceful evolution of instrument design and reuse of data acquisition and processing codes. CIMA could provide solutions for data acquisition and transport for the monitoring function, and as an abstraction for the adapter layer of the Expedition Grid.

IU's Linked Environments for Atmospheric Discovery (LEAD) [22] project is a multi-disciplinary effort involving 9 institutions and more than 100 scientists, students and technical staff. LEAD is creating an integrated, scalable framework in which meteorological analysis tools, forecast models, and data repositories can operate as dynamically adaptive, on-demand, grid-enabled systems that a) change configuration rapidly and automatically in response to weather; b) respond to decision-driven inputs from users; c) initiate other processes automatically; and d) steer remote observing technologies to optimize data collection for the problem at hand. Although mesoscale meteorology is the particular domain to which these concepts are being applied, the methodologies and infrastructures being developed are extensible to others, including polar science. LEAD is building on the same service-oriented Cyberinfrastructure that is being proposed for much of the work described here. LEAD is also one of the TeraGrid gateway projects.

7.0 Prior Results

Marlon Pierce, Dennis Gannon, Geoffrey Fox, Beth Plale (NSF ANI-0330613): "Middleware for Grid Portal Development," NSF NMI Program 9/1/03 - 8/31/06 $3M (total). Other collaborating institutions included University of Chicago, National Center for Supercomputing Applications, the University of Texas, the University of Michigan, and San Diego State University. Called the Open Grid Computing Environment collaboration, this project provides a standard software platform for Web portals for e-Science. This project's activities and contributions are described in more detail in the text. The team's key publication is [43].

Linda Hayden: "CI-TEAM Implementation Project: Cyberinfrastructure for Remote Sensing of Ice Sheets." 01/01/07 – 12/31/08 PI: L. Hayden NSF OCI-0636361. The goal for this project is to create a virtual classroom and a TeraGrid gateway for ice-sheet research and education. The major effort will be to utilize the research performed by the NSF Science and Technology Center for Remote Sensing of Ice Sheets (CReSIS), for educational purposes, through collaborations with the Minority-Serving Institutions Cyberinfrastructure Institute (MSI-C(I)2), and the Association of Computer and Information Science/Engineering Departments at Minority Institutions (ADMI).

M. A. McRobbie, C. A. Stewart, J.C. Huffman, and R. Bramley: "AVIDD – a system for Analysis and Visualization of Instrument-Driven Data flows" NSFGrant No. 0116050, 2001-2003.. $1,800,000). The Analysis and Visualization of Instrument-Driven Data flows (AVIDD) system includes distributed Linux clusters, storage systems, and visualization facilities. AVIDD was designed to meet the needs of data-intensive science applications. AVIDD is a key part of IU's overall cyberinfrastructure efforts and IU's participation in the TeraGrid. A full report on the results of the AVIDD project is available online at [58].

References

[1] Rignot, E., and P. Kanagaratnam (2006), Changes in velocity structure of the Greenland Ice Sheet. *Science 311*, 986-990.

[2] Rignot, E., D. Braaten, S.P. Gogineni, W.B. Krabill, and J.R. McConnell (2004), Rapid ice discharge from southeast Greenland glaciers. *Geophysical Research Letters 31*, L10401, doi 1029/2004GL019474.

[3] CNN Web Site: "Ancient ice shelf breaks free from Canadian Artic": http://www.cnn.com/2006/TECH/science/12/29/canada.arctic.ap/index.html.

[4] Intergovernmental Panel on Climate Change Web Site: http://www.ipcc.ch/.

[5] United Nations Environment Program (UNEP), "Potential impact of sea-level rise on Bangladesh-Climate Change" http://www.grida.no/climate/vital/33.htm.

[6] The Center for Remote Sensing of Ice Sheets (CReSIS) Web Site: http://www.cresis.ku.edu.

[7] International Polar Year (IPY) Web Site: http://www.ipy.org/

[8] Atkins, D. E., K. K. Droegemeier, S. I. Feldman, H. Garcia-Molina, M. L. Klein, D. G. Messerschmitt, P. Messina, J. P. Ostriker, and M. H. Wright (2003), "Revolutionizing Science and Engineering Through Cyberinfrastructure." Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Available from http://www.nsf.gov/cise/sci/reports/atkins.pdf.

[9] Bement, Arden L. (2007), Cyberinfrastructure: the Second Revolution, the Chronicle of Higher Education Volume 53, Issue 18, Page B5 January 5.

[10] Connor, Steve, "If we fail to act, we will end up with a different planet", *The Independent*, January 1, 2007. Available from http://news.independent.co.uk/environment/article2116874.ece.

[11] Hansen, Jim, "Global Warming: Is There Still Time to Avoid Disastrous Human -Made Climate Change? i.e. Have We Passed a ' Tipping Point '?" Discussion on 23 April 2006 National Academy of Sciences, Washington, DC Available from http://www.columbia.edu/~jeh1/nas_24april2006.pdf.

[12] Broecker, W.S. (1994), Massive iceberg discharges as triggers for global climate change. *Nature 372*, 421-424.

[13] Kienast, M., T.J.J. Hanebuth, C. Pelejero, and S. Steinke (2003), Synchroneity of meltwater pulse 1a and the bolling warming: New evidence from the South China Sea. *Geology 31*, 67-70.

[14] Church, J.A., and J.M. Gregory (2001), Changes in Sea Level. In: *Climate Change 2001: The scientific basis (eds. Houghton, J.T. and others).* Cambridge U.K.: Cambridge University Press, 639-693.

[15] Thomas, R.H., W. Abdalati, E. Frederick, W.B. Krabill, S. Manizade, and K. Steffen (2003), Investigation of surface melting and dynamic thinning on Jakobshavn Isbræ, Greeenland. *Journal of Glaciology 49*, 231-239.

[16] Van der Veen, C.J. (2001), Greenland ice sheet response to external forcing. *Journal of Geophysical Research 106*, 34,047-34,058.

[17] Vaughan, D.G. (1993), Implications of the break-up of Wordie Ice Shelf, Antarctica for sea level. *Antarctic Science 5*, 403-408.

[18] Zwally, K.J., W. Abdalati, T. Herring, K. Larson, J. Saba, and K. Steffen (2002), Surface melt-induced acceleration of Greenland ice-sheet flow. *Science 297*, 218-222.

[19] The Association of Computer/Information Sciences and Engineering Departments at Minority Institutions (ADMI): http://www.admi.us/

[20] Bramley, R., Chiu, K., Devadithya, T., Gupta, N., Hart, C., Huffman, J.C., Huffman, K.L., Ma, Y., McMullen, D.F. (2006), *Instrument Monitoring, Data Sharing and Archiving Using Common Instrument Middleware Architecture (CIMA)*. Journal of Chemical Information and Modeling, 46(3) p.1017-25, May-June.

[21] Devadithya, T., Chiu, K., Huffman, K., McMullen, D.F. (2005), *The Common Instrument Middleware Architecture: Overview of Goals and Implementation*. Proceedings of IEEE International Conference on e-Science and Grid Computing (e-Science 2005), December 5-8, Melbourne, Australia.

[22] Plale, Beth, Dennis Gannon, Jerry Brotzge, Kelvin Droegemeier, James F. Kurose, David McLaughlin, Robert Wilhelmson, Sara Graves, Mohan Ramamurthy, Richard D. Clark, Sepi Yalda, Daniel A. Reed, Everette Joseph, V. Chandrasekar (2006), *CASA and LEAD: Adaptive Cyberinfrastructure for Real-Time Multiscale Weather Forecasting*, IEEE Computer 39(11): 56-64. Available at: http://kkd.ou.edu/CASA-LEAD-IEEE.pdf.

[23] Minority-Serving Institution Cyberinfrastructure Empowerment Coalition http://www.educationgrid.org/ NSF CI Team project building on Minority-Serving Institutions (MSI) Cyberinfrastructure (CI) Institute

[24] Marlino, M. R., T. R. Sumner, and M. J. Wright (2004). *Geoscience Education and Cyberinfrastructure*. Report of a workshop sponsored by the National Science Foundation (NSF), April 19-20. Boulder, CO: Digital Library for Earth System Education (DLESE) Program Center; University Corporation for Atmospheric Research (UCAR), 43p. Available at: http://www.dlese.org/documents/reports/GeoEd-CI.html

[25] Second Workshop: Geoscience Education and Public Outreach Network (GEPON) report from The Geoscience Outreach Workshop held March 22 - 24, 2006 at UNAVCO in Boulder, Colorado http://www.gepon.org/workshop2.html

[26] Open Grid Forum: http://www.ogf.org

[27] Fox, G. C., Messina, P., Williams, R. (1994), *Parallel Computing Works*!, Morgan Kaufmann, San Mateo, CA.

[28] Galip Aydin, Ahmet Sayar, Harshawardhan Gadgil, Mehmet S. Aktas, Geoffrey Fox, Sunghoon Ko, Hasan Bulut,and Marlon E. Pierce (2006), *Building and Applying Geographical Information System Grids*.Special Issue on Geographical information Systems and Grids based on GGF15 workshop, Concurrency and Computation: Practice and Experience, to be published. Available from http://grids.ucs.indiana.edu/ptliupages/publications/GISGrids_Concurrency_submitted.pdf .

[29] *QuakeSim* homepage at JPL http://www.quakesim.org.

[30] Semantic Grid Community Portal http://www.semanticgrid.org

[31] *The Sourcebook of Parallel Computing* edited by Jack Dongarra, Ian Foster, Geoffrey Fox, William Gropp, Ken Kennedy, Linda Torczon, and Andy White, Morgan Kaufmann, November 2002.

[32] Alley, R.B., and R.A. Bindschadler (eds.) (2001), *The West Antarctic Ice Sheet: Behavior and environment.* Washington DC: American Geophysical Union, Antarctic Research Series vol. 77, 296 pp.

[33] De Angelis, H., and P. Skvarca (2003), Glacier surge after ice shelf collapse. *Science 299*, 1560-1562.

[34] Gogineni, S., D. Tammana, D. Braaten, C. Leuschen, T. Akins, J. Legarsky, P. Kanagaratnam, J. Stiles, C. Allen and K. Jezek (2001), “Coherent Radar Ice Thickness Measurements over the Greenland Ice Sheet,” *Journal of Geophysical Research (Climate and Physics of the Atmosphere)*, vol. 106, no. D24, pp. 33,761-33,772.

[35] Gogineni, S., J. Paden, T. Akins, C. Allen, P. Kanagaratnam, D. Braaten, and K. Jezek (2006), “Synthetic Aperture Radar Imaging of Ice-bed Interface,” AGU Fall Meeting, San Francisco, CA, December 11-14.

[36] Jezek, K., and M. Drinkwater (2006), "Global Interagency IPY Polar Snapshot Year (GIIPSY), Eos, Vol. 87, No. 50, 12 December..

[37] Joughin, I., W. Abdalati, and M. Fahnestock (2004), Large fluctuations in speed on Greenland’s Jakobshavn Isbræ glacier. *Nature 432*, 608-610.

[38] Paden, J. (2006), “Synthetic Aperture Radar for Imaging the Basal Conditions of the Polar Ice Sheets,” Ph.D. dissertation, The University of Kansas.

[39] Rott, H., W. Rack, P. Skvarca, and H. De Angelis (2002), Northern Larsen Ice Shelf, Antarctica: further retreat after collapse. *Annals of Glaciology 34*, 277-282.

[40] Thomas, R.H., and PARCA Investigators (2001), Program for Arctic Regional Climate Assessment (PARCA): Goals, key findings, and future directions. *Journal of Geophysical Resea5rch, 106*, 33,691-33,705.

[41] Vaughan, D.G., G. Marshall, W.M. Connolley, C. Parkinson, R. Mulvaney, D.A. Hodgson, J.C. King, C.J. Pudsey, J. Turner, and E. Wolff (2003), Recent rapid regional climate warming on the Antarctic Peninsula. *Climatic Change 60*, 243-274.

[42] The Open Science Grid Project Page: http://www.opensceinceGrid.org/

[43] Alameda, Jay, Marcus Christie, Geoffrey Fox, Joe Futrelle, Dennis Gannon, Mihael Hategan, Gopi Kandaswamy, Gregor von Laszewski, Mehmet A. Nacar, Marlon Pierce, Eric Roberts, Charles Severance, Mary Thomas, (2006), *The Open Grid Computing Environments collaboration: portlets and services for science gateways*. DOI: 10.1002/cpe.1078. Available from http://www3.interscience.wiley.com/cgi-bin/fulltext/113391287/PDFSTART.

[44] The TeraGrid Web Site: http://www.teragrid.org.

[45] The Open Science Grid Web Site: http://www.opensciencegrid.org/.

[46] The TeraGrid Sciences Gateways Program: http://www.teragrid.org/programs/sci_gateways/.

[47] The Grid Computing Environments 2006 Workshop: http://www.cogkit.org/GCE06/GCE06/GCE06%20-%20Home.html.

[48] The Open Grid Computing Environments Web Site: http://www.collab-ogce.org.

[49] Department of Defense MIL-STD-810F Test Method Standard: http://www.dtc.army.mil/navigator.

[50] Morjan, Peter, Rodgers, Greg, Aiken, Ross, Turner, George, Hancock, Dave, Feinswog, Laurie, “Distributed Image Management (DIM) for Cluster Administration.” Available from http://hdl.handle.net/2022/517.

[51] Cohen, Ronald E., “High Performance Computing Requirements for the Computational Solid Earth Sciences.” GEO PROSE Web Site. Available from http://www.geo-prose.com/computational_SES.html.

[52] Hayden, L., Gogineni, P., Fox, G., Cyberinfrastructure for Remote Sensing of Ice Sheets, TeraGrid 2007 conference proceedings, Madison, WI (under review).

[53] Hayden, L., Development of Cyberinfrastructure Partnerships Dedicated to Remote Sensing of Ice Sheets, IEEE- IGARSS Proceedings, Barcelona, Spain, July 2007 (under review).

[54] M. R. Marlino, T. R. Sumner, and M. J. Wright, (2004). Geoscience Education and Cyberinfrastructure. Report of a workshop sponsored by the National Science Foundation (NSF), April 19-20. Boulder, CO: Digital Library for Earth System Education (DLESE) Program Center; University Corporation for Atmospheric Research (UCAR), 43p. Available at: http://www.dlese.org/documents/reports/GeoEd-CI.html.

[55] Pfirman, S., and the AC-ERE, 2003, Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century, A report summarizing a 10-year outlook in environmental research and education for the National Science Foundation, 68 pp.

[56] Top500 Supercomputing Sites: http://www.top500.org/.

[57] TeraGrid Resource Providers: http://www.teragrid.org/userinfo/hardware/resources.php.

[58] McRobbie, Michael, et al., "Creation of the AVIDD Date Facility: A Distributed Facility for Managing, Analyzing and Visualizing Instrument-Driven-Data (AVIDD)." Indiana University: Bloomington IN (2003). Available from: https://scholarworks.iu.edu/dspace/bitstream/2022/477/1/Scholarworks+AVIDD+Final+Reportpdf